# Modality and Spacing

## Technical Report

**December 2024** *
**Primordial Machine Vision Systems, Inc.**
**Greg Kreider**

**Abstract**

Spacing, the difference between consecutive order statistics, reflects changes in the modality of the data. Local increases correspond to transitions between modes, while consistent spacing comes from within, with a value related to the distribution's scale. We look at four ways of using spacing to detect multi-modality. The first detects local peaks and flats after passing the spacing through a low-pass filter to reduce noise. We develop null distribution models of each feature to evaluate their significance, and also add a non-parametric bootstrap test. The second inverts a known test for increasing slope to find level sections that would identify a mode. The third performs three non-parametric tests on runs in the signed difference of the spacing taken over a larger interval, which is equivalent to running a rectangular filter over the spacing. The fourth combines change point detectors run on the spacing to find the transitions between any modes. We evaluate these approaches for their accuracy, consistency, and sensitivity. Some work even if the original data is discrete or taken with limited resolution, and they allow us to test not only uni- versus multi-modality but also to say where the modes or the transitions between them lie.

Spacing is the difference between consecutive points after sorting, which are the order statistics. The differences are smallest over the center of a distribution and increase the further one goes into a tail. Even if data is taken from more than one variate, the same trends hold and the combined spacing can reflect the underlying modes. A modality detector, then, would look for transitions between these behaviors, seen as a larger spacing between consistent regions. We will evaluate four different approaches, two using feature detectors for local peaks and flats and two that look for changes in general. Spacing usually has a high variability and some smoothing is needed to suppress noise before looking for features. The first approach will use a low-pass filter for smoothing, and we will develop two ways to test the significance of the peaks and flats, one with a model of the features' distribution and the other a non-parametric bootstrap test drawn from the sample. The combination of filter and model amounts to a kernel density test for a conservatively chosen null distribution. The second approach is to use the same feature detectors on the interval spacing, or difference between order statistics separated by a gap or lag, which is equivalent to applying to the spacing a rectangular or running mean filter whose size is the interval. We will use three tests on runs of increasing or decreasing interval spacings, by way of symbolic or Markov chain analysis or permutation of the sample within the feature. One advantage of either approach is that the filtering smooths any steps

*udpated April and September, 2025

in the spacing caused by discrete or limited precision values. The other two approaches do not search for specific features, but instead look for points where the behavior of the data changes. This problem has generated a rich variety of solutions for problems in statistical process control and outlier detection and trend analysis. The third approach combines any number of existing changepoint detectors to identify transitions between flat and peak or the boundaries of the features. The fourth approach inverts an existing test that identifies regions of non-zero slope, with the idea that the level sections that result will correspond to flats. Evaluating these four tests will show that the change between variates in the draw can be located at a given significance level, and that flats when they appear do correspond to individual modes. The tests will have trouble where the modes are not well-separated or are similar, or where one small variate lies in the tail of a larger.

Multi-modality detectors group into a few categories. Parametric detectors compare the data to a single, known distribution and look for deviations from the expected density, for example in higher moments [26] [10]. Another class fits the data to a mixture of known distributions, for example assuming the sample is a combination of normal variates and calculating the mean and standard deviation of each using a mixture model [6] [20], or combining a basis set of modified distributions [32]. Non-parametric detectors do a kernel density estimation, and then look for features (bumps) in the result. The critical bandwidth [44] and excess mass [16] tests are two such approaches. The key problem here is to determine the bandwidth of the estimating filter: too coarse and features will be damped or lost; too fine and the result will be noisy. The mode tree [31] and SiZer [5] tools help select the best bandwidth of the kernel estimator. Other non-parametric approaches use properties of the distributions, for example that the slope of the density increases below a mode and decreases above [49] [18].

So as not to obscure the development and evaluation of the four approaches in the main text, many side comments have been placed as detailed notes after the bibliography. The key results are: the feature models (6) and (8) after low-pass filtering of the spacing; the processing of the non-zero slope statistic (9) to find level sections [Detail 19]; runs tests (13) and (15) on the signed difference of the interval spacing; permutation and bootstrap tests for feature significance in (17), (18), and (19); and the fusion of changepoint detectors described in Section 5.

# 1 Spacing and Modality

Following the notation of Pyke [38], spacing is defined as $D_i = T_i - T_{i-1}$, where $T_j$ is the $j$'th order statistic and $i$ is the index of the upper point such that $2 \leq i \leq n$ with $n$ the number of data points. In general the spacing forms an upright U as the index varies, with the sides rising as the distance between order statistics increases the further one goes into the tails, and the bottom stable over many points. An asymmetric distribution will have unequal sides and a minimum shifted away from the center. A one-sided distribution will have one arm of the U. The density of the spacing $f_{D_i}$ for a distribution or cumulative density function $F(x)$ and density function $f(x)$ is

$$f_{D_i}(y) = \frac{n!}{(i-2)!(n-i)!} \int_{-\infty}^{\infty} \{F(x)\}^{i-2} \{1 - F(x+y)\}^{n-i} f(x) f(x+y) dx \tag{1}$$

The first and second moments of this density give the expected spacing and its variance. We have simple expressions of the expected spacing for only a few distributions. For a uniform variate over (0,1) it is [38]

$$E\Big\{D_{i,unif}\Big\} = \frac{1}{n+1} \tag{2}$$

For an exponential variate with rate $\lambda$ [38, (2.9)]

$$E\Big\{D_{i,exp}\Big\} = \frac{1}{\lambda(n-i+1)} \tag{3}$$

This has a minimum at $i = 2$ and a maximum at the other end, in the tail, where the expected spacing is $n-1$ times larger. A logistic variate with mean $\mu$ and standard deviation $\sigma$ has [24, (17)]

$$E\Big\{D_{i,logis}\Big\} = \frac{\sigma n}{(i-1)(n-i+1)} \tag{4}$$

This is smallest at $i = (n/2)+1$, with the largest expected spacing in either tail being $n^2/4(n-1)$ times larger. The spacing remains within a factor $\alpha$ of the minimum over a range of $(n\alpha+1)/(1+\alpha)$ and $1+n\sqrt{\alpha/(1+\alpha)}$ points, respectively. For example, with $n = 100$ an exponential variate has 10 points starting at $i = 2$ over which the spacing is within 10% of its minimum, and the maximum at $i = n$ is 99 times larger. A logistic variate is flat over 31 points, and its maximum spacing is 25.3 times the minimum. The curves indeed resemble a U, with a wide base and steep sides.

Equations for the variance of the spacing of these distributions are found in [25].

Spacing is generated simply by sorting the data and taking the difference between adjacent points. The sort step amounts to estimating the inverse distribution function with as quantile the fraction $i/n$, and the difference an approximation of its derivative. This is the basis of the quantile estimator [24].

We can look at the spacing for multi-modal distributions in two ways. First, a combination of two or more of these spacings will give rise to regions of consistent values near the modes where the spacing is close to minimum, separated by regions with larger values where the distributions tail off. Second, and the original idea behind this approach, is to consider a histogram of the data. Larger counts within fixed-width bins correspond to a greater density or smaller spacing, assuming points are evenly distributed within the bins. Bumps in the histogram become regions of smaller spacing within modes, and increased spacing matches the anti-modes or gaps or dips between bumps.

It is easy enough to write down density and distribution functions when combining more than one mode; the cumulative density function is the sum of the contributing CDF's weighted by the fraction of points from each. An analytic solution to the spacing's density or its moments is not within reach, however. Most distributions involve exponentials or powers of the argument, and these cannot be untangled when shifted in the combined function, either during the integration of (1) or when inverting the distribution function for the quantile estimator. [Detail 1] Numeric integration is of course possible, but does not provide us with any insight into how the features develop as parameters of the modes change. We could try modeling the combined spacing using quadratic polynomials based on (4). The fit would be good where one variate is dominant, for example in the center of the modes if they are well separated, but would break down where they interact. The location or height of the increased spacing between the two variates cannot be captured due to this breakdown. We will not be able to attack an analysis of the spacing directly, and will have to resort to simulations and other general approaches.

We can, however, make general, quantitative statements about the behavior of the spacing. Within a flat its value will depend on the distribution's scale, the $\lambda$ or $\sigma$ that remains in the minimum spacing. The length of the flat will depend on the size of the draw. A mode therefore becomes more difficult to detect as the draw size decreases or if there is less difference in variance to other draws. An increase in the spacing depends on the interaction of the tails, set by the growth there and how far into the tail the transition occurs. A peak becomes more difficult to detect if two variates are too close or

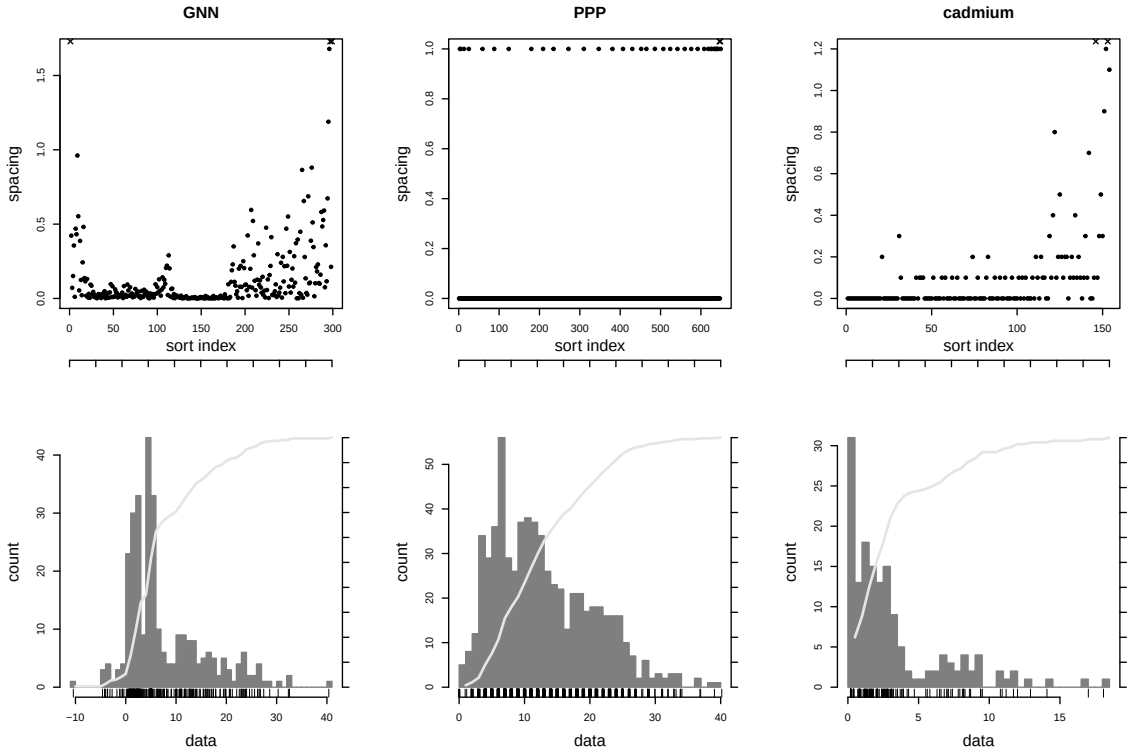



Figure 1: Spacing (top row) and histograms (bottom) for the demonstration data. The curve in the bottom plot is the cumulative density, with quantile axes to the right of the histogram and below the spacing.

have similar scales. Large samples also produce smaller spacings, assuming the range of the data is the same, and decrease the size of the peaks. Overall the spacing of a multi-modal distribution will resemble the U of a single variate, with a strong initial and final tail from the outer variates bounding local features and with modal effects appearing in the base. We cannot compensate for these edge effects — there is no consistent pattern — and they can hide small draws.

Figure 1 shows three data sets that we will use to demonstrate using spacing to detect multi-modality. The first row plots the spacing from one draw, and the second shows a histogram of the data. It includes a cumulative density curve to align values in the two graphs, since the $x$ axis is not the same. The extra axis underneath the spacing matches the right axis in the cumulative density. For example, in the leftmost graphs the middle index 150 falls halfway on the extra axis, which when followed across from the midpoint of the right axis intersects the cumulative density curve in the middle of the narrow, right peak, at a value near 5.

The left and middle graphs are examples of

$$\begin{array}{lll} 90 \times \mathrm{G}(2,1) & 60 \times \mathrm{N}(5,\, 0.1) & 150 \times \mathrm{N}(11,\, 9) \qquad \text{(GNN)} \\ 200 \times \mathrm{P}(6) & 250 \times \mathrm{P}(12) & 200 \times \mathrm{P}(22) \qquad \text{(PPP)} \end{array}$$

where $\mathrm{G}(r,\, \lambda)$ is the Gamma distribution with shape $r$ and rate $\lambda$, $\mathrm{N}(\mu,\, \sigma)$ the Normal distribution with mean $\mu$ and standard deviation $\sigma$, and $\mathrm{P}(\lambda)$ the Poisson distribution with scale $\lambda$. The first example comes from [14] and has two narrow, close peaks in the histogram, with the third draw

providing a large background, offset to the right. Each draw generates a region of consistent spacing, with a few points with larger spacing between the first two marking the gap between them. The $y$ axis is limited to show these features, and points that fall outside the graph are marked with x's at the top. The Poisson draws in the second example generate integers and the data contains repeated values. The spacing is 1 at the step when the value changes and 0 for the repetitions. Here we want to identify changes in the density or frequency of the steps. The loose pattern between indices 200 and 400 corresponds to a bump in the histogram between 7 and 17 and the large gap after index 100 to the change of the first two variates. The third example is the concentration of cadmium (in mg per kg of soil, or parts per million) in topsoil samples, provided in the "meuse" data of the R package *sp* [39]. The values were taken with a limited resolution, to one decimal place, which can be seen by the discrete values of the spacing. The data has many ties with zero spacing in the bulk of the histogram, which changes near index 100 to consistently larger spacings, as found in the second broad mode. The largest spacings that fall outside the graphs, at both the start and end of GNN or at the end of the one-sided PPP and cadmium data, are the edge effects from the tails of the total sample.

[Detail 2] describes the expected spacing for GNN and PPP, and its variation seen over many trial draws.

## 2 Features after Low-Pass Filtering

Although the spacing in the first example of Figure 1 shows differences between the modes, the other two are less clear. Low-pass filtering will bring out change in spacing. It also will smooth the variance in the spacing, which can be substantial, as seen to the right side of the GNN example. Because we treat the data as a collection of points, ignoring temporal or spatial ordering and instead using an index, we can work with discrete systems. Finite impulse response (FIR) filters are ideal for this case, as there are no uncompensated poles to affect stability. [17] has a comprehensive overview of the common filter kernels and their performance. We consider several possible filters: the Bartlett or triangular filter, cosine filters with Hanning or Hamming coefficients, Gaussian kernels, and Kaiser-Bessel windows. We reject the rectangular or running mean filter out of hand as its sidelobe reduction is too small, which means too many high frequency components will remain after filtering. For simplicity we also ignore cosine filters of more than degree 2, which includes the Blackman, Nuttall, and Blackman-Harris windows. Although these have good sidelobe suppression, they have wide main lobes. Gaussian FIR kernels, preferred by some for the nice property that the number of local maxima decreases as the filter's bandwidth increases [44], have infinite support and must be truncated. They trade off a faster sidelobe reduction for wider main lobe, minimizing the time-bandwidth product. Kaiser windows, designed following the procedure of [34, Section 7.4.3], maximize the equivalent noise bandwidth figure of merit, concentrating the kernel in the frequency domain.

[17] concludes that the Kaiser filter is best in general. [Detail 3] In practice there is not much difference in the good filters. The rectangular filter has too little smoothing, and the Blackman and higher degree cosine filters have too much, damping the filtered signal. [Detail 4] In terms of the responsiveness of the filter to generating features, the Kaiser seems most conservative, followed by the Bartlett, Hamming, Hanning, and Gaussian or Blackman.

These kernels are symmetric and time-invariant, so we can delay the input by half the width to make the filter causal. There is no answer about how to handle the situation where the filter extends beyond the data, and we choose to not run it over the border, dropping points that are partially covered. Other common options are to repeat the first and last points, or to mirror the data, but

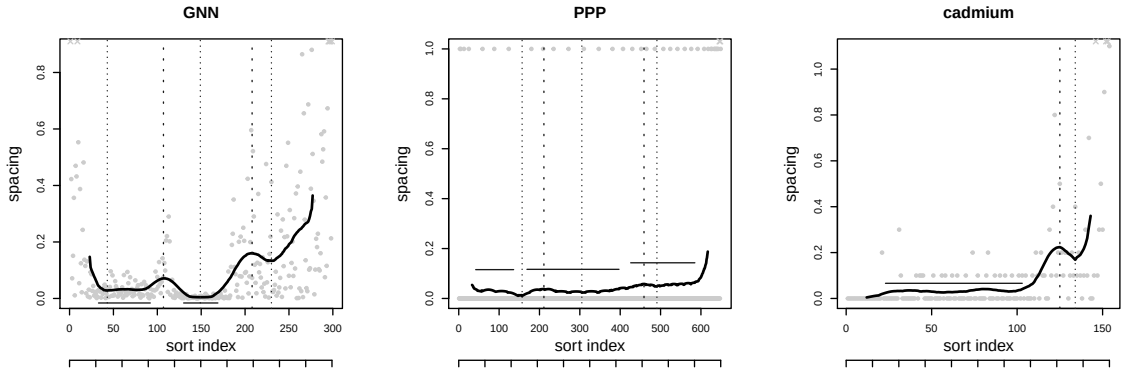



Figure 2: The low-pass response of the demonstration data, from a Kaiser filter whose kernel size is 15% of the data. Local maxima are marked with vertical dashed lines, minima with dotted. Horizontal bars indicate local flats.

neither captures the expected growth of the spacing in the tails. One could extrapolate the increase after making assumptions about the growth. Ignoring the borders does have the cost of not finding features at the edges, so that side draws can be lost.

Figure 2 shows the low-pass response for each of the demonstration data. The behavior we hope to see appears in the curves. In the first example the local maxima match the transitions between variates and there are flats for the first two draws. The third is too wide to produce a long, stable region and the filtered signal only shows the growth of the tail. The smoothed value within each flat reflects the different scales of the first two variates, smaller in the second. The low-pass filter smooths the gap between steps in the second example, with small local maxima at indices 200 and 450 marking the change between draws. It has three level sections, again at different heights set by the density of the step changes. The third example has a clear peak and consistent spacing to its left. The potential flat below index 20 is lost because it is only partially covered by the filter.

Choosing a low-pass filter involves two parameters: the kernel to use, and its size. We will typically express the size as a fraction $f_{lp}$ of the data size $n$, which for spacing will be one less than the raw size; the kernel width will be rounded to an integer, and we allow even sizes which center the filter on the half grid. The kernel size determines how much smoothing is done. Small kernels respond quickly to changes in the data but may not adequately remove variations. They emphasize peaks at the cost of flats but may not remove enough noise. Large kernels remove the local structure at the potential cost of making features indistinct. Before we look at how to choose the filter, we first need to define our local features and model their occurrence.

## 2.1 Local Extrema Detector

Local extrema are those that are smaller or larger than the neighboring data points on either side, with two additional requirements. First, two neighboring points that are within a factor of 0.001 of their average are treated as equal, and if these are locally extreme, the extrema is placed in the middle of the plateau of like values. Second, extrema are ignored if they are insignificant, because the height of the peak to either adjacent minimum is either below a global threshold, a fraction $f_h$ of all the data's range, or below a local threshold, a fraction $f_{h,rel}$ of the average of the two extrema. In other words, for any sequence of two extrema, be it minimum-maximum or maximum-minimum,

let their values be $l$ and $m$. Then ignore the first extremum either if $|l - m| \leq f_h(\max(x) - \min(x))$ or if $|l - m|/((m + l)/2) \leq f_{h,rel}$. The largest maximum and smallest minimum are never ignored. Potentially ignored extrema are removed one by one starting with the smallest, merging them into the larger neighbor. This can change the decision to keep the neighbor. The first and last data point are always marked as extrema. No conditions are placed on the locations of the extrema, as long as the height is large enough. They may even be adjacent points. The detector reports the position of each extremum, a flag whether it is a maximum or minimum, and the data value. [Detail 5] contains pseudo-code for the algorithm.

We will define peaks as local maxima with minima to either side, ignoring the first and last maxima at the initial and final tails. The left height $h_l$ is the difference between the value at the maximum and the prior minimum, the right height $h_r$ to the next minimum. For some tests we will define the peak's extension at some fraction of the height, similar to full-width-at-half-maximum (FWHM). This avoids a problem when a minimum falls in the middle of a flat, making the feature appear much wider than it really is.

## 2.2 Null Distribution of Peaks

To judge the significance of a peak, we compare its prominence to peaks found in unimodal draws from a base distribution. This forms a parametric test, but to mitigate the effect we will create a conservative model. Given the base distribution and a filter, we can vary the size of the draw $n$ and the kernel size $f_{lp}$. For each combination of these variables we characterize the features over a very large number of trials to determine their distribution. We define the peak by the larger height to either minima, standardized by the standard deviation $\sigma_{lp}$ of the low-pass spacing over the draw in order to remove scaling effects within the null distribution.

$$h_{peak} = \max(h_l, h_r)/\sigma_{lp} \tag{5}$$

Our model will predict the probability corresponding to the standardized height, or the critical value for a quantile. The model itself will be chosen to best fit the data, not for theoretical reasons.

Our first decision is to choose the base distribution. Thirteen variates group into four categories depending on the size and number of peaks they generate. [Detail 6] Two of these groups are outliers. Uniform and beta variates generate many more and larger peaks. Were we to pick either as the base distribution, we would rarely rate peaks in other variates as significant; because we are making no assumptions about the actual distribution of the data, a test based on these two would be much too conservative. An asymmetric version of the Weibull, with scale parameter $a = 2$ and shape parameter $b = 4$, has the next most peaks, represents the other variates well, and is a good basis for the model. The other draws will appear somewhat less significant, with their 0.975 or 0.99 quantile modeled at the Weibull 0.95 level.

The preferred low-pass filter, the Kaiser, produces fewer, smaller peaks than the alternatives, at a fixed kernel size. [Detail 7] Other filters will generate more peaks from the same data which, because the spacing of the draws should not have local maxima, are undesirable. The quantiles of the standardized peak heights are constant multiples $f_{filter}$ of the Kaiser's values at $q = 0.99$, making a simple correction possible. The Bartlett quantiles are a factor 1.086 larger, the Hamming 1.122, the Hanning 1.155, and the Blackman and Gauss 1.24.

Given the base Weibull distribution and Kaiser filter, we next try to fit the peak height to the draw size $n$ and fractional kernel window $f_{lp}$. The height is proportional to $\log n$ for a fixed quantile

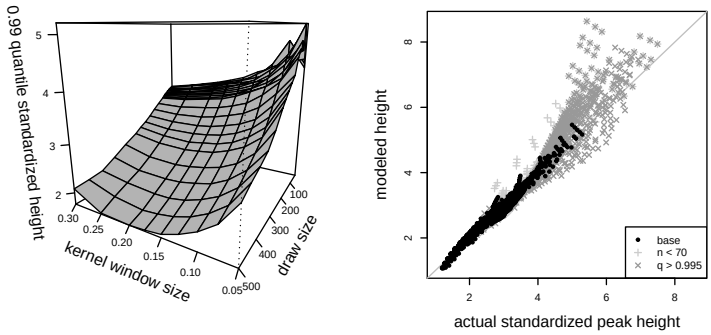



Figure 3: Actual 99th quantile of standardized peak height as draw size and kernel window change (left), and modeled $h_{peak}$ over all combinations of $n$, $f_{lp}$, and $q$ against actual.

$q$ and $w$, but there is no pattern to the fits when they vary. Instead, we model the distribution of the peak heights using the inverse Gaussian or Wald distribution. [Detail 8] Each of the Wald's parameters, the location $\mu$ and shape $\lambda$, is fit to functions involving $\log n$ and $f_{lp}$ or its logarithm. [Detail 9] The model over-estimates the height quantile at small $n$ or large heights, and a correction is needed. This takes the form of a linear scaling of the logarithm of the prediction, where the constant offset is a cubic function of $f_{lp}$ and the slope depends on $f_{lp}$ and its inverse. Letting $F_{Wald}$ be the Wald cumulative density function (21), the probability of seeing a peak of standardized height $h_{peak}$ is

$$p = F_{Wald}\left(10^{((h_{peak}/f_{filter})-b_{adj})/m_{adj}}, \mu, \lambda\right) \tag{6}$$

with

$$\begin{aligned} b_{adj} &= -0.2305 + 11.8716 f_{lp} - 46.9360 f_{lp}^2 + 85.0096 f_{lp}^3 \\ m_{adj} &= 4.2412 - 7.2054 f_{lp} + (0.1547/f_{lp}) \\ \mu &= (5.8158 + 2.4152 \log f_{lp}) - (1.9704 - 1.0131 \log f_{lp}) \log n \\ \lambda &= (-2.0204 + 49.7357 f_{lp}) + (2.6034 - 19.5195 f_{lp}) \log n \end{aligned}$$

The critical value at quantile $q$ is

$$h_{peak} = f_{filter}\left(b_{adj} + m_{adj} \log F_{Wald}^{-1}(1-q, \mu, \lambda)\right) \tag{7}$$

Clearly there is no theoretical basis to the model. It is chosen for its fit to the data.

This model, complicated as it is, matches actual heights for the features of interest (Figure 3). The surface is smooth except for very small windows and draws. The modeled $h_{peak}$ is less accurate for small heights, indicated by crosses in the right comparison plot, and large quantiles, marked by x's. [Detail 10] The latter are less important because the variation will still be judged significant at normal thresholds. We can see this in the confidence bands in the quantile comparison chart (Figure 4). The left graph in this figure is the equivalent of the height comparison, where again the small draws are highlighted with a different marker type. The right graph shows the fraction of variants within the indicated confidence bands. False positives, modeled values above and to the left of a threshold, occur in the next smaller simulated quantile at a 10% rate. False negatives, below and

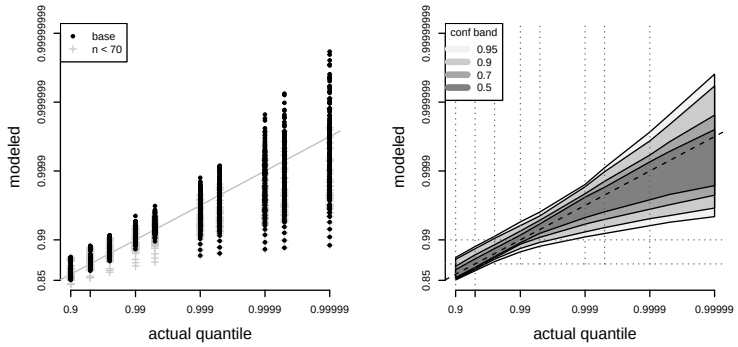



Figure 4: Predicted quantile of standardized heights against actual for all combinations of $n$, $f_{lp}$, and $q$ (left), and confidence bands.

to the right, grow as the threshold increases because the confidence bands widen. They are 2% at the 0.95 level, but 14% at the 0.99. The median modeling error, as a fraction of the actual quantile, is less than 5% except for small draws and the very largest. The interquartile range is within 10% for all window sizes, draws between 60 and 500 points, and quantiles up to 0.99995. Figure 26 of [Detail 10] breaks down the error as each model parameter varies.

The model, then, has limitations. It is inaccurate for small data sets with fewer than 60 points and for windows smaller than 5% of the data or larger than 30%. The analysis included windows up to 40% of the data, but the models here are too inaccurate to be useful. The wide confidence bands at $q = 0.95$ encourage using a tighter significance threshold for this test to avoid mis-classifying peaks as insignificant. A threshold of 0.975 or 0.99 is better. Basing the model on an asymmetric Weibull variate is a conservative choice, but it will fail on some data, notably if drawn from a uniform distribution, which is often used as a stress test when evaluating modality detectors.

## 2.3 Local Flat Detector

Flats are defined in terms of a ripple $\delta_{ripple}$ that is a fraction of the range of the data $x$. [Detail 11] Around each point the detector determines the interval between indices *st* and *end* where the data remains within the ripple bounds. It allows one outlier point outside the bounds to either side of the source point. It then takes the longest flat at each point, skipping flats shorter than a relative fraction $f_{L,rel}$ of the data or an absolute cut-off $L_{abs}$. The detector reports the start and endpoints of the flat, its height $h_{flat}/\sigma$ scaled by the signal's standard deviation and its length $l_{flat}$. By nature the height will be close to the ripple. [Detail 12] contains pseudo-code for the algorithm. The flats around each data point will overlap, and this approach returns the longest. The detector will also report the shorter if the uncovered portion meets the length requirement. The complexity of this algorithm is $O(n\bar{l}_{flat})$ using the average length. A flat can extend over a substantial fraction of the data, especially for very large data sets, and the complexity is then $O(n^2)$.

## 2.4 Null Distribution of Flats

Developing models of the flat height follows the procedure used for peaks. The process is more straightforward because the parameters can be fit directly in terms of the three setup parameters, the draw size $n$, the fractional kernel size $f_{lp}$, and the quantile $q$. The intermediate step to fit to the Wald distribution is not needed. The process is more complicated, though, because the lengths vary with the base distribution [Detail 13] and filter kernel. [Detail 14] For the peaks we could use scaling factors to account for the filter and distribution, but the variation for flats is larger and we will need several models. They will all have the same form, but still will require separate fits.

This is an exercise in fitting the data, rather than building models from theoretical considerations. The regressions were done choosing as few terms as possible, first by determining the general variation with each of the three setup parameters, and then allowing them to interact. The length depends on $q$, $q^2$, and a logistic $\log q/(1-q)$. It varies linearly with $f_{lp}$, and linearly and quadratically with $n$. The combined model is

$$l_{flat} \propto (c_{flp} + f_{lp}) \times (c_n + n + n^2) \times (c_q + q + q^2 + \log q/(1-q)) \tag{8}$$

where the $c$ are constants. Multiplying out, this model has 24 terms.

Some simplification is possible. Not all coefficients, typically half or even fewer, are significantly different from 0, per a $t$ test, and these can be dropped by backwards reduction with a small penalty in the overall performance. The length model can also be changed to use the absolute window size $nf_{lp}$ rather than the relative, again with a small penalty. However, these adjustments must be done individually for specific combinations of filter and base distribution. The framework still uses the full 24, with a substantial fraction of the coefficients set to zero. By default we will use the full models.

The accuracy of the models is better than for the peaks (Figure 5). The $q = 0.99$ surface shows the transition from a quadratic dependence on $n$ at the back to a linear at the front, while simplifying the small curvature with $f_{lp}$ to a line. The fit between the actual and predicted is good for all combinations of the setup parameters, as seen in the right graph. [Detail 15] The median modeled values are within 5% of the actual for all parameter combinations, and the inter-quartile ranges are larger for small draw sizes or the very largest kernel windows. These results are for draws from a logistic variate and use the Kaiser filter kernel. The confidence bands for the predicted quantiles, with the same layout as Figure 4, are shown in Figure 6, and are consistent across the range of quantiles measured. The mis-classification rates are smaller than for the peak model.

Although the model is good, it depends on which distribution and filter is used to generate critical values. The choice of filter is perhaps less important because the Kaiser model is conservative, but not dramatically so: a flat passing the Kaiser 0.95 level will correspond to the 0.99 level for other filters. The models are much more sensitive to the base distribution. There is little overlap in the critical values between base distributions, and we cannot pick one that is somewhat conservative but that will work in general. Flats are the inverse of peaks, and the distributions that generate many peaks will create fewer flats. Therefore there is not a single model for judging flats. We can ignore five distributions that generate unusually few or many flats: uniform, beta, exponential, F, and Wald variates. Three distributions span the range of critical values, the Weibull, logistic, and Gumbel, and we add the normal for the typical assumption made about the underlying data. By default we will use the logistic model, whose critical values are in the middle of these four options.

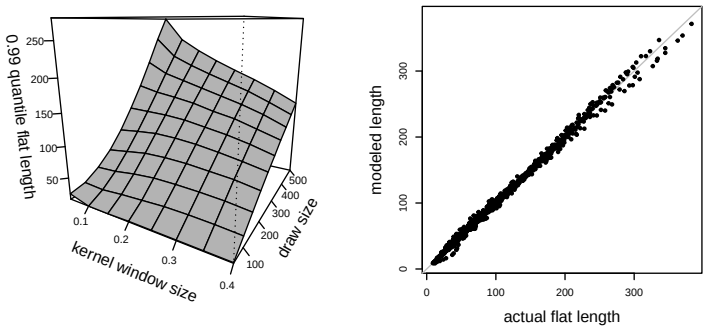



Figure 5: Actual 99th quantile of length as draw size and Kaiser kernel window change (left), and accuracy of model for $l_{flat}$ (right) over all combinations of $n$, $f_{lp}$, and $q$.

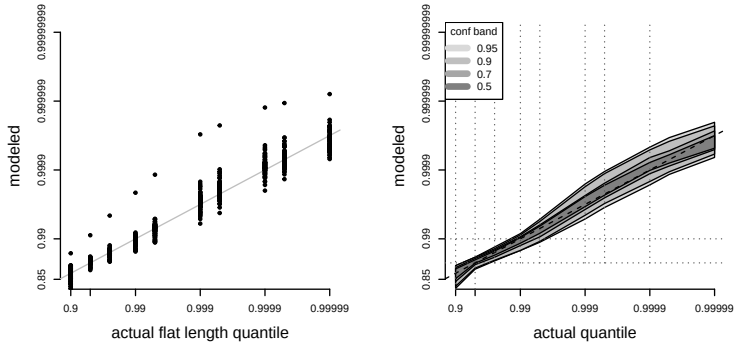



Figure 6: Predicted quantile of flat length against actual for all combinations of $n$, $f_{lp}$, and $q$ (left), and confidence bands (right). Outliers in the left graph are for the smallest $n$ (50) and $f_{lp}$ (0.05).

## 2.5 Filter Choice

Low-pass filtering of the spacing is equivalent to kernel density estimation. [Detail 16] The kernel width determines the bandwidth, and the problem of choosing the best bandwidth has been explored in depth, as it forms the basis of Silverman's test for multi-modality [44]. One common approach is to vary the bandwidth and watch the change in modes that appear, either their location or by the result of some test. The mode tree [31], for example, traces in a 2D plot the position of modes and anti-modes as the bandwidth increases, linking their locations. The links can be marked according to a test statistic that measures the area above the highest adjacent anti-mode [30], similar to the peak height above its bounding minima that appears in our model. This technique has also been applied to tracking maxima between wavelet filters of different resolutions [4].

The mode tree can be used directly for peaks, shading or coloring the local maxima according to their probability. Tracking the feature between different window sizes amounts to picking the nearest neighbor while preventing crossing a local minimum. This last condition tends to produce traces that dwindle out as the window grows, rather than seeing features merge, when a maximum and a

minimum disappear as a pair. Matching flats between window sizes is difficult because they may overlap and their ends may not be stable. They are simply displayed as a bar along their length, again shaded or colored to indicate their significance. [Detail 17] discusses these charts for the PPP example.

These tools say more about the stability of features than they are able to pick a specific window size. They provide bounds on the window size: too small generates many insignificant peaks; too large a few long, significant flats. In practice windows of 10% to 15% of the data seem to work well.

# 3 Level Section Test

Rather than tracking the position of extrema as the bandwidth changes, the SiZer approach identifies regions of increasing or decreasing values, in other words testing if the local slope differs from zero or is flat [5]. [14] instead tests for non-zero slope at many bandwidths at the same time. Their statistic, adapted to Pyke's notation, is

$$\begin{aligned} T_{jk} &= \sum_{i=j+1}^{k-1} \beta\left(\frac{T_i - T_j}{T_k - T_j}\right) \\ &= -(k-j) \sum_{i=j+1}^{k} \beta\left(\frac{i-j-1/2}{k-j}\right) \frac{D_i}{T_k - T_j} \end{aligned} \tag{9}$$

where $\beta(x) = 2x - 1$ is the Wilcoxon score function giving $T_{jk}$ zero mean and variance $(k-j-1)/3$. $j$ is the starting index and $k$ the ending of an interval in the data. If $T_{jk}$ is greater than the statistic for a uniform draw of the same size, then the data is increasing over the interval. If less, decreasing. These directions apply in the original data and will reverse for spacing. The test procedure adds a correction based on the fraction of the data covered by the interval, to suppress short intervals which will be more common. Critical values, with or without correction, are determined by repeated sampling of uniform draws to get the $T_{jk}$ distribution. Although the reference implementation of this test in the R *modehunt* package [40] does this sampling, one can also model the value as a function of the data size and significance level. [Detail 18]

This test can identify several intervals for any starting index, and adjacent indices can source overlapping intervals. [13] defines the minimal set of intervals as those that contain no other as a proper subset. Even this set contains many intervals that substantially overlap, and we can further reduce them by taking the outer endpoints of two intervals whose common segment is a large fraction of either length. We merge intervals that overlap by 75%.

The simplification of the intervals can be seen in Figure 7, run on the GNN example. It shows the test statistic in the left graph and projects the results onto the data in the right. The lower right of the first plot thresholds the test statistic at the $\alpha = 0.05$ level, with the starting index $j$ along rows and ending index $k$ along columns. It is a mirror of a heat map of the raw $T_{jk}$ in the upper left, with axes swapped. The minimal common subset is bounded by the leftmost significant point along each row and the topmost point in each column. These give the smallest end point and largest starting point, respectively. The reference implementation in the *modehunt* library searches first vertically to find the starting index, then left for the ending. These intervals are marked with dots; there are 26. The additional overlap condition then combines nearby points into a final set of 6, marked with large circled crosses and drawn atop the spacing as sloped line segments. Two segments partially but not completely cross the two flats, meeting at the peak between them. A third starts in the second flat

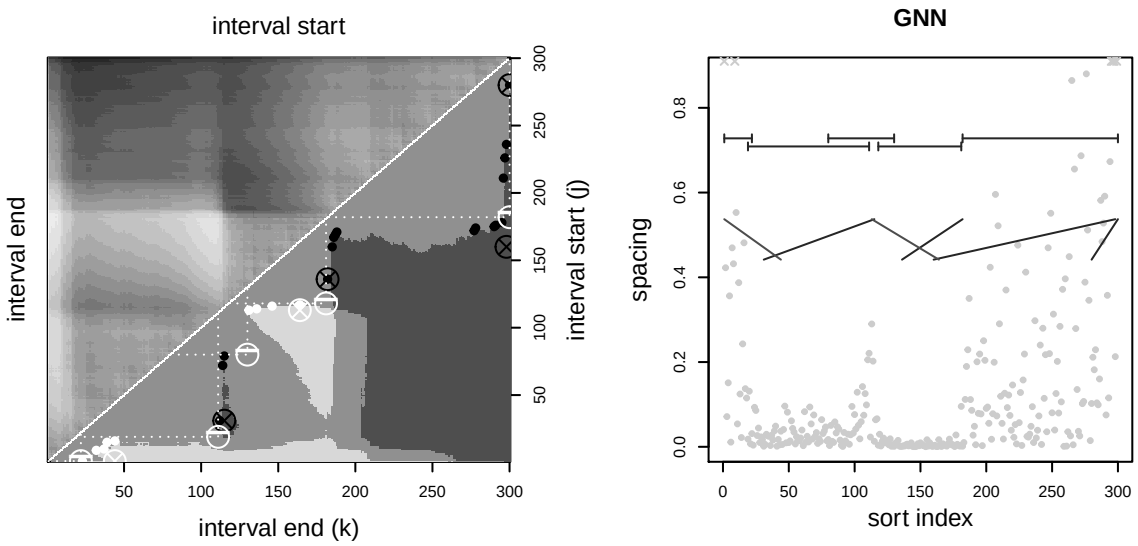



Figure 7: The left graph demonstrates the slope and level tests for the GNN example, with the raw $T_{jk}$ in the upper left and thresholded in the lower right. The right graph marks intervals of non-zero slope and level stretches.

and crosses the wide variate. Two more segments mark the tails, and a spurious result lies within the second flat, responding to its right end. The sloping segments are not localized to single features.

The thresholded image shows there is an additional result available, in the space that does not contain significant slopes. Call this inversion a level test. A level section is the longest interval without a significant result, again combining proper subsets. Graphically this amounts to picking the largest empty triangles on the diagonal, marked with stippled lines in the left plot. The lower right corner of the triangle gives the longest common interval, marked with a circled dash and drawn atop the spacing as flat segments. The algorithm starts the search for a new triangle in the first row above the sloped region that determined the end point of the level section, but may back off the starting index. [Detail 19] These sections may therefore overlap, as happens in the first anti-mode between indices 80 and 131. The results in Figure 7 show this identifies the three modes, plus the left-side initial tail. A larger $\alpha$ for the threshold critical value shrinks the neutral area and creates more, smaller sections. Both tests are sensitive to steps and outliers in the data and tend to be noisy, even with the additional overlap condition. The problem is worst with discrete data; each unique value in the PPP sample triggers either test.

# 4 Interval Spacing

Variants of spacing have been used in mode and slope tests. Venter [47] estimates the mode of data with the average of the order statistics placed before and after a point with the smallest difference, where the separation depends on the polynomial order of the peak. Proschan and Pyke [37] scale the exponential spacing by the denominator of (3) in order to test if data has a monotonically increasing failure rate. By removing the dependence on the index $i$ the normalized spacing should be a constant and a signs test over pairs will be normally distributed. [Detail 20] The test statistic (9) sums all

the differences in spacing to determine if that data has a positive slope.

Let us call the difference in order statistics over distances greater than one "interval spacing". Extending Pyke's notation used in Section 1, let $i$ be the upper index and $w$ the width of the interval, with $w \leq i \leq n$ and

$$D_{i,w} = T_i - T_{i-w} \tag{10}$$

Alternatives are a center point $j = i - (w/2)$ and radius $r_n = w/2$, as used in [47, (2.1)], or start index $j = i - w + 1$ and endpoint $k = i$ of the range as used in (9). The center representation matches the indexing used by the low-pass filters, where an even kernel size, centered on a half grid, shifts downward.

The density of the interval spacing

$$f_{D_{i,w}}(y) = S_1 \int_{-\infty}^{\infty} \{F(x)\}^{i-w-1} \{F(x+y) - F(x)\}^{w-1} \{1 - F(x+y)\}^{n-i} f(x) f(x+y) dx \tag{11}$$

$$S_1 = \frac{n!}{(i-w-1)!(w-1)!(n-i)!}$$

follows from the joint density of two order statistics [50, (8) and (31), with $r = i - w$ and $r' = i$]. Expanding [38, (2.4)] gives the same equation. [Detail 21] This and its moments can be solved for uniform, exponential, and logistic variates, but the results are neither simple nor can be separated for several variates. Multi-modal draws again cannot be analyzed.

The interval spacing is the sum of individual spacings over the width.

$$\begin{aligned} \sum_{j=0}^{w-1} D_{i-j} &= (T_i - T_{i-1}) + (T_{i-1} - T_{i-2}) + \ldots + (T_{i-w+2} - T_{i-w+1}) + (T_{i-w+1} - T_{i-w}) \\ &= T_i - T_{i-w} = D_{i,w} \end{aligned} \tag{12}$$

Such a sum is the same as a rectangular filter FIR kernel, without scaling by the width of the interval. Interval spacing is therefore a low-pass filtered version of the spacing. The wider main lobe of the rectangular filter compared to others permits the use of smaller intervals. In Figure 8 the interval width is 25 for the GNN example, 30 for the PPP, and 15 for the cadmium data, compared to filter widths of 45, 98, and 23 respectively in Figure 2. Because the kernel is not normalized by its size, it amplifies the signal and does poorly at smoothing any steps created if the data is taken with limited precision, or if it is made of discrete values. This is noticeable in the PPP graph. The small sidelobe suppression of the rectangular filter means that the interval spacing will be rougher, although conversely it may preserve some small high-frequency features. [Detail 22] We can use the same detectors to find local extrema and flats, perhaps with relaxed relative height or ripple parameters to accommodate the data's precision or the interval spacing's roughness. The features are marked as usual in the example, using a relative ripple specification $\delta_{ripple}$ for the flats of 0.03, 0.20, and 0.05 respectively, and peak height $f_h$ 0.20 for the PPP. In general features from both approaches will agree. We must account for the different indexing convention as the centered low-pass filter aligns better with feature positions than the endpoint of an interval does.

## 4.1 Tests for Runs and Features

Given a feature, peak or flat, found in any source, low-pass filter or interval spacing, we would like a test to evaluate its significance. Many existing test frameworks cannot be used to do this, however,

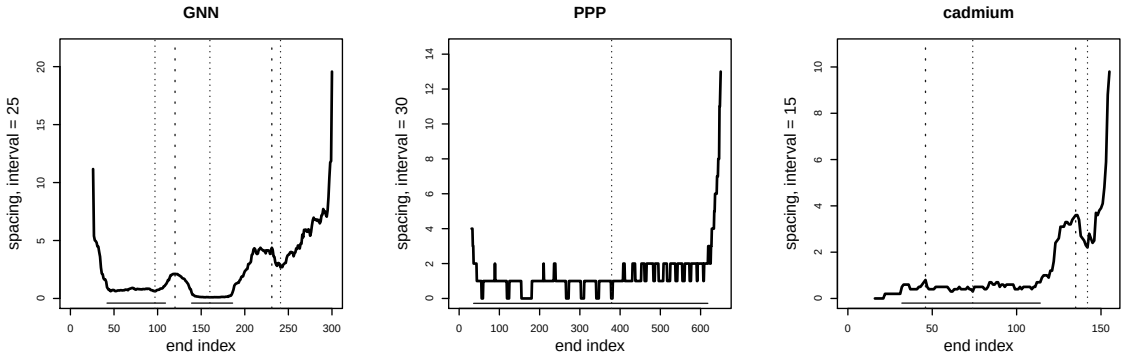



Figure 8: Interval spacing applied to examples. Extrema and flats marked as in Figure 2.

because we cannot assume a distribution. Notably, the spacing is not normal. Non-parametric tests offer a general solution. One type reduces or simplifies the data so that a distribution becomes valid, for example rank comparisons. Another approach heavily quantizes the data, for example using the sign of the point to point difference to mark local increases, decreases, or ties, and combines these into patterns or runs in the symbols. We will use two tests, one known and one new. A third type of non-parametric test takes the single sample as representative of the underlying distribution and scrambles it to see how often similar features arise. We will adapt these known permutation and bootstrapping techniques to the problem.

Runs are repeated sequences of the same symbol after simplifying the data and have long been used to check for trends. They are characterized by their length, which is related to the number of runs because the sample has a fixed size. Kaplanksy and Riordan extended the combinatorial analysis of Wald and Wolfowitz [48] to any set of symbols. The number of runs over $s$ symbols with population $a_i$ in each is distributed normally with expected value and variance [22, (12) and (13)]

$$\begin{aligned} E\{U\} &= 1 + \frac{2\alpha_2}{\alpha_1} \\ V\{U\} &= \frac{2\alpha_2(2\alpha_2 - \alpha_1) - 6\alpha_1\alpha_3}{\alpha_1^2(\alpha_1 - 1)} \end{aligned} \tag{13}$$

where

$$\begin{aligned} \alpha_1 &= a_i + aj + a_k && = a_1 + a_2 + a_3 \\ \alpha_2 &= \sum_{i=1}^{s-1} \sum_{j=i+1}^{s} a_i aj && = a_1a_2 + a_1a_3 + a_2a_3 \\ \alpha_3 &= \sum_{i=1}^{s-2} \sum_{j=i+1}^{s-1} \sum_{k=j+1}^{s} a_i aj a_k && = a_1a_2a_3 \end{aligned}$$

in general and specifically for three symbols representing the sign of the difference, allowing ties. [42] gives another general counting method.

Such runs tests are based on combinatorial counting of shuffles of the data and depend on individual points being independent. For example, the length of a run with two symbols would be a binomial variate. Our data, however, is not independent: the interval spacing or low-pass filtering introduce

correlations. Our second, new test uses a Markov chain to capture these interdependencies. Let $\mathbf{T}$ be the transition matrix, estimated normally as the frequency of symbol pairs in the data [3]. If the combinations create a chain with order greater than one, then the symbol set can be expanded to capture the longer sequences. [Detail 23] We build a new transition matrix $\mathbf{R}$ modeling the runs by splitting $\mathbf{T}$ into its diagonal $\mathbf{A}$ and off-diagonal $\mathbf{B}$; $\mathbf{T} = \mathbf{A} + \mathbf{B}$. $\mathbf{A}$ advances the run length while $\mathbf{B}$ resets it for the next different symbol. To model a run of length $L$

$$\mathbf{R} = \begin{array}{c} \begin{matrix} \text{next} \\ \text{length} \end{matrix} \\ \begin{matrix} \\ \\ \\ \\ \\ \\ \end{matrix} \end{array} \begin{array}{c} \begin{matrix} 1 & 2 & 3 & 4 & & \text{L} & \text{L+1} \end{matrix} \\ \begin{bmatrix} \mathbf{B} & \mathbf{A} & 0 & 0 & & 0 & 0 \\ \mathbf{B} & 0 & \mathbf{A} & 0 & & 0 & 0 \\ \mathbf{B} & 0 & 0 & \mathbf{A} & & 0 & 0 \\ & & & & \ddots & & \\ \mathbf{B} & 0 & 0 & 0 & & \mathbf{A} & 0 \\ \mathbf{B} & 0 & 0 & 0 & & 0 & \mathbf{I} \end{bmatrix} \end{array} \tag{14}$$

The identity matrix $\mathbf{I}$ absorbs any runs longer than $L$. The probability of the longest run is

$$\begin{aligned} P\{\text{len} \le L\} &= 1 - \mathbf{w}\ \mathbf{r}_{1,L,N} \\ P\{\text{longest run} = L\} &= P\{\text{len} \le L\} - P\{\text{len} \le L-1\} = \mathbf{w}\left(\mathbf{r}_{1,L-1,N} - \mathbf{r}_{1,L,N}\right) \end{aligned} \tag{15}$$

$\mathbf{r}$ is the upper right $s \times s$ sub-matrix after running the chain over $N$ steps, which will be the size of the feature. $\mathbf{w}$ is an $s \times 1$ initial weighting vector which can be the stationary state of $\mathbf{T}$ in general or a unit vector if starting from a single state. Evaluating $\mathbf{R}^N$, we find

$$\mathbf{r}_{1,L,N} = \mathbf{A}^L + \sum_{j=1}^{L} \mathbf{A}^{j-1}\mathbf{B}\mathbf{r}_{1,L,N-j} \tag{16}$$

subject to

$$\mathbf{r}_{1,L,n} = \begin{cases} 0 & n \le L \\ \mathbf{A} & n = L \end{cases}$$

$\mathbf{A}$ is diagonal so the powers in (16) are simply the diagonal elements raised to that power.

These two tests check the collection of symbols that make a run, but we can also step back and look at how patterns of runs form features. The simplest peak would be made of two long runs rising then falling. If these two get broken by opposing runs the height would decrease, depending on the number and length of the interruptions. The statistics of the number of runs follows from a combinatorial counting of shuffles of the symbols, and we can do the same for the height using the runs identified within a feature. [Detail 24] We evaluate the height of the permutations of the runs, reconstructing the signal with a cumulative sum and taking its range $H_{perm}$ as the feature height. [Detail 25] The probability of the observed height is

$$P\{\text{peak height} \ge H\} = \#\{H_{perm} \ge H\}/N_{perm} \tag{17}$$

Of course the number of permutations $N_{perm}$ will be too large to check all exhaustively and we will have to use a random sample. Note that the total sum of the runs equals the difference between the last and first point and is always the same for each permutation. $H_{perm}$ cannot be smaller than this, nor can it be smaller than the longest run. We do not impose any restrictions on the permutations other than preventing adjacent runs of the same direction or symbol, which avoids their merging. By its nature the distribution of permuted heights is coarse. The values are discrete and

the counts increase quickly, creating ties with a broad range of significance levels for a given result. The probabilities in (17) are conservative and other strategies such as taking the mid-point of the level [27] may be more appropriate.

The permutation test for peaks works because the pattern of runs within the permutation determines the height, and the signed difference is a natural simplification such as made in other non-parametric tests. It will not apply to flats, which cannot contain long runs. The adjacency prohibition will force runs to mix, limiting the feature height and making the flats appear commonplace.

We next generalize the runs permutations by returning to the raw signal and using as building blocks its difference. With the spacing as basis this amounts to the second derivative of the order statistics. We no longer take the subset of points within the feature but use the whole sample, because otherwise the test is biased since the feature is chosen for not being representative of the whole. Flats would contain too many small steps and peaks too many large, making either feature appear more likely than it is. We draw steps to fill the feature, reconstruct the signal with a cumulative sum, and determine the range, just as for the permutation test. [Detail 25] Call this an excursion test with feature height $H_{excur}$, although it is really just a bootstrap. The feature's probability after $N_{trial}$ draws is

$$P\{\text{peak height} \geq H\} = \#\{H_{excur} >= H\}/N_{trial} \tag{18}$$

$$P\{\text{flat height} \leq H\} = \#\{H_{excur} <= H\}/N_{trial} \tag{19}$$

again adapted from (17). The test can be done with either the interval spacing or after low-pass filtering. Because the spacing increases strongly in the tails, which would bias the test towards large features, we remove the $N_{top}$ difference from the set if they occur in the first or last $N_{top}/2$ points. Typically $N_{top} = 10$ is sufficient to remove the largest outliers.

# 5 Change-Point Detectors

As we have said, low-pass filtering amounts to kernel density estimation, and the feature detectors are essentially bump hunting. The approach so far has looked at ways to evaluate these features, both parametrically with the models, combinatorially with the runs tests and level sections, and non-parameterically with the data-based excursion test. But, can we analyze the spacing without using feature detectors?

Changepoint analysis has a rich history stretching back to the development of statistical process control techniques. The detectors look for some change in the data, be it distributional parameters such as location or scale, or local trends such as slope or drift, or outliers. Tests range from two-sample consistency checks to cumulative sums to informational criteria and similarity metrics to Bayesian probability, performed on dynamic partitions of fixed data sets or continuous updates of streaming data. The literature is large, and there are more than 25 maintained R libraries, some providing multiple test strategies. [33] and [45] are good recent surveys of the field, and [46] compares the performance of many approaches.

Our experience is that the detectors vary considerably in the quality of their results and their implementation. Although a few are generally stable [Detail 26], none has been completely trustworthy. All produce unreasonable results for some test cases. The detectors are individually noisy and often inconsistent in their placement of changepoints. Some algorithms deal poorly when the spacing contains trends, for example at the edges of flats, while others have problems with regions of

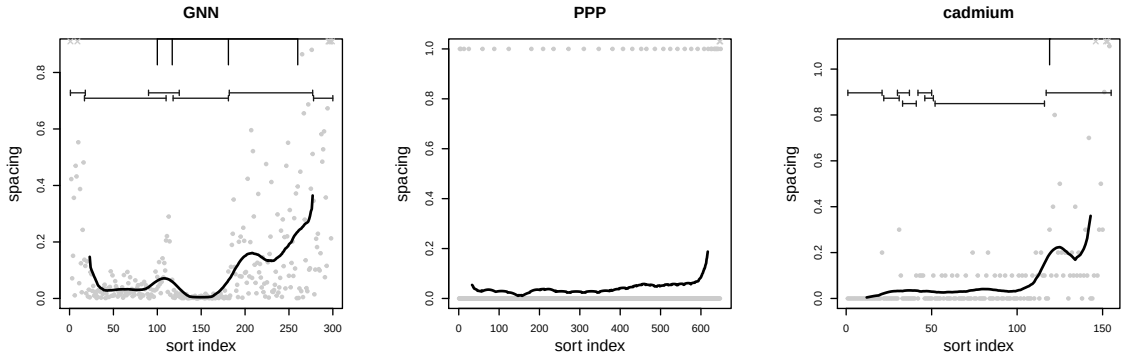



Figure 9: Changepoints and level sections in the examples. Changepoints are marked as tics at the top of each graph, and level sections as bars. Both tests would trigger at each step of the PPP data, the points at spacing 1.0, and are not marked in that graph.

consistent spacing, especially when this appears as a change in local variance but not mean. Some methods have problems in regions of increased spacing, firing multiple times. Many cannot handle discrete data, triggering at each step in value. Few detectors assign a significance level to their results, often taking a threshold level as an input parameter, or even a proxy such as the average run length, without judging the quality of those chosen.

The problem becomes one of classifier fusion, merging the outputs of many detectors. Here the variety of results limits our options. In general there are three approaches [41], depending on whether a test statistic or metric accompanies each decision, or if classifiers assign a result to one of many categories, or if there is a single binary decision. Most of the changepoint detectors do not provide the detailed information for the first and the second does not apply, which leaves only the third option. Classifier fusion based on binary outputs is done by a majority voting scheme [52]. Our voting algorithm [Detail 27] assumes that positions have some variability and treat clustered changepoints as one when voting.

The changepoint analysis runs on the spacing, without any filtering. It successfully identifies the transition between regions of consistent spacing (Figure 9). [Detail 27] shows the voting leading to these results. The analysis does have limitations. The changepoints do not locate features exactly, especially peaks. They do better at the edge of sharp transitions. Several of the detectors, especially those based on process control parameters, require a minimum sample size to generate stable data statistics before looking for deviations. This prevents the detection of changes not only at the start of the data, similar to our handling of the partial overlap of a low-pass kernel, but also after a changepoint. In the worst case the detector fires repetitively at the gap rate. The algorithms also do poorly on discrete data. The PPP data has integer steps, and the detectors, unable to smooth them, signal each as a change. Finally, there is no indication of the quality of the result, so the analysis can only guide the identification of modes. We could try to quantify each changepoint, for example performing a two-sample test [12] to each side to see if the distribution has changed. But these are the kinds of tests already used in the detectors, such as the *cpm* library, and using them separately makes no sense: they are already part of the ensemble.

The endpoints of level sections form a set of changepoints and could be incorporated into the voting. Figure 9 shows a general agreement between the two, and also demonstrates the noisiness of the level detector, at the start of the cadmium data, and a looseness in the alignment to individual

changepoints in the GNN example.

# 6 Evaluation

## 6.1 Examples

[Detail 28] defines the parameters for the feature detectors and filters used to analyze each of our examples.

Table 1 lists the five local maxima in our examples, marked in Figure 2 with dashed lines. The data values can be matched to notches in the histograms in Figure 1. The height model considers the three peaks in the PPP and cadmium examples, plus the second in the GNN example, significant at the 0.05 level. The height of the first GNN peak is too small to pass because the low-pass filter does not respond quickly enough to the clear but narrow increase in the spacing, although it is also insignificant with a 10% kernel size. The excursion test accepts the right GNN peak but no others.

Table 1: Low-Pass Peaks

| | | | | peak model | | | excursion | |
|---|---|---|---|---|---|---|---|---|
| ID | filter | location | data | $h_{peak}$ | side | signif | $H$ | signif |
| Example 1 (GNN) | | | | | | | | |
| 1 | LP | 107 (43 – 149) | 3.4 | 0.860 | R | 0.387 | 0.067 | 0.793 |
| 2 | LP | 208 (149 – 230) | 9.0 | 2.020 | L | 0.044 | 0.157 | 0.028 |
| Example 2 (PPP) | | | | | | | | |
| 3 | LP | 211 (157 – 305) | 8.6 | 1.292 | L | 0.023 | 0.027 | 0.657 |
| 4 | LP | 459 (305 – 491) | 17.0 | 1.655 | L | 0.007 | 0.035 | 0.655 |
| Example 3 (cadmium) | | | | | | | | |
| 5 | LP | 125 (12 – 134) | 6.05 | 2.865 | L | 0.026 | 0.310 | 0.609 |

The interval spacing tests in Table 2 give different results. The ID column corresponds to the equivalent peak in Table 1, although the location is now the end of the interval rather than the center of the filter. The peaks are marked in Figure 8. The Kaplansky-Riordan runs statistics test accepts all but the left GNN peak at the 0.05 significance level. The longest run from Markov Chain modeling accepts the first GNN peak, but the other features are rougher and do not consistently rise or fall. The large step to the left of the second cadmium peak appears smaller with the signed runs than it does in the figure, with the raw spacing. This can also be seen in the feature height for the permutation test, which is small in the cadmium example. Only the second GNN peak passes this test at the 0.05 level. The excursion test would generally produce higher probabilities; 0.495 and 0.212 for GNN peaks 1 and 2, and 0.977 but 0.076 for the cadmium peaks 4 and 5. Because the minima fall in the middle of some flats, we consider each feature to extend to 90% of its full height, in effect moving the minima to the edge of the flat. This shortens the feature and decreases its probability. If we were to take the full width out to the marked minima, then the excursion test results increase, to 0.702, 0.385, 0.988, and 0.468.

Table 2: Interval Spacing Peaks

| | | | | runs count | | longest run | | permutation | |
|---|---|---|---|---|---|---|---|---|---|
| ID | filter | location | data | $U$ | signif | $L$ | signif | $H$ | signif |
| Example 1 (GNN) | | | | | | | | | |
| 1 | Diw | 120 (97 – 160) | 4.803 | −3.630 | 0.000 | 21 | 0.008 | 26 | 0.331 |
| 2 | Diw | 231 (160 – 241) | 12.607 | −2.106 | 0.018 | 13 | 0.093 | 31 | 0.040 |
| Example 2 (PPP) | | | | | | | | | |
| | | not run | | | | | | | |
| Example 3 (cadmium) | | | | | | | | | |
| | Diw | 46 (16 – 74) | 1.143 | −1.203 | 0.115 | 9 | 0.120 | 6 | 0.663 |
| 5 | Diw | 135 (74 – 142) | 7.850 | −1.841 | 0.033 | 6 | 0.134 | 12 | 0.561 |

Flats (Table 3) match in the low-pass filter and interval spacing. In the GNN and cadmium examples they overlap by more than half, covering at least 65% of the length of one of the pair. For the PPP example the flat extends over the sample because the interval is only large enough to cover one or two steps, so the interval spacing is less than two except in the tails. This one flat overlaps the three separate low-pass flats split by the two peaks. The cadmium flat passes the length model. The length model marginally rejects the long PPP flat at the 0.05 level, and neither of the other two come close to acceptance. The excursion test accepts all of the flats with either filter at the 0.05 level, except the left and right low-pass PPP flats. With 5000 bootstrap samples and values rounded to three decimal places, excursion test probabilities of 0.000 mean no samples had so small a height.

Table 3: Flats

| | | | | flat model | | excursion | |
|---|---|---|---|---|---|---|---|
| ID | filter | location | data | $l_{flat}$ | signif | H | signif |
| Example 1 (GNN) | | | | | | | |
| 1 | LP | 33 – 92 | 0.88 – 2.64 | 60 | 0.145 | 0.0169 | 0.008 |
| 1 | Diw | 42 – 109 | 1.09 – 3.63 | 68 | | 0.5571 | 0.000 |
| 2 | LP | 130 – 169 | 4.90 – 5.07 | 40 | 0.430 | 0.0167 | 0.050 |
| 2 | Diw | 139 – 186 | 4.94 – 5.68 | 48 | | 0.4297 | 0.000 |
| Example 2 (PPP) | | | | | | | |
| 3 | Diw | 36 – 618 | 3.7 – 26.6 | 583 | | 2.0000 | 0.000 |
| 3.1 | LP | 42 – 136 | 4.0 – 6.7 | 95 | 0.549 | 0.0175 | 0.325 |
| 3.2 | LP | 169 – 397 | 7.4 – 14.1 | 229 | 0.066 | 0.0175 | 0.015 |
| 3.3 | LP | 426 – 585 | 15.3 – 24.1 | 160 | 0.240 | 0.0177 | 0.081 |
| Example 3 (cadmium) | | | | | | | |
| 4 | LP | 23 – 103 | 0.36 – 2.95 | 81 | 0.000 | 0.0167 | 0.000 |
| 4 | Diw | 32 – 114 | 0.71 – 3.63 | 83 | | 0.5000 | 0.000 |

Checking for changepoints and level sections was not done on the PPP example, because both would identify each step in the values as a feature. Changepoints (Figure 9 and Table 4) in the GNN example mark the edges of the flats, surrounding the peak in the narrow gap and the increase in spacing to the background draw. There is also a changepoint in the trailing tail. The cadmium example generates one changepoint at the transition to the background sample rate, close to the local maximum created by the step in the spacing. These points, however, do not match the peaks. Identifiers in the table are assigned only if within 10 points of the low-pass peak.

Table 4: Changepoints

| ID | source | location | data | votes | level end |
|---|---|---|---|---|---|
| Example 1 (GNN) | | | | | |
| 1 | Di | 100 | 2.9 | 6 | |
| | Di | 117 | 4.7 | 7 | 118 |
| | Di | 182 | 5.2 | 8 | 182 |
| | Di | 261 | 17.2 | 5 | |
| Example 3 (cadmium) | | | | | |
| 5 | Di | 119 | 4.1 | 3 | 117 |

Changepoints generally align to the ends of the level sections (Figure 9 and Table 5). Indeed, the level section test could be treated as another detector and added to the voting list, especially since some changepoint algorithms similarly test regression segments. The level sections include matches to the flats, extending beyond the GNN features and overlapping those in the cadmium data, but there are also additional shorter sections.

Table 5: Level Sections

| ID | source | location | data | length |
|---|---|---|---|---|
| Example 1 (GNN) | | | | |
| 1 | raw | 17 – 110 | 0.2 – 3.8 | 95 |
| | raw | 90 – 125 | 2.6 – 4.9 | 37 |
| 2 | raw | 118 – 181 | 4.8 – 5.2 | 65 |
| | raw | 182 – 277 | 5.3 – 22.3 | 97 |
| | | + 2 short sections at each tail with length $\leq 23$ | | |
| Example 3 (cadmium) | | | | |
| 3 | raw | 52 – 116 | 1.3 – 3.8 | 65 |
| | raw | 117 – 155 | 3.9 – 18.1 | 39 |
| | | + 6 short sections before index 52 with length $\leq 21$ | | |

For an additional example using real-world data, we will analyze the focal depth of earthquakes under Mount St. Helens before its eruption [15]. Magma collected in a resevoir 12 to 7 km below the mountain, and fed to the summit crater through a 50 m conduit to the surface vent system [35]. Starting in March, 1980, two months before the eruption, seismic activity was concentrated within the top of the conduit and in the vent system [28], but there were also deeper earthquakes around the resevoir. The histogram (Figure 10) seems to have three modes for the deep, shallow, and surface earthquakes, separated by anti-modes near $-4$ km and $-0.5$ km. The source paper makes a logarithmic scaling of the depth to emphasize these. This equalizes the histogram, spreading the large peak at the shallowest depths and stretching its neighbor so that the maximum counts are roughly the same. The comb at the right of the histogram comes from quantization of the data, which was taken to two decimal places. If we invert the $x$ axis and take $-\log(-depth)$, we do not change the order of the data. The scaling does change the spacing, with larger peaks and shorter flats. This will affect the evaluation of a feature's significance.

The effect of the scaling is obvious at the shallow anti-mode. With the raw depths there is a small local maximum here, so small that the peak detector does not fire in the interval spacing although a bulge is visible, but the peak is large in the logarithmic depth. The height of the deeper anti-mode also increases somewhat, improving its significance level. The deeper low-pass peak is significant below 0.001 in both the height model and excursion test (Table 6). The shallower peak is also significant at this level after scaling, again for either test.

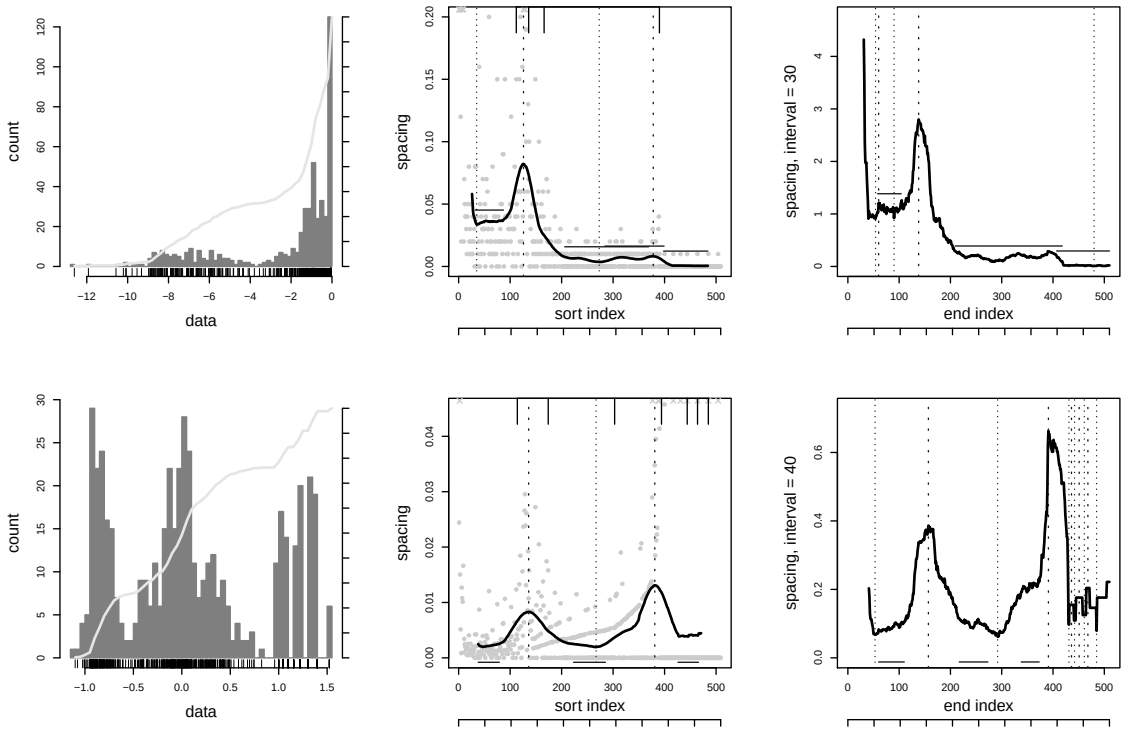



Figure 10: Earthquake focal depths under Mount St. Helens. Top graphs use the raw depths, bottom the logarithm. The left graph has the data histogram, the middle the low-pass filter with 10% Kaiser kernel, and the right the interval spacing with width 30. Features marked as before.

Table 6: Low-Pass Peaks for Earthquake Depths

| | | | | peak model | | | excursion | |
|---|---|---|---|---|---|---|---|---|
| ID | filter | location | data | $h$ | side | signif | $H$ | signif |
| Example 4 (focal depth) | | | | | | | | |
| 1 | LP | 126 (35 – 273) | −4.21 | 3.647 | R | 0.000 | 0.079 | 0.000 |
| 2 | LP | 378 (273 – 484) | −0.27 | 0.355 | R | 0.316 | 0.008 | 0.991 |
| Example 4 ($-\log(-depth)$) | | | | | | | | |
| 1 | LP | 136 (38 – 267) | −2.84 | 2.128 | R | 0.002 | 0.006 | 0.002 |
| 2 | LP | 381 (267 – 471) | −0.24 | 3.723 | L | 0.000 | 0.011 | 0.000 |

In the interval spacing (Table 7) the deeper peak passes the two runs tests but not the permutation. Despite the logarithmic scaling reducing the number of runs, such that the runs statistic test fails and the feature height is smaller, the peak passes both the longest run check at a better level, and passes the permutation test. The shallower peak is significant at 0.004 in the permutation test but both runs tests reject it. The scaling shifts the position of the deeper anti-mode upward, but the shallower location is stable. Table 7 also lists a spurious, very deep peak in the raw depth that all tests reject, and does not include four peaks in the logarithmic depth that correspond to the discrete comb in the histogram and which the low-pass filter smooths into a flat.

Table 7: Interval Spacing Peaks for Earthquake Depths

| | | | | runs count | | longest run | | permutation | |
|---|---|---|---|---|---|---|---|---|---|
| ID | filter | location | data | $U$ | signif | $L$ | signif | $H$ | signif |
| Example 4 (focal depth) | | | | | | | | | |
| | Diw | 60 (54 – 90) | −8.29 | 0.058 | 0.523 | 6 | 0.174 | 9 | 0.291 |
| 1 | Diw | 138 (90 – 480) | −2.70 | −6.495 | 0.000 | 13 | 0.015 | 56 | 0.090 |
| Example 4 ($-\log(depth)$) | | | | | | | | | |
| 1 | Diw | 157 (53 – 292) | −2.07 | −1.106 | 0.134 | 15 | 0.003 | 41 | 0.017 |
| 2 | Diw | 391 (292 – 431) | −0.11 | −0.260 | 0.397 | 6 | 0.263 | 35 | 0.004 |

We see more variability from scaling in the flats, both their number and position. The raw depths contain three, where the weak shallow peak is large enough to separate flats in both the low-pass and interval spacing. This split is natural given the larger peak produced by the logarithmic scaling. The central interval flat splits into two overlapping features in the low-pass spacing (Table 8); a larger ripple specification would merge them. The middle and shallow flats are significant below 0.01 in the excursion test for either filter; the flats are too short to pass the length test. The deep flat, which only partly overlaps between the two filters, is not significant at the 0.05 level. The flats in the logarithmic depth are subsets of the unscaled versions. Here the central interval spacing flat splits into two. These two pieces have different spacings, suggesting they represent a denser shallower mass at −1 km and a broader distribution at −0.4 km. They also appear less significant, with higher probabilities in all tests. None have significant lengths.

Table 8: Low-Pass Flats for Earthquake Depths

| | | | | flats model | | excursion | |
|---|---|---|---|---|---|---|---|
| ID | filter | location | data | $l_{flat}$ | signif | H | signif |
| Example 4 (focal depth) | | | | | | | |
| 1 | LP | 32 – 87 | −8.25 – −6.27 | 56 | 0.625 | 0.004 | 0.075 |
| 1 | Diw | 58 – 104 | −7.33 – −5.56 | 47 | | 0.310 | 0.266 |
| 2.1 | LP | 206 – 369 | −1.27 – −0.35 | 164 | 0.058 | 0.004 | 0.000 |
| 2.2 | LP | 284 – 399 | −0.88 – −0.10 | 116 | 0.224 | 0.004 | 0.002 |
| 2 | Diw | 209 – 418 | −1.27 – −0.08 | 210 | | 0.200 | 0.000 |
| 3 | LP | 398 – 484 | −0.10 – −0.05 | 87 | 0.418 | 0.004 | 0.011 |
| 3 | Diw | 407 – 510 | −0.09 – −0.03 | 104 | | 0.190 | 0.001 |
| Example 4 ($-\log(depth)$) | | | | | | | |
| 1 | LP | 38 – 79 | −8.08 – −6.57 | 42 | 0.698 | 0.001 | 0.057 |
| 1 | Diw | 60 – 110 | −7.11 – −5.38 | 51 | | 0.029 | 0.000 |
| 2.1 | LP | 223 – 285 | −1.19 – −0.87 | 63 | 0.547 | 0.001 | 0.005 |
| 2.1 | Diw | 217 – 273 | −1.22 – −0.91 | 57 | | 0.030 | 0.000 |
| 2.2 | Diw | 338 – 373 | −0.52 – −0.33 | 36 | | 0.038 | 0.038 |
| 3 | LP | 426 – 466 | −0.08 – −0.05 | 41 | 0.705 | 0.001 | 0.055 |

Changepoints (Table 9) surround the anti-modes. Three bracket and match the deeper peak, at depths of −5.35, −2.87, and −1.82 km, and one one marks the shallower, at −0.11 km. The logarithmic scaling matches three of these peaks. Three changepoints, not listed in the table, mark the comb, at −0.04, −0.05, and −0.06 km.

Table 9: Changepoints

| ID | source | location | data | votes | level end |
|---|---|---|---|---|---|
| Example 4 (focal depth) | | | | | |
| 1 | Di | 112 | −5.35 | 4 | |
| | Di | 136 | −2.87 | 4 | |
| 2 | Di | 166 | −1.82 | 4 | |
| 3 | Di | 390 | −0.11 | 5 | 398 |
| Example 4 ($-\log(depth)$) | | | | | |
| 1 | Di | 114 | −5.13 | 4 | 118 |
| 2 | Di | 174 | −1.66 | 4 | 175 |
| | Di | 303 | −0.79 | 3 | 309 |
| 3 | Di | 394 | −0.10 | 4 | 398 |

The longest level sections generally align between the raw and logarithmic depth (Table 10), but we see some variation in their endpoints. There are several short sections not listed in the table, one for the deep tail with depth below −9.43 km and eight above −0.10 km. These sections have the same ends in both scalings, and at most have a width of 21. The level sections in the depth do not match changepoints, but the agreement is better with the logarithmic scaling. Neither matches up well with flats.

Table 10: Level Sections

| ID | source | location | data | length |
|---|---|---|---|---|
| Example 4 (focal depth) | | | | |
| 1 | raw | 9 – 128 | −9.10 – −3.90 | 120 |
| 2 | raw | 95 – 154 | −5.90 – −2.20 | 60 |
| | raw | 155 – 206 | −2.20 – −1.30 | 52 |
| 3 | raw | 207 – 315 | −1.30 – −0.70 | 109 |
| 4 | raw | 270 – 397 | −0.90 – −0.10 | 128 |
| Example 4 ($-\log(depth)$) | | | | |
| 1 | raw | 8 – 118 | −9.43 – −4.90 | 111 |
| 2 | raw | 105 – 185 | −5.54 – −1.50 | 81 |
| 3 | raw | 175 – 308 | −1.62 – −0.73 | 134 |
| 4 | raw | 309 – 383 | −0.73 – −0.22 | 75 |
| | raw | 366 – 397 | −0.37 – −0.10 | 32 |

## 6.2 Accuracy and Stability

The literature about modality testing defines more than 60 artificial distributions of different complexity that are meant to check the performance of the tests under development. They include draws from a single variate and combinations of draws that form distorted single mode densities, as well as true multi-modal setups. Because these samples come from defined variates, we can analyze the density for the location and depth of the sample's modes and anti-modes, allowing us to evaluate the accuracy of our four approaches: features in the low-pass and interval spacing, level sections, and changepoints. The variation in the features found will give us an indication of the stability of the analysis. [Detail 35] defines the samples and plots the features' locations. [Detail 36] quantifies the alignment, counting how many features match to each mode and anti-mode. [Detail 37] presents average counts of the generated features and test consistency. [Detail 38] summarizes all results, grouping samples by their complexity.

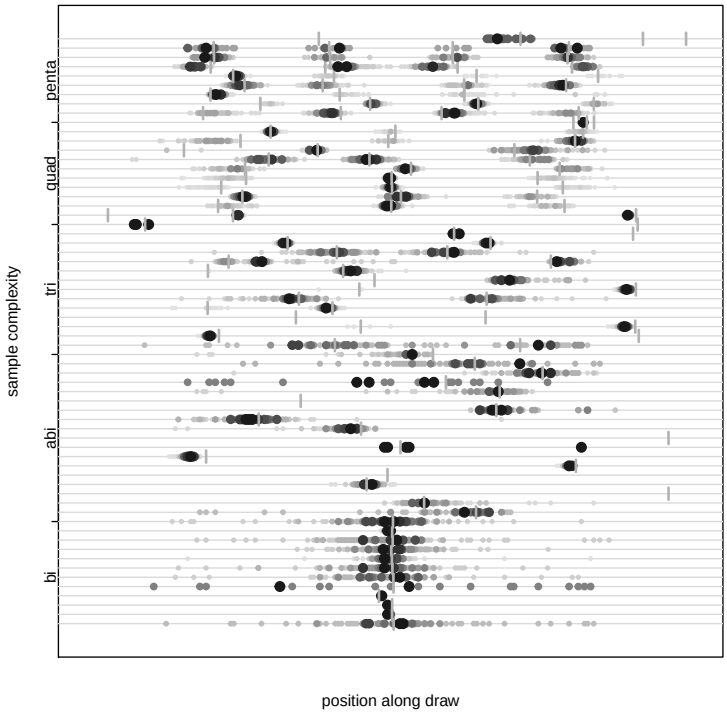



Figure 11: Greyscale histograms of position of significant low-pass peaks in literature samples grouped by number of modes.

We introduce the selectivity to characterize the placement of features. Letting $N_m$ be the number aligned to a mode and $N_{am}$ to an anti-mode, where alignment for a point feature means closest and for a flat means overlap without covering a second extremum,

$$\text{selectivity} = \frac{N_{am} - N_m}{N_{am} + N_m} \tag{20}$$

The metric ranges from $+1$ when all features match the gaps to $-1$ at the modes. Over all literature examples, the selectivity for significant low-pass peaks is $+0.80$, in the interval spacing $-0.04$, and for changepoints $-0.13$. The low-pass peaks do mark anti-modes, but interval peaks are balanced and do not. Changepoints are not in principle tied to mode or anti-mode, and should mark the transition between them, which happens. Flats in either spacing align to modes, with selectivity $-1.0$.

Another way to characterize this is to calculate the distance between a peak and an anti-mode, ignoring any aligned with modes. In the low-pass spacing the average index separation is 1.87, with 95% of the peaks within 5 points of the anti-mode and 99% within 16; the tail is long and the largest placement error is 70. The interval peaks do not align as well. Their average separation is 2.99, with 95% within 10 points and 99% within 27.

We can visualize the stability of the analysis by looking at the position of the peaks over 200 draws of each sample (Figure 11). These are histograms, one sample per line, where larger counts are plotted as larger, darker dots. The greyscale coding is done per sample and not on an absolute scale, so we cannot compare counts or colros directly. Instead, we are looking for well-formed spots without

trails to either side, and how they align to the position of the anti-modes, marked as tics along each line. The samples have been grouped by the number of modes along the $y$ axis, with 'abi' meaning asymmetric bi-modal. Similar charts for detected low-pass peaks or in the interval spacing are much more dispersed. [Detail 35]

The second, third, and fourth lines from the bottom show stable peaks at the anti-mode, while the other bi-modal samples are more marginal, with more dispersion of the peaks, albeit still located near the correct position. Several asymmetric bi-modal and tri-modal samples also have clearly defined peaks, with a handful of exceptions. The more complicated samples tend to have one or two stable features, but also miss as many anti-modes. This speaks to the difficulty of interpreting the results, because the samples deliberately include modes that are hard to resolve. We must rely on other simulations to find where the spacing analysis works and where it breaks down.

## 6.3 Sensitivity

Without being able to model the spacing, we can neither check the accuracy of the tests against a multi-modal setup, nor their stability. Instead we must simulate these situations and look at how the tests respond. The simplest setup is a draw from two normals, where we can change the offset between them by shifting the means, or the width or standard deviation, or the size of the draws. [Detail 29 and those following] In this section we look at what features each test accepts as significant. We also perform similar checks changing one draw of a tri-modal setup. [Detail 34]

The base of comparison will be draws from $250 \times \mathrm{N}(0, 1)$ and $250 \times \mathrm{N}(3, 1)$. This is a variant of the example found in [Detail 1] and has been used several times in the literature [Detail 35], as sample N2 and, in smaller draws, as M1 and H1. We vary one parameter of the second normal for 1000 repetitions of draws, calculating the spacing and passing it through a Kaiser low-pass filter with a kernel covering 15% of the data, then running the peak and flat detectors and gathering the probability of each feature per the model or excursion test. We use the default parameters for the detectors given in [Detail 28], tightening $\delta_{ripple} = 0.02$ for interval flats.

We expect that the increase in spacing at the anti-modes will decrease as the separation decreases, or said differently, for a small second mean. It will also decrease when the two draw sizes are imbalanced and the larger draw dominates the other, or when the width or standard deviation of the second normal gets too large and the larger variation blurs the boundary between the two distributions. We expect two flats in the center of each normal. They will grow with the separation and become more prominent if the width decreases. Imbalanced draw sizes should shrink those in the smaller draw and lengthen those in the larger.

The position of features is consistent with the modes and anti-mode. To see this we plot histograms of the location of each point feature for each variation, or of the number of interval features covering each draw point. Three grey levels mark counts of detected features at 90%, 50% and 10% of the maximum, and a contour line encloses 75% of the significant features. We use the height model to judge peaks and the excursion test for flats, passing at the 0.05 level. Heavy solid lines mark the local maximum in the spacing, solved numerically for each variation. These lines stop if the two variates cannot be distinguished, for example at a separation of 2.2 or a standard deviation of 2.0. Heavy dashed lines mark the local minima, or modes. They merge when the separating peak disappears.

The low-pass peak location drifts away from the ideal position into the smaller draw as the sizes become imbalanced, or into the second draw when its width increases (Figure 12). The detected peaks follow the expected position for all variations. The significant peaks follow the anti-mode and show

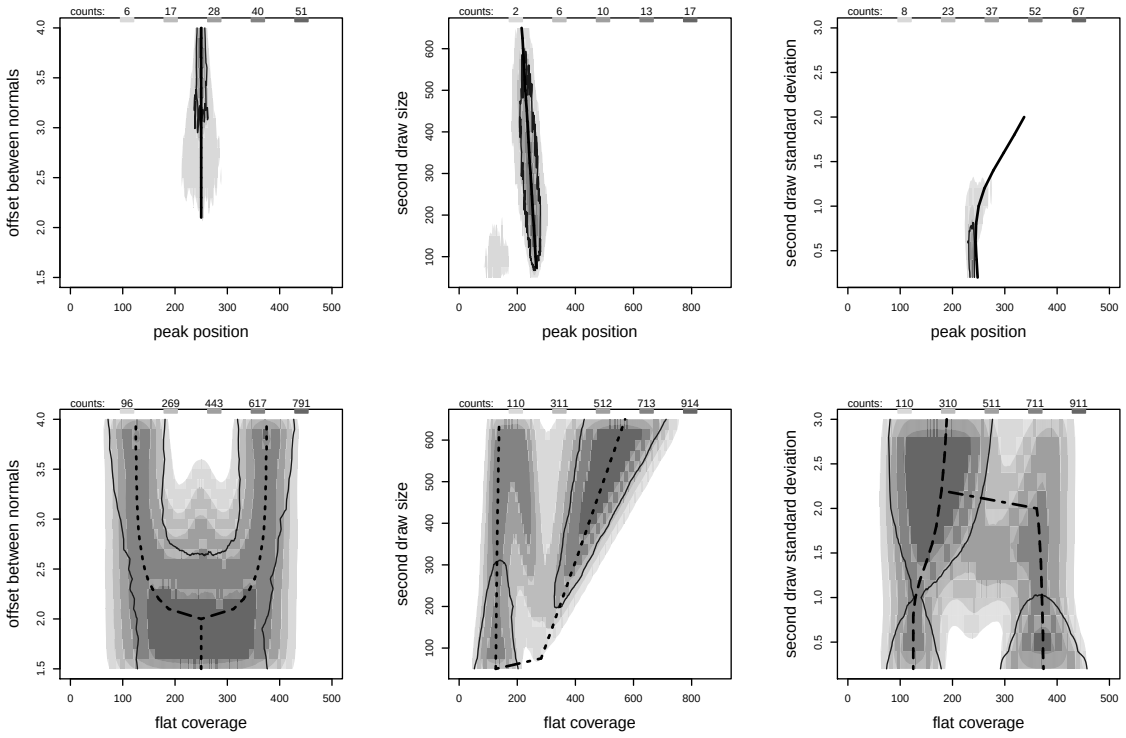



Figure 12: Location of peaks (top) and spans of flats (bottom) in five tiers while changing the second draw ($y$ axis). Heavy lines mark expected locations. Contour lines enclose 75% of the significant features.

which situations cannot be distinguished, for separations below 3.0 and standard deviations above 0.8. Wider contours for smaller offsets indicate an increasing uncertainty in the peak's location. This effect is much smaller for the standard deviation variants. There are some second peaks that appear in the first mode when the second becomes much smaller.

The detected flats not only follow the minima in the spacing, but also appear in the middle of the peaks. This can be seen in the transition of the offset from 2.0 to 3.0, or when the second draw is larger than 350, or the transition of the width from 1.2 to 1.7. Detected flats appear in the second draw even when its standard deviation is above 2.0 and the two modes cannot be distinguished. The countours show that testing rejects these spurious flats. Significant flats track both modes when the separation is more than 2.8, the second draw balances the first with 200 to 300 points, or the standard deviation is below 1.0.

Figure 13 plots the average number of low-pass features found at a significance level of 0.05 over the trials. The left graphs vary the second draw's mean, the middle its size, and the right its width.

The peak tests confirm the expectations. The height model (solid curve in top row) finds a significant peak in all trials for separations larger than 2.7, or standard deviation smaller than 1.2, or if the second draw size balances the first, pulling 200–325 samples. These values correspond to peak heights of 0.00259, 0.00217, and 0.00352. Beyond these thresholds the average number of peaks drops steadily and the test can no longer reliably detect the bi-modal situation. The drop as the draw size varies would be symmetric if the average was plotted against the ratio of the draw sizes.

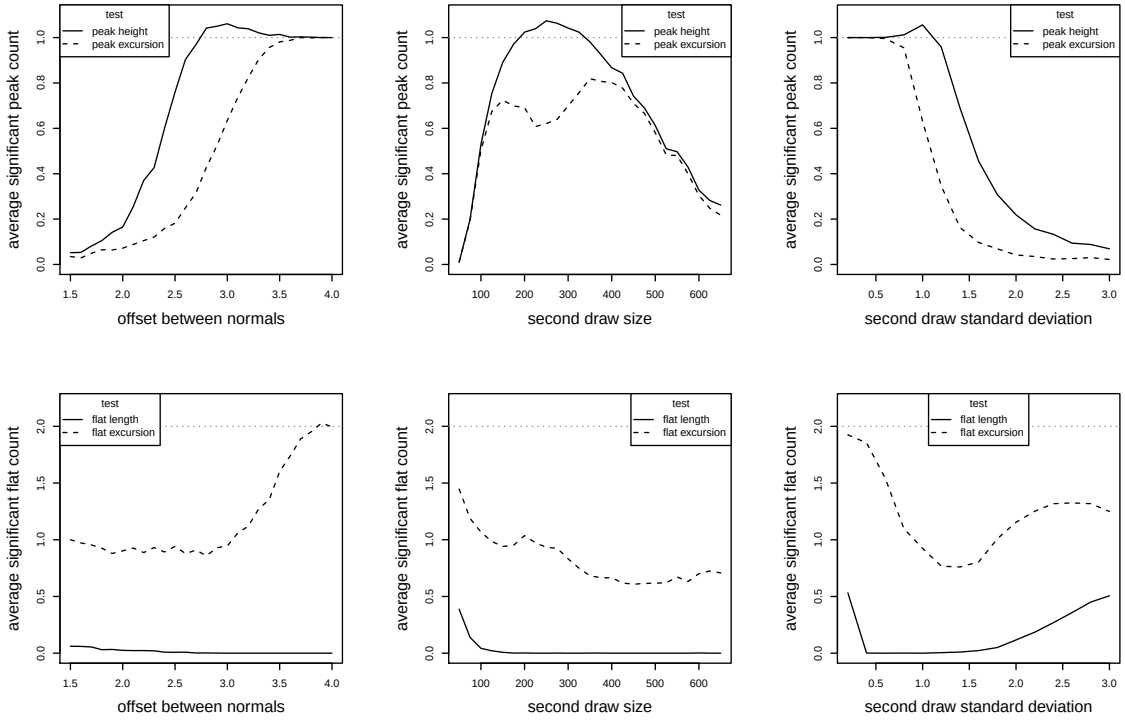



Figure 13: Average number of low-pass features (top row peaks, bottom flats) at a significance level of 0.05 for variations of two normal draws.

The excursion test is more conservative than the peak model, finding a peak if the separation is 3.2 or larger or the standard deviation 0.8 or smaller, equivalent to peak heights of 0.00752 and 0.00766. Under the base conditions it finds a peak in 63% of the trials, and the variation of the draw size is capped at this level. That is, we would have to change the base mean or standard deviation to have the count reach 100%.

The flat detector identifies many potential features, between 3 and 5 per trial, but the tests reject most of them. The excursion test (dashed curve in the bottom row) finds one flat for all variations of the size and when the draws cannot be distinguished, for separations smaller than 3.0 and standard deviations above 0.7. It then transitions to passing two flats at the largest offsets and smallest widths. The length model is insensitive, rejecting all candidates except for the smallest draw sizes or extreme widths.

In summary, the peak tests are able to resolve the two distributions, with the excursion test proving more conservative. The flat tests only succeed in general with a more extreme setup, either with a larger separation between the distributions, or smaller widths. The excursion test here is more the liberal.

Figure 14 plots the significant feature counts from the interval spacing tests. We use $250 \times N(3.75, 1)$ as the base second draw because the sensitivity of these tests are lower and teh anti-mode must be deeper. The range of variations will shift accordingly. The base peak height is 0.0165. The maximum number of features does not change except for flats with large second draws. We use an

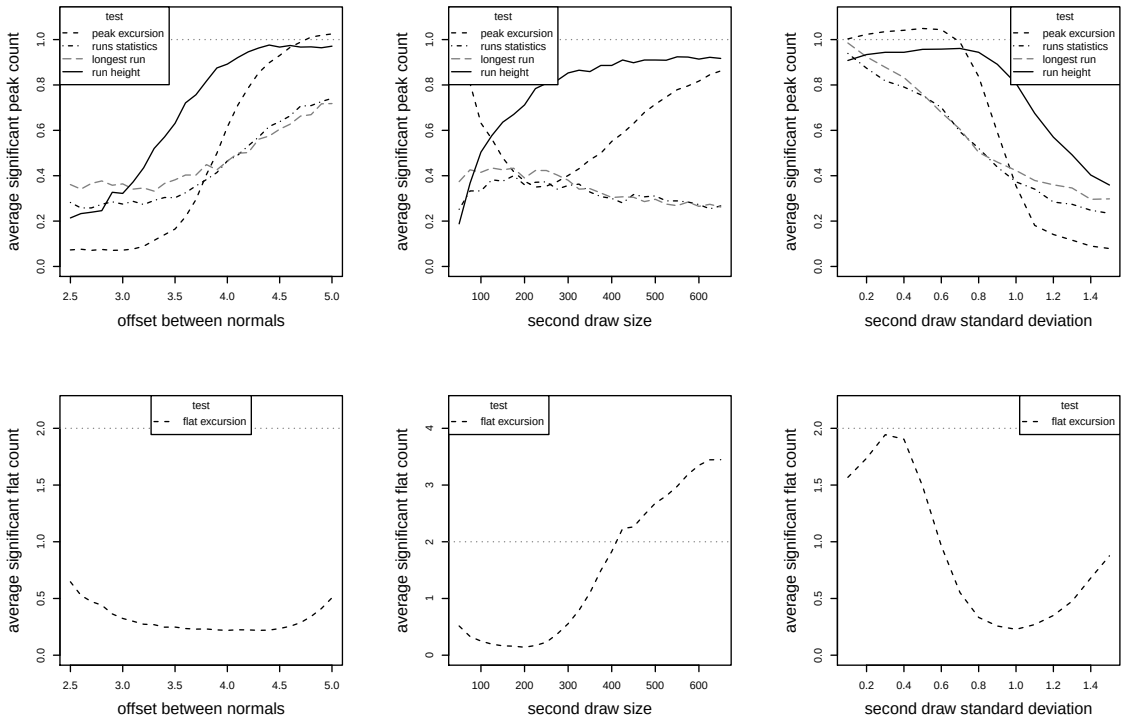



Figure 14: Average number of interval spacing features (top row peaks, bottom flats) at a significance level of 0.05 for variations in the draw from two normals.

interval of 10% of the sample size, giving the same main lobe width as the 15% Kaiser filter.

The excursion test applied to the raw interval spacing has the same sensitivity it shows in the low-pass data for the separation and width variants, although it requires a larger offset or smaller standard deviation. Against the draw size changes the behavior is different, not falling to eventually to zero for unbalanced draws but accepting one peak. The rejection rate with balanced draws is 30% greater, which we explain with the greater variability in the interval spacing caused by poor sidelobe suppression, which would diminish the significance of any peak. That would also imply that flats would be judged more significant, and the results bear this out, keeping in mind that the interval spacing has many fewer flats. [Detail 29] The acceptance rate of the peaks is close to 100%.

Runs in the signed difference of the interval spacing show even less sensitivity. The transition region of the run height permutation test is broader than the excursion test for the separation and width variants, although the center of the transitions has shfited left, accepting smaller peaks. The runs statistics and longest runs tests respond even slower, with the mid-point at larger offsets or smaller widths than the excursion test. The runs tests have a background acceptance rate of 30–40% when the draws cannot be distinguished, equivalent to a high false positive rate. We can accept peaks at the 0.01 level, which shifts the curves down without changing their shape, so that the lowest count is zero but the highest is 0.8. This trades false positives for false negatives.

In summary, the interval spacing tests are less sensitive than the low-pass, requiring larger differences between the draws and responding more slowly to changes. Accepting the runs tests at a 0.01

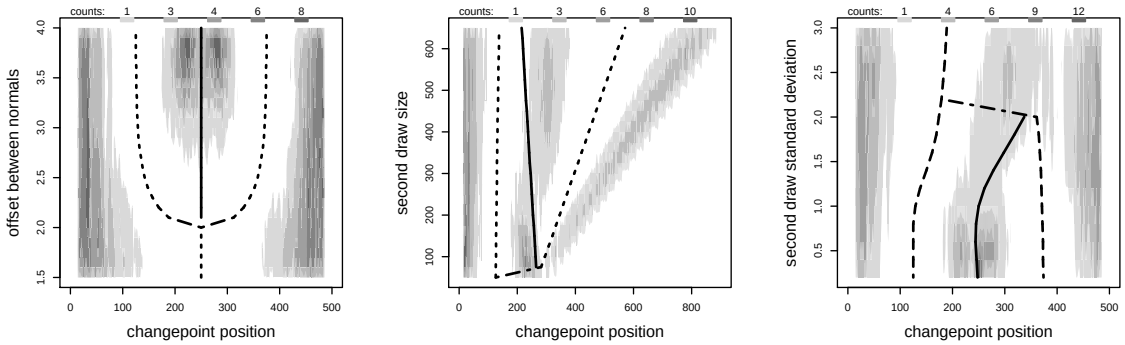



Figure 15: Greyscale coded counts of changepoint positions for each second draw variant. Heavy solid line is the expected maximum spacing, dashed lines minima.

level seems necessary to avoid false positives; the resulting higher false negative rate will still allow them to complement the other tests.

Only when the second draw's standard deviation is small, below 0.5, do we see flats in each mode. A draw of more than 400 points produces one or two flats within the mode. The low counts come from detecting few flats rather than rejecting those found. [Detail 29]

Due to the run time of all the changepoint detectors, we use them only on the first 400 repetitions of each variant. Figure 15 is a grey-scale coded map of the number of changepoints at each location ($x$) for each variation ($y$). Darker colors indicate larger counts. The changepoints are neither localized nor stable. Heavy lines in the graphs mark the modes and anti-mode, as in Figure 12. Changepoints avoid the modes and, looking closely, also fall to either side of the anti-mode. If the modes are broad, supporting flats, then the changepoints mark their edges. They border the peak when the draws can be distinguished, for separations above 3.0 and standard deviations below 1.0, where the peak height is 0.00522, and disappear just outside this range. However, they continue to mark the side of the anti-mode even if the width is above 1.5, responding to the step in variance. The graphs also show that changepoints appear in the strong tails at the start and end of the spacing; in the draw size variants this becomes a sloping line to the right as the total draw grows.

The details have similar plots for the location of detected and significant peaks and flats.

## 6.4 Consistency

The details look at how consistently tests evaluate the same feature, or matching features between the two spacings. The low-pass models and excursion test roughly agree, are significantly associated with each other, and rank pairwise features the same. The probabilities the tests assign to a feature vary roughly linearly, but there is a substantial difference in the acceptance rate. This we anticipated from the shifts in the sensitivity curves. The excursion test passes flats more readily than the models, but peaks less so. Low-pass and excursion flat tests are not associated, but they order features similarly. There is no relationship between the assigned probabilities. Nor are the results of the interval and runs tests consistent. The low-pass and interval spacing tests can be treated as independent checks.

Level sections cover nearly all points of a trial, as a consequence of their definition, and many are

small. We get more sections than changepoints or flats, so matching the features is difficult. We find half to three-quarters of the changepoints align with an endpoint of a section, but that there are three or four times as many level endpoints. [Detail 30] [Detail 33] The level sections form a superset of flats, which are less than half as long and completely covered by a section. They are therefore not good at identifying modes, and cannot replace the flat detectors.

For each sample from the literature [Detail 37] includes consistency checks between features in the low-pass and interval spacing, and between changepoints and level sections. [Detail 38] summarizes the results by the complexity of the samples. 74% of detected peaks match their position, but the rate drops to 40% of the significant peaks. The interval spacing contains twice as many features, and it seems that testing does not pass the matching peaks equally. This might come from ringing shifting peaks, with the accepted ones not aligned to the low-pass maxima. Flats do better, with 90% in the interval spacing having a match in the low-pass spacing. The rate in the other direction is 131%, meaning the interval flats are shorter, with more than one overlapped on average.

## 6.5 Modality Testing

We can compare these results to the performance of existing unimodality tests. For each variation of the second draw's parameters we run 200 repetitions on four established procedures, counting the number of trials judged unimodal at the 0.05 significance level. We use the R package *diptest* for the Hartigans' dip test [18]. The *multimode* package [2] provides the Silverman critical bandwidth test [44], which we use without any corrections, and the excess mass test per [1]. For the folding test we use *Rfolding* [43].

The top half of Figure 16 plots the number of trials accepting the null hypothesis that the data has only one mode. The excess mass test is the most lenient. It begins to separate the modes at a separation above 2.0 or standard deviation below 1.4, when a peak begins to appear in the ideal density. It reliably detects them when the separation is 3.0 or more (a peak height of 0.00552), or the second draw's standard deviation is below 0.8 (0.00766). The critical bandwidth test is next, needing a $+0.2$ larger separation (0.00752) or $-0.1$ smaller standard deviation (0.0100). The dip test is most conservative, requiring a $+0.4$ greater separation (0.0103) or $-0.1$ smaller standard deviation (0.0100). All three tests have sharp transitions. The folding test splits its output, deciding first if data is uni- or multi-modal, and then assigning a $p$ value. We plot the total of the uni-modal trials at the 0.05 level and the multi-modal trials that fail at that level. The transition between the two judgements is not smooth and happens to lie at the base variant. The folding test behaves similarly to the dip but shifts by $-0.4$ separation or $-0.2$ standard deviation, until it switches its decision to accept multi-modality, when then shifts significantly in the other direction. In other words, the multi-modal component is much more conservative, requiring larger differences between the modes, than the other tests.

The bottom half of Figure 16 counts the uni-modal results using peak tests. It combines curves from Figures 44 and 45. [Detail 30] The low-pass height model is more liberal than the excess mass test, even at the 0.01 level. The excursion test at a 0.05 level performs similarly to the dip test, although its transition begins earlier, rejecting the unimodal condition more often when the separation is below 3.0 or the standard deviation above 0.9, and is not as sharp. The excursion peak test in the interval spacing, mirroring Figure 14 about the $y$ axis, has a transition similar to the folding test, centered about a separation shifted by $+1.0$ or $-0.2$ smaller standard deviation.

None of the four existing tests resolve the two variates when the draw size is imbalanced; that they partially succeed when the draws are balanced is due to the base variant. The transition is

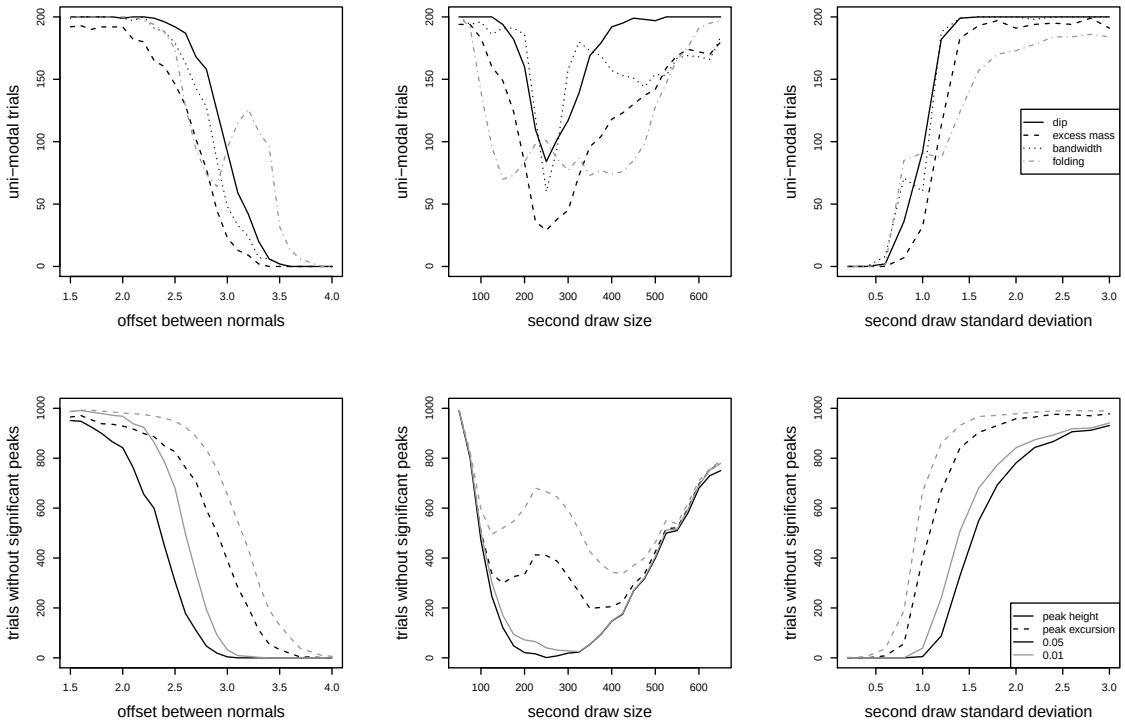



Figure 16: Resolution of unimodality tests (top) as bi-normal parameters change, and equivalent low-pass tests (bottom).

sharp. Once again the excess mass test is most liberal and the dip test most conservative. In contrast the low pass tests do respond to the imbalance, the height model completely and the excursion test partially. That the height test rejects unimodality almost completely when the draws are balanced reflects the other two parameters at the operating point, namely that the separation is large enough for the model to separate the modes. The excursion test succeeds as often as the dip and bandwidth tests for the balanced draw but still responds when they are not, like the height test.

As a second check we can run each test on draws from a single distribution and count how many find multiple modes, a false positive rate. We take points from a normal variate with unit standard deviation and zero mean, or from a uniform over the range zero to one. Evaluating at the 0.05 and 0.01 significance level, we expect no more than that fraction of the trials to reject the null uni-modal hypothesis. Table 11 has the failing rate with underlined values meeting the expectation.

The dip and critical bandwidth tests have appropriate failure rates on the uniform draws, and no false positives on the normal. The excess mass test, on the other hand, is tuned to normal variates and generates an appropriate failure rate for them, but rejects unimodality for uniform draws at too high a rate. The folding test also does well for the normal draw, with no false positives, but not for the uniform, where half the trials are classified as multi-modal.

If we report multi-modality when a sample contains significant peaks, the height model meets and the excursion test slightly exceeds the expected failure rate for the normal variates. Both find a significant peak in almost all uniform samples. This is not unexpected, given how much different

the feature distributions are for uniform variates. The signed run statistics and permutation tests at the 0.01 level correspond to 5% error rates. These tests should only be used at 0.01 and with the understanding that the actual significance is 0.05. The interval spacing tests do better on the uniform trials than the low-pass features, but still require acceptance at the 0.01 level.

Table 11: Uni-modal Rejection Rates

| | normal N(0,1) draws | | uniform U(0,1) draws | |
|---|---|---|---|---|
| | 0.05 level | 0.01 level | 0.05 level | 0.01 level |
| existing tests | | | | |
| Hartigans' Dip | 0.000 | 0.000 | 0.044 | 0.014 |
| folding | 0.000 | 0.000 | 0.482 | 0.484 |
| critical bandwidth | 0.000 | 0.000 | 0.042 | 0.002 |
| excess mass | 0.028 | 0.004 | 0.220 | 0.074 |
| spacing low-pass feature tests | | | | |
| peak height model | 0.029 | 0.003 | 0.990 | 0.968 |
| peak excursion | 0.067 | 0.020 | 0.966 | 0.897 |
| interval spacing tests | | | | |
| peak excursion | 0.120 | 0.102 | 0.196 | 0.038 |
| run height permutation | 0.084 | 0.010 | 0.208 | 0.052 |
| signed run statistics | 0.180 | 0.048 | 0.456 | 0.140 |
| longest signed run | 0.210 | 0.088 | 0.588 | 0.206 |

# 7 Summary

Changes in the spacing of data reflect changes in the underlying distribution. Ignoring edge effects at the very smallest and largest values, the steep sides of a curve shaped in general like a U, we look for two local features. Increases in the spacing correspond to a transition between distributions and multi-modality. Around a mode the spacing is stable over a series of points. The average value over the flat reflects the scale of the underlying distribution. However, the variance of the spacing is generally high and it is hard to identify changes in it. We do see modality effects in the spacing, in examples with separated or overlapping variates, but the problem is extracting and confirming these effects. We have looked at four different approaches to detecting features.

The simplest is to run the spacing through a low-pass filter; this is equivalent to kernel density techniques. We have built parametric models for the probability of peaks and flats that depend on the feature height or length, amount of data, and the low-pass filter kernel size. The peak tests are liberal and although accurate at the 0.05 level, they should be used at the 0.01 level to control false positives. The flat tests are conservative and can be used at the 0.05 level. The models have been chosen for a conservative null distribution, and have too high a false positive rate for uniform unimodal draws. An alternative is to use a bootstrap a.k.a. excursion test drawn from the low-pass filtered signal. This is a non-parametric test that is competitive with other current multi-modal checks, in terms of its sensitivity. It favors flats, and so the recommended acceptance levels are swapped compared to the models: 0.05 for peaks and 0.01 for flats.

Interval spacing is equivalent to running a rectangular or running mean filter over the spacing and suffers from that filter's limitations, with less suppression of sidelobes or high frequency components. It creates a rougher signal, which can require relaxing the feature detector parameters, especially for flats. Peaks are often squared off, less rounded than their low-pass counterparts, and ringing at the top makes the location of the maximum less precise. An excursion test is biased towards accepting

flats and needs a larger separation between modes than the low-pass equivalent to distinguish them. Converting the interval spacing to the sign of its difference and noting whether it increases, decreases, or is constant, allows three more non-parametric tests. A permutation of the runs within a feature is a very conservative test, requiring larger separations between the modes to accept them and responding slowly to such changes. The test has a high fixed offset that can be moved from accepting false features to rejecting real ones by requiring a significance level of 0.01. Runs of the signed difference of the interval spacing can be evaluated for their count by Kaplansky-Riordan statistics or by the longest run distributed according to a Markov chain model. Either are less sensitive than the permutation test. Flats occur rarely in the interval spacing, but there is a trade-off between the roughness of the signal and the ripple parameter. The ripple should be chosen on a case-by-case basis. An excursion test will accept almost any detected flat. It can be used at the 0.05 level.

Interval spacings are part of the Dümbgen and Walther test statistic for regions of non-zero slope within the sample. The test can be inverted to identify level sections, regions with no statistically significant slope. The test is noisy. It generates many intervals and is sensitive to the usual variations in a draw. These sections, by design of the algorithm, fill the sample, with 20% of the data covered by more than one. They appear in the tails at the start and end of the data, with those at the start favored by the way the algorithm grows the intervals. The sections are long, extending beyond a mode, and one will overlap any significant flat. The test cannot be used with discrete data, as each step in the values seeds a section.

A test for trends or slopes in the data is also an example of a changepoint detector, a broad group of algorithms that search for changes in the data, either in its distribution statistics or regression trends or outliers, using parametric and non-parametric techniques in stream or batch processing. Because of the large variety of tests and the information they provide, only the simplest classifier fusion approach is possible, a majority vote over each algorithm's changepoints. The vote does not provide a significance level for each result, unfortunately, so the list of points supplements the feature tests. Changepoints do identify the boundaries between peaks and flats, often incompletely and on only one side. They do not correspond directly to the modes or anti-modes and cannot be used to locate such changes, except in the strongest cases, such as large changes in the data's variance where they mark the edges of a flat. Level sections could be added as a changepoint test, but they are much more plentiful and their endpoints do not match up well. They would likely be removed from the vote because the detectors would appear to be noisy.

These tests are fairly independent of each other and a positive result in any should be considered evidence for a feature. The low-pass models and excursion test are weakly correlated. Testing does screen many of the detected features, passing less than 70%, and those peaks found significant follow the modality, albeit incompletely, as seen in the large number of examples taken from other modality studies. Low-pass peaks identify the transition between modes but the selectivity of interval peaks and changepoints is lower and their placement accuracy poorer. Detected peaks in the low-pass and interval spacing may match, but those considered significant do not. The acceptance rate for flats is high, above 70%, and they are confined to modes, with a selectivity below $-0.9$. Flats in both spacings align, limited by the rarity of the interval features.

Using spacing to detect modality does have some limitations. It only works for one dimensional data, as there is no spacing or sorting order that can be defined in higher dimensions. One could work with the local density, but that has not proven particularly reliable as a clustering strategy. The size of the filter or interval width sets an upper limit on the analysis of discrete data, as there must be some transitions between values to smooth over. Too many samples with the same value, a problem for larger data sets, means a gradual increase in the spacing with no local maxima and long flats. But the data must also contain enough points in a feature, peak or flat, to cause a change in the

filtered signal, or for the changepoint algorithms to react. Too few points and the tests will not be able to separate a signal from the noise, the classic trade-off in selecting a bandwidth. A mode must contain also contain enough points to support a flat. This will be more than the minimum length requirement, because the signal must settle at the edges. The low-pass kernel size and minimum span for changepoints means we cannot analyze spacing changes at the start and end of data. This is not just a problem in artificially generated test examples, where placing small variates at the edges of the main distribution is a common trick, but also appears in real data. These tails are in general hard to analyze, because the spacing changes strongly within them, dominating any local features. The interval spacing has some success picking out local features under these conditions. Finally, extremely large data sets, on the order of tens of thousands to millions of points, will suppress features because the spacing increments become very small, possibly limited by the resolution of the data into quasi-discrete values, and the features exist on a large scale. They can still be seen by tweaking the parameters of the detection algorithm, for example looking for much smaller peaks because the filtered data is so smooth, but this may push the boundary of where the feature models are valid. There are also implementation problems, including an $O(n^2)$ algorithm for extending flats and even more complex changepoint detectors.

# 8 *Dimodal* Package

The results presented in this report come from internal development software. *Dimodal* is the public implementation, available as an R package. In addition to a consistent interface to the data and test results, it makes several changes to the feature detectors and test algorithms. Although the results afterward are similar to the performance of the original version, they do differ in enough details to require new simulations of the bi-modal and literature variants, which will be presented in papers describing the package. The notable changes include

for feature definition,

- Defining the relative height as $|x_1 - x_2|/((|x_1| + |x_2|)/2)$ to handle zero crossings.
- Always treating values within double precision tolerance as equal when building runs, despite the epsilon argument.
- Re-working the mid-quantile approximation, including a strategy that avoids interpolation for non-discrete/quantized values.

for testing features,

- Defining separate acceptance levels for all tests.
- Removing the flat aspect ratio model.

for excursion and permutation tests,

- Changing the definition of the peak height for excursions and the run height permutation tests, as the maximum signal above the lower of the start and finish values. This penalizes negative-going draws and improves the performance of the tests.

- Using a new procedure to generate alternating permutations, based on the frequency of symbols, and making an exhaustive check of the permutations if they are few enough.
- Counting half the exactly matching permutation and excursion heights as a simple mid-quantile estimate.
- Adding the support range height for peaks as an analysis parameter and taking the peak extent between supports for the excursion test.

for changepoint detection,

- Several updates for changepoint libraries. *Dimodal* adds the SdEwma, TsSdEwma, and KnnCad detectors from the *otsad* package. It supports the API change to the *anomaly* package. It uses the endpoints of segments from the *anomaly* and *breakfast* detectors directly, without handling bands.
- Using the level section detector as a changepoint detector.

for feature detection,

- A new search strategy for flats using two segtrees to bound the ripple and move quickly outward, reducing the $O(n^2)$ scan. This approach has substantial overhead, however, and becomes efficient only above 10k points.
- A new voting strategy for flats similar to $z$ buffering, tracking exposed segments and again avoiding an $O(n^2)$ step.
- Allowing any number of outliers within a flat, although the default is still one.
- Using a minheap to efficiently store and update local maxima while removing minor peaks. Edge cases at the start and end of the data require special handling.

# References


[1] Ameijeiras-Alonso, J., Crujeiras, R. M., and Rodríguez-Casal, A. (2016). Mode testing, critical bandwidth and excess mass. *Test* **28**, 3 (Sept.), 900–919.

[2] Ameijeiras-Alonso, J., Crujeiras, R. M., and Rodríguez-Casal, A. (2021). multimode: An R package for mode assessment. *Journal of Statistical Software* **97**, 9 (Mar.), 1–32.

[3] Anderson, T. W. and Goodman, L. A. (1957). Statistical inference about Markov chains. *The Annals of Mathematical Statistics* **28**, 1 (Mar.), 89–110.

[4] Bigot, J. (2005). A scale-space approach with wavelets to singularity estimation. *ESAIM: Probability and Statistics 9*, 143–164. doi: 10.1051/ps:2005007.

[5] Chaudhuri, P. and Marron, J. S. (2000). Scale space view of curve estimation. *The Annals of Statistics* **28**, 2, 408–428.

[6] Chen, J. and Kalbfleisch, J. D. (2005). Modified likelihood ratio test in finite mixture models with a structural parameter. *Journal of Statistical Planning and Inference 129*, 93–107. doi: 10.1016/j.jspi.2004.06.041.

[7] Cheng, M.-Y. and Hall, P. (1999). Mode testing in difficult cases. *The Annals of Statistics 27*, 1294–1315.

[8] Chung, K. L. (1976). Excursions in Brownian motion. *Arkiv för Matematik* **14**, 1–2 (Dec.), 155–177.

[9] Davies, P. L. and Kovac, A. (2004). Densities, spectral densities and modality. *The Annals of Statistics* **32**, 3, 1093–1136.

[10] DeCarlo, L. T. (1997). On the meaning and use of kurtosis. *Psychological Methods* **2**, 3, 292–307.

[11] Dorea, C. C. Y., Goncalves, C. R., and Resende, P. A. A. (2014). Simulation results for Markov model selection: AIC, BIC and EDC. In *Proceedings of the World Congress on Engineering and Computer Science*. Vol. **II**. WCECS, San Francisco.

[12] Dowd, C. (2020). A new ECDF two-sample test statistic. arXiv:2007.01360v1.

[13] Dümbgen, L. (2002). Application of local rank tests to nonparametric regression. *Journal of Nonparametric Statistics 14*, 511–537. doi: 10.7892/boris.73750.

[14] Dümbgen, L. and Walther, G. (2008). Multiscale inference about a density. *The Annals of Statistics* **36**, 4, 1758–1785. doi: 10.1214/07-AOS521.

[15] Duong, T., Cowling, A., Koch, I., and Wand, M. P. (2008). Feature significance for multivariate kernel density estimation. *Computational Statistics and Data Analysis 52*, 4225–4242. doi: 10.1016/j.csda.2008.02.035.

[16] Fisher, N. I. and Marron, J. S. (2001). Mode testing via the excess mass estimate. *Biometrika* **88**, 2, 499–517.

[17] Harris, F. J. (1978). On the use of windows for harmonic analysis with Discrete Fourier Transforms. *Proceedings of the IEEE* **66**, 1 (Jan.), 51–83.

[18] Hartigan, J. A. and Hartigan, P. M. (1985). The dip test of unimodality. *The Annals of Statistics* **13**, 1, 70–84.

[19] Haynes, K., Fearnhead, P., and Eckley, I. A. (2017). A computationally efficient nonparametric approach for changepoint detection. *Statistics and Computing* **27**, 5 (Sept.), 1293–1305. doi: 10.1007/s11222-016-9687-5.

[20] Holzmann, H. and Vollmer, S. (2008). A likelihood ratio test for bimodality in two-component mixtures — with application to regional income distribution in the EU. *AStA Advances in Statistical Analysis* **92**, 1, 57–69.

[21] Inclan, C. and Tiao, G. C. (1994). Use of cumulative sums of squares for retrospective detection of changes of variance. *Journal of the American Statistical Association* **89**, 427 (Sept.), 913–923.

[22] Kaplansky, I. and Riordan, J. (1945). Multiple matching and runs by the symbolic method. *The Annals of Mathematical Statistics 16*, 272–277.

[23] Keogh, E., Lin, J., Lee, S.-H., and van Herle, H. (2006). Find the most unusual time series subsequence: algorithms and applications. *Knowledge and Information Systems 11*, 1–27.

[24] Kreider, G. (2023a). Expected spacing. *Communications in Statistics - Theory and Methods* **53**, 23, 8286–8296. doi: 10.1080/03610926.2023.2281265.

[25] Kreider, G. (2023b). Spacing estimator. Tech. rep., Primordial Machine Vision Systems, Inc., Lyndeborough, NH. Dec. https://www.primachvis.com/.

[26] Larkin, R. P. (1979). An algorithm for assessing bimodality vs. unimodality in a univariate distribution. *Behavior Research Methods and Instrumentation* **11**, 4, 467–468.

[27] Ma, Y., Genton, M. G., and Parzen, E. (2011). Asymptotic properties of sample quantiles of discrete distributions. *Annals of the Institute of Statistical Mathematics 63*, 227–243.

[28] Malone, S. D. (1990). Mount St. Helens, then 1980 re-awakening and continuing seismic activity. *Geoscience Canada* **17**, 3 (Sept.), 146–150.

[29] Marron, J. S. and Wand, M. P. (1992). Exact mean integrated squared error. *The Annals of Statistics* **20**, 2, 712–736.

[30] Minnotte, M. C. (1997). Nonparametric testing of the existence of modes. *The Annals of Statistics* **25**, 4, 1646–1660.

[31] Minnotte, M. C. and Scott, D. W. (1993). The mode tree: a tool for visualization of nonparametric density features. *Journal of Computational and Graphical Statistics 2*, 51–68.

[32] Mukhopadhyay, S. (2017). Large-scale mode identification and data-driven sciences. *Electronic Journal of Statistics* **11**, 1, 215–240. arXiv:1509.06428v4.

[33] Niu, Y. S., Hao, N., and Zhang, H. (2015). Multiple change-point detection: a selective overview. *Statistical Science* **31**, 4 (Nov.), 611–623. arXiv:1512.04093v1.

[34] Oppenheim, A. V. and Schafer, R. W. (1989). *Discrete-Time Signal Processing.* Prentice Hall, Englewood Cliffs, New Jersey.

[35] Pallister, J. S., Hoblitt, R. P., Crandell, D. R., and Mullineaux, D. R. (1992). Mount St. Helens a decade after the 1980 eruptions: magmatic models, chemical cycles, and a revised hazards assessment. *Bulletin of Volcanology* **54**, 2, 126–146.

[36] Pierre-Jean, M., Rigaill, G., and Neuvial, P. (2014). Performance evaluation of DNA copy number segmentation methods. *Briefings in Bioinformatics* **16**, 4, 600–615. doi: 10.1093/bib/bbu026.

[37] Proschan, F. and Pyke, R. (1967). Tests for monotone failure rate. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability.* Vol. **3**. 293–312.

[38] Pyke, R. (1965). Spacings. *Journal of the Royal Statistical Society, Series B* **27**, 3, 395–449.

[39] Rikken, M. G. J. and van Rijn, R. P. G. (1993). Soil pollution with heavy metals — an inquiry into spatial variation, cost of mapping and the risk evaluation of copper, cadmium, lead and zinc in the floodplains of the Meuse west of Stein, the Netherlands. Tech. rep., Utrecht University, Department of Physical Geography. Doctoraalveldwerkverslag.

[40] Rufibach, K. and Walther, G. (2010). The block criterion for multiscale inference about a density, with applications to other multiscale problems. *Journal of Computational and Graphical Statistics* **19**, 1, 175–190.

[41] Ruta, D. and Gabrys, B. (2000). An overview of classifier fusion methods. *Computing and Information Systems 7*, 1–10.

[42] Shaughnessy, P. W. (1981). Multiple runs distributions: Recurrences and critical values. *Journal of the American Statistical Association* **76**, 375 (Sept.), 732–736.

[43] Siffer, A., Termier, A., Fouque, P. A., and Largouet, C. (2018). Are your data gathered? The folding test of unimodality. In *24th ACM SIGKDD International Conference on Knowledge Discovery*. London, 2210–2218. doi: 10.1145/3219819.3219994.

[44] Silverman, B. W. (1981). Using kernel density estimates to investigate multimodality. *Journal of the Royal Statistical Society, Series B* **43**, 1, 97–99.

[45] Truong, C., Oudre, L., and Vayatis, N. (2020). Selective review of offline change point detection methods. *Signal Processing 167*. arXiv:1801.00718v3.

[46] van den Burg, G. J. J. and Williams, C. K. I. (2022). An evaluation of change point detection algorithms. arXiv:2003.06222v3.

[47] Venter, J. H. (1967). On estimation of the mode. *The Annals of Mathematical Statistics 38*, 1446–1455.

[48] Wald, A. and Wolfowitz, J. (1940). On a test whether two samples are from the same population. *The Annals of Mathematical Statistics 11*, 147–162.

[49] Wegman, E. J. (1970). Maximum likelihood estimation of a unimodal density function. *The Annals of Mathematical Statistics 6*, 2169–2174.

[50] Wilks, S. S. (1948). Order statistics. *Bulletin of the American Mathematical Society* **54**, 1 (Jan.), 6–50.

[51] Xu, L., Bedrick, E. J., Hanson, T., and Restrepo, C. (1024). A comparison of statistical tools for identifying modality in body mass distibutions. *Journal of Data Science 12*, 175–196.

[52] Xu, L., Kryzyżak, A., and Suen, C. Y. (1992). Methods of combining multiple classifiers and their applications to handwriting recognition. *IEEE Transactions on Systems, Man, and Cybernetics* **22**, 3 (May), 418–435.

# Detailed Notes

### Detail 1 Bimodal Spacing Example

We combine two normal draws, one of 150 points from N(0, 0.5) and the other of 250 points from N(5, 2). Figure 17 shows the relationship between the distribution function and the spacing. In the middle graph the light curve is the actual CDF, the sum of each draw's CDF weighted by the draw size. The dark curve is an inverse CDF formed by summing the expected spacing. In the right graph the dark curve is the expected spacing from numeric integration and the light curve is the quantile estimator of the spacing, here calculated as the numeric derivative of the actual inverse CDF.

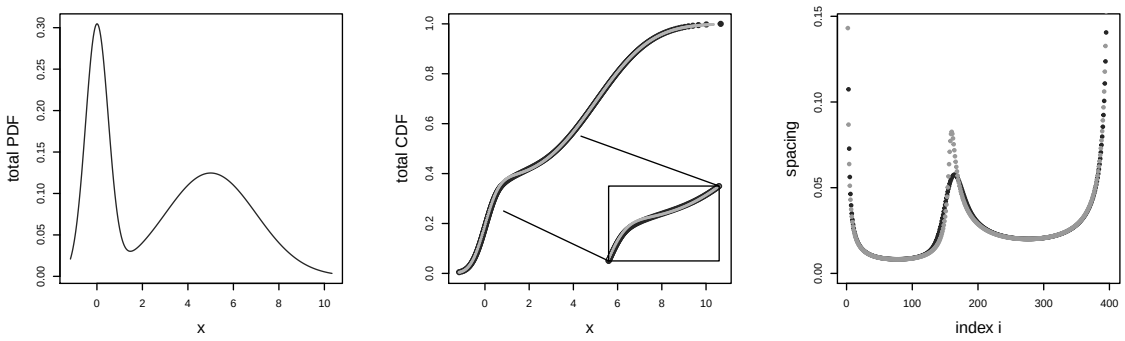



Figure 17: Density of the bimodal draws (left), the distribution (middle), and spacing (right). The light curves are the actual CDF (middle) and the derivative of its inverse (right), called the quantile estimator. The dark curves show the expected spacing from numeric integration (right), and reconstructed distribution function (middle).

In the CDF graph the slope inside each mode, around the inflection points at $x = 0$ and $x = 5$, depends inversely on the standard deviation. The inter-mode transition at $x = 1.46$ shows a flattening of the distribution as we enter the right tail of the first draw while still in the left tail of the second. This corresponds to the increase in the spacing at $i = 163$. Within each mode the slope is fairly constant, giving a flat spacing at indices $i = 65$ and $i = 275$, with the vertical offset between them corresponding to difference in the standard deviations. The greatest difference between the actual and estimated spacing occurs in the transition, shown in the inset by way of the actual and reconstructed CDF.

### Detail 2 Spacing of the Examples

The GNN and PPP examples have known distributions and we can calculate the expected spacing by integrating numerically.

Figure 18 presents the density for the GNN example and the spacing, with the expected value from numeric integration of (1) drawn as dots and the 25% and 99% quantiles from 50 thousand simulated draws drawn as dotted and dashed lines. The expected spacing falls between the median and 75% quantile, fitting best to the 62%. The expected spacing does not show much of a peak between the variates. These transitions will not occur exactly at the boundary of each draw, because the third variate is very wide and will contribute many points to the first two modes, shifting the transitions

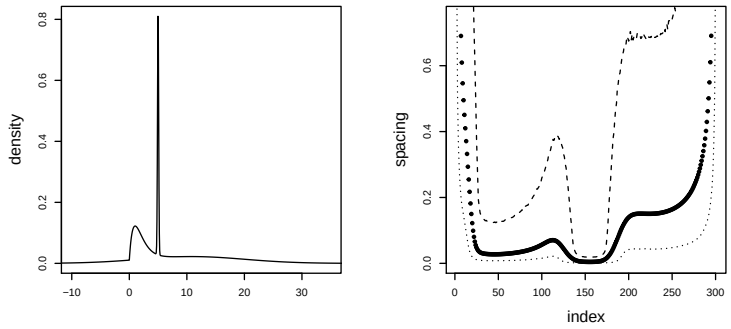



Figure 18: Density of the GNN example (left) and spacing (right).

right on the graph. The local peaks are small, falling at indices 112 and 212 with the second not at all obvious in the graph. Their heights are 0.00028 and 0.000038. Comparing this to the size of the narrow normal variate, 60 points, and of the gamma, 150 points, the shift is clear. The growth of the peaks at the upper quantiles is rapid, with the left peak height at 90% half that shown in the 99% quantile. The local peaks, especially the lower, are prominent in 1% of the trials. The tight draw creates the flat in the middle indices, but it not hard to see a second to its left, and possibly one to its right that could incorporate the second peak. The differences in the spacing over each of these regions reflect the scales of the three draws.

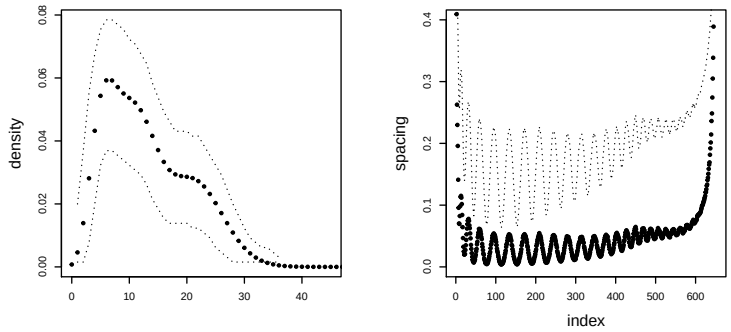



Figure 19: Density of the PPP example (left) and simulated spacing (right).

The left graph of Figure 19 shows the density of the PPP example as dots, since Poisson variates are discrete. The dotted lines show the 1% and 99% quantiles after 50 thousand simulated draws. Multiplying the density by the total draw size gives a value that is repeated that many times, so that the two maximum dots represent sequences of 38 each 38 points long. The main text identifies 7 and 17 as the key values, which correspond to the maximum width and the shelf to the right. From the same trials we also calculate the spacing. However, this has only two values, 0 or 1, depending on whether the draw's value changes at that index. We plot on the right the average spacing as dots, and the standard deviation as a dotted line; it makes no sense to derive a quantile. The beats come from the variance in the number of each value drawn, so the "wavelength" corresponds to the density times the total sample size.

The order statistics behind the spacing only hold for continuous variates. If we were to use a discrete sum instead of the integral in (1) we would have an expected spacing that would look like an upside-down version of the density when plotted on a log scale. We cannot predict the expected spacing for discrete distributions, and must depend on averaging the steps, either here from the repeated samples or during the analysis from filtering, to detect the changes. The low-pass spacing in the main text looks like the spacing in this figure, without the beats.

### Detail 3 Filter Evaluation

The conclusion to [17] introduces a figure of merit to compare filters, but does not provide values in the comparison table. The figure of merit is the difference between the equivalent noise bandwidth and 3 dB bandwidth, as a fraction of the 3 dB bandwidth. Acceptable values will be between 0.04 and 0.055. Table 12 gives the figure of merit for the filters we have discussed.

Table 12: Figures of Merit

| | |
|---|---|
| rectangular | 0.124 |
| Bartlett | 0.039 |
| Hanning ($\alpha = 2$) | 0.042 |
| Hamming | 0.046 |
| Gaussian ($\alpha = 3$) | 0.058 |
| Kaiser ($\alpha = 3.5$) | 0.055 |
| Blackman (exact) | 0.033 |

The $\alpha$ chosen for the Kaiser filter corresponds to a ripple of 0.01. The overall recommendation in the text is based on the number of peaks found and their height. [Detail 8] It follows the figure of merit except for the rank of the Bartlett filter. The two filters that fall outside the acceptable range, the Gaussian and Blackman, are at the bottom of the recommended list.

### Detail 4 Filter Comparison

Figure 20 shows the range of filter behavior on the GNN example at two different kernel sizes. The 15% Kaiser kernel, drawn as a thin dark line, generates an smooth curve that captures the peak at index 100 and the flats to either side. At a window sized to 10% of the data a second peak begins to appear at index 200, the middle flat is more sharply defined, the first flat becomes uneven, and the curve is less smooth. The Blackman-Harris filter with 15% window, drawn in medium grey, is smooth, has a more prominent peak at 100 and wider flats, but also has more ringing, visible in the left flat and the rise to the right of the second flat. In the 10% window it shows significant ringing, although the middle flat does not change. The second peak at 200 is just one of several local maxima. The 15% rectangular filter, drawn in light grey, does not produce a smooth curve, instead generating many high-frequency steps atop ringing. The peak at index 100 is less prominent, and the flats narrower and noisier. At 10% width we get substantial ringing.

The Nuttall filter, a fourth degree cosine filter, behaves like the Blackman-Harris, with even more exaggerated ringing. The Bartlett, Hamming, and Hamming filters lie in-between the Kaiser and Blackman-Harris, with the peak response increasing and flats length decreasing in that order, but without the ringing.

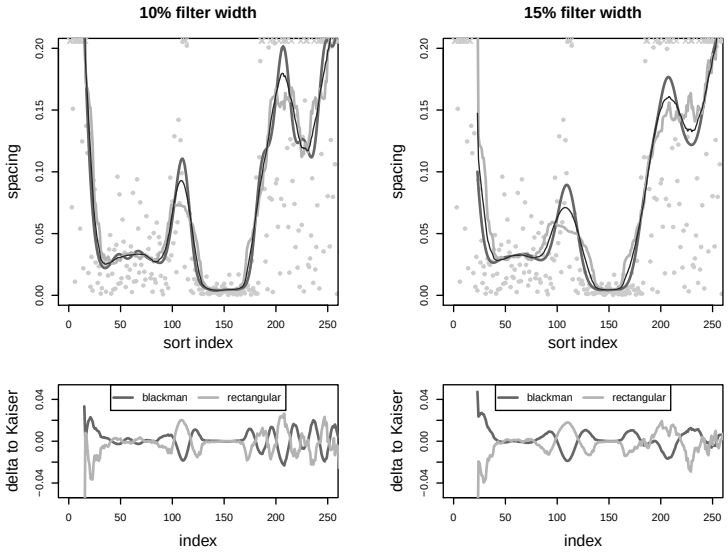



Figure 20: Performance of low-pass filters on the GNN example, for two different kernel widths. Base Kaiser is draw as a thin curve.

### Detail 5 Local Extrema Detection Algorithm

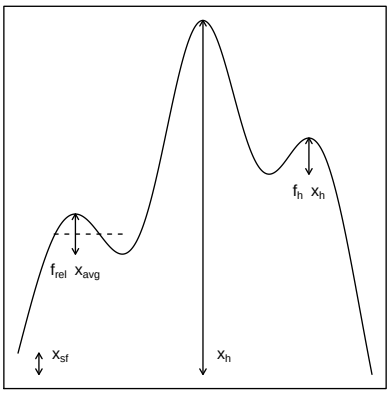



Figure 21: Heights used to screen local extrema.

Listing 1 provides pseudo-code for finding local minima and maxima, ignoring small side peaks. We drop a peak if its height off the adjacent minimum is less than a fraction $f_h$ of the range of the entire data, or if the feature height is less than a fraction $f_{h,rel}$ of the average of the extrema (Figure 21). The first condition imposes a global minimum peak height, the second is local and relative. Default values are $f_h = 0.05$ and $f_{h,rel} = 0.15$ to avoid numerical precision effects. The algorithm treats values within the relative fraction $f_{h,tie}$ of their average as the same and places a maximum in the middle of a sequence of ties. The default value is 0.001. The feature may have an offset $x_{sf}$ between the start and finish; it need not be an excursion. For each extrema the algorithm reports its position and value.

| Listing 1: Local Extrema Detector | |
|---|---|
| Input: | data $x$ of $n$ points, fractional thresholds $f_h$, $f_{h,rel}$, $f_{h,tie}$ |
| 1: | let $x_h \leftarrow (\max(x) - \min(x))f_h$ |
| 2: | create array $pos[1 \dots n] \leftarrow 1 \dots n$ |
| 3: | for $j_{st}$ in $1 \dots n$: |
| 3a: | $j \leftarrow j_{st}$ |
| 3b: | while $j \leq (n-1)$ and $(2\|x[j+1] - x[j]\|/(x[j+1] + x[j]) <= f_{h,tie})$: $j \leftarrow j+1$ |
| 3c: | if $j_{st} < j$ : |
| 3c1: | $x[j_{st} \dots j] \leftarrow x[j_{st}]$ |
| 3c2: | $pos[j_{st} \dots j] \leftarrow (j_{st} + j)/2$ |
| 4: | let $\mathcal{I}_{min} \leftarrow \{i \mid x[i] < x[i-1], x[i] < x[i+1]\}$, |
| 5: | let $\mathcal{I}_{max} \leftarrow \{i \mid x[i] > x[i-1], x[i] > x[i+1]\}$ |
| 6: | let $\Delta x_i = \|x[i+1] - x[i]\|$ |
| 7: | let $\overline{x_i} = (x[i+1] + x[i])/2$ |
| 8: | while $\mathcal{I}_{min} \cup \mathcal{I}_{max} \neq \varnothing$: |
| 8a: | $\mathcal{I}_{drop} \leftarrow \{i \mid (i \in \mathcal{I}_{min} \cup \mathcal{I}_{max}) \cap ((\Delta x_i \leq x_h) \cup (\Delta x_i \leq f_{h,rel}\overline{x_i}))$ |
| 8b: | if $\mathcal{I}_{drop} = \varnothing$ then break |
| 8c: | $i_{drop} \leftarrow argmin_{i \in \mathcal{I}_{drop}} \Delta x_{i-1}$ |
| 8d: | $\mathcal{I}_{min} \leftarrow \mathcal{I}_{min} \setminus i_{drop}$ ; $\mathcal{I}_{max} \leftarrow \mathcal{I}_{max} \setminus i_{drop}$ |
| 9: | report minima in $\mathcal{I}_{min}$: position $pos[\mathcal{I}_{min}]$, value $x[\mathcal{I}_{min}]$ |
| 10: | report maxima in $\mathcal{I}_{max}$: position $pos[\mathcal{I}_{max}]$, value $x[\mathcal{I}_{max}]$ |

Step 1 sets the global minimum height. Steps 2 and 3 handle rough maxima by collapsing nearly equal plateaus, whose relative roughness is within $f_{h,tie}$, to the mid-point. Steps 4 and 5 identify the local extrema, and steps 6 and 7 define the local change and average. Note that $x$ need not be sorted and $\Delta x_i$ is not the spacing. Step 8 does the screening, calculating the peak height to either side in 8a and picking the smallest peak in 8c. Drops can chain, so by removing the smallest first we may keep other intermediate peaks.

### Detail 6 Peak Height Distribution by Variate

We run 100 thousand trials of draws of 200 points of many random variates, passing the spacing of each through a low-pass Kaiser filter with 30 point (15%) window, and registering local maxima per the feature detector. [Detail 5] Peaks require minima to either side, so that maxima at the start or end of the data are ignored. We track the number of maxima and the height to each adjacent minimum, scaled by the signal's standard deviation. We use gamma, Gumbel, logistic, exponential, normal, Weibull, beta, and uniform variates. All have default parameters, or zero location and unit scale, except the Weibull draw with scale $a = 2$ and shape $b = 4$, the gamma with shape $r = 2$ and rate $\lambda = 4$, the beta with shapes $r = 2$ and $s = 2$, the F with 10 and 6 degrees of freedom, and the Rayleigh with scale $\sigma = 10$. The tables presenting feature counts, both peaks and flats, may omit a column in order to fit the page if the counts are small; any missing values may be recovered by bringing the row total to the number of trials.

The grouping was initially done by the clusters of curves in Figure 22, and can be seen in both the average number of peaks and their height in Table 13. Two are outliers. Uniform variates generate the most and largest peaks, and beta variates are still noticeably above the others. The third group contains the normal, Rayleigh, and Weibull variates, with somewhat higher counts, 1.2, and larger heights, 2.50 at the 0.99 quantile. The remaining variates form the fourth group, and have less than

1 peak per trial and a height around 2.0. Three in this group, the exponential, F, and Wald variates, have heights that transition into the third group at large quantiles.

Table 13: Peak Count per Base Distribution

| | | number of peaks | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | group | 0 | 1 | 2 | 3 | 4 | 5 | mean |
| beta | 2 | 5859 | 38920 | 43757 | 10872 | 587 | 5 | 1.61 |
| chi-squared | 4 | 48369 | 41845 | 9171 | 603 | 12 | | 0.62 |
| exponential | 4–3 | 81783 | 16983 | 1208 | 25 | 1 | | 0.19 |
| F | 4–3 | 99543 | 454 | 3 | | | | 0.005 |
| gamma | 4 | 56317 | 36535 | 6795 | 347 | 6 | | 0.51 |
| Gumbel | 4 | 48601 | 41760 | 9120 | 517 | 2 | | 0.62 |
| logistic | 4 | 36728 | 48995 | 13549 | 811 | 7 | | 0.78 |
| normal | 3 | 16908 | 52079 | 27842 | 3117 | 54 | | 1.17 |
| Rayleigh | 3 | 17576 | 50537 | 28237 | 3573 | 76 | 1 | 1.18 |
| t | 4 | 52075 | 40146 | 7431 | 346 | 2 | | 0.56 |
| uniform | 1 | 400 | 12852 | 45984 | 34716 | 5887 | 160 | 2.33 |
| Wald | 4–3 | 97025 | 2929 | 46 | | | | 0.03 |
| Weibull | 3 | 14383 | 50661 | 31000 | 3887 | 69 | | 1.25 |

Ignoring the first two groups, the Weibull setup is the most conservative. A peak of some standardized height will have the lowest significance level if treated as coming from the Weibull (2, 4) distribution (Table 14). For example, the critical value 1.9260 at $q = 0.95$ falls into the $q = 0.975$ bucket for the group 3 and 4–3 transition variates, and into $q = 0.99$ or $q = 0.995$ for group 4. The spread in significance levels narrows at larger $q$ as the curves in Figure 22 converge. At $q = 0.99$ the Weibull critical value, 2.6060, appears in the 0.995 – 0.999 buckets for groups 3 and 4, except for the logistic and $t$ variates, where it is significant at $q = 0.9995$.

Table 14: Critical Peak Heights per Base Distribution

| | | quantile | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | group | 0.90 | 0.95 | 0.975 | 0.99 | 0.995 | 0.999 | 0.9995 |
| beta | 2 | 2.7490 | 3.1328 | 3.4319 | 3.7245 | 3.8987 | 4.2171 | 4.3182 |
| chi-squared | 4 | 1.1066 | 1.4142 | 1.7383 | 2.1905 | 2.4851 | 2.9792 | 3.1574 |
| exponential | 4–3 | 1.3079 | 1.6735 | 2.0225 | 2.3822 | 2.6094 | 2.9352 | 3.0439 |
| F | 4–3 | 1.3275 | 1.7296 | 2.0469 | 2.5017 | 2.5833 | 2.8827 | 2.8917 |
| gamma | 4 | 1.1013 | 1.4190 | 1.7481 | 2.1626 | 2.4544 | 2.9155 | 3.0354 |
| Gumbel | 4 | 1.0923 | 1.3644 | 1.6721 | 2.0956 | 2.4114 | 2.9503 | 3.1279 |
| logistic | 4 | 1.1021 | 1.3335 | 1.5718 | 1.8812 | 2.1062 | 2.5627 | 2.7430 |
| normal | 3 | 1.4605 | 1.7518 | 2.0370 | 2.3893 | 2.6263 | 3.0551 | 3.2137 |
| Rayleigh | 3 | 1.4224 | 1.7242 | 2.0513 | 2.4885 | 2.7884 | 3.2620 | 3.3896 |
| t | 4 | 1.0047 | 1.2174 | 1.4390 | 1.7328 | 1.9497 | 2.3929 | 2.5867 |
| uniform | 1 | 3.7098 | 3.9664 | 4.1641 | 4.3798 | 4.5224 | 4.8167 | 4.9319 |
| Wald | 4–3 | 1.4397 | 1.8104 | 2.1459 | 2.4507 | 2.6628 | 3.0509 | 3.1233 |
| Weibull | 3 | 1.6073 | 1.9260 | 2.2391 | 2.6060 | 2.8581 | 3.3200 | 3.4886 |

Were we to use either the uniform or beta variates as the base of our model, the critical values would be too high for the other distributions. The $q = 0.90$ critical value for a uniform variate is 3.7098 and lies beyond the $q = 0.9995$ quantile of other variates. The beta $q = 0.90$ critical value passes at $q = 0.999$ for most of groups 3 and 4. Choosing the Weibull (2, 4) distribution for our peak model is a compromise between the ability to evaluate peaks and a guarantee that the test can

be applied to any data. Its $q = 0.95$ critical value, 1.9260, corresponds to the same quantile for five variates, including the normal and exponential, to $q = 0.975$ for three, and to $q = 0.99$ for $t$ and logistic draws.

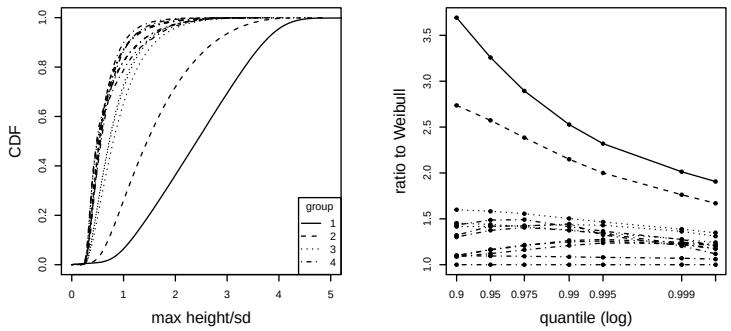



Figure 22: Distribution of standardized peak heights for different variates (left) and ratio of height quantiles to Weibull (right). Line styles are the same in both graphs. Groups are identified in the text.

The left side of Figure 22 shows the distribution function of the standardized peak heights for all variates. The right graph plots the ratio of the standardized height to that of the Weibull, for each column in the table above. The values converge at the highest quantiles, but for the third and fourth groups are relatively constant in the region of interest, to 0.99. If another base distribution is chosen, quantiles can be scaled by their ratio to the Weibull in Table 14.

### Detail 7 Peak Height Distribution by Filter

We run 100 thousand trials, drawing 200 points from a Weibull distribution with scale $a = 2$ shape $b = 4$, and calculating the spacing. Each draw is run through a low-pass filter with different kernel and width 30, or 15% of the draw. The quantiles of the standardized heights of all peaks generated in the trials are plotted on the left of Figure 23. The Kaiser filter gives the most conservative results. It generates the fewest and smallest peaks (Table 15) and smallest critical values (Table 16). The filters have the same order whether ranked by the number of peaks or their height. The Gaussian and Blackman filters create similar peaks.

Table 15: Peak Count per Filter Type

| | number of peaks | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | mean |
| Kaiser | 14383 | 50661 | 31000 | 3887 | 69 | | | 1.25 |
| Bartlett | 8196 | 42520 | 40008 | 8844 | 425 | 7 | | 1.51 |
| Hamming | 4760 | 33314 | 44795 | 15657 | 1445 | 29 | | 1.76 |
| Hanning | 3516 | 28555 | 45486 | 19945 | 2415 | 83 | | 1.89 |
| Blackman | 638 | 11492 | 37521 | 36531 | 12408 | 1362 | 47 | 2.53 |
| Gauss | 504 | 10070 | 35711 | 37806 | 14056 | 1785 | 67 | 2.60 |

The sensitivity of the filters is roughly constant, up to $q = 0.99$, as the right graph of Figure 23 shows. Beyond this quantile the curves converge and the scaling drops off toward unity. At $q = 0.95$

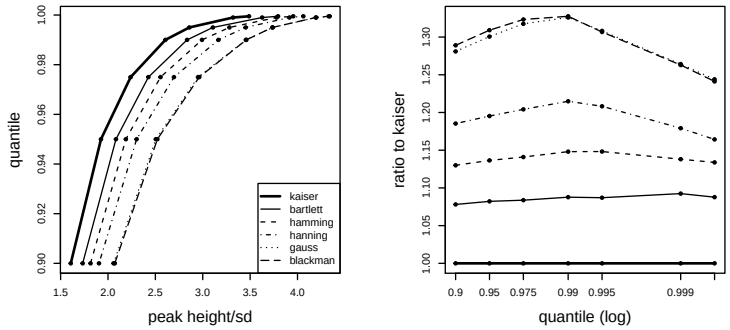



Figure 23: Quantiles of standardized peak heights for different filters with kernel window size 30 (15%) (left) from the base Weibull distribution, and ratio to Kaiser results (right). Line styles are the same in both graphs.

these scaling factors are 1.082 for the Bartlett filter, 1.136 for the Hamming, 1.195 for the Hanning, 1.301 for the Gauss, and 1.309 for the Blackman.

Table 16: Critical Peak Heights per Filter Type

| | quantile | | | | | | |
|---|---|---|---|---|---|---|---|
| | 0.90 | 0.95 | 0.975 | 0.99 | 0.995 | 0.999 | 0.9995 |
| Kaiser | 1.6073 | 1.9260 | 2.2391 | 2.6060 | 2.8581 | 3.3200 | 3.4886 |
| Bartlett | 1.7328 | 2.0843 | 2.4266 | 2.8346 | 3.1068 | 3.6270 | 3.7947 |
| Hamming | 1.8161 | 2.1888 | 2.5547 | 2.9920 | 3.2818 | 3.7784 | 3.9555 |
| Hanning | 1.9052 | 2.3021 | 2.6962 | 3.1660 | 3.4533 | 3.9150 | 4.0617 |
| Blackman | 2.0719 | 2.5214 | 2.9630 | 3.4597 | 3.7348 | 4.1929 | 4.3230 |
| Gauss | 2.0588 | 2.5050 | 2.9501 | 3.4555 | 3.7388 | 4.1971 | 4.3399 |

### Detail 8 Fitting the Peak Height Distribution

We draw 200 points from the base Weibull distribution in each of 50 thousand trials, pass them through a Kaiser filter with 0.15 window, and register the standardized height of local maxima to the adjacent minima. We then fit the CDF to ten distributions and measure the squared difference between the ideal and actual CDF. The left graph of Figure 24 shows the agreement for the four best fits, and the middle graph the density with the same fit parameters.

The right graph in Figure 24 shows the fit error is stable over many repetitions.

The parameters for the best fits in Table 17 are named according to the arguments of the distribution function in R or the package *extraDistr*. The Wald distribution, a.k.a. the inverse Gaussian, provides the best fit, with the log-normal a close second, rising a bit quicker and having a bit lower peak. The next-best fits, the Gumbel and gamma, do not match the tail of the density well. Although not presented here, the same results, including the ranking of the fits and the average parameter values, hold if we draw samples from a normal variate.

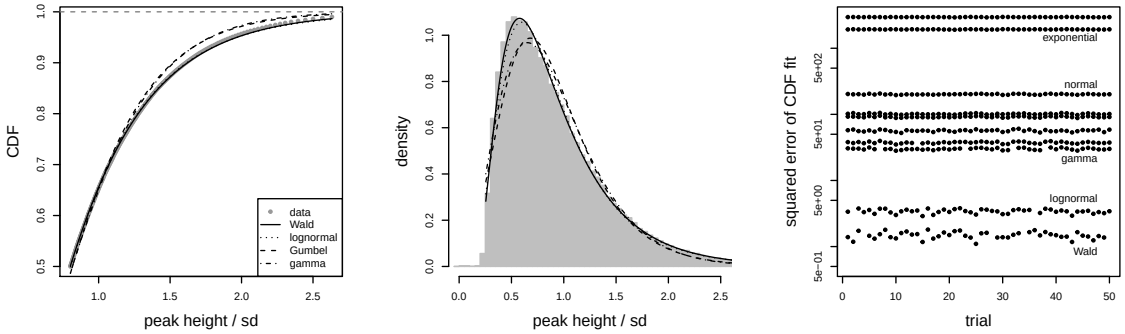



Figure 24: Best fits to distribution of standardized peak heights (left) and corresponding densities (center). Line styles are the same in both graphs. Fit accuracy is stable (right).

Table 17: Distribution Fits

| distribution | error | rank | parameters | | | |
|---|---|---|---|---|---|---|
| beta | 56.58 | 5 | shape1 $r$ | 2.994 | shape2 $s$ | 11.624 |
| chi-squared | 2956.84 | 10 | df | 1.458 | | |
| exponential | 1927.76 | 9 | rate $\lambda$ | 0.866 | | |
| gamma | 29.94 | 3 | shape $r$ | 3.672 | scale $1/\lambda$ | 0.245 |
| Gumbel | 37.42 | 4 | mu $\mu$ | 0.682 | sigma $\sigma$ | 0.373 |
| lognormal | 3.42 | 2 | meanlog $\log \mu$ | $-0.217$ | sdlog $\log \sigma$ | 0.545 |
| normal | 201.24 | 8 | mean $\mu$ | 0.845 | sd $\sigma$ | 0.432 |
| Rayleigh | 101.40 | 7 | sigma $\sigma$ | 0.706 | | |
| Wald | 1.48 | 1 | mu $\mu$ | 0.933 | lambda $\lambda$ | 2.826 |
| Weibull | 91.34 | 6 | shape $a$ | 2.106 | scale $b$ | 0.991 |

### Detail 9 Modeling the Wald Height Parameters

For a variety of draw sizes $n$ from the base Weibull distribution we find local peaks in the spacing after filtering with a Kaiser kernel of size $w = f_{lp}n$. We fit the Wald or inverse Gaussian CDF, with its two parameters, to the distribution of the standardized peak height, derived from one half to one million trials of draws. The Wald density and distribution functions are

$$f_{Wald}(x, \mu, \lambda) = \sqrt{\frac{\lambda}{2\pi x^3}} e^{-\lambda(x-\mu)^2/2\mu^2 x}$$
$$F_{wald}(x, \mu, \lambda) = \Phi\left(\frac{x}{\mu - 1}\sqrt{\lambda/\mu}\right) - e^{2\lambda/\mu}\Phi\left(\frac{x}{\mu + 1}\sqrt{\lambda/\mu}\right) \tag{21}$$

Each parameter forms a surface over $n$ and $f_{lp}$ (Figure 25) that can be modeled with two linear regressions. Against $n$ the fit has the form $b + m \log n$, where the logarithm is to base 10. The coefficients are themselves linear regressions, for the Wald $\mu$ against $\log f_{lp}$ and for $\lambda$ against $f_{lp}$. All together the model is

$$\mu = (5.8158 + 2.4152 \log f_{lp}) - (1.9704 - 1.0131 \log f_{lp}) \log n \tag{22}$$
$$\lambda = (-2.0204 + 49.7357 f_{lp}) + (2.6034 - 19.5195 f_{lp}) \log n \tag{23}$$

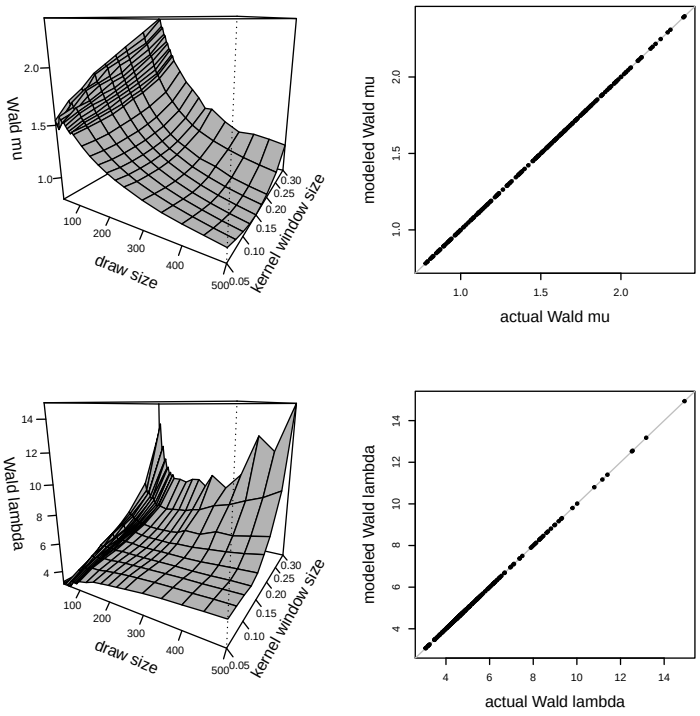



Figure 25: Wald parameters $\mu$ (top row) and $\lambda$ (bottom) fitting the peak heights. Surface plots (left) show parameters as draw and filter window size changes. Comparison plots (right) show fit values against modeled.

A model based on an absolute window size $w = f_{lp}n$ can be built, but it still needs terms with just $f_{lp}$ and $n$ to achieve a good fit, so little is gained. In other words, a model using only the absolute window size will not work.

The right plots in Figure 25 compare the actual values of $\mu$ and $\lambda$ against the modeled. The results fall on the 45 degree line for all combinations of $n$ and $f_{lp}$, indicating a good match.

### Detail 10 Accuracy of Peak Height Model

Figure 26 shows the accuracy of the critical value, the difference in the modeled $h_{peak}$ as a fraction of the actual height. Positive errors mean the model underestimates the height. The interquartile error lies within 5%, marked with solid horizontal lines, for a subset of the parameter space, namely kernel windows around $f_{lp} \approx 0.20$, draw sizes between 100 and 350 points, and the important quantile levels from 0.95 to 0.999. The errors remain within 10% for all window sizes, draws between 60 and 500 points, and quantiles up to 0.99995. The median relative error is less than 5% except for small draws and the very largest. Figure 27 shows the parameter combinations that produce more than 10% errors. These cluster on the edges of the simulation range.

In the main text Figure 4 plots the modeled quantiles against the actual. The area above and to

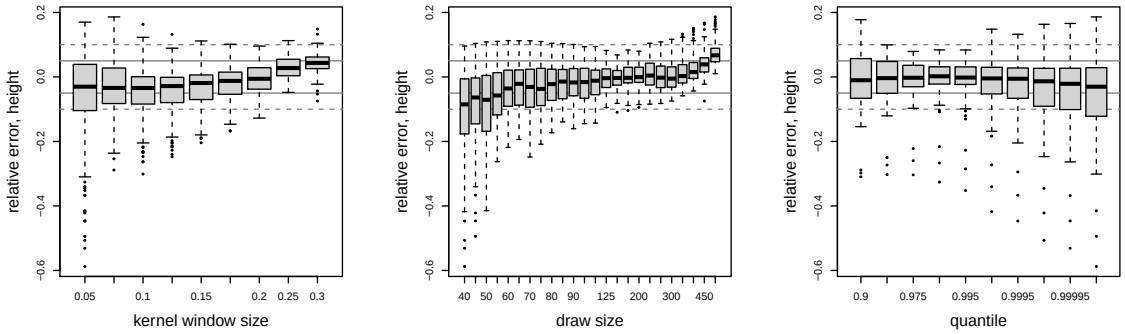



Figure 26: Error of modeled height critical value as fraction of actual, per model parameter. Horizontal lines mark 10% (dashed) and 5% (solid) error levels.

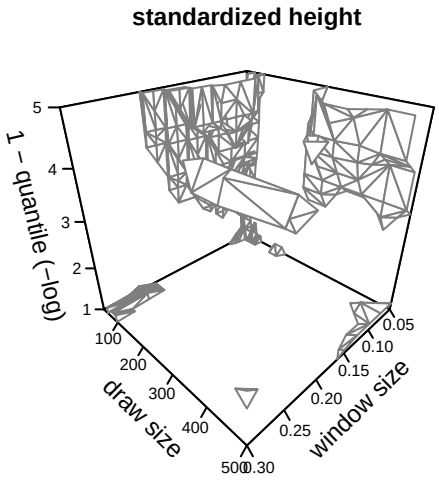



Figure 27: Boundary of the error in the modeled critical value of the standardized height relative to the actual. Model parameters inside the surface are within 10% of the actual.

the left of the ideal match, the 45° line, are the parameter combinations generating false positives, points whose modeled quantile is above the threshold when the actual is below. Below and to the right the combinations give false negatives. Table18 quantifies these counts. For example, 12.0% (26 of 216) combinations at the 0.90 quantile appear in the model to be significant at the 0.95 level, but there are no points at the 0.99 level. The errors are largest in adjacent buckets. The 0.99 level has more false negatives because the range of the error grows with the quantile, the model underestimating more combinations.

Table 18: Peak Model Mis-Classifications

| actual | vs. $q = 0.95$ | | | vs. $q = 0.99$ | | |
|---|---|---|---|---|---|---|
| 0.90 | FP | 12.0% | (26) | FP | 0.0% | |
| 0.95 | | | | FP | 0.0% | |
| 0.975 | FN | 1.4% | (3) | FP | 8.8% | (19) |
| 0.99 | FN | 0.5% | (1) | | | |
| 0.995 | FN | 0.0% | | FN | 9.3% | (20) |
| 0.9999 | FN | 0.0% | | FN | 1.4% | (3) |
| 0.9999+ | FN | 0.0% | | FN | 3.7% | (8) |

**Detail 11 Ideal Flats**

The expected spacing has a flat around the minimum value, in the center of a two-sided distribution or the start of a one-sided. We can compare the size of this section against flats created by draws from the distribution, holding three points in mind. First, we can analytically handle only exponential and logistic variates, using (3) and (4) in Section 1. Other distributions require numerically integrating the first moment of the density of the spacing [38, (2.7)], which is

$$E\Big\{D_i\Big\} = \frac{n!}{(i-2)!(n-i)!}\int_0^\infty y dy \int_{-\infty}^\infty \{F(x)\}^{i-2}\{1-F(x+y)\}^{n-i} f(x)f(x+y)dx \tag{24}$$

with $f(x)$ the variate's density and $F(x)$ its distribution. Second, the flat definition uses the ripple of the signal's range, which differs from the $\alpha$ scaling of the minimum spacing discussed in Section 1. If $D_{i,min}$ is the smallest spacing and $D_{i,max}$ the largest, then the flat will extend to the index $i$ where

$$\begin{aligned} D_{i,flat} &= D_{i,min} + \delta_{ripple}\left(D_{i,max} - D_{i,min}\right) \\ &= \delta_{ripple} D_{i,max} + \left(\delta_{ripple} - 1\right) D_{i,min} \end{aligned} \tag{25}$$

Here we use the total ripple above the minimum rather than centering it as is done in the detector. [Detail 12] This will not change the flat's interval but does place the source point to one side, between the minimum and maximum. Finally, although the expected spacing is already smooth and filtering is not needed, we still must account for the points that we would ignore that the kernel partially overlaps. The first valid index is $i_{lead} = (nf_{lp} + 1)/2$. The last is $i_{tail} = n + 1 - i_{lead}$.

In general the strategy to determine the ideal flat length is to get the minimum and maximum expected spacing, scale the difference by the ripple to get the flat height, and find the indices corresponding to this. For exponential and logistic variates we have simple expressions to do this, but for all others the work must be done numerically, by integration to get the total bounds and by root finding to get the interval.

For logistic draws the minimum spacing occurs at the middle index $i_{min} = i_{mid} = (n+1)/2$ and the maximum at $i_{max} = i_{lead}$ or $i_{tail}$. The flat height follows from (25) and (4) using these indices, giving

$$\begin{aligned} D_{i,flat,logis} &= \sigma n \frac{(1-\delta_{ripple})(i_{max}-1)(n-i_{max}+1) + \delta_{ripple}(i_{min}-1)(n-i_{min}+1)}{(i_{min}-1)(n-i_{min}+1)(i_{max}-1)(n-i_{max}+1)} \\ &= 4\sigma n \frac{(1-\delta_{ripple})(nf_{lp}-1)((2-f_{lp})n+1) + \delta_{ripple}(n-1)(n+1)}{(n-1)(n+1)(nf_{lp}-1)((2-f_{lp})n+1)} \end{aligned} \tag{26}$$

Inverting (4) gives the indices where the expected spacing has this value. By symmetry there are two on either side of the minimum,

$$i = \frac{n+2}{2} \pm \frac{1}{2}\sqrt{(n+2)^2 - 4(n+1+\sigma n/D_{i,flat,logis})} \tag{27}$$

The flat length is the difference between these, plus one to make the ends inclusive.

$$l_{flat,logis} = 1 + \sqrt{n^2 - 4\sigma n/D_{i,flat,logis}} \tag{28}$$

Exponential draws proceed similarly. The minimum spacing occurs at the first valid point, $i_{min} = i_{lead}$, and the maximum at $i_{max} = i_{tail}$. Using (3) to convert these to the expected spacings and substituting into (25), the flat height is now

$$D_{i,flat,exp} = \frac{2}{\lambda}\frac{(2nf_{lp}+1) + \delta_{ripple}(2-3f_{lp})n}{(2nf_{lp}+1)((2-f_{lp})n+1)} \tag{29}$$

The flat length stretches from $i = 1$ to the index corresponding to this spacing, or

$$l_{flat,exp} = n + 1 - \frac{1}{\lambda D_{i,flat,exp}} \tag{30}$$

Using the standard setup for many of the simulations, let the draw size $n = 200$, the filter kernel cover $f_{lp} = 0.15$ of it, and the ripple limit be $\delta_{ripple} = 0.05$. Table 19 shows that the flats cover a third of the draw. The lengths in (28) and (30) match the numeric results. Other distributions can produce even longer flats. Figure 28 shows the expected spacing for a Gumbel variate, an asymmetric distribution. The small points at the ends, which extend off the graph, are those partially covered by the filter and ignored. The maximum spacing at $i = 185$, 0.0652, is larger than the spacing at the smallest valid index $i = 16$, 0.0256. The minimum spacing, 0.0136, lies at $i = 73$. The flat height is thus $0.05(0.0652 - 0.0136) = 0.0026$ and points within this range of the minimum, from $i = 37$ through $i = 118$, are marked in the graph in black. The flat length is $118 - 37 + 1 = 82$.

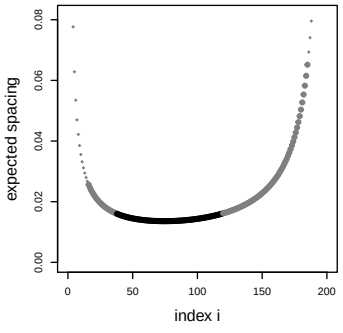



Figure 28: Expected spacing for a Gumbel variate, with ignored indices marked with small points and those within the flat in black.

Table 19: Ideal Flat Width

| | numeric | analytic |
|---|---|---|
| beta | 57 | |
| exponential | 64 | 63.2 |
| gamma | 91 | |
| Gumbel | 82 | |
| logistic | 67 | 67.6 |
| normal | 65 | |
| Rayleigh | 78 | |
| t | 71 | |
| Weibull | 65 | |

### Detail 12 Local Flats Detection Algorithm

Listing 2 provides pseudo-code for finding flats. The test parameters are the ripple $\delta_{ripple}$ as a fraction of the full data range and the minimum length requirements, $L_{abs}$ in absolute data points and $f_{L,rel}$ as a fraction of the data length. The return values include the start and endpoints of the flat, its length, and the standardized height. One could consider an aspect ratio, but this adds little with a height fixed by the ripple.

Step 2 of the algorithm defines the maximum ripple height and minimum flat length. Step 4 calculates the interval within the ripple specification, centered at each data point. $j$ is the second point on either side that fails, which allows for a single outlier. The endpoints are adjusted by one or two depending to get back within the ripple limit, depending on whether the second point also lies outside the limit. Because the scan starts at each point, if a flat covers a large fraction of the data the algorithm's performance will degrade towards $O(n^2)$. In step 5 we register the source point of the longest flat that overlaps each data point. This *src* array contains the longest flat plus the uncovered ends of overlapped intervals. If the signal rapidly moves away from the flat in these adjacent sections then they will be ignored as being too short, but if they are part of a long, gradual drift then the uncovered portion will be added as a separate flat. The algorithm reports the original length of these, not just the uncovered portion. The reported flats can therefore overlap. The number of times a source point appears in the array equals the length of the uncovered portion and must meet the minimum length.

| Listing 2: Local Flats Detector | |
|---|---|
| Input: | data $x$ of $n$ points, ripple $\delta_{ripple}$, length $L_{abs}$, fractional length $f_{L,rel}$ |
| 1: | calculate ripple specification $\Delta \leftarrow (\max(x) - \min(x))\delta_{ripple}/2$ |
| 2: | calculate length requirement $L \leftarrow \max(nf_{L,rel}, L_{abs})$ |
| 3: | create arrays $st[1 \dots n]$, $end[1 \dots n]$, $src[1 \dots n]$ |
| 4: | for $i$ in $1 \dots n$: |
| 4a: | $st[i] \leftarrow$ second $j$ in $i \dots 1$ with $x[j] < x[i] - \Delta$ or $x[i] + \Delta < x[j]$ |
| 4b: | $end[i] \leftarrow$ second $j$ in $i \dots n$ with $x[j] < x[i] - \Delta$ or $x[i] + \Delta < x[j]$ |
| | note: back $st[i]$, $end[i]$ to last index inside ripple bounds |
| 5: | for $i$ in $1 \dots n$: |
| 5a: | $src[i] \leftarrow argmax_{j \text{ in } 1 \dots n}(end[j] - st[j] + 1)$ |
| 6: | for $s$ in unique($src$): |
| 6a: | $l_s \leftarrow \#\{src = s\}$ |
| 6b: | if $L \le l_s$: |
| 6b1: | $x_{min} \leftarrow \min(x[st[s] \dots end[s]])$ |
| 6b2: | $x_{max} \leftarrow \max(x[st[s] \dots end[s]])$ |
| 6b3: | length $l_{flat} \leftarrow end[s] - st[s] + 1$ |
| 6b4: | height $h_{flat} \leftarrow (x_{max} - x_{min})/\sigma(x)$ |
| 6b5: | report flat at $s$ between $st[s]$, $end[s]$, with $l_{flat}$, $h_{flat}$ |

### Detail 13 Flat Length Distribution by Variate

We run 100 thousand trials of draws of 200 points of different random variates, passing the spacing through a low-pass Kaiser filter of width 30 (15% window) and identifying flats per the feature detector. [Detail 12] All distributions use default parameters, with zero location and unit scale, except the Weibull draws with scale $a = 2$ and shape $b = 4$, the gamma with shape $r = 2$ and rate $\lambda = 4$, the beta with shapes $r = 2$ and $s = 2$, the F with 10 and 6 degrees of freedom, and the

Rayleigh with scale $\sigma = 10$. Although the null distribution will include flats of any length above the minimum, the shortest are too numerous to be interesting, and we ignore those spanning fewer than 5 points. Table 20 counts the trials generating flats of minimum length 15, and Table 21 those of length 30. Table 20 omits counts of 8 or 9 flats to fit the page. The variates are placed in the table in the order of the critical values, not by the feature count.

Table 20: Flat Count per Base Distribution

| | | number of flats, $l \geq 15$ | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | group | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | mean |
| uniform | 3 | 49683 | 33980 | 12613 | 3060 | 565 | 89 | 10 | | 0.71 |
| beta | 3 | 14774 | 28538 | 27887 | 18002 | 7971 | 2350 | 422 | 53 | 1.85 |
| Weibull | 2 | 496 | 3652 | 13226 | 27501 | 30534 | 18476 | 5357 | 699 | 3.65 |
| normal | 2 | 187 | 1994 | 9612 | 24995 | 33589 | 21917 | 6724 | 922 | 3.87 |
| Rayleigh | 2 | 84 | 766 | 4914 | 17677 | 32280 | 29297 | 12414 | 2342 | 4.32 |
| logistic | 2 | 7 | 802 | 7789 | 26213 | 36153 | 22359 | 5970 | 672 | 3.92 |
| t | 1 | 1 | 1910 | 11943 | 30406 | 33707 | 17423 | 4103 | 477 | 3.67 |
| Gumbel | 1 | | 1295 | 10165 | 27028 | 33649 | 20786 | 6153 | 864 | 3.85 |
| chi-squared | 1 | | 977 | 9554 | 25549 | 33580 | 21796 | 7319 | 1139 | 3.93 |
| gamma | 1 | | 1405 | 12912 | 28156 | 32008 | 18972 | 5672 | 819 | 3.75 |
| exponential | 4 | | 1398 | 25827 | 42677 | 22657 | 6346 | 1005 | 86 | 3.10 |
| Wald | 4 | | 9921 | 53282 | 30008 | 5981 | 742 | 64 | 2 | 2.35 |
| F | 4 | | 37015 | 50218 | 11384 | 1284 | 95 | 6 | | 1.77 |

Table 21: Long Flat Count per Base Distribution

| | | number of flats, $l \geq 30$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | group | 0 | 1 | 2 | 3 | 4 | 5 | 6 | mean |
| uniform | 3 | 98654 | 1302 | 42 | 2 | | | | 0.01 |
| beta | 3 | 88644 | 10475 | 843 | 38 | | | | 0.12 |
| Weibull | 2 | 50155 | 37796 | 10652 | 1306 | 87 | 4 | | 0.63 |
| normal | 2 | 39925 | 42125 | 15524 | 2285 | 137 | 4 | | 0.81 |
| Rayleigh | 2 | 24045 | 41281 | 26814 | 7094 | 718 | 48 | | 1.19 |
| logistic | 2 | 14946 | 42887 | 33181 | 8316 | 649 | 19 | 2 | 1.37 |
| t | 1 | 6688 | 36926 | 41612 | 13400 | 1313 | 59 | 2 | 1.66 |
| Gumbel | 1 | 4609 | 30560 | 43711 | 18506 | 2468 | 143 | 3 | 1.84 |
| chi-squared | 1 | 3317 | 26624 | 44657 | 21757 | 3442 | 193 | 10 | 1.96 |
| gamma | 1 | 1751 | 23465 | 46410 | 24127 | 3997 | 243 | 7 | 2.06 |
| exponential | 4 | 57 | 15994 | 54585 | 25169 | 3879 | 305 | 10 | 2.18 |
| Wald | 4 | | 34415 | 53248 | 11200 | 1077 | 56 | 4 | 1.79 |
| F | 4 | | 59807 | 35164 | 4673 | 340 | 14 | 2 | 1.46 |

The variates that generate peaks do not create many flats. The top five distributions by average peak count [Detail 6] are the bottom five by longest flat: uniform, beta, Weibull, Rayleigh, and normal. The converse does not hold, as the top five by longest length are not at the bottom of the peak list, although they are in the bottom half. Naturally there are fewer long flats, but the distributions vary in their ability to produce them, and the ranking changes. For example, the Rayleigh distribution generates the most shorter flats, but ranks ninth for longer features. The base distributions for peaks generate fewer long flats, so they cannot be used to evaluate flats without over-estimating their rarity.

As we did with the peaks, we can group the variates by the distribution of the flat length (Figure 29). There are again four different behaviors, but not with the same split as we made for the

peaks. The tables give the group assignments. Group 4 contains the exceptionally active variates, with distribution curves in the graphs far to the right. Their critical values at the 0.05 significance level would pass at 0.001 for the other groups. Group 3 has the smallest critical values, and is too insensitive in the other direction. The features they find significant at the 0.01 level would fail at 0.10 for any other variate. We see a gap between the sets of curves for the other two groups, and a larger spread or less consistency within each group than we have with the peaks. Group 1, with larger critical values, is more conservative than Group 2. Although the individual ranks in the flat count tables are scrambled compared to the peak heights, as groups they are more consistent. Flat group 1 matches peak group 4, flat group 2 corresponds to peak group 3, and the transitional peak group 4–3 maps to the outlier flat group 4.

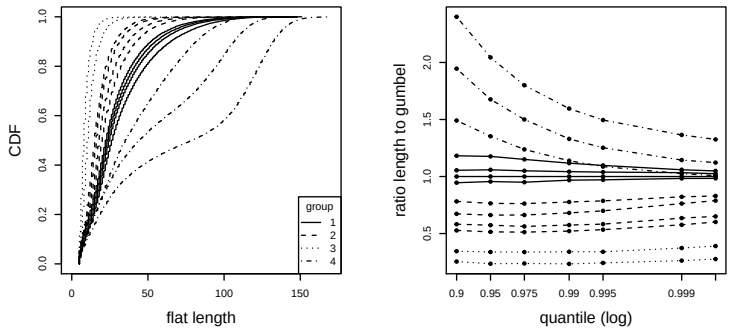



Figure 29: Distribution of flat length for different uni-modal variates (left) and ratio of length quantiles to Gumbel (right). Line styles are the same in both graphs. Groups are identified in the text.

Figure 29 also plots the quantiles of each value as a fraction of the Gumbel variate. Except for group 4, the ratios are stable, which would allow us to use scaling of the critical values to convert between base distributions. The spread between the ratios is larger than for the peaks.

Critical values for the flat length are given in Table 22.

Table 22: Critical Flat Lengths per Base Distribution

| | | quantile | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | group | 0.90 | 0.95 | 0.975 | 0.99 | 0.995 | 0.999 | 0.9995 |
| uniform | 3 | 14 | 16 | 19 | 22 | 25 | 31 | 34 |
| beta | 3 | 19 | 23 | 27 | 32 | 35 | 44 | 48 |
| Weibull | 2 | 29 | 35 | 41 | 49 | 55 | 68 | 74 |
| normal | 2 | 32 | 39 | 45 | 54 | 60 | 75 | 80 |
| Rayleigh | 2 | 37 | 45 | 53 | 64 | 72 | 90 | 97 |
| logistic | 2 | 43 | 52 | 61 | 73 | 81 | 97 | 102 |
| t | 1 | 52 | 65 | 76 | 91 | 100 | 116 | 121 |
| Gumbel | 1 | 55 | 68 | 80 | 94 | 103 | 118 | 123 |
| chi-squared | 1 | 58 | 72 | 84 | 98 | 107 | 122 | 126 |
| gamma | 1 | 65 | 80 | 92 | 105 | 113 | 125 | 129 |
| exponential | 4 | 82 | 92 | 99 | 107 | 112 | 121 | 124 |
| Wald | 4 | 107 | 114 | 120 | 125 | 129 | 135 | 138 |
| F | 4 | 132 | 139 | 144 | 150 | 154 | 161 | 163 |

The distributions in Figure 29 do not group as tightly as they do for the peak height (Figure 22), which makes choosing a base variate for testing difficult. Critical values of the flat length are less stable. To quantify this, count in which quantile bin each $q = 0.99$ critical value would fall. For example, this length is 64 for Rayleigh draws, which would be at $q = 0.975$ in logistic and $q = 0.995$ in normal or Weibull variates. Summing these over all thirteen distributions indicates how the critical values disperse (Table 23). The peak heights are stable, with most 0.99 critical values mapping remaining in the 0.975 – 0.995 range. The high counts in the first and last column come from the outlier distributions, the uniform and beta variates. For flats almost all critical values map to these outer columns. They change significantly between distributions.

Table 23: Stability of Flat Models

| | $\leq 0.90$ | 0.90 | 0.95 | 0.975 | 0.99 | 0.995 | 0.999 | 0.9995 |
|---|---|---|---|---|---|---|---|---|
| peak height | 22 | 5 | 14 | 37 | 27 | 23 | 4 | 24 |
| flat length | 50 | 11 | 8 | 9 | 6 | 11 | 7 | 54 |

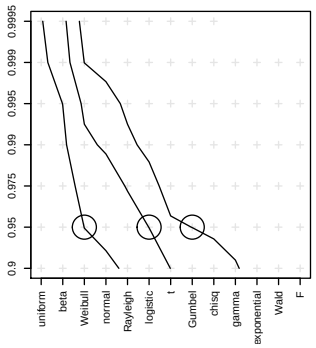



Figure 30: Length quantiles corresponding to the 95% critical for Weibull (left), logistic (middle), and Gumbel (right) variates.

Figure 30 tries to depict this variation for the critical values of the Weibull, logistic, and Gumbel variates. Their $q = 0.95$ critical values are marked at the appropriate quantile for each other draw. The curves are steep, showing that the quantiles are not stable and that one variate cannot stand in for another, as we were able to do with the peak height. Were we to generate an equivalent graph for the peak height, the curves would be horizontal, with a slope approximately the reciprocal of that for the flats.

The ranking of the variates by critical value has the same order as seen in ideal flat length [Detail 11], with two exceptions: the logistic and Rayleigh variates swap, and the exponential ideal length is short compared to the other distributions while its critical length is longer. The ideal lengths map to consistent quantiles, to $q = 0.95$ – $0.975$ for Group 1 (gamma, Gumbel, t), to $q = 0.995$ for Group 2 (Weibull, normal, Rayleigh) except $q = 0.975$ for logistic draws, to $q = 0.9995$ for Group 3 (beta), and to $q \leq 0.90$ for Group 4 (exponential). The quantiles follow the order of the curves in Figure 29, increasing towards the upper left corner of the graph.

**Detail 14 Flat Length Distribution by Filter**

We run 100 thousand trials, drawing 200 points from a Gumbel distribution with $\mu = 0$ and $\sigma = 1$ for each and passing the spacing through six different low-pass filters of width 30 (15%). The underlying data is the same for each trial, so we can see the effect of only the kernel change. The filtered spacings are then passed to the flat detector, ignoring flats less than 5 points in length. Table 24 counts the number of flats with length at least 15 and Table 25 at least 30. The equality in the counts in these tables for the Kaiser row and those in [Detail 13] for the Gumbel is no coincidence, for the same random seed was used before the draws.

Table 24: Flat Count per Filter Type

| | number of flats, $l \geq 15$ | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | mean |
| Kaiser | 0 | 1295 | 10165 | 27028 | 33649 | 20786 | 6153 | 864 | 57 | 3.85 |
| Bartlett | 5 | 660 | 6156 | 21108 | 33108 | 26236 | 10472 | 2064 | 182 | 4.19 |
| Hamming | 17 | 662 | 5632 | 19824 | 32702 | 27120 | 11432 | 2370 | 232 | 4.25 |
| Hanning | 82 | 1055 | 6430 | 19893 | 31733 | 26508 | 11603 | 2431 | 251 | 4.22 |
| Blackman | 465 | 3131 | 11166 | 23140 | 29255 | 21555 | 8986 | 2026 | 264 | 3.90 |
| Gauss | 602 | 3944 | 12375 | 24444 | 28515 | 20116 | 7975 | 1816 | 205 | 3.79 |

Table 25: Long Flat Count per Filter Type

| | number of flats, $l \geq 30$ | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | mean |
| Kaiser | 4609 | 30560 | 43711 | 18506 | 2468 | 143 | 3 | 1.84 |
| Bartlett | 9194 | 34849 | 38931 | 15010 | 1925 | 88 | 3 | 1.66 |
| Hamming | 13400 | 37883 | 35282 | 12011 | 1366 | 55 | 3 | 1.50 |
| Hanning | 21199 | 40328 | 29277 | 8368 | 787 | 41 | | 1.27 |
| Blackman | 36418 | 39193 | 19378 | 4506 | 488 | 17 | | 0.94 |
| Gauss | 40905 | 38242 | 16791 | 3685 | 363 | 15 | | 0.84 |

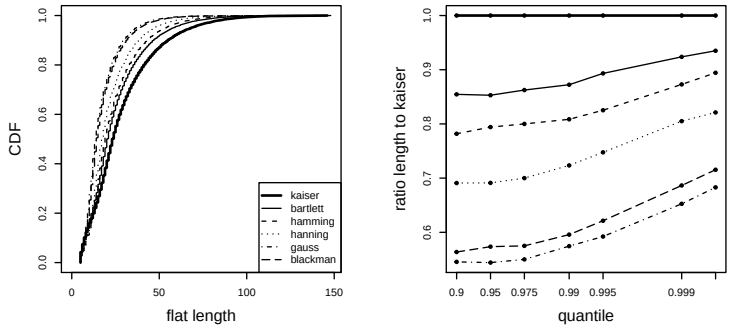



Figure 31: Distribution of flat length for different low-pass filters applied to uni-modal Gumbel (left) and ratio of quantiles to the Kaiser result (right). Line styles are the same in both graphs.

Figure 31 plots the distribution of the flat length for each filter, and Table 26 gives the values. They form a consistent set. The Kaiser kernel is most conservative, producing the fewest flats and the largest critical values at a given significance level. The Bartlett and Hamming filters are roughly the same, as are the Gauss and Blackman; the Hannning filter falls in-between. This agrees with the number of flats generated. The ratio of the critical values at each significance level to that of the Kaiser filter is not stable, so separate models will be needed for each kernel. However, the critical

values are consistent, with the 5% level for the Kaiser passing at the 1%–5% level for all but the Gaussian. In other words, a flat generated by another filter will appear somewhat less significant in the Kaiser model, but the difference is small enough that the Kaiser filter is a good choice. As with the variates, the filters that generate the most peaks, the Gauss and Blackman, have the fewest flats. The $l \geq 15$ counts do not reflect the trends in the $l \geq 30$ counts or the critical values, and this may indicate some instability in the repeatability of the results. Those flats may be too short.

Table 26: Critical Flat Lengths per Filter Type

| | quantile | | | | | | |
|---|---|---|---|---|---|---|---|
| | 0.90 | 0.95 | 0.975 | 0.99 | 0.995 | 0.999 | 0.9995 |
| Kaiser | 55 | 68 | 80 | 94 | 103 | 118 | 123 |
| Bartlett | 47 | 58 | 69 | 82 | 92 | 109 | 115 |
| Hamming | 43 | 54 | 64 | 76 | 85 | 103 | 110 |
| Hanning | 38 | 47 | 56 | 68 | 77 | 95 | 101 |
| Blackman | 31 | 39 | 46 | 56 | 64 | 81 | 88 |
| Gauss | 30 | 37 | 44 | 54 | 61 | 77 | 84 |

### Detail 15 Accuracy of Flat Models

The relative error of the flat models is the difference between the predicted critical value and the actual as a fraction of the actual. It is positive if the model underestimates the value. We vary the draw size from 50 to 500 in steps of 50 and the fractional kernel window size from 0.05 to 0.40 in steps of 0.05. Quantiles are measured from 0.9 to 0.99999, including the half values starting at 0.95, and 0.975. This is two more columns of critical values than used to develop the models. [Detail 14] The relative error is within 5% over the interquartile range for most of the model parameter space. The surfaces in Figure 32 mark the 5% and 10% errors over the parameter test space. Such errors are limited to the edges of the box; the central volume has a smaller error. The model is inaccurate for sections along the minimum kernel width, for combinations of small kernel and draw size at quantiles below 0.95, and for the smallest kernel and draw values for all quantiles. At 10% three surfaces remain along the back pillar over all quantiles, the combination of moderate draws and largest filter width, and small draws and kernels.

Figure 33 shows the range of error for each model parameter separately, allowing the other two to change. The interquartile range is mostly within the 5% limit (solid horizontal lines). The model occasionally runs out to 10%, notably for the largest kernels and smallest draws.

Figure 6 in the main text plots the modeled quantiles against the actual. The area above and to the left of the ideal match are parameter combinations generating false positives, modeled quantiles that pass the test level but actually have a smaller quantile. Those below and to the right are false negatives. Table 27 tallies the mis-classifications for the key significance levels. The 0.95 level has only one false positive, or in other words there is one combination with an actual quantile of 0.90 that passes the 0.95 critical value. The model considers eight combinations at the 0.99 quantile to be false negatives, accepting them at the 0.995 level.

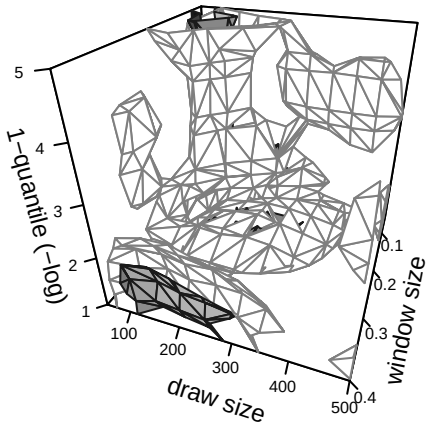



Figure 32: Boundary of the model error. Model parameters inside the surface are within 5% (light) and 10% (dark) of the actual.

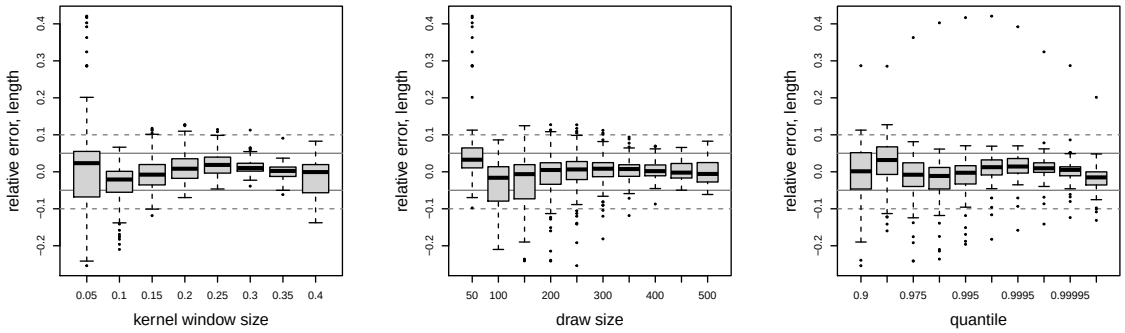



Figure 33: Error of model critical value as fraction of actual, per model parameter. Bars mark the median, boxes the inter-quartile range, and whiskers extend the error 50% further.

Table 27: Flat Model Mis-Classifications

| | Height | | | | | |
|---|---|---|---|---|---|---|
| actual | vs. $q = 0.95$ | | | vs. $q = 0.99$ | | |
| 0.90 | FP | 1.3% | (1) | FP | 0.0% | |
| 0.95 | | | | FP | 1.3% | (1) |
| 0.975 | FN | 0.0% | | FP | 1.3% | (1) |
| 0.99 | FN | 0.0% | | | | |
| 0.995 | FN | 0.0% | | FN | 10.0% | (8) |
| 0.9999 | FN | 0.0% | | FN | 0.0% | |

### Detail 16 Kernel Density Estimation

Figure 34 shows the low-pass filtered spacing is equivalent to a kernel density estimate of the data, here for the GNN example. The kernel density estimate was made with the R density() function, with a triangular kernel and bandwidth determined by the SJ method. The filter also uses a Bartlett or triangular kernel with window width 0.075, or 23 points, chosen to match the curves. The density calculation is done by transforming into the Fourier domain before performing the convolution, while the low-pass filter operates directly on the data. To convert the density estimate to spacing, invert its value after scaling by the number of points to place them at quantiles:

$$x'_{kde}[i] = n \sum_{j=1}^{i} y_{kde}[j] / \sum_{k=1}^{n} y_{kde}[k]$$
$$y'_{kde}[i] = 1/n \cdot y_{kde}[i]$$

The kernel density shows a more ringing than the low-pass filter, especially for indices beyond 250, and does not reach the central flat. Increasing its bandwidth would dampen the ringing, improving the match in the background at the right side of the graph, but would also remove the flat and the peak at index 102. Decreasing the bandwidth would better fit these two features but would create many false peaks in the background region. The same comments apply when using a Gaussian kernel: the kernel density estimate is close to the filter, but differs in details.

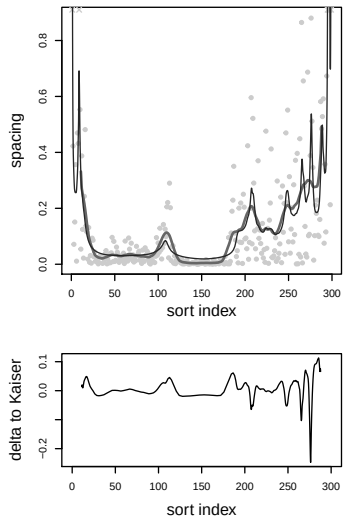



Figure 34: Low-pass filtered spacing (light curve) and kernel density equivalent (dark, thin).

### Detail 17 Varying Kernel Width

The graphs in Figure 35 show two possibilities of adapting the mode tree to each of our features. A direct application of the idea produces the graph for peaks on the left. Local minima or anti-modes

are drawn as dashed lines, and local maxima as solid lines shaded by the probability of the feature per the tests in this article. For flats we instead draw a line along the length of the feature, again shaded by its probability. In both graphs the filter window size increases along the $y$ axis, shown in absolute units on the left axis and as the fraction $f_{lp}$ on the right. We use as data the PPP example, as filter the Kaiser kernel, and for tests the Weibull critical values for peaks and logistic base distribution for flats.

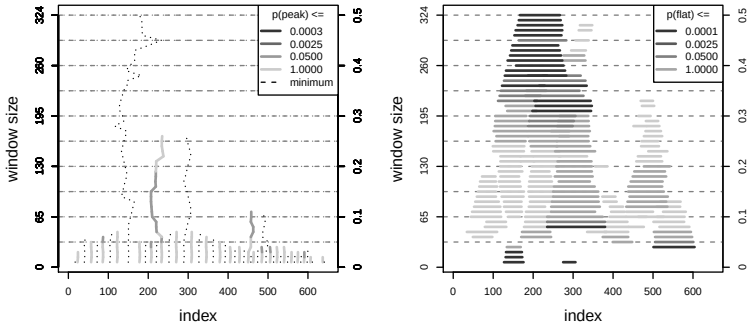



Figure 35: Peaks (left) and flats (right) in the PPP example as the filter window changes.

The left chart shows one stable peak near index 210 at a significance level of 0.05 or better that persists to $f_{lp}$ 25%. It is bounded on either side by a stable minimum. The second peak near index 450 appears in the 10% window at the 0.05 level. These positions agree with the sizes of the first two modes in the draw. Windows smaller than 5% of the data are rough and produce many weak, false peaks. This behavior seems to apply in general and is the main reason for limiting the model to larger windows.

A 50% window smooths the data into a rising curve that is only flat between the first two modes, but this feature is significant at the 0.0001 level. The flat grows as the window shrinks, and the effect of ignoring data partly covered by the filter is clear as the left edge shifts outward, eventually exposing the first mode. At $f_{lp} = 0.30$ the flat begins to divide into the first two modes and a third flat appears in the last. Flats in the first two modes continue to dividie as the filter becomes smaller, although the split is not clear: there is a small flat that appears in the peak at index 210 and obscures the two. It is easier to follow the flats at the minima in the spacing. The flat in the third mode also becomes significant with a clear gap at the peak. Below $f_{lp} = 0.05$ the flats disappear, leaving one spur in the first mode.

### Detail 18 Modeling the Slope Test Critical Value

The critical value for the slope test in Section 3 requires calculating the test statistic on many trials of uniform draws, typically 100 thousand in the reference implementation. We can instead use a model generated by varying the data size $n$ and significance level $\alpha$. We collect the critical value with and without correction for one million trials, varying $n$ between 50 and 5000 in 41 steps of increasing size. We sort the results and take quantiles at $\alpha$ 1, 2.5, and 5 times powers of ten between $-5$ and 0, and at $1 - \alpha$.

Fitting the surfaces proceeds by trial-and-error; there is no theoretical basis for the terms chosen.

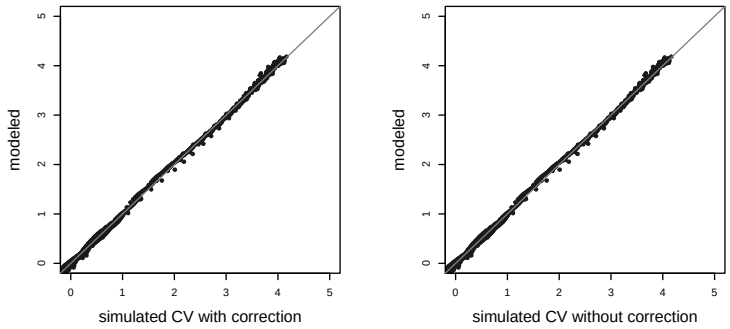



Figure 36: Modeled critical values plotted against simulated for the corrected slope test statistic critical value (left) and uncorrected (right).

Regression gives

$$\begin{aligned} cv_{corrected} = \quad & 1.45638 + 3.05017 \times 10^{-5} n - 0.91411\alpha + 2.09873 \times 10^{-4} \alpha n \\ & -4.43401 \times 10^{-8} \alpha n^2 + 1.27914 \times 10^{-8} (\alpha n)^2 - 29.1864/n \\ & -0.51588 * \log(\alpha) + 0.16782 * \log(1-\alpha) \end{aligned} \tag{31}$$

$$\begin{aligned} cv_{uncorrected} = \quad & 3.87589 + 1.14578 \times 10^{-4} n - 0.74516\alpha + 3.49072 \times 10^{-4} \alpha n \\ & -6.46512 \times 10^{-8} \alpha n^2 + 5.81907 \times 10^{-9} (\alpha n)^2 - (1.0185/n \\ & -0.43067 * \log(\alpha) + 0.12376 * \log(1-\alpha) \end{aligned} \tag{32}$$

Figure 36 plots the simulated values against the modeled. There is still some regular error visible, an interaction between $n$ and $\alpha$ that has no clear pattern, but the overall fit is good. The residual standard error of the regression for the corrected CV is 0.04525 and for the uncorrected 0.06232. The residuals are not normally distributed in the tails; the models tend to over-estimate the critical value.

### Detail 19 Level Section Algorithm

The level section detector in Listing 3 takes as an input an upper diagonal matrix $ttij$ with the slope test results. Rows correspond to the interval start points $j$ and columns to the end points $k$, with $j < k$. Values are $-1$ if $T_{ij} < -cv$, $+1$ if $T_{ij} > cv$, and 0 otherwise, meaning the interval holds no significant slope. Let $n$ be the number of data points. Use 1-based indexing for arrays.

Step 1 prepares two vectors, $firstk$ that holds the leftmost sloping endpoint for each starting index, and *slope* with the sign of the slope. Steps 2 and 3 define the first non-sloping triangle on the diagonal. Step 4b searches inside and moves the endpoint inward and left to the earliest sloping interval. The next section will begin above this triangle in step 4d, with step 4e moving the start point downward to the first row with an opposite slope. This allows level sections to overlap.

```
Listing 3: Level Section Detector
Input:  thresholded slope test statistic ttij of size n × n
1       let firstk, slope be vectors of length n
1a          for j in 1...n:
1a1             k ← j + 1
1a2             while k ≤ n and ttij[j, k] = 0 :
                    k ← k + 1
1a3             firstk[j] ← k
1a4             slope[j] ← ttij[j, k]
1b          firstk[n] ← 0
2       kend ← firstk[1]
3       jst ← 1
4       while kend < n:
4a          jend ← jst
4b          for j in jst...firstk[jst]:
4b1             if firstk[j] < kend:
                    jend ← j
                    kend ← firstk[j]
4c          register level section jst, kend − 1
4d          jst ← kend
4e          if jst < n and slope[jend] ≠ slope[kend]:
4e1             while 1 < jst and slope[jst − 1] = slope[kend]:
                    jst ← jst − 1
5       return registered levels
```

**Detail 20 Normalized Spacing**

This can also be done for logistic variates, multiplying the spacing by $(i-1)(n-i+1)$ to compensate for both tails. The drop-off in the tails of other distributions differs from this and the spacing is not quadratic in $i$. There is no generic correction for the growth of the tails.

**Detail 21 Joint Density from Pyke**

Substitute into [38, (2.4)] $k_1 = i - w$, $t_1 = x$, $k_2 = i$, $t_2 = x + y$, $k_3 = n + 1$, $t_3 = \infty$, $t_0 = -\infty$.

**Detail 22 Interval Spacing vs. Low-Pass Filter**

The interval spacing is rougher than a Kaiser filter over the same data (Figure 37). The ratio of the widths that match the results is the ratio of the equivalent noise bandwidth of the filters [17, Table 1], where we take the Kaiser time-bandwidth product to be $\alpha = 3.4\pi$. The figures center the interval spacing curves by shifting them by half the interval.

**Detail 23 Low-Pass Filtering and Markov Order**

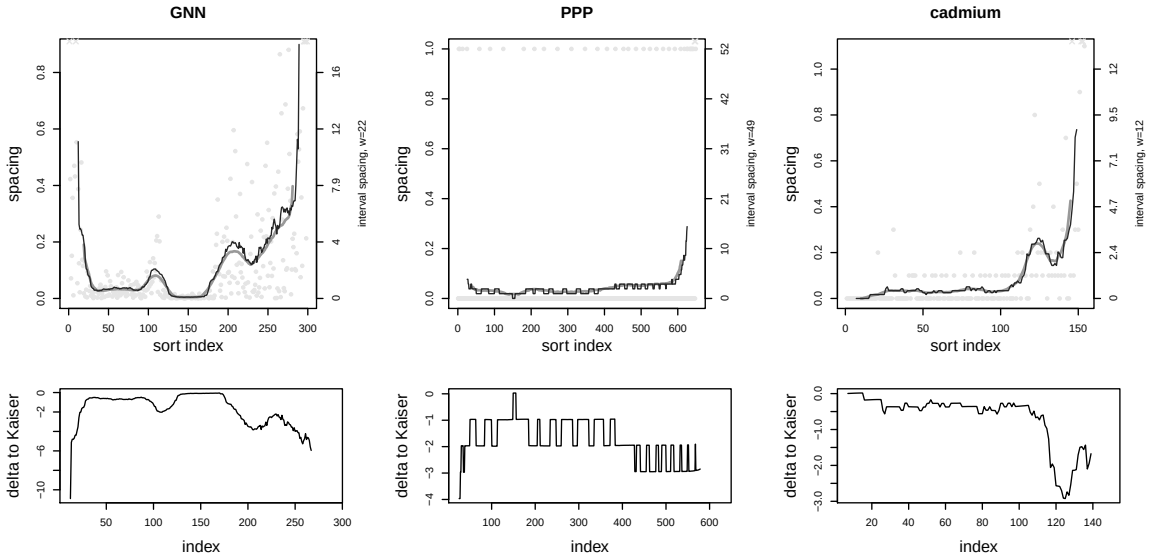



Figure 37: Comparing the low-pass Kaiser filter with 0.125 window (light curve) to the interval spacing with 0.075 window (dark). Raw data drawn as points.

Using the efficient determination criterion (EDC) metric to judge the order of a Markov chain [11], we find low-pass filtering can increase the interdependence of the data (Table 28). The order of the GNN example, base tri-modal variant, and log of the quake depth increases at a Kaiser kernel spanning 10% of the data, the quake data at 12.5%, and the bi-modal base variant at 5%. Below the 10% width the order does not show consistent behavior. The order of the interval spacing is lower, only increasing for the GNN example and inconsistently for the log earthquake depths.

Table 28: Order of Markov Chains

| | | Low-Pass Filter | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| window | raw | 2.5% | 5% | 7.5% | 10% | 12.5% | 15% | 17.5% | 20% |
| GNN | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | 2 |
| PPP | 1 | 2 | 2 | 1 | 1 | 1 | 1 | 1 | 1 |
| cadmium | 1 | 2 | 1 | 1 | 1 | 1 | 1 | 2 | 2 |
| quake | 1 | 1 | 1 | 1 | 1 | 2 | 3 | 3 | 2 |
| log quake | 1 | 1 | 2 | 1 | 2 | 2 | 2 | 2 | 2 |
| bi-modal | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 3 | 2 |
| tri-modal | 1 | 2 | 1 | 1 | 2 | 2 | 2 | 2 | 2 |
| | | | | | Interval Spacing | | | | |
| GNN | | 1 | 1 | 2 | 2 | 2 | 2 | 2 | 2 |
| PPP | | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| cadium | | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| quake | | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| log quake | | 1 | 2 | 1 | 2 | 2 | 1 | 1 | 2 |
| bi-modal | | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| tri-modal | | 1 | 1 | 1 | 1 | 1 | 2 | 1 | 1 |

### Detail 24 Excursion Heights

The bounds of features in the signal, be they the minima to either side of a peak or the ends of a flat, define the subset of the data that we use for the permutation and excursion tests. The feature will have an offset $H_{sf}$ between its starting and ending points independent of any shuffling by permutation, and an overall range $H_{rng}$. Shuffling and sampling may introduce new extrema not present in the original feature, or may shift the position of the peak. For a draw pool we can use either the difference between adjacent spacings or the quantized, signed difference version with values $+1$, $-1$, or 0 for ties. Such quantization may distort the signal, amplifying differences smaller than the average step and shrinking those larger. Figure 38 shows two examples. The sharp peak to the left side of the first graph has been reduced because the raw signal contains a few relatively large steps that are compressed by the uniform sign. Smaller peaks in the middle and right are amplified because the raw signal contains many small, similar steps. In the second graph, the first large decrease at index 83 shrinks and the side lobe merges into and actually becomes the peak, delaying the falling edge. The start-to-finish height in both examples has also changed.

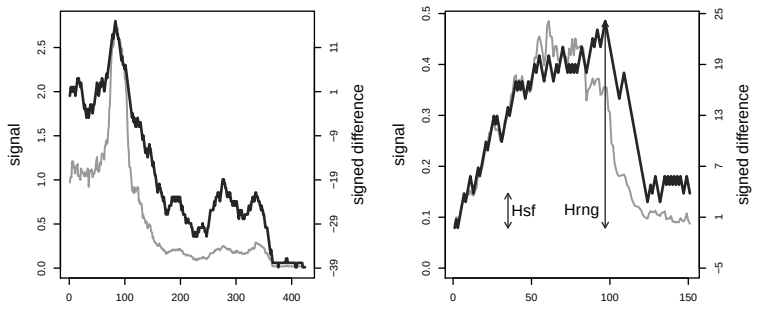



Figure 38: Using signed differences (dark curve) can distort features in the raw signal (light). Curves have been scaled to have the same range.

An analogy may be drawn with the various Brownian trajectories. In general this theory will not apply because the distribution of the steps will not be normal; indeed, for spacing it cannot be. Still, if the feature is fairly symmetrical and has a minimal start-to-finish height and no subsidiary peaks, the distribution of the permutations is close to a Brownian excursion [8, (4.5)] As an example, consider the distribution of heights from 100k permutations of the two significant peaks found in a modified version of the N5 literature sample:

$$\begin{array}{ccc} 50 \times \mathrm{N}(-2,\, 0.1) & 50 \times \mathrm{N}(-0.75,\, 0.1) & 500 \times \mathrm{N}(0,\, 3) \\ 50 \times \mathrm{N}(0.75,\, 0.1) & 50 \times \mathrm{N}(2,\, 0.1) & \end{array}$$

There are four modes plus a wide background (Figure 39). We use an interval of 25 to generate the spacing, and use runs in the signed difference to build each permutation. Peak 2 fits an excursion of length 2320, for a Gaussian random walk with standard deviation 0.0904. Peak 4 fits an excursion of length 645 with standard deviation 0.7339. The remaining peaks do not follow an excursion height distribution. Perhaps the quality of the fit reflects the significance of the feature.

### Detail 25 Permutation and Excursion Tests

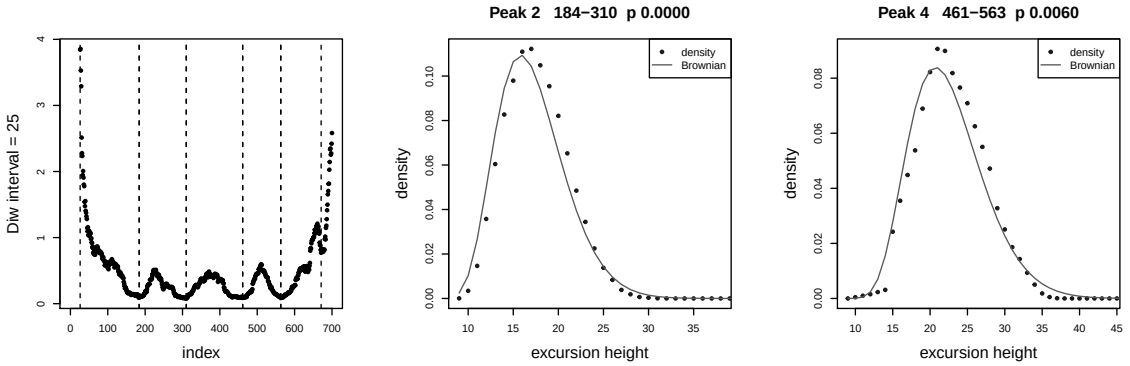



Figure 39: Interval spacing ($w = 25$) of quad-modal example, and fit of Brownian excursion to the height permutation distribution for the two significant peaks.

Listing 4 details the permutation test for runs within a feature and Listing 5 the excursion test built from the point-by-point difference. The framework for both is the same, although the specific operations within each step differ. Step 1 defines the test value, and Steps 2 and 3 prepare the sample set. Step 4 performs the sample and determines the height of the reconstructed feature. Step 5 uses these heights to measure the probability of the feature.

In the run permutation test we will permute the set of runs found in the signed difference of the signal. Step 2 takes the sign of the point-by-point difference of the signal within the feature, reducing the values to $-1$, 0, and $+1$. Step 3 breaks the sequence of symbols into runs, storing the length $l$ and value $v$ of each. In Step 4a we take a random permutation of this set adding the requirement that no two runs adjacent in the draw can have the same sign. If there are only two symbols this is easy to guarantee: the permutation alternates. There can be at most one more run in one symbol than in the other, and if this is the case that run must start the permutation; within each symbol we can simply shuffle the lengths and values. If there are the same number of runs in both symbols then we pick at random one symbol to start, and interweave the shuffled subsets after. If there are three or more symbols then we create a random permutation and remove adjacent same symbols by swapping them with the closest position whose neighbors would meet the requirement. Step 4b reconstructs the signal using a cumulative sum of the run length multiplied by its direction, and Step 4c determines the height or largest range of the result. Step 5 returns the probability of generating the feature, which is the fraction of trials that result in at least as great a height.

| Listing 4: Height Run Permutation Test | |
|---|---|
| Input: | signal $x$, feature index bounds $j$, $k$ inclusive, trial count $N_{perm}$ |
| 1: | let feature height $H \leftarrow \max(x[j:k]) - \min(x[j:k])$ |
| 2: | let signed signal $S \leftarrow \text{sign}(x[i] - x[i-1])$ for $1 \leq i \leq \lvert x \rvert$ |
| 3: | let runs $R \leftarrow \text{runs}(S)$ with length $l$, value $v = -1, 0, +1$ |
| 4: | repeat $N_{perm}$ times: |
| 4a: | $r_{perm} \leftarrow \text{permute}(R)$ such that no adjacent runs have same sign |
| 4b: | for $i' \leftarrow 1$ to $\lvert R \rvert$: |
| 4b1: | $xperm[i'] = \sum_{i=1}^{i'} l[r_{perm}[i]] \cdot v[r_{perm}[i]]$ |
| 4c: | $h_{perm} \leftarrow \max(x_{perm}) - \min(x_{perm})$ |
| 5: | return $p_{feat} \leftarrow \#(h_{perm} \geq H)/N_{perm}$ |

In the excursion test we draw from the raw signed differences. These are taken in Step 2 over

the entire signal, not just the feature, and next we remove the largest $N_{top}$ values if they occur for the first or last $N_{top}/2$ points, as discussed in the main text. The order function in Step 3 stores the rank of each value and in the loop in Step 4 we remove those points whose rank is in the largest $N_{top}$ values. Step 5 creates vectors $s$ to hold the sample and $x_{excur}$ for the reconstructed signal. These are filled in Steps 6a and 6b. The random function returns an integer between the bounds, inclusive and uniformly; we sample with replacement. Step 6b is a cumulative sum. Step 6c determines the height or largest range of the result. The feature's probability in Step 7 depends on whether we are simulating a peak, in which case it is the fraction of trials whose reconstructed height is larger than the feature's, or a flat, when it is the fraction with a smaller height.

Listing 5: Height Excursion Test

Input: signal $x$, feature index bounds $j$ to $k$ inclusive, trial count $N_{excur}$, boolean $isPeak$, drop count $N_{top}$

1: let feature height $H \leftarrow \max(x[j:k]) - \min(x[j:k])$

2: let signed signal $S \leftarrow x[i] - x[i-1]$ for $j+1 \le i \le k$

3: let $o \leftarrow \text{order}(S)$

4: for $i \leftarrow 1$ to $N_{top}/2$:

4a: if $|x| - N_{top} \le o[i]$:

$S \leftarrow S \setminus S[o[i]]$

4b: if $|x| - N_{top} \le o[|x| - 1 - i]$:

$S \leftarrow S \setminus S[o[|x| - 1 - i]]$

5: create arrays $s[1 \ldots (k-j+1)]$, $x_{excur}[1 \ldots (k-j+1)]$

6: repeat $N_{excur}$ times:

6a: for $i \leftarrow 1$ to $k - j + 1$:

6a1: $s[i] \leftarrow S[\text{random}(1, |S|)]$

6b: for $i' \leftarrow 1$ to $k - j + 1$:

6b1: $x_{excur}[i'] \leftarrow \sum_{i=1}^{i'} s[i]$

6c: $h_{excur} \leftarrow \max(x_{excur}) - \min(x_{excur})$

7: if $isPeak$:

return $p_{feat} \leftarrow \#(h_{excur} \ge H)/N_{excur}$

else

return $p_{feat} \leftarrow \#(h_{excur} \le H)/N_{excur}$

### Detail 26 Changepoint Detection Evaluation

We have evaluated fifteen changepoint libraries on five artificial and twelve actual data sets with multimodal data. This is neither a rigorous nor a complete comparison because the variability we have seen in the performance of any detector is too great: we do not believe a general, stable conclusion is possible. Our goal is instead to make a rough recommendation of a minimal set of algorithms. The comparison was based on the accuracy of locating spacing features, the number of changepoints found, and the run time. Scoring best were the non-parametric PERT library *changepoint.np* [19], the Iterative Cumulative Sum of Squares package *ICSS* [21], and joint segmentation library *jointseg* [36]. These involve three different approaches. The first uses a dynamic programming algorithm to find regions that minimize a penalty function based on the distribution function of the data, with changepoints at the boundaries between regions. The second is a process control technique based on the cumulative sum of a metric going out-of-bounds at a changepoint. The third uses either consistent regressions to segment the data or a binary segmentation algorithm, with changepoints again falling at the boundaries.

### Detail 27 Changepoint Voting

Listing 6 gives the voting algorithm for combining different changepoint detectors into a common list. Each package may contain several independent methods $M$ or decision criteria, whose results we combine as a union. It takes a number of extra parameters to screen the library results and to merge nearby points. If a detector is unusually noisy, labeling more than the fraction $f_{pt}$ of the data as changepoints, then we will ignore those results. We will also ignore detectors whose point count is outside the quantile range $q_{vote}$. This might be the inter-quartile values (0.25, 0.75) or a looser (0.10, 0.90) range; it automatically restricts the number of libraries taken into the voting and represents a trade-off between rejecting outlier detections and accepting their variety. When fusing the points from the detectors, we limit their separation using an absolute value $\Delta_{max}$ and a relative fraction $f_\Delta$ of the data.

| Listing 6: Fusion of Changepoint Detectors | |
|---|---|
| Input: | spacing $x$, changepoint detectors $C$, |
| | maximum changepoint fraction $f_{pt}$, point count range $q_{vote}$, |
| | maximum separation relative $f_\Delta$, absolute $\Delta_{max}$ |
| 1: | for each detector $C_i$ in $C$: |
| 1a: | for each method $M_j$ in $C_i$: |
| 1a1: | $p_{raw}[i,j] \leftarrow M_j(x)$ |
| 1a2: | if $f_{pt}\lvert x\rvert \leq \lvert p_{raw}[i,j]\rvert$ then clear $p_{raw}[i,j]$ |
| 1b: | $p[i] \leftarrow$ remove_nearby(sort($p[i,j]$), $\Delta = 2$) |
| 2: | $N_{pt} \leftarrow$ quantile($\{\lvert p[i]\rvert\}, q_{vote}$) |
| 3: | for $i$ in $1 \ldots \lvert p\rvert$: |
| 3a: | if ($\lvert p[i]\rvert \leq \min(N_{pt})$) or ($\max(N_{pt}) \leq \lvert p[i]\rvert$) then clear $p[i]$ |
| 4: | $p_{common} \leftarrow$ remove_nearby($p, \Delta = \min(\text{round}(f_\Delta\lvert x\rvert), \Delta_{max})$) |
| 5: | for $i \leftarrow 1$ to $\lvert p_{common}\rvert$: |
| 5a: | $N_{near} \leftarrow \#\,(\lvert p_{common}[i] - p\rvert \leq \Delta)$ |
| 5b: | if $0.5\lvert p\rvert \leq N_{near}$: |
| 5b1: | report $p_{common}[i]$ |

| Listing 6a: remove_nearby (Intra-library Point Simplification) | |
|---|---|
| Input: | points list $p$, maximum separation $\Delta$ |
| 1: | if $\lvert p\rvert \leq 1$ then return $p$ |
| 2: | for $i \leftarrow 2$ to $\lvert p\rvert$: |
| 2a: | $j \leftarrow i$ |
| 2b: | while $(p[j] - p[j-1]) < \Delta$ and $j \leq \lvert p\rvert$: |
| 2b1: | $j \leftarrow j + 1$ |
| 2c: | if $i \leq j$: |
| 2c1: | $p[i] \leftarrow$ mean($p[i:j]$) |
| 2c2: | delete $p[(i+1):j]$ |
| 3: | return $p$ |

Step 1 builds a set of points $p_{raw}$ for each method that each detector provides. If there are more points than allowed by the fraction $f_{pt}$, then Step 1a2 drops that result. Step 1b combines the per-method results in $p_{raw}$ into a detector changepoint list $p$ using the helper routine remove_nearby with a hard-wired radius of 2. This function, in Listing 6a, scans the list of points and considers those within some distance $\Delta$ to be the same. The test chains, and all points that are clustered are replaced by their center of mass in Step 6a.2c. This process is similar to single linkage for hierarchical clustering, and other methods of grouping would also be possible. $p$ then ends up a union of the

points from each method, with close points merged and noisy methods ignored. In the algorithm $|p|$ represents the number of detectors generating points, and $|p[i]|$ the number of points from detector $i$. Steps 2 and 3 screen the libraries by converting $q_{vote}$ to actual counts and ignoring libraries outside the range. Step 4 builds the potential final changepoint list $p_{common}$ as the union of all detectors' points, using the separation parameters to condense clusters to a point. Step 5 does the actual voting. For each potential changepoint it counts the number of libraries that are within the separation limit and then, if this is at least half the number of detectors with points, it records the changepoint.

By default we take $f_{pt} = 0.05$ and the inter-quartile range as $q_{vote}$. Separations are at most $f_\Delta = 0.02$ of the data or $\Delta_{max} = 10$.

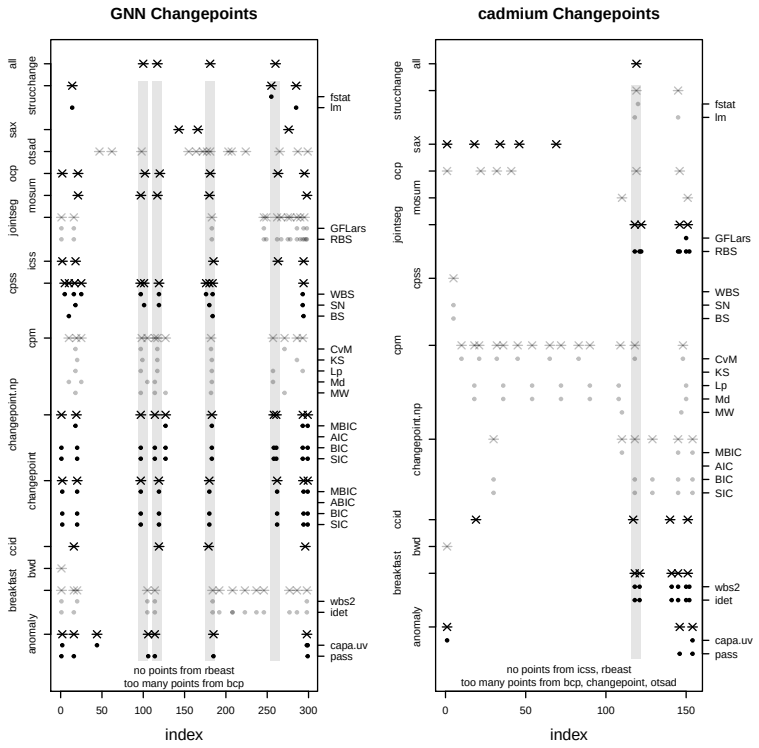



Figure 40: Changepoints for the GNN (left) and cadmium (right) examples. Grey points not used in voting. Grey bands mark the final selection and fusion width.

Figures 40 and 41 shows the changepoint voting for our examples. The variants within each detector are named on the right axis and marked with dots. These combine by union into a set for the library and are labeled on the left and marked with crosses. The variants are clearly related but not duplicates. Libraries that generate too many or few changepoints do not contribute to the voting and are greyed out. The *breakfast* package for the GNN spacing, and *cpm* package for the others, are especially noisy. The voting matches nearby points and keeps those appearing in a majority of the remaining detectors: 5 of 9 for the GNN example, 2 of 4 for the cadmium and 3 of 5 and 2 of 4 for the earthquake depths. The final selections are marked in the top "all" row, with grey bars giving the "nearby" range. This is $\pm 6$ for the GNN example, $\pm 3$ for the cadmium data, and $\pm 10$ for the earthquake depths. Some points that seem to lie at the edges of these ranges fall just outside.

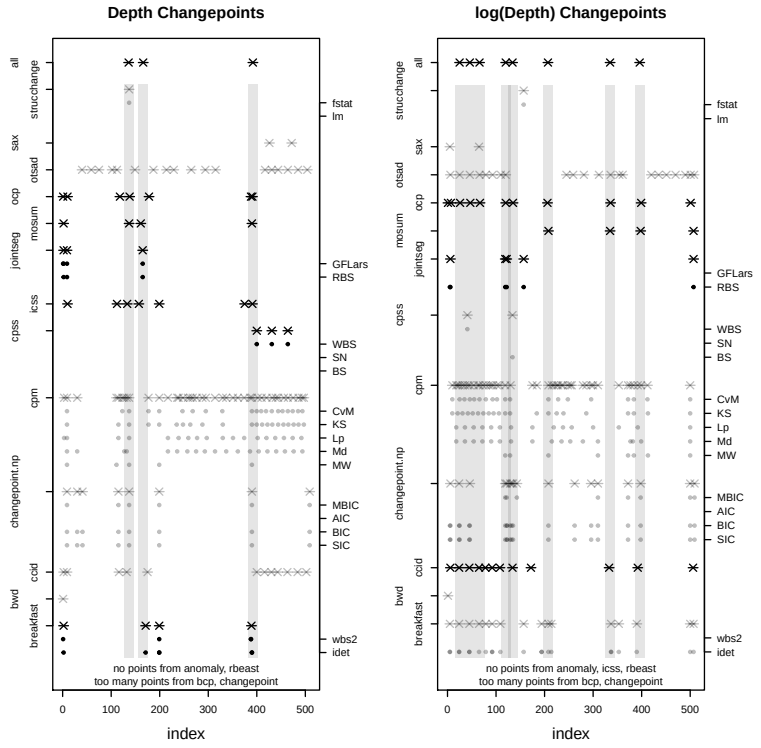



Figure 41: Changepoints for the earthquake depth examples, based on raw values (left) and log scaling (right).

### Detail 28 Analysis Parameters for the Examples

We use a Kaiser low-pass filter spanning 15% of the data for the GNN, cadmium, and log earthquake depth examples, For the PPP and raw earthquake depth examples we use a 10% Kaiser filter so that the second peaks appear. The interval spacing is set to 25 for the GNN example, 30 for the PPP and raw earthquake depth data, 15 for the cadmium example, and 40 for the log depths. These are in the range of 5% to 10% of the sample sizes.

In all examples we use $f_h = 0.05$ and $f_{h,rel} = 0.15$ for the peak detector. The excursion tests use 5000 permutations or resamples, and the largest $N_{top} = 10$ steps are ignored from the sample if they occur among the first or last 5 points. The width of a peak extends to 90% of its height, rather than all the way to the minimum. The flat detector uses by default $\delta_{ripple} = 0.05$ and $f_{L,rel} = 0.05$, with a minimum length of $L_{abs} = 30$. However, this changes with the example. The PPP data needs a larger ripple of $\delta_{ripple} = 0.10$ to merge the many small flats separated by the steps into one. The interval spacing is much rougher and requires tuning. For the GNN sample we set $\delta_{ripple} = 0.03$, for the PPP $\delta_{ripple} = 0.20$ and $f_h = 0.20$, and for the cadmium data $\delta_{ripple} = 0.05$. For both the low-pass and interval spacing analyses of the earthquake data we use all defaults.

Changepoints are based on voting using the following R libraries: *anomaly* (pass and capa.uv tests, with post-processing of bands); *bcp* (bcp test keeping posterior probabilities above 0.5); *breakfast* (idetect_seq and wbs2 solutions, with post-processing of bands); *ccid* (detect.ic test); *changepoint.np* (PELT with SIC, BIC, AIC, and MBIC penalties); *cpm* (Mann-Whitney, Mood, Lepage, Komogorov-

Smirnov, and Cramer-von-Mises tests); *cpss* (cpss.meanvar test with WBS algorithm); *ICSS* (ICSS test); *jointseg* (jointSeg test with RBS and GFLars methods); *mosum* (multiscale bottomUp test); *ocp* (onlineCPD test); *otsad* (pewma test); *Rbeast* (beast test); *strucchange* (breakpoints test for constant and F statistics models). An internal version of the HotSAX anomaly detector [23] was also used, for 15 libraries in total. Changepoint parameters by default are $f_{pt} = 0.05$, $q_{vote} = (0.25, 0.75)$ or the inter-quartile range, $f_\Delta = 0.02$, and $\Delta_{max} = 10$. For the methods that return bands, which we've called intervals or sections, we first merge them if they overlap or are separated by a small gap $\Delta_{max}$ wide, and then collapse short bands with width less than $\Delta_{max}/2$ to their midpoint.

We set a significance level $\alpha = 0.5$ for the critical value in the level test.

**Detail 29 Bi-Modal Variants**

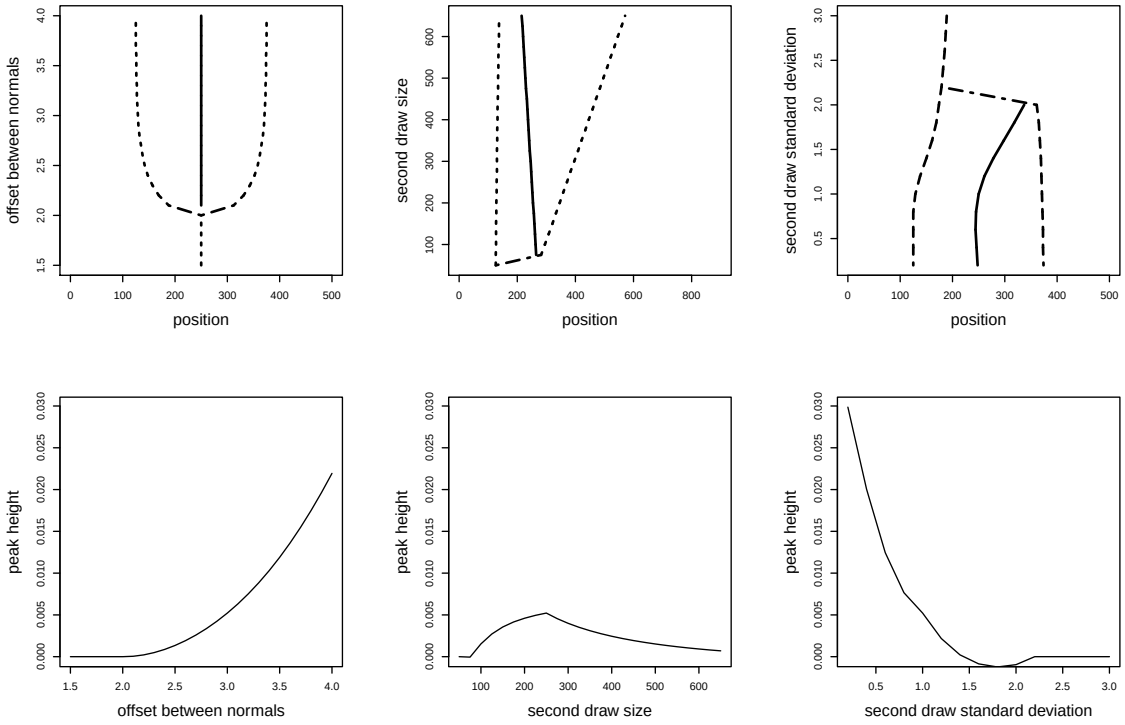



Figure 42: Location of gaps (solid) and modes (dashed) in bi-modal draw variants (top), and peak height at gap (bottom).

The base setup uses two normal draws

$$250 \times \mathrm{N}(0, 1) \quad 250 \times \mathrm{N}(3, 1)$$

where the second parameter is the standard deviation. We vary the mean, draw size, and standard deviation of the second variate. Larger separations will make it easier to distinguish the two modes by increasing the peak at the anti-mode. An imbalance in the draw size will shift the anti-mode towards the smaller draw and decrease the peak while making more room for a flat in the larger variate. A

smaller width increases the peak and supports a flat by creating a series of tightly spaced points. Rather than numerically solving the expected spacing for each variant, we estimate the position of the extrema by finding the inflection points of the variant's total distribution function. There will be three, two minima astride a maximum. To convert positions to indices we use the distribution function as a fraction of the total draw. We then numerically integrate the expected spacing at these points to get the height difference. The larger is the height of the peak. Figure 42 plots the location of the modes (solid lines) and anti-modes (dashed) for each variant, and the peak heights at the gaps.

Figure 43 counts the number of features of each type detected for each variation. Peaks are much more common in the interval spacing, because its poor sidelobe suppression leaves high-frequency components. The behavior also differs from the low-pass peaks, which follow the expected curves for each variation: with small separations or large widths, when the two draws cannot be distinguished, or with larger draw imbalances, which reduce the peak height. Flats on the other hand are much more common in the low-pass spacing, again because the rougher interval data prevents them from forming. While we find more than one low-pass flat per draw, for interval flats this only happens with a large second draw. Again the transitional behavior for the two filters differs in each variation.

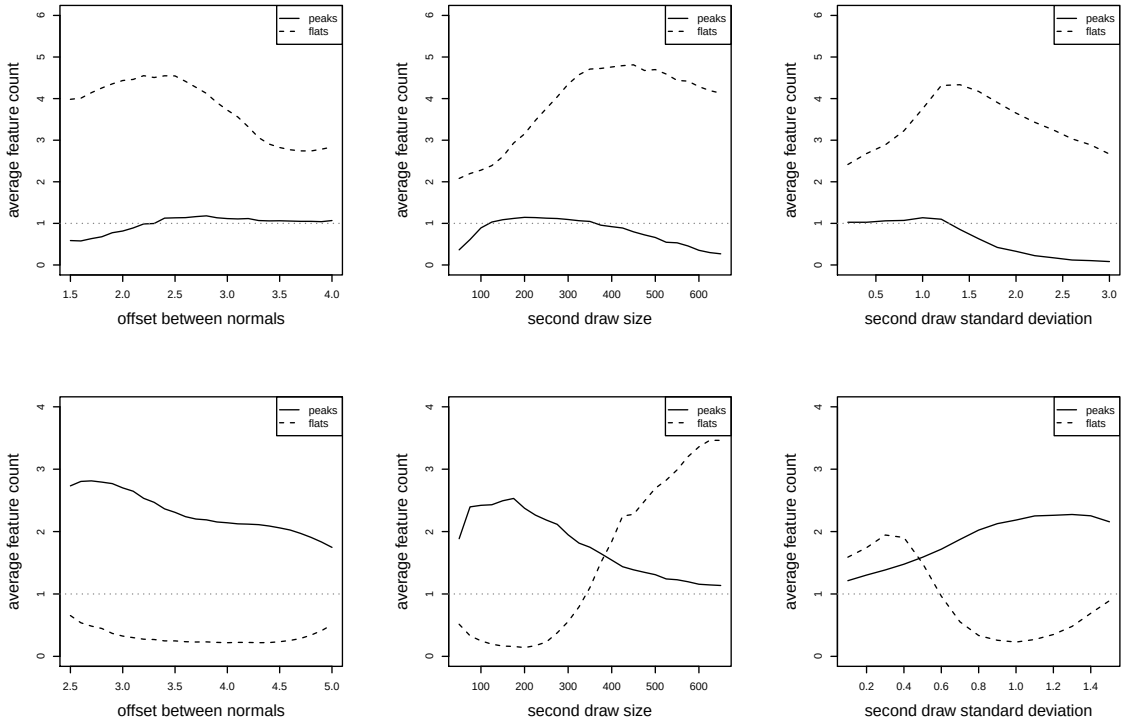



Figure 43: Number of features detected in the low-pass (top) and interval spacing (bottom) for variations in the second draw mean (left), size (middle), and standard deviation (right).

The analysis of these variations has been divided into three parts. [Detail 30] studies how the peak and flat features change in the low-pass spacing, and at changepoint and level sections in the raw spacing. [Detail 31] looks at excursion tests in the interval spacing. [Detail 32] analyzes the performance of the runs tests in the signed difference of the interval spacing.

### Detail 30 Analysis of Low-Pass Bi-Modal Variants

The main text summarizes what happens with the peak and flat tests on average as the parameters of the bi-modal setup change, either the separation between the two normal distributions (mean), the size of the draws, or a difference in the distribution's width (standard deviation). Here we show how many features are detected and accepted by each low-pass test for each variant, their position against the ideal, and a classification error rate. We also describe the changepoints and level sections in the raw spacing.

Figure 44 is effectively a stacked bar plot with lines separating regions of different numbers of significant features, as labelled. Solid and dashed lines alternate, with dark and light grey representing the acceptance level. Peaks appear as the separation increases from 2.0 to 3.0 (left graph). To the left of the boundary the test cannot distinguish the two modes; it considers the set-up to be uni-modal. The tighter 0.01 significance level is equivalent to a 0.3 larger separation. Unbalancing the draw sizes (middle graph) produces smaller peaks, which leads to none significant being found. The curve looks asymmetric, but that would disappear it were plotted as the ratio of the larger draw size to the smaller. Changing the significance level has little effect on the behavior. Increasing the width of the second draw (right graph) blurs the boundary between the two modes and the peak disappears as the standard deviation increases from 1.0 to 2.0. The tighter significance level shifts follows the same transition, but at a standard deviation 0.1 smaller. In all three graphs a second peak can appear during the transition, in up to 10% of the trials, which causes the average in Figure 13 in the main text to rise above 1.0.

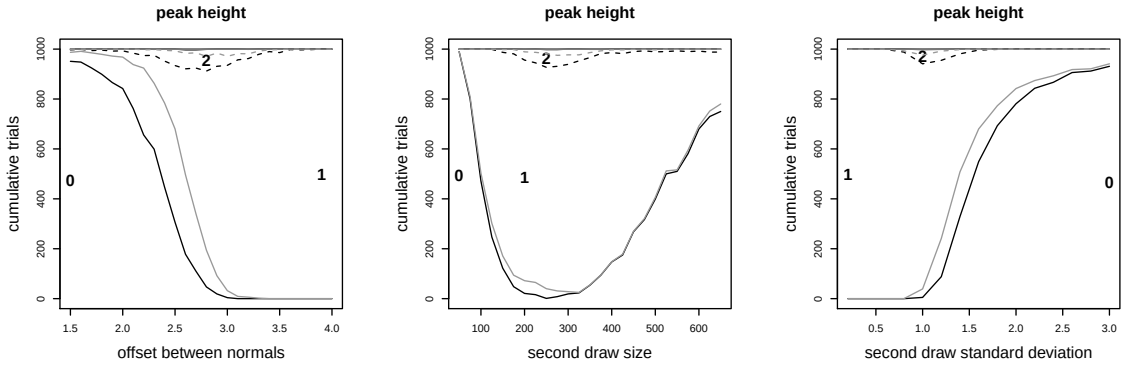



Figure 44: Number of peaks accepted by the height model at significance levels 0.05 (dark) and 0.01 (light) as the separation (left), draw size (middle), and standard deviation (right) of the second normal changes.

The excursion test applied to peaks is more conservative than the height model. Its behavior is generally the same (Figure 45), with an extra separation of 0.5 needed to find peaks, or a standard deviation 0.5 smaller. The transition in the first is a little broader, not reaching either limit as quickly, while the second is sharper. The second transitional peaks do not appear. The test's sensitivity is locally lower when the draws are balanced, with more significant peaks found with a slight imbalance in either direction in draw sizes, before all potential peaks are rejected at large imbalances. In this last region, with the second draw below 150 or above 250, the separation between no and one peak is the same as the height model test. A tighter significance leads to similar shifts as the height model: an extra separation of 0.3 or a standard deviation 0.1 lower. It accentuates the difference between the balanced and slightly imbalanced draws.

The height model results are roughly consistent if plotted against the actual peak height (Fig-

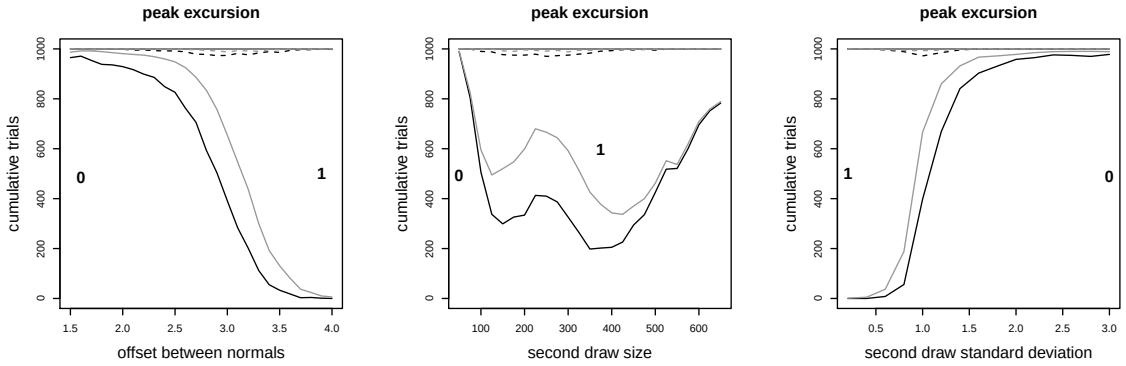



Figure 45: Number of significant peaks using the excursion test, with the same layout as Figure 44.

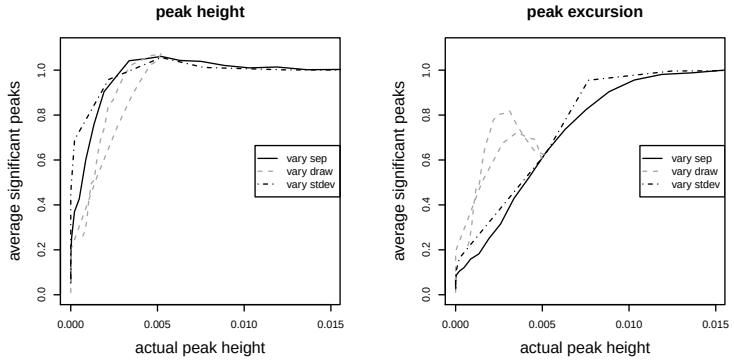



Figure 46: Average number of significant peaks, against the actual peak height from numeric integration of the spacing (Figure 42).

ure 46). The curves are similar, but the pass rates are more sensitive to changes in the mean than to the standard deviation, for both tests. The size variations have two branches taking the larger ratio of the second draw size to the first or its inverse. The excursion test appears to be sensitive to the specific setup and not just the peak height. It responds more slowly to shifts in the mean and draw size than the height model does, but has a sharp transition against the standard deviation. A small width would reduce the variability of the spacing, decreasing the probability of a peak.

The length (Figure 47) model accepts few flats, except when the second standard deviation is small and the draw is tight enough to support a flat, or when it is large and the draws merge. It also accepts flats when the second draw is small and affects the first mode least. No flats pass at the 0.01 level.

If the peak excursion test is more conservative than the model, the flat excursion test is more liberal, for all variations (Figure 48). It regularly finds multiple flats, up to four, over the trials. There are very broad transitions to more flats with larger separations, imbalanced draws, or standard deviations away from the nominal value. The test accepts the most flats when the second draw is larger than the first, not when balanced. The width variations are also asymmetrical, the count slowly dropping for large widths. Using an 0.01 significance level is equivalent to accepting one fewer flat, in general, although this is mitigated for the largest second draw or width.

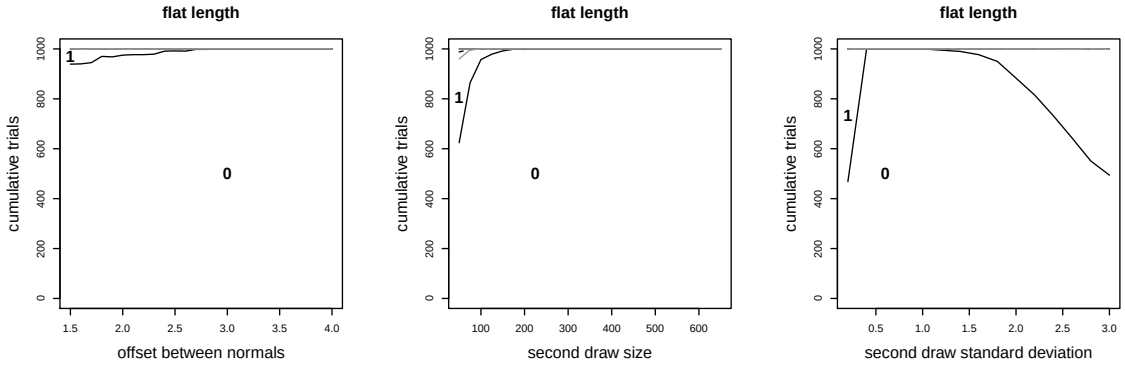



Figure 47: Number of flats accepted by the length model. Same layout as Figure 44.

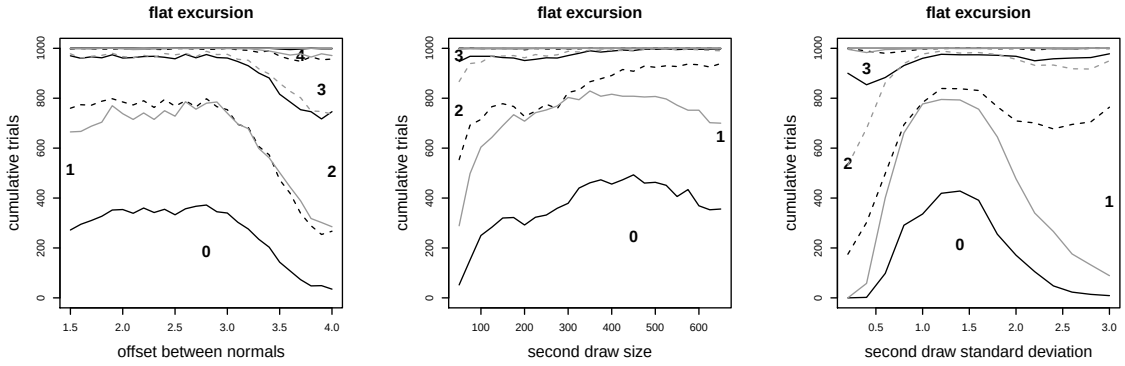



Figure 48: Number of significant flats using the excursion test. Same layout as Figure 44.

Such raw counts of the number of significant features found reflect the sensitivity of the test, but the results are incomplete in two ways. They say nothing about the detected features, which include those found insignificant, or the accuracy of the test. The positional plots and accompanying discussion in the main text (Figure 12) show they follow the expected modes and anti-mode, and that testing correctly restricts spurious detections, eliminating sometimes complicated secondary features.

The contours using the excursion test are essentially the same as those from the height model. The significant feature contour stops short of the detected, which follows the ideal position. This suggests the tests are conservative and could be relaxed while maintaining accuracy. The excursion test does a good job of limiting flats to just the minima. It rejects those at the maxima, in the smaller draw, and in the larger width. This accuracy argues against relaxing the significance level.

The length test accepts a small subset of the significant excursion flats (Figure 49). These flats appear when the separation is too small to distinguish the draws, or when the standard deviation of one draw is much smaller than the other and the higher variance determines the maximum ripple height, or when the second draw is much smaller than the first.

Comparing the significance of the features found against their expected positions allows us to judge the accuracy of the test and assign results to a confusion matrix. Define a true positive as a significant peak whose location lies within 5% of the sample of the ideal maximum. For flats the span

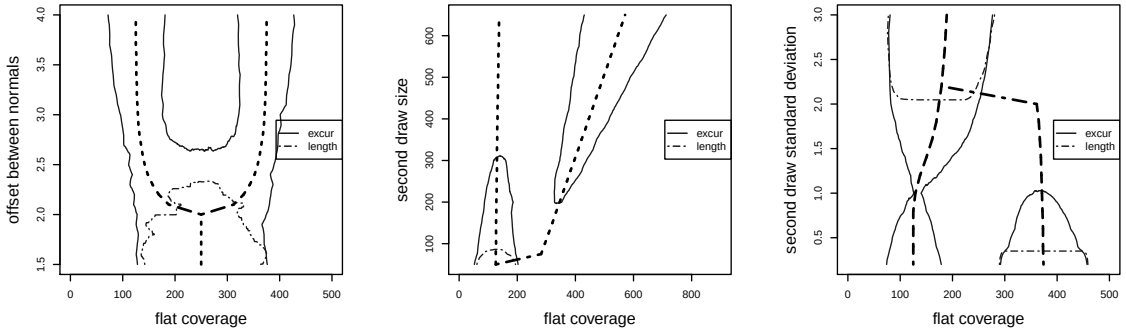



Figure 49: Contour lines enclose 75% of the significant flats at the 0.05 level for each test. Heavy lines mark expected minima.

must include a minimum of the spacing. A false positive will be a significant peak not positioned correctly, or a flat not holding a minimum. A true negative will be a detected but insignificant feature not near an ideal maximum or not including a minimum. A false negative will be a detected feature located at the right position but whose test is not considered significant. It will also include trials without any significant features. This classification scheme is still arbitrary in the sense that we cannot specify which situations should generate features, but allows us to investigate the limits of detection.

The confusion matrix assignment rates will differ from the passing test counts. Those treat all failing tests as one, rather than counting the number of failures separately. Trials that generate one insignificant feature have as many passes, zero, as those that generate five. The mis-match is bigger for flats than for peaks. In less than 10% of the trials do the detectors find multiple insignificant peaks, so the division between the negative and positive bins will be similar to the border between no and one significant peak. Most trials instead find four or five insignificant flats, and the true and false negative rates will dominate.

Our rate plots will show the stacked or cumulative fractions of the trials for each variation of the bi-modal setup. Solid lines separate true from false, and dashed lines negative from positive. For peaks, where the tests find only one feature per trial, the dashed line will correspond to the boundary in Figures 44 and 45 between no and one flat, but for flats it will correspond to the division between one and two flats. From bottom to top the curves divide the space into true and false negatives, then false and true positives.

Figure 50 shows the confusion matrices for the peak height test. The false errors form a wide band about the transition. False negatives disappear when the modes are well-defined and can be distinguished: separated, balanced, or with small widths. Trials yielding no significant peaks account for the largest false negative rates, when the second draw is larger than 400 points or its width between 1.5 and 2.0. The false positive rate is generally 20–30%, disappearing only for imbalanced draws. This suggests the variability of the peak's location beyond the 5% alignment requirement is constant.

Although the overall behavior of the excursion test is similar to the height model (Figure 51), there are many subtle differences. It produces fewer false positives and more false negatives, for all variations and at both significance levels (not plotted). This is the conservative nature of the test. It

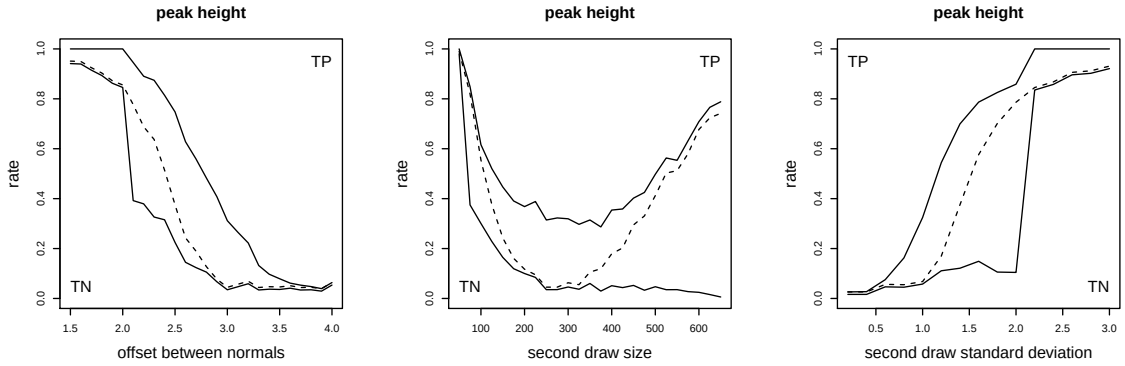



Figure 50: Confusion matrix rates for the peak height model at the 0.05 significance level, varying separation (left), draw size (middle), and width (right).

considers fewer peaks to be significant even if the higher FN rate suggests the peaks that are detected are correct, or at least correctly positioned. The sudden rise of true negatives when the standard deviation reaches 2.0 or the separation 2.0 is present in both tests.

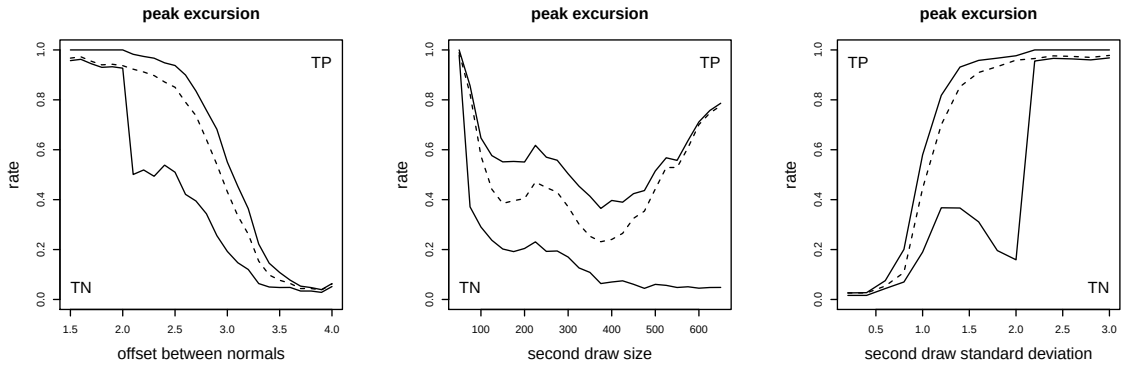



Figure 51: Confusion matrix rates for the peak excursion test at the 0.05 level.

We can check the consistency of the evaluation of a feature by two tests, either graphically by comparing their probabilities, numerically with confusion matrix metrics, or statistically by looking at their correlation and ordering. For the first we plot the test probabilities against each other and look for a clear trend, ideally along the 45° diagonal. Differing significance judgements will be seen by dividing the graph at the 0.05 or 0.01 level into quadrants, so that points in the first and third agree and the second and fourth do not. Defining a sensitivity-like metric for points in the second check, let

$$\begin{aligned} \text{disagree}_y &= \frac{Q_2}{Q_2 + Q3} \\ \text{disagree}_x &= \frac{Q_4}{Q_3 + Q_4} \\ \text{agree} &= \frac{Q_1 + Q_3}{Q_1 + Q_2 + Q_3 + Q_4} \end{aligned} \tag{33}$$

for the test on the $x$ or $y$ axis. We use Spearman's $\rho$ statistic to check correlation for the third check,

and the Wilcoxon rank test to see if the two tests consider one point larger or smaller than another.

The low-pass peak height model follows the excursion test (Figure 52); excursion probabilities of 0 are plotted on the left despite the log scale. There is a moderate matching when the draw separation is 3.0; a regression line has slope 0.86, standard error 2.73, and $R^2$ 0.37. As the separation increases the height model divides the peaks into two sets. Those that are insignificant match the excursion test with a clear trend, although a slope less than one means the height model is more sensitive. Those that are significant show less matching, with a smaller regression slope and $R^2$ going to zero. Agreement rises from 62% when the offset is 3.0 to 97% at 3.5, and the disagreement by the excursion test drops from 40% to 4%. The height model passes every feature that the excursion model accepts, but the converse is not true. This is more an effect of the movement of the results into the third quadrant. Probabilities from the excursion test are always larger than the worst from the height model among the significant features. $\rho$ decreases from 0.641 at the 3.0 offset to 0.361 at 4.0, but these translate to a significance level of 0.0000; the two tests are associated. The two tests consistently order peaks at all offsets, with the Wilcoxon test rejecting a difference at the 0.0000 level.

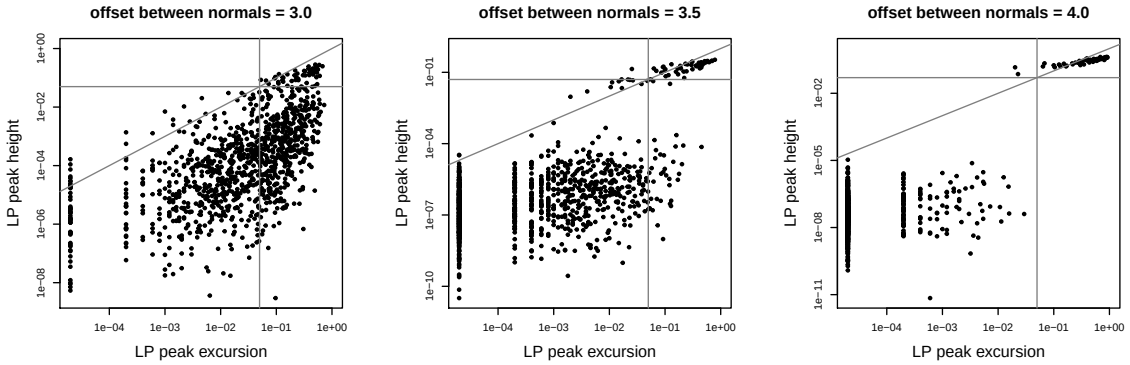



Figure 52: Probabilities of same peaks from excursion test ($x$ axis) and height model ($y$) for individual trials at three separations of the two normal draws. Lines mark 0.05 significance level and ideal match.

Table 29 counts the number of each feature in the quadrants of Figure 52, adding a split at the 0.01 level. At a separation of 3.0 the height model predominantly accepts proposed peaks at the 0.01 level, but the excursion test splits them roughly equally. At 3.5 the excursion test accepts most of the detected features at the 0.01 level, and at 4.0 the shift is complete: except for a small core of 66 peaks that are rejected by both tests, only eight do not fall in the 0.01 bin for either test. The height model passes every feature that the excursion model accepts, but the converse is not true.

Table 29: Peak Test Comparison for Low-Pass Spacing

| | LP height model | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | 3.0 separation | | | 3.5 separation | | | 4.0 separation | | |
| LP excursion | 0.01 | 0.05 | insig | 0.01 | 0.05 | insig | 0.01 | 0.05 | insig |
| insig | 375 | 51 | 54 | 33 | 2 | 47 | 0 | 0 | 66 |
| 0.05 | 266 | 11 | 0 | 98 | 9 | 2 | 6 | 0 | 2 |
| 0.01 | 356 | 2 | 0 | 871 | 1 | 0 | 994 | 0 | 0 |

Because the length model passes so few features, the confusion matrices only generate TN and FN rates, with the dashed line dividing the positive buckets pinned to the top of the graphs (Figure 53). There are few false positives, even as the standard deviation changes and some significant flats appear. The true negative rates generally vary between 50% and 75%, which agrees with the surplus of flats

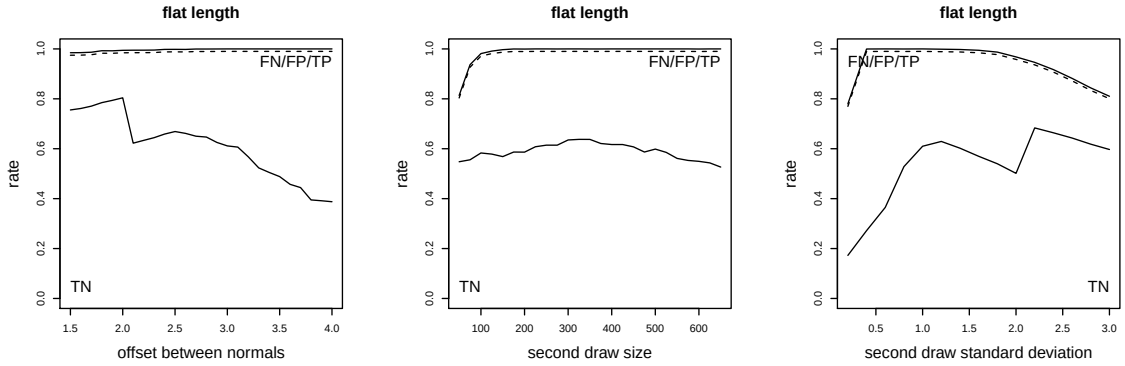



Figure 53: Confusion matrix rates for the flat length model test at the 0.05 level.

detected, two to three times higher than the two modes that are available.

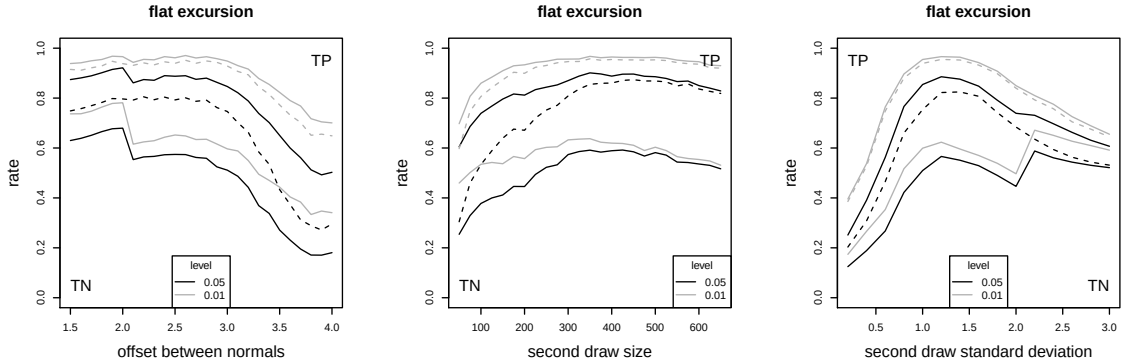



Figure 54: Confusion matrix rates for the flat excursion test while varying second draw, at 0.05 (dark) and 0.01 (light) significance levels.

The excursion test does accept some of the proposed flats, and so it does have TP and FP buckets (Figure 54). With four or five flats detected per trial and only two modes, we would expect a true negative rate of 50% or 60%, and a true positive rate in the remainder. In general the test achieves this for the rejected flats, but only at half the rate for those accepted. In other words, the test passes a flat in one of the two modes. It does better with larger separations and small standard deviations. The false positive rate is largest for the separation variants and when the second draw is small, but is in general less than the false negative rate, which again implies the test is rejecting flats that are correctly positioned. The false positive rate disappears when the significance level is 0.01 while all curves shift upward by 10%, representing the rejection of more features. The consistency of the flat model and excursion tests is included with the comparison to interval flats. [Detail 31]

Figure 55 counts the number of changepoints found. The curves peak at the base conditions and the changepoint count increases away from these parameters. From Figure 15 in the main text, the increase for larger separations comes from changepoints around the local peak, and for smaller separations from additional points in the tails. The count increases for large second draws as more changepoints appear in its tail, but the first, fixed mode generates them when the draw size shrinks. At small standard deviations changepoints appear at the sharper transition, and they are scattered

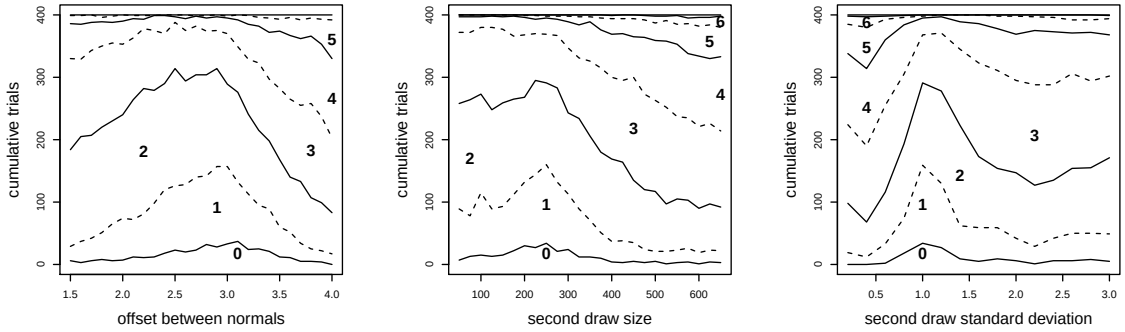



Figure 55: Number of changepoints found per draw parameter variant.

throughout the second draw when its width increases. We see a limited number of changepoints that fall between the start or end of a flat and a peak, which is the boundary we are interested in marking. Of the 400 trials of the two draw variants where we gather both features and changepoints, we see more such changepoints when it is easier to distinguish the two draws (Table 30). Half of the repetitions have two appropriate changepoints when the separation is 4.0, and half have one when the second standard deviation is 0.8 or better, or the second draw is large.

Table 30: Changepoint Placement

| | boundary changepoint count | | | | | | boundary changepoint count | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | 0 | 1 | 2 | 3 | 4 | | 0 | 1 | 2 | 3 | 4 |
| separation | | | | | | standard deviation | | | | | |
| 2.5 | 356 | 35 | | | | 0.4 | 40 | 209 | 133 | 17 | 1 |
| 3.0 | 267 | 108 | 24 | 1 | | 0.8 | 94 | 187 | 103 | 16 | |
| 3.5 | 79 | 145 | 155 | 20 | 1 | 1.2 | 311 | 77 | 12 | | |
| 4.0 | 7 | 64 | 236 | 86 | 7 | 1.6 | 286 | 104 | 10 | | |
| second draw size | | | | | | | | | | | |
| 100 | 253 | 140 | 7 | | | | | | | | |
| 200 | 249 | 131 | 20 | | | | | | | | |
| 300 | 246 | 127 | 27 | | | | | | | | |
| 400 | 172 | 187 | 38 | 3 | | | | | | | |
| 500 | 217 | 161 | 16 | 6 | | | | | | | |

The number of level sections per variant (Figure 56) is higher than the number of changepoints. The smallest counts drift towards smaller offsets and larger standard deviations, rather than curves peaking at the base setup. The counts stabilize at the largest separations but not the largest and smallest widths, unlike the changepoints. Steadily more sections cover the sample as the second draw increases, so additional sections are being created there. Most sections are short, seen in Figure 57 by the shift of the median length towards the 25% quantile. The median length seems stable except in the transition region as the two draws become distinguishable. The 10% and 25% quantiles do not vary. The 75% quantile length ranges between 150 and 225 for the separation and width variants, but grows steadily with the larger draw until it stabilizes beyond 400 points. The sections overlap substantially. Figure 58 plots the average number of data points covered by more than one section, ignoring those overlaps that are more than 60% of either section. The count is sensitive to this cut-off, which balances keeping large common segments and complete coverage. The overlap grows during the transitions in the variants.

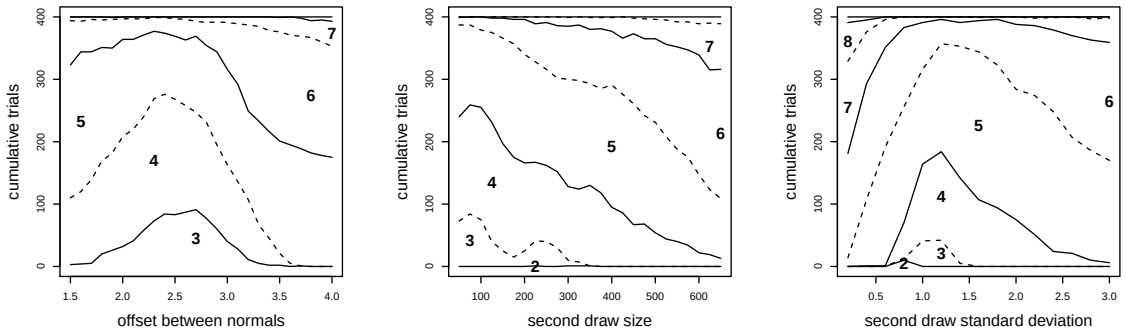



Figure 56: Number of level sections found per draw parameter variant.

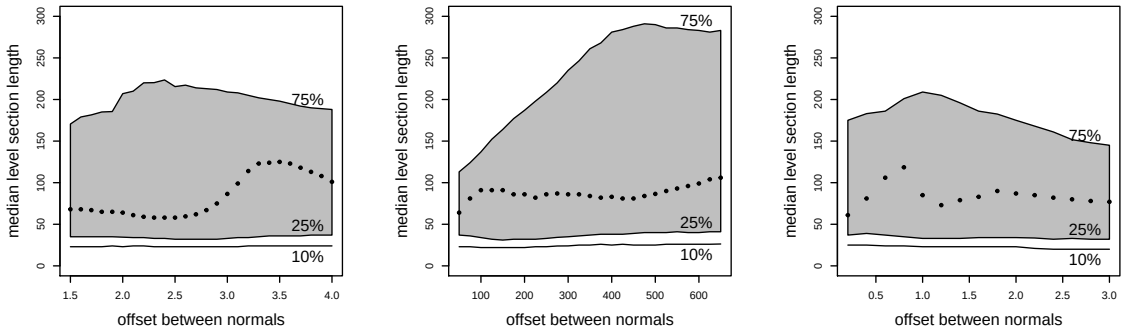



Figure 57: Median level section length and inter-quartile range (grey) per draw variation.

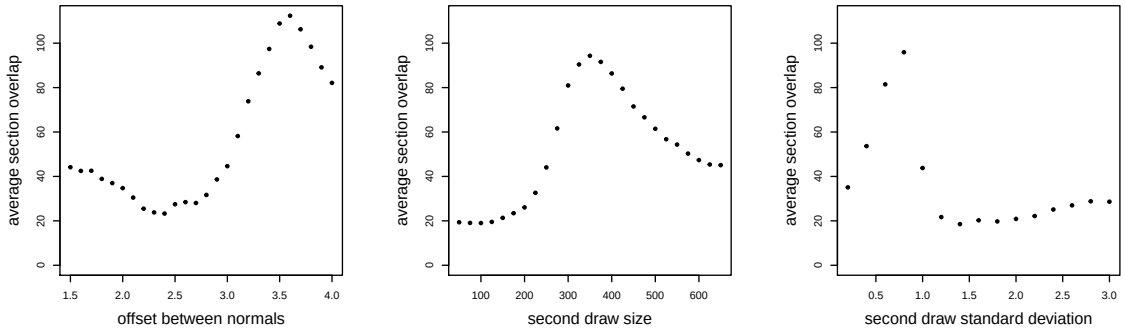



Figure 58: Level section overlap per draw variation.

Alignment between changepoints and level section endpoints, defined as lying within 5 points of each other, shows little effect from the draw variations (Figure 59). We calculate per trial the fraction of each that align, and plot the average over all 400 repetitions. The changepoint alignment changes a little over the variations, increasing until the separation is 3.0 and then stabilizing, decreasing with increasing draw size although stable for imbalances in the other direction, and stepping down at

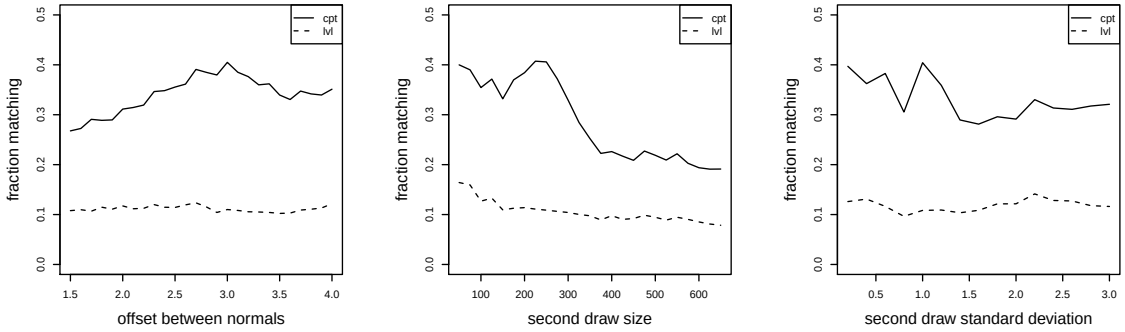



Figure 59: Fraction of changepoints and level section endpoints that align, per second draw variation.

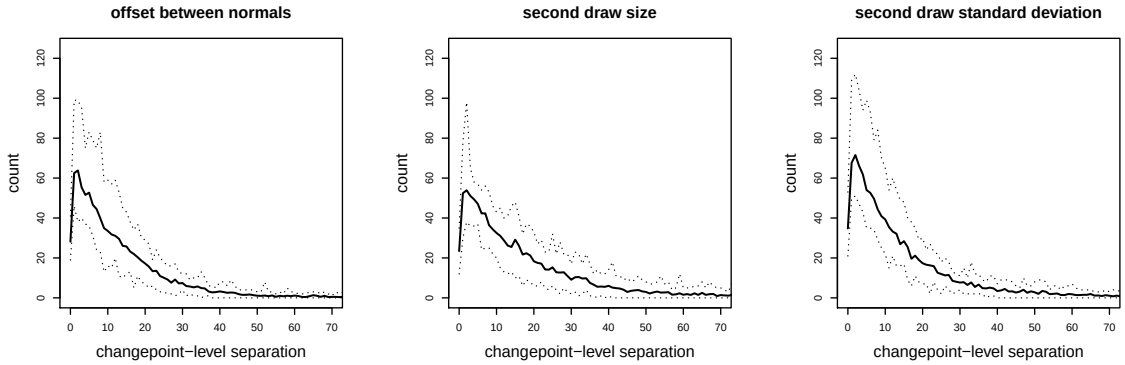



Figure 60: Number of changepoints per separation to the nearest level endpoint, averaged over all variants and trials. Dotted lines bound the counts.

a standard deviation above 1.0. The endpoint alignment is stable, and smaller because there are roughly three times fewer changepoints. There is no change to the results if we loosen the alignment criterion to 10 points. [Detail 33] uses a similar analysis on another set of variations.

Instead of counting the number of matches, we can instead measure the smallest separation. Figure 60 plots the average count of these minimum separations over the variants and all trials and changepoints. The histograms peak at a separation of one, but tail off fairly slowly, especially considering that there are only 500 points drawn in the base setup. The dotted lines are the smallest and largest counts over the variants. We see only a weak agreement between the two tests.

[Detail 33] looks at the matching of the two in another sample set where we can control the stability of the features.

Level sections form a superset of flats. Define overlapping as either a flat being completely within a level section, or when the common segment is more than 75% of the length of the flat and the section. Then we find that almost all flats are overlapped by a level section, with fewer than a tenth not being covered when the two draws are separated by 4.0. At the base separation of 3.0 this increases to a third of the flats not being overlapped, even as the number of number of flats increases. These flats are small, which can be seen in two ways. With fewer level sections than flats, there are multiple

flats per section and they must be shorter. Alternatively, a quarter to a fifth of the level section is occupied by each flat, with the flats consuming a majority of the level section in only a handful of trials. The multiplicity increases as the separation shrinks. Figure 61 demonstrates these conclusions for a separation of 4.0 and 3.0, measured over 400 trials. The bars in the left graph indicate the number of uncovered flats, and in the middle the number of overlapping level sections. The right chart is a histogram of the fraction of the level section length that each flat occupies. The flats are much shorter than the level sections, so covering means complete overlap rather than mutual.

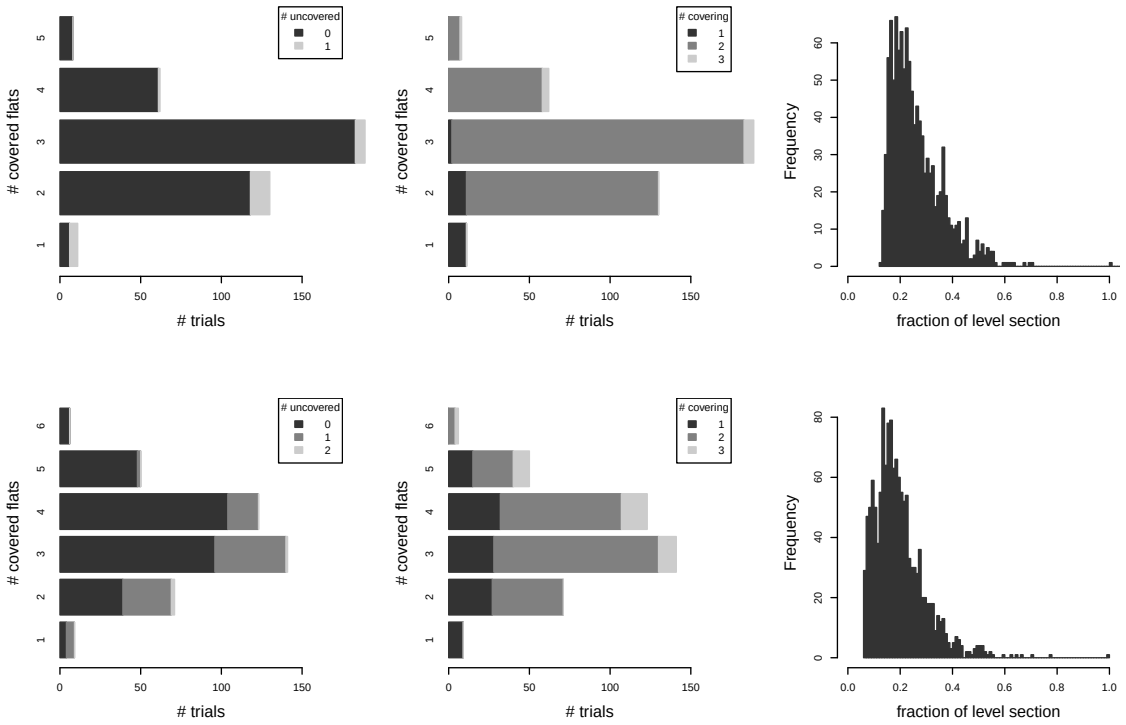



Figure 61: Overlap of flats by level sections for the two normal draws separated by 4.0 (top) and 3.0 (bottom), showing (left) how many flats are not covered, (middle) how many level sections cover the flats, and (right) the fraction of each level section occupied by a flat.

Were we to introduce a tight mode, by setting the standard deviation of the second draw to 0.4, we would decrease the number of flats and improve the match to the overlapping level sections, with fewer flats per level section occupying a larger fraction of their length.

### Detail 31 Analysis of Interval Spacing Bi-Modal Variants

We use different ranges for the variations of the interval spacing tests to account for their lower sensitivity. The base separation is increased to 3.75 from 3.0 and varied over the range 2.5 to 5.0, because the tests require a larger offset to distinguish the modes. The standard deviation base value decreases to 0.75 from 1.0 and varies between 0.1 and 1.5. We do not change the draw size variations. We use an interval of 10% of the total draw. Flats are found with a fractional ripple of $\delta_{ripple} = 0.02$. Features are evaluated with an excursion test. [Detail 32] looks at the runs tests.

Peaks in the interval spacing occur more often and are much more widely distributed than the low-pass features (Figure 62). Similar trends appear: peaks follow the expected position, although there is an offset to one side, especially for the separation variations. This is because we report the feature peak at the end index of the interval. Shifting it to the middle would move the feature to its expected position. As the two draws merge the peak's position is much less stable. We cannot see if peaks disappear when the draws become indistinguishable, because these conditions — separation smaller than 2.2 or width above 1.5 — are now outside the variations. We see many detections around the minima. Peaks are accepted at larger offsets than in the low-pass spacing, above 3.75. Detected peaks also appear in the second mode when its draw is small. The test rejects peaks when draws are balanced. Detected peaks exist for larger standard deviations than needed for the low-pass tests, but their acceptance stops at the same width, 0.75. The interval spacing seems to be noisier than the low-pass, generating more proposed peaks, but the excursion test picks only those that are located correctly, while rejecting the many proposals away from the actual peak. Note too the vertical lines at the left and right sides of the offset and width plots. These peaks lie at the edge of the data.

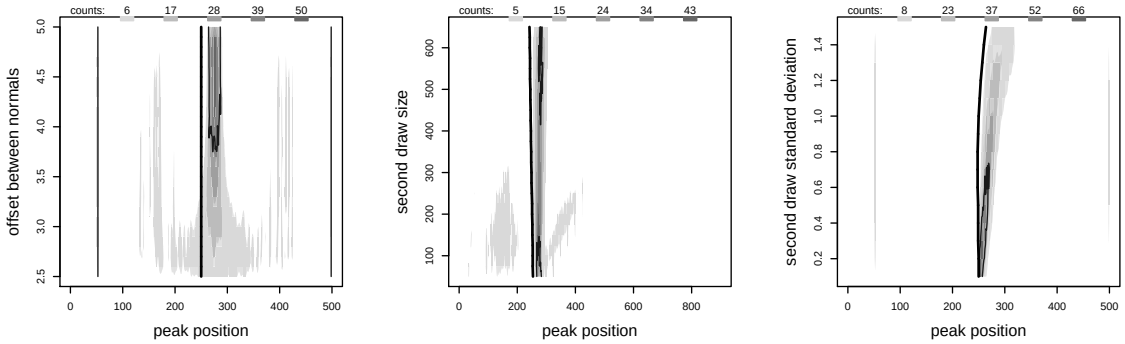



Figure 62: Counts of locations of peaks for second draw variations. Heavy line marks the expected location. Contour line encloses 75% of the significant peaks per the excursion test.

Comparing the counts in Figure 63 to the low-pass result in Figure 13, we find the same transition when accepting peaks, except for a shift in the mean by $+1.0$ or standard deviation by $-0.9$. The trials that generate two significant peaks mostly include one of the edges among them, visible at the edges of the separation and width graphs of Figure 62, or one of the unbalanced modes. The transition for the separation variants corresponds to an increase in the number of significant peaks located accurately between the modes, marked true positive in the bottom graph. The error rate, or sum of the false positives and negatives, is steady at 25%, but at the transition it shifts from correctly placed but insignificant maxima, or false negatives, to incorrectly placed but significant cases, or false positives. The false positive rate stays small for the standard deviation variants, in contrast, with false negatives becoming true positives. As we saw with the low-pass filtered data, the base operating point produces fewer significant peaks when the draws are balanced. The errors are split evenly between false negatives and positives, except when the second draw is small and the false negative rate goes to zero. A peak in a mode is responsible for the false positive.

Figure 64 is the equivalent of Figure 46 for the interval spacing tests including runs, plotting the number of significant peaks per trial and variant against the actual peak height, solved numerically. The run permutation height accepts more peaks of a given height, but has a large false positive rate when there are no peaks. A tighter acceptance level shifts the curves down, replacing the false positives by false negatives at the same rate. The longest run test is much less sensitive. From the size variations it seems the excursion and permutation tests are also affected by the number of points

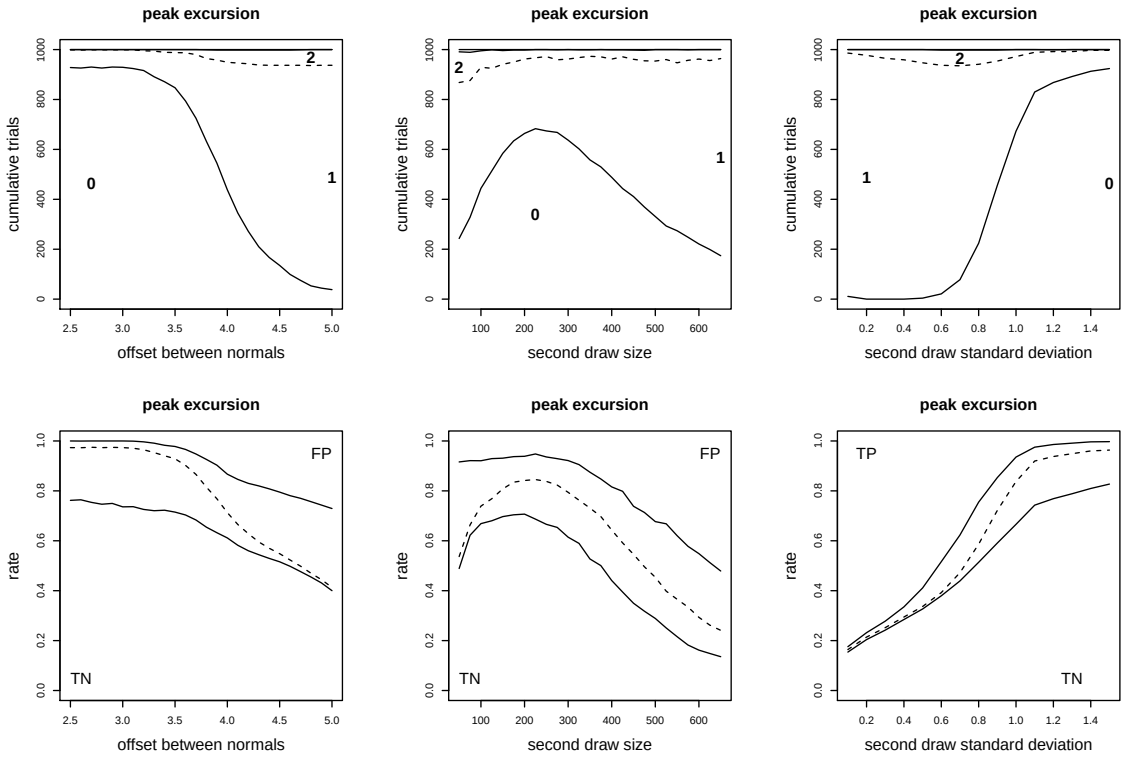



Figure 63: Number of significant peaks in the interval spacing per the excursion test (top, 0.05 level) and confusion matrix rates (bottom).

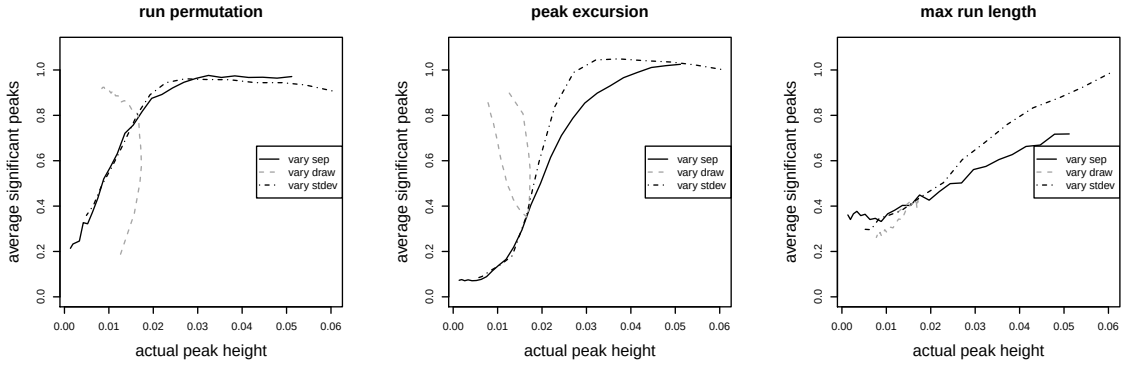



Figure 64: Average number of significant peaks for each distribution variation, expressed as the actual peak height from numeric integration of the spacing, for three tests.

and not just the actual feature height. Changing the total draw affects the bootstrap source, which will change how often features can be re-created.

The changed ranges of the draw variations remove the regions where the two draws cannot be distinguished, and compared to the low-pass results there are always two distinct minima (Figure 65). Still, some combinations of all variations produce a single significant flat, at the 0.05 level according to the excursion test. The transition to two significant flats occurs for the same separation as the low-pass test, below 2.6, but the flat in the peak between the two modes disappears. The raw contours remain distinct along both minima. Most detected flats pass as significant; this is also true in the draw size variations. But flats only occur in large second draws, above 400 points; those in the first mode for small draws have largely disappeared. Similarly, flats exist in the second draw when its standard deviation is below 0.8, and rarely appear in the first mode, and then not significantly.

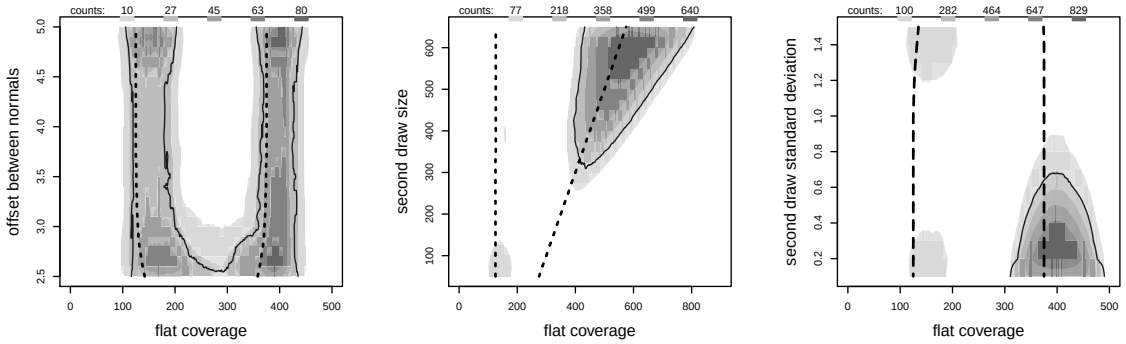



Figure 65: Spans of flats for draw variations. Heavy lines mark the expected minima in the spacing. Contour line encloses 75% of the significant flats per the excursion test.

Despite the correct location of the flats about the modes, the number that are accepted at the 0.05 significance level is much lower (Figure 66). The test is selective, but accurate. Up to 80% of the trials yield no significant flats. There is a high rate of accepting more than two flats, which when combined with the correct location means that more than one flat is appearing per mode. This happens at the largest imbalances in the draw size, or more weakly at small and large standard deviations away from the point where the modes can be distinguished. In other words, a well-defined mode seems to support multiple flats, and tweaks to the algorithm's parameters to detect longer flats or merge shorter may be needed. The confusion matrices are dominated by false positives, with true positives only appearing at more than a 20% rate when the standard deviation is very tight. The false negative rate is largest during the transition in mode detectability.

We see better agreement between the low-pass flat models and excursion test than we did for the peaks, but little to the interval excursion (Figure 67). The trend in the left graph is clear, with regression $R^2 = 0.664$ and 0.231. The slopes are much less than one, meaning the model is less sensitive, and points lying in only the first and second quadrant mean it is also more conservative. All model probabilities are above 0.05 and the disagreement is 100%. An agreement rate of 75% reflects only insignificant peaks. Spearman $\rho = 0.805$ and 0.427 shows the model results are associated with the excursion test at the 0.0000 level, and they order the peaks the same. The interval spacing is rougher than the low-pass due to the filter, and generates only a quarter the flats. Of these just 5% of the features, or 67, align, defined as overlapping by half their length. The interval excursion test accepts all flats; the rough signal makes any flat that appears seem unlikely. Spearman's $\rho = 0.253$ means the test results do not associate, at the $p = 0.26$ level, but they have a consistent order, at the

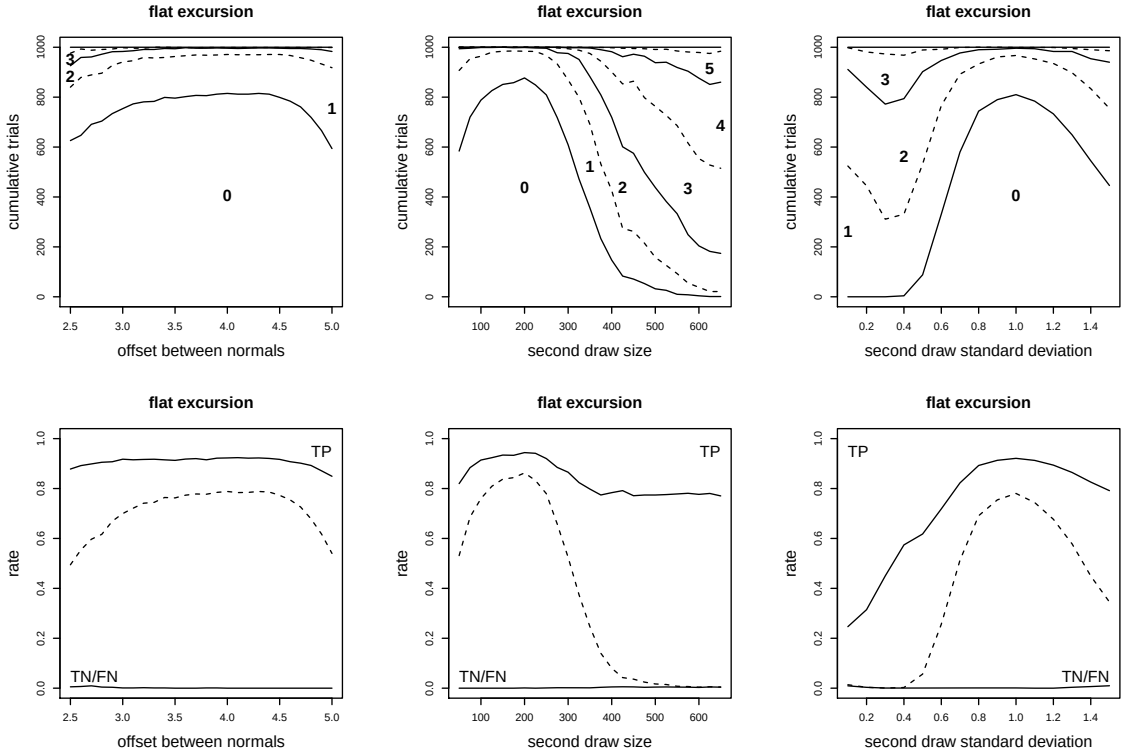



Figure 66: Number of significant flats per the excursion test (top, 0.05 level) and confusion matrix rates (bottom) in the interval spacing.

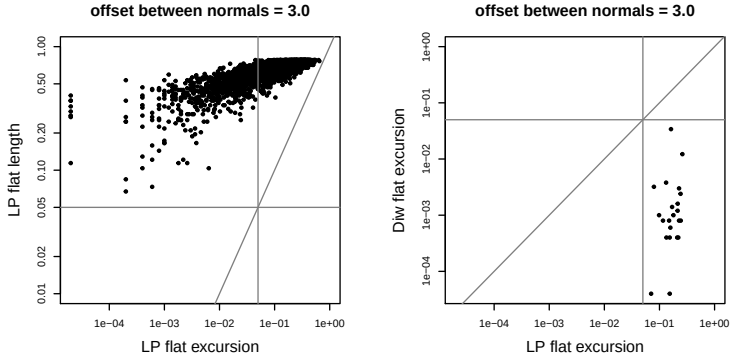



Figure 67: Probabilities of flats per length against low-pass excursion test (left), or of interval spacing excursion test (right), for draws from the two normals.

0.0000 level from the Wilcoxon test.

Table 31 counts the classification of each matching flat. The length and aspect ratio models are the same. The low-pass excursion test accepts 18% of the features, and all are rejected by the models. Against the interval excursion test we see the opposite behavior. Here the low-pass test rejects all but two matching flats, and the interval test accepts 90% at the 0.05 level.

Table 31: Flat Test Consistency

| | lp flat model | | | Diw excursion | | |
|---|---|---|---|---|---|---|
| lp excursion | 0.01 | 0.05 | insig | 0.01 | 0.05 | insig |
| insig | 0 | 0 | 4386 | 46 | 12 | 7 |
| 0.05 | 0 | 0 | 705 | 2 | 0 | 0 |
| 0.01 | 0 | 0 | 272 | 0 | 0 | 0 |

**Detail 32 Response of Signed Difference of Interval Spacing**

We evaluate the three tests available on runs in the signed difference of the interval spacing: the run height permutation test over the proposed feature; the Kaplansky-Riordan summary statistics of the number of runs; and the length of the longest run using a Markov chain model. We run these tests on the same draws as the excursion test in [Detail 31], and use them to evaluate peaks and not flats. In general they behave the same as the excursion test, with important differences in their ability to distinguish two normals as we vary the second distribution's parameters. They do not quickly decide on whether a peak is significant, compared to the low-pass filter tests, and do not come to a complete decision. An advantage of this low sensitivity is that the tests rarely accept more than one peak. The feature counts do not switch completely between no trials and all trails even when the curves plateau, so there is a bias in the test. Changing the acceptance level will shift this bias.

The run height test has a false positive rate of 20% at the 0.05 level when the two draws are not distinguishable, at separations below 3 or standard deviations above 1.2 (Figure 68). At the 0.01 level the curves shift vertically, so that 95% are rejected, but they then fail to accept 20% when the modes are well-defined. The transition from not recognizing a peak to accepting one occurs over a separation of 1.5, or a change in the standard deviation of 0.6, which is slower than for the excursion test. It occurs steadily as the size of the second draw increases, and again we see the trade-off when choosing the acceptance level. The dashed curves in the confusion matrix graphs show a change from false negatives when the modes are not well-separated, to false positives when they are. The false negative rate is stable over the standard deviation variations, while the false positive rate decreases sharply at the smaller widths, and over the draw size variations.

The runs statistics test (Figure 69) has the same 20% fixed acceptance rate when the draws cannot be distinguished, but an even slower change than the run height test as the separation increases. This produces a similar but softer switch from false negatives to false positives. Over the standard deviation variations the acceptance is sharper and the judgements include more false negatives, again at a consistent rate. The test is unaffected by draw size changes, and false negatives grow with the size of the second draw. Accepting features at the 0.01 level shifts the counts upward by 200 trials, trading off the false positives for more false negatives when the draws can be distinguished. The shape of the transition does not change, just the background acceptance rate.

The longest runs test performs essentially the same as the runs statistics. The counts are 10% lower but otherwise have the same form. The error classifications are also the same, as is the shift to

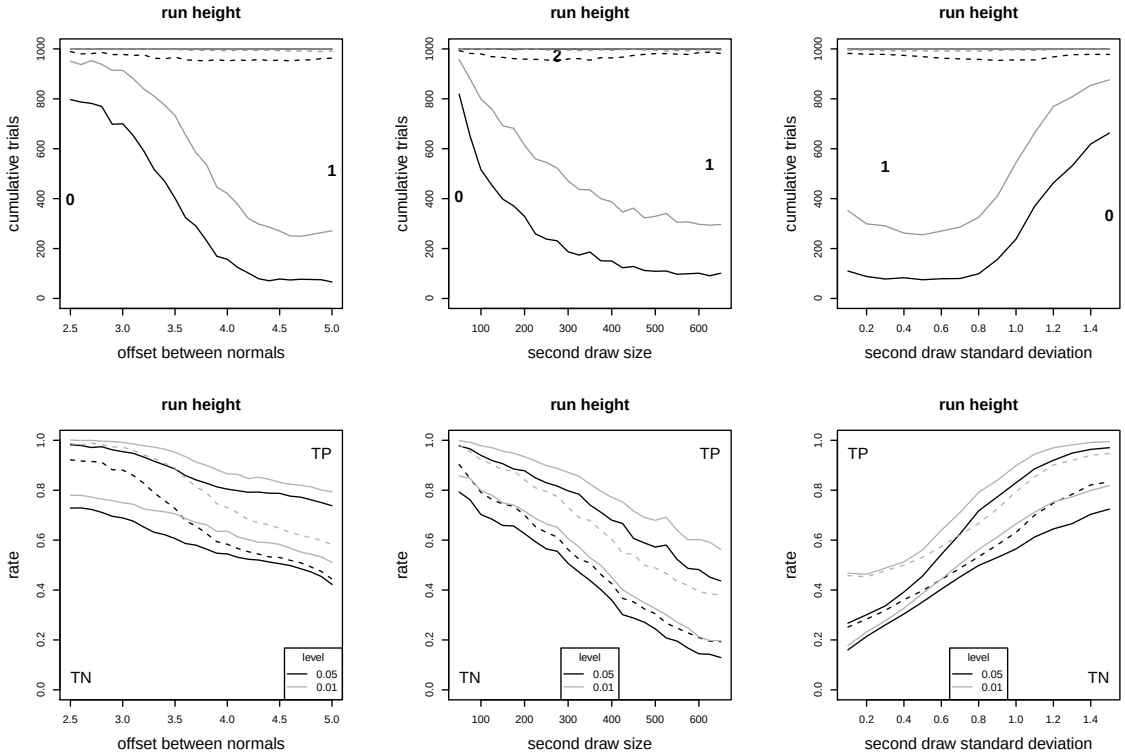



Figure 68: Number of significant peaks (top, 0.05 level in dark, 0.01 in grey) and confusion matrix rates (bottom) using the run height permutation test.

false negative predictions when judging at the 0.01 level.

The run permutation and peak excursion tests in the interval spacing do not follow the low-pass excursion test (Figure 70). The interval spacing contains twice as many peaks, 2188 against 1052, but only 466 align within 25 points. Neither test's probabilities match the low-pass excursion, with $R^2$ values of 0.008 and 0.006 showing no clear trend. The points in both graphs are above the 45° line, meaning the low-pass test is more liberal and that disagreements are found in the second quadrant. The run permutation test disagreement is 24%, the interval excursion test 68%. A $\rho = 0.034, p = 0.45$ for the permutation test and $\rho = 0.033, p = 0.47$ for the interval excursion test means neither is associated with the low-pass excursion test. The ordering of features from either test, however, is consistent, at the 0.0000 level by the Wilcoxon test. If the draw separation was increased to 4.0 we would start to see two clusters of peaks accepted by the low-pass excursion test. There would still be no trend between individual probabilities, only a horizontal gap without any peaks, much as a vertical gap appears in Figure 52.

The low-pass excursion test accepts almost all of the common peaks at the 0.01 significance level, but this is not true for the run height permutation and excursion test on the interval spacing (Table 32). The permutation test has a 75.9% agreement rate, but disagrees for 24.3% of the accepted low-pass accepted peaks. The interval excursion test's agreement rate is 32.9%, and its disagreement rate 67.8%. The low-pass excursion test has small disagreement rates, of 0.3% and 1.3%, so any significant peak in the interval spacing also passes if it appears in the low-pass spacing. However, this matching condition holds less than half the time.

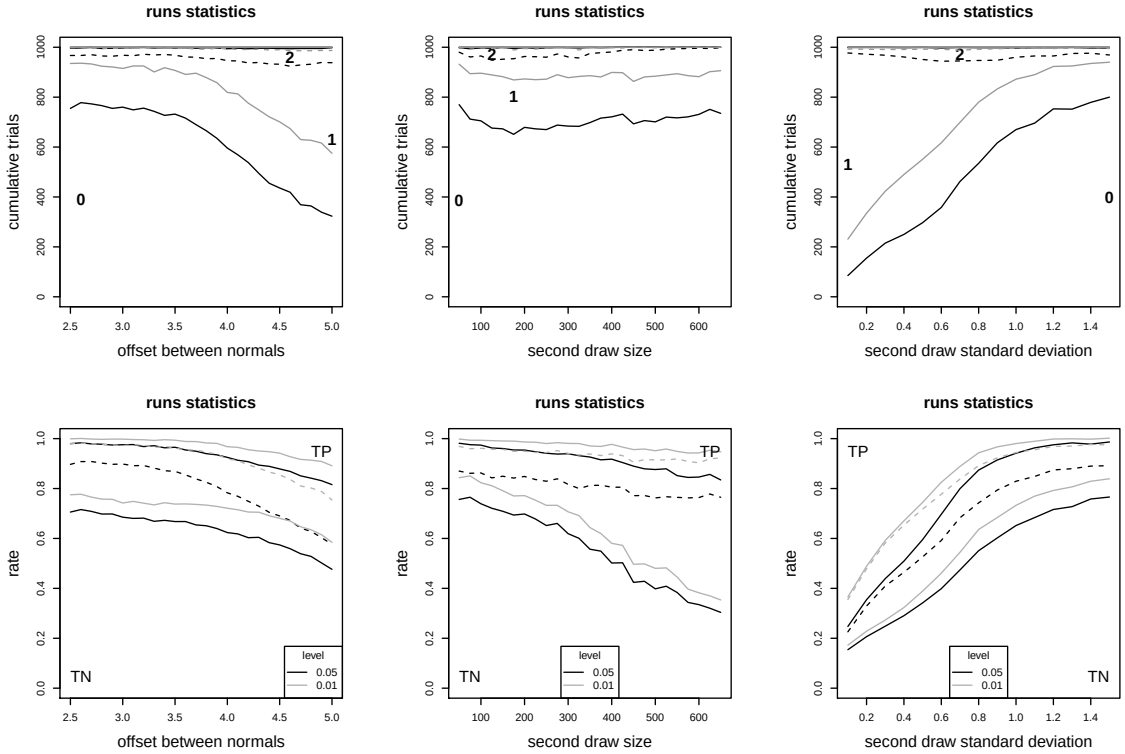



Figure 69: Number of significant peaks (top, 0.05 level in dark, 0.01 in grey) and confusion matrix rates (bottom) using the Kaplansky-Riordan runs statistics test.

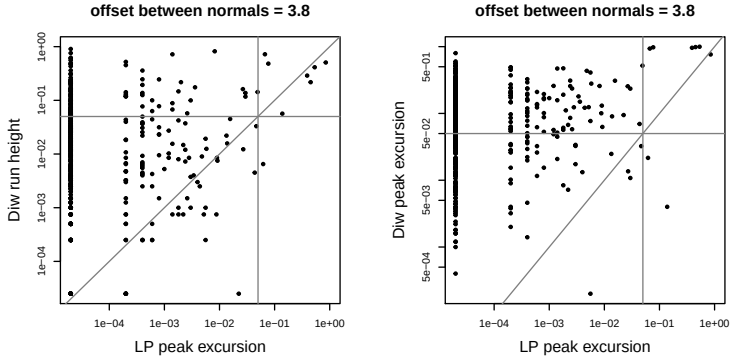



Figure 70: Probabilities of matching peaks from low-pass excursion test and interval spacing run height (left) and excursion (right) tests for the two normal draws. Lines mark 0.05 significance level and ideal match.

Table 32: Peak Test Comparison between Spacings

| | Diw run height | | | Diw excursion | | |
|---|---|---|---|---|---|---|
| lp excursion | 0.01 | 0.05 | insig | 0.01 | 0.05 | insig |
| insig | 1 | 0 | 7 | 1 | 1 | 6 |
| 0.05 | 2 | 5 | 4 | 0 | 4 | 7 |
| 0.01 | 225 | 139 | 115 | 49 | 105 | 325 |

There is no trend when comparing the peak probabilities from each test against each other (Figure 71). The $R^2$ of a regression on each chart is low, 0.09 for the runs statistics, 0.04 for the longest run, and 0.21 for the peak excursion test, although the Spearman test does indicate an association at the 0.0000 level, with $\rho = 0.362, 0.276,$ and 0.632 respectively. The Wilcoxon test also shows the probabilities of the tests are ordered the same, at the 0.0000 level. The agreement rate for the three tests ranges between 64% and 73%, while the disagreement with the permutation test is 48–68%, and the disagreement by the permutation test 24–38%. The counts for the runs statistics and longest runs tests in Table 33 are very close, but the tests compared to each other show no trend and have an agreement rate of 74% and disagreement rate of 60%.

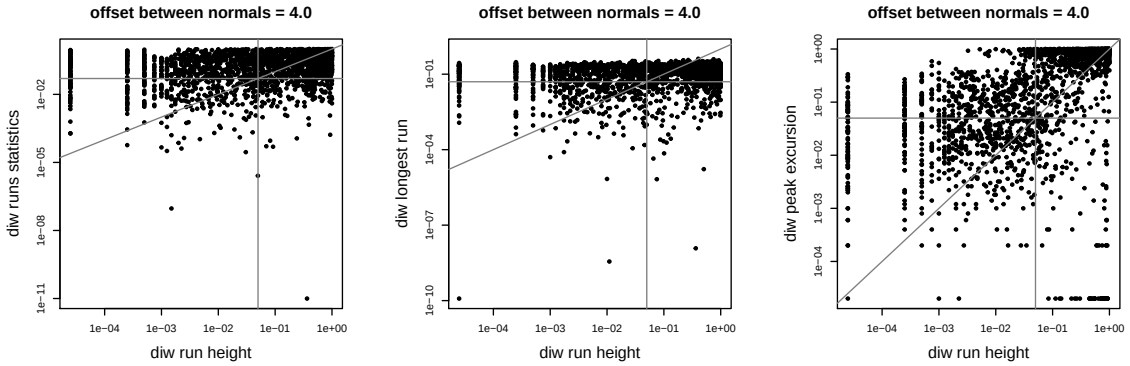



Figure 71: Probabilities of same peaks from run length statistics test (left) and longest run (middle) compared to height permutation test, and against a height excursion test (right). Lines mark 0.05 significance levels and ideal match.

Table 33: Run Length Test Comparison

| | run statistics | | | longest run | | | excursion | | |
|---|---|---|---|---|---|---|---|---|---|
| run height | 0.01 | 0.05 | insig | 0.01 | 0.05 | insig | 0.01 | 0.05 | insig |
| insig | 48 | 126 | 1075 | 53 | 121 | 1075 | 86 | 61 | 1102 |
| 0.05 | 41 | 45 | 219 | 35 | 66 | 204 | 48 | 87 | 170 |
| 0.01 | 97 | 108 | 382 | 66 | 123 | 398 | 129 | 204 | 254 |

### Detail 33 Changepoint and Level Section Alignment

It appears from the examples that some changepoints and the endpoints of level sections align. This seems to be true, although it depends on the stability of the changepoints. Consider a draw with three tight modes and a background to one side that creates a separate boundary, based on the N5 test data in [9]; there are in principle seven changepoints bounding the first three draws and at the separation.

$$250 \times \mathrm{N}(2,\, 3) \quad 50 \times \mathrm{N}(-2,\, \sigma) \quad 50 \times \mathrm{N}(-0.75,\, \sigma) \quad 50 \times \mathrm{N}(0.75,\, \sigma)$$

With a standard deviation $\sigma = 0.1$ the three modes form flats that have sharp transitions to the boundaries. These are easy to detect and the changepoints should be stable. At $\sigma = 0.3$ the modes still create a peak in the low-pass filtered spacing but their flats blend into the background, and the changepoint algorithms have trouble finding any difference.

Repeating the draw 500 times, the average number of changepoints drops, from 5.66 for $\sigma = 0.1$, to 4.24 for $\sigma = 0.2$, to 2.66 for $\sigma = 0.3$. The number of unique endpoints for level sections in the same trials goes from 15.1 to 14.6 to 10.1; this considers close endpoints from slightly overlapping sections to be the same. There are many more level sections than changepoints, forced in part by their covering the total sample and by the tendency of the algorithm to create short sections. This discrepancy means we must consider matching rates separately. Let us count the number of changepoints that are close, within 5 points, of any endpoint, and vice versa. Figure 72 plots histograms of the fraction of each that match, and Table 34 shows the average fraction, the most common, and the 90% confidence interval for each variant. At $\sigma = 0.1$ most changepoints align with the level sections and the matched endpoint rate scales by the 3:1 ratio between them, distributed more in a bell curve over the confidence interval. At $\sigma = 0.2$ the changepoints align less well with the level sections, with the mean fraction dropping to 0.60 and values spread between 0 and 1 and peaking at 0.5. The matching endpoint rate also shifts lower, with the mean still holding the 3:1 ratio. At $\sigma = 0.3$ the smaller number of changepoints leads to a sparse quantization of the fractions in the histogram, which are now evenly spread between 0 and 1 with a significant number of trials with no matches. The mean endpoint ratio is 4:1, the same as for the overall counts.

In summary, half to three quarters of the changepoints will match endpoints of level sections, but there will be three or four times as many endpoints. Level sections, if added to the voting procedure, might well end up excluded from the final vote based on this discrepancy and the $q_{vote}$ range.

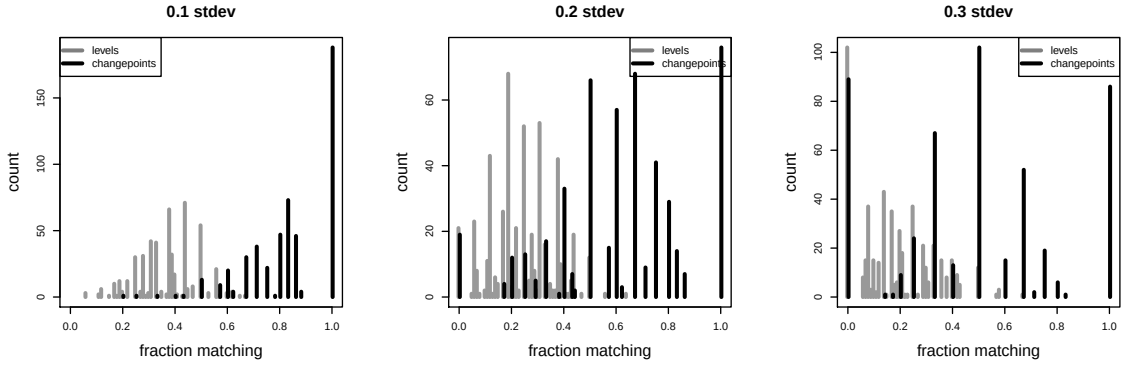



Figure 72: Frequency of fraction of changepoints or level section endpoints close to the other, within 5 points, as the standard deviation of the draws changes.

Table 34: Fractional Matching of Changepoints and Level Sections

| | changepoints | | | endpoints | | |
|---|---|---|---|---|---|---|
| $\sigma$ | mean | most common | 90% CI | mean | most common | 90% CI |
| 0.1 | 0.82 | 1.0 | 0.5, 1.0 | 0.34 | 0.4 | 0.14, 0.50 |
| 0.2 | 0.60 | 0.5 | 0.2, 1.0 | 0.21 | 0.2 | 0.00, 0.40 |
| 0.3 | 0.57 | 0.0 | 0.0, 1.0 | 0.15 | 0.0 | 0.00, 0.38 |

**Detail 34 Sensitivity to Tri-Modal Changes**

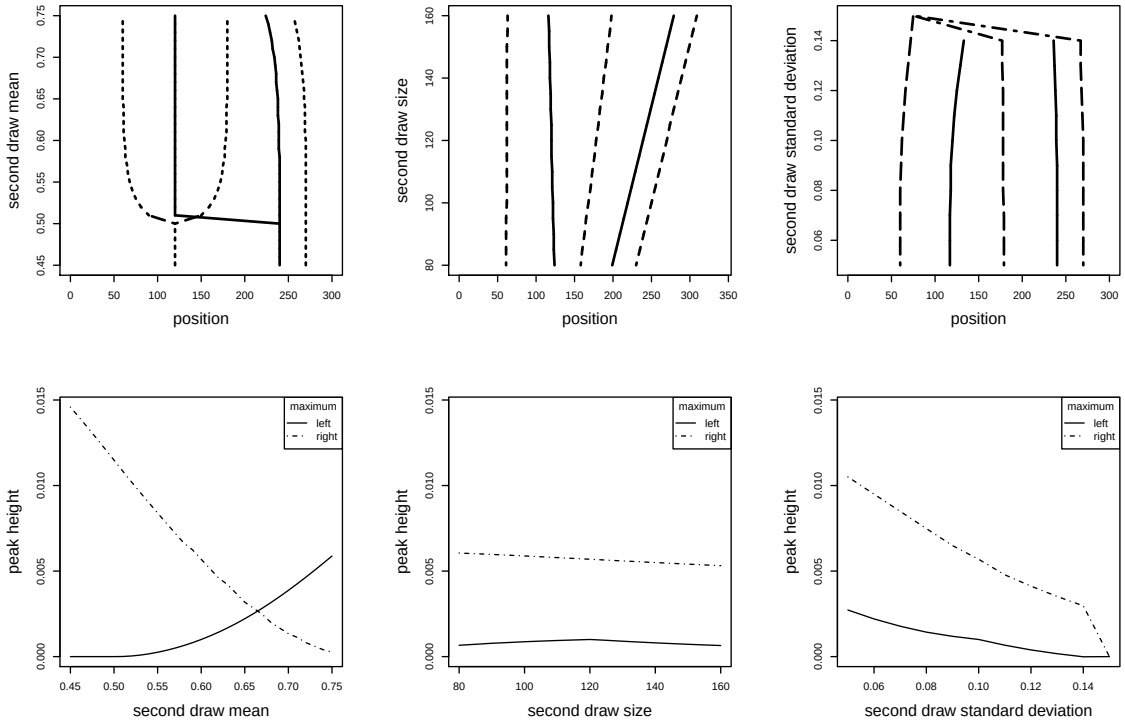



Figure 73: Location of minima (solid) and maxima (dashed) in tri-modal draw variants (top), and peak height at anti-modes (bottom).

This base setup uses three normal draws, two balanced and a small tight distribution to the side. Modifying literature sample F4 [16, Table 2], we double the size of the third draw so it is larger than the filter width and minimum flat length. F4 allocates draws in the ratio 9:9:2, which now becomes 2:2:1. Scaled to 300 points the base setup is

$$120 \times \mathrm{N}(0.3, 0.1) \quad 120 \times \mathrm{N}(0.6, 0.1) \quad 60 \times \mathrm{N}(0.9, 0.02)$$

The difference between the first two draws is marginal, while there is a clear break between the second and third. We will vary the parameters of the second draw. When its mean is 0.45 we expect the middle draw to overlap the first or lower variate, and at 0.75 the third or upper. Changing its draw size between 80 and 160 will lessen the peak height to either side as the number of points increases, while making room for a longer flat. Increasing its standard deviation from 0.05 to 0.15 will make the middle mode less distinct. Separating the first two draws is harder than the second pair. Figure 73 shows position of the modes (solid lines) and anti-modes (dashed) and the expected peak heights at the gaps, calculated as we did for the bi-modal variants. The solid line in the height graphs corresponds to the left anti-mode, the dashed the right. The first two variates cannot be distinguished if the separation is less than 0.50.

We will generate the sample 1000 times and use the same detector parameters as we did for the bi-modal experiments. We set $L_{abs} = 30$ and for the interval spacing $\delta_{ripple} = 0.02$.

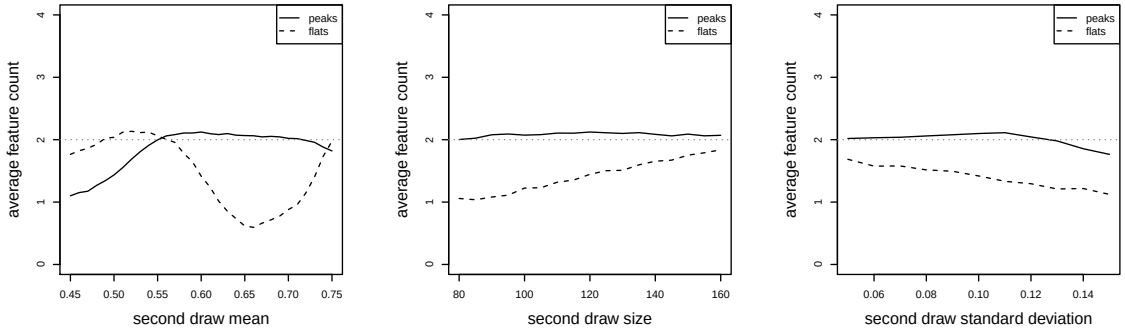



Figure 74: Number of features detected while varying the middle draw position (left), size (middle), and standard deviation (right). Averages are over 1000 repetitions.

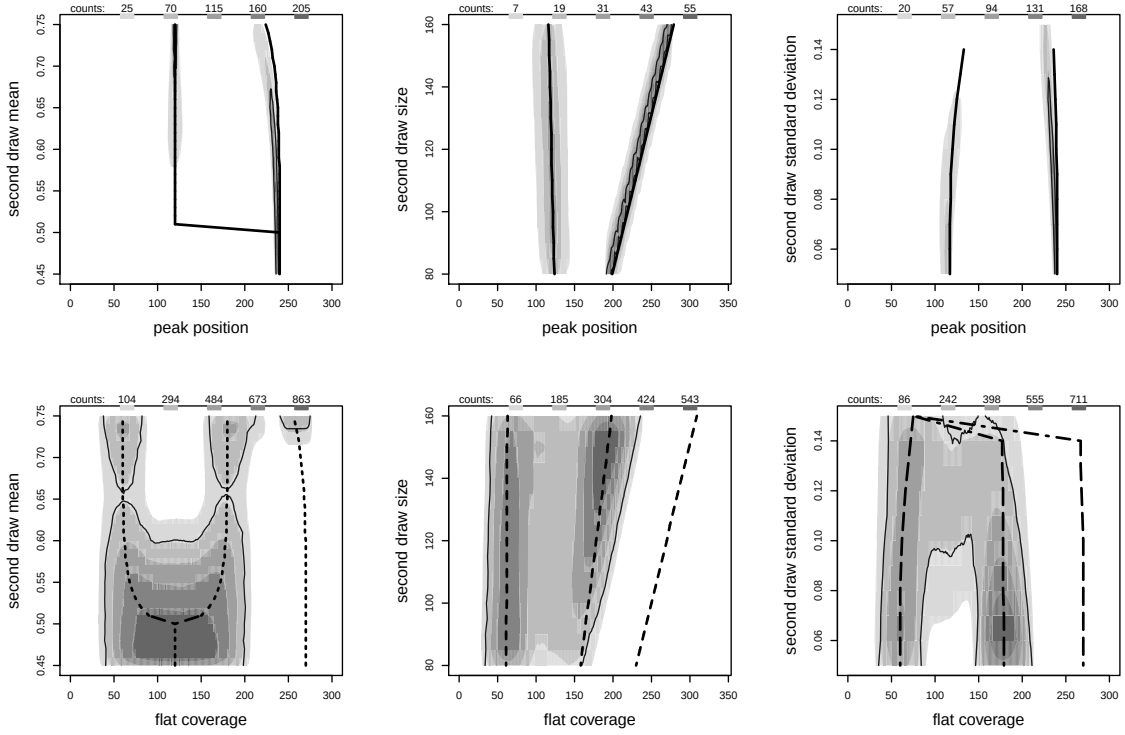



Figure 75: Count of peak locations (top) and flats spans (bottom) over second draw variations. Heavy lines are expected positions of spacing maxima (solid) and minima (dashed). Contour line encloses 75% of the significant features.

In general the tests find two peaks separating the three modes and two flats in the first two draws (Figure 74). Were we to loosen the length requirement we would find up to 5 flats. Plotting the position of the detected peaks (Figure 75), we see a strong feature between the second and third draws that weakens for large separations and widths. The peak between the first and second is stable over the second draw size, but disappears when the offset is less than 0.57 or the standard deviation

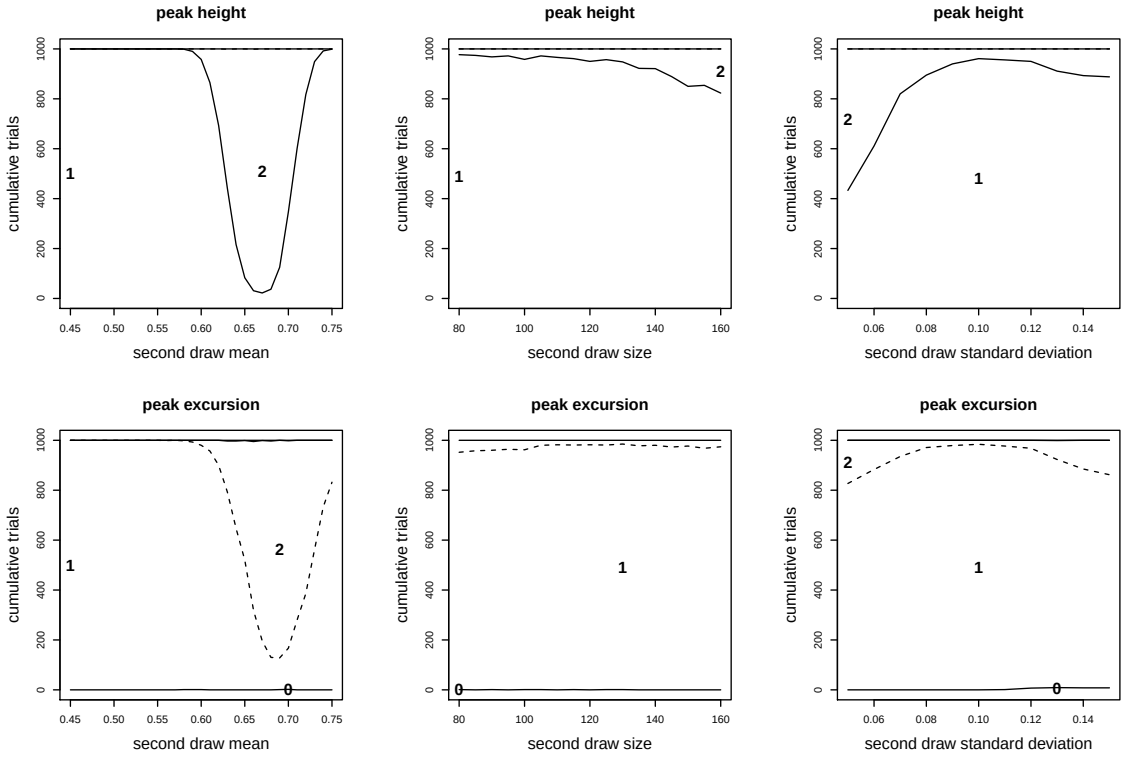



Figure 76: Number of peaks accepted by the height model (top) and peak excursion tests (bottom) at a significance level of 0.05.

above 0.12. The significant peak contour comes from the height model. It follows to the side of the right maximum, towards the central draw. A significant peak switches from the second to the first maximum as the separation increases, as expected, but none appear at the first maximum for the other two variations because the base setup cannot resolve it.

Flats follow the modes or minima, seen in the bottom graphs, where the excursion test provides the significant contour. The third draw does end up too small to support a flat except when the offset is above 0.72 and the peak moves into the second draw, giving more room in the third. The first and second modes have flats for separations above 0.63 and widths below 0.10. On the other side of these values the two draws merge and generate one flat spanning both, even if during the transition the detected features still follow both modes. Actually, we will see that there are multiple slightly overlapping flats in this region that average over all runs and appear as one long feature. Although the detected flats cover the modes as the draw size changes, the significance contour from the excursion test covers both. Once again this does not mean that there is one significant flat, but that the endpoints spread out over the middle so the contour cannot distinguish the two flats that actually exist.

Both positional figures show that detected features align to modes and anti-modes. The features locate modality changes.

The acceptance rates of the height and peak excursion tests are similar (Figure 76). They consider the peaks at each anti-mode significant at a mean of 0.68, although the excursion test is a little more

conservative, rejecting one peak in 10% of the trials and requiring a slightly larger mean. It does not accept two peaks for the draw size variations beyond a 5% rate, while the height models accept up to 20% for the largest second draw. It is slower to accept a second peak as the standard deviation drops below 0.08. The peak between the second and third draws is strongest and accepted by both tests. The difference between one and two peaks comes from the evaluation of the first anti-mode. The significant peaks are not consistently predicted by the peak height. The region of two peaks in the separation variants corresponds to a height of 0.00265, but this does not capture the occasional second peak with draw sizes above 120 or standard deviations above 0.12. A significant peak is lost at separations above 0.70, when the second pair of draws loses definition.

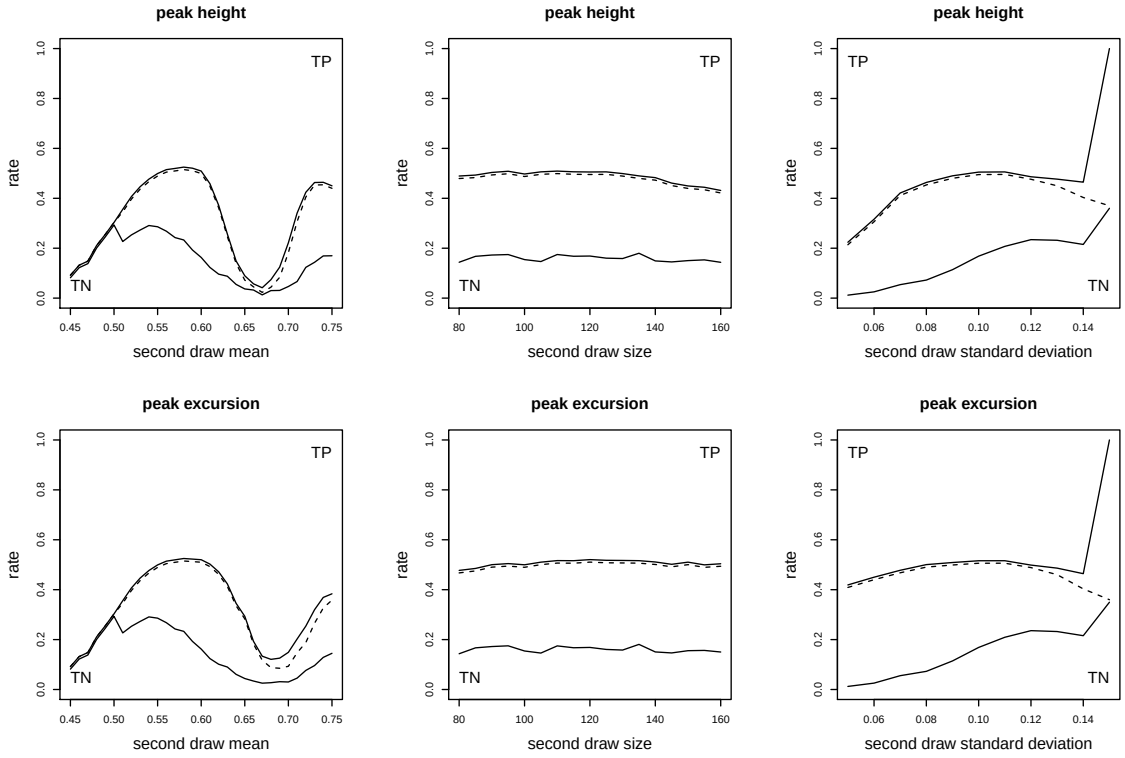



Figure 77: Confusion matrix rates for peak height model (top) and excursion tests (bottom) while varying second draw.

Assessing accuracy as described in [Detail 30], both the height model and excursion tests (Figure 77) produce very few false positives, or peaks they consider significant but are not aligned with their expected position. The false negative rate, where detected peaks do have their expected position but are not considered significant, is roughly 30%. It disappears at moderate means when there is enough separation between the draws that both peaks are found.

The flat model picks up the long flat when the first two draws cannot be distinguished for means below 0.57, but otherwise rejects any detected features (Figure 78). Their accuracy is therefore split between true and false negatives. High true negative rates, up to 40% for separations smaller than 0.50 or standard deviations above 0.10, represent the correct rejection of small flats that do not include a minimum. The high false negative rate includes short flats that do overlap a mode.

The excursion test behaves quite differently. It has the fewest significant flats when the draws are

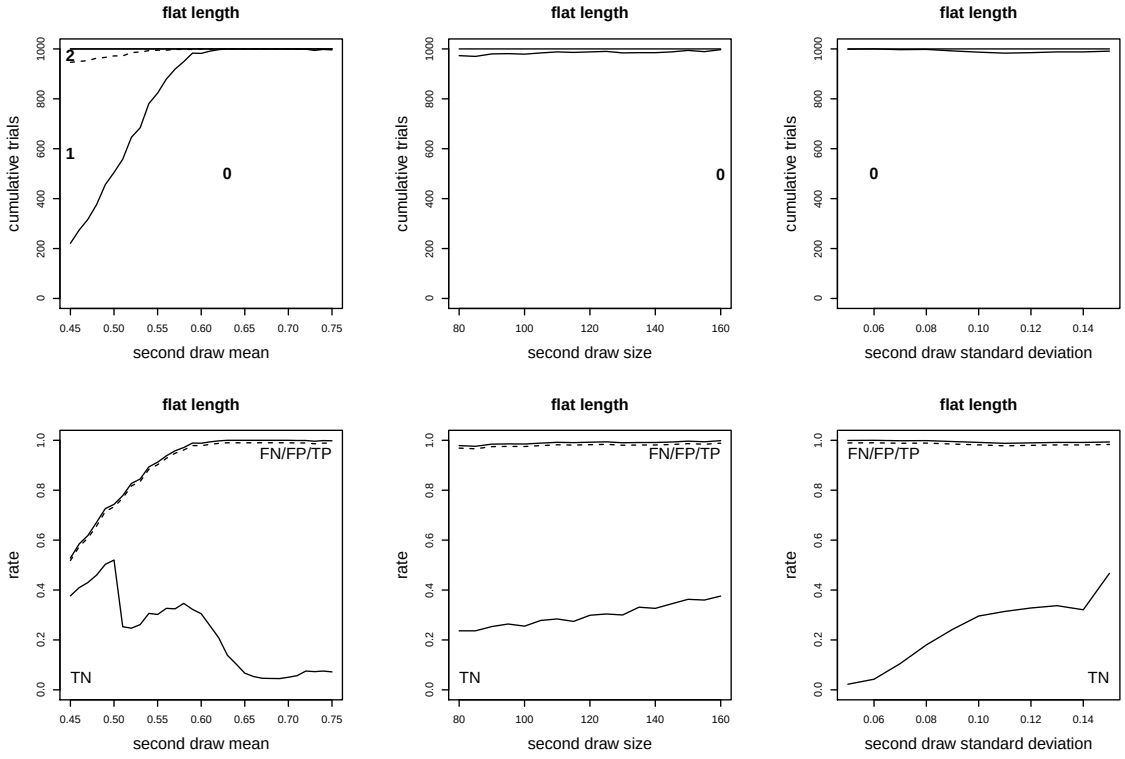



Figure 78: Number of flats per length model (top, 0.05 level) and confusion matrix rates (bottom).

best distinguishable, for means between 0.6 and 0.75, but this reflects the smaller number of detected flats (Figure 79). There is no evidence of flats being accepted for means below 0.55, where the model starts passing features because of the long length; we see instead more rejection for means below 0.50. This leads on average to the lower count seen in Figure 74. The flats also overlap in general, which is smoothed over the trials into the uniform count in Figure 75. The test accepts steadily more flats as the draw size increases or the standard deviation decreases. Flat counts above one where the contour lines in Figure 75 span the first two draws mean there are two or three shorter features within the span, as we saw with the aspect ratio at small offsets. Here we see small flats appearing as the second draw grows or when the second width is larger than 0.10. Only when the offset is above 0.65 or the standard deviation is below 0.10 do flats represent the modes. Because the excursion test does accept many of the proposed features, false negatives become false positives. The small fraction of flats it rejects are correctly positioned, and account for the false negatives as the first two modes separate, for means above 0.80. In the opposite direction, the accepted short flats away from the modes that are marked true negatives by the models become false positives, at a rate up to 50%.

In summary, the detector finds many short flats when two modes are hard to distinguish, at small separations or large width or any draw size. One may cover a minimum, but others will not. Because the length model rejects most of these features the two groups translate into false and true negatives. But the excursion test accepts most, and so they become true and false positives.

The run height permutation test accepts more peaks than the excursion test, and their position follows the expected location (Figure 80). The peak between the second and third draw is strong and significant for all variants. When the first and second draws can be distinguished, for a mean

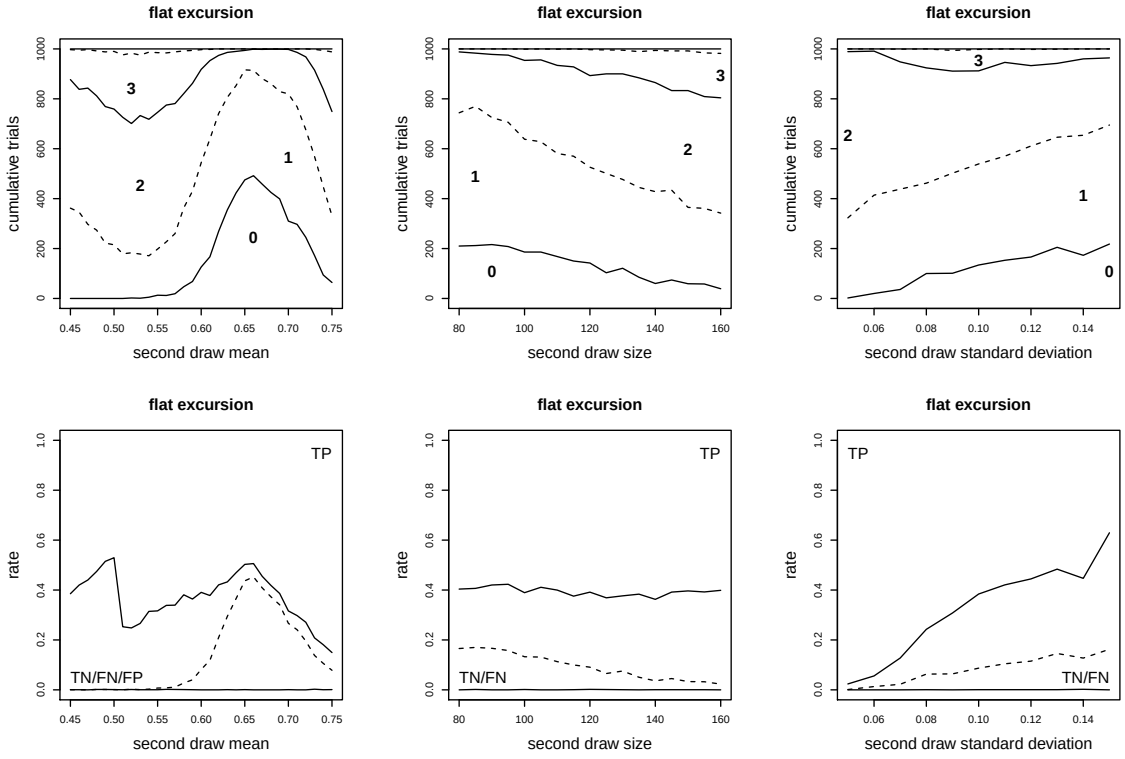



Figure 79: Significant flat counts by the flat excursion test (top, 0.05 level) and confusion matrix rates (bottom).

above 0.67 or standard deviation below 0.08, a peak appears and is accepted as significant, although it moves towards the second draw. When the first pair cannot be distinguished, the peak's location washes out, even into the modes. This also happens for all draw size variations. We would not want to relax the acceptance level to avoid passing such features. Flats are very limited, appearing when the first two draws merge, or when the second draw is very tight with a width below 0.7, or at the extremes of the draw size. The accepted features follow the detected.

### Detail 35 Feature Position in Datasets from the Literature

The literature contains a large number of examples that have been used to evaluate different approaches to analyzing modality. Many are created to be difficult to resolve, with small modes against a large background or with little separation. We run the spacing tests on these distributions to evaluate the number of features found and their position. Because the samples are known combinations of random variates, mostly normally distributed, we can determine from the ideal density functions the actual location of the modes and the gaps or anti-modes between them. In practice this is done by evaluating the total density on a very fine grid and using the peak detector, rather than numerically solving for the extrema. We generate each sample 200 times and find and evaluate the features in the low-pass and interval spacing, the changepoints, and the level sections. The test setup follows [Detail 28], with default parameters for the feature detectors and changepoint voting algorithm. Flats must be at least 30 point long. Significance of peaks and flats is at the 0.01 level. The samples are

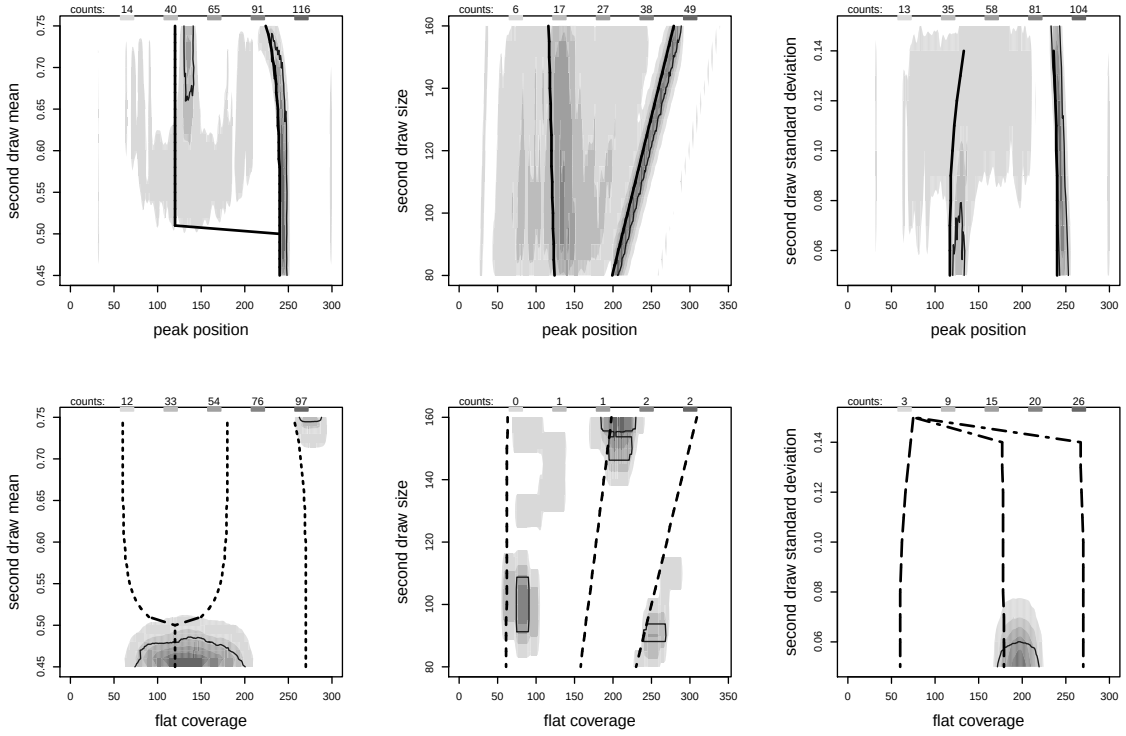



Figure 80: Interval spacing peak locations (top) and flat coverage (bottom) over second draw variants. Heavy lines mark the expected position. Contour line encloses 75% of the significant features.

mostly defined with draw sizes for individual variates as fractions of the total draw, which we set to 200 points but increase if needed to give 30 points to the smallest variate. Note this is the same as the minimum flat length, so we cannot expect to find these features in such variates, even if their distribution has a small width, which it usually does; we probably should have set a target of 40 points, or allowed shorter flats. We aimed for 200 points because this seems to be closer to the size of real-world data than a much larger sample, despite the potential loss of flats.

For each feature we store its position and significance. We test peaks and flats in the low-pass spacing against the models and as excursions. Excursion tests are also used on the interval features, with additionally three tests on the signed runs for peaks, the Kaplansky-Riordan statistics, the maximum run length from a Markov chain model based on the trial sample, and a permutation test on the peak height. Changepoints and level sections do not have a significance level. When comparing the position against the ideal location, we convert from an index to a value using the cumulative density from that trial, picking the value in the sorted raw data at that index. The actual draw therefore imposes some inaccuracy in the conversion, particularly if there are large gaps between the variates; a better strategy would use the mid-point of the cumulative density [27]. This happens in samples W14, A19, and A20.

In this detail we present an overview of the results, in the form of graphs visualizing the position of the features. [Detail 36] compares the location of the features against the known, ideal positions. [Detail 37] looks at the number of each feature found, and how often the low-pass and interval spacing agree. The data used for all three details is the same. Only its analysis differs. [Detail 38] summarizes

the results by the complexity of the samples. It introduces several new metrics to compare the different approaches.

The samples are grouped by the number of distinct modes they contain, from one uni-modal density to five in penta-modal setups. The number of modes may differ from the number of variates if they are placed to provide background noise or to alter the tails of the total draw. The bi-modal group is further divided into asymmetric (abi-) and symmetric (bi-) samples. The other classes do not make this distinction.

The D series is defined in [32], excluding the uni-modal cases. The M series comes from [30]. M1, M2, and M3 are bi-modal normal draws defined in Table 1, and M4 is a tri-modal normal draw in Table 2. The F samples come from [16, Table 2], with F1 labeled Contaminated, F2 Small Blip, F3 Asymmetric Bimodal, F4 Trimodal, and F5 Four Modes. F1 and F4 have 300 points, F2 600, and F5 800. [7, eq. (3.1), (3.2), and (3.3)] define the C draws, with 800, 400, and 500 points respectively. All are mixtures with a small second normal draw added as a shoulder to the main normal. All appear uni-modal. The H series is found in [20, Section 3] as alternatives a, b, c, and d. [9, Section 3.1] defines the N series, multi-modal distributions that use normal variates, and in Section 3.7 P1, based on Poisson variates. N4 draws 2000 points. The X series are the multi-modal entries in [51, Table 1], starting in the fourth row and numbered in order. X6 has 296 points and X7 300. G1, from [14], is the GNN example in the main text. [29, Table 1] defines the W examples, with numbering following the table entries. W5, W9, and W10 draw 300 points. [1] defines the A data, with 'A' used in the name instead of 'M'. Draws A1 through A10 are unimodal. A11 through A20 are bi-modal, although there are often additional draws added for shoulders and other discontinuities; these have been carefully placed for smoothness. A21 through A25 are tri-modal. A14, A22, and A23 draw 300 points, A13 600, and A20 1600.

The examples in the literature include only a few with four or five draws. In order to get a better idea of how the spacing behaves in these cases and its stability, we define several new examples. In general the combined density shows well-separated modes that should be detected as peaks in the spacing, but this does not mean the features will be found significant, especially at the 0.01 level. Often flats will not be seen, despite all individual draws containing at least 30 points, the minimum flat length. The modes are sharply defined, enough to support changepoints and level sections. Samples K1 through K9 have four modes, K10 through K15 five (Table 35). Using the notation from the main text, they are built from normal variates N($\mu$, $\sigma$) that have mean $\mu$ and standard deviation $\sigma$, uniform variates U($a$, $b$) that have density $1/(b-a)$ between $a$ and $b$, binomial variates B($N$, $p$) that have success probability $p$ for $N$ trials, gamma variates G($r$, $\lambda$) that have shape $r$ and rate $\lambda$, and Weibull variates W($a$, $b$) that have scale $a$ and shape $b$. Plots of each of these densities accompany those of the literature samples in this detail.

The graphs we use to track feature positions are greyscale-encoded histograms. At each index of the sample we count the number of features over the 200 trials, either the point location of peaks and changepoints or if the index is between the start and end of flats. We plot dots whose size and darkness reflects the counts. Each line in the graph is encoded separately, so that black corresponds to each maximum count and not some absolute scale. Level sections are not included in the graphs because they span the entire sample and would appear as a solid dark line. The graphs count the number of significant figures in the low-pass (LP) and interval (Diw) spacing, using the best test result for each. Diw feature positions are shifted by half the interval to align with the filter. Underneath the two peak histograms, a second unlabeled line counts the detected features. This allows us to see which pass testing, or the selectivity of the tests. Some peaks may be reliably detected but rejected, while others, presumably better defined in the data, are robust to testing. The rejection rate of flats is much lower and we do not include separate plots for the detected flats. Vertical solid lines in these

graphs mark the modes, and dashed lines the gaps or anti-modes.

Table 35: K Sample Definitions

| | | | | | |
|---|---|---|---|---|---|
| K1 | 60 N(−1.5, 0.25) | 140 N(0, 1) | 140 N(4, 1) | 60 N(5.5, 0.25) | |
| K2 | 50 N(−5, 0.5) | 50 N(−2.5, 0.75) | 50 N(1, 1.25) | 50 N(4.5, 1) | |
| K3 | 75 N(−2, 1) | 50 N(−0.75, 0.2) | 50 N(0.75, 0.2) | 75 N(2, 1) | |
| K4 | 75 N(−2.5, 1) | 50 N(−0.75, 0.2) | 50 N(0.75, 0.2) | 75 N(2.5, 1) | |
| K5 | 100 N(−2.5, 1.5) | 50 N(−0.75, 0.3) | 50 N(0.5, 0.2) | 75 N(2, 1) | |
| K6 | 45 N(0, 0.2) | 30 N(1, 0.15) | 55 N(2, 0.2) | 40 N(3, 0.25) | 130 U(−1.5, 4.5) |
| K7 | 40 W(2, 0.75) | 100 W(4, 7) | 50 W(5, 2) | 60 W(8, 4) | |
| K8 | 60 N(−2, 2) | 40 W(2, 1) | 40 N(7, 1) | 60 G(14, 4) | |
| K9 | 60 B(300, 0.04) | 40 B(300, 0.12) | 60 B(300, 0.20) | 40 B(300, 0.29) | |
| K10 | 60 N(−1.5, 0.25) | 140 N(0, 1) | 50 N(2, 0.25) | 140 N(4, 1) | 60 N(5.5, 0.25) |
| K11 | 70 N(−5, 0.2) | 45 N(−3.5, 0.5) | 45 N(−1, 0.75) | 45 N(2, 0.75) | 50 N(5, 1) |
| K12 | 50 N(−3.5, 0.2) | 50 N(−2, 0.5) | 50 N(0, 0.75) | 50 N(2.5, 1) | 50 N(5, 1.25) |
| K13 | 55 N(0, 0.2) | 30 N(1, 0.15) | 60 N(2, 0.25) | 40 N(3, 0.2) | 50 N(4, 0.15) |
| | 115 U(−1.5, 5) | | | | |
| K14 | 60 B(200, 0.05) | 40 B(200, 0.15) | 60 B(200, 0.25) | 50 B(200, 0.35) | 40 B(200, 0.45) |
| K15 | 75 N(−5, 1) | 50 N(−3, 0.2) | 100 N(0, 2) | 50 N(3, 0.2) | 75 N(5, 1) |

When reading the graphs, we are looking for points that group around the gaps and flats at the modes without extending to the gaps. Features may shift from these expected positions if the sample's variations have unbalanced draw sizes or large differences in their width. Small dots mean a feature is stable, while trails indicate a weak or spurious one. We expect more dispersal in the detected rates, as many of these proposed features are rejected by testing. These graphs cannot distinguish individual flats, so counts smeared horizontally may represent short overlapping intervals or long features. In general, there will not be flats in the interval spacing because it is usually too rough. Concentrations of points at the edges of the plot, remembering that 10% to 15% of the sample is outside the filters, indicate spurious features in the tails of the total draw, those in the arms of the U.

Figure 81 plots densities of the 19 bi-modal symmetric samples, and Figure 82 the features. The separation between the two variates is the key variable for resolving them. M1 and M2, separated by 3.0 and 4.0 respectively, have been studied in [Detail 30] and [Detail 31], where we saw the second could be cleanly resolved. This analysis bears that out. The position of peaks and changepoints are much more variable in M1 than M2. The changepoints and interval peaks move from the tails in M1 towards the gap, but do not stabilize as well as the low-pass peaks. The low-pass flats also span the gap in M1, while in M2 they are cleanly separated within each mode. Sample H1 is the same as M1, so differences between the two results give us an indication of the stability of the tests over the 200 runs. In Figure 82 H1 falls between M1 and M2, with a little less spread in the low-pass flat and the beginning of a move by the changepoints off the tails, but the peak behavior does not change. X1 has a wider separation than M1 but also larger standard deviation. The two effects mostly cancel, and its analysis resembles H1.

Some other differences can be seen. Flats in the interval spacing, when they occur, are shorter than their low-pass cousins. They share the same position. Flats in general fall within the modes, except for D4 where the separation is so small that the two merge and the flat covers the gap. Peaks in the low-pass spacing are better defined, with less dispersion and smaller dots, than the interval peaks. The mode is often detected in the interval spacing, for example in D4 and W6, but there is a large variation in the position of such peaks and they do not always survive acceptance testing. Indeed, the difference between the detected and significant counts shows that testing does not pass interval

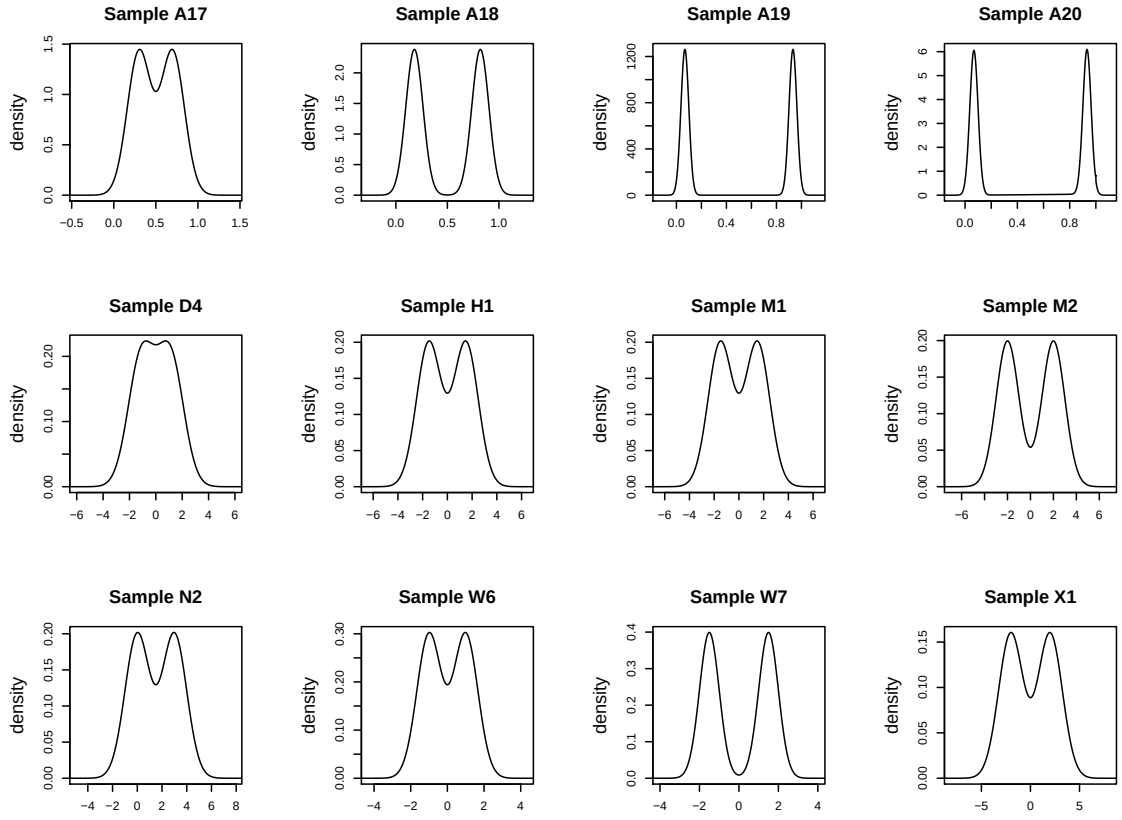



Figure 81: Symmetric bi-modal samples.

peaks uniformly, with those at the gap in D4 and W6 rejected while peaks at the tails pass. The difference between the detected and significant counts of the low-pass peaks is much more consistent.

Overall, A18, A19, A20, M2, and W7 are cleanly resolved. Table 36 counts the number of significant features that are well-defined. Underlined counts are marginal, meaning they are not completely stable. All told, the low-pass spacing finds the anti-mode in 8 of 12 samples (2 marginally) and the interval spacing 4. Changepoints match the anti-mode in 4 samples. The low-pass spacing has significant flats in both modes of 8 samples (3 marginal) and a flat in one mode in another. The interval spacing places flats over both modes in 2 samples, and over one in 1.

Table 36: Clean Bi-Modal Feature Counts

| | | A17 | A18 | A19 | A20 | D4 | H1 | M1 | M2 | N2 | W6 | W7 | X1 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| peak | LP | | 1 | 1 | 1 | | 1 | | 1 | 1 | | 1 | |
| | Diw | | 1 | 1 | 1 | | | | | | | 1 | |
| CPT | | 1 | | 1 | 1 | | | | | | | 1 | |
| flat | LP | | 2 | 2 | 2 | | 1 | | 2 | 2 | 2 | 2 | 2 |
| | Diw | | | 2 | 2 | | | | | 1 | | | |

*underlined counts are marginal*

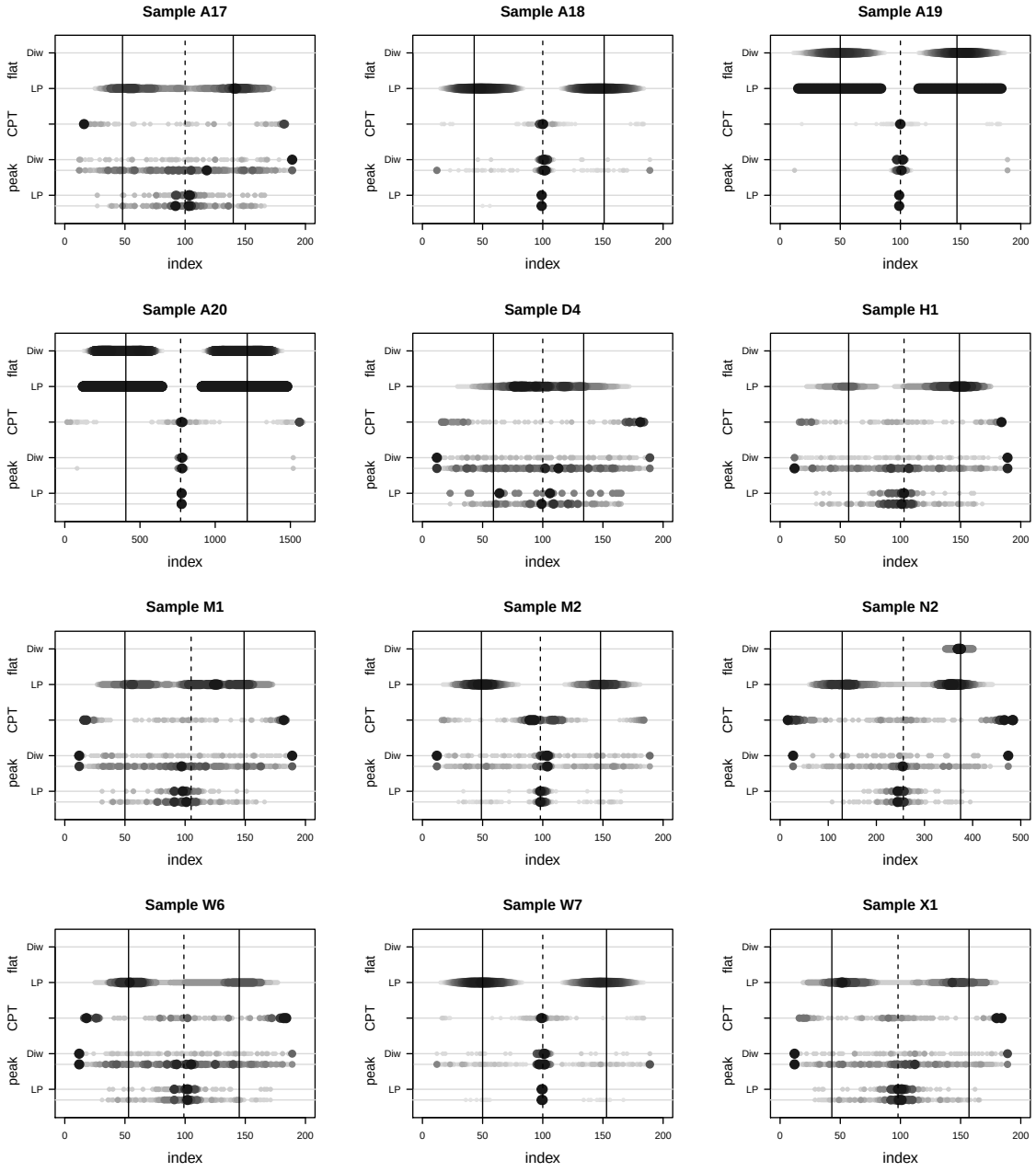



Figure 82: Feature positions in the bi-modal samples. Darkness and size of points correspond to frequency of position/span. Second lines for the LP and Diw peaks are detected rates. Vertical solid lines mark modes, dashed gaps.

A larger group of samples have two modes that are not laid out symmetrically (Figure 83). Many combine two normal variates of different widths, which creates unequal tails at the start and end of the total draw. A13, D5, M3, and X4 are examples with very large differences. Many also have an imbalance in the draw size; for example, A13 and F2 place a small variate deep in the tail of the main draw. Other samples such as A12 and A15 add background variates to extend the tails.

Overall, the point features reliably find the gap in samples A13, A16, D5, and F2 (Figure 84). F3, H2, H3, and X4 are not as clean but are generally correct for the low-pass peaks and changepoints; the interval spacing is much less stable. The asymmetries in the samples prevent flats from identifying both modes, except for A14 and marginally A15 and W8, but one is cleanly marked in most of the others. This is usually the larger draw, although in X4 and F3 it is the tighter variate.

Peaks detected in the interval spacing often mark the gap, for example in H2 and X2, but there is a high background spread over the entire draw. Testing mostly rejects those peaks we would like to keep, while those in the tails survive, for example in samples A15, D7, H4, M3, and X3. Changepoints do better, although the accuracy may be off or unstable, spread over a range of indices, as seen in A14 and H2. If they fail, they tend to mark the tails of the total draw rather than interior indices. Small draws define the upper tail of samples A13 and F2, and these are lost to the low-pass filter because there are too few points to create a local maximum. The interval spacing does identify these peaks, as do the changepoints. Testing does a good job passing the low-pass peaks that are correctly placed at the gap, except for A15 and H4.

Flats in the interval spacing are a subset of the low-pass. In sample D7 they lie in the first mode, which begins at index 1. The low-pass flats extend over or include the gap in samples A11 and X3, and fail to separate the variates in H4. They weakly mark the second mode in A12, D7, and X4, at a much lower rate than the dominant variate.

Table 37 gives the number of features accurately found. The low-pass tests correctly identify the anti-modes in 10 of 18 samples (2 marginally) while the interval spacing peaks succeed in 6 (1 marginal). Changepoints mark the gap in 10 samples (3 marginal). Significant low-pass flats cover both modes in 2 samples (1 marginal) and one mode in 12 (3 marginal). The interval spacing places a flat over a mode in 8 samples (1 marginal).

Table 37: Clean Asymmetric Bi-Modal Feature Counts

| | | A11 | A12 | A13 | A14 | A15 | A16 | D5 | D7 | F2 |
|---|---|---|---|---|---|---|---|---|---|---|
| peak | LP | | 1 | | 1 | | 1 | 1 | | |
| | Diw | | | 1 | 1 | | 1 | 1 | | 1 |
| CPT | | | 1 | 1 | 1 | | 1 | 1 | | 1 |
| flat | LP | | 1 | 1 | 2 | | 1 | 1 | 1 | 1 |
| | Diw | | | 1 | | | | 1 | 1 | 1 |

| | | F3 | H2 | H3 | H4 | M3 | W6 | X2 | X3 | X4 |
|---|---|---|---|---|---|---|---|---|---|---|
| peak | LP | 1 | 1 | 1 | | 1 | | 1 | | 1 |
| | Diw | | | | | | | | | 1 |
| CPT | | 1 | | 1 | | | | 1 | | 1 |
| flat | LP | 1 | 1 | 1 | | 1 | 2 | 1 | | 1 |
| | Diw | 1 | | 1 | 1 | | | | | 1 |

*underlined counts are marginal*

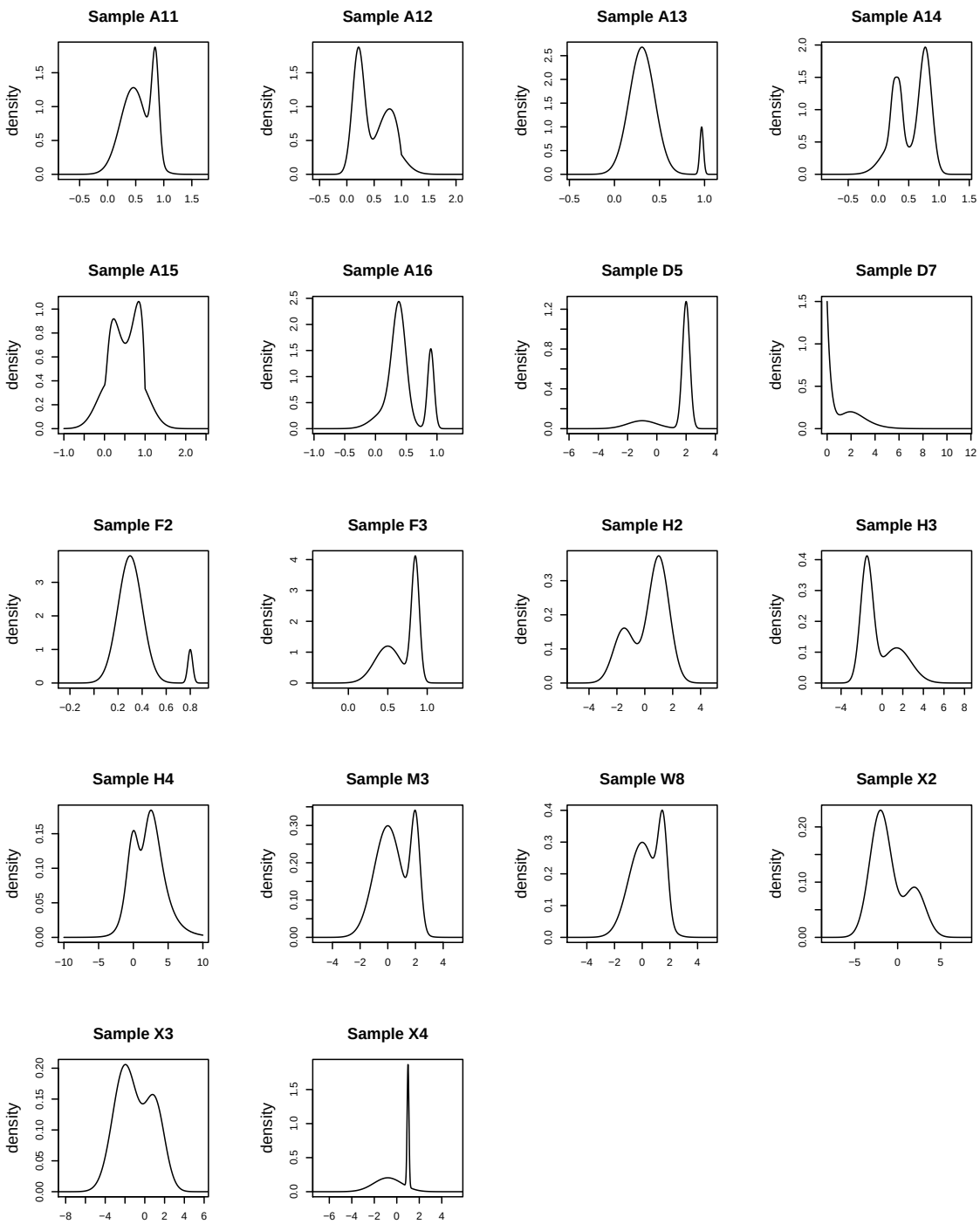



Figure 83: Asymmetric bi-modal samples.

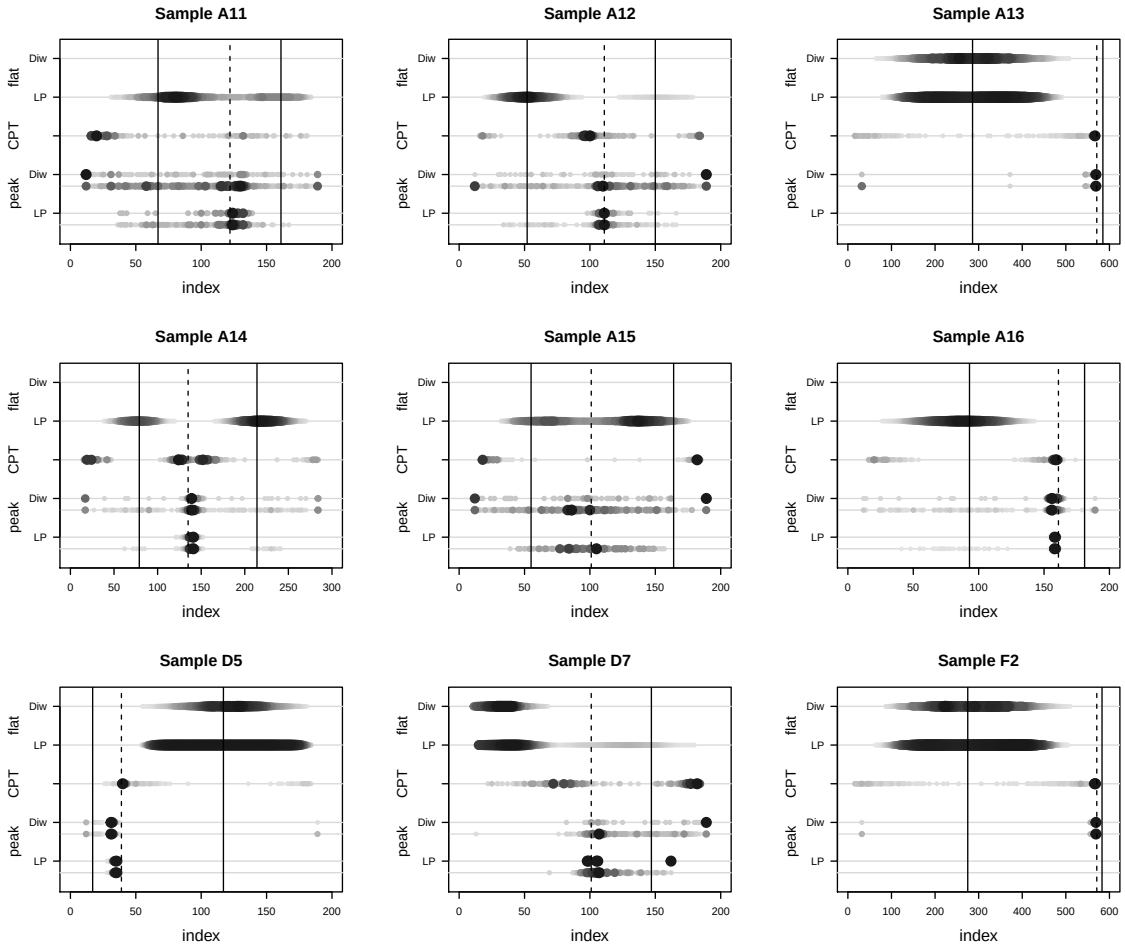



Figure 84: Feature positions in the asymmetric bi-modal samples. Darker points correspond to more frequent counts. Second lines under the LP and Diw peaks are detected rates. Vertical solid lines mark modes, dashed gaps.

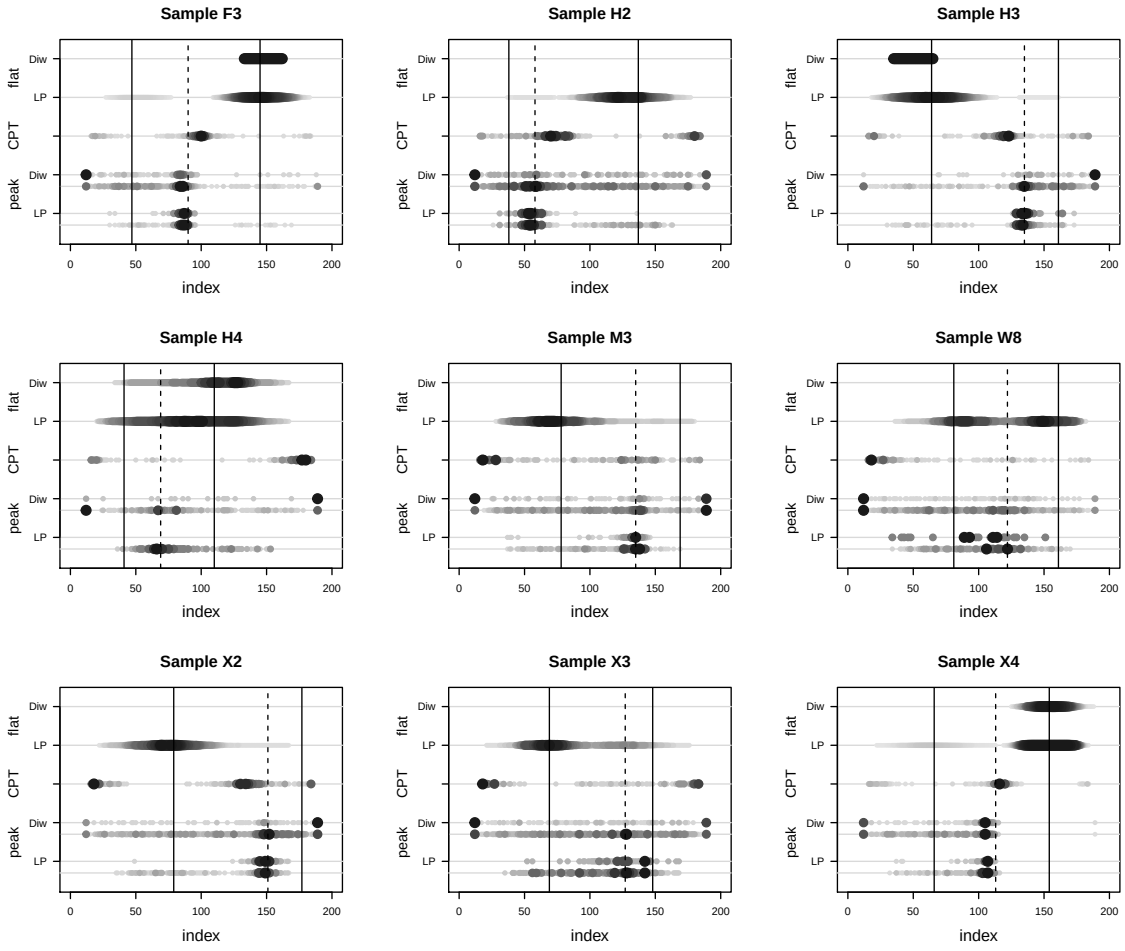



Figure 84: Feature positions in the asymmetric bi-modal samples, continued.

The tri-modal samples (Figure 85) are not completely resolved, and it is hard to make general statements about the results. Symmetric distributions, for A24, D8, W9, X5, and X7, have symmetric feature positions. The presence of a variate to the side of two symmetric draws, in A23, A25, and F4, disturbs the feature locations, so that flats are pulled into the second mode in A23, or out of it in A25.

Sample X5, a symmetric, well-separated distribution, has the cleanest result (Figure 86). The low-pass flats fall within the first mode more often than the second or third, and the interval spacing finds the first gap less often than the second.

Both gaps are found by the point features in X7, although with reservations. The imbalanced draw sizes push them towards the first and third modes and the changepoints towards the center, while the second gap does not appear among the low-pass peaks. The low-pass spacing has peaks at both gaps in sample P1, although the discrete draws mean the changepoints and interval peaks respond to each step in value. It also finds the gaps in W9 and D8, while the detected interval peaks are rejected as insignificant; some changepoints lie at the gaps, although more occur in the tails. Changepoints find both gaps in A22.

More commonly the point features identify one gap, consistently in A22, A23, A25, F4, and M4. The other gap is often detected, for example in A25 and M4, but rejected by testing. Sample X6 is an intermediate case, with low-pass peaks matching the first gap, and interval peaks and changepoints the second while also having a weak association with the first. G1 is also inconsistent. There are significant interval peaks and changepoints between the modes, although shifted away from the gap. The low-pass spacing contains matching peaks that are rejected by testing, leaving significant features to the right of the second mode that roughly match another group of changepoints.

Samples A21 and A24 are not resolved. Peaks exist in both spacings in A24, but testing rejects all.

Changepoints and interval peaks often lie in the tails of the sample, at the left or right edge of the histograms. This happens in A21, A24, D8, G1, P1, and W9. Many of the other samples occasionally have tailing peaks in the interval spacing, marked with weak dots.

In samples X5 and D8 the low-pass flats lie within the three modes, although not equally despite the symmetry of the set-up. Samples W9, F4, and G1 have low-pass flats at two modes, and A22, A25, M4, X6, and X7 at one. In A23 and P1 they extend over two modes. Flats cannot resolve A21 and A24. Interval flats are a subset of those in the low-pass spacing for samples A22 and G1. In P1 we see different behavior between the spacings for the first time, with an interval flat appearing between the modes but also extending weakly over the second gap.

Table 38 classifies the results. The low-pass spacing tests find significant peaks at both anti-modes in 5 of 14 samples (1 marginal), the interval peaks in 1, and changepoints in 3. The low-pass peaks identify one anti-mode in 8 samples, the interval peaks in 9 (2 marginal), and changepoints in 7 (2 marginal). The low-pass spacing does somewhat better overall, and this naturally applies to flats because there are so few in the interval spacing. There are low-pass flats covering all three modes in 1 sample, two modes in 3, and in just one mode in 9 (1 marginal). The interval spacing has a flat over one mode in just 3 samples (1 marginal). There is no clear correlation between successfully matching flats with modes and peaks to gaps, except in X5.

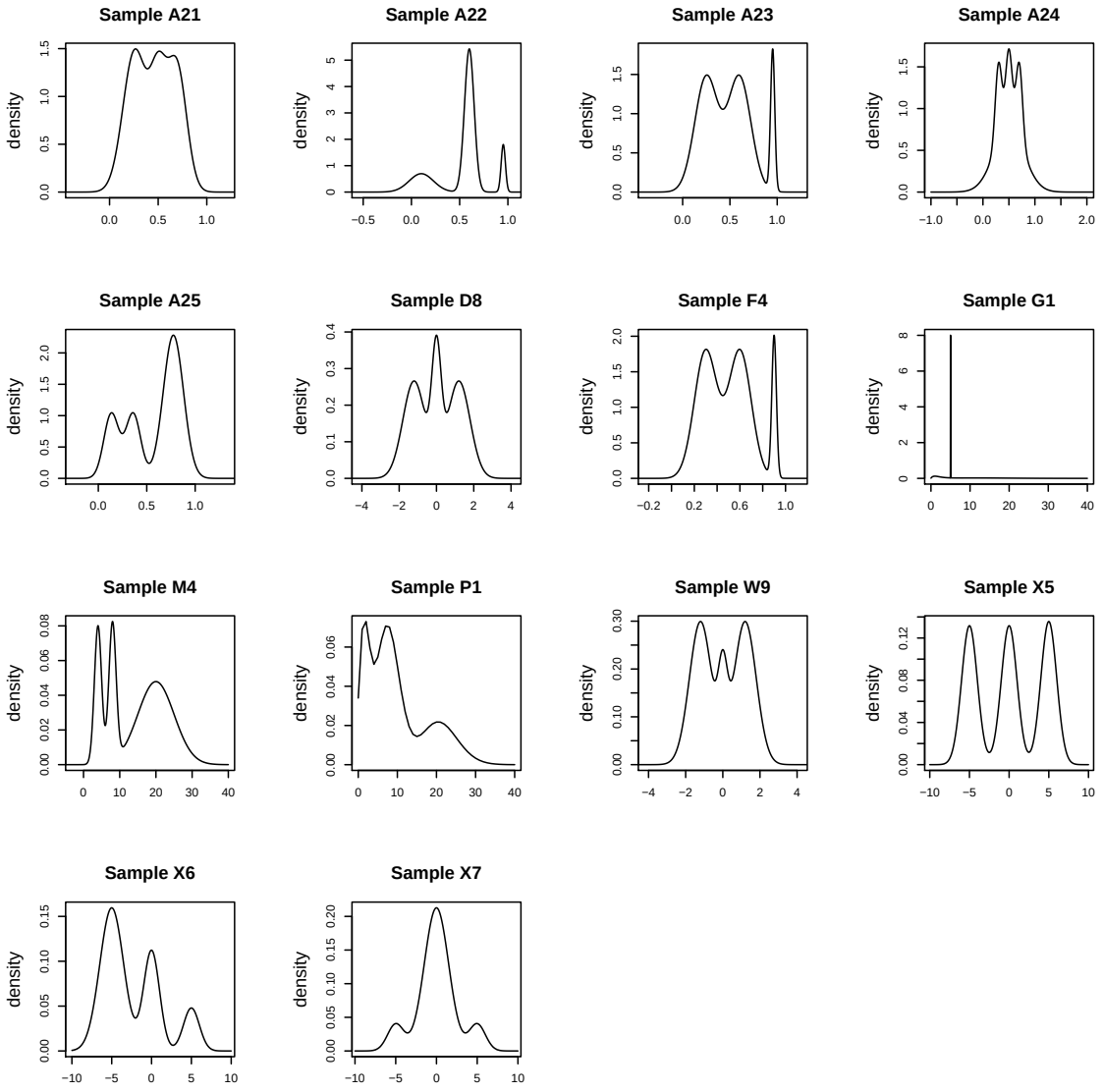



Figure 85: Tri-modal samples.

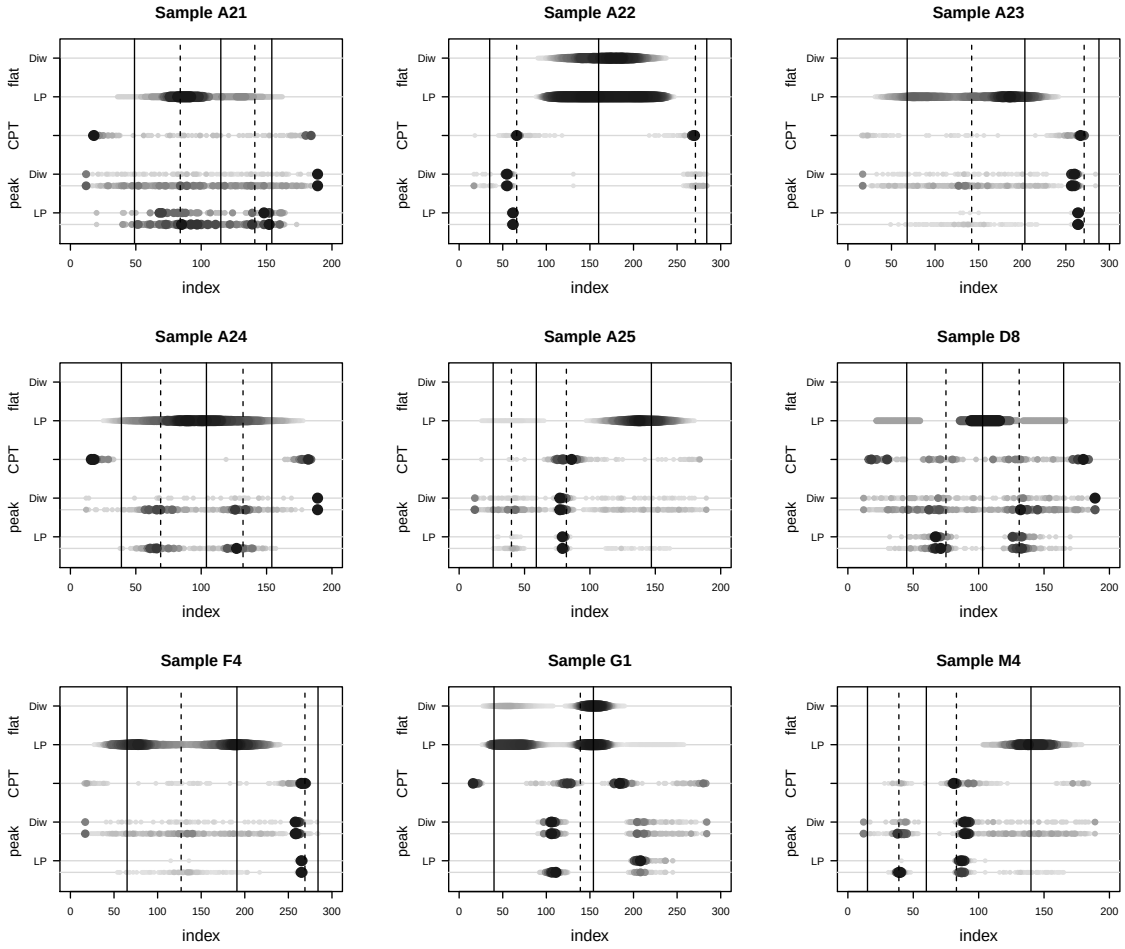



Figure 86: Feature positions in the tri-modal samples. Point darkness and size corresponds to count of trials. Peak lines include detected rates underneath. Vertical solid lines mark modes, dashed gaps.

Table 38: Clean Tri-Modal Feature Counts

| | | A21 | A22 | A23 | A24 | A25 | D8 | F4 |
|---|---|---|---|---|---|---|---|---|
| peak | LP | 2 | 1 | 1 | | 1 | 2 | 1 |
| | Diw | | 1 | 1 | | 1 | | 1 |
| CPT | | | 2 | 1 | | 1 | | 1 |
| flat | LP | 1 | 1 | | 1 | 1 | 1 | 2 |
| | Diw | | 1 | | | | | |

| | | G1 | M4 | P1 | W9 | X5 | X6 | X7 |
|---|---|---|---|---|---|---|---|---|
| peak | LP | 1 | 1 | 2 | 2 | 2 | 1 | 1 |
| | Diw | 1 | 1 | 1 | | 1 | 1 | 2 |
| CPT | | 1 | 1 | 1 | | 2 | 1 | 2 |
| flat | LP | 2 | 1 | 1 | 2 | 3 | 1 | 1 |
| | Diw | 1 | | 1 | | | | |

*underlined counts are marginal*

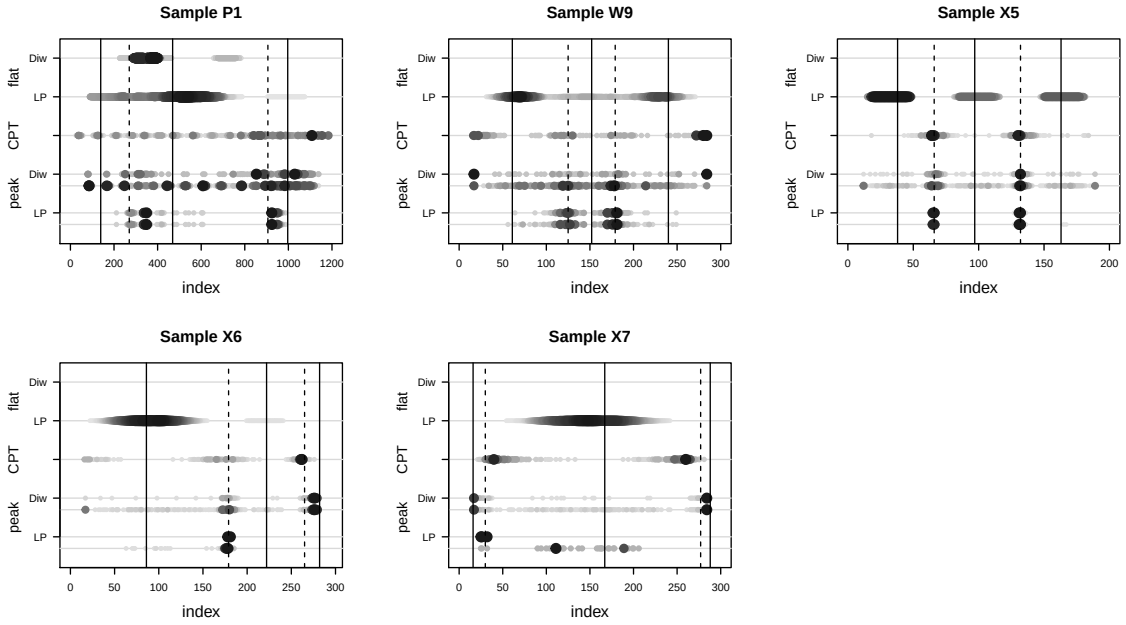



Figure 86: Feature positions in the tri-modal samples, continued.

The stability of the feature positions for the quad-modal samples, with densities shown in Figure 87, is better than we saw with the tri-modal. Point features are dots in the count graphs (Figure 88) and the flats have repeatable endpoints that generally do not extend beyond a mode. Yet having cleaner results does not mean finding features at all modes and gaps. K5 comes the closest. Flats in the low-pass spacing occur at each mode, with the small narrow variate with mean 0.5 the weakest. Changepoints find all three gaps, with some occasionally placed in the tails. The spacing contains significant peaks at the second gap, and weakly the other two.

We find partial matches of the ideal gaps and modes in the other samples but there is no overall pattern. The interval peaks agree with the low-pass in the first six samples, but fail to match up in the remaining five. There is a tendency to find interval peaks at the tails, for example in K3 and K5, which may be a small drop at the first or last index creating a local maximum, which the greater smoothing of the low-pass filters would remove. We see a larger variation in the position of peaks detected in the interval spacing than in the low-pass. We see some examples of the acceptance tests preferring detected peaks and creating a denser dot in the significant counts, for example the low-pass peak at the third gap of K7, or the interval peaks at the two tail ends of K5. The interval peaks in K9 are scattered throughout the sample because the variates generate discrete values.

Changepoints lie at the gaps, with some dispersion over the runs possible, as at the second gap of K1 or K5. They can identify different gaps than either type of peak, for example being most focussed on the first gap in K8 or F5.

Flats are well-defined at the modes in general. They appear in the largest draws of F5 and N4. The latter also has a flat in the interval spacing but not the low-pass in the fourth, tight variate; the second and third draws are at the minimum size of 30, too small to contain a flat because a filter will soften a sharp transition, eating into the feature. Flats cannot distinguish between the third and fourth modes of K2 and K6 and extend over the gap between them. In K8 they reach from the third

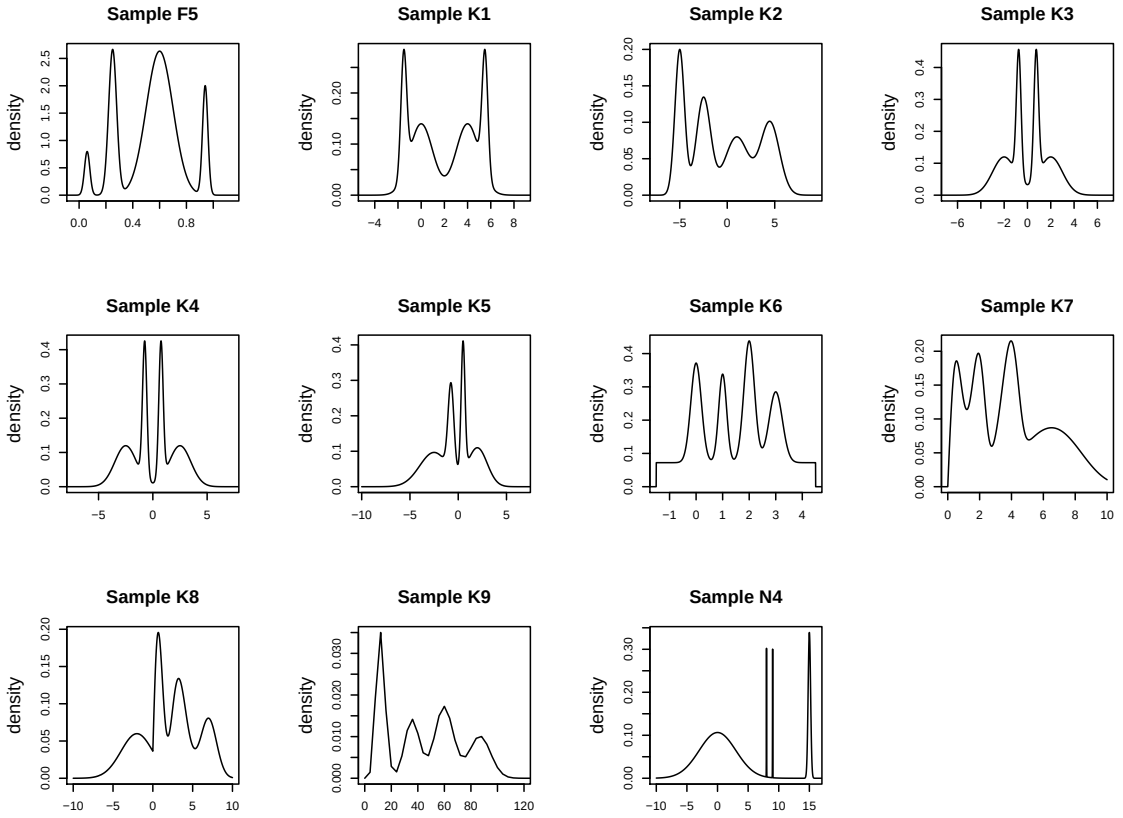



Figure 87: Quad-modal samples.

mode toward the second, and in K7 a flat covers the first gap. Nothing in the density of these four samples would lead one to suspect that the flats would extend beyond a mode. The flats from sample K1 reach outward from the large second and third modes over the gaps, explained by the imbalance in draw sizes. The flats do well matching the modes of samples K3, K4, and K5, although not equally strong.

Table 39: Clean Quad-Modal Feature Counts

| | | F5 | K1 | K2 | K3 | K4 | K5 | K6 | K7 | K8 | K9 | N4 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| peak | LP | 2 | 1 | 2 | 1 | 1 | 1 | 3 | 2 | 1 | 1 | 1 |
| | Diw | 3 | 1 | 2 | 1 | 3 | 2 | | | | 3 | 1 |
| CPT | | 3 | 1 | 2 | 2 | 1 | 3 | 2 | 2 | 1 | 1 | 1 |
| flat | LP | 1 | | | 2 | 2 | 3 | | 1 | 1 | 1 | 1 |
| | Diw | 1 | | | | | | | | | | 2 |

*underlined counts are marginal*

Table 39 counts the number of features clearly found in each sample. They must be well-defined in the plot and properly bounded, so that flats do not extend over gaps. Marginal features are not as concentrated, but we ignore any high background uncertainty, for example in the K2 changepoints. Much weaker histogram dots, such as the low-pass peaks in K4 at the first and third anti-mode, do

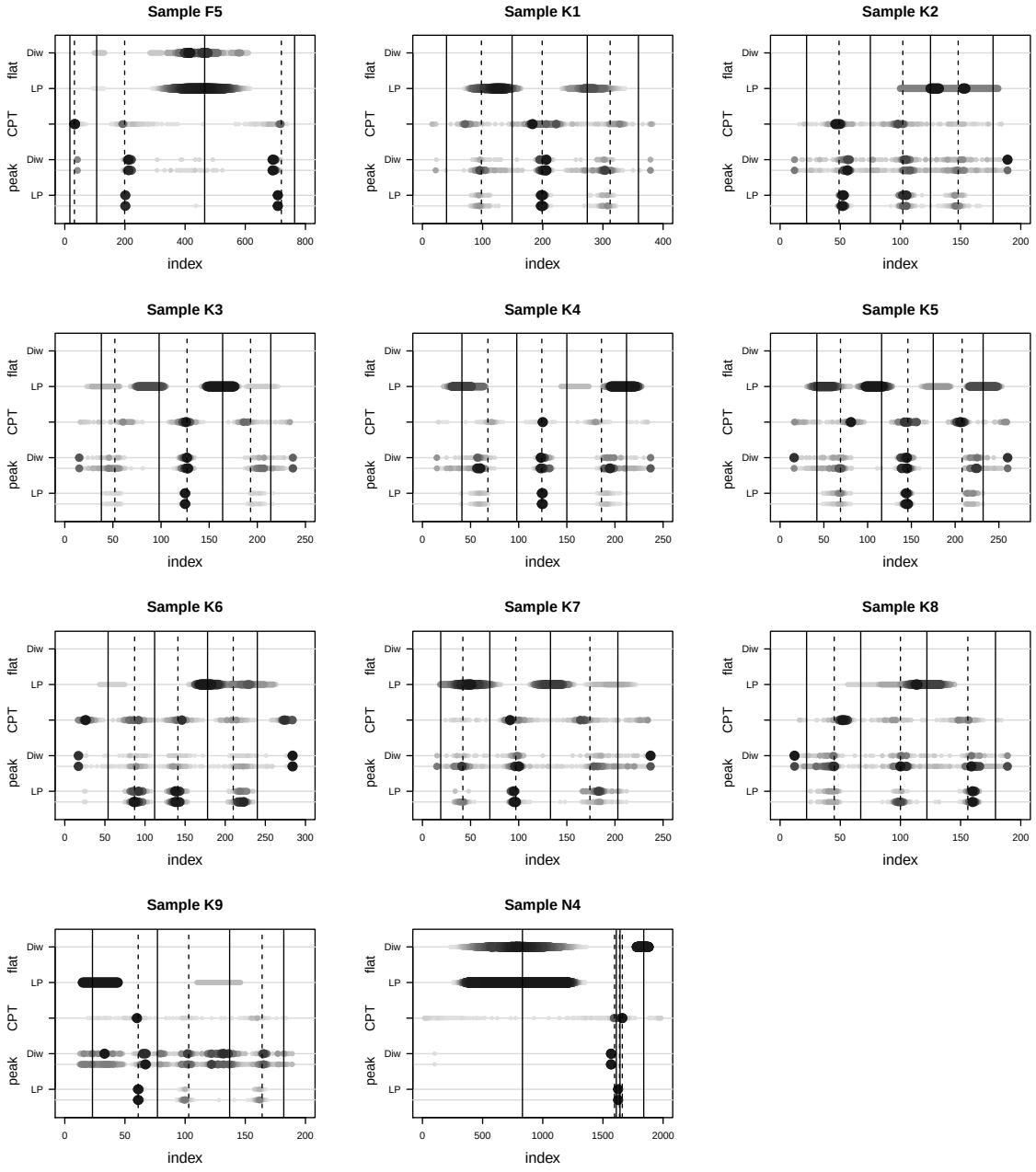



Figure 88: Feature positions in the quad-modal samples. Point size and darkness represents frequency over all runs. The second line underneath the LP and Diw peaks are detected rates. Solid vertical lines mark modes, dashed gaps.

not count. To a degree these counts are subjective and the analyses in [Detail 36] and [Detail 37] will quantify them.

All three anti-modes are matched by low-pass peaks in 1 of 11 samples, by interval peaks in 3 (1 marginal), and by changepoints in 2. Low-pass peaks identify two of the anti-modes in 3 samples, interval peaks in 2 (1 marginal), and changepoints in 4. One anti-mode is located in 7 samples by low-pass peaks, for 3 samples in the interval spacing, and in 5 (1 marginal) by changepoints. The low-pass spacing has no flats covering all four modes, over three modes in 1 sample, over two modes in 2, and over one mode in 5. The interval spacing has very few significant flats, covering two modes in 1 sample and one mode in another.

Figure 89 shows the densities of the samples with five modes. Four are symmetric, K10, K15, N5, and W10. K10 is the quad-modal K1 with an extra variate added at mean 0, and K15 is based on K4, with a larger separation in the middle and a fifth variate added. Samples N5 and W10 are built from the same variates, with 5/3 more points in each draw for N5 than W10; their densities are the same. Although K12 is built from five variates, there is no gap between the fourth and fifth.

Point features are reliable over the 200 runs (Figure 90), forming solid dots, with more spread in the changepoints than the peaks. They lie at the gaps, although not consistently. For example, in K11 the low-pass spacing occasionally has a peak at the first gap, while this is the strongest feature

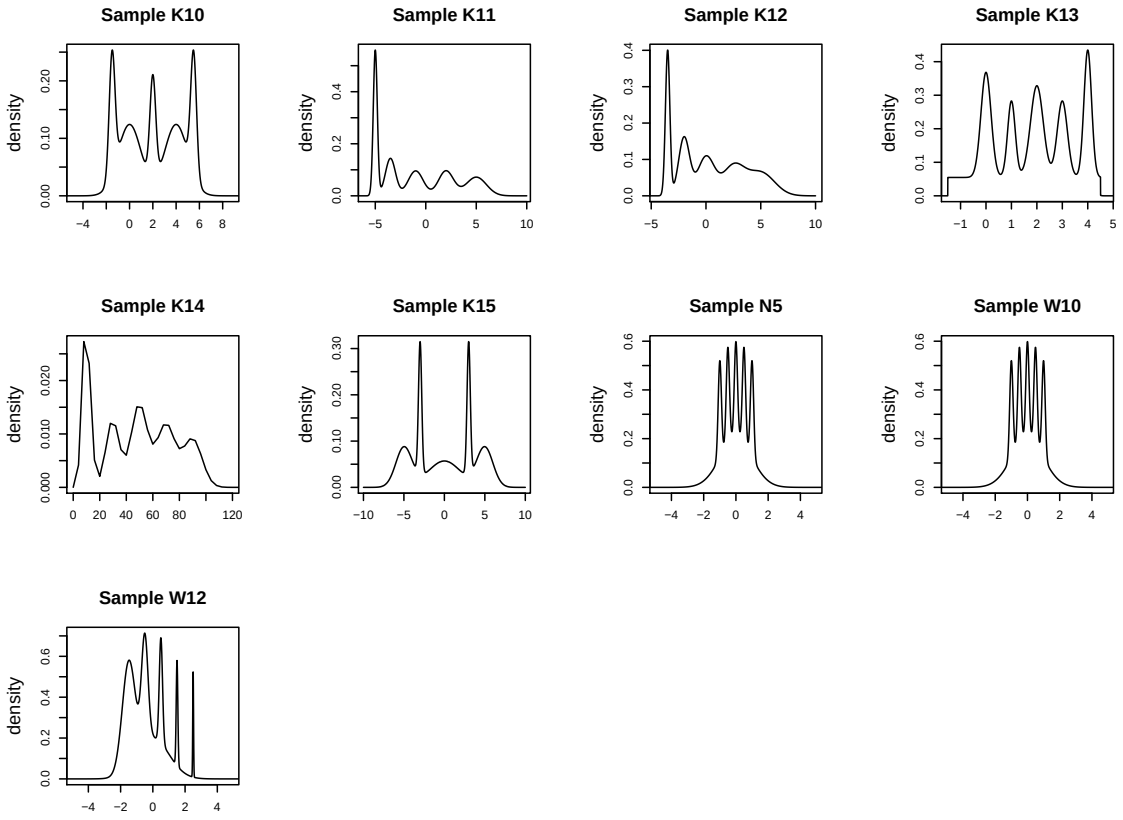



Figure 89: Penta-modal samples.

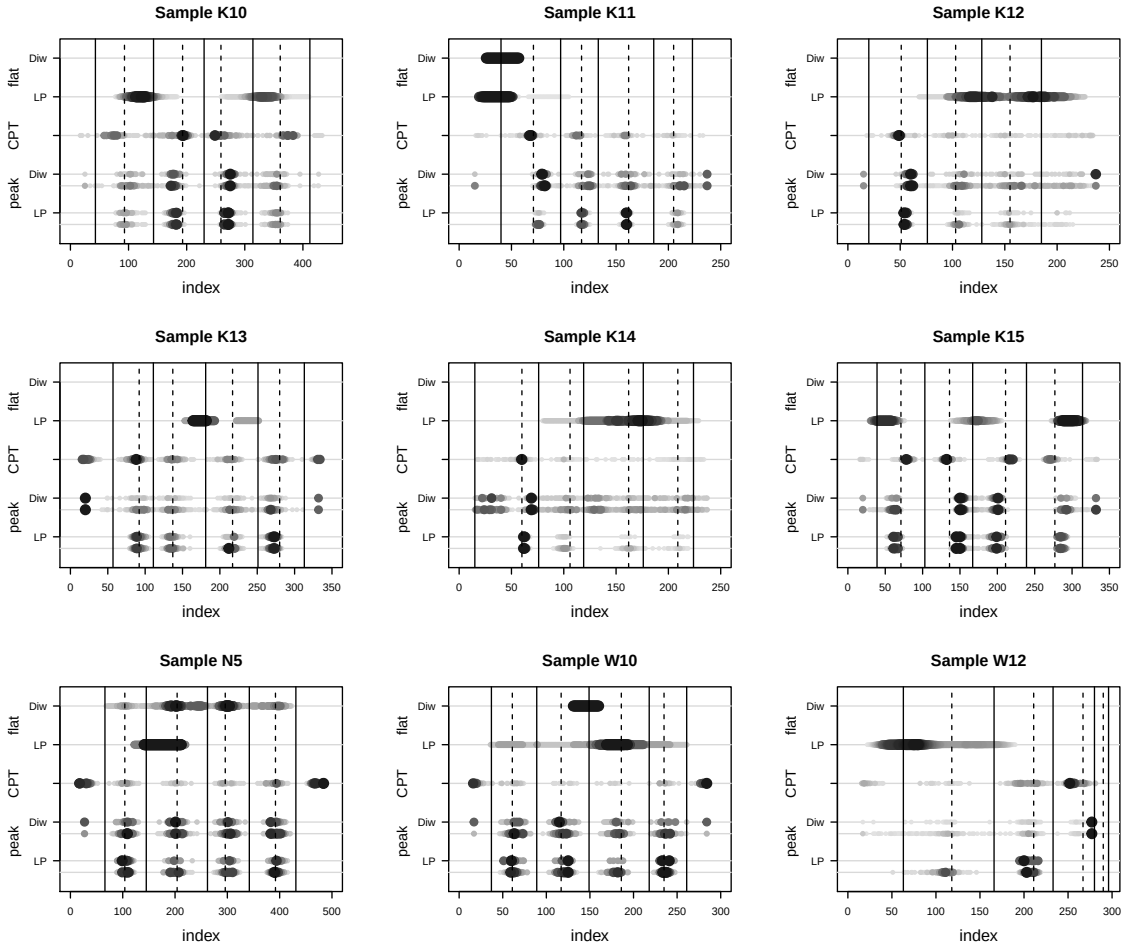



Figure 90: Feature positions in the penta-modal samples. Darker points occur more frequently. Second plots underneath each peak are for detected rates. Vertical solid lines mark modes, dashed gaps.

in the interval spacing and changepoints. In K15 the low-pass peaks pick out all four gaps, but the interval spacing finds the first and fourth in fewer trials. Samples K12, K14, and W12 seem difficult to resolve, as the peaks identify only one gap. K14 has discrete variates, so the interval peaks are dispersed except at the first gap. The point features preserve the symmetry of the distribution. Although not all gaps are found equally often, we do see the same pattern left and right of the middle mode. Changepoints trigger in tails of K13, N5, and W10; we also have some trials with interval peaks at the tails of several other samples. The acceptance rate of testing is not equal in either the low-pass or interval spacing. Examples include the low-pass peak at the first gap of K11 or the middle two gaps of K13 for either spacing. The larger draws of N5 compared to W10 do not affect the position of any of the point features, but do improve the uniformity of the counts.

Flats are less sharply defined over the trials. Even when the endpoints are stable, such as in samples K10 and N5, the flats can extend outside the mode over a gap. The former is the same stretching seen in the quad-modal K1. Why the N5 stretching occurs is not clear, as it breaks the symmetry of the sample even as the interval flats preserve it. The same occurs in W10. In samples K12, K14, and W12 we see flats extending over two modes and the intervening gap, indicating the

spacing increase at the gap is near the ripple specification. Confirming this, there are no point features at these gaps. While there is a flat in the first mode of K11, it disappears in K12 because the draw is smaller and the effects of the starting tail and clipping by the filter size reduce the constant section to less than the flat length.

The summary of features in Table 40 reflects this wide variety of results. One sample, K15, comes closest to being resolved. Two, the twins N5 and W10, are nearly so. K10 and K13 are partially resolved, and in the others we identify only one mode change.

Table 40: Clean Penta-Modal Feature Counts

| | | K10 | K11 | K12 | K13 | K14 | K15 | N5 | W10 | W12 |
|---|---|---|---|---|---|---|---|---|---|---|
| peak | LP | 2 | 2 | 1 | 3 | 1 | 4 | 3 | 3 | 1 |
| | Diw | 2 | 1 | 1 | | 1 | 2 | 4 | 2 | 1 |
| CPT | | 4 | 1 | 1 | 2 | 1 | 4 | | | 1 |
| flat | LP | | 1 | | 1 | | 3 | | 1 | |
| | Diw | | 1 | | | | | 2 | 1 | |

*underlined counts are marginal*

W14 is the only sample with a chirp, ever-smaller draws of ever-narrowing normal variates at ever-shrinking separations (Figure 91). The sixth draw is so small that it does not register, so the graph only shows four gaps and five modes. Because the separations are absolute, devoid of any data, the point features are stable and appear as sharp dots. Changepoints locate each gap, doing better as the width decreases. The low-pass spacing has peaks at the first two, but only the first passes

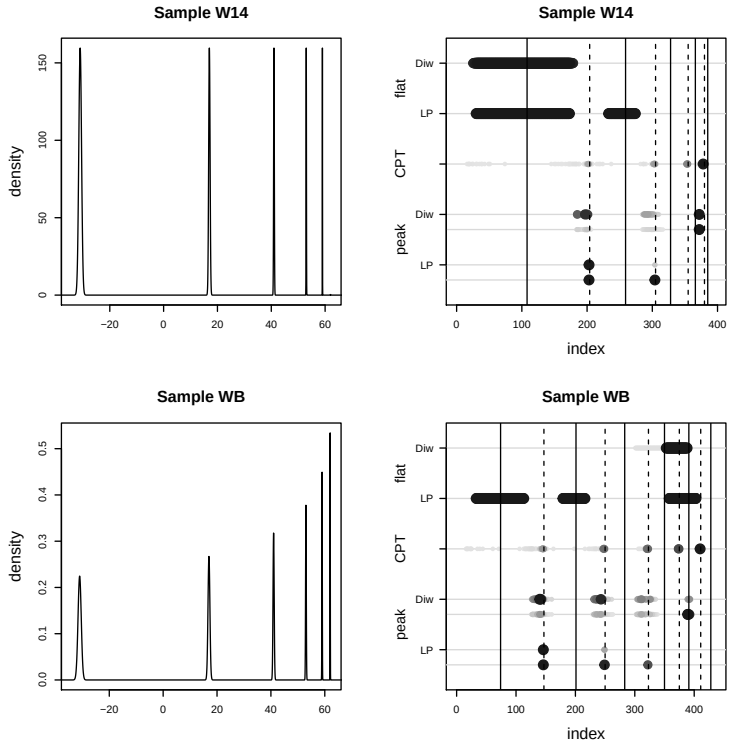



Figure 91: Chirp samples (left) and feature positions (right).

acceptance testing. The interval spacing has a clear peak at the fourth gap, and occasionally at the first two. Testing rejects these unequally, weakening the rate at the fourth gap until it is more equal to the first. Note that the peaks lie between the fourth mode and gap. Here we run into the problem that the separation between variates is so large that there are no points generated between them, so that the translation from index to coordinate rounds off the actual gap towards the mode. The first draw is large enough to support flats in both spacings, while the second mode only appears in the low-pass flats. The absolute separations make for stable endpoints, and the flats appear as solid bars that cut-off immediately. The third draw is larger than the minimum flat length, 50 against 30, but smoothing eats away at the boundary and blocks the feature.

We add a variant WB of the chirp to correct the draw sizes and make the sixth variate more prominent. The draws $i = 0 \ldots 5$ are

$$\left(800 \cdot 2^{(5-i)/2}/31\right) N\left(65 - (96/2^i), 2^{-0.75(i+1)}\right)$$

The means have not changed. The draw sizes range between 26 and 146 instead of 6 and 203. The standard deviations are larger, moreso for the fifth and sixth draw.

The changepoint behavior in this variant does not change. They pick out each gap, more often for the narrower variates. The interval peaks match the first three gaps at weakening rates, with one at the fourth or fifth gap shifted to the fifth mode. The low-pass spacing has peaks at the first three gaps, but only the first is considered significant. Low-pass flats again lie within the first two modes, and a third appears at the fourth but extends over the adjacent gaps indicating the filter's smoothing is removing these peaks. There is no interval flat at the first mode, but the extended flat also occurs in the interval spacing. Overall the analysis finds this variant easier to resolve and the results better capture the actual structure, although still incompletely.

The literature also contains a large number of uni-modal samples, often with skewed tails or superimposed variates that do not separate (Figure 92). These are useful for checking the analysis for false features. We expect flats within the mode and no peaks.

All samples have low-pass flats at the modes (Figure 93). Several, such as A1, A6, C2, and C3, show variability in the endpoints. Only W5, and maybe A9, are truly stable. Most samples also contain flats in the interval spacing. These are generally subsets of the low-pass flats. Samples C2 and W5 are notably stable. Still, in several we see the roughness of the interval spacing affecting the flats. Samples A1 and A9 seem to have more than one flat, and A10 clearly does.

The changepoints and interval peaks avoid the mode. Any that appear are limited to the sample's tail, at the first and/or last indices of the total draw. We do see false positives in the low-pass spacing, in A2, A4, A6, and C2, as well as peaks that lie towards the tails. Remember though that the greyscale encoding is set per sample and that a black dot may represent only a few trials. We see large rejection rates of peaks in either spacing, and not always equally.

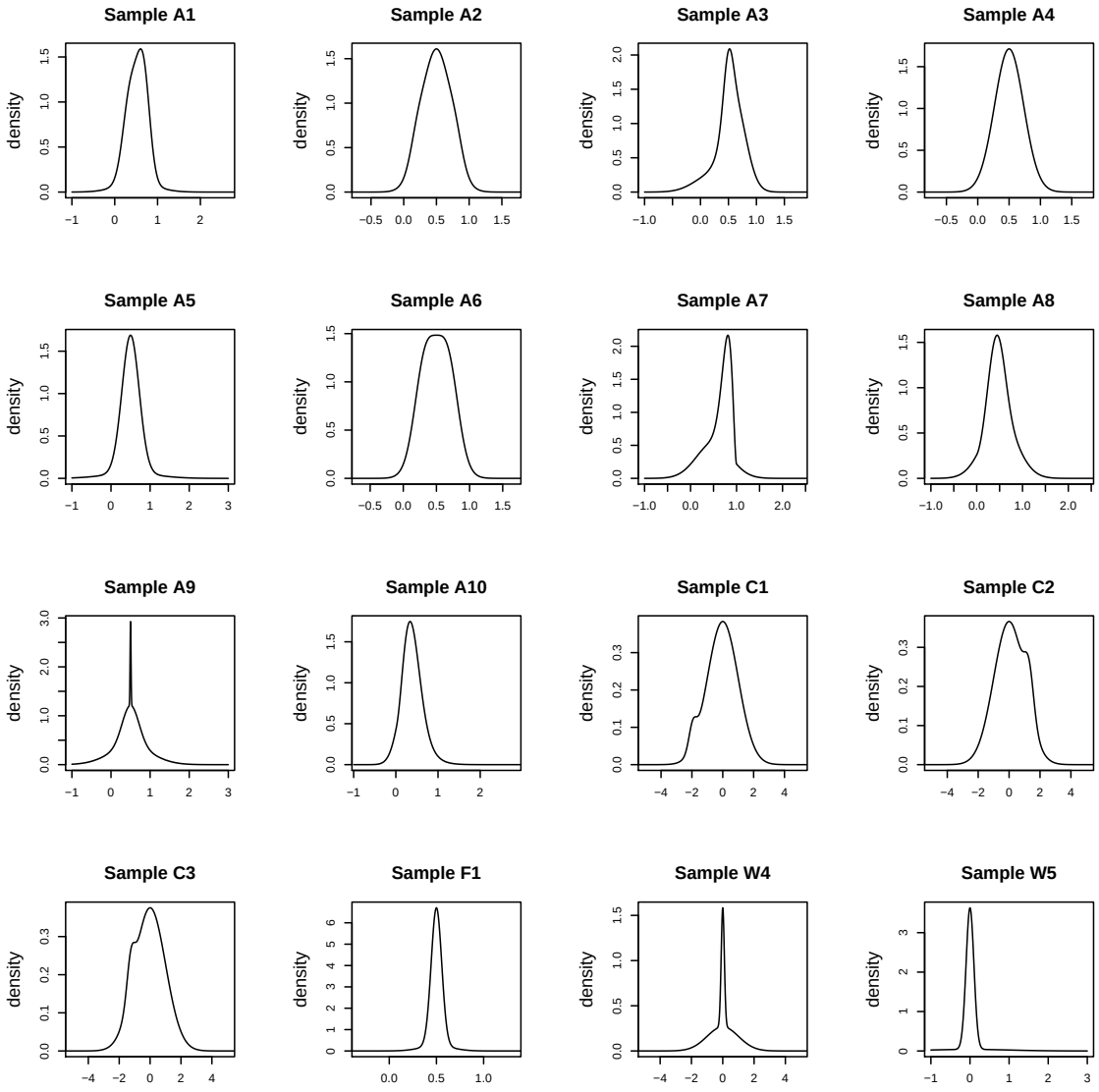



Figure 92: Uni-modal samples.

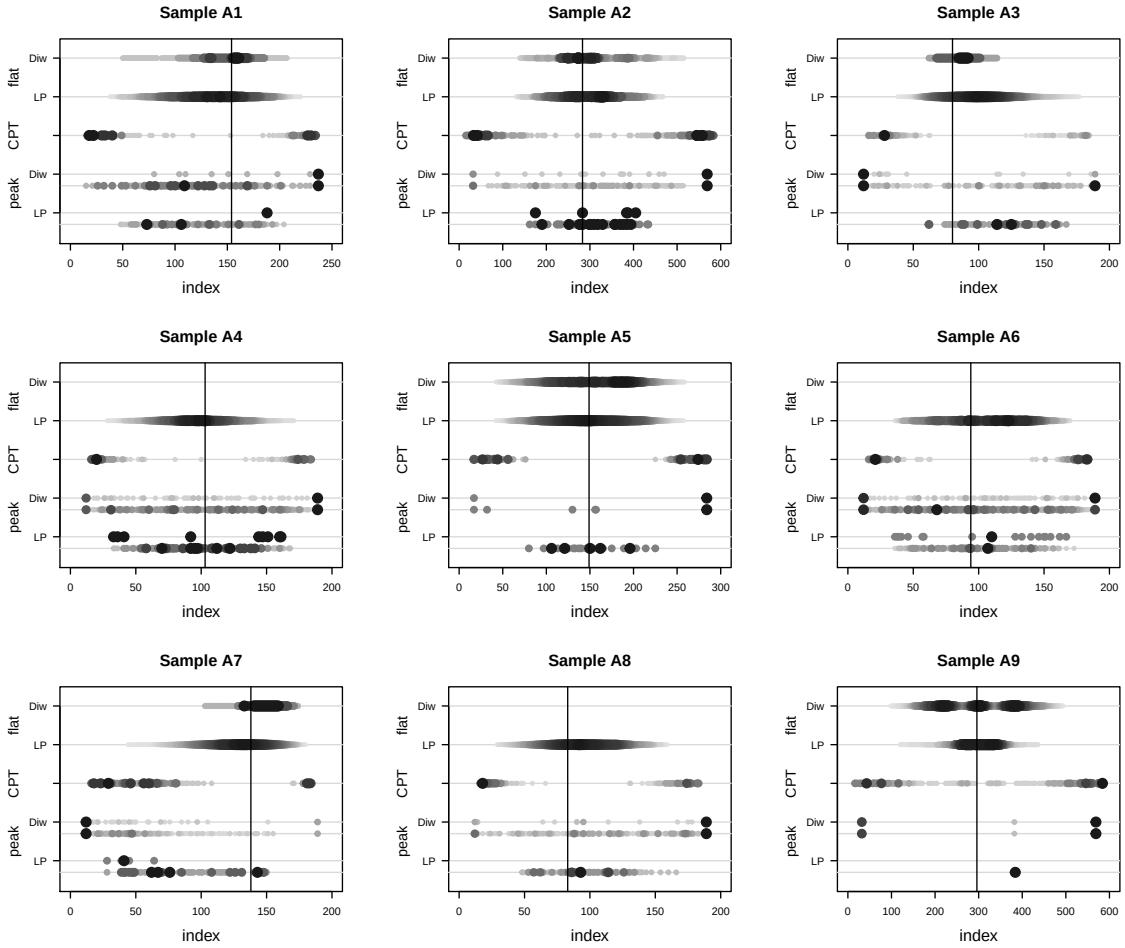



Figure 93: Feature positions in the unimodal samples. Darkness and size of points corresponds to frequency of position. Second lines for peaks are detected rates. Vertical solid lines mark ideal modes, dashed gaps.

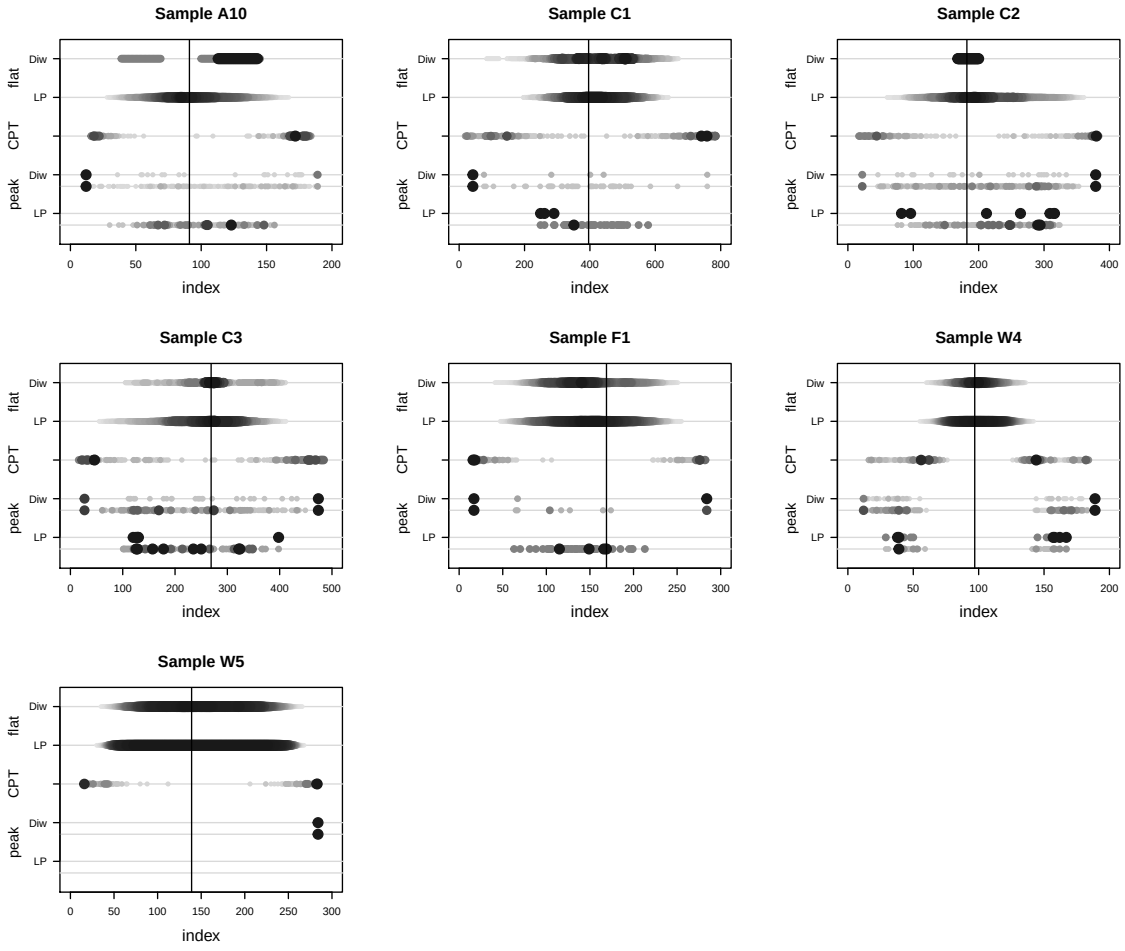



Figure 93: Feature positions in the unimodal samples, continued.

**Detail 36 Feature Placement in Literature Datasets**

To quantify the placement accuracy we track to which modes and anti-modes in the ideal density features align. Point features, the peaks and changepoints, match to the closest mode or gap, without regard for the size of the separation. There may be multiple points aligned to the same extremum. Interval features, the flats and level sections, cover a mode or gap if they lie completely within it, with boundaries set mid-way between mode and anti-mode, without regard for the location of inflection point. Coverage does not imply that the feature includes the mode or gap itself. It may lie to one side, but it may not include cross the mid-point. Because the level sections often respond to the tails of the total sample, we define this alignment as those intervals that lie before or after the outer modes without covering them. The counts for sections in the first and last mode do not include the tailing sections. The left and right tails correspond to the initial and final, or leading and trailing, points in the sample, in the steep arms of the U. Extended intervals include at least one mode and gap. They are not confined to a single ideal feature. Significance is again at the 0.01 level for all tests.

The table for each sample presents total counts over the same 200 runs as used in [Detail 35] and [Detail 37]. Extrema are numbered from left to right. Blanks replace zero counts for legibility. The first row of each table has the total number of each feature found over the trials. Divide the counts by 200 to get the average number of features per run. We used each draw itself to convert from the index of the features to the absolute coordinate used in the density functions. The resolution of the drawn values therefore imposes some inaccuracy in the conversion, particularly if there are large gaps, for example in samples with two well-separated variates.

**D Samples**

The symmetry in the D4 sample can be seen in the peaks of either method or in the low-pass flats, which are balanced between the two modes. The changepoints and level sections are not symmetric, favoring the second and first modes, respectively. By a small margin, the low-pass peaks lie most often in the gap, but this is suppressed in the interval spacing and changepoints. There is a large rejection of the detected peaks by either method, roughly 90%, but only 30% of the low-pass flats are declared insignificant. The significant peaks mostly align to a mode so the accuracy is poor. The tails account for 90% of the level sections. Both flats and level sections often fall outside a mode.

| D4 (bi) | | significant | | | | | detected | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | CPT | peak | | flat | | level | peak | | flat | |
| | | LP | Diw | LP | Diw | | LP | Diw | LP | Diw |
| # features | 183 | 22 | 134 | 42 | 0 | 651 | 314 | 963 | 61 | 0 |
| left tail | | | | | | 206 | | | | |
| mode 1 | 84 | 8 | 55 | 6 | | 18 | 101 | 368 | 8 | |
| gap 1 | 7 | 5 | 26 | | | | 118 | 238 | | |
| mode 2 | 92 | 9 | 53 | 5 | | 8 | 95 | 357 | 8 | |
| right tail | | | | | | 177 | | | | |
| extended | | | | 31 | | 242 | | | 45 | |

Sample D5 shows a difference between its two modes, with peaks matching the first and flats and level sections covering the second. All low-pass features and the interval flats pass testing; 20% of the interval peaks are rejected, which is still a high pass rate. The low-pass peaks also often match the gap, avoiding the tight second draw but alignment to the mode means their accuracy is poor. The changepoints mark the second mode or the transition to it. Flats capture the second mode. The starting tail accounts for a higher fraction of the first draw's sections than the ending tail. One third

of the level sections cover more than a mode or gap.

| D5 (abi) | | significant | | | | | detected | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | CPT | peak | | flat | | level | peak | | flat | |
| | | LP | Diw | LP | Diw | | LP | Diw | LP | Diw |
| # features | 365 | 199 | 219 | 200 | 164 | 901 | 199 | 271 | 200 | 164 |
| left tail | | | | | | 40 | | | | |
| mode 1 | 17 | 121 | 184 | | | 41 | 121 | 223 | | |
| gap 1 | 183 | 78 | 33 | | | | 78 | 33 | | |
| mode 2 | 165 | | 2 | 200 | 164 | 348 | | 15 | 200 | 164 |
| right tail | | | | | | 199 | | | | |
| extended | | | | | | 273 | | | | |

The density for sample D7 has only one mode and gap. The first draw is largest at the first index, so flats and level sections will appear to align to the gap, and there is no left tail. Although the low-pass spacing has one peak per draw and the interval spacing two, testing rejects almost all. The low-pass peaks and changepoints match the gap, or first variate, but the interval peaks are balanced between gap and mode. The flats are found significant, but cover the gap rather than the mode. There is one significant flat in the low-pass spacing per trial, but only one in five in the interval spacing. Two thirds of the level sections are confined to the gap, or cover the second tail.

| D7 (abi) | | significant | | | | | detected | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | CPT | peak | | flat | | level | peak | | flat | |
| | | LP | Diw | LP | Diw | | LP | Diw | LP | Diw |
| # features | 312 | 5 | 45 | 229 | 41 | 714 | 213 | 425 | 266 | 41 |
| gap 1 | 187 | 4 | 19 | 180 | 41 | 306 | 153 | 192 | 192 | 41 |
| mode 1 | 125 | 1 | 26 | 14 | | 8 | 60 | 233 | 25 | |
| right tail | | | | | | 195 | | | | |
| extended | | | | 35 | | 205 | | | | |

The distribution of features between modes and gaps mirrors the symmetry of sample D8, where the first and third variates are identical and placed evenly around the narrow, smaller middle draw. Except for the low-pass spacing and level sections, the counts in the first and third mode, or first and second gap, are similar. The low-pass peaks strongly match the gaps for good accuracy, while the interval peaks are spread evenly except in the middle mode. The runs generate many peaks, two in the low-pass spacing and five in the interval, that are largely rejected, leaving one peak of either kind. Flats essentially do not occur. The level sections are long and rarely confined to a single gap or mode. The imbalance between the first and third modes comes from unequal coverage of the tails, with many more sections in the leading tail.

| D8 (tri) | | significant | | | | | detected | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | CPT | peak | | flat | | level | peak | | flat | |
| | | LP | Diw | LP | Diw | | LP | Diw | LP | Diw |
| # features | 174 | 213 | 202 | 5 | 0 | 628 | 402 | 1057 | 6 | 0 |
| left tail | | | | | | 166 | | | | |
| mode 1 | 50 | 13 | 41 | 1 | | 16 | 37 | 220 | 1 | |
| gap 1 | 22 | 97 | 50 | | | | 166 | 237 | | |
| mode 2 | 24 | | 16 | 1 | | | 6 | 119 | 1 | |
| gap 2 | 24 | 97 | 49 | | | 1 | 174 | 256 | | |
| mode 3 | 54 | 6 | 46 | | | 23 | 19 | 225 | 1 | |
| right tail | | | | | | 47 | | | | |
| extended | | | | 3 | | 375 | | | 3 | |

**M Samples**

The low-pass peaks in the M1 sample match the gap, with testing suppressing almost all aligned with a mode. The accuracy is therefore good. The peaks in the interval spacing are distributed between the three extrema, with the fewest matching the gap. Testing of these features is particularly aggressive, rejecting 85%, but those passing are still scattered and the accuracy is poor. The separation of the two variates in M1 is marginal for resolving the modes, and this can be seen in the small imbalances between the mode counts, which do not reflect the symmetry of the draws as well as D4 does. The changepoints are balanced but are distributed like the interval peaks, matching the modes. There are few flats in the low-pass spacing. The long level sections span the total draw. The large imbalance between the tails may reflect a bias in the algorithm, when the search for the next section begins at the end of the previous, cutting short the last section. This is not symmetric.

| M1 (bi) | | significant | | | | | detected | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | CPT | peak | | flat | | level | peak | | flat | |
| | LP | LP | Diw | LP | Diw | | LP | Diw | LP | Diw |
| # features | 19 | 113 | 156 | 157 | 0 | 599 | 305 | 1034 | 23 | 0 |
| left tail | | | | | | 196 | | | | |
| mode 1 | 5 | 9 | 60 | 57 | | 21 | 61 | 357 | 7 | |
| gap 1 | 1 | 101 | 43 | 35 | | | 197 | 336 | 1 | |
| mode 2 | 3 | 3 | 53 | 65 | | 13 | 47 | 341 | 3 | |
| right tail | | | | | | 82 | | | | |
| extended | 10 | | | | | 287 | | | 12 | |

Sample M2 has a larger separation between the two variates than M1 does, and a comparison with those results shows how this makes identification of the bi-modality easier. There are fewer detected low-pass peaks aligned to a mode and the testing is more selective, passing all in the gap. The detected interval peaks avoid the gap, but again testing is selective and those that are correctly positioned survive at a higher rate. Some significant interval peaks remain aligned with modes, so the accuracy here is moderate. The changepoints have doubled and strongly match the gap. Still, as many lie closer to a mode. There are a few more flats in the low-pass spacing, all significant and contained in one of the modes. Most level sections overlap both a mode and the gap, and there is a 2:1 imbalance between the left and right tails.

| M2 (bi) | | significant | | | | | detected | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | CPT | peak | | flat | | level | peak | | flat | |
| | | LP | Diw | LP | Diw | | LP | Diw | LP | Diw |
| # features | 430 | 204 | 193 | 44 | 0 | 875 | 274 | 886 | 44 | 0 |
| left tail | | | | | | 187 | | | | |
| mode 1 | 106 | 2 | 46 | 23 | | 30 | 36 | 327 | 23 | |
| gap 1 | 224 | 200 | 92 | | | | 200 | 236 | | |
| mode 2 | 100 | 2 | 55 | 21 | | 50 | 38 | 323 | 21 | |
| right tail | | | | | | 97 | | | | |
| extended | | | | | | 511 | | | | |

The M3 sample has a deep gap, like M2. The significant low-pass peaks, present in half the runs, match the gap, while the interval peaks split between the gap and first, broader mode. Changepoints also split the two, but lie more often in the mode, in a ratio similar to that for the detected interval peaks. Flats in the low-pass spacing, mostly significant, cover the first mode. Half the level sections are long, and the short fall in the leading tail, not the trailing.

| M3 (abi) | | significant | | | | | detected | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | CPT | peak | | flat | | level | peak | | flat | |
| | | LP | Diw | LP | Diw | | LP | Diw | LP | Diw |
| # features | 252 | 90 | 111 | 45 | 0 | 519 | 330 | 885 | 59 | 0 |
| left tail | | | | | | 216 | | | | |
| mode 1 | 143 | 9 | 43 | 34 | | 28 | 120 | 444 | 40 | |
| gap 1 | 74 | 81 | 45 | 1 | | 1 | 204 | 320 | 2 | |
| mode 2 | 35 | | 23 | 1 | | 5 | 6 | 121 | 3 | |
| right tail | | | | | | 11 | | | | |
| extended | | | | 9 | | 258 | | | 14 | |

For M4 the detected peaks in both spacings match the gaps. Only low-pass peaks in the second gap survive testing, although the first two variates duplicate the M2 setup. The significant interval peaks are not as consistently placed, half in the second gap and a quarter each in the first gap and third mode. Their accuracy is good overall. The changepoints also focus on the second gap or the modes to either side at half the rate. They occasionally resolve the first pair of draws. There are a few low-pass flats, all significant, covering the third, broader mode. Three quarters of the level sections are long, and here the short ones lie in the right tail.

| M4 (tri) | | significant | | | | | detected | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | CPT | peak | | flat | | level | peak | | flat | |
| | | LP | Diw | LP | Diw | | LP | Diw | LP | Diw |
| # features | 446 | 200 | 179 | 22 | 0 | 894 | 451 | 879 | 22 | 0 |
| left tail | | | | | | 32 | | | | |
| mode 1 | 12 | | 9 | | | 39 | 1 | 41 | | |
| gap 1 | 32 | 1 | 31 | | | | 198 | 185 | | |
| mode 2 | 76 | | 2 | | | | | 14 | | |
| gap 2 | 223 | 199 | 95 | | | | 200 | 220 | | |
| mode 3 | 103 | | 42 | 21 | | 34 | 52 | 419 | 21 | |
| right tail | | | | | | 145 | | | | |
| extended | | | | 1 | | 644 | | | 1 | |

**F Samples**

Sample F1 is unimodal and all features would match the single mode.

Sample F2 has a deep gap that cannot be found in the spacings. The interval spacing has a few peaks, mostly significant, that do match the gap when they occur. Two of the three changepoints per draw match the first, large mode, and the other matches the gap. The second variate is too small to affect either spacing. The sample does generate flats, one or two in the low-pass spacing and four or five in the interval, that are mostly accepted as significant. These fall within the first mode. Most of the level sections are contained within that mode or the leading tail, but not the trailing.

| F2 (bi) | | significant | | | | | detected | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | CPT | peak | | flat | | level | peak | | flat | |
| | | LP | Diw | LP | Diw | | LP | Diw | LP | Diw |
| # features | 612 | 0 | 22 | 227 | 740 | 1333 | 0 | 25 | 302 | 875 |
| left tail | | | | 9 | | 448 | | | 54 | 367 |
| mode 1 | 437 | | 6 | 218 | 740 | 362 | | 7 | 248 | 508 |
| gap 1 | 168 | | 18 | | | 4 | | 18 | | |
| mode 2 | 7 | | | | | 136 | | | | |
| right tail | | | | | | | | | | |
| extended | | | | | | 383 | | | | |

There is one significant peak in the spacings for sample F3 in about half the trials. The low-pass peak matches the gap, the interval peak splits between the first mode and gap. Testing rejects half the peaks seen in the low-pass spacing and more than two-thirds of the interval peaks, targeting those aligned with the mode. The low-pass accuracy is good, the interval moderate. The changepoints match the gap most often, at twice the rate of either mode. There is a flat in the low-pass spacing in each run, most significant, covering the tight second mode. Short level sections prefer the second mode too, although three times as many lie in the tails. There are almost as many long sections as short.

| F3 (abi) | | significant | | | | | detected | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | CPT | peak | | flat | | level | peak | | flat | |
| | | LP | Diw | LP | Diw | | LP | Diw | LP | Diw |
| # features | 294 | 122 | 139 | 155 | 1 | 773 | 251 | 616 | 175 | 1 |
| left tail | | | | | | 178 | | | | |
| mode 1 | 55 | 5 | 59 | 10 | | 27 | 38 | 325 | 14 | |
| gap 1 | 160 | 117 | 72 | | | 2 | 196 | 231 | | |
| mode 2 | 79 | | 8 | 141 | 1 | 99 | 17 | 60 | 156 | 1 |
| right tail | | | | | | 111 | | | | |
| extended | | | | 4 | | 356 | | | 5 | |

The spacings contain peaks at both gaps in sample F4. The detected interval peaks are equally distributed throughout the sample, except in the third mode and favoring the first gap. After testing one significant peak in both spacings remains closest to the second gap, with a third of the interval peaks scattered in and between the first two modes. Changepoints also most often mark the second gap, or the first pair of modes at a third the rate. Half the low-pass flats are significant, with a third each in the first two modes and the remainder extended. The level sections are long, except for those covering the left tail.

| F4 (tri) | | significant | | | | | detected | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | CPT | peak | | flat | | level | peak | | flat | |
| | | LP | Diw | LP | Diw | | LP | Diw | LP | Diw |
| # features | 385 | 188 | 237 | 157 | 0 | 908 | 426 | 963 | 349 | 0 |
| left tail | | | | | | 205 | | | | |
| mode 1 | 66 | | 25 | 50 | | 18 | 23 | 222 | 99 | |
| gap 1 | 23 | 2 | 30 | 7 | | 1 | 188 | 297 | 32 | |
| mode 2 | 77 | | 30 | 41 | | 2 | 29 | 258 | 111 | |
| gap 2 | 211 | 186 | 149 | | | 1 | 186 | 180 | | |
| mode 3 | 8 | | 3 | | | 42 | | 6 | | |
| right tail | | | | | | 2 | | | | |
| extended | | | | 59 | | 637 | | | 107 | |

Both spacings pick up the second and third gaps in sample F5, with testing accepting the detected peaks as significant. The accuracy is good. Changepoints match the first gap at half the rate, as well as the third, broadest mode. Two low-pass flats are found in the third mode and one in the second. Testing accepts most in the third but none in the second. This also holds for interval flats but at a tenth the rate. The first and fourth modes are too small, despite their being very tight normal variates. Short level sections cover all but the first mode, with one in the third mode and in half the trials one in the second and fourth. There are not enough points from the first and fourth modes for level sections not to include the maxima themselves, so very few sections cover just the tails. Three quarters of the sections are extended.

| F5 (quad) | | significant | | | | | detected | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | CPT | peak | | flat | | level | peak | | flat | |
| | | LP | Diw | LP | Diw | | LP | Diw | LP | Diw |
| # features | 1002 | 393 | 385 | 274 | 24 | 1925 | 394 | 438 | 556 | 24 |
| left tail | | | | | | 1 | | | | |
| mode 1 | 27 | | | | | 5 | | | | |
| gap 1 | 146 | | 1 | | | | | 1 | | |
| mode 2 | 76 | | 6 | 1 | 1 | 119 | | 6 | 198 | 1 |
| gap 2 | 291 | 200 | 194 | | | 16 | 200 | 200 | | |
| mode 3 | 193 | | 7 | 273 | 23 | 222 | 1 | 32 | 358 | 23 |
| gap 3 | 262 | 193 | 177 | | | 19 | 193 | 199 | | |
| mode 4 | 7 | | | | | 115 | | | | |
| right tail | | | | | | | | | | |
| extended | | | | | | 1420 | | | | |

**C Samples**

All C examples are uni-modal. This analysis does not apply to them.

**H Samples**

Sample H1 is the same as M1, and the same comments hold. Differences in the counts reflect the variation of the draws. The features in the low-pass spacing, peaks and flats, are stable. We see some flats that have shortened and cover the second mode. There are a quarter more significant peaks in the interval spacing, with most added to the gap. There are 15% more changepoints, also at the gap. We get 5% more level sections, mostly covering the right tail.

H1 (bi)

| | | significant | | | | | detected | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | CPT | peak | | flat | | level | peak | | flat | |
| | | LP | Diw | LP | Diw | | LP | Diw | LP | Diw |
| # features | 171 | 117 | 187 | 22 | 0 | 630 | 315 | 1052 | 23 | 0 |
| left tail | | | | | | 187 | | | | |
| mode 1 | 61 | 7 | 65 | 5 | | 29 | 55 | 358 | 5 | |
| gap 1 | 44 | 105 | 58 | | | | 203 | 353 | | |
| mode 2 | 66 | 5 | 64 | 11 | | 13 | 57 | 341 | 11 | |
| right tail | | | | | | 100 | | | | |
| extended | | | | 6 | | 301 | | | 7 | |

Sample H2 has a significant peak in the low-pass spacing in most trials that matches the gap. Significant interval peaks are not well placed, scattered between the modes and gap. Testing for both rejects half of the detected low-pass peaks and 85% of the interval peaks, mostly those matching the second mode. The accuracy of the low-pass detectors is good, of the interval poor. Changepoints match mostly the gap and second mode. There are flats in the low-pass spacing in half the runs, covering the second mode. Half the level sections are long. A fifth of the short sections cover the modes, with the rest in the tails, more often the trailing.

H2 (abi)

| | | significant | | | | | detected | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | CPT | peak | | flat | | level | peak | | flat | |
| | | LP | Diw | LP | Diw | | LP | Diw | LP | Diw |
| # features | 304 | 174 | 131 | 78 | 0 | 749 | 301 | 850 | 83 | 0 |
| left tail | | | | | | 121 | | | | |
| mode 1 | 36 | 25 | 30 | | | 36 | 27 | 153 | | |
| gap 1 | 132 | 147 | 47 | | | | 164 | 292 | | |
| mode 2 | 136 | 2 | 54 | 73 | | 49 | 110 | 405 | 77 | |
| right tail | | | | | | 173 | | | | |
| extended | | | | 5 | | 370 | | | 6 | |

The peaks in the two spacings and the changepoints align differently to the modes and gap in sample H3. There is one peak in the low-pass spacing, found significant only a third of the time but matching the gap. There are two or three peaks in the interval spacing, 85% rejected, and aligning with the gap or second, smaller but wider mode. There are one or two changepoints per trial matching the first mode or gap. The common theme is alignment to the gap, but not in a dominant fashion. There is one flat in the low-pass spacing that is usually significant. It covers the first, tighter mode. The short level sections that exist also fall within this mode, above and beyond the left tail. Still, half the sections are long.

H3 (abi)

| | | significant | | | | | detected | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | CPT | peak | | flat | | level | peak | | flat | |
| | | LP | Diw | LP | Diw | | LP | Diw | LP | Diw |
| # features | 298 | 60 | 66 | 145 | 1 | 711 | 214 | 528 | 194 | 1 |
| left tail | | | | | | 204 | | | | |
| mode 1 | 122 | | 6 | 143 | 1 | 83 | 42 | 104 | 176 | 1 |
| gap 1 | 131 | 52 | 26 | | | 1 | 147 | 213 | | |
| mode 2 | 45 | 8 | 34 | | | 9 | 25 | 211 | | |
| right tail | | | | | | 49 | | | | |
| extended | | | | 2 | | 365 | | | 18 | |

Sample H4 generates a peak in both the low-pass and interval spacing, but testing rejects almost all as insignificant. The detected low-pass peaks match the gap most often, or the second mode at half the rate. The interval peaks also match the gap, although they also spread more into both modes. A changepoint in each run matches the second mode. The low-pass spacing contains a flat that testing usually rejects. Any significant flats that remain cover the second mode, but more are long. A third of the level sections are extended, and most short sections correspond to the tails. Spacing cannot resolve the modality of this sample.

| H4 (abi) | | significant | | | | | detected | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | CPT | peak | | flat | | level | peak | | flat | |
| | | LP | Diw | LP | Diw | | LP | Diw | LP | Diw |
| # features | 189 | 0 | 24 | 52 | 19 | 752 | 218 | 242 | 288 | 22 |
| left tail | | | | | | 189 | | | | |
| mode 1 | 36 | | 3 | | | 13 | 26 | 50 | 10 | |
| gap 1 | 10 | | 5 | | | 1 | 132 | 106 | | |
| mode 2 | 143 | | 16 | 10 | 11 | 11 | 60 | 86 | 107 | 13 |
| right tail | | | | | | 255 | | | | |
| extended | | | | 42 | 8 | 283 | | | 171 | 9 |

**N and P Samples**

The N2 sample has the same setup as M1 and H1, but with 250 points drawn per variate rather than 100. The position of the peaks, flats, and changepoints respect the symmetry of the sample. The peaks in the low-pass spacing, almost all significant, match the gap and the accuracy is good. Testing rejects more than half found in the interval spacing, equally among those matching a mode or gap. The interval peaks align with a gap but the accuracy is poor. Changepoints on the other hand align with a mode four times as often as they do to the gap; there are two per trial. The larger draws double the number of flats detected in the low-pass spacing compared to H1 or M1, falling inside a mode or a tail, but also to the gap in half the runs. Almost all fail testing, especially those in the gap and tails, leaving a few covering a mode. The level sections do not respect the symmetry. Few short sections remain outside the tails, preferring the first mode. About half the sections are long.

| N2 (bi) | | significant | | | | | detected | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | CPT | peak | | flat | | level | peak | | flat | |
| | | LP | Diw | LP | Diw | | LP | Diw | LP | Diw |
| # features | 410 | 203 | 51 | 53 | 2 | 939 | 228 | 541 | 746 | 2 |
| left tail | | | | | | 263 | | | 77 | |
| mode 1 | 166 | 5 | 15 | 20 | | 47 | 14 | 144 | 173 | |
| gap 1 | 73 | 195 | 24 | 1 | | 2 | 200 | 266 | 131 | |
| mode 2 | 171 | 3 | 12 | 20 | 2 | 6 | 14 | 131 | 192 | 2 |
| right tail | | | | | | 217 | | | 59 | |
| extended | | | | 12 | | 404 | | | 114 | |

Sample N4 has one peak in the low-pass spacing per run, significant, which shifts into the second mode from the first gap. This is due to the large difference in the draw size between the first and second variates, which pushes the local maxima into the smaller draw. In one-quarter of the runs the peak matches the large gap between the second and third modes. The interval spacing also has one peak, significant, that matches the first gap. The interval accuracy is better than the low-pass. Changepoints match the modes and first gap, and align to the second and third modes at half the

rate. The low-pass spacing has one flat, significant, that lies within the first mode; the interval spacing divides this into three or four flats, with a third to one side in the initial tail. There is also a significant interval flat in the fourth mode. Both the first and fourth mode are much larger than the minimum flat length, while the second and third are too small, despite the narrow width of their variates. Testing accepts the detected flats in either spacing. N4 generates thirteen level sections on average. A third are long. Short sections lie in all but the third mode, and cover the gap. As many fall in the tails, especially the initial.

| N4 (quad) | | significant | | | | | detected | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | CPT | peak | | flat | | level | peak | | flat | |
| | | LP | Diw | LP | Diw | | LP | Diw | LP | Diw |
| # features | 943 | 200 | 201 | 200 | 929 | 2681 | 200 | 201 | 212 | 929 |
| left tail | | | | | 271 | 635 | | | 10 | 271 |
| mode 1 | 234 | | 1 | 200 | 458 | 251 | | 1 | 202 | 458 |
| gap 1 | 235 | | 200 | | | 137 | | 200 | | |
| mode 2 | 127 | 156 | | | | 109 | 156 | | | |
| gap 2 | 9 | 44 | | | | | 44 | | | |
| mode 3 | 72 | | | | | | | | | |
| gap 3 | 36 | | | | | | | | | |
| mode 4 | 230 | | | | 200 | 400 | | | | 200 |
| right tail | | | | | | 212 | | | | |
| extended | | | | | | 937 | | | | |

The N5 sample contains five symmetric draws atop a sixth broad background, which does not contribute a mode but does evenly reduce the depth of the second and third gaps. The spacings accurately detect peaks at each gap, but testing rejects two out of three. The significant interval peaks are spread equally among the four gaps, but the low-pass peaks are sensitive to the smaller depth of the middle two and match at a higher rate to the first and fourth gap. The changepoints match the first and fifth mode, and at a lower rate the adjacent first and fourth gaps. They identify a central gap or mode much less often. Low-pass flats have the reverse pattern, with rates increasing toward the center of the sample for those within an extremum, but two-thirds are extended. Testing rejects all of these as insignificant. The interval spacing contains many fewer flats, in one-quarter of the trials, but all are significant. The short ones scatter between mode and gap, but half are long. The level sections also are long. One third correspond to the tails, moreso the initial than final.

| N5 (penta) | | significant | | | | | detected | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | CPT | peak | | flat | | level | peak | | flat | |
| | | LP | Diw | LP | Diw | | LP | Diw | LP | Diw |
| # features | 544 | 266 | 287 | 2 | 53 | 1597 | 783 | 765 | 939 | 53 |
| left tail | | | | | | 282 | | | | |
| mode 1 | 160 | | 9 | | | 22 | | 13 | 2 | |
| gap 1 | 50 | 76 | 67 | | | | 192 | 180 | 5 | |
| mode 2 | 11 | | 1 | | 2 | 2 | | 3 | 102 | 2 |
| gap 2 | 22 | 44 | 65 | | 6 | | 197 | 175 | 30 | 6 |
| mode 3 | 14 | | 2 | | 8 | 2 | | 6 | 135 | 8 |
| gap 3 | 29 | 62 | 72 | | 3 | | 200 | 191 | 17 | 3 |
| mode 4 | 25 | | 3 | | 4 | | | 6 | 87 | 4 |
| gap 4 | 61 | 84 | 68 | | 1 | | 194 | 181 | 11 | 1 |
| mode 5 | 172 | | | | | 4 | | 10 | 5 | |
| right tail | | | | | | 208 | | | | |
| extended | | | | 2 | 29 | 1077 | | | 545 | 29 |

The P1 sample contains discrete values and the usual warnings about the changepoint and level section algorithms triggering at each unique value apply. The high feature counts show this. We will ignore them. The interval spacing also responds to the steps, with eighteen peaks per trial, but testing reduces this to one or two significant peaks that match the second gap or third mode. The low-pass spacing contains two peaks, both significant, that match the first or second gap. There are six flats in the low-pass spacing per run, with a third accepted as significant. They lie mostly within the second mode, or in the first at half the rate. However, a third of the significant flats are long. There are three flats in the interval spacing distributed between the first two modes and their gap, but almost all are rejected by testing.

| P1 (tri) | | significant | | | | | detected | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | CPT | peak | | flat | | level | peak | | flat | |
| | | LP | Diw | LP | Diw | | LP | Diw | LP | Diw |
| # features | 2768 | 362 | 325 | 343 | 17 | 5804 | 362 | 3631 | 1176 | 560 |
| left tail | | | | | | 200 | | | | |
| mode 1 | 408 | 41 | 25 | 58 | | 400 | 41 | 527 | 205 | 120 |
| gap 1 | 193 | 151 | 43 | 33 | 9 | 400 | 151 | 320 | 160 | 203 |
| mode 2 | 631 | 11 | 23 | 140 | 3 | 1000 | 11 | 998 | 344 | 187 |
| gap 2 | 630 | 158 | 101 | | | 1200 | 158 | 791 | 60 | 2 |
| mode 3 | 906 | 1 | 133 | | | 600 | 1 | 995 | 120 | 16 |
| right tail | | | | | | 1604 | | | 4 | |
| extended | | | | 112 | 5 | 400 | | | 283 | 32 |

**X Samples**

The notes for the X1 sample in [Detail 35] say that it is equivalent to M1 and H1, with the larger separation between the variations compensated by their larger width. The counts over the 200 runs are also similar. The acceptance rate of the low-pass peaks is lower but still matched to the gap. The acceptance rate of the interval peaks is higher, and they are again distributed equally among the gap and modes. There are more changepoints, now equally distributed where in M1 and H1 they align with the modes. Despite the wider draws the interval spacing and changepoints seem to be more sensitive to the larger gap. Flats and level sections appear at the same rate.

| X1 (bi) | | significant | | | | | detected | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | CPT | peak | | flat | | level | peak | | flat | |
| | | LP | Diw | LP | Diw | | LP | Diw | LP | Diw |
| # features | 208 | 163 | 113 | 19 | 0 | 621 | 313 | 1062 | 20 | 0 |
| left tail | | | | | | 183 | | | | |
| mode 1 | 75 | 13 | 38 | 6 | | 24 | 64 | 370 | 6 | |
| gap 1 | 71 | 145 | 39 | 1 | | | 200 | 351 | 1 | |
| mode 2 | 62 | 5 | 36 | 7 | | 18 | 49 | 341 | 8 | |
| right tail | | | | | | 75 | | | | |
| extended | | | | 5 | | 321 | | | 5 | |

Sample X2 has two variates with the same width and spacing but unequal draw size. The low-pass spacing has a peak matching the first, larger mode or gap in half the trials, but testing rejects the peak aligned with the mode; the accuracy is good. Almost all of the interval spacing peaks are rejected, with the significant peak, one in half the runs, distributed between the modes and gap. The changepoints mirror the detected low-pass peaks, aligning to both the first mode or gap. There is sometimes a flat in the low-pass spacing that is usually significant and falls within the first mode. The level sections are mostly extended, with a few remaining within the first mode above those corresponding to the tail. There seems to be a bias towards sections at the start of each draw, with fewer for the right tail, despite both variates having the same width.

| X2 (abi) | | significant | | | | | detected | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | CPT | peak | | flat | | level | peak | | flat | |
| | | LP | Diw | LP | Diw | | LP | Diw | LP | Diw |
| # features | 293 | 129 | 93 | 70 | 0 | 680 | 278 | 806 | 80 | 0 |
| left tail | | | | | | 221 | | | | |
| mode 1 | 117 | 1 | 32 | 64 | | 61 | 124 | 405 | 73 | |
| gap 1 | 153 | 120 | 36 | | | 2 | 145 | 249 | | |
| mode 2 | 23 | 8 | 25 | | | 8 | 9 | 152 | | |
| right tail | | | | | | 12 | | | | |
| extended | | | | 6 | | 376 | | | 7 | |

The X3 sample has a smaller separation between its variates than X1 and X2, and the draw sizes are not as unbalanced as X2. The affected low-pass peaks match the gap or the first mode, but testing rejects 80% of them, removing the link with the mode and reducing that to the gap to a quarter of the trials. The interval spacing is noisy with five peaks per run spread throughout, but most are not significant. Those that pass are also distributed so the accuracy is poor. A changepoint per run ties to a mode, or occasionally the gap. The low-pass spacing has a flat in a quarter of the trials, usually significant. Half fall within the first mode, a third are extended. Half the level sections also are long, and the remainder are associated with the tails.

| X3 (abi) | | significant | | | | | detected | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | CPT | peak | | flat | | level | peak | | flat | |
| | | LP | Diw | LP | Diw | | LP | Diw | LP | Diw |
| # features | 176 | 68 | 148 | 41 | 0 | 658 | 321 | 984 | 51 | 0 |
| left tail | | | | | | 213 | | | | |
| mode 1 | 80 | 8 | 66 | 23 | | 26 | 124 | 422 | 25 | |
| gap 1 | 29 | 49 | 43 | 4 | | 1 | 158 | 311 | 4 | |
| mode 2 | 67 | 11 | 39 | | | 33 | 39 | 251 | 1 | |
| right tail | | | | | | 100 | | | | |
| extended | | | | 14 | | 285 | | | 21 | |

The significant low-pass peaks in sample X4 match the gap, testing having eliminated those aligned with the first mode but also a majority of those at the gap. The significant interval peaks also match the gap, with testing strongly but not completely suppressing those aligned with the first mode. Neither spacing contains peaks aligned with the second mode, but they do have flats there, most found significant. Few flats are long. The changepoints match the gap at twice the rate of either mode. There are short level sections, one per trial, in the second mode, and one in each tail. The remaining sections are extended.

| X4 (abi) | | significant | | | | | detected | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | CPT | peak | | flat | | level | peak | | flat | |
| | | LP | Diw | LP | Diw | | LP | Diw | LP | Diw |
| # features | 287 | 76 | 161 | 236 | 90 | 893 | 259 | 599 | 245 | 90 |
| left tail | | | | 1 | | 210 | | | 2 | |
| mode 1 | 71 | 8 | 47 | 25 | | 20 | 86 | 389 | 31 | |
| gap 1 | 147 | 68 | 113 | | | | 173 | 209 | | |
| mode 2 | 69 | | 1 | 200 | 90 | 153 | | 1 | 200 | 90 |
| right tail | | | | | | 196 | | | | |
| extended | | | | 10 | | 314 | | | 12 | |

X5 has a symmetric layout, with three easily separated variates, as the results show. There are two peaks in the low-pass spacing, significant, that match each gap. The interval spacing has four peaks, at each gap and distributed between the modes, but testing reduces this to one significant, split between the gaps and some to the modes. There are two changepoints at each gap, and one split between the modes, favoring the central, second draw. The low-pass spacing has few flats, but testing passes them, divided equally between the modes. The level sections are long. One short section covers the first or third mode, above and beyond those corresponding to tails.

| X5 (tri) | | significant | | | | | detected | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | CPT | peak | | flat | | level | peak | | flat | |
| | | LP | Diw | LP | Diw | | LP | Diw | LP | Diw |
| # features | 604 | 374 | 212 | 34 | 0 | 1110 | 402 | 794 | 35 | 0 |
| left tail | | | | | | 112 | | | | |
| mode 1 | 51 | | 14 | 15 | | 61 | | 136 | 16 | |
| gap 1 | 201 | 190 | 85 | | | | 200 | 189 | | |
| mode 2 | 96 | | 17 | 9 | | 4 | | 135 | 9 | |
| gap 2 | 198 | 184 | 80 | | | | 200 | 194 | | |
| mode 3 | 58 | | 16 | 10 | | 100 | 2 | 140 | 10 | |
| right tail | | | | | | 20 | | | | |
| extended | | | | | | 813 | | | | |

Sample X6 is an unbalanced draw with a smaller gap between the first two variates and an imbalance in size between the second pair. The low-pass spacing has one detected peak that is not significant. It matches the first gap. The interval spacing has two or three peaks, one at the first gap and one split between the first and third mode; the draw imbalance may shift the peak at the second gap into the mode. Half of the interval peaks are significant but the accuracy is poor. Those matching the third mode all survive testing, while only a third at the first gap do, and none at the first mode. Changepoints match the two gaps and, at a lower rate, the first and second mode. There are two flats detected in the low-pass spacing. Those in the second mode are rejected by testing, leaving a strong association with the first mode. Few flats extend over a mode and gap, but most of the level sections do. Some short sections lie in the first mode, but many more lie in the initial tail.

| X6 (tri) | | significant | | | | | detected | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | CPT | peak | | flat | | level | peak | | flat | |
| | | LP | Diw | LP | Diw | | LP | Diw | LP | Diw |
| # features | 562 | 2 | 265 | 203 | 0 | 1060 | 214 | 540 | 355 | 0 |
| left tail | | | | | | 229 | | | | |
| mode 1 | 137 | | 8 | 192 | | 44 | 14 | 116 | 259 | |
| gap 1 | 175 | 2 | 72 | | | 4 | 200 | 210 | 2 | |
| mode 2 | 84 | | 2 | | | 1 | | 24 | 76 | |
| gap 2 | 160 | | 20 | 5 | | | | 21 | | |
| mode 3 | 6 | | 163 | | | 8 | | 169 | | |
| right tail | | | | | | 2 | | | | |
| extended | | | | 6 | | 772 | | | 18 | |

X7 is another symmetric sample where the side draws match the minimum flat length. There are few peaks in the low-pass spacing, but in the interval spacing they are distributed symmetrically between the extrema, primarily matching the second mode, but also the first and third at a third the rate. Testing rejects half of the peaks at the second mode, passing the others unchanged, meaning accuracy is poor. The changepoints are distributed symmetrically, aligning to the first and second gap and, at a lower rate, the second mode. The low-pass spacing does have two flats inside the second mode, most significant. The short level sections also fall within this mode, although 85% are extended. There is a very slight asymmetry in the counts, with a few more sections in the first mode and tail than the third.

| X7 (tri) | | significant | | | | | detected | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | CPT | peak | | flat | | level | peak | | flat | |
| | | LP | Diw | LP | Diw | | LP | Diw | LP | Diw |
| # features | 573 | 3 | 152 | 281 | 0 | 843 | 35 | 300 | 389 | 0 |
| left tail | | | | | | 12 | | | | |
| mode 1 | 4 | 1 | 50 | | | 12 | 1 | 51 | | |
| gap 1 | 203 | 2 | 22 | | | 1 | 2 | 29 | | |
| mode 2 | 159 | | 12 | 276 | | 105 | 32 | 145 | 377 | |
| gap 2 | 204 | | 19 | | | 9 | | 24 | | |
| mode 3 | 3 | | 49 | | | 1 | | 51 | | |
| right tail | | | | | | 1 | | | | |
| extended | | | | 5 | | 702 | | | 12 | |

**G Sample**

The low-pass spacing has one peak matching the gap in sample G1, and in half the runs one at the second mode. The significant low-pass peaks match only the second mode in a third of the runs, with testing rejecting all aligned to the gap. The interval spacing splits detected peaks between the gap and second mode, favoring the latter, and most pass, although the rejection rate is higher at the mode. One or two changepoints mark the second mode and one the gap, so they emphasize the second mode more than the interval spacing does. The low-pass spacing has a flat in the first mode, usually significant, and one in half the trials in the second mode, which is rejected. Half of these flats are extended. The interval spacing contains a flat, all significant, that is usually long, or covers the first mode in a third of the trials. Half of the level sections are short and correspond to tails, twice as many for the trailing than leading. The remaining sections are long.

| G1 (tri) | | significant | | | | | detected | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | CPT | peak | | flat | | level | peak | | flat | |
| | | LP | Diw | LP | Diw | | LP | Diw | LP | Diw |
| # features | 582 | 67 | 348 | 388 | 260 | 1227 | 289 | 454 | 532 | 260 |
| left tail | | | | | | 222 | | | | |
| mode 1 | 92 | | 8 | 160 | 60 | 45 | | 8 | 181 | 60 |
| gap 1 | 184 | | 166 | | | 6 | 199 | 187 | | |
| mode 2 | 306 | 67 | 174 | | | | 90 | 259 | | |
| right tail | | | | 8 | | 405 | | | 85 | |
| extended | | | | 220 | 200 | 549 | | | 266 | 200 |

**W Samples**

Samples W4 and W5 are uni-modal, so this analysis does not apply to them.

W6 is a symmetric two variate sample, with smaller separation and widths than the base M1 and H1 setup, or equivalent X1. The results are similar, with the spacing peaks and flats respecting the underlying symmetry, some variation in the changepoints, and a large difference between the tail coverage of the level sections. Peaks in the low-pass spacing match the gap, with testing rejecting two-thirds. The detected interval peaks are distributed equally between the modes and gaps, but only 15% survive testing, with a slight final preference for peaks aligned with a mode. The changepoints also match the modes more strongly than the gap. Half of the level sections extend over gap and mode, the other half cover the two tails, preferring the initial. The few flats detected in the low-pass spacing are significant. Most are short and cover the modes equally.

| W6 (tri) | | significant | | | | | detected | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | CPT | peak | | flat | | level | peak | | flat | |
| | | LP | Diw | LP | Diw | | LP | Diw | LP | Diw |
| # features | 159 | 124 | 156 | 21 | 0 | 603 | 323 | 1020 | 23 | 0 |
| left tail | | | | | | 190 | | | | |
| mode 1 | 53 | 11 | 57 | 7 | | 16 | 60 | 340 | 7 | |
| gap 1 | 36 | 103 | 44 | 1 | | 1 | 206 | 333 | 2 | |
| mode 2 | 70 | 10 | 55 | 7 | | 13 | 57 | 347 | 7 | |
| right tail | | | | | | 91 | | | | |
| extended | | | | 6 | | 292 | | | 7 | |

The W7 sample has the same separation as M1 and H1 but a smaller standard deviation. These results show this setup is easier to resolve. There are more significant peaks in the low-pass and interval spacing. They match the gap. The interval spacing has less than half the detected peaks than in M1, but a somewhat better acceptance rate, although two-thirds are still rejected, overwhelmingly from the modes. One changepoint in every trial matches the gap, while one aligns with one of the modes, splitting equally between them. This is the inverse behavior seen in M1. There are one or two flats in the low-pass spacing, located equally in the modes. There is one short level section in each mode in half the trials, above those corresponding to tails. Those covering a mode favor the second, although the left tail still dominates. The other tests respect the symmetry of the setup.

| W7 (bi) | | significant | | | | | detected | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | CPT | peak | | flat | | level | peak | | flat | |
| | | LP | Diw | LP | Diw | | LP | Diw | LP | Diw |
| # features | 374 | 200 | 159 | 307 | 0 | 941 | 225 | 462 | 307 | 0 |
| left tail | | | | | | 182 | | | | |
| mode 1 | 83 | 3 | 15 | 158 | | 94 | 13 | 140 | 158 | |
| gap 1 | 199 | 197 | 135 | | | 2 | 197 | 190 | | |
| mode 2 | 92 | | 9 | 149 | | 147 | 15 | 132 | 149 | |
| right tail | | | | | | 93 | | | | |
| extended | | | | | | 423 | | | | |

Few of the peaks in either the low-pass or interval spacing for sample W8 survive testing. The significant interval peaks match mostly the first mode, but also the gap and second mode. The detected low-pass peaks match the gap most often, then the first mode, while this is reversed in the interval spacing. Testing passes low-pass peaks from the first mode and gap equally. The changepoint matches the first mode. The low-pass spacing has a flat in half the runs, but it is usually not found significant and most are extended. Most of the level sections are also long. The sections appear in the left tail much more often than the right.

| W8 (abi) | | significant | | | | | detected | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | CPT | peak | | flat | | level | peak | | flat | |
| | | LP | Diw | LP | Diw | | LP | Diw | LP | Diw |
| # features | 170 | 25 | 117 | 40 | 0 | 557 | 327 | 817 | 90 | 0 |
| left tail | | | | | | 215 | | | | |
| mode 1 | 136 | 11 | 58 | 4 | | 10 | 123 | 402 | 22 | |
| gap 1 | 17 | 13 | 36 | 1 | | | 178 | 282 | 4 | |
| mode 2 | 17 | 1 | 23 | 9 | | 10 | 26 | 133 | 17 | |
| right tail | | | | | | 55 | | | | |
| extended | | | | 26 | | 267 | | | 47 | |

The spacing and changepoint results respect the symmetry of the W9 sample, but the level sections show a small imbalance between the left and right tail coverage. The detected low-pass peaks match the two gaps, with testing rejecting a third of them equally. Detected interval peaks are distributed except to the second mode, but testing rejects 90% of them equally. The changepoints also avoid the small second mode, and match the gaps at half the rate of the first and third mode. The low-pass spacing has a flat in half the trials, usually significant. Half extend over gap and mode, the remaining few fall equally within the first and last modes. Except for the tails, the level sections are also extended. There is a small preference for the initial tail.

| W9 (tri) | | significant | | | | | detected | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | CPT | peak | | flat | | level | peak | | flat | |
| | | LP | Diw | LP | Diw | | LP | Diw | LP | Diw |
| # features | 214 | 248 | 94 | 74 | 0 | 734 | 338 | 914 | 94 | 0 |
| left tail | | | | | | 206 | | | | |
| mode 1 | 71 | 3 | 19 | 24 | | 12 | 15 | 213 | 26 | |
| gap 1 | 37 | 118 | 17 | | | | 150 | 203 | 2 | |
| mode 2 | 6 | 19 | 12 | | | | 25 | 110 | | |
| gap 2 | 27 | 105 | 28 | 2 | | | 138 | 205 | 3 | |
| mode 3 | 73 | 3 | 18 | 21 | | 9 | 10 | 183 | 23 | |
| right tail | | | | | | 142 | | | | |
| extended | | | | 27 | | 365 | | | 40 | |

The symmetry of sample W10 is clear to see in the alignment of the detected peaks, evenly distributed between the four gaps. The low-pass spacing has a peak at each gap, although few are significant. The interval spacing peaks also are spread among the gaps, with one of four considered significant. They scatter a bit into the modes. The one or two changepoints also respect the symmetry but match the first or last mode. There are some flats, few significant, that are not confined to a gap or mode. Except for the level sections that cover the tails equally in each run, the majority are long.

| W10 (penta) | | significant | | | | | detected | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | CPT | peak | | flat | | level | peak | | flat | |
| | | LP | Diw | LP | Diw | | LP | Diw | LP | Diw |
| # features | 271 | 60 | 160 | 11 | 1 | 1187 | 759 | 765 | 55 | 1 |
| left tail | | | | | | 213 | | | | |
| mode 1 | 85 | | 7 | | | 16 | 1 | 22 | | |
| gap 1 | 25 | 22 | 35 | | | | 183 | 164 | | |
| mode 2 | 8 | | 5 | | | 1 | | 17 | 1 | |
| gap 2 | 10 | 14 | 43 | | | | 195 | 176 | 1 | |
| mode 3 | 6 | | 2 | | 1 | | | 10 | 5 | 1 |
| gap 3 | 7 | 6 | 35 | | | | 192 | 173 | | |
| mode 4 | 4 | | 5 | | | 2 | | 19 | 1 | |
| gap 4 | 26 | 18 | 23 | | | | 188 | 167 | | |
| mode 5 | 100 | | 5 | | | 16 | | 17 | | |
| right tail | | | | | | 198 | | | | |
| extended | | | | 11 | | 741 | | | 47 | |

W12 has four variates that get progressively narrower from left to right, atop a draw centered at the left that forms the first mode. This variate shifts the gaps right towards the modes, especially the fourth and fifth. The gaps are not centered in the separation but fall close to the mode, which means the matching will have a bias: the gaps appear to be wider on the left side with the minimum lying close to the right boundary, and the modes on the right. The test results show little consistency in their ability to resolve this setup. The low-pass spacing has peaks at the first and second gap, but they are not considered significant. The interval spacing also has peaks at these gaps, and at the fourth mode. Testing accepts 40% of them, keeping those at the fourth mode and half at the second gap, while rejecting most at the first gap. One changepoint appears aligned to the third mode, and in half the trials one to the second mode or second gap, and in a quarter the trials to the first mode or third gap. What is missing in all three tests are features aligned to the fifth mode or fourth gap, or the first gap. There are two flats detected in the low-pass spacing, one within the first or second mode and one extending over gaps and modes. Half pass testing, most long or covering the first mode. One level section corresponds to the initial left tail, but all others are long and not confined to an extremum.

| W12 (penta) | | significant | | | | | detected | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | CPT | peak | | flat | | level | peak | | flat | |
| | | LP | Diw | LP | Diw | | LP | Diw | LP | Diw |
| # features | 488 | 30 | 247 | 162 | 0 | 1129 | 401 | 635 | 375 | 0 |
| left tail | | | | | | 201 | | | | |
| mode 1 | 48 | | 4 | 59 | | 18 | 15 | 61 | 112 | |
| gap 1 | 18 | | 27 | 2 | | | 185 | 199 | 14 | |
| mode 2 | 72 | | 3 | 19 | | 4 | 2 | 19 | 98 | |
| gap 2 | 81 | 30 | 87 | | | | 199 | 198 | | |
| mode 3 | 213 | | 4 | | | 12 | | 11 | | |
| gap 3 | 48 | | 19 | | | | | 38 | | |
| mode 4 | 8 | | 103 | | | | | 109 | | |
| gap 4 | | | | | | | | | | |
| mode 5 | | | | | | 1 | | | | |
| right tail | | | | | | | | | | |
| extended | | | | 82 | | 893 | | | 151 | |

W14 is the chirp, five narrow, well-spaced draws. Especially in this sample we see the problem of transforming indices back to absolute coordinates, because the wide gaps will have no data, causing points to map to the edge of the mode. There is no way to get a coordinate in-between the modes to match the position of a gap. Overall, all detectors easily find the first two modes. Changepoints also pick up the third and fourth, and the level sections the fifth. Significant low-pass flats occur in the first two modes. They also exist in the interval spacing, although testing rejects most, leaving a significant flat in a third of the runs in the first mode. Half of the interval flats are long, but all are rejected. A short level section appears in the first, second, and fifth modes, and in half the trials the third, in addition to those corresponding to the tails. Half of the sections are extended.

| W14 (chirp) | | significant | | | | | detected | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | CPT | peak | | flat | | level | peak | | flat | |
| | | LP | Diw | LP | Diw | | LP | Diw | LP | Diw |
| # features | 841 | 231 | 238 | 400 | 68 | 2189 | 400 | 600 | 400 | 800 |
| left tail | | | | | | 227 | | | | |
| mode 1 | 225 | 200 | 134 | 200 | 68 | 281 | 200 | 198 | 200 | 200 |
| gap 1 | | | | | | | | | | |
| mode 2 | 228 | 31 | 74 | 200 | | 233 | 200 | 184 | 200 | 200 |
| gap 2 | | | | | | | | | | |
| mode 3 | 203 | | 6 | | | 91 | | 18 | | |
| gap 3 | | | | | | | | | | |
| mode 4 | 185 | | 24 | | | 7 | | 200 | | |
| gap 4 | | | | | | | | | | |
| mode 5 | | | | | | 318 | | | | |
| right tail | | | | | | 75 | | | | |
| extended | | | | | | 957 | | | | 400 |

## A Samples

Samples A1 through A10 are uni-modal and outside this analysis.

The significant low-pass peaks find the gap in sample A11 in a quarter of the trials, while the significant interval peaks split between the gap and first mode in half the runs. There is a changepoint, also in half the trials, matching the first mode. Testing rejects 80% of the detected peaks in either spacing, especially those matching the first mode. But this still leaves interval peaks scattered between modes and gap. There is a flat in the low-pass spacing in half the trials, usually significant. Half are short, falling within the first mode, while the other half extend further. The short level sections correspond to the tails, favoring the initial with its broader distribution, but again half are long.

| A11 (abi) | | significant | | | | | detected | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | CPT | peak | | flat | | level | peak | | flat | |
| | | LP | Diw | LP | Diw | | LP | Diw | LP | Diw |
| # features | 162 | 61 | 136 | 51 | 0 | 558 | 319 | 783 | 82 | 0 |
| left tail | | | | | | 217 | | | | |
| mode 1 | 110 | 9 | 69 | 23 | | 19 | 122 | 417 | 35 | |
| gap 1 | 30 | 52 | 49 | 1 | | | 187 | 294 | 2 | |
| mode 2 | 22 | | 18 | 8 | | 11 | 10 | 72 | 17 | |
| right tail | | | | | | 37 | | | | |
| extended | | | | 19 | | 274 | | | 28 | |

The peaks in the spacings match the gap in sample A12 most strongly, although the interval spacing also places a peak in each mode in each run. The rejection rate from testing is high, passing peaks at the gap and suppressing those at the modes. The significant interval peaks are distributed and their accuracy poor. Changepoints align to the gap or first mode equally, or the second mode at half the rate. There is one low-pass flat in half the runs, usually significant, that lies within the first mode. The level sections outside the tails are long. The few short sections not in a tail, 6% of all, align to the mode.

| A12 (abi) | | significant | | | | | detected | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | CPT | peak | | flat | | level | peak | | flat | |
| | | LP | Diw | LP | Diw | | LP | Diw | LP | Diw |
| # features | 274 | 131 | 93 | 69 | 0 | 808 | 270 | 689 | 93 | 0 |
| left tail | | | | | | 195 | | | | |
| mode 1 | 107 | | 19 | 62 | | 52 | 43 | 160 | 77 | |
| gap 1 | 119 | 125 | 41 | | | 1 | 201 | 283 | | |
| mode 2 | 48 | 6 | 33 | 5 | | 11 | 26 | 246 | 8 | |
| right tail | | | | | | 126 | | | | |
| extended | | | | 2 | | 423 | | | 8 | |

The second draw in A13 is too small to support flats, so there is one significant in the low-pass spacing and three or four in the interval spacing that lie within the first mode. Most detected flats are significant, except those marking the initial tail, which are usually rejected. Two short level sections split the first mode, above and beyond the two in the left tail, and one section appears in the second mode, not tied to a tail. The low-pass spacing has no peaks, the interval a few, most significant, that match the gap. There are two changepoints that align to the first mode and one to the gap. The tests do not pick up the difference between the modes, but the first is large enough to support multiple flats and level sections.

| A13 (abi) | | significant | | | | | detected | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | CPT | peak | | flat | | level | peak | | flat | |
| | | LP | Diw | LP | Diw | | LP | Diw | LP | Diw |
| # features | 592 | 0 | 38 | 246 | 672 | 1339 | 0 | 46 | 331 | 762 |
| left tail | | | | 14 | | 451 | | | 65 | 328 |
| mode 1 | 418 | | 10 | 232 | 672 | 364 | | 15 | 266 | 434 |
| gap 1 | 170 | | 28 | | | 2 | | 31 | | |
| mode 2 | 4 | | | | | 147 | | | | |
| right tail | | | | | | | | | | |
| extended | | | | | | 375 | | | | |

The points match the gap in sample A14 best: a single significant peak in the low-pass spacing, a significant interval peak in half the runs, and a changepoint. The interval spacing has a second detected peak spread equally between the modes and gap, but testing rejects it. There is a second changepoint matching the first mode, and in half the runs one aligned to the second. There are one or two flats, usually significant, that lie within the modes and favor the second. A short level section lies in the second mode in half the trials, in addition to sections in each tail. There are also two or three long sections.

| A14 (abi) | | significant | | | | | detected | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | CPT | peak | | flat | | level | peak | | flat | |
| | | LP | Diw | LP | Diw | | LP | Diw | LP | Diw |
| # features | 518 | 200 | 129 | 243 | 0 | 1057 | 221 | 356 | 291 | 0 |
| left tail | | | | | | 239 | | | | |
| mode 1 | 188 | | 14 | 95 | | 33 | 5 | 68 | 118 | |
| gap 1 | 211 | 200 | 98 | | | | 200 | 208 | 1 | |
| mode 2 | 119 | | 17 | 147 | | 96 | 16 | 80 | 171 | |
| right tail | | | | | | 178 | | | | |
| extended | | | | 1 | | 511 | | | 1 | |

The low-pass peaks for sample A15 primarily match the gap, but none are significant. The interval peaks also match the gap at twice the rate of either mode, but testing rejects 85%, leaving one occasional peak distributed between the modes and gap. What looks to be good accuracy for the detected peaks in either spacing disappears during testing. One changepoint per run marks a mode, split between the two. If a flat exists in the low-pass spacing, significant in half the trials, then it is extended, or at a lower rate covers one of the modes. Short level sections correspond to both tails; the large background variate generates values beyond the two modes and creates the tails. The remaining sections are long.

| A15 (abi) | | significant | | | | | detected | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | CPT | peak | | flat | | level | peak | | flat | |
| | | LP | Diw | LP | Diw | | LP | Diw | LP | Diw |
| # features | 227 | 0 | 76 | 83 | 0 | 727 | 303 | 568 | 140 | 0 |
| left tail | | | | | | 209 | | | | |
| mode 1 | 132 | | 22 | 10 | | 24 | 57 | 161 | 20 | |
| gap 1 | 3 | | 30 | 5 | | | 200 | 280 | 10 | |
| mode 2 | 92 | | 24 | 21 | | 21 | 46 | 127 | 33 | |
| right tail | | | | | | 188 | | | | |
| extended | | | | 47 | | 285 | | | 77 | |

The gap in sample A16 is found in both spacings and the changepoints. The significant low-pass peaks match the gap, although there are also a few detected peaks aligned to the first, larger mode. In half the runs the interval spacing has a significant peak that matches the gap, with testing rejecting most of the peaks in the modes, especially the first. Two changepoints match the gap and first mode, the latter at a higher rate. There is one low-pass flat, significant, that falls within the first mode. There is also a short level section in that mode, and in half the trials one in the second. Like A15, the background added to the first variate creates one or two sections in the initial tail, but there are none in the trailing. Two sections are long.

| A16 (abi) | | significant | | | | | detected | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | CPT | peak | | flat | | level | peak | | flat | |
| | | LP | Diw | LP | Diw | | LP | Diw | LP | Diw |
| # features | 447 | 200 | 116 | 232 | 0 | 987 | 240 | 344 | 232 | 0 |
| left tail | | | | | | 319 | | | | |
| mode 1 | 243 | | 21 | 232 | | 163 | 40 | 139 | 232 | |
| gap 1 | 195 | 200 | 84 | | | 3 | 200 | 169 | | |
| mode 2 | 9 | | 11 | | | 79 | | 36 | | |
| right tail | | | | | | | | | | |
| extended | | | | | | 423 | | | | |

Sample A17 is another draw from two normals, with a smaller separation and standard deviation than M1, H1, N2, X1, or W6. It is a little harder to resolve, but the results respect the symmetry of the setup. The detected peaks in either spacing are about the same, but testing accepts fewer cases. The significant low-pass peaks match the gap, while those in the interval spacing are evenly distributed between the gap and modes. A changepoint appears in half the trials, avoiding the gap. Occasionally the low-pass spacing has a flat, significant, in one of the modes or extended. Half of the level sections are long. The other half correspond to a tail, the initial twice as often as the final.

| A17 (bi) | | significant | | | | | detected | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | CPT | peak | | flat | | level | peak | | flat | |
| | | LP | Diw | LP | Diw | | LP | Diw | LP | Diw |
| # features | 123 | 83 | 109 | 34 | 0 | 589 | 330 | 1043 | 35 | 0 |
| left tail | | | | | | 201 | | | | |
| mode 1 | 56 | 9 | 29 | 10 | | 12 | 62 | 341 | 11 | |
| gap 1 | 17 | 63 | 38 | 1 | | 1 | 195 | 361 | 1 | |
| mode 2 | 50 | 11 | 42 | 11 | | 13 | 73 | 341 | 11 | |
| right tail | | | | | | 103 | | | | |
| extended | | | | 12 | | 259 | | | 12 | |

A18 is symmetric and easily resolved. The low-pass spacing has one peak, significant, that matches the gap. There are one or two interval peaks, but testing rejects one. The significant peak matches the gap, or one of the modes at a third the rate, so its accuracy is moderate. There are one or two changepoints nearly as strongly aligned to the modes as to the gap. Each mode holds a significant low-pass flat. There is also a short level section associated with each, above those covering the tails, 2:1 in favor of the initial. One or two long level sections split the rest of the sample.

| A18 (bi) | | significant | | | | | detected | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | CPT | peak | | flat | | level | peak | | flat | |
| | | LP | Diw | LP | Diw | | LP | Diw | LP | Diw |
| # features | 316 | 200 | 95 | 388 | 0 | 976 | 202 | 288 | 389 | 0 |
| left tail | | | | 3 | | 183 | | | 3 | |
| mode 1 | 91 | 23 | 14 | 191 | | 172 | 25 | 79 | 191 | |
| gap 1 | 135 | 177 | 61 | | | | 177 | 123 | | |
| mode 2 | 90 | | 20 | 194 | | 233 | | 86 | 195 | |
| right tail | | | | | | 85 | | | | |
| extended | | | | | | 303 | | | | |

A19 is even better separated than A18, but the large separation of the two very tight variates

means the gap contains no points, which distorts the mapping of feature indices to coordinates, moving the gap to the edge of the first mode. We see this by having no point features in the gap and in the lack of low-pass peaks for the second mode. The interval peak splits between the modes, with testing rejecting one per trial. The changepoint also splits between the modes but shows a preference for the first. In contrast, flats are symmetric, since they fall within the mode and are unaffected by the separation. There is a significant flat in each mode, at a somewhat lower rate in the interval spacing. There are short level sections in each mode at equal rates, plus those corresponding to the tails, preferring the initial by 2:1. A third of the sections are long.

| A19 (bi) | | significant | | | | | detected | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | CPT | peak | | flat | | level | peak | | flat | |
| | | LP | Diw | LP | Diw | | LP | Diw | LP | Diw |
| # features | 265 | 200 | 67 | 400 | 272 | 1052 | 200 | 212 | 400 | 345 |
| left tail | | | | | | 190 | | | | |
| mode 1 | 167 | 200 | 34 | 200 | 130 | 219 | 200 | 112 | 200 | 169 |
| gap 1 | | | | | | | | | | |
| mode 2 | 98 | | 33 | 200 | 142 | 272 | | 100 | 200 | 176 |
| right tail | | | | | | 94 | | | | |
| extended | | | | | | 277 | | | | |

A20 adds two beta variates to the A19 sample to generate some points in the separation. This helps align features to the gap. Although there is still an imbalance favoring the second mode, it is not as strong as in A19. The low-pass peak, significant, now matches the gap. The interval peak, also significant, matches the gap or second mode equally, while the occasional alignment to the first mode is unchanged. The changepoints are not symmetric, with two aligning to the second mode and one splitting between the first mode and gap. The flats in either spacing remain balanced, falling within each mode. Any detected are usually significant; testing rejects all long interval flats. The sample generates many level sections, mostly short. Two or three cover the second mode and two the first, plus two in each tail. None of the flats or sections lie in the gap. Relatively few level sections, just 15%, are long.

| A20 (bi) | | significant | | | | | detected | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | CPT | peak | | flat | | level | peak | | flat | |
| | | LP | Diw | LP | Diw | | LP | Diw | LP | Diw |
| # features | 632 | 200 | 202 | 400 | 406 | 2568 | 200 | 203 | 400 | 597 |
| left tail | | | | | 2 | 436 | | | | 2 |
| mode 1 | 125 | | 32 | 200 | 202 | 391 | | 33 | 200 | 202 |
| gap 1 | 99 | 177 | 86 | | | 8 | 177 | 86 | | |
| mode 2 | 408 | 23 | 84 | 200 | 201 | 5878 | 23 | 84 | 200 | 201 |
| right tail | | | | | 1 | 761 | | | | 2 |
| extended | | | | | | 385 | | | | 190 |

Point features disperse among the extrema of sample A21. Accuracy in either spacing is poor. The significant low-pass peak, appearing in every other trial, matches the first and deepest gap half the time, but distributes between the other gap and modes the other half. The significant interval peak matches the first or third mode, but the counts for the gaps and second mode are also high. Testing rejects most of the detected peaks in both spacings. The strongest alignments are those rejected the least. Changepoints when they occur, fewer than one per run, match the first and third modes, with lower counts for the gaps and second mode, similar to the values from the low-pass spacing. When

a flat appears in the low-pass spacing, usually significant, it is long and covers mode and gap. The level sections are also long, except those corresponding to the tails, which are unbalanced 2:1 towards the initial.

| A21 (tri) | | significant | | | | | detected | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | CPT | peak | | flat | | level | peak | | flat | |
| | | LP | Diw | LP | Diw | | LP | Diw | LP | Diw |
| # features | 157 | 104 | 183 | 17 | 0 | 538 | 369 | 1167 | 21 | 0 |
| left tail | | | | | | 183 | | | | |
| mode 1 | 66 | 10 | 55 | | | 18 | 62 | 292 | | |
| gap 1 | 16 | 50 | 37 | | | | 136 | 276 | 1 | |
| mode 2 | 13 | 11 | 25 | | | | 74 | 202 | | |
| gap 2 | 11 | 18 | 19 | | | | 65 | 168 | | |
| mode 3 | 51 | 15 | 47 | | | 4 | 32 | 229 | | |
| right tail | | | | | | 73 | | | | |
| extended | | | | 17 | | 260 | | | 20 | |

The second mode of sample A22 is large and tight enough to contain a flat that appears in both the low-pass and interval spacing, always significant. Two short level sections also appear within it. The third draw is too small to generate a flat despite being narrow, being equal to the minimum length, but does contain one level section. One short section lies at the start, mostly in the tail rather than the first mode. Half the sections are long The low-pass spacing has one peak, significant, that matches the first gap. One interval peak of two detected passes testing, and is distributed, matching the first mode most often. Changepoints align to the first gap or the second mode in each run.

| A22 (tri) | | significant | | | | | detected | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | CPT | peak | | flat | | level | peak | | flat | |
| | | LP | Diw | LP | Diw | | LP | Diw | LP | Diw |
| # features | 516 | 200 | 173 | 206 | 182 | 1404 | 200 | 376 | 206 | 182 |
| left tail | | | | | | 99 | | | | |
| mode 1 | 25 | 29 | 113 | | | 46 | 29 | 212 | | |
| gap 1 | 197 | 171 | 25 | | | 1 | 171 | 30 | | |
| mode 2 | 234 | | 14 | 206 | 182 | 393 | | 32 | 206 | 182 |
| gap 2 | 44 | | 4 | | | | | 17 | | |
| mode 3 | 16 | | 17 | | | 169 | | 85 | | |
| right tail | | | | | | | | | | |
| extended | | | | | | 696 | | | | |

Sample A23 has two symmetric variates and a third, small and very tight, deep in the tail of the second draw. The significant point features match this second gap. There are also high counts aligned with the first two modes for the changepoints or interval peaks. Testing is strict, rejecting half the detected low-pass peaks, especially those at the first gap, and three quarters in the interval spacing, also most strongly at the first gap but also the modes surrounding it. There is one flat in the low-pass spacing, half of the time in one of the first two modes, the other half extended. The third draw is too small to support a flat. Testing rejects half the detected flats. The level sections are long, with a quarter covering the left tail. There is a small imbalance in the changepoint and low-pass flat counts between the first and second mode, possibly an effect of the third variate.

| A23 (tri) | | significant | | | | | detected | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | CPT | peak | | flat | | level | peak | | flat | |
| | | LP | Diw | LP | Diw | | LP | Diw | LP | Diw |
| # features | 391 | 201 | 256 | 168 | 0 | 871 | 436 | 1010 | 335 | 0 |
| left tail | | | | | | 206 | | | | |
| mode 1 | 64 | | 32 | 28 | | 21 | 34 | 261 | 85 | |
| gap 1 | 13 | 4 | 31 | 7 | | | 180 | 309 | 33 | |
| mode 2 | 92 | | 39 | 48 | | 3 | 25 | 251 | 97 | |
| gap 2 | 213 | 197 | 152 | | | 1 | 197 | 187 | | |
| mode 3 | 9 | | 2 | | | 37 | | 2 | | |
| right tail | | | | | | 2 | | | | |
| extended | | | | 85 | | 601 | | | 120 | |

Although the low-pass and interval spacing have peaks matching both gaps in sample A24, a large number also align to one of the three modes. The accuracy is poor. Testing rejects all of those in the low-pass spacing and 90% of the interval peaks. A changepoint matches the first or third mode at equal rates; all features respect the symmetry of the sample. There is one flat in the low-pass spacing that is significant in half the trials, but it is long and extends beyond a mode. The level sections divide equally between the tails and longer intervals.

| A24 (tri) | | significant | | | | | detected | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | CPT | peak | | flat | | level | peak | | flat | |
| | | LP | Diw | LP | Diw | | LP | Diw | LP | Diw |
| # features | 189 | 0 | 44 | 91 | 0 | 717 | 337 | 432 | 179 | 0 |
| left tail | | | | | | 210 | | | | |
| mode 1 | 95 | | 3 | | | 17 | 20 | 40 | 1 | |
| gap 1 | | | 6 | | | | 129 | 144 | 1 | |
| mode 2 | 1 | | 6 | | | | 35 | 49 | | |
| gap 2 | | | 12 | | | | 141 | 144 | | |
| mode 3 | 93 | | 17 | 1 | | 7 | 12 | 55 | 3 | |
| right tail | | | | | | 203 | | | | |
| extended | | | | 90 | | 280 | | | 174 | |

The third variate in sample A25 is much larger and wider than in A24, and this appears in the matching and coverage counts. There is one significant peak at the second gap in both spacings. Detected peaks mark the first gap but fail testing. The interval spacing also finds two additional peaks at the modes, which testing rejects. There is a changepoint at the second gap, and one in the second or third mode surrounding it. Half the runs have a low-pass flat, significant, in the third mode. Most level sections are long, although there are some short sections in the first and last modes, above those for the tails, which are biased towards the trailing end.

| A25 (tri) | | significant | | | | | detected | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | CPT | peak | | flat | | level | peak | | flat | |
| | | LP | Diw | LP | Diw | | LP | Diw | LP | Diw |
| # features | 382 | 215 | 223 | 111 | 0 | 828 | 374 | 898 | 112 | 0 |
| left tail | | | | | | 50 | | | | |
| mode 1 | 1 | 1 | 30 | | | 40 | 8 | 107 | | |
| gap 1 | 5 | 14 | 24 | | | | 116 | 177 | | |
| mode 2 | 68 | 1 | 18 | | | | 2 | 87 | | |
| gap 2 | 217 | 199 | 108 | | | | 199 | 193 | | |
| mode 3 | 91 | | 43 | 105 | | 88 | 49 | 334 | 105 | |
| right tail | | | | | | 138 | | | | |
| extended | | | | 6 | | 512 | | | 7 | |

**K Samples**

The symmetry of sample K1 is clear in the matching and coverage counts. The low-pass and interval spacings detect peaks at the three gaps, most strongly with the second. The interval spacing has a peak in the second or third mode, not accepted. Testing also suppresses the number of significant peaks at the first or second gap. The peaks are accurately placed, although the first and third gap counts are weak. Changepoints fall off from the central gap, with half as many aligned to the second or third mode as to the second gap, and half again to the outer modes or gaps. The low-pass spacing has two flats, many within the second or third mode, but with more than half extending over a gap and mode. Testing rejects most, especially those located within the modes, leaving long flats. Most level sections are also long, other than those corresponding to tails. There is a strong 4:1 imbalance between the left and right tails.

| K1 (quad) | | significant | | | | | detected | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | CPT | peak | | flat | | level | peak | | flat | |
| | | LP | Diw | LP | Diw | | LP | Diw | LP | Diw |
| # features | 658 | 338 | 240 | 57 | 0 | 1181 | 448 | 736 | 400 | 0 |
| left tail | | | | | | 192 | | | | |
| mode 1 | 67 | | 1 | | | 9 | | 7 | 20 | |
| gap 1 | 60 | 65 | 30 | | | | 112 | 164 | 4 | |
| mode 2 | 90 | 5 | 8 | 5 | | | 9 | 95 | 61 | |
| gap 2 | 217 | 200 | 151 | | | | 200 | 202 | | |
| mode 3 | 98 | 7 | 6 | 2 | | 2 | 12 | 96 | 68 | |
| gap 3 | 62 | 61 | 40 | | | | 115 | 161 | 1 | |
| mode 4 | 64 | | 4 | | | 13 | | 11 | 25 | |
| right tail | | | | | | 52 | | | | |
| extended | | | | 50 | | 913 | | | 221 | |

Peaks in the spacings match the extrema of sample K2 differently. Those in the low-pass spacing align with the gaps, with the third harder to detect. The interval spacing aligns most strongly with the gaps, but also each of the modes at half the rate. Testing does affect the distribution. It rejects some of the low-pass features matching the second gap, half of those the first and almost all of the third; put differently, significant low-pass peaks match the second gap most often, then the first, then the third at a noticeably lower rate. Just one interval peak of five passes testing, with those matching the third and fourth mode and the gap between suppressed the most. The significant interval peaks mark the gaps, but also disperse into the modes at half the rate. Its accuracy is poor. Changepoints

largely avoid the third and fourth mode and their gap, matching the first gap and second mode most often. There are no flats, and the level sections are long, not confined to one mode or gap. Relatively few sections correspond to the left tail.

| K2 (quad) | | significant | | | | | detected | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | CPT | peak | | flat | | level | peak | | flat | |
| | | LP | Diw | LP | Diw | | LP | Diw | LP | Diw |
| # features | 475 | 338 | 201 | 3 | 0 | 613 | 559 | 1029 | 3 | 0 |
| left tail | | | | | | 63 | | | | |
| mode 1 | 77 | | 16 | | | 41 | | 81 | | |
| gap 1 | 116 | 111 | 44 | | | | 200 | 188 | | |
| mode 2 | 105 | | 21 | | | 1 | | 102 | | |
| gap 2 | 89 | 159 | 48 | | | | 184 | 198 | | |
| mode 3 | 34 | 3 | 17 | | | | 18 | 125 | | |
| gap 3 | 35 | 65 | 36 | | | | 155 | 214 | | |
| mode 4 | 19 | | 19 | | | 4 | 2 | 121 | | |
| right tail | | | | | | 2 | | | | |
| extended | | | | 3 | | 502 | | | 3 | |

The feature positions reflect the symmetry of K3, although there a little spread in the alignment of the detected peaks, moreso than in other symmetric samples. Almost all peaks in the low-pass spacing are significant, matching the central second gap most, but also the left or right modes and gaps. Testing rejects two interval peaks, leaving one significant that matches the second gap and at half the rate the first or fourth mode, with some aligning to the remaining. The tests suppress peaks aligned to the outer gaps most strongly. There are three changepoints per run, falling off slowly to either side of the central gap. The few flats in the low-pass spacing are significant and at equal rates cover the second or third mode. Most level sections are long, with a few in the left tail.

| K3 (quad) | | significant | | | | | detected | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | CPT | peak | | flat | | level | peak | | flat | |
| | | LP | Diw | LP | Diw | | LP | Diw | LP | Diw |
| # features | 589 | 298 | 240 | 34 | 0 | 1210 | 311 | 671 | 34 | 0 |
| left tail | | | | | | 135 | | | | |
| mode 1 | 33 | 23 | 42 | | | 27 | 24 | 137 | | |
| gap 1 | 77 | 22 | 19 | | | | 27 | 93 | | |
| mode 2 | 115 | | 8 | 12 | | 11 | | 20 | 12 | |
| gap 2 | 140 | 199 | 98 | | | | 200 | 159 | | |
| mode 3 | 117 | | 13 | 16 | | 16 | | 22 | 16 | |
| gap 3 | 71 | 38 | 16 | | | | 43 | 85 | | |
| mode 4 | 36 | 16 | 44 | | | 5 | 17 | 155 | | |
| right tail | | | | | | 25 | | | | |
| extended | | | | 6 | | 991 | | | 6 | |

The gap between the first and second, and third and fourth modes of K4 is larger than in K3, and the alignment of points to the gaps is stronger. The low-pass spacing now matches any gap in more than half the runs. Detected peaks in the interval spacing fall in the outer modes and gaps, but also mark the central gap in half the runs. Testing results in a fairly uniform distribution, so accuracy is poor. The strong association with the second gap in K3 has disappeared. Changepoints match the three gaps best, then to the second and third mode. Overall, though, the changepoints really

avoid the first and fourth mode, spreading equally between the remaining extrema and preferring the central gap a bit. The level sections do not change compared to K3.

| K4 (quad) | | significant | | | | | detected | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | CPT | peak | | flat | | level | peak | | flat | |
| | | LP | Diw | LP | Diw | | LP | Diw | LP | Diw |
| # features | 583 | 481 | 266 | 10 | 0 | 1201 | 500 | 827 | 10 | 0 |
| left tail | | | | | | 138 | | | | |
| mode 1 | 12 | 20 | 27 | 1 | | 29 | 20 | 147 | 1 | |
| gap 1 | 103 | 128 | 59 | | | | 136 | 160 | | |
| mode 2 | 98 | 9 | 20 | | | 6 | 9 | 42 | | |
| gap 2 | 155 | 189 | 53 | | | | 189 | 91 | | |
| mode 3 | 93 | 2 | 26 | 1 | | 3 | 2 | 67 | | |
| gap 3 | 95 | 124 | 40 | | | | 135 | 139 | | |
| mode 4 | 27 | 9 | 41 | | | 3 | 9 | 181 | | |
| right tail | | | | | | 15 | | | | |
| extended | | | | 8 | | 1007 | | | 8 | |

The significant low-pass peak picks out the three gaps in sample K5, favoring the central one. Testing rejects half of the detected peaks that align with this gap while accepting all others. The detected interval peaks match the second gap most often, but one also appears in the first and last mode and gap. Testing passes a third of the peaks at the central gap and a fifth at the edges. Changepoints associate most strongly with the second gap and third mode, but still match the other modes and gaps. Most of the level sections are long; those that are short favor the left tail over the right. The low-pass spacing has a flat in half the trials, but most are rejected as insignificant.

| K5 (quad) | | significant | | | | | detected | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | CPT | peak | | flat | | level | peak | | flat | |
| | | LP | Diw | LP | Diw | | LP | Diw | LP | Diw |
| # features | 534 | 228 | 191 | 24 | 0 | 1142 | 347 | 672 | 82 | 0 |
| left tail | | | | | | 176 | | | | |
| mode 1 | 58 | 20 | 33 | 2 | | 15 | 20 | 138 | 2 | |
| gap 1 | 70 | 51 | 22 | | | | 56 | 116 | | |
| mode 2 | 77 | | 5 | 8 | | 1 | | 14 | 25 | |
| gap 2 | 108 | 87 | 69 | | | | 200 | 187 | | |
| mode 3 | 111 | | | 3 | | 1 | | 2 | 12 | |
| gap 3 | 73 | 58 | 25 | | | | 59 | 84 | | |
| mode 4 | 37 | 12 | 37 | | | 4 | 12 | 131 | | |
| right tail | | | | | | 21 | | | | |
| extended | | | | 11 | | 924 | | | 43 | |

We detect peaks at each gap of K6. Testing rejects a third of them, moreso from those matching the third gap, but the accuracy is good. The interval spacing also has peaks at each gap, but in half the trials there is an additional peak matching the first and fourth modes. Testing rejects more than half the detected peaks, primarily from the gaps, equalizing the rate they align with the outer modes and gaps. There are four changepoints, two in the first and fourth modes and two distributed among the other extrema, less so for the second mode or third gap. The level sections are long and do not fall within a mode. Fewer than a quarter correspond to tails and favor the leading over the trailing.

| K6 (quad) | | significant | | | | | detected | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | CPT | peak | | flat | | level | peak | | flat | |
| | | LP | Diw | LP | Diw | | LP | Diw | LP | Diw |
| # features | 808 | 422 | 307 | 18 | 0 | 1320 | 592 | 794 | 18 | 0 |
| left tail | | | | | | 194 | | | | |
| mode 1 | 237 | 2 | 58 | 1 | | 18 | 2 | 84 | 1 | |
| gap 1 | 90 | 157 | 62 | | | | 200 | 186 | | |
| mode 2 | 51 | | 5 | | | 1 | | 29 | | |
| gap 2 | 104 | 168 | 62 | | | | 200 | 181 | | |
| mode 3 | 92 | | 5 | 10 | | | | 22 | 10 | |
| gap 3 | 62 | 95 | 51 | | | | 190 | 187 | | |
| mode 4 | 172 | | 64 | | | 26 | | 105 | | |
| right tail | | | | | | 132 | | | | |
| extended | | | | 7 | | 949 | | | 7 | |

K7 has four Weibull variates, which limits the lower tail of the total sample because there are no negative values; we see no level sections here. The detected low-pass peaks match the gaps, least often the first. Testing rejects two-thirds of them unequally, keeping most matching the third mode at the cost of those at the second gap. The interval peaks also match the three gaps, plus the fourth mode. Testing accepts those aligned to the second gap most, but leaves significant peaks at half the rate at the third gap and fourth mode. The changepoints align to the right half of the sample, best matched to the second and third gap, but also to the third and fourth mode. The low-pass flats that exist, usually significant, are mostly long, with the short confined to the third mode. The level sections are the same.

| K7 (quad) | | significant | | | | | detected | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | CPT | peak | | flat | | level | peak | | flat | |
| | | LP | Diw | LP | Diw | | LP | Diw | LP | Diw |
| # features | 514 | 158 | 176 | 61 | 0 | 782 | 497 | 828 | 76 | 0 |
| left tail | | | | | | 1 | | | | |
| mode 1 | 7 | | 7 | | | 9 | 1 | 41 | | |
| gap 1 | 12 | 3 | 20 | | | | 147 | 179 | | |
| mode 2 | 42 | | 6 | | | 1 | | 33 | | |
| gap 2 | 139 | 41 | 67 | | | | 200 | 203 | | |
| mode 3 | 97 | | 1 | 22 | | 4 | 2 | 33 | 29 | |
| gap 3 | 133 | 88 | 33 | | | | 119 | 160 | | |
| mode 4 | 84 | 26 | 42 | 1 | | 13 | 28 | 179 | 2 | |
| right tail | | | | | | 50 | | | | |
| extended | | | | 38 | | 704 | | | 45 | |

Peaks and changepoints do not have a consistent alignment with the extrema of K8. Both the low-pass and interval spacing detect peaks at the second and third gap, but after testing only the low-pass peak matching the third gap is repeatedly significant. The interval spacing also has a peak matching the first mode that testing mostly rejects. At lesser rates there are low-pass peaks matching the first mode, and interval peaks at all but the second mode. There is a changepoint at the first gap, and others spread between the second and third mode and third gap. The level sections are long, with 15% covering the initial left tail.

| K8 (quad) | | significant | | | | | detected | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | CPT | peak | | flat | | level | peak | | flat | |
| | | LP | Diw | LP | Diw | | LP | Diw | LP | Diw |
| # features | 429 | 278 | 274 | 11 | 0 | 881 | 475 | 878 | 12 | 0 |
| left tail | | | | | | 119 | | | | |
| mode 1 | 16 | 68 | 86 | | | 26 | 68 | 213 | | |
| gap 1 | 169 | 29 | 27 | | | | 30 | 79 | | |
| mode 2 | 50 | | 5 | | | 1 | | 22 | | |
| gap 2 | 39 | 30 | 69 | | | | 195 | 212 | | |
| mode 3 | 64 | | 15 | 1 | | | 5 | 69 | 1 | |
| gap 3 | 84 | 151 | 51 | | | | 177 | 201 | | |
| mode 4 | 7 | | 21 | | | 4 | | 82 | | |
| right tail | | | | | | | | | | |
| extended | | | | 10 | | 731 | | | 11 | |

Sample K9 has discrete data, which creates a level section at each step; we ignore these features. The number of changepoints is not unreasonably high and mostly align to the first mode or gap, or weakly with the third. They do not match the second gap. The steps also create a large number of interval peaks, although testing rejects three quarters of them equally among the extrema, leaving as significant a peak matching the third mode and in half the trials one for the first mode, but also at a high rate with the other modes and gaps. The accuracy is poor. The low-pass filtering does work with discrete data, and we find peaks at each gap in it. Testing accepts the one matching the first mode, and weakly the third. The handful of flats that exist cover the first mode.

| K9 (quad) | | significant | | | | | detected | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | CPT | peak | | flat | | level | peak | | flat | |
| | | LP | Diw | LP | Diw | | LP | Diw | LP | Diw |
| # features | 466 | 317 | 469 | 6 | 0 | 1589 | 562 | 1935 | 6 | 0 |
| left tail | | | | | | 347 | | | | |
| mode 1 | 144 | 9 | 111 | 5 | | 529 | 9 | 540 | 5 | |
| gap 1 | 123 | 190 | 27 | | | 1 | 191 | 125 | | |
| mode 2 | 24 | 3 | 54 | | | 18 | 10 | 202 | | |
| gap 2 | 40 | 34 | 55 | | | 1 | 185 | 220 | | |
| mode 3 | 51 | 1 | 154 | 1 | | 57 | 4 | 523 | 1 | |
| gap 3 | 79 | 80 | 44 | | | | 163 | 210 | | |
| mode 4 | 5 | | 24 | | | 1 | | 115 | | |
| right tail | | | | | | 1 | | | | |
| extended | | | | | | 634 | | | | |

K10 has added a variate in the middle of sample K1 that points avoid. The low-pass spacing has detected and found significant peaks at the second and third gaps, and in half the trials at the first and fourth. The interval spacing matches the same peaks, the first and fourth more strongly, and detects peaks at the second and fourth modes. Testing rejects most, leaving a weak association with the middle gaps. The changepoints match the inner gap pair most often and the outer at half the rate. There are also changepoints at all five modes, most strongly at the third, the others equally. All counts respect the symmetry of the sample. There is a low-pass flat in more than half the trials, usually accepted as significant, but these flats are long and not contained by a single mode. The level sections are also extended, with a quarter covering the tails, the initial much more often than the final.

| K10 (penta) | | significant | | | | | detected | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | CPT | peak | | flat | | level | peak | | flat | |
| | | LP | Diw | LP | Diw | | LP | Diw | LP | Diw |
| # features | 691 | 578 | 248 | 88 | 0 | 954 | 585 | 833 | 143 | 0 |
| left tail | | | | | | 185 | | | | |
| mode 1 | 55 | | 1 | | | 21 | | 10 | | |
| gap 1 | 74 | 81 | 23 | | | | 85 | 138 | 1 | |
| mode 2 | 48 | 10 | 11 | 14 | | | 10 | 78 | 19 | |
| gap 2 | 130 | 196 | 76 | | | | 196 | 195 | | |
| mode 3 | 91 | | | | | 1 | | 1 | | |
| gap 3 | 117 | 195 | 79 | | | | 195 | 187 | | |
| mode 4 | 37 | 12 | 19 | 4 | | | 13 | 85 | 14 | |
| gap 4 | 86 | 84 | 37 | | | | 86 | 135 | 3 | |
| mode 5 | 53 | | 2 | | | 10 | | 4 | 1 | |
| right tail | | | | | | 26 | | | | |
| extended | | | | 70 | | 711 | | | 105 | |

The peaks detected in the low-pass spacing match the gaps in sample K11, while in the interval spacing they match all but the first gap. Except for the first mode, however, interval peaks align to all extrema in at least a third of the runs; there is a large background rate. Testing rejects most of these features. Half of the low-pass peaks are significant, with higher rejection rates for the fourth and particularly the first gap. The accuracy is good, if uneven. One quarter of the interval peaks pass testing, leaving a rather uniform allocation between all but the third, fourth, and fifth modes and fourth gap. Accuracy is poor. The changepoints are equally dispersed. There is one flat in the low-pass spacing, significant, that lies within the first mode. Most level sections are long, although there are some confined to the first mode, above those for the left tail. Level sections do not appear in the right tail.

| K11 (penta) | | significant | | | | | detected | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | CPT | peak | | flat | | level | peak | | flat | |
| | | LP | Diw | LP | Diw | | LP | Diw | LP | Diw |
| # features | 541 | 377 | 202 | 173 | 1 | 815 | 700 | 938 | 173 | 1 |
| left tail | | | | | | 122 | | | | |
| mode 1 | 95 | | 2 | 171 | 1 | 63 | | 15 | 171 | 1 |
| gap 1 | 130 | 29 | 26 | | | | 175 | 67 | | |
| mode 2 | 72 | | 45 | | | | 7 | 153 | | |
| gap 2 | 77 | 138 | 29 | | | | 189 | 135 | | |
| mode 3 | 35 | 1 | 19 | | | | 1 | 91 | | |
| gap 3 | 69 | 139 | 31 | | | | 200 | 170 | | |
| mode 4 | 27 | | 12 | | | | | 66 | | |
| gap 4 | 32 | 70 | 27 | | | | 127 | 167 | | |
| mode 5 | 4 | | 11 | | | 3 | 1 | 74 | | |
| right tail | | | | | | 2 | | | | |
| extended | | | | 2 | | 625 | | | 2 | |

Although sample K12 contains five draws, the fifth appears as a shoulder to the fourth and there are only three gaps and four modes. The point features are not consistent in their location. The low-pass spacing has peaks at the three gaps, plus some matching the fourth mode, but testing rejects most, leaving a significant peak at the first gap in half the trials and much weaker associations with

the other gaps. The interval spacing has the most peaks aligned to the fourth mode, strong counts at the second and third gaps, and the rest scattered. Testing rejects 85% of the detected peaks, with the significant interval peaks distributed rather uniformly, somewhat more strongly aligned to the fourth mode or first gap. The changepoints also match most strongly the first gap and first and fourth modes, but some are distributed over all extrema. The activity in the fourth mode might correspond to the buried hidden transition to the fifth draw. There are short level sections within the first and fourth modes, above those lying in the tails, but the majority of the sections extend over multiple gaps and modes. Short sections favor the trailing tail.

| K12 (penta) | | significant | | | | | detected | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | CPT | peak | | flat | | level | peak | | flat | |
| | | LP | Diw | LP | Diw | | LP | Diw | LP | Diw |
| # features | 436 | 146 | 161 | 33 | 0 | 837 | 564 | 1006 | 47 | 0 |
| left tail | | | | | | 57 | | | | |
| mode 1 | 93 | | 5 | | | 54 | | 14 | | |
| gap 1 | 113 | 99 | 38 | | | | 196 | 111 | | |
| mode 2 | 44 | 1 | 23 | | | | 1 | 111 | | |
| gap 2 | 37 | 19 | 21 | | | | 156 | 174 | | |
| mode 3 | 31 | | 12 | | | | 8 | 94 | | |
| gap 3 | 32 | 20 | 18 | | | | 151 | 192 | | |
| mode 4 | 86 | 7 | 44 | 8 | | 30 | 52 | 310 | 11 | |
| right tail | | | | | | 91 | | | | |
| extended | | | | 25 | | 605 | | | 36 | |

Sample K13 adds a fifth variate to the four in K6 after tweaking their size and width, and extends the uniform background to cover it. Counts will shift because of these tweaks, but the general behavior is the same. The low-pass peaks match the four gaps, with testing rejecting half of those aligned to the second and fourth. The low-pass peaks avoid the modes. The interval peaks match the gaps, particularly the second and third, but one also marks the modes, the second and fourth most strongly. Testing rejects two-thirds of the detected peaks, but leaves the same pattern of alignment and poor accuracy. There is one changepoint matched to the first mode, another two that split between the four gaps, and one divided between the remaining modes. The level sections extend over multiple extrema, beyond the one that corresponds to the left tail.

| K13 (penta) | | significant | | | | | detected | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | CPT | peak | | flat | | level | peak | | flat | |
| | | LP | Diw | LP | Diw | | LP | Diw | LP | Diw |
| # features | 885 | 597 | 321 | 4 | 0 | 1036 | 747 | 902 | 4 | 0 |
| left tail | | | | | | 200 | | | | |
| mode 1 | 214 | | 38 | | | 4 | | 51 | | |
| gap 1 | 124 | 188 | 49 | | | | 191 | 145 | | |
| mode 2 | 34 | | 11 | | | | | 70 | | |
| gap 2 | 96 | 125 | 47 | | | | 162 | 171 | | |
| mode 3 | 60 | | 3 | 3 | | | | 31 | 3 | |
| gap 3 | 70 | 85 | 45 | | | | 194 | 181 | | |
| mode 4 | 37 | 5 | 31 | | | | 5 | 82 | | |
| gap 4 | 138 | 194 | 76 | | | | 195 | 146 | | |
| mode 5 | 112 | | 21 | | | 7 | | 25 | | |
| right tail | | | | | | 41 | | | | |
| extended | | | | 1 | | 784 | | | 1 | |

Like K9, K14 has discrete data, which disrupts the level section algorithm and peaks in the interval spacing. The significant interval peaks match the first and third mode most strongly, but also align to all but the first gap. The low-pass spacing does smooth the steps between data points enough to match the first three gaps, and weakly the fourth. Testing accepts a peak matching the first gap as significant, and weakly one at the second gap, but any at the third and fourth gaps are insignificant. The total number of changepoints is not unreasonable; the algorithms do not fire at each step. There is a changepoint at the first gap, and in half the trials one at the first mode. In the other half of the trials a changepoint matches another extrema, avoiding the third mode and gap.

| K14 (penta) | | significant | | | | | detected | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | CPT | peak | | flat | | level | peak | | flat | |
| | | LP | Diw | LP | Diw | | LP | Diw | LP | Diw |
| # features | 444 | 278 | 402 | 27 | 0 | 1520 | 564 | 2048 | 27 | 0 |
| left tail | | | | | | 89 | | | | |
| mode 1 | 103 | | 90 | | | 853 | | 463 | | |
| gap 1 | 173 | 196 | 9 | | | | 196 | 49 | | |
| mode 2 | 25 | 4 | 54 | | | 13 | 6 | 229 | | |
| gap 2 | 44 | 57 | 34 | | | 2 | 170 | 214 | | |
| mode 3 | 12 | | 91 | 2 | | 51 | 1 | 337 | 2 | |
| gap 3 | 8 | 9 | 27 | | | 10 | 138 | 218 | | |
| mode 4 | 21 | | 42 | | | 9 | 10 | 204 | | |
| gap 4 | 36 | 11 | 26 | | | 1 | 42 | 204 | | |
| mode 5 | 22 | 1 | 29 | | | 5 | 1 | 130 | | |
| right tail | | | | | | 2 | | | | |
| extended | | | | 25 | | 485 | | | 25 | |

The results show more variability than expected from the symmetric K15 setup. Peaks in the low-pass spacing, all significant, match the four gaps or central, third mode, but prefer the second gap to the third. Interval peaks, mostly significant, match the third mode or, at a lower rate, the second or third gap; these are balanced. An additional peak falls within the outer pairs of modes and gaps. Changepoints center around the second and fourth modes and their gaps. The low-pass spacing may contain a flat, usually significant, that is extended. Level sections are also long, except

those covering the left tail. The sections ignore the right tail.

| K15 (penta) | | significant | | | | | detected | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | CPT | peak | | flat | | level | peak | | flat | |
| | | LP | Diw | LP | Diw | | LP | Diw | LP | Diw |
| # features | 836 | 479 | 507 | 69 | 0 | 1239 | 479 | 865 | 77 | 0 |
| left tail | | | | | | 139 | | | | |
| mode 1 | 21 | 28 | 34 | 1 | | 32 | 28 | 107 | 1 | |
| gap 1 | 107 | 80 | 41 | | | | 80 | 112 | | |
| mode 2 | 201 | | | | | 5 | | | | |
| gap 2 | 94 | 110 | 90 | | | | 110 | 105 | | |
| mode 3 | 15 | 95 | 158 | 1 | | | 95 | 198 | 2 | |
| gap 3 | 139 | 71 | 97 | | | | 71 | 114 | | |
| mode 4 | 162 | | | | | 5 | | | | |
| gap 4 | 80 | 89 | 33 | | | 1 | 89 | 105 | | |
| mode 5 | 17 | 6 | 54 | | | 3 | 6 | 124 | | |
| right tail | | | | | | 11 | | | | |
| extended | | | | 67 | | 1043 | | | 74 | |

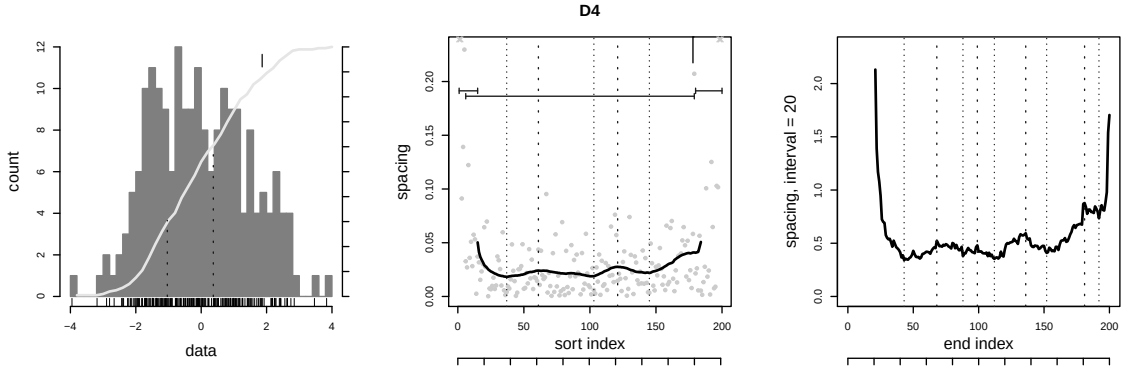
**Detail 37 Feature Counts and Consistency in Literature Datasets**

We check the performance, stability, and consistency of the modality tests for the samples defined in [Detail 35]. For the first we gather the number of features found and count those significant at the 0.01 level according to any test, including models and excursions for the low-pass spacing and excursions and runs for the interval spacing. For the second we generate 200 different draws and collect the mean and 90% confidence interval for the number of features, significant or not. For the third we compare the features found from both filters using difference matching criteria depending on the type of feature. Peaks must lie within 10 points of each other. Flats overlap if the common segment is 70% of the length of each or if one is completely covered by another, even if much shorter. Changepoints within the raw spacing are well-placed if they mark the transition between mode and gap and fall between a detected peak and the edge of a flat. This is not a pure check, since it also depends on the peaks and flats found. Endpoints of level sections must match changepoints within 5 points. Level sections will overlap flats if they either completely cover the flat or if the common segment is 70% of the length of either. These requirements — 5 or 10 points, 70% overlap — are arbitrary but substantial, representing a real alignment of two features.

We summarize the measurements in a table for each sample and plot a histogram and the low-pass and interval spacing for one draw as an example. These use the annotations from the main text: dashed lines for peaks, dotted for minima, horizontal lines for flats, tic marks along the top for changepoints, and bars near the top for level sections. The changepoints and level sections, which come from the raw data without filtering, are only marked on the low-pass spacing chart. Low-pass peaks, changepoints, and flats are also marked on the histograms, the latter as bars at the top. Of course the plots may not represent the average behavior over all 200 runs, but they give an idea of the draw distribution and features found, despite the variation in the spacing. The draws are identical to those analyzed in [Detail 35] and [Detail 36].

**D Samples**

D4 (Figure 94) draws from two normals separated by 2.2, which as we have seen is too small



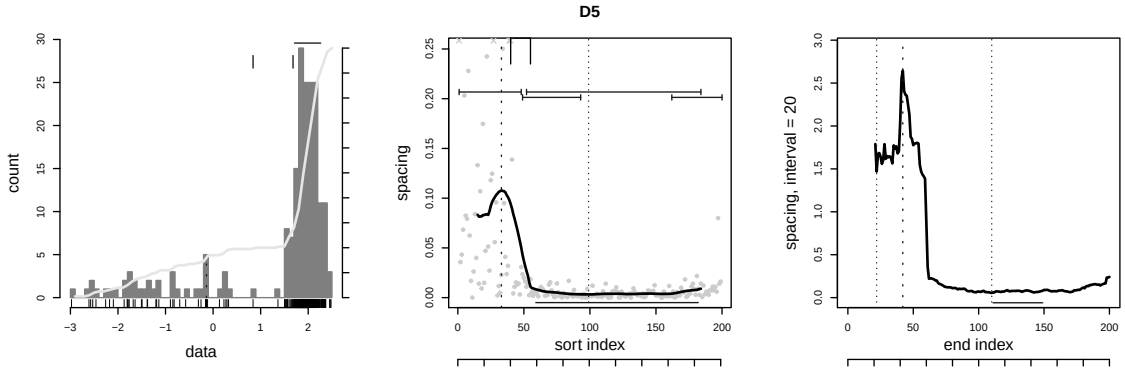

| D4 (bi) | peaks | | flats | | CPT | level | |
|---|---|---|---|---|---|---|---|
| | LP | Diw | LP | Diw | | | |
| mean count | 0.1 | 0.7 | 0.2 | 0 | 1.0 | 3.3 | |
| 90% CI | 0–1 | 0–3 | 0–1 | | 0–2 | 2–4 | |
| detected | 1.6 | 4.8 | 0.3 | 0 | | | |
| matching | 0 | | 0 | | 0 | 0.4 | 0.2 |
| 90% CI | | | | | | 0–1 | 0–1 |
| detected | 1.4 | | 0 | | placed | CPT | flat |

Figure 94: Bi-modal dataset D4.

to resolve the modes. The interval spacing does better at identifying a significant peak, although the detector finds many more potential maxima; the 90% confidence interval extends two significant peaks above the average. One or two of the detected peaks match between the two approaches, but the rate for significant peak is zero because there are so few in the low-pass spacing. Few flats are detected in the low-pass spacing, and none in the interval spacing. The tests accept most of the proposed low-pass features. The changepoint in the figure lies at the side as the spacing increases in the tail, and if the one that appears on average over the trials is located elsewhere, it is likely from noise. The lack of peaks in the low-pass spacing prevent checking placement. One level section spans both modes in the figure, and two mark the tails. One endpoint aligns with the changepoint in half the trials, and the lack of low-pass flats means no section can match them.

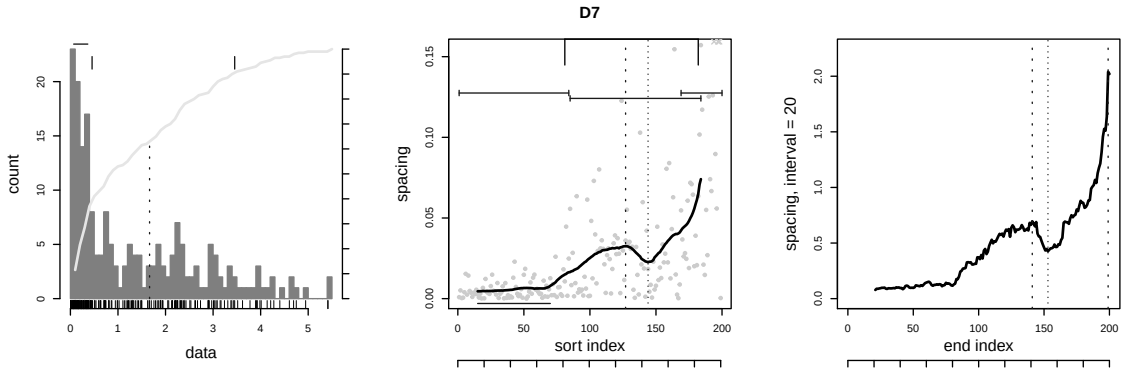
D5 (Figure 95) draws from two normals that are well separated but have different widths, with most points coming from the tighter variate. There is a clear difference between the narrow peak and the broad background to one side. A significant peak appears in the gap between the two in both the low-pass and interval spacing, and its position matches. A second maximum appears in half the trials in the interval spacing, but the confidence interval is tight about the average counts. A significant flat appears in the larger draw, and again for both methods is stable and matches, with a tight confidence interval and no other detected flats. Two changepoints appear at the transition between the draws, and these fall between the peak and flat. There are many level sections, with one endpoint lining up with a changepoint. One overlaps the significant flat.

D7 (Figure 96) samples from two unequal gamma variates, creating a tight and broad lobe, like D5. One peak appears on average in the low-pass spacing and two in the interval spacing, although in the example in the figure the second falls near the end of the data in the last bin of points. Neither is judged significant, however, and so the only matches are for a detected peak. One flat is found in



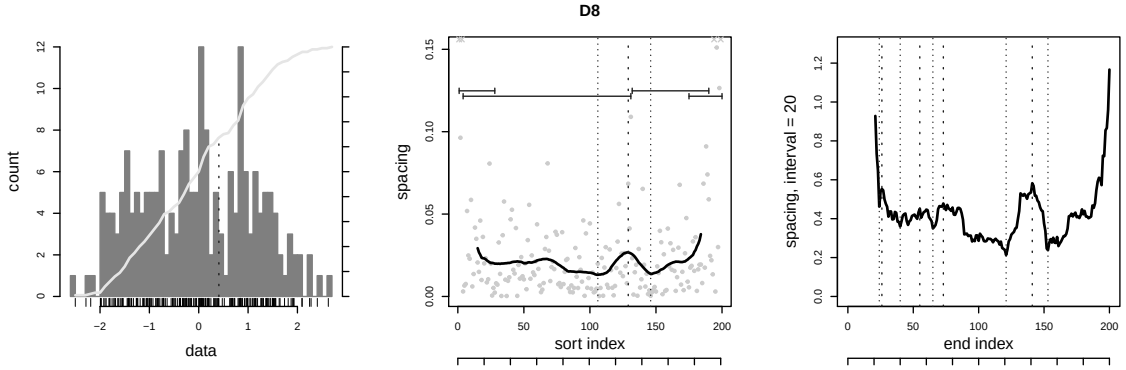

| D5 (abi) | peaks | | flats | | CPT | level | |
|---|---|---|---|---|---|---|---|
| | LP | Diw | LP | Diw | | | |
| mean count | 1.0 | 1.1 | 1.0 | 0.8 | 1.9 | 4.5 | |
| 90% CI | 1–1 | 1–2 | 1–1 | 0–2 | 1–3 | 3–6 | |
| detected | 1.0 | 1.4 | 1.0 | 0.8 | | | |
| matching | 0.9 | | 0.8 | | 1.4 | 0.7 | 0.9 |
| 90% CI | 0–1 | | 0–2 | | 1–2 | 0–2 | 0–1 |
| detected | 1.0 | | 0.8 | | placed | CPT | flat |

Figure 95: Asymmetric bi-modal dataset D5.




| D7 (abi) | peaks | | flats | | CPT | level | |
|---|---|---|---|---|---|---|---|
| | LP | Diw | LP | Diw | | | |
| mean count | 0 | 0.2 | 1.1 | 0.2 | 1.6 | 3.6 | |
| 90% CI | | 0–1 | 0–2 | 0–1 | 0–3 | 3–5 | |
| detected | 1.1 | 2.1 | 1.3 | 0.2 | | | |
| matching | 0 | | 0.2 | | 0.8 | 0.6 | 0.9 |
| 90% CI | | | 0–1 | | 0–2 | 0–2 | 0–2 |
| detected | 0.9 | | 0.2 | | placed | CPT | flat |

Figure 96: Asymmetric bi-modal dataset D7 from gamma draws.




| D8 (tri) | peaks | | flats | | CPT | level | |
|---|---|---|---|---|---|---|---|
| | LP | Diw | LP | Diw | | | |
| mean count | 1.1 | 1.0 | 0 | 0 | 0.9 | 3.1 | |
| 90% CI | 0–2 | 0–5 | | | 0–3 | 2–5 | |
| detected | 2.0 | 5.3 | 0 | 0 | | | |
| matching | 0.2 | | 0 | | 0 | 0.2 | 0 |
| 90% CI | 0–1 | | | | | 0–1 | |
| detected | 1.7 | | 0 | | placed | CPT | flat |

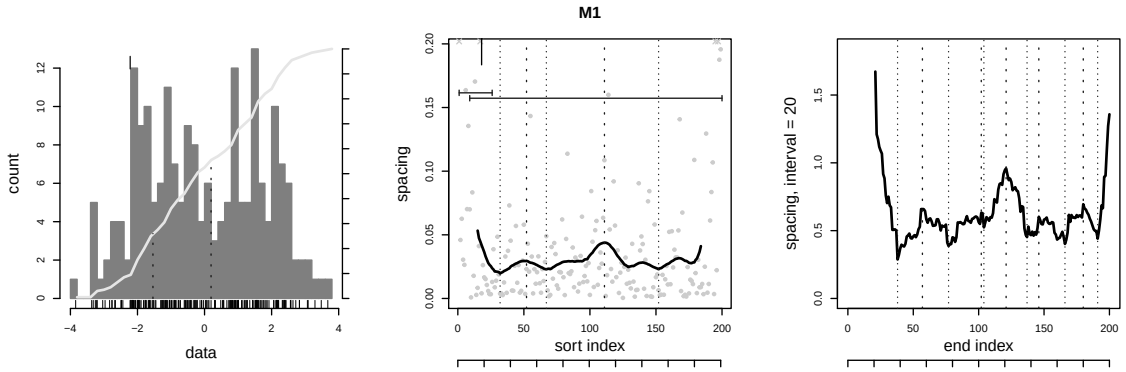
Figure 97: Tri-modal dataset D8.

the low-pass spacing, and occasionally one in the interval spacing. The detected flats are accepted as significant, and when one is found in the interval spacing it agrees with its low-pass cousin. Any flat in the interval spacing overlaps or is covered by a low-pass feature. One or two changepoints occur, at the boundary of the flat and in the increase in the spacing at the tail end. The boundary point is placed correctly. Level sections span the two draws and the tails. In half the trials one of the endpoints agrees with a changepoint, but the others do not. A section matches or covers most low-pass flats.

D8 (Figure 97) draws from three normals, two large samples symmetrically placed about a tight middle. The separation is too small to reliably distinguish the modes. The interval spacing is much rougher than the low-pass filtered spacing, which leads to a much higher peak detection rate that tests reduce to one significant peak. In the figure this does fall between two modes. Although the low-pass detected peaks match one from the interval spacing, only a few significant peaks align. The interval position is noisy, seen in the large confidence interval. The roughness also prevents the detection of any flats, and this is also true in the low-pass spacing. On average there is one changepoint, but we cannot judge its placement because of the lack of flats. In the figure the peak does separate two of the level sections, which also trigger in both tails. The changepoint only rarely aligns with the ends of the level sections. Overall, the peak tests seem to be able to resolve only one of the mode changes, but not both.

### M Samples

The M1 example (Figure 98) is our bi-modal test case with a separation of 3.0 but a smaller total draw of 200 points. We have seen this offset should be enough to distinguish the two variates. One or two peaks are detected in the low-pass spacing, and in half the trials one of these passes the



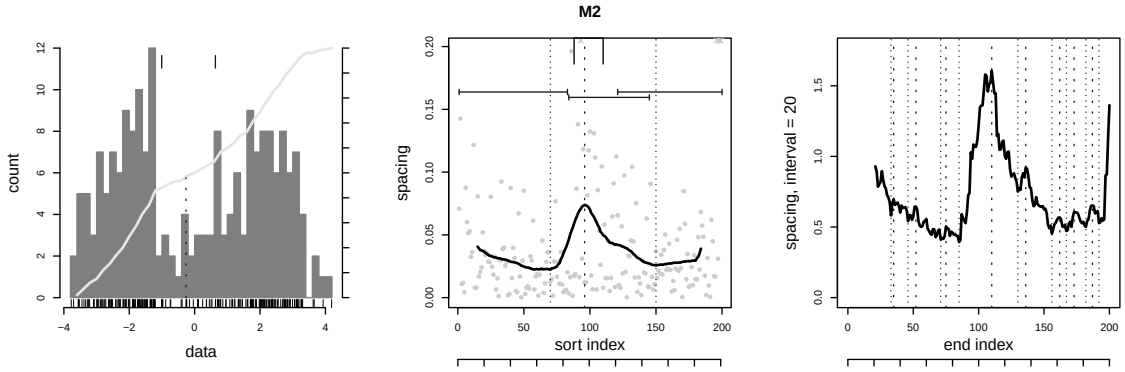

| M1 (bi) | peaks | | flats | | CPT | level | |
|---|---|---|---|---|---|---|---|
| | LP | Diw | LP | Diw | | | |
| mean count | 0.6 | 0.8 | 0.1 | 0 | 0.9 | 3.0 | |
| 90% CI | 0–1 | 0–4 | 0–1 | | 0–3 | 2–5 | |
| detected | 1.5 | 5.2 | 0.1 | 0 | | | |
| matching | 0.1 | | 0 | | 0 | 0.2 | 0.1 |
| 90% CI | 0–1 | | | | | 0–1 | 0–1 |
| detected | 1.3 | | 0 | | placed | CPT | flat |

Figure 98: Bi-modal dataset M1 with narrow separation.

acceptance level. The interval spacing is much rougher, producing many more peaks that the tests again screen to one significant; the confidence interval is large, however. In the figure the largest peak does align to the anti-mode, but the smaller ones are noise in the draw. Although the detected low-pass peaks align with those from the interval spacing, the significant peaks mostly do not. Flats only occasionally appear in the low-pass spacing, and those that do are judged significant. There is on average one changepoint, although the confidence interval is wide, and in the figure is found at the initial trailing edge. The level sections may do a better job of distinguishing the modes, as their endpoints come close to the largest peak and span both draws; extra sections would come at the starting and ending tails. They occasionally match an changepoint but cover the few significant flats.

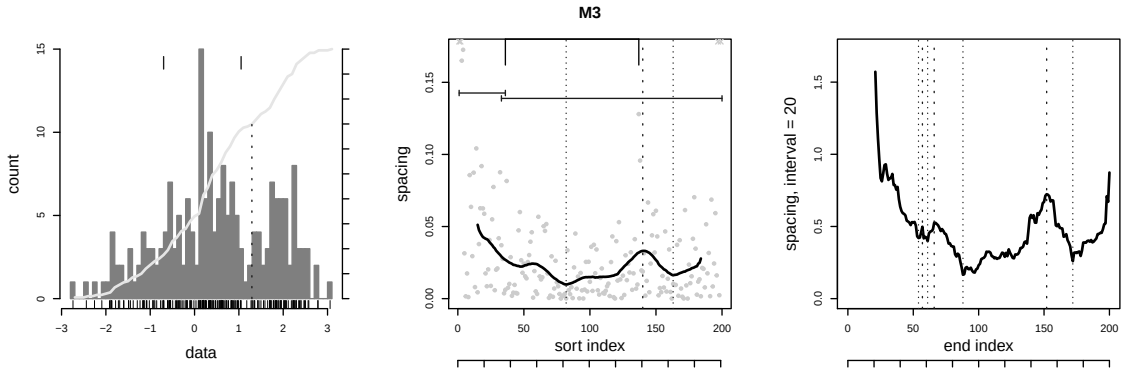
M2 (Figure 99) is also our bi-modal test case, with a separation of 4.0. The interval spacing is rough and generates five potential peaks per trial, while the smoother low-pass spacing reduces this to one or two. Testing finds one significant peak in either method, although the interval spacing's confidence interval is large. These align in half the trials, with better matching between the larger number of detected features. The interval spacing is too rough to generate flats, but also in the low-pass signal finding one is rare; when found, the tests accept the detected flat. The algorithms agree on one changepoint, with a moderate spread in the confidence interval. The absence of a flat limits the placement check, as it does the overlap by a level section. There are three sections on average, which in the figure span the two draws and the inter-modal region. An endpoint usually matches one changepoint.

The M3 sample (Figure 100) is an asymmetric bi-modal setup, with a large, broad draw separated by a gap of 2.0 from a smaller, narrower one. The performance falls between M1 and M2, with the difficult separation of the first eased by the imbalance between the variates. The number of detected peaks is the same for either method as in the symmetric case, but the tests reject more,



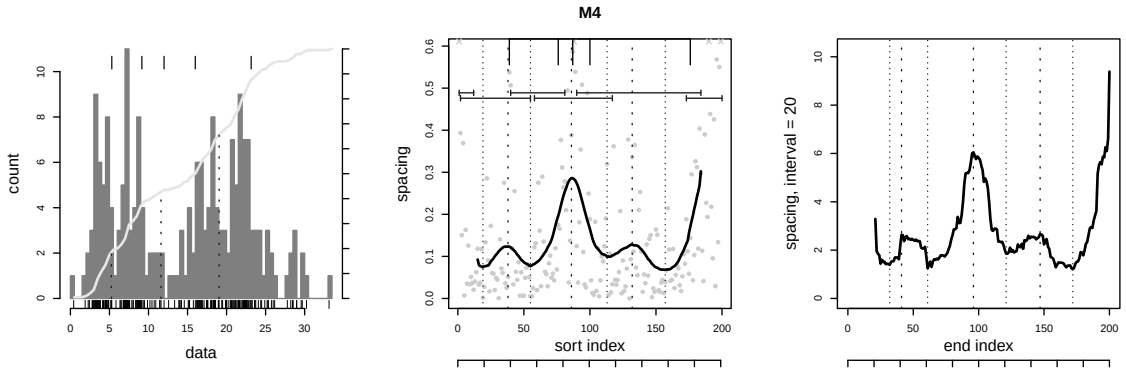

| M2 (bi) | peaks | | flats | | CPT | level | |
|---|---|---|---|---|---|---|---|
| | LP | Diw | LP | Diw | | | |
| mean count | 1.0 | 1.0 | 0.2 | 0 | 2.1 | 4.4 | |
| 90% CI | 1–1 | 0–4 | 0–1 | | 1–4 | 3–5 | |
| detected | 1.4 | 4.4 | 0.2 | 0 | | | |
| matching | 0.4 | | 0 | | 0.2 | 0.7 | 0.2 |
| 90% CI | 0–1 | | | | 0–1 | 0–2 | 0–1 |
| detected | 1.2 | | 0 | | placed | CPT | flat |

Figure 99: Bi-modal dataset M2.




| M3 (abi) | peaks | | flats | | CPT | level | |
|---|---|---|---|---|---|---|---|
| | LP | Diw | LP | Diw | | | |
| mean count | 0.5 | 0.6 | 0.2 | 0 | 1.2 | 2.6 | |
| 90% CI | 0–1 | 0–2 | 0–1 | | 0–3 | 2–4 | |
| detected | 1.6 | 4.4 | 0.3 | 0 | | | |
| matching | 0.1 | | 0 | | 0.1 | 0.3 | 0.2 |
| 90% CI | 0–1 | | | | 0–1 | 0–1 | 0–1 |
| detected | 1.5 | | 0 | | placed | CPT | flat |

Figure 100: Asymmetric bi-modal dataset M3.




| M4 (tri) | peaks | | flats | | CPT | level | |
|---|---|---|---|---|---|---|---|
| | LP | Diw | LP | Diw | | | |
| mean count | 1.0 | 0.9 | 0.1 | 0 | 2.2 | 4.5 | |
| 90% CI | 1–1 | 0–3 | 0–1 | | 1–4 | 2–6 | |
| detected | 2.3 | 4.4 | 0.1 | 0 | | | |
| matching | 0.4 | | 0 | | 0.1 | 1.0 | 0.1 |
| 90% CI | 0–1 | | | | 0–1 | 0–2 | 0–1 |
| detected | 1.8 | | 0 | | placed | CPT | flat |

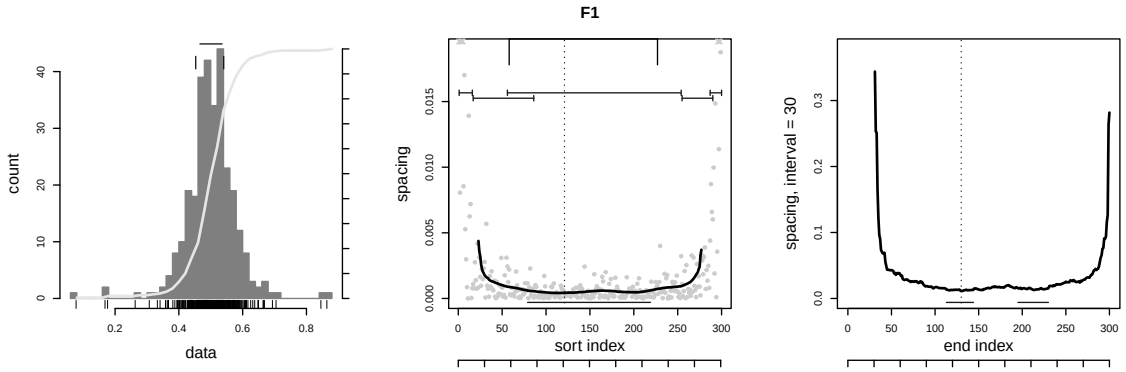
Figure 101: Tri-modal dataset M4.

passing one peak in half the trials. The peak does fall between the modes, but its position in the two methods varies too much for the alignment check. Flats are uncommon in the low-pass spacing but the acceptance rate is high. They do not appear in the interval spacing. In one third the trials the one changepoint matches an endpoint of the three level sections. The figure has two changepoints, one near the peak and one in the initial tail, that bound the first variate. The four level sections split the broad draw in two and cover the narrow one, with poorly defined boundaries and a section marking the initial tail. The number of flats limits the section coverage rate.

M4 (Figure 101) has two small normal draws separated by 4.0 and a larger, shifted background. Both spacing methods find similar numbers of detected peaks as the bi-modal examples, but they pass the one to the background because the first two are too small. The interval peaks have a moderate confidence interval. The figure does have a peak in the gap at index 90, and one at a spurious dip in the background towards index 150. Matching between the two methods is similar to the M2 value. The draws are too small for flats. There are two changepoints on average, with a moderate spread; in the figure they bound the background transition and mark the first anti-mode. The level section test is noisy. In the figure they overlap considerably, with one section spanning each mode plus the starting and ending tails. Only one endpoint matches a changepoint, despite the large number of endpoints. The flat count limits its coverage rate.

## F Samples

F1 (Figure 102) has two superimposed normal variates with no separation and different standard deviations; the total sample has been scaled to 300 points for the narrow draw. It appears as a uni-modal distribution. This is clear in the figure, where there is no increase in the spacing. The wider draw is not obvious in the spacing because it sets the two tails. Few trials generate a possible



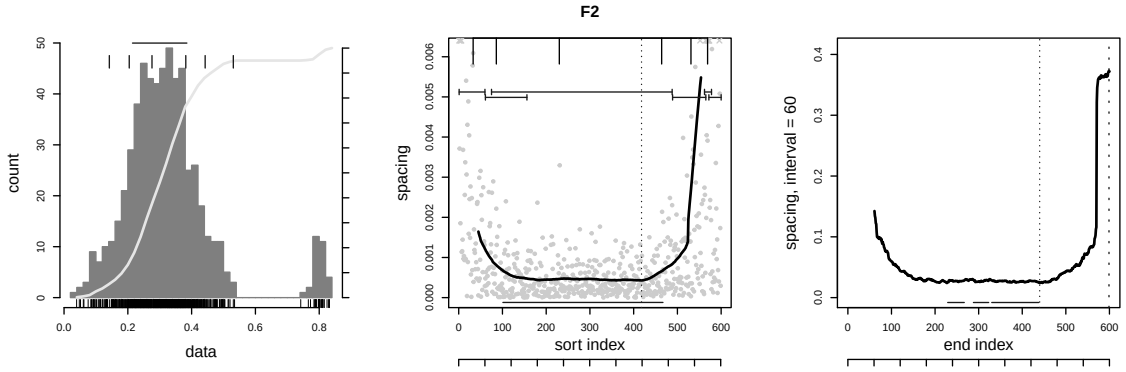

| F1 (uni) | peaks | | flats | | CPT | level | |
|---|---|---|---|---|---|---|---|
| | LP | Diw | LP | Diw | | | |
| mean count | 0 | 0 | 0.6 | 1.2 | 1.4 | 4.9 | |
| 90% CI | | | 0–1 | 0–3 | 0–3 | 4–6 | |
| detected | 0.2 | 0.1 | 2.1 | 1.3 | | | |
| matching | 0 | | 1.0 | | 0 | 0.7 | 0.6 |
| 90% CI | | | 0–2 | | | 0–2 | 0–1 |
| detected | 0 | | 1.1 | | placed | CPT | flat |

Figure 102: Uni-modal dataset F1.

peak in the low-pass spacing, and no peaks with either method are found to be significant. On the other hand, two flats do appear in the low-pass spacing, one of which is significant in half the trials, and one appears in the interval spacing that is accepted. The flats from the two methods do mostly match. That the matching rate is higher than the number of low-pass flats means there are multiple interval flats for each low-pass, as seen in the figure. The draw generates one or two changepoints that bound the narrow draw. Five level sections span the sample or tails. A section overlaps any flat. Half the changepoints align with one of the endpoints. In the figure the level sections and changepoints do pick out the edges of the tight draw, the shoulders from the broader draw, and the tails.

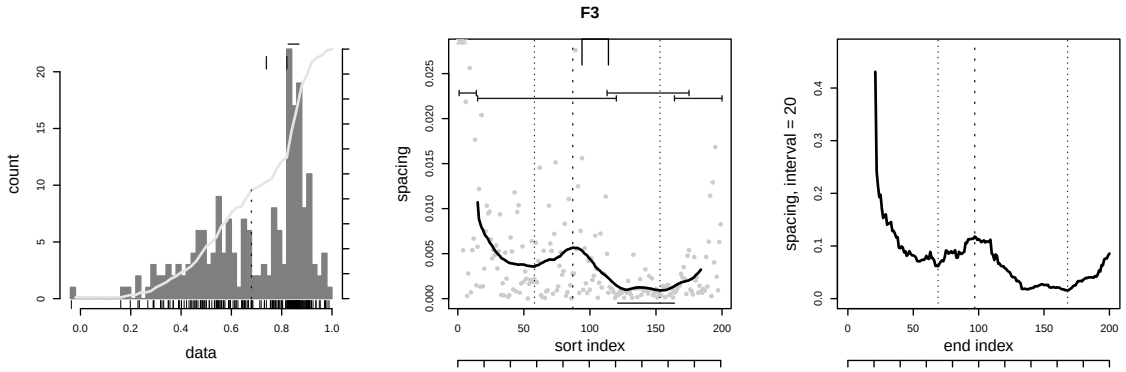
Sample F2 (Figure 103) has a very small normal draw well-separated from the main mode, whose size is 19 times larger; there are 600 points in the total draw. This buries the small draw far in the end tail and any peak from the separation is lost as an edge effect, either outside the filter or in the increased spacing in the tails. The interval spacing has no dead zone and occasionally a local maxima appears and is judged significant due to the large drop into the main draw. This mode instead creates a flat, often split into many pieces by noise, especially in the interval spacing; different parameters for the detector would merge them into a larger feature. Most flats detected are accepted as significant. The matching rate is a little less than the number of interval spacing flats but more than the low-pass count, which means the interval flats are subsets of the low-pass, as seen in the figure. We get four changepoints per trial, with a moderate confidence interval. In the figure these seem to be noise or related to the two tails. Each sample also reliably generates a large number of level sections that cover the main mode with one large interval and trigger repeatedly in both tails. Only one endpoint matches a changepoint despite the large number of both, and the flat and level section from the main mode match.

The F3 sample (Figure 104) is an asymmetric bi-modal. The detector notes one peak in the



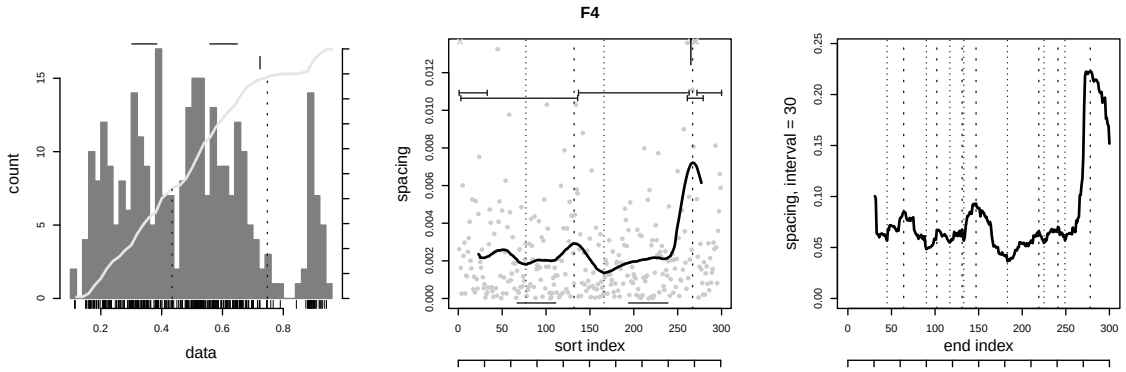

| F2 (abi) | peaks | | flats | | CPT | level | |
|---|---|---|---|---|---|---|---|
| | LP | Diw | LP | Diw | | | |
| mean count | 0 | 0.1 | 1.2 | 3.7 | 3.8 | 6.7 | |
| 90% CI | | 0–1 | 1–2 | 2–5 | 2–6 | 6–8 | |
| detected | 0 | 0.1 | 1.5 | 4.5 | | | |
| matching | 0 | | 3.5 | | 0 | 1.2 | 1.0 |
| 90% CI | | | 2–5 | | | 0–2 | 0–1 |
| detected | 0 | | 4.0 | | placed | CPT | flat |

Figure 103: Bi-modal dataset F2.




| F3 (abi) | peaks | | flats | | CPT | level | |
|---|---|---|---|---|---|---|---|
| | LP | Diw | LP | Diw | | | |
| mean count | 0.6 | 0.7 | 0.8 | 0 | 1.5 | 3.9 | |
| 90% CI | 0–1 | 0–2 | 0–1 | | 1–3 | 3–5 | |
| detected | 1.3 | 3.1 | 0.9 | 0 | | | |
| matching | 0.2 | | 0 | | 0.8 | 0.7 | 0.8 |
| 90% CI | 0–1 | | | | 0–2 | 0–1 | 0–1 |
| detected | 1.1 | | 0 | | placed | CPT | flat |

Figure 104: Asymmetric bi-modal dataset F3.



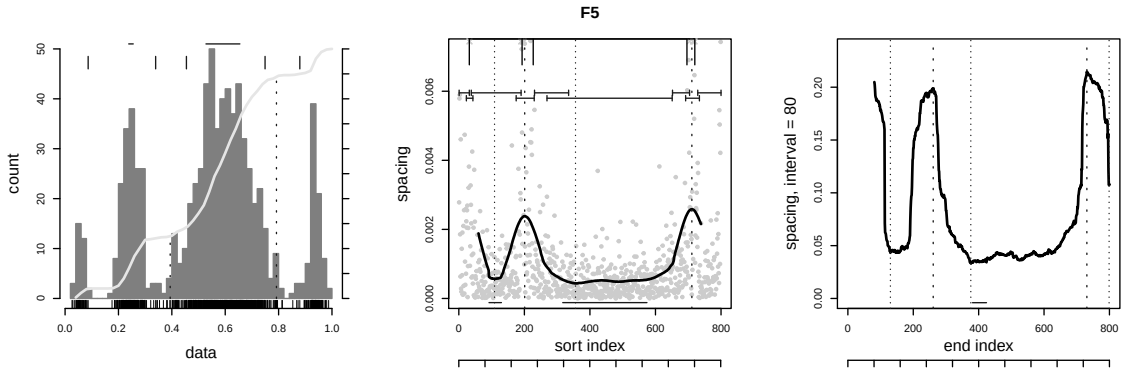

| F4 (tri) | peaks | | flats | | CPT | level | |
|---|---|---|---|---|---|---|---|
| | LP | Diw | LP | Diw | | | |
| mean count | 1.0 | 1.2 | 0.7 | 0 | 2.0 | 4.5 | |
| 90% CI | 0–1 | 0–2 | 0–2 | | 1–3 | 3–6 | |
| detected | 2.1 | 4.8 | 1.8 | 0 | | | |
| matching | 0.8 | | 0 | | 0.9 | 1.1 | 0.8 |
| 90% CI | 0–1 | | | | 0–2 | 0–2 | 0–2 |
| detected | 1.7 | | 0 | | placed | CPT | flat |

Figure 105: Tri-modal dataset F4.

low-pass spacing and three in the rougher interval spacing, but testing reduces this to one significant peak in the majority, but not all, of the trials. The two methods generally do not align, although they do when comparing detected peaks. The significant peaks are repeatable. The narrow variate generates one flat in the low-pass spacing that passes the significance test, but the interval spacing is too uneven. There are one or two changepoints; in the figure they occur when the spacing's variation reduces at the narrow draw, and at the increased spacing at the peak. Either position would be correctly placed between the peak and flat. The placement rate is limited by the number of flats. The four level sections include the two draws and the tail ends, with the changepoint at the transition matching an endpoint in half the trials and any flat covered by a section.

F4 (Figure 105) is a tri-modal distribution, with two equal draws separated by three times their width and a narrow variate that is off to a side. The total draw is 300 points. The detector finds two peaks in the low-pass spacing, with the separation to the third variate creating a significant feature, but with too small a gap between the equal draws to pass. Five peaks are detected in the rougher interval spacing, of which one is significant. The detected peaks match but only a third of the significant do, so the position of the large gap seems to be noisy in the interval spacing. The low-pass spacing contains two flats for the equal draws, one of which passes the significance threshold. There are two changepoints, one of which is placed correctly. The figure contains one that matches with the main peak. The four or five level sections span the modes and tail ends. One of the changepoints, in the figure at the main peak, matches an endpoint. The sections overlap any significant flat.

Four well-spaced normal variates of different widths form the F5 sample (Figure 106). The smallest draw lies furthest left and the total sample contains 800 points to compensate for its small size. This is not enough to avoid edge effects from either filter, and the two peaks that reliably appear in either the low-pass or interval spacing separate the three larger draws. Few false peaks are detected, and




| F5 (quad) | peaks | | flats | | CPT | level | |
|---|---|---|---|---|---|---|---|
| | LP | Diw | LP | Diw | | | |
| mean count | 2.0 | 2.0 | 1.4 | 0.1 | 5.4 | 9.6 | |
| 90% CI | 2–2 | 1–3 | 1–2 | 0–1 | 4–7 | 9–11 | |
| detected | 2.0 | 2.2 | 2.7 | 0.1 | | | |
| matching | 0.5 | | 0.1 | | 3.3 | 1.3 | 1.3 |
| 90% CI | 0–1 | | 0–1 | | 2–5 | 0–3 | 1–2 |
| detected | 0.6 | | 0.1 | | placed | CPT | flat |

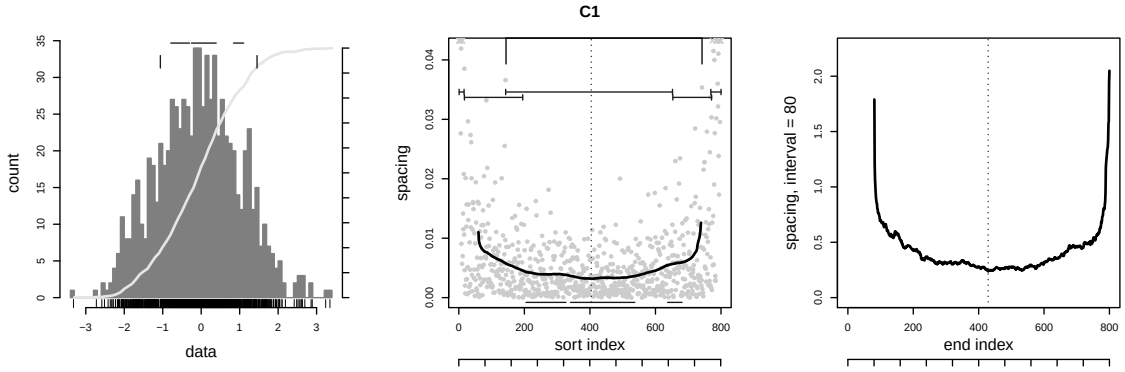
Figure 106: Quad-modal dataset F5.

the tests accept all the detected. The matching of detected or significant peaks is poor, however. Three flats appear in the low-pass spacing, and one or two are judged significant. The figure has two for the central variates. Only one appears in the interval spacing, passes, and matches. In the figure it seems that a larger ripple allowance is needed. The five changepoints do bound the transitions between modes, reflected in the high correct placement rate. The level section detector is noisy, reliably producing ten level sections that trigger multiple overlapping times in the transition regions, while also spanning the modes and two tails. Their endpoints do not align well with the changepoints, matching only a third, despite the high number of both. They do cover any flat.

## C Samples

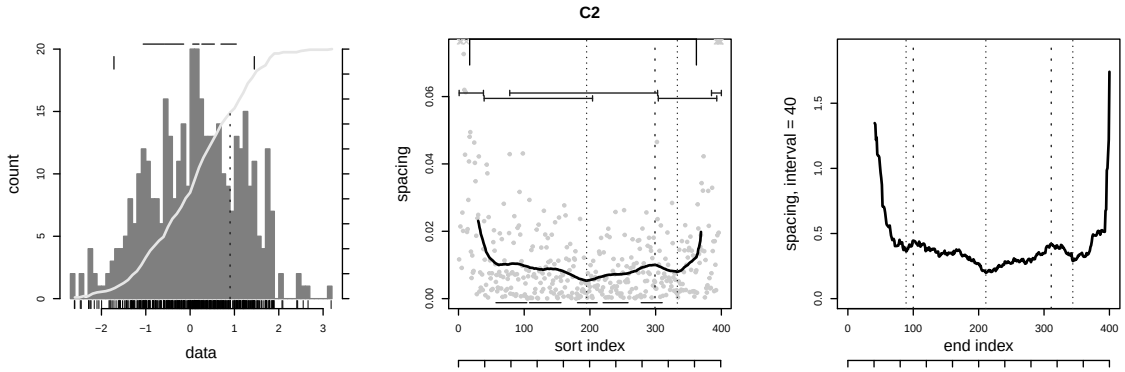
The C1 sample (Figure 107) has a total of 800 points. The small side draw produces points on the shoulder between $-2$ and $-1$. Peaks rarely appear in either the low-pass or interval spacing, and are never accepted as significant. Three or four flats are detected in the low-pass spacing, but testing rejects almost all of them. The interval spacing is smooth enough to generate one or two flats with a wide confidence interval, and they are accepted as significant. Low-pass flats cover these shorter features, with the matching rate reflecting the interval count. There is a wide variation in the number of changepoints, averaging three; the figure has one at each tail. There are also six level sections. Given the large counts of these last two, the match of one changepoint in half the trials to a section endpoint is low. The flat counts limit the coverage rate.

C2 (Figure 108) has a total draw size of 400 points. Although its density is unimodal, the small, narrow side draw forms a shoulder near 1.0, which appears in the histogram as a dip with a corresponding peak in the spacing. The peak is reliably detected in the low-pass spacing, but does not pass the acceptance tests. The interval spacing contains two, with an occasionally significant peak,



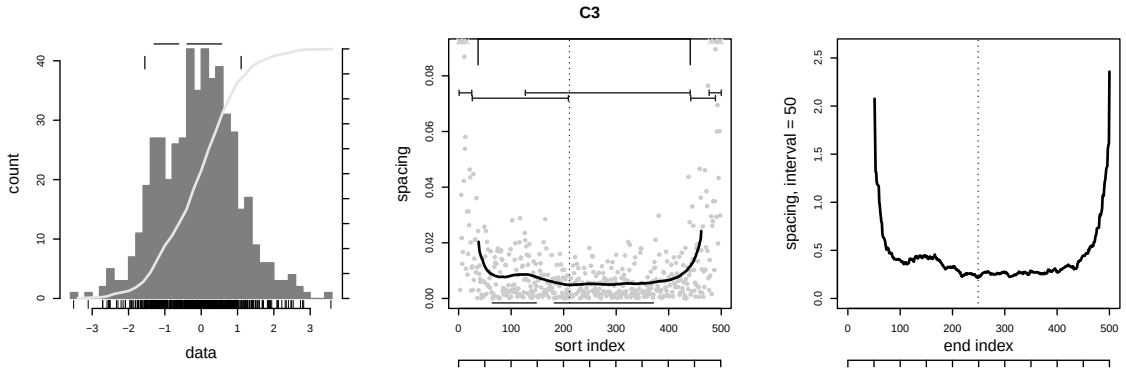

| C1 (uni) | peaks | | flats | | CPT | level | |
|---|---|---|---|---|---|---|---|
| | LP | Diw | LP | Diw | | | |
| mean count | 0 | 0 | 0.2 | 1.5 | 2.8 | 5.8 | |
| 90% CI | | | 0–1 | 0–4 | 1–5 | 5–7 | |
| detected | 0.2 | 0.2 | 3.4 | 1.5 | | | |
| matching | 0 | | 1.1 | | 0.1 | 0.5 | 0.2 |
| 90% CI | | | 0–3 | | | 0–2 | 0–1 |
| detected | 0 | | 1.1 | | placed | CPT | flat |

Figure 107: Uni-modal dataset C1.




| C2 (uni) | peaks | | flats | | CPT | level | |
|---|---|---|---|---|---|---|---|
| | LP | Diw | LP | Diw | | | |
| mean count | 0 | 0.2 | 0.4 | 0 | 1.6 | 4.5 | |
| 90% CI | | 0–1 | 0–1 | | 0–3 | 3.9–6 | |
| detected | 0.9 | 1.8 | 3.6 | 0 | | | |
| matching | 0 | | 0 | | 0.1 | 0.5 | 0.4 |
| 90% CI | | | | | 0–1 | 0–2 | 0–2 |
| detected | 0.4 | | 0 | | placed | CPT | flat |

Figure 108: Uni-modal dataset C2.




| C3 (uni) | peaks | | flats | | CPT | level | |
|---|---|---|---|---|---|---|---|
| | LP | Diw | LP | Diw | | | |
| mean count | 0 | 0.2 | 0.4 | 0.2 | 2.2 | 4.9 | |
| 90% CI | | 0–1 | 0–1 | 0–1 | 1–4 | 4–6 | |
| detected | 0.7 | 1.3 | 4.3 | 0.2 | | | |
| matching | 0 | | 0.1 | | 0.1 | 0.6 | 0.4 |
| 90% CI | | | 0–1 | | 0–1 | 0–2 | 0–1 |
| detected | 0.2 | | 0.1 | | placed | CPT | flat |

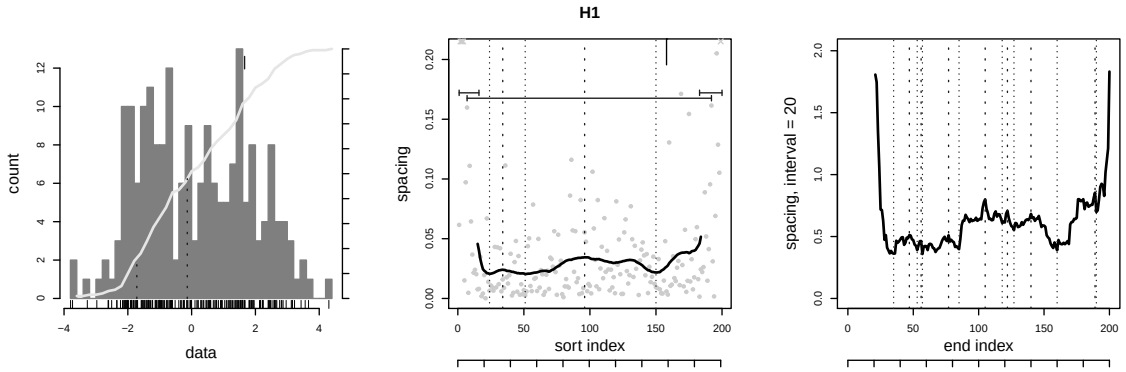
Figure 109: Uni-modal dataset C3.

one of which matches the low-pass feature in half the trials. The low-pass spacing is smooth enough to support several flats, located in the main draw but not spanning it, with one passing in half the trials. The level sections cover the modes better, splitting at the peak, but overlapping. A section covers any significant flat. There are one or two changepoints per run with a moderate spread in the count, whose placement check is limited by the flats. One aligns with a section endpoint in half the trials.

A typo during testing divided the tight standard deviation by ten, and this created a tight sequence with small spacing. The number of points was too small to pull the low-pass spacing into a flat, but depending on the draw the spacing in the larger variate was steady and supported a flat. The dip in spacing was enough to create a significant peak at the transition, and was also sharp enough to create changepoints at each side.

C3 (Figure 109) has a total draw size of 500 points. The shoulder near $-1.0$ might have a small dip in the histogram. One peak usually appears in the low-pass spacing, but the test does not judge it significant. The peak appears more often in the interval spacing, which occasionally accepts it. As with C2 the low-pass spacing contains four flats, one of which is significant in half the trials and covered by a level section. These flats divide up the main mode, and in the figure a second appears in the shoulder. Flats do not appear in the interval spacing despite a clean signal; this suggests relaxing the ripple parameter. There are five additional level sections that divide up the modes, taking the starting and ending tails into account. There are two changepoints at the tails, with one matching an endpoint in half the trials.

**H Samples**



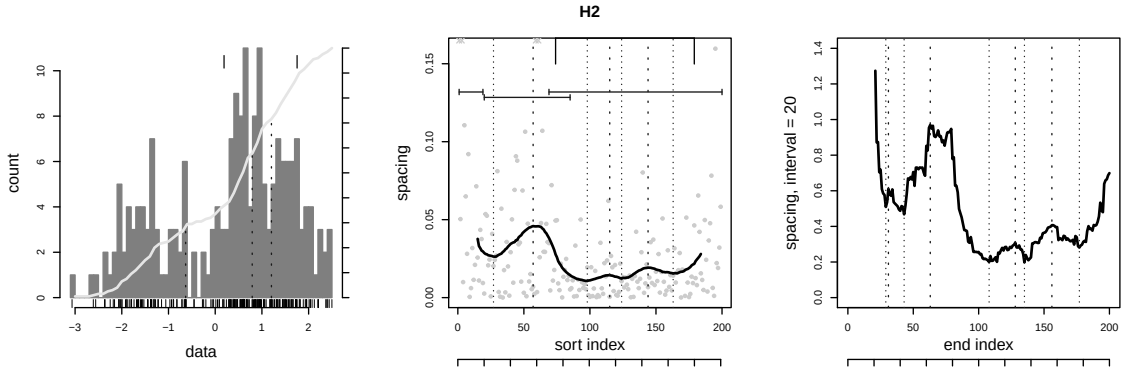

| H1 (bi) | peaks | | flats | | CPT | level | |
|---|---|---|---|---|---|---|---|
| | LP | Diw | LP | Diw | | | |
| mean count | 0.6 | 0.9 | 0.1 | 0 | 0.9 | 3.1 | |
| 90% CI | 0–1 | 0–6 | 0–1 | | 0–3 | 2–5 | |
| detected | 1.6 | 5.3 | 0.1 | 0 | | | |
| matching | 0.1 | | 0 | | 0 | 0.3 | 0.1 |
| 90% CI | 0–1 | | | | | 0–1 | 0–1 |
| detected | 1.4 | | 0 | | placed | CPT | flat |

Figure 110: Bi-modal dataset H1.

H1 (Figure 110) duplicates M1, a draw from two normals separated by 3.0. The discussion for M1 applies. The average measurements differ from the M1 values by 0.1 at most, giving an idea of the stability of the testing procedure. A comparison with Figure 98 hints at the variation possible in the histogram and spacing.

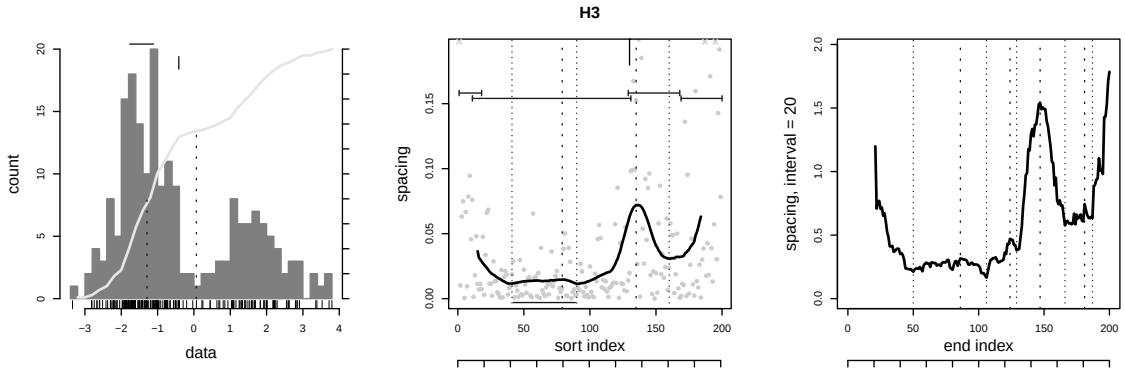
H2 (Figure 111) combines two normals with unbalanced draw sizes to make an asymmetric bi-modal distribution. There are still 200 points in total in the sample. The low-pass spacing contains one or two peaks, with the maximum at the transition between the draws significant and stable. The interval spacing is rougher, with four peaks detected per run, but testing limits this to one significant peak in 70% of the trials, albeit with a wide spread. The detected low-pass peaks align with the interval spacing maxima, but this alignment does not persist into the accepted peaks. The noise visible in the figure in the mesa around the peak may explain this lack. The interval spacing does not support any flats, while the low-pass spacing finds one in 40% of the trials, with any detected flat passing the tests. There are one or two changepoints found, one near the transition between modes; the judgement about its placement is limited by the presence of a flat. There are four level sections that span the two draws and the two tails, covering a flat if one exists. A changepoint matches an endpoint in half the runs.

The H3 sample (Figure 112) is an asymmetric bi-modal setup where the normals have different widths and the draw sizes are slightly imbalanced. The second draw forms a smaller bump than the H2 example, and this is reflected in the measurements. There are fewer detected peaks in either spacing and fewer test as significant, dropping the rate to one peak per three trials. The counts are stable. The detected peaks in the low-pass and interval spacing still match. One flat is now reliably found in the low-pass spacing in the tighter draw, and usually the tests judge it significant. There is a changepoint located at the inter-mode transition and placed between a peak and flat. Four



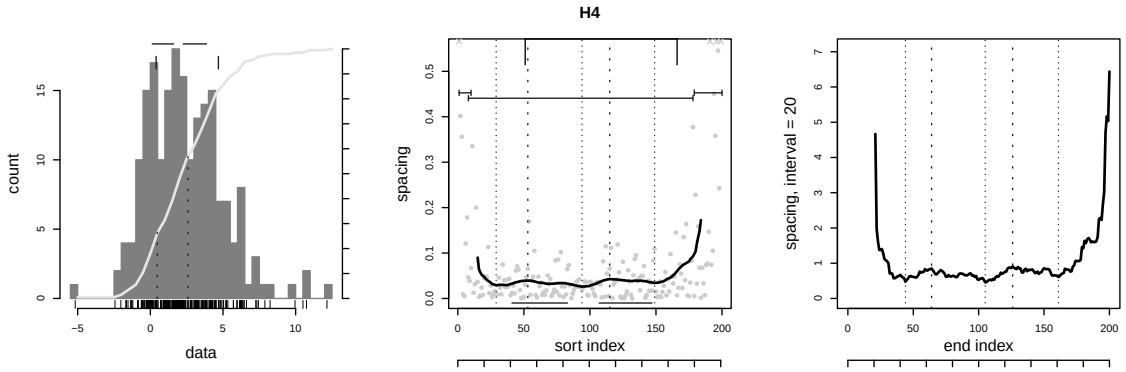

| H2 (abi) | peaks | | flats | | CPT | level | |
|---|---|---|---|---|---|---|---|
| | LP | Diw | LP | Diw | | | |
| mean count | 0.9 | 0.7 | 0.4 | 0 | 1.6 | 3.7 | |
| 90% CI | 0–1 | 0–3 | 0–1 | | 1–3 | 3–5 | |
| detected | 1.5 | 4.2 | 0.4 | 0 | | | |
| matching | 0.2 | | 0 | | 0.3 | 0.7 | 0.4 |
| 90% CI | 0–1 | | | | 0–1 | 0–2 | 0–1 |
| detected | 1.3 | | 0 | | placed | CPT | flat |

Figure 111: Asymmetric bi-modal dataset H2.




| H3 (abi) | peaks | | flats | | CPT | level | |
|---|---|---|---|---|---|---|---|
| | LP | Diw | LP | Diw | | | |
| mean count | 0.3 | 0.3 | 0.7 | 0 | 1.5 | 3.6 | |
| 90% CI | 0–1 | 0–1 | 0–1 | | 0–3 | 3–5 | |
| detected | 1.1 | 2.6 | 1.0 | 0 | | | |
| matching | 0.1 | | 0 | | 0.8 | 0.7 | 0.7 |
| 90% CI | 0–1 | | | | 0–2 | 0–2 | 0–1 |
| detected | 0.9 | | 0 | | placed | CPT | flat |

Figure 112: Asymmetric bi-modal dataset H3.




| H4 (abi) | peaks | | flats | | CPT | level | |
|---|---|---|---|---|---|---|---|
| | LP | Diw | LP | Diw | | | |
| mean count | 0 | 0.1 | 0.2 | 0.1 | 1.0 | 3.8 | |
| 90% CI | | 0–1 | 0–1 | 0–1 | 0–2 | 3–5 | |
| detected | 1.1 | 1.2 | 1.4 | 0.1 | | | |
| matching | 0 | | 0.1 | | 0 | 0.6 | 0.3 |
| 90% CI | | | 0–1 | | | 0–2 | 0–1 |
| detected | 0.7 | | 0.1 | | placed | CPT | flat |

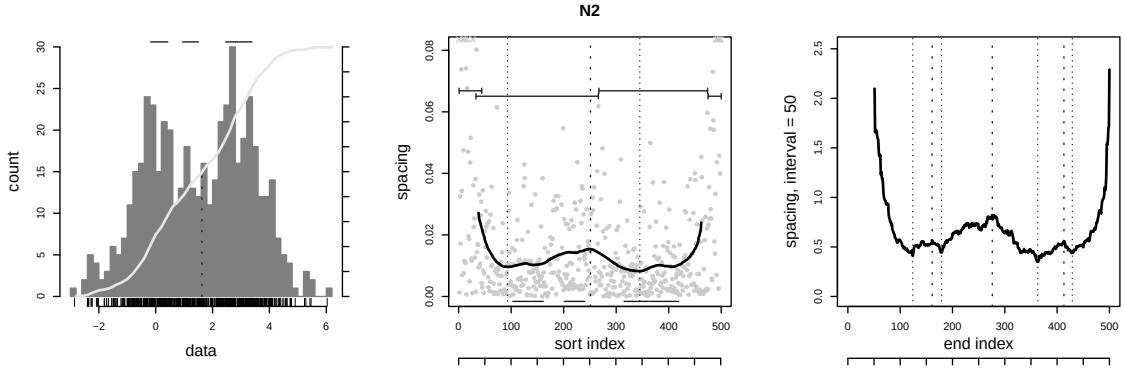
Figure 113: Asymmetric bi-modal dataset H4 with $t$ variates.

level sections cover the two draws and tails, with one covering the significant flat, and a changepoint usually aligns with an endpoint.

The H4 sample (Figure 113) combines two $t$ variates in an asymmetric bi-modal distribution. The local minimum in the distribution's density lies at 1.09. The two modes cannot be seen in the figure, which has two dips in the histogram, the left corresponding to 1.09. Two flats overlap the minor peaks and the main level section spans the entire sample. One changepoint lies near the left dip, the other occurs at the start of the trailing tail. Peaks do not resolve this setup. They are detected in both the low-pass and interval spacing and generally match, but they do not pass any test. One or two flats appear in the low-pass spacing, but rarely pass. In the figure it looks like the flat detection parameters may need relaxing for the interval spacing. There is one changepoint that usually matches up with one of the four flats' endpoints. The stability of these results is good, with confidence intervals a little wider than rounding about the average.

### N and P Samples

Sample N2 (Figure 114), like samples M1 and H1, is our bi-modal test case with the base draw size, 500 points in total, and separation of 3.0. The draw size is larger than M1 and H1. The makes the peak in the inter-mode transition more stable, with fewer detected maxima and more in the low-pass spacing accepted as significant. The acceptance rate is lower for the interval spacing, however, although the confidence interval is much tighter and there is no change in the matching of significant peaks. The detected peaks do not align as well. The larger draws increase the number of flats in the low-pass spacing, yet only one out of ten pass the tests. In the figure two flats share the first mode, so draw variation is important. There is an additional changepoint, which correspondingly shifts the confidence interval, and one of the changepoints lies between the peak and a flat in half the trials.

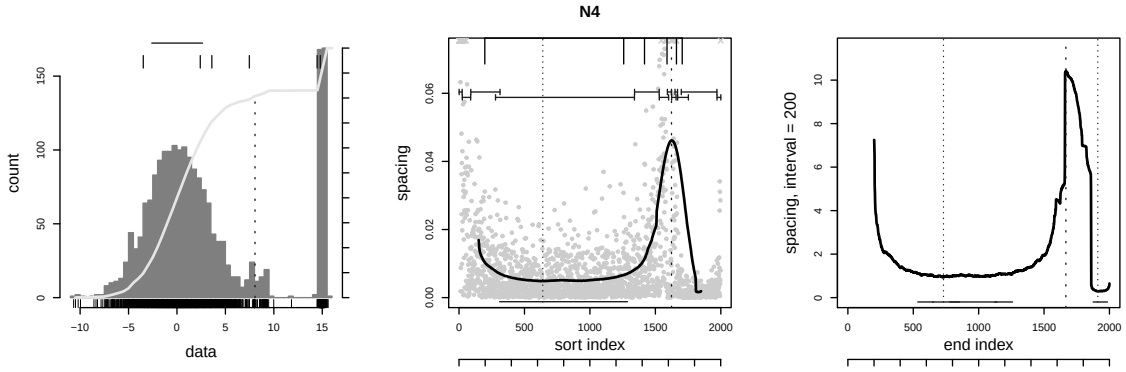



| N2 (bi) | peaks | | flats | | CPT | level | |
|---|---|---|---|---|---|---|---|
| | LP | Diw | LP | Diw | | | |
| mean count | 1.0 | 0.3 | 0.3 | 0 | 2.1 | 4.7 | |
| 90% CI | 1–1 | 0–1 | 0–1 | | 0.9–4 | 3–6 | |
| detected | 1.1 | 2.7 | 3.7 | 0 | | | |
| matching | 0.1 | | 0 | | 0.5 | 0.6 | 0.2 |
| 90% CI | 0–1 | | | | 0–2 | 0–2 | 0–1 |
| detected | 0.4 | | 0 | | placed | CPT | flat |

Figure 114: Bi-modal dataset N2.

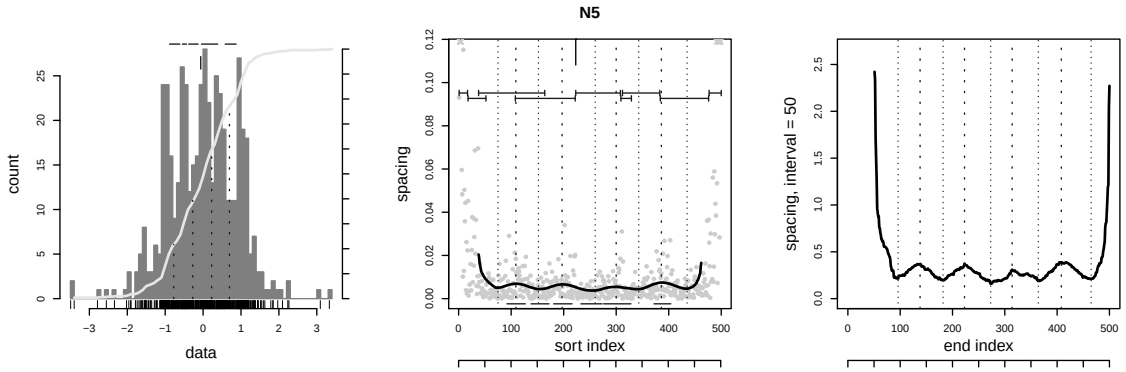
There are two additional level sections that mirror the behavior with the flats; the five would cover one mode with two sections, the other with a third, and the two tails with the fourth and fifth.

The N4 sample (Figure 115) is a quad-modal distribution. Beside a standard normal variate sit two small, narrow, closely spaced draws, and a well-separated fourth variate. The total draw is 2000 points, of which 30 are in each of the two smallest draws and 340 in the separated. The figure shows a peak set by the transition to the fourth variate. The two small draws between 7 and 10 cannot be separated from the larger spacing increase to the third anti-mode, smoothed by the filters. There is therefore one significant peak that is easily and stably found in both the low-pass and interval spacing. The two methods agree about the peak's location. In the figure this peak actually splits the two tight draws. The low-pass spacing contains one significant flat covering the main draw; as seen in the figure, the fourth is lost or cut short when we ignore points partially overlapped by the filter. The interval spacing does not have this dead zone and a significant flat appears in the fourth draw. Level sections also do not have a dead zone and can find this mode. Three or four significant interval flats cover the main draw; tweaking the ripple specification would merge them into one. The two methods do agree. The matching rate is higher than the number of significant or detected flats in the low-pass spacing, so each low-pass flat covers multiple interval flats. The changepoints find the transition from the main to the small draws, and from the third to the fourth. They may separate the tight pair. The point list is noisy, however, with a very wide confidence interval. Because the peak lies in the gap, two changepoints are correctly placed. The level section algorithm also responds to the two small draws by creating many regions around the transition. There are thirteen sections that are stable across trials, with many small sections around the gap, but also many at the start of the sample. There are separate small sections for each member of the pair, but these are hard to distinguish among all the others. The concentration of the endpoints in the gap helps match them to two changepoints. One long level section spans the main draw and corresponds to the flat.



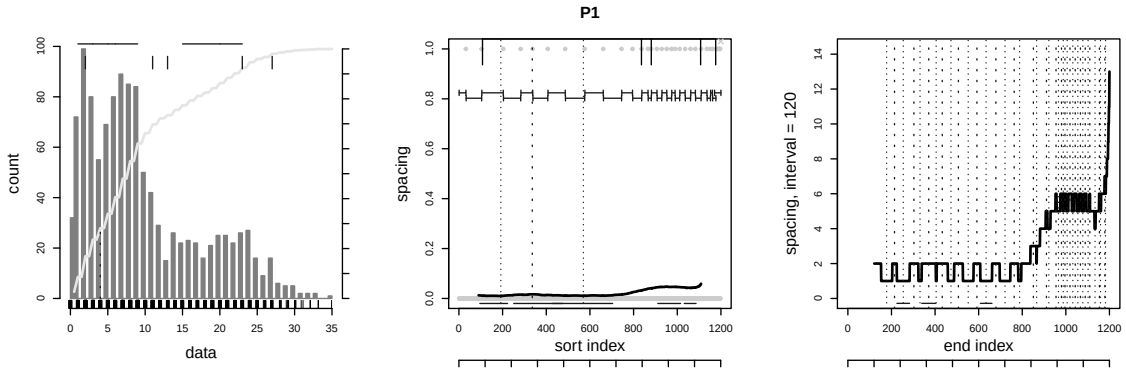

| N4 (quad) | peaks | | flats | | CPT | level | |
|---|---|---|---|---|---|---|---|
| | LP | Diw | LP | Diw | | | |
| mean count | 1.0 | 1.0 | 1.0 | 4.6 | 5.0 | 13.4 | |
| 90% CI | 1–1 | 1–1 | 1–1 | 3–6 | 2–9 | 12–15 | |
| detected | 1.0 | 1.0 | 1.1 | 4.6 | | | |
| matching | 0.9 | | 3.3 | | 1.8 | 2.0 | 0.9 |
| 90% CI | 0–1 | | 2–5 | | 0–4 | 0–4 | 0–1 |
| detected | 0.9 | | 3.3 | | placed | CPT | flat |

Figure 115: Quad-modal dataset N4.




| N5 (penta) | peaks | | flats | | CPT | level | |
|---|---|---|---|---|---|---|---|
| | LP | Diw | LP | Diw | | | |
| mean count | 1.1 | 1.4 | 0 | 0.3 | 3.3 | 8.0 | |
| 90% CI | 0.9–2 | 0–3 | | 0–1 | 1–5 | 5.9–10 | |
| detected | 3.0 | 2.9 | 4.9 | 0.3 | | | |
| matching | 0.3 | | 0 | | 1.1 | 1.7 | 0 |
| 90% CI | 0–1 | | | | 0–3 | 0–4 | |
| detected | 1.0 | | 0 | | placed | CPT | flat |

Figure 116: Penta-modal claw dataset N5.

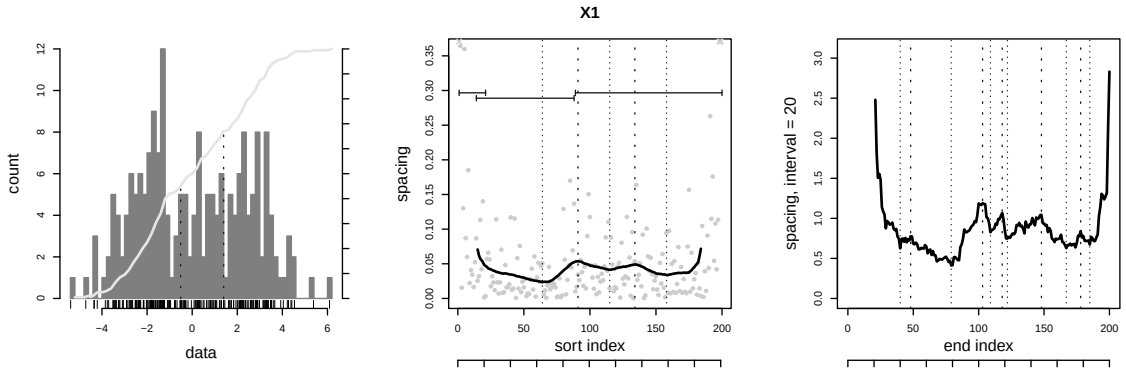



| P1 (tri) | peaks | | flats | | CPT | level | |
|---|---|---|---|---|---|---|---|
| | LP | Diw | LP | Diw | | | |
| mean count | 1.8 | 1.7 | 1.7 | 0.1 | 13.8 | 29.0 | |
| 90% CI | 1–2 | 0–4 | 0–3 | 0–1 | 2–27 | 28–30 | |
| detected | 1.8 | 18.2 | 5.9 | 2.8 | | | |
| matching | 0.2 | | 0.3 | | 6.3 | 13.5 | 0 |
| 90% CI | 0–1 | | 0–2 | | 1–14 | 2–27 | |
| detected | 1.2 | | 1.5 | | placed | CPT | flat |

Figure 117: Discrete tri-modal dataset P1 with Poisson draws.

N5 (Figure 116) has five narrow draws that are well-spaced for their width, plus a sixth variate forming a broad background located between the second and third draw of the claw. It is a larger version of sample W10, with 500 points in total with half in the background draw and 50 in each of the narrow. The figure shows peaks between the narrow draws, flats that correspond to some of the modes and some of the peaks, one changepoint at the edge of one of the draws, and level sections that roughly cover the peak-to-peak regions, plus sections in both tails. The measurements over the trials tell a different story. Both spacing methods contain three peaks, only one of which is significant, with the location of a third of the features, significant or not, matching. Similarly, flats appear consistently in the low-pass spacing; the five could match each mode or, as in the figure, also include the inter-mode transitions. But none are judged significant. The larger amplitude interval spacing prevents a flat from appearing except in one out of three runs, but in those cases it is accepted as significant. There are three changepoints on average, with a moderate spread between trials. The one with a correct placement may be misleading as there is no consideration in the check that the peak and flat surrounding it are adjacent. There are eight level sections with a wide confidence interval. Two of the changepoints match an endpoint, a rate higher than normal.

The PPP example in the main text is based on P1 (Figure 117). PPP is a little more challenging, with a smaller separation to the third Poisson draw and more equal draw sizes. The total draw in P1 follows the text and has 1200 points. As with PPP the low-spacing spacing has on average two peaks, both found significant. The interval spacing triggers at each discrete step. Any such feature passes the run length test at the 0.05 level, but two peaks on average pass at 0.01. The low matching of the significant peaks says the interval spacing judgement is spurious, and the matching of most detected peaks only reflects the number of interval peaks. There are six flats in the low-pass spacing stretching over several steps smoothed away by the filter. Two are accepted as significant, covering the first two draws; the flat in the third variate is too short to pass the tests. Interval flats are spurious, related




| X1 (bi) | peaks | | flats | | CPT | level | |
|---|---|---|---|---|---|---|---|
| | LP | Diw | LP | Diw | | | |
| mean count | 0.8 | 0.6 | 0.1 | 0 | 1.0 | 3.1 | |
| 90% CI | 0–1 | 0–2 | 0–1 | | 0–3 | 2–5 | |
| detected | 1.6 | 5.3 | 0.1 | 0 | | | |
| matching | 0.1 | | 0 | | 0 | 0.2 | 0.1 |
| 90% CI | 0–1 | | | | | 0–1 | 0–1 |
| detected | 1.4 | | 0 | | placed | CPT | flat |

Figure 118: Bi-modal dataset X1.

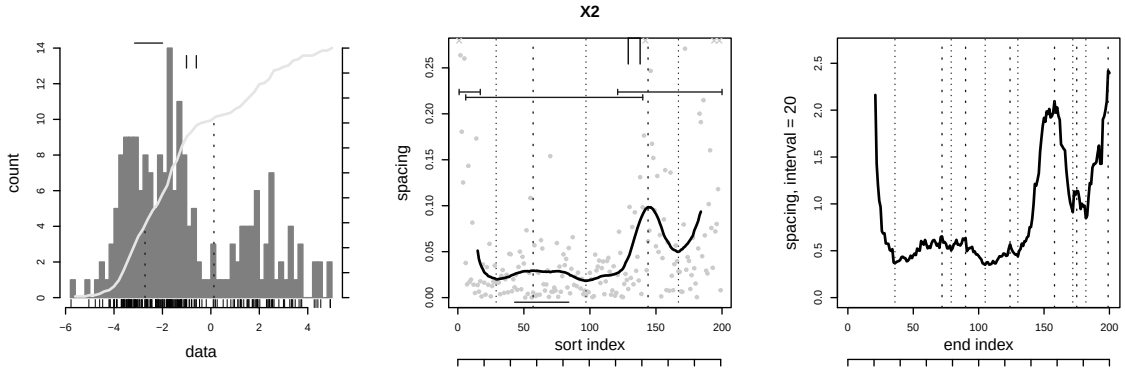
to the step width, and rejected. The changepoints and level section algorithms also trigger on each discrete step and cannot be used.

### X Samples

X1 (Figure 118) is a bi-modal normal draw with a separation of 4.0 and larger standard deviations than used in our variations. These changes cancel, and the overall performance is the same as our base variation with a separation of 3.0, as measured in the M1 and H1 samples.

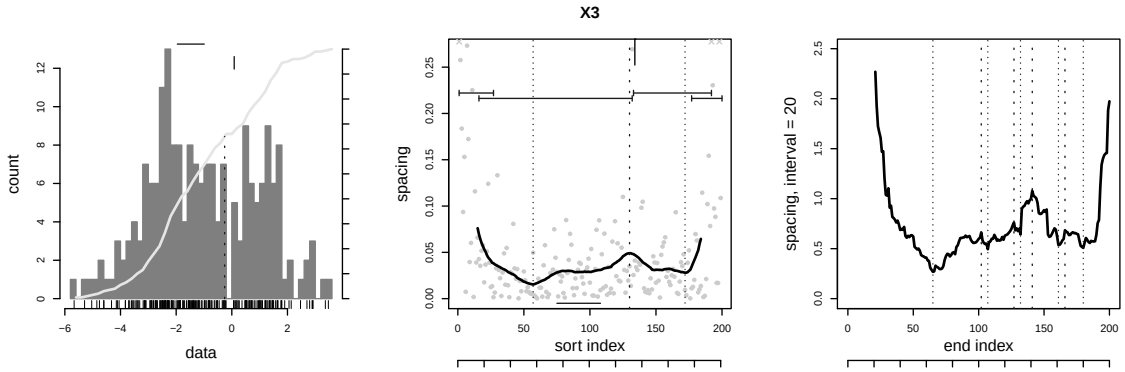
X2 (Figure 119) is an asymmetric bi-modal setup, with the same normal parameters as used in X1 but a 3:1 imbalance in the draw sizes. This makes the difference between the modes much clearer. But there are slightly fewer peaks detected and accepted in the low-pass spacing, and one peak fewer detected in the interval spacing, although the acceptance rate is the same. The imbalance increases the number of flats in the low-pass spacing, and most that are detected are accepted as significant. There are no flats in the interval spacing. The number of changepoints has risen to one or two, and the number of level sections has increased slightly. They agree better with the changepoints, and cover all significant flats. In the figure the main peak correctly divides the two draws, the level sections cover each, a flat appears in the larger draw, and the changepoints capture the increased spacing in the inter-mode transition. The interval spacing is rough, which generates many local maxima and blocks flat detection with the default parameters.

Sample X3 (Figure 120) has unequal draw sizes and differing standard deviations, making an asymmetric bi-modal distribution that has a smaller difference between its modes than X2 does. The annotations in the figure are similar to X2, with a peak in the low-pass spacing between the variates that is also captured by a changepoint, a flat in the larger draw to the side of the minimum, and level



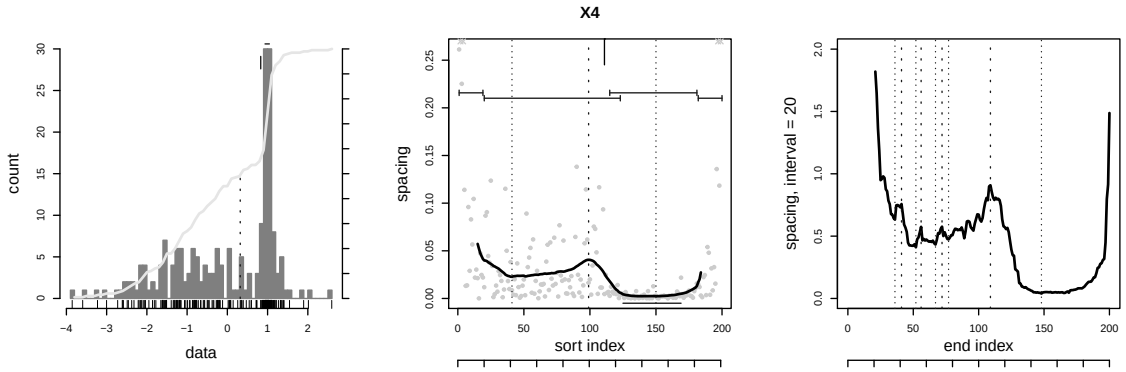

| X2 (abi) | peaks | | flats | | CPT | level | |
|---|---|---|---|---|---|---|---|
| | LP | Diw | LP | Diw | | | |
| mean count | 0.6 | 0.5 | 0.3 | 0 | 1.5 | 3.4 | |
| 90% CI | 0–1 | 0–2 | 0–1 | | 0–3 | 3–5 | |
| detected | 1.4 | 4.0 | 0.4 | 0 | | | |
| matching | 0.1 | | 0 | | 0.3 | 0.5 | 0.3 |
| 90% CI | 0–1 | | | | 0–1 | 0–2 | 0–1 |
| detected | 1.2 | | 0 | | placed | CPT | flat |

Figure 119: Asymmetric bi-modal dataset X2.




| X3 (abi) | peaks | | flats | | CPT | level | |
|---|---|---|---|---|---|---|---|
| | LP | Diw | LP | Diw | | | |
| mean count | 0.3 | 0.7 | 0.2 | 0 | 0.9 | 3.3 | |
| 90% CI | 0–1 | 0–4 | 0–1 | | 0–2 | 2–5 | |
| detected | 1.6 | 4.9 | 0.3 | 0 | | | |
| matching | 0.1 | | 0 | | 0 | 0.2 | 0.2 |
| 90% CI | 0–1 | | | | | 0–1 | 0–1 |
| detected | 1.4 | | 0 | | placed | CPT | flat |

Figure 120: Asymmetric bi-modal dataset X3.



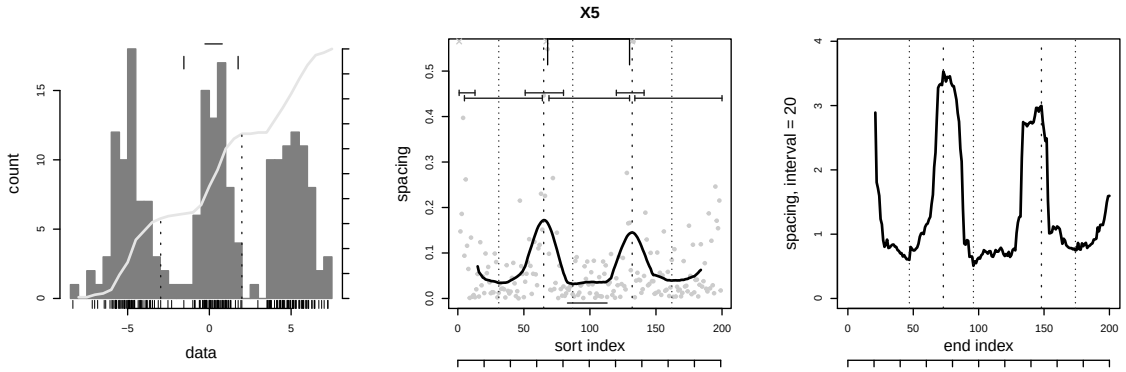

| X4 (abi) | peaks | | flats | | CPT | level | |
|---|---|---|---|---|---|---|---|
| | LP | Diw | LP | Diw | | | |
| mean count | 0.4 | 0.8 | 1.2 | 0.5 | 1.4 | 4.5 | |
| 90% CI | 0–1 | 0–2 | 1–2 | 0–1 | 0.9–3 | 3–6 | |
| detected | 1.3 | 3.0 | 1.2 | 0.5 | | | |
| matching | 0.2 | | 0.3 | | 1.0 | 0.8 | 1.2 |
| 90% CI | 0–1 | | 0–1 | | 0–2 | 0–2 | 1–2 |
| detected | 1.2 | | 0.3 | | placed | CPT | flat |

Figure 121: Asymmetric bi-modal dataset X4.

sections that cover the two modes. There are one or two peaks detected in the low-pass spacing, but most are rejected as insignificant; the peak is less prominent. The detector finds five peaks in the rougher interval spacing, one that is accepted significant but with a wide confidence interval. The detected peaks match, limited by the low-pass count, but significant peaks do much worse. A flat appears in the low-pass spacing in one out of three trials, and is usually accepted as significant. There is one changepoint and three level sections, but they have only a weak agreement. All significant flats are covered by a level section.

The X4 sample (Figure 121) adds a small, narrow normal draw to a broad background. The low-pass spacing contains one or two peaks, the interval spacing three, and a third of each are accepted as significant. The detected peaks match, limited by the lower rate in the low-pass spacing, but only half the significant peaks do, which implies that different peaks are being accepted. The low-pass spacing contains one flat, the interval spacing one in every two trials. All flats detected are accepted as significant. The two methods mostly agree, with the matching rate limited by the interval spacing flats. In the figure the flat falls within the tight draw. There are one or two changepoints per trial, with one falling between a peak and flat and therefore marking the transition between the draws. There are four or five level sections, two spanning the modes and two in the tails. One changepoint aligns with the level endpoints, which in the figure overlap slightly at the transition. The level sections also overlap any significant flat.

X5 (Figure 122) has three standard normal draws separated by 5.0. The results indicate this is easy to resolve. The two peaks between the modes are detected and accepted as significant in the low-pass spacing. The interval spacing contains more peaks, four, but fewer significant, one with a moderate confidence interval. Peaks from the two methods match, the significant limited by those found in the interval spacing and the detected by the low-pass. Flats rarely appear, and never in the



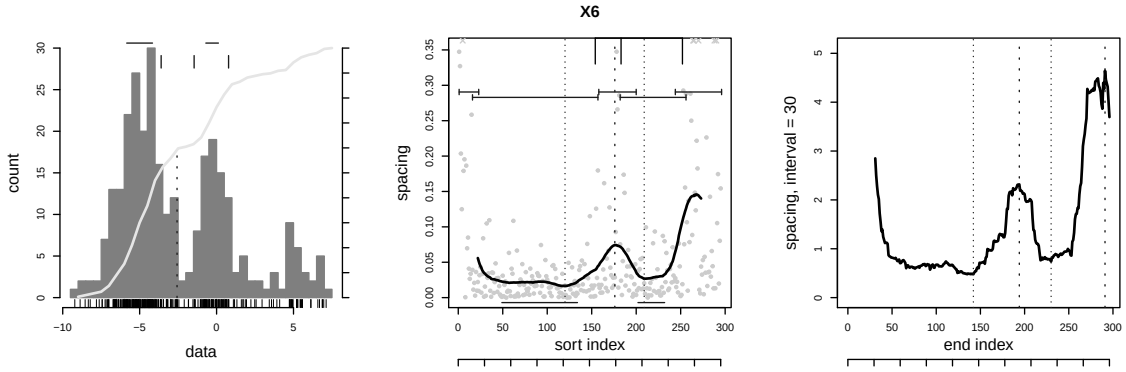

| X5 (tri) | peaks | | flats | | CPT | level | |
|---|---|---|---|---|---|---|---|
| | LP | Diw | LP | Diw | | | |
| mean count | 1.8 | 1.1 | 0.2 | 0 | 3.0 | 5.5 | |
| 90% CI | 1–2 | 0–3 | 0–1 | | 2–4 | 1.9–7 | |
| detected | 2.0 | 4.0 | 0.2 | 0 | | | |
| matching | 0.7 | | 0 | | 0.3 | 1.4 | 0.2 |
| 90% CI | 0–2 | | | | 0–2 | 0–3 | 0–1 |
| detected | 1.7 | | 0 | | placed | CPT | flat |

Figure 122: Tri-modal dataset X5.

interval spacing. The figure shows one in the middle variate, with the tails preventing the filtered spacing from reaching a stable value. There are three changepoints whose alignment is limited by the number of flats; in the figure they coincide with the peaks. The five or six level sections span the modes, transitions, and tails, with a wide spread in the confidence interval. Half the changepoints match a level endpoint. A level section overlaps all significant flats.

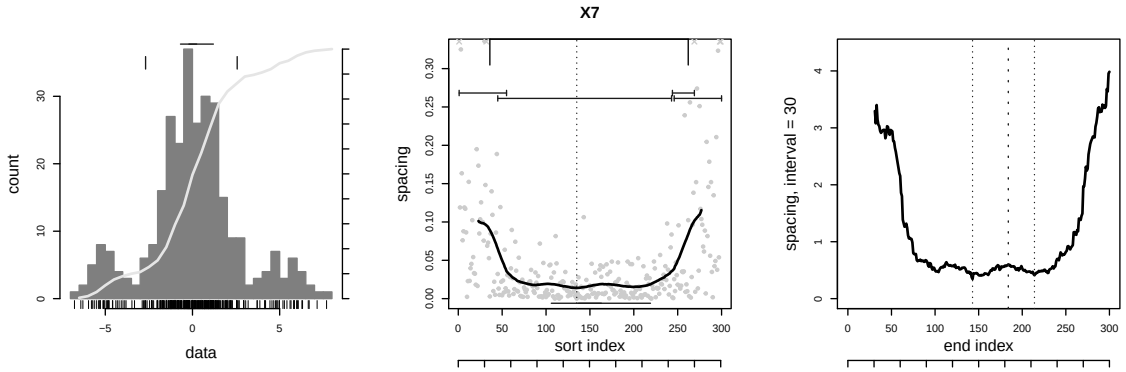
The X6 sample (Figure 123) is an unequal tri-modal distribution with a large, broad normal variate aside two standard normals, one small. There are 300 points drawn in total so the smallest has 32. The separation between each mode is 5.0, making it easy to distinguish the draws. The low-pass spacing contains one peak, the interval spacing two or three. In the figure the third variate is too small and at the edge of the draw, so the filter cuts off the peak while the interval runs and includes it. Tests reject all low-pass peaks as insignificant, but pass half of the interval spacing features. Half the detected peaks match between the two methods. The low-pass spacing contains two flats in the first two draws, with the third lost in the edge; one of these is accepted as significant. There are no flats in the interval spacing, but the figure suggests more are possible with a larger ripple. The spacing contains three changepoints bounding the first inter-mode transition and the larger side of the second. One of the changepoints falls between a peak and flat. There are five level sections that cover the first two modes and the transition between them, plus the two tails. One or two changepoints match level endpoints, and the level sections cover all significant flats.

X7 (Figure 124) is a symmetric tri-modal distribution with two small lobes aside a somewhat wider main variate. The total draw is 300 points because of the side lobes. Although the 5.0 separation should make it easy to distinguish the draws, the measurements show this does not happen. The side lobes are located in the tails and lost as edge effects during filtering. This means that they cannot be detected even if the total draw continued to scale, say to 500 points, because the filter size is a fraction



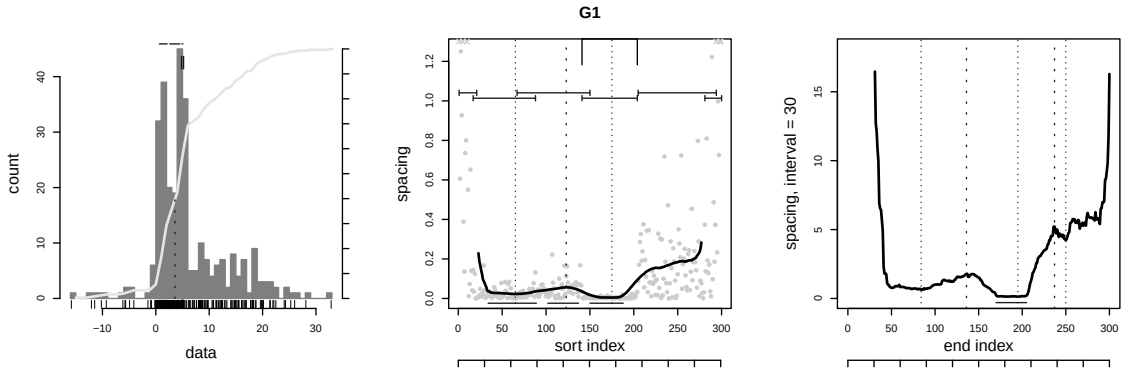

| X6 (tri) | peaks | | flats | | CPT | level | |
|---|---|---|---|---|---|---|---|
| | LP | Diw | LP | Diw | | | |
| mean count | 0 | 1.3 | 1.0 | 0 | 2.8 | 5.3 | |
| 90% CI | | 0–2 | 0–2 | | 1–4 | 4–7 | |
| detected | 1.1 | 2.6 | 1.8 | 0 | | | |
| matching | 0 | | 0 | | 1.1 | 1.3 | 1.0 |
| 90% CI | | | | | 0–2 | 0–3 | 0–2 |
| detected | 0.6 | | 0 | | placed | CPT | flat |

Figure 123: Tri-modal dataset X6.




| X7 (tri) | peaks | | flats | | CPT | level | |
|---|---|---|---|---|---|---|---|
| | LP | Diw | LP | Diw | | | |
| mean count | 0 | 0.8 | 1.3 | 0 | 2.8 | 4.2 | |
| 90% CI | | 0–2 | 0.9–2 | | 2–4 | 3–6 | |
| detected | 0.2 | 1.4 | 1.8 | 0 | | | |
| matching | 0 | | 0 | | 0 | 0.8 | 1.3 |
| 90% CI | | | | | | 0–2 | 1–2 |
| detected | 0.1 | | 0 | | placed | CPT | flat |

Figure 124: Tri-modal dataset X7.



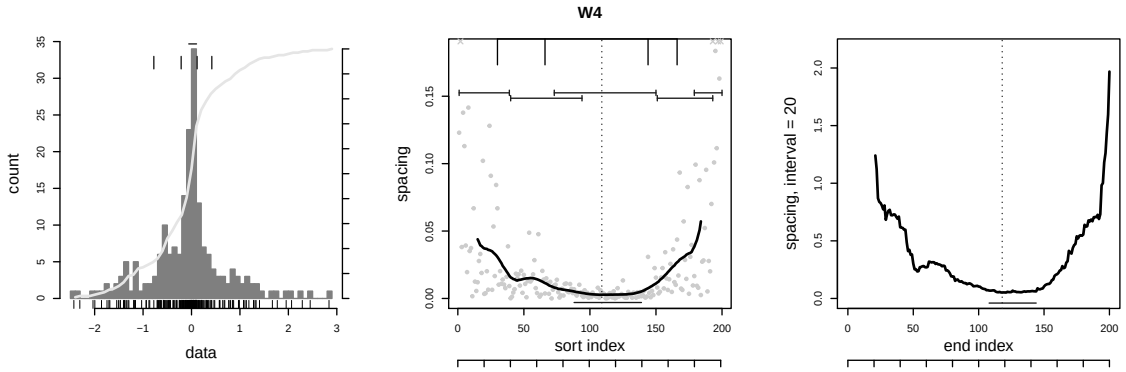

| G1 (tri) | peaks | | flats | | CPT | level | |
|---|---|---|---|---|---|---|---|
| | LP | Diw | LP | Diw | | | |
| mean count | 0.3 | 1.7 | 1.9 | 1.3 | 3.0 | 6.1 | |
| 90% CI | 0–1 | 1–3 | 1–2 | 1–2 | 2–4 | 5–7 | |
| detected | 1.4 | 2.3 | 2.7 | 1.3 | | | |
| matching | 0.2 | | 0.8 | | 1.8 | 2.3 | 1.9 |
| 90% CI | 0–1 | | 0–2 | | 1–3 | 1–4 | 1–2 |
| detected | 1.1 | | 0.8 | | placed | CPT | flat |

Figure 125: Tri-modal dataset G1 with gamma and normal draws.

of the total and grows accordingly. Instead, the ratio between the small draw size and the main would have to increase to make the transition visible, for example from 30:240:30 to 45:210:45. As it is, the low-pass spacing contains no peaks, while the interval spacing, which better accommodates edge effects, finds one or two, one of which is significant. The detected peaks partially align, limited by the low count in the interval spacing. The low-pass spacing does contain two flats that span the main variate, with most passing the acceptance test, while the default detection parameters prevent any flat from appearing in the interval spacing. The changepoint algorithms do pick up the inter-mode transitions, plus a third changepoint on average, with one matching to a level section endpoint. There are four level sections, one spanning the main variate and two in the tails, plus an extra. The large section overlaps any significant flats.

### G Sample

G1 mixes a gamma variate with a tight and a broad normal variate to produce an asymmetric tri-modal distribution with 300 points in total. The tight draw is clear in Figure 125 between indices 150 and 200, with sharp boundaries to the region and different average spacing levels to the left and right. Changepoints mark these boundaries, and the level sections show the three modes plus tails. There are one or two peaks in the low-pass and interval spacing. Tests usually accept both interval peaks but only occasionally the low-pass. Matching between the methods is limited by the low-pass results but is otherwise good. There are three flats in the low-pass spacing, two of which are considered significant, and one in the interval spacing, which is accepted, corresponds to the tight draw. This agrees with the low-pass result in half the trials. There are usually three changepoints, of which the two that bound the middle variate are correctly placed. The six level sections are stable, and two of the changepoints match an endpoint. Level sections cover the significant flats.



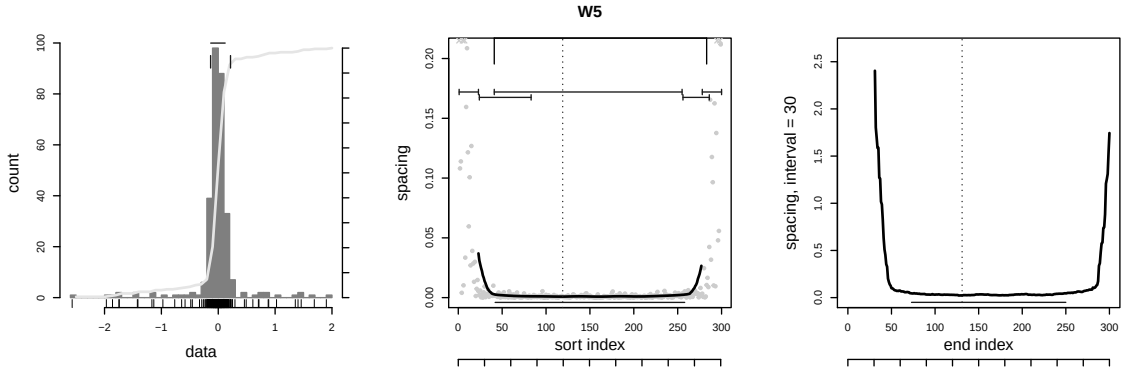

| W4 (uni) | peaks | | flats | | CPT | level | |
|---|---|---|---|---|---|---|---|
| | LP | Diw | LP | Diw | | | |
| mean count | 0.1 | 0.3 | 1.0 | 0.3 | 2.1 | 4.8 | |
| 90% CI | 0–1 | 0–2 | 1–1 | 0–1 | 1–4 | 4–6 | |
| detected | 0.2 | 1.4 | 1.0 | 0.3 | | | |
| matching | 0 | | 0.2 | | 0.2 | 1.2 | 1.0 |
| 90% CI | | | 0–1 | | 0–1 | 0–3 | 1–1 |
| detected | 0.2 | | 0.2 | | placed | CPT | flat |

Figure 126: Uni-modal dataset W4.

### W Samples

W4 has two superimposed normal draws, one small and tight, the other large and broad, forming a univariate distribution. The spacing in Figure 126 shows the tight draw in the center with a larger spacing to either side and a second step upward in the tails. A peak rarely appears in the low-pass spacing, though if it does it is accepted as significant half the time. The interval spacing contains one or two peaks, aligned with the detected low-pass peaks but mostly rejected by testing. The tight draw generates one significant low-pass flat in every trial run, while the interval spacing occasionally contains a matching feature, which again is marked as significant every time it appears. A level section covers the flat. There are two changepoints on average, with a moderate confidence interval. They bound the tight draw. The five level sections pick up the tiered spacing, covering the central peak, the flanks to either side, and the tails. One of the changepoints matches a level endpoint.

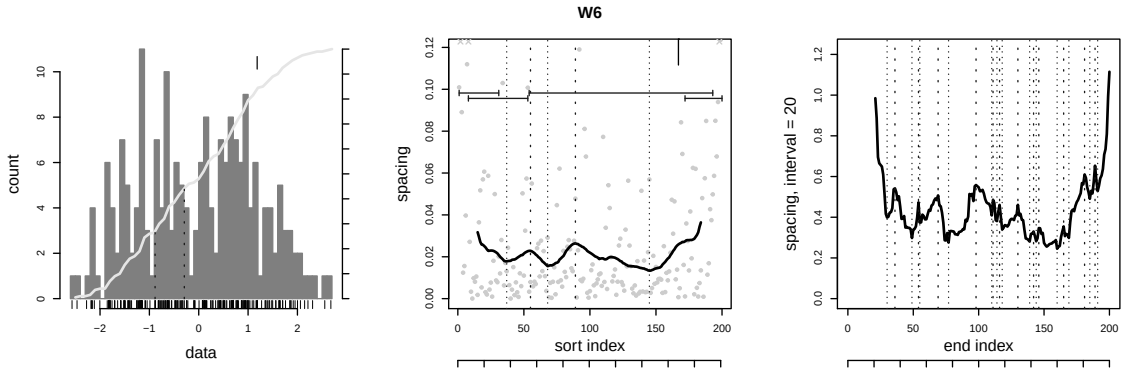
Like W4, W5 (Figure 127) superimposes broad and tight normal variates. For this distribution, the tight draw is the largest, by a 9:1 ratio, compared to the 1:2 ratio in W4. There are therefore no flanks in the spacing, just the central peak and tails. The total draw has 300 points. Either spacing is smooth in the figure, with a long flat in the tight draw and matching level section, and changepoints and level sections marking the tails. The confidence intervals indicate this behavior is stable. There are no peaks in either the low-pass or interval spacing, and one flat, or two in the interval spacing in half the runs. All flats are significant and covered by level sections. The two methods agree; in this case the matching rate is higher than the number of low-pass flats, so some low-pass flats must split into two shorter subsets in the interval spacing. There are two changepoints and five level sections whose endpoints align with one changepoint on average. A level section covers the low-pass flat, so the extra level sections must still be able to mark the broader draw.




W5 (uni)

| | peaks | | flats | | CPT | level | |
|---|---|---|---|---|---|---|---|
| | LP | Diw | LP | Diw | | | |
| mean count | 0 | 0 | 1.0 | 1.4 | 1.9 | 5.4 | |
| 90% CI | | | 1–1 | 1–3 | 0–3 | 5–7 | |
| detected | 0 | 0 | 1.0 | 1.5 | | | |
| matching | 0 | | 1.2 | | 0 | 0.7 | 0.9 |
| 90% CI | | | 1–2 | | | 0–2 | 0–1 |
| detected | 0 | | 1.4 | | placed | CPT | flat |

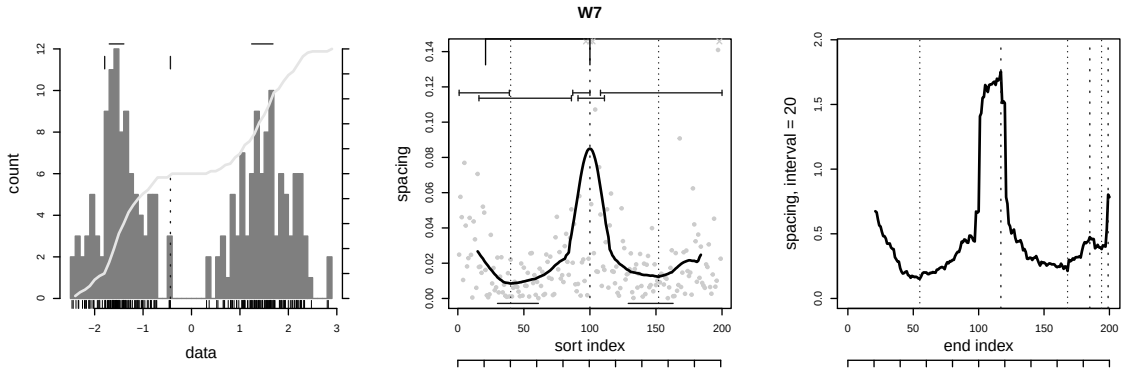
Figure 127: Uni-modal dataset W5.




W6 (tri)

| | peaks | | flats | | CPT | level | |
|---|---|---|---|---|---|---|---|
| | LP | Diw | LP | Diw | | | |
| mean count | 0.6 | 0.8 | 0.1 | 0 | 0.8 | 3.0 | |
| 90% CI | 0–1 | 0–4 | 0–1 | | 0–3 | 2–5 | |
| detected | 1.6 | 5.1 | 0.1 | 0 | | | |
| matching | 0.1 | | 0 | | 0 | 0.2 | 0.1 |
| 90% CI | 0–1 | | | | | 0–1 | 0–1 |
| detected | 1.3 | | 0 | | placed | CPT | flat |

Figure 128: Bi-modal dataset W6.



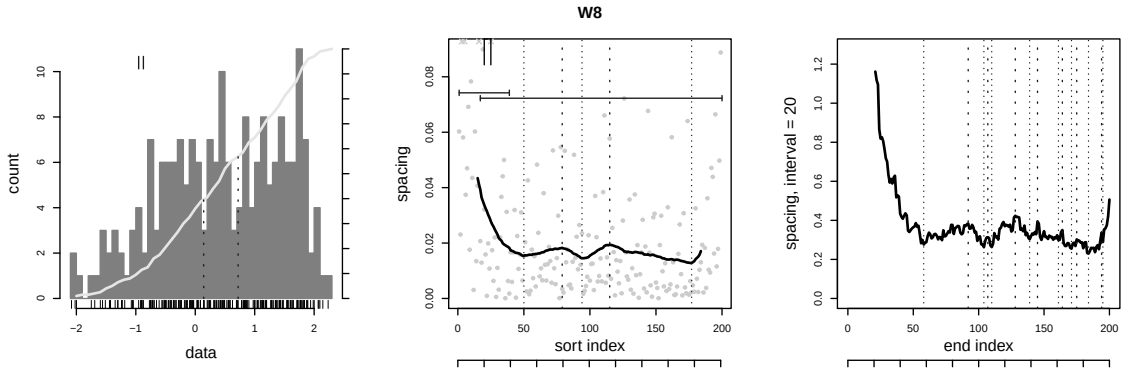

| W7 (bi) | peaks | | flats | | CPT | level | |
|---|---|---|---|---|---|---|---|
| | LP | Diw | LP | Diw | | | |
| mean count | 1.0 | 0.8 | 1.5 | 0 | 1.9 | 4.7 | |
| 90% CI | 1–1 | 0–2 | 0.9–2 | | 1–3 | 4–6 | |
| detected | 1.1 | 2.3 | 1.5 | 0 | | | |
| matching | 0.6 | | 0 | | 1.3 | 1.0 | 1.5 |
| 90% CI | 0–1 | | | | 0–2 | 0–2 | 0.9–2 |
| detected | 0.9 | | 0 | | placed | CPT | flat |

Figure 129: Bi-modal dataset W7.

The W6 sample (Figure 128) is a bi-modal pair of normal variates, separated by 2.0 and with a standard deviation of 2/3. This is a critical setup, partially resolved and subject to noise. The low-pass spacing has one or two peaks, of which one passes the tests in half the trials. The interval spacing is very rough and has five peaks, but only one is judged significant, with a large confidence interval. The detected low-pass peaks agree with those from the interval spacing but the significant peaks rarely match. This suggests the low-pass peaks form a subset of the interval features and that testing does not select within the subset. There are no flats in either method. One changepoint marks a tail. The level finding algorithm flags three sections, with a moderate spread; in the figure the longest overlaps the true inter-mode transition at index 100. Matching to a changepoint is poor.

W7 (Figure 129) is a bi-modal normal draw. Separated by 3.0 and with a standard deviation of 0.5, it is easier to resolve than W6 or M1/H1. The low-pass spacing contains one peak repeatably, accepted as significant. The interval spacing has two, with one usually accepted. The detected peaks match but the alignment is somewhat worse for the significant. This might reflect noise in the interval peak's position, not sharp in the figure. The second interval peak comes from outliers in the second draw that are separated from the main lobe. The low-pass spacing contains flats in one or two of the draws, with the detected features passing the acceptance tests. There are no flats in the interval spacing. The spacing contains two changepoints, usually placed in the transition between the draws. Five level sections cover the two variates, the transition, or the tails. The changepoint at the peak matches an endpoint. Sections cover all significant flats, with a confidence interval that matches that of the flats.

Sample W8 (Figure 130) is an asymmetric bi-modal distribution that is not cleanly separated. One or two peaks may appear in the low-pass spacing, but tests reject all of them. The interval spacing is not smooth and generates four peaks, one of which passes testing in half the trials. Detected peaks



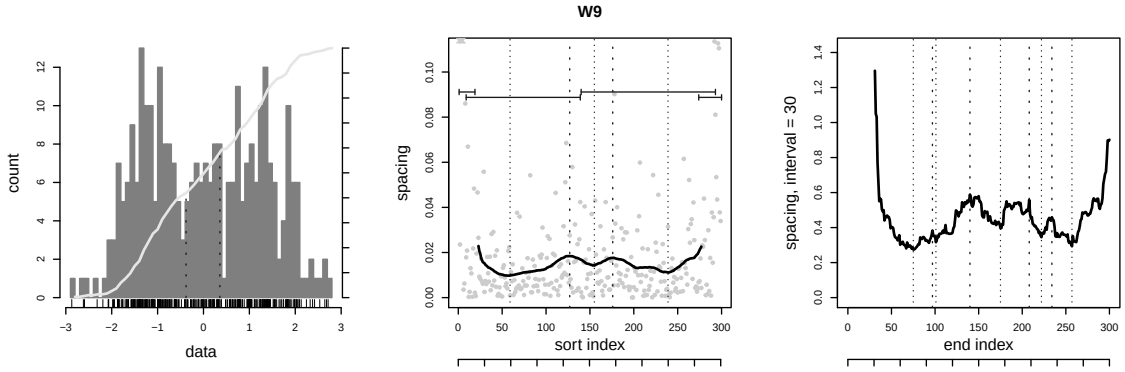

| W8 (abi) | peaks | | flats | | CPT | level | |
|---|---|---|---|---|---|---|---|
| | LP | Diw | LP | Diw | | | |
| mean count | 0.1 | 0.6 | 0.2 | 0 | 0.9 | 2.8 | |
| 90% CI | 0–1 | 0–2 | 0–1 | | 0–2 | 2–4 | |
| detected | 1.6 | 4.1 | 0.5 | 0 | | | |
| matching | 0 | | 0 | | 0 | 0.4 | 0.2 |
| 90% CI | | | | | | 0–1 | 0–1 |
| detected | 1.4 | | 0 | | placed | CPT | flat |

Figure 130: Asymmetric bi-modal dataset W8.

match up, and there are too few significant low-pass peaks to say if they align. A flat may appear in the low-pass spacing but usually testing rejects it; the interval spacing contains no flats. There is one changepoint, in the figure located at the edge of the starting tail. Three level sections span the data and the two tails. In every other run the changepoint matches an endpoint. A level section overlaps any significant flat that may exist.

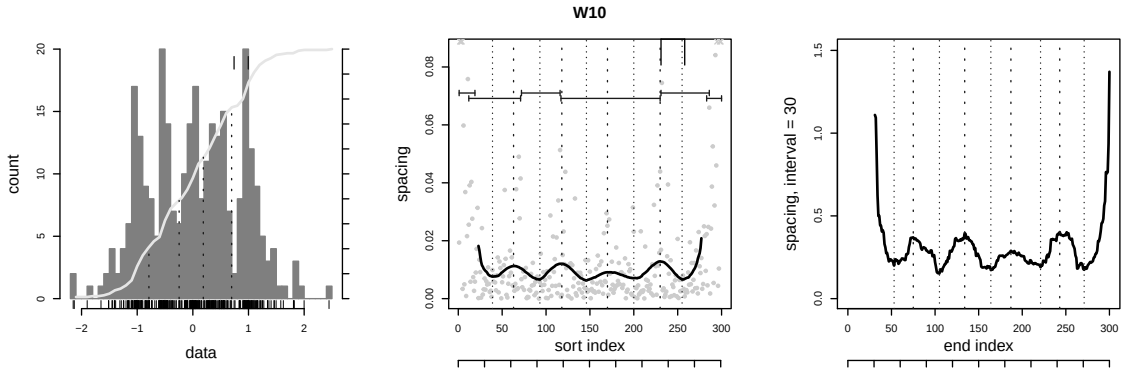
W9 (Figure 131) contains three normal variates, a small narrow central draw surrounded by two wider and much larger draws. There are 300 points in total so that the smallest draw has 30 points. The spacing can resolve this setup, although the results are marginal. There are usually two peaks in the low-pass spacing, and testing accepts one of them as significant. The interval spacing has five peaks, with one passing in every other trial. The detected low-pass peaks usually align with those in the interval spacing, but the significant peaks do not match. The low-pass spacing has a flat in every other run, found significant, but the interval spacing has none. There is one changepoint that occasionally aligns with an endpoint of the four level sections found. In the figure one large section covers one mode, and the other large section covers two; they do not cleanly distinguish the modes. A section will cover any flat found significant.

W10 (Figure 132) is a penta-modal distribution, equivalent to N5 but with 3/5 the draw size, making each tooth in the comb 30 points wide, for a total of 300 points. Compared to the N5 results both spacing treatments detect all four anti-modes, but the smaller draws and peak heights mean that tests accept few in the low-pass spacing and just one in the interval spacing, an acceptance rate that is lower. The matching between detected peaks is better, however. In the figure the increased spacing between each tooth is clear. The draws are too small for flats. The changepoints algorithms have a harder time, finding one or two points at half the rate in N5, and the lack of flats means there are no correct placements. There are six level sections that in the figure stretch between peaks,



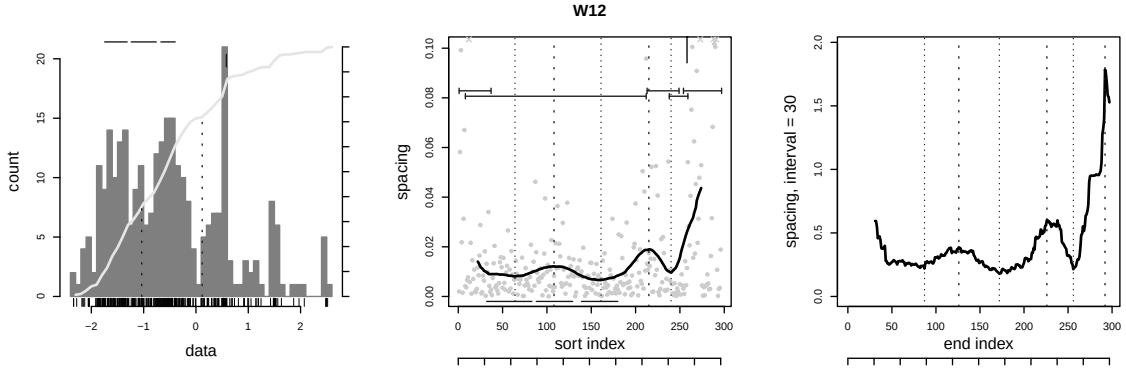

| W9 (tri) | peaks | | flats | | CPT | level | |
|---|---|---|---|---|---|---|---|
| | LP | Diw | LP | Diw | | | |
| mean count | 1.2 | 0.4 | 0.4 | 0 | 1.1 | 3.7 | |
| 90% CI | 0–2 | 0–2 | 0–1 | | 0–3 | 2–5 | |
| detected | 1.7 | 4.7 | 0.5 | 0 | | | |
| matching | 0.1 | | 0 | | 0.1 | 0.3 | 0.4 |
| 90% CI | 0–1 | | | | 0–1 | 0–1 | 0–1 |
| detected | 1.2 | | 0 | | placed | CPT | flat |

Figure 131: Tri-modal dataset W9.




| W10 (penta) | peaks | | flats | | CPT | level | |
|---|---|---|---|---|---|---|---|
| | LP | Diw | LP | Diw | | | |
| mean count | 0.2 | 0.8 | 0.1 | 0 | 1.6 | 5.9 | |
| 90% CI | 0–1 | 0–2 | | | 0–4 | 4–8 | |
| detected | 3.8 | 3.9 | 0.2 | 0 | | | |
| matching | 0.1 | | 0 | | 0.1 | 1.0 | 0.1 |
| 90% CI | 0–1 | | | | 0–1 | 0–3 | 0–1 |
| detected | 2.5 | | 0 | | placed | CPT | flat |

Figure 132: Penta-modal dataset W10.



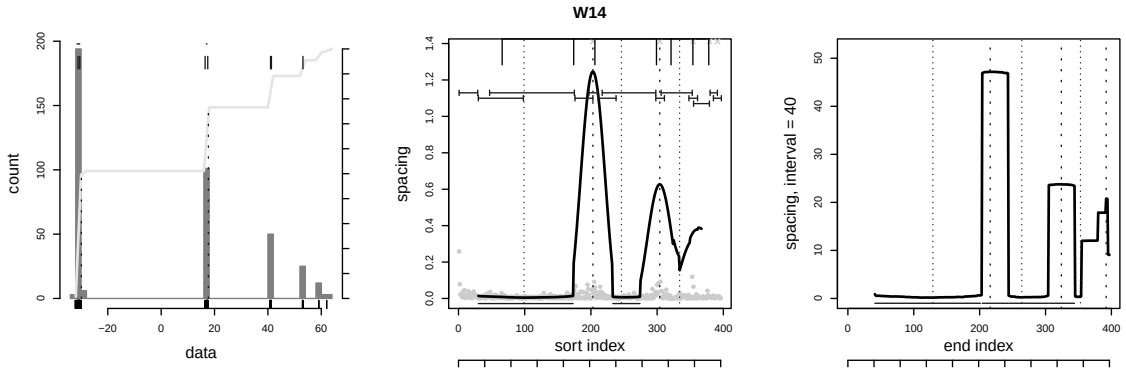

| W12 (penta) | peaks | | flats | | CPT | level | |
|---|---|---|---|---|---|---|---|
| | LP | Diw | LP | Diw | | | |
| mean count | 0.1 | 1.2 | 0.8 | 0 | 2.4 | 5.6 | |
| 90% CI | 0–1 | 0–3 | 0–2 | | 1–4 | 4–7 | |
| detected | 2.0 | 3.2 | 1.9 | 0 | | | |
| matching | 0.1 | | 0 | | 0.7 | 1.2 | 0.8 |
| 90% CI | 0–1 | | | | 0–1 | 0–3 | 0–2 |
| detected | 1.2 | | 0 | | placed | CPT | flat |

Figure 133: Penta-modal dataset W12.

combining the two variates with the shallowest separation at index 200. Allowing for one or two level sections in the tails, they cover four or five of the teeth. A section covers any significant flat.

W12 (Figure 133) is another claw with five tight, small normal variates atop a standard normal background. The draw size and standard deviation of the teeth decrease by a factor of two stepping from left to right. The background is offset so the overall distribution is asymmetric. There are 297 data points. Test results are mixed. The low-pass spacing contains two peaks, neither found significant. The interval spacing contains three, one of which reliably passes. The smallest draws on the right side of the figure disappear as edge effects of the filter, while the interval spacing, which does not have a dead zone, is able to pick up the last tooth as the largest, and most significant, peak. The low-pass spacing contains two flats on average, one that is significant, which correspond to the left two teeth; there is a third flat in the figure in the transition between them. The interval spacing contains no flats. There are two or three changepoints; the placement check is limited by the flat. The six level sections are not enough to identify every tooth, assuming they also include both tails; in the figure they do not separate the two left teeth. Half the changepoints, or one, match a level section endpoint. A section covers all significant flats.

W14 (Figure 134) is a chirp, a series of draws of decreasing size, width, and separation. In total there are six variates in the distribution, with a total of 397 points. The sizes of the three smallest teeth, respectively 25, 12, and 6 points, are below our minimum of 30. The histogram shows the six teeth with a large gap between them and decreasing count. The separation determines the height of the peaks in the interval spacing with a width equal to the interval until it covers more than one variate. The draw size fixes the space between them, with the spacing negligible compared to the separation. The interval spacing has square peaks because the transition is abrupt, while thse are rounded by the low-pass filter.




| W14 (chirp) | peaks | | flats | | CPT | level | |
|---|---|---|---|---|---|---|---|
| | LP | Diw | LP | Diw | | | |
| mean count | 1.2 | 1.2 | 2.0 | 0.3 | 4.2 | 10.9 | |
| 90% CI | 1–2 | 0–2 | 2–2 | 0–1 | 3–5 | 10–13 | |
| detected | 2.0 | 3.0 | 2.0 | 4.0 | | | |
| matching | 0.8 | | 1.0 | | 2.1 | 5.0 | 2.0 |
| 90% CI | 0–1 | | 1–1 | | 1–3 | 3–7 | 2–2 |
| detected | 1.9 | | 1.0 | | placed | CPT | flat |

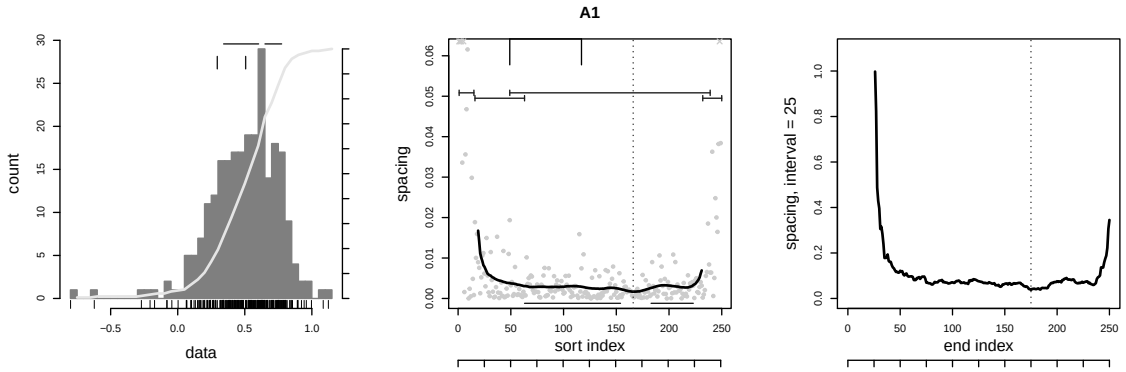
Figure 134: Chirp dataset W14.

The low-pass spacing resolves the first three teeth, the interval spacing the first four because it does not cut-off filtering on the right. The gap between the first two teeth is significant. The position of the peaks in both methods, detected or significant, match. The first teeth are also large enough draws to form flats, two in the low-pass spacing, both found significant, and four in the interval spacing, few passing. The difference in the integral steps in the interval spacing and the smoothed edges of the low-pass filter mean that only the largest, first flat matches. Changepoints pick up three of the teeth, bounding each side of the variate. The number of flats determine how many are correctly placed. The level section algorithm does best at finding the small teeth, identifying 11 sections fairly reliably. All changepoints match an endpoint, and a section covers all flats.

Scaling the draw sizes by 5 to bring the smallest to 32 points makes only small changes to the results. The changepoints pick up the fifth tooth, flats in the interval spacing find the transition between the third and fourth, and there are more level sections within each mode. The relative kernel size in the low-pass spacing prevents changes to its features. Since the modes are so sharply bounded we could reduce the kernel window to 1% of the data, accurately capturing the transitions. This has peaks between all six teeth even with the base draw sizes, but only the first is significant, the second sometimes so. There are also flats for the first three variates, all significant; the fourth through sixth draws are too small and fall below the minimum flat length. Scaling the draws and using the smaller window adds a significant flat at the fourth tooth, and makes the peak between the second and third variates more stably significant.

## A Samples

The single mode in sample A1 (Figure 135) is built from two overlapping normal draws that create an asymmetric distribution, plus a small third draw between them that increases the tails. This small

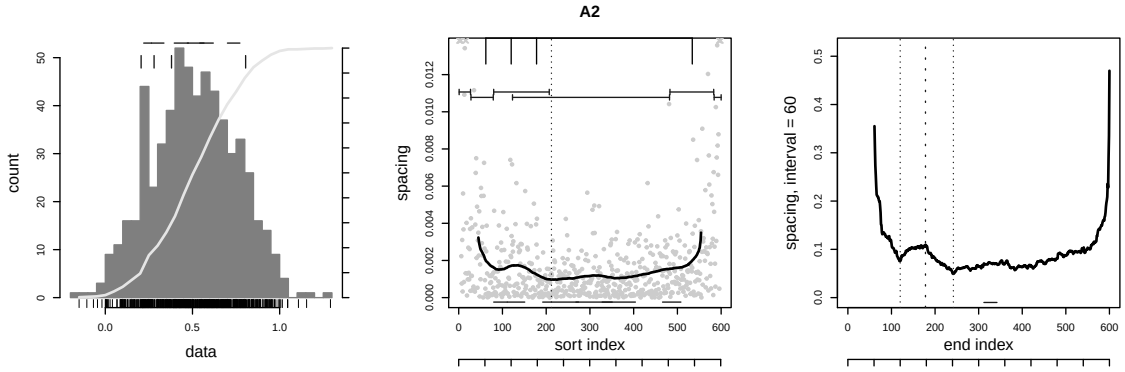



| A1 (uni) | peaks | | flats | | CPT | level | |
|---|---|---|---|---|---|---|---|
| | LP | Diw | LP | Diw | | | |
| mean count | 0 | 0.1 | 0.3 | 0.1 | 1.1 | 4.2 | |
| 90% CI | | 0–1 | 0–1 | | 0–2 | 3–5 | |
| detected | 0.9 | 0.8 | 1.8 | 0.1 | | | |
| matching | 0 | | 0 | | 0 | 0.6 | 0.3 |
| 90% CI | | | | | | 0–2 | 0–1 |
| detected | 0.4 | | 0 | | placed | CPT | flat |

Figure 135: Uni-modal dataset A1.

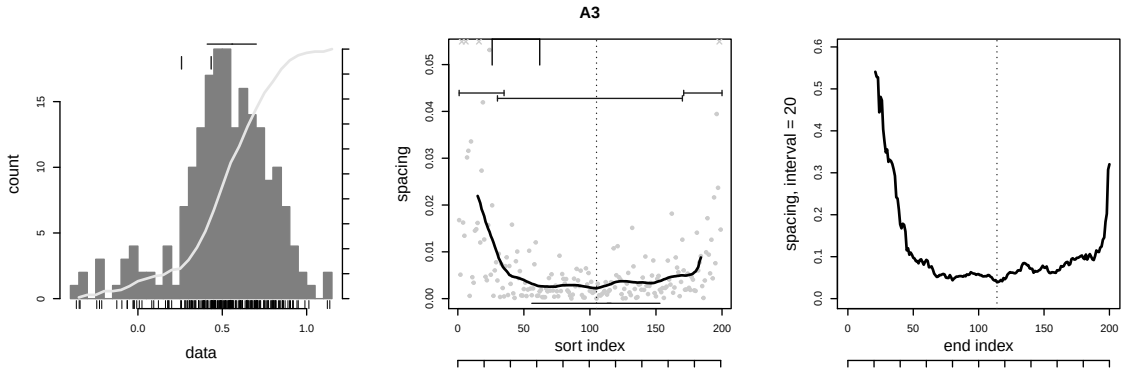
variate creates several points with larger spacing at index 110 in the figure, and points with elevated spacing towards the tails. There are 250 points in the total draw. Both the low-pass and interval spacing contain one peak on average whose position matches in half the runs. Tests reject it as significant, though. The low-pass spacing contains two flats in the mode, one passing the acceptance tests in a third of the trials. The interval spacing has no flats, although in the figure the signal seems clean. The sample generates four level sections. In the figure one spans the mode and the others are in the tails, including one across a shelf from the third draw at the left side. An endpoint matches the changepoint found in half the trials, and one of the sections overlaps any significant flat.

Two very small draws are added to either flank of the main normal variate in Sample A2 (Figure 136) to adjust the transition to the tails. There are 600 points in the total draw so these two draws have 30 points. The interval spacing contains one peak on average, the low-pass spacing a peak in half the runs, but their position does not match. In the figure the maximum falls between the first side draw and main variate. Tests reject all detected peaks. The low-pass spacing contains four flats which tests reject as significant, the interval spacing one in half the trials which is accepted. Flats match between the two methods in a third of the trials; this rate is higher than the significant low-pass count, so some of the low-pass flats cover more than one interval feature. The interval spacing seems clean enough to allow a looser ripple parameter. There are five level sections which in the figure span the first two draws and the transition between them, plus two for the tails. The sections and changepoints on the left correspond to a region of increased spacing. But this does not come from the small draw, which is only 30 points, and it must be an artefact of the particular draw. The lack of a flank on the right supports this interpretation. Some section covers all significant low-pass flats. There are two changepoints with a moderate spread in the confidence interval, but in only half the trials does one align with the endpoint of a level section.



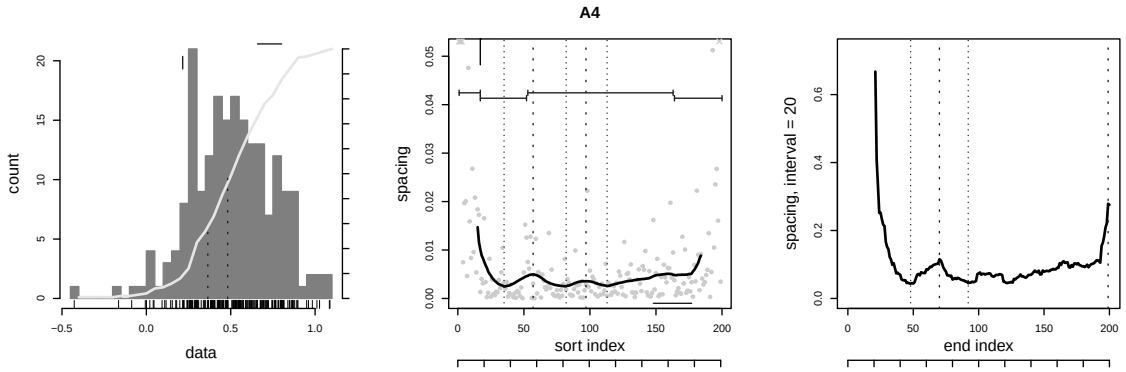

| A2 (uni) | peaks | | flats | | CPT | level | |
|---|---|---|---|---|---|---|---|
| | LP | Diw | LP | Diw | | | |
| mean count | 0 | 0.1 | 0.2 | 0.5 | 2.2 | 5.3 | |
| 90% CI | | 0–1 | 0–1 | 0–2 | 1–4 | 4–6 | |
| detected | 0.4 | 1.0 | 4.2 | 0.5 | | | |
| matching | 0 | | 0.3 | | 0.1 | 0.5 | 0.2 |
| 90% CI | | | 0–2 | | 0–1 | 0–1 | 0–1 |
| detected | 0.1 | | 0.3 | | placed | CPT | flat |

Figure 136: Uni-modal dataset A2.




| A3 (uni) | peaks | | flats | | CPT | level | |
|---|---|---|---|---|---|---|---|
| | LP | Diw | LP | Diw | | | |
| mean count | 0 | 0.1 | 0.8 | 0 | 1.6 | 4.2 | |
| 90% CI | | 0–1 | 0–2 | | 0.9–3 | 3–5 | |
| detected | 0.3 | 0.8 | 1.4 | 0 | | | |
| matching | 0 | | 0 | | 0 | 0.9 | 0.8 |
| 90% CI | | | | | | 0–2 | 0–2 |
| detected | 0.2 | | 0 | | placed | CPT | flat |

Figure 137: Uni-modal dataset A3.



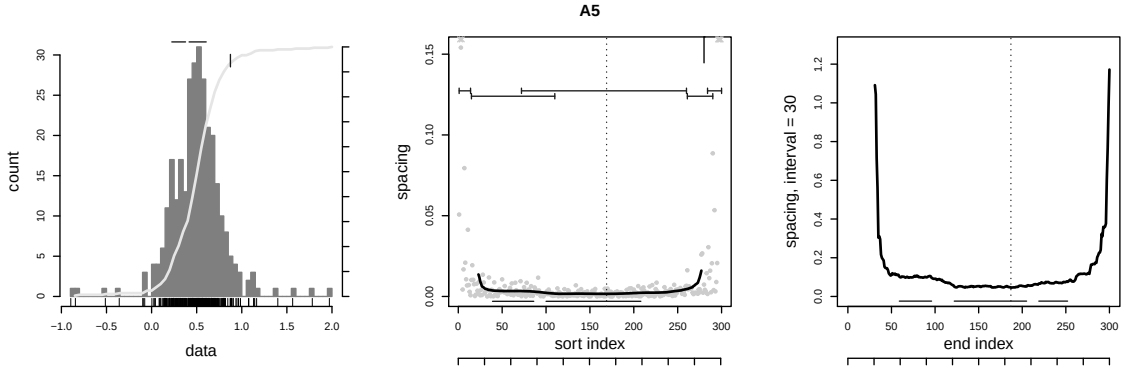

| A4 (uni) | peaks | | flats | | CPT | level | |
|---|---|---|---|---|---|---|---|
| | LP | Diw | LP | Diw | | | |
| mean count | 0 | 0.4 | 0.4 | 0 | 1.0 | 3.7 | |
| 90% CI | | 0–2 | 0–1 | | 0–2 | 3–5 | |
| detected | 1.2 | 3.1 | 0.7 | 0 | | | |
| matching | 0 | | 0 | | 0 | 0.5 | 0.4 |
| 90% CI | | | | | | 0–2 | 0–1 |
| detected | 1.0 | | 0 | | placed | CPT | flat |

Figure 138: Uni-modal dataset A4.

Sample A3 (Figure 137) is an asymmetric distribution built from three draws, with a broad tail to the left and a sharp fall-off to the right. The interval spacing contains one peak, the low-pass spacing occasionally one which aligns with its interval counterpart. Tests reject the detected features. The low-pass spacing contains one or two flats, half of which pass testing. The interval spacing has no flats. One of the four level sections covers this significant flat. There are one or two changepoints, one matching the endpoint of a level section. The lack of peaks reduces the correct placement rate to zero. In the figure a changepoint and level section ends mark the first draw, perhaps because the standard deviation changes despite the location in the leading tail. There is no obvious flank to the side of the mode, and no unusual effects in either smoothed curve.

A4 (Figure 138) is a single normal variate. The low-pass spacing contains one peak that does not pass testing. The interval spacing is rougher and contains three peaks, one which matches the low-pass peak. The tests accept one interval peak in half the trials. There is one flat in the low-pass spacing in most runs, which tests pass half the time. The flat in the figure is a short segment in the right tail, not in the main mode. The interval spacing contains no flats. Four level sections span the mode, two in the tails and the peak dividing another two. They cover any significant flat. There is one changepoint at a tail, which matches a section endpoint in half the runs. There are no significant peaks to check its placement. Any peak, or subdivision of the sample by an interval, is a false positive, the result of natural variations in the draw.

Sample A5 (Figure 139) has two superimposed normal variates, adding a small wide draw to the main lobe. There are 300 points in the total draw. The figure shows a region of dense spacing between indices 100 and 200, flanked by regions of 75 points with a larger spacing, then the quickly growing tails. The flanks are artefacts of the draw, because the small draw is only 30 points. The low-pass spacing has two flats, one found significant, and the interval spacing one with a wider confidence



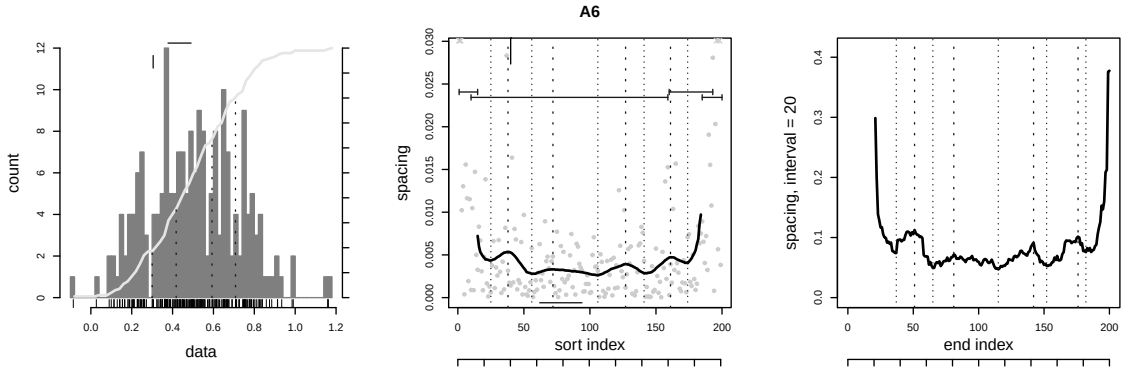

| A5 (uni) | peaks | | flats | | CPT | level | |
|---|---|---|---|---|---|---|---|
| | LP | Diw | LP | Diw | | | |
| mean count | 0 | 0 | 0.8 | 1.2 | 1.3 | 5.0 | |
| 90% CI | | | 0–2 | 0–3 | 0–3 | 4–6 | |
| detected | 0.2 | 0 | 2.1 | 1.3 | | | |
| matching | 0 | | 0.9 | | 0 | 0.6 | 0.8 |
| 90% CI | | | 0–2 | | | 0–2 | 0–2 |
| detected | 0 | | 1.0 | | placed | CPT | flat |

Figure 139: Uni-modal dataset A5.

interval. In the figure these match the central mode and flanks. The significant flats in the two methods match. The five level sections cover the main lobe and the starting and ending tails. They are not as precisely located as the flats, overlapping and short on the right flank. A section covers the significant low-pass flat. There is one changepoint, matching a section endpoint in half the runs. Neither spacing contains a peak.

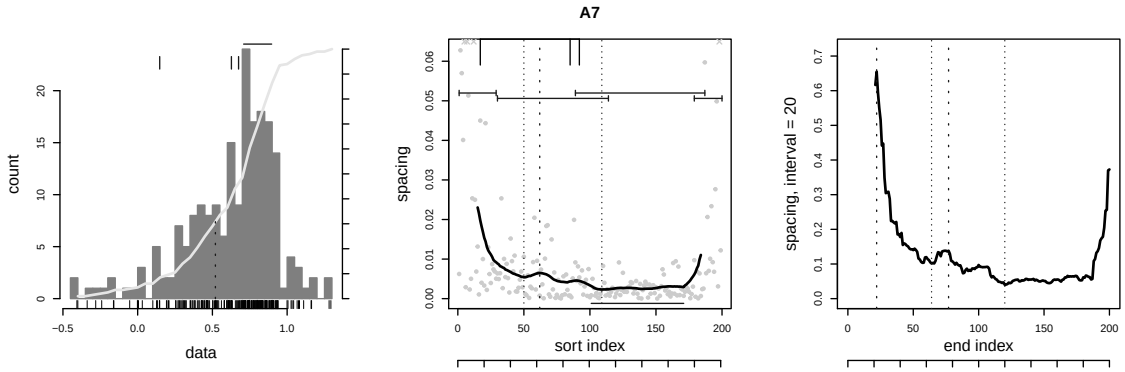
The two draws placed to each side of a larger normal variate in sample A6 (Figure 140) flatten or widen the distribution. In effect this is a bi-modal setup with a third draw added to fill in the gap. Despite a smooth merging of the density, variation in the draws is enough to identify the transitions between the three. There are one or two peaks in the low-pass spacing that do not pass testing, and four in the rougher interval spacing, one that passes in half the trials. The two methods do match detected peaks. The low-pass spacing contains no or one flat, which is accepted as significant half the time that it appears. The three or four level sections do better at covering the mode, with two smaller sections in the tails. A section covers any significant low-pass flat. There is one changepoint which aligns with a section endpoint in half the runs. It does not lie between a peak and flat, however.

Sample A7 (Figure 141) adds a beta variate to a normal to create an asymmetric uni-modal distribution, with a long tail to the left and a sharp transition between the draws at the upper edge of the beta variate. The low-pass spacing has a peak in half the runs that testing rejects. The interval spacing has one or two peaks, one found significant in a third of the trials. The detected features in the two methods align and in the figure identify the transition between the two draws. There is one flat in the low-pass spacing that tests mostly accept as significant, but no flats in the interval spacing. In the figure the flat corresponds to the beta draw, a clearly denser region. The four level sections cover the two draws plus the two tails. All significant flats are covered by a section, so the sections identify the two draws. There is one changepoint on average. Half the time it matches an



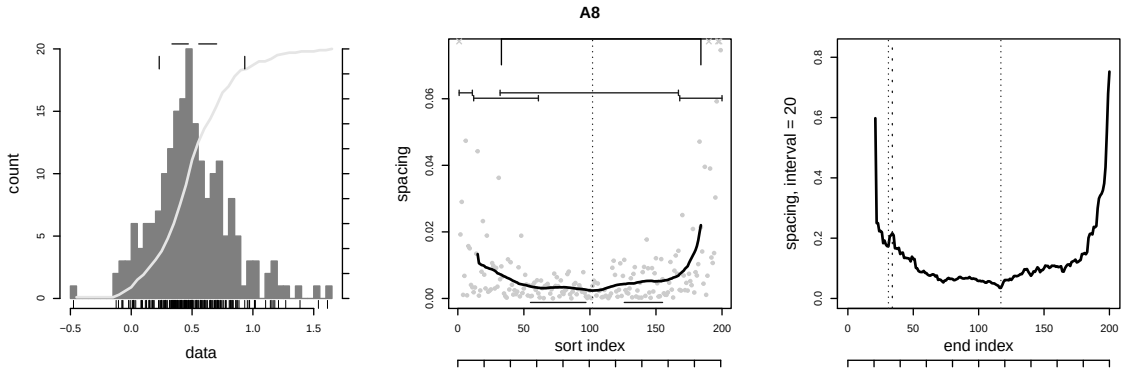

| A6 (uni) | peaks | | flats | | CPT | level | |
|---|---|---|---|---|---|---|---|
| | LP | Diw | LP | Diw | | | |
| mean count | 0.1 | 0.5 | 0.3 | 0 | 0.8 | 3.4 | |
| 90% CI | 0–1 | 0–2 | 0–1 | | 0–2 | 2–5 | |
| detected | 1.5 | 4.2 | 0.5 | 0 | | | |
| matching | 0 | | 0 | | 0 | 0.4 | 0.3 |
| 90% CI | | | | | | 0–1 | 0–1 |
| detected | 1.3 | | 0 | | placed | CPT | flat |

Figure 140: Uni-modal dataset A6.




| A7 (uni) | peaks | | flats | | CPT | level | |
|---|---|---|---|---|---|---|---|
| | LP | Diw | LP | Diw | | | |
| mean count | 0 | 0.3 | 1.1 | 0 | 1.2 | 4.3 | |
| 90% CI | | 0–1 | 0–2 | | 0–2 | 4–5 | |
| detected | 0.4 | 1.4 | 1.3 | 0 | | | |
| matching | 0 | | 0 | | 0.1 | 0.5 | 1.1 |
| 90% CI | | | | | 0–1 | 0–2 | 0–2 |
| detected | 0.3 | | 0 | | placed | CPT | flat |

Figure 141: Uni-modal dataset A7.



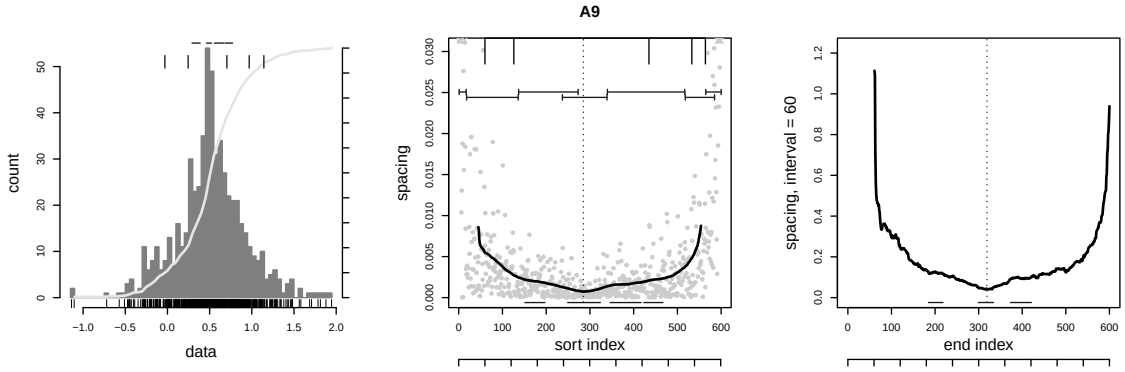

| A8 (uni) | peaks | | flats | | CPT | level | |
|---|---|---|---|---|---|---|---|
| | LP | Diw | LP | Diw | | | |
| mean count | 0 | 0.1 | 0.7 | 0 | 1.5 | 4.1 | |
| 90% CI | | 0-1 | 0–2 | | 0–3 | 3–5 | |
| detected | 0.6 | 1.2 | 1.2 | 0 | | | |
| matching | 0 | | 0 | | 0 | 1.0 | 0.7 |
| 90% CI | | | | | | 0–2 | 0–2 |
| detected | 0.4 | | 0 | | placed | CPT | flat |

Figure 142: Uni-modal dataset A8.

endpoint of a level section. Overall the features correspond to the layout of the two draws, although the example is designed to hide the difference.

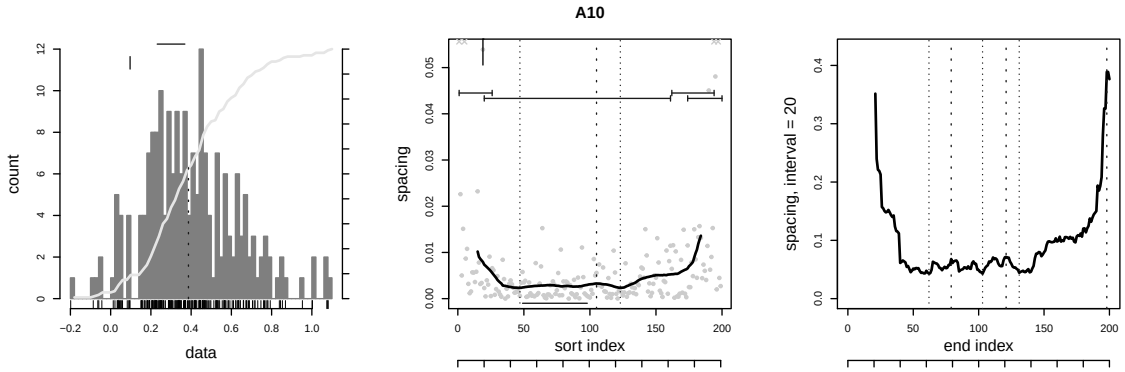
A8 (Figure 142) is another sample that combines two different draws, a normal and Weibull variate, to create a slightly asymmetric distribution. Half the runs generate a peak in the low-pass spacing, although testing rejects it, and one peak in the interval spacing, also rejected. The detected peaks in the two methods mostly align. The low-pass spacing has one flat, passing tests half the time, but the interval spacing has a larger amplitude which curves the floor of the curve and prevents any flats. There are four level sections, one spanning the distribution and three around the tails. They cover any flat found. There are one or two changepoints, one which aligns with an endpoint of a level section. Without significant peaks its placement cannot be checked. These results say the spacing does not detect the individual draws.

Sample A9 (Figure 143) superimposes three normal variates of decreasing width. It has 600 points in the total draw so that the third, smallest, and very narrow draw has 30 points. The spacing contains five different sections corresponding to the left and right sides of the two variates, plus one for the third draw, which can be seen as regions of different variation. The low-pass spacing does not reach a stable value in the middle sections. But there are flats in each, although tests reject them as significant. The interval spacing also contains three flats which are accepted, despite being small. Two of the three detected flats match between the two methods. The high matching rate of significant flats despite having only a few in the low-pass spacing means the low-pass features must be long and cover two interval flats. The level sections are more fragmented, with seven or eight appearing, repeatably. They do cover any significant flat in the low-pass spacing. Changepoints are also able to mark the boundaries between the three draws, with three or four appearing. They do not generally line up with a section endpoint, however. There are no peaks in either spacing.



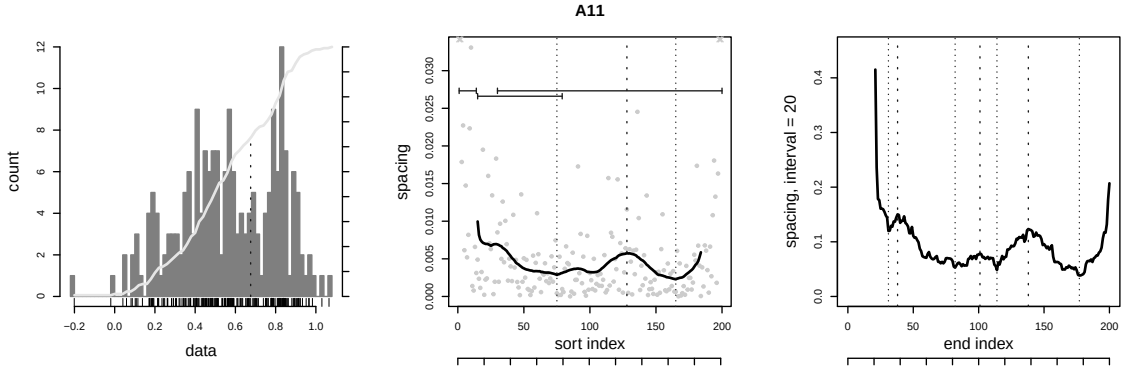

| A9 (uni) | peaks | | flats | | CPT | level | |
|---|---|---|---|---|---|---|---|
| | LP | Diw | LP | Diw | | | |
| mean count | 0 | 0.1 | 0.2 | 3.1 | 3.5 | 7.4 | |
| 90% CI | | 0–1 | 0–1 | 1–5 | 2–6 | 6.9–8 | |
| detected | 0 | 0.1 | 3.3 | 3.3 | | | |
| matching | 0 | | 1.9 | | 0 | 0.6 | 0.2 |
| 90% CI | | | 1–3 | | | 0–2 | 0–1 |
| detected | 0 | | 2.1 | | placed | CPT | flat |

Figure 143: Uni-modal dataset A9.




| A10 (uni) | peaks | | flats | | CPT | level | |
|---|---|---|---|---|---|---|---|
| | LP | Diw | LP | Diw | | | |
| mean count | 0 | 0.2 | 0.5 | 0 | 1.4 | 3.9 | |
| 90% CI | | 0–1 | 0–1 | | 0–3 | 3–5 | |
| detected | 0.7 | 1.3 | 1.2 | 0 | | | |
| matching | 0 | | 0 | | 0 | 0.8 | 0.5 |
| 90% CI | | | | | | 0–2 | 0–1 |
| detected | 0.5 | | 0 | | placed | CPT | flat |

Figure 144: Uni-modal dataset A10.



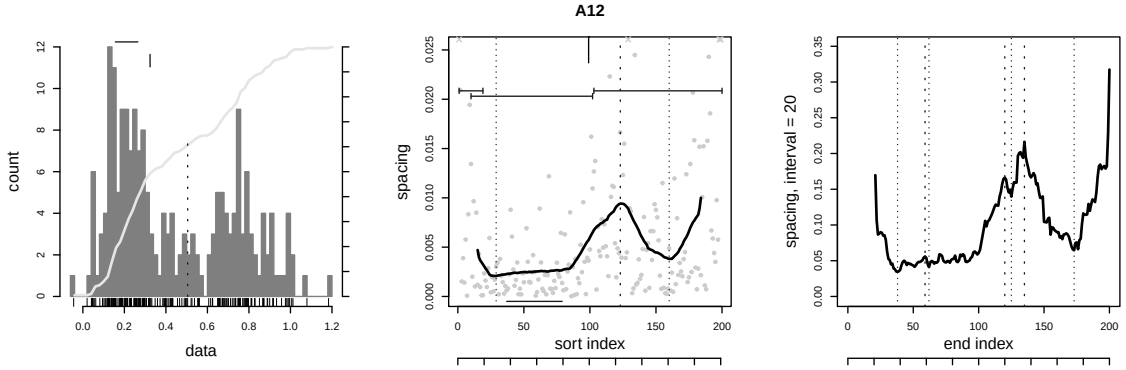

| A11 (abi) | peaks | | flats | | CPT | level | |
|---|---|---|---|---|---|---|---|
| | LP | Diw | LP | Diw | | | |
| mean count | 0.3 | 0.7 | 0.3 | 0 | 0.8 | 2.8 | |
| 90% CI | 0–1 | 0–3 | 0–1 | | 0–3 | 2–4 | |
| detected | 1.6 | 3.9 | 0.4 | 0 | | | |
| matching | 0.1 | | 0 | | 0.1 | 0.2 | 0.3 |
| 90% CI | 0–1 | | | | | 0–1 | 0–1 |
| detected | 1.4 | | 0 | | placed | CPT | flat |

Figure 145: Asymmetric bi-modal dataset A11.

Combining a normal and a gamma variate makes sample A10 (Figure 144) an asymmetric unimodal draw, with a larger tail to the right. The gamma density is larger than the normal for $x > 0.58$, which corresponds to index 160. In the figure this is about the position of the step increase in the spacing to the right. The low-pass spacing contains one peak at the transition between the draws, although no test accepts it as significant. The interval spacing contains one or sometimes two peaks, most rejected, but which generally align with the low-pass peak. There is one flat in the low-pass spacing passing tests in half the runs that identifies the normal draw in the figure. There is no flat in the interval spacing. The four level sections contain one spanning most of the total draw, with small sections in the tails; the sections do not separate the two variates. They cover any significant flat. There are one or two changepoints in the tails, half of which match a section endpoint.

Sample A11 (Figure 145) is an asymmetric bi-modal distribution built from two normals with different standard deviations and unequal draw sizes. The low-pass spacing contains one or two peaks which are occasionally significant. The interval spacing is rougher, with four peaks, one of which is significant, although the count's confidence interval is broad. The low rate of acceptance prevents matching between significant peaks, but the raw detections do align. In less than half the trials do we find a flat in the low-pass spacing, but it is accepted when it exists. There is one changepoint that does not align to a level section endpoint. Three level sections cover any flat. The spacing does not resolve this setup in general, although the peak in the figure does.

A12 (Figure 146) contains two asymmetric modes. A normal draw forms the first, and a combination of a normal and beta variate forms the second. The low-pass spacing contains one or two peaks, one of which passes the acceptance test; in the figure it falls in the inter-mode transition. The interval spacing has three or four peaks, one significant in half the trials. The significant peaks do not match, but the detected peak does. This implies the location of the selected peaks are not stable,



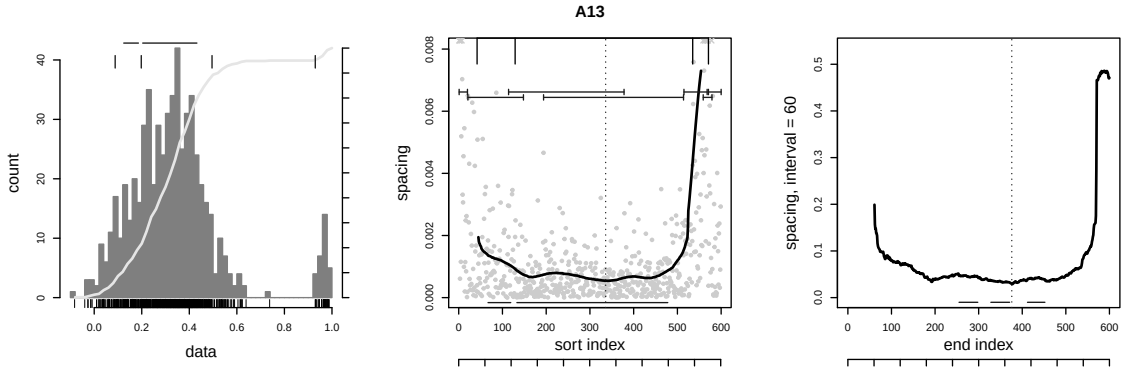

| A12 (abi) | peaks | | flats | | CPT | level | |
|---|---|---|---|---|---|---|---|
| | LP | Diw | LP | Diw | | | |
| mean count | 0.7 | 0.5 | 0.3 | 0 | 1.3 | 4.0 | |
| 90% CI | 0–1 | 0–2 | 0–1 | | 0–3 | 3–5 | |
| detected | 1.4 | 3.4 | 0.5 | 0 | | | |
| matching | 0.1 | | 0 | | 0.3 | 0.6 | 0.3 |
| 90% CI | 0–1 | | | | 0–1 | 0–2 | 0–1 |
| detected | 1.1 | | 0 | | placed | CPT | flat |

Figure 146: Asymmetric bi-modal dataset A12.

most likely do to noise in the interval spacing around the maximum. In half the trials there is a flat in the low-pass spacing that is usually accepted as significant, but there are no flats in the interval spacing, although one seems to exist in the figure, implying different detector parameters may be needed. There is one changepoint near the transition and its placement rate is limited by the number of flats. The four level sections span the two modes and the two tails. Half the changepoints match an endpoint. A section covers any flat judged significant.

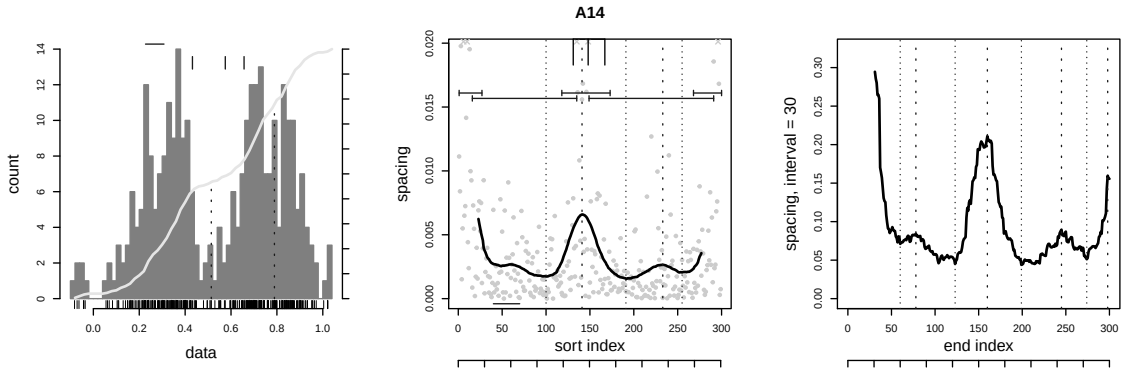
A13 (Figure 147) is a mix of two normal variates, one wide and the other narrow, with a large separating gap. The draw sizes are unbalanced in a 19:1 ratio, so the total sample has 600 points to give 30 to the second draw. Even with so many points the small narrow variate falls outside the low-pass kernel and is ignored, and there is no maximum. The interval spacing may occasionally contain a peak, but much more often there is not enough of a trailing tail to create a two-sided local maximum. The main lobe does support flats, two in the low-pass spacing and four in the interval spacing. Tests accept most of the detected flats, and they align. The matching rate is higher than the number of low-pass flats so the interval spacing flats are shorter and subsets of the low-pass. The spacing has four changepoints, two that bound the gap between the variates and two in the leading tail. The level section result is noisy but stable, with several covering the the large variate and one or two corresponding to the second draw. The splitting of level sections means that not all cover a significant flat. In summary, the second draw is too small and set too far to the side to be found by peaks or flats, but the changepoints and level sections do respond to it.

Sample A14 (Figure 148) is an asymmetric bi-modal distribution built from four draws of unequal size and different normal parameters. The right lobe comes from one variate, the left from three, two narrow normals placed symmetrically about the mean of the third. The two narrow draws are small, leading to a total draw of 300 points. There is one maximum at the inter-mode transition in



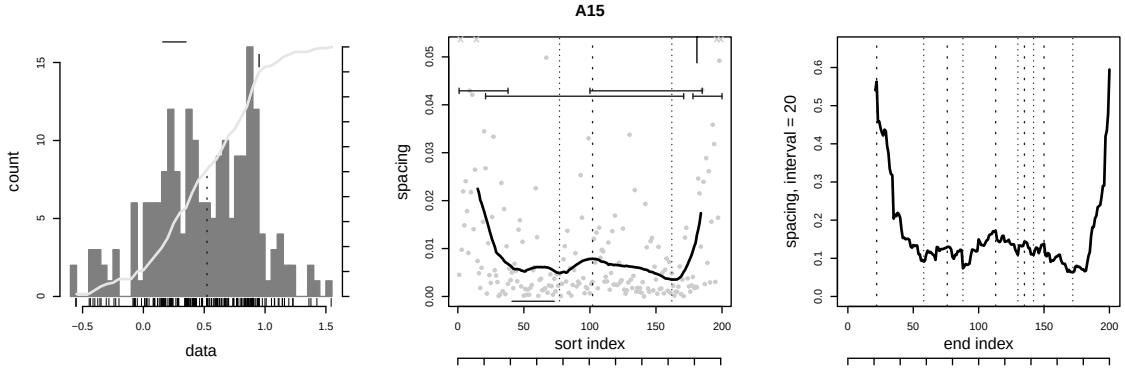

| A13 (abi) | peaks | | flats | | CPT | level | |
|---|---|---|---|---|---|---|---|
| | LP | Diw | LP | Diw | | | |
| mean count | 0 | 0.1 | 1.2 | 3.5 | 3.6 | 6.7 | |
| 90% CI | | 0–1 | 1–2 | 2–5 | 2–6 | 6–8 | |
| detected | 0 | 0.2 | 1.7 | 3.9 | | | |
| matching | 0 | | 3.2 | | 0 | 1.1 | 1.0 |
| 90% CI | | | 1–5 | | | 0–3 | 0–2 |
| detected | 0 | | 3.5 | | placed | CPT | flat |

Figure 147: Bi-modal dataset A13.




| A14 (abi) | peaks | | flats | | CPT | level | |
|---|---|---|---|---|---|---|---|
| | LP | Diw | LP | Diw | | | |
| mean count | 1.0 | 0.6 | 1.2 | 0 | 2.7 | 5.3 | |
| 90% CI | 1–1 | 0–1 | 0–2 | | 1–4 | 4–6 | |
| detected | 1.1 | 1.7 | 1.5 | 0 | | | |
| matching | 0.4 | | 0 | | 1.3 | 1.1 | 1.2 |
| 90% CI | 0–1 | | | | 0–3 | 0–2 | 0–2 |
| detected | 0.8 | | 0 | | placed | CPT | flat |

Figure 148: Asymmetric bi-modal dataset A14.



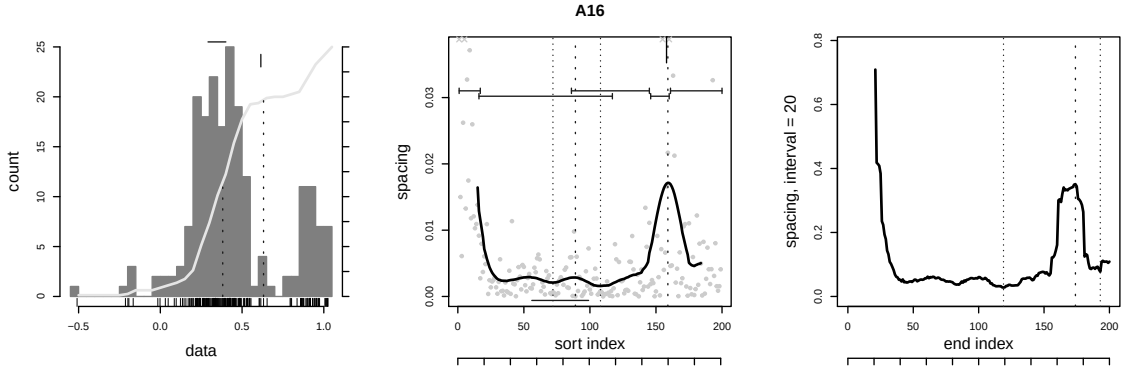

| A15 (abi) | peaks | | flats | | CPT | level | |
|---|---|---|---|---|---|---|---|
| | LP | Diw | LP | Diw | | | |
| mean count | 0 | 0.4 | 0.4 | 0 | 1.2 | 3.6 | |
| 90% CI | | 0–2 | 0–1 | | 0–2 | 3–5 | |
| detected | 1.5 | 2.8 | 0.7 | 0 | | | |
| matching | 0 | | 0 | | 0 | 0.8 | 0.4 |
| 90% CI | | | | | | 0–2 | 0–1 |
| detected | 1.2 | | 0 | | placed | CPT | flat |

Figure 149: Asymmetric bi-modal dataset A15.

the low-pass spacing, accepted as significant by the peak tests. The confidence interval shows this is stable. There are two peaks in the interval spacing, one of which is significant in half the trial runs. This matches the low-pass position, so the extra detected peak must be spurious. Flats appear in the low-pass spacing in one or two of the modes, and the tests accept them as significant. The interval spacing is too rough for flats. There are three changepoints on average, with the placement rate limited by the flat. In the figure all three appear within the transition regime. The spacing has five level sections spanning the modes, the tails, and the transition. There are enough sections clustered around the peak that one changepoint matches an endpoint. Level sections cover all significant flats. The spacing resolves the large change in modality but does not break apart the mixture in the left mode.

Sample A15 (Figure 149) is an unequal bi-modal distribution built from four draws. The basis of each mode is a normal variate with the same parameters and draw size. The sample adds a gamma draw to the left mode and a beta to the right to change the tails. The low-pass spacing has one or two peaks that match one of three maxima found in the interval spacing. None of the low-pass features are significant, but one of the interval peaks is in half the runs. One mode contains a low-pass flat that is usually considered significant. There is one changepoint on average, located in the figure at the trailing tail. Four level sections appear, with the changepoint matching one of their endpoints. The figure shows one section covering the second mode, but another spans both. The peak in the inter-mode transition is not strong enough to break the longer span. Two sections cover the tails. Any significant flat is covered by a level section. The spacing does not clearly establish this sample as a bi-modal distribution, nor does it resolve the change in either tail.

A16 (Figure 150) is an unequal bi-modal distribution from two normal variates with a third normal carefully placed to create a large initial tail in the left mode. The setup is easily resolved, but not



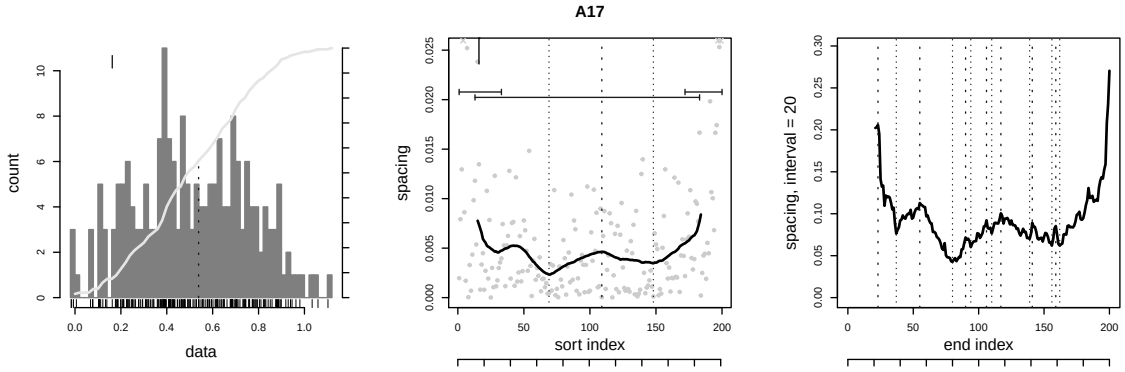

| A16 (abi) | peaks | | flats | | CPT | level | |
|---|---|---|---|---|---|---|---|
| | LP | Diw | LP | Diw | | | |
| mean count | 1.0 | 0.6 | 1.2 | 0 | 2.2 | 4.9 | |
| 90% CI | 1–1 | 0–1 | 1–2 | | 1–3 | 4–6 | |
| detected | 1.2 | 1.7 | 1.2 | 0 | | | |
| matching | 0.5 | | 0 | | 0.9 | 1.7 | 1.1 |
| 90% CI | 0–1 | | | | 0–2 | 0–3 | 1–2 |
| detected | 1.1 | | 0 | | placed | CPT | flat |

Figure 150: Asymmetric bi-modal dataset A16.

the change in the tail. There is one stable peak in the low-pass spacing that is significant, and two can appear in the interval spacing, one of which passes its tests in half the trials. The position of the peaks align, so the second in the interval spacing is spurious. There is one flat in the low-pass spacing that is significant, and none in the interval spacing. The figure suggests a relaxed ripple specification might allow for more flats. The flat is covered by one of the five level sections, which split the main mode in two. In the figure sections break at a small noise peak in the main mode near index 90, and between the two modes. They do not correspond to the extension of the tail, visible in the knee of the cumulative density near data value 0.1. There are two changepoints, one placed in the transition between the peak and flat. One matches an endpoint of a level section.

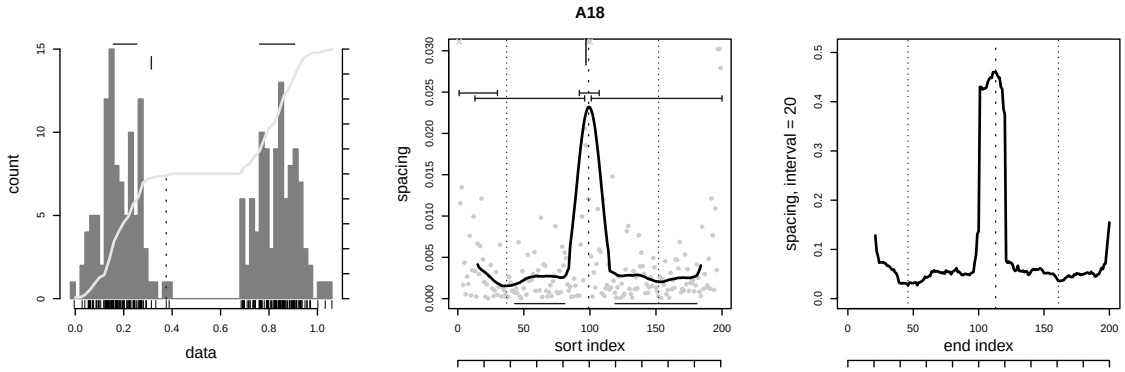
Sample A17 (Figure 151) is a balanced draw from two normals with the same standard deviation. The separation is 2.8 times the standard deviation, so we expect a marginal resolution of the setup. The measurements bear this out. There are one or two peaks in the low-pass spacing, one of which passes the tests in half the trials. The interval spacing finds more, five, but also only accepts one half the time. The detected peaks match, but those found significant do not. The low-pass spacing occasionally contains a flat, found significant, but the interval spacing does not. Three level sections span the total draw and appear in the two tails. They do not resolve the modes. The large span covers any flat that appears. There is a changepoint in half the runs that does not match a section endpoint. In the figure the changepoint lies in the initial tail and does not mark the transition between modes.

Sample A18 (Figure 152) draws from two identical normal variates separated by 7.6 standard deviations. This is an easy test case. The low-pass spacing repeatably contains one peak found significant and a flat in each mode, also both significant. The interval spacing has one or two peaks with one passing tests in half the runs, and no flats. The peaks in both methods match. A relaxed ripple specification would detect the flats that seem to exist in the figure. There are one or two



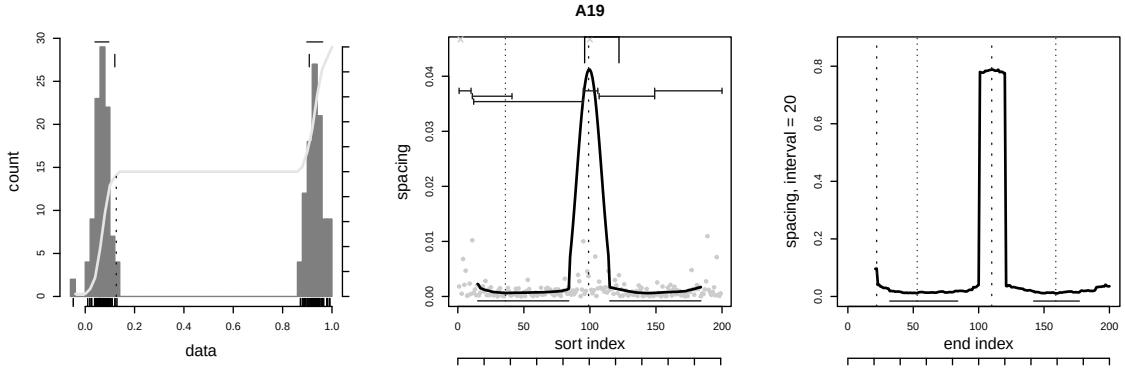

| A17 (bi) | peaks | | flats | | CPT | level | |
|---|---|---|---|---|---|---|---|
| | LP | Diw | LP | Diw | | | |
| mean count | 0.4 | 0.5 | 0.2 | 0 | 0.6 | 2.9 | |
| 90% CI | 0–1 | 0–2 | 0–1 | | 0–2 | 2–4 | |
| detected | 1.6 | 5.2 | 0.2 | 0 | | | |
| matching | 0.1 | | 0 | | 0 | 0.2 | 0.2 |
| 90% CI | 0–1 | | | | | 0–1 | 0–1 |
| detected | 1.4 | | 0 | | placed | CPT | flat |

Figure 151: Bi-modal dataset A17.




| A18 (bi) | peaks | | flats | | CPT | level | |
|---|---|---|---|---|---|---|---|
| | LP | Diw | LP | Diw | | | |
| mean count | 1.0 | 0.5 | 1.9 | 0 | 1.6 | 4.9 | |
| 90% CI | 1–1 | 0–1 | 1–2 | | 1–3 | 4–6 | |
| detected | 1.0 | 1.4 | 1.9 | 0 | | | |
| matching | 0.4 | | 0 | | 1.5 | 1.1 | 1.9 |
| 90% CI | 0–1 | | | | 1–3 | 0–3 | 1–2 |
| detected | 0.9 | | 0 | | placed | CPT | flat |

Figure 152: Bi-modal dataset A18.



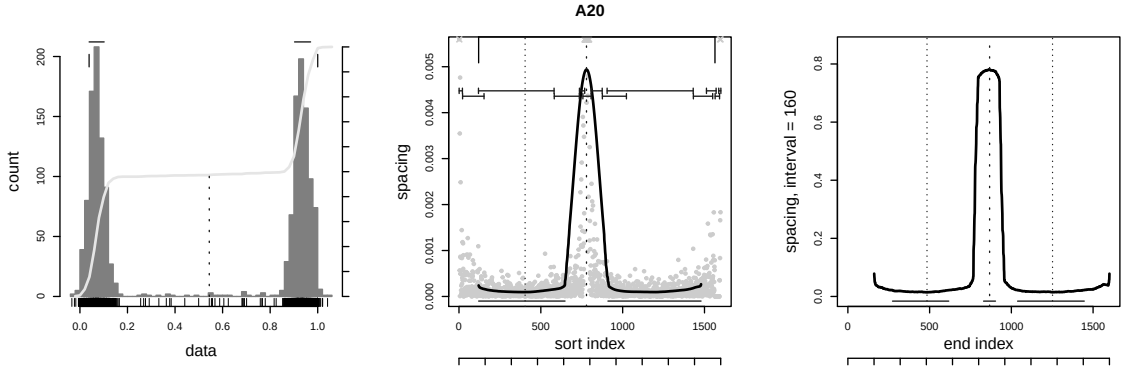

| A19 (bi) | peaks | | flats | | CPT | level | |
|---|---|---|---|---|---|---|---|
| | LP | Diw | LP | Diw | | | |
| mean count | 1.0 | 0.3 | 2.0 | 1.3 | 1.3 | 5.3 | |
| 90% CI | 1–1 | 0–1 | 2–2 | 0–2 | 1–2 | 4–7 | |
| detected | 1.0 | 1.1 | 2.0 | 1.7 | | | |
| matching | 0.3 | | 1.1 | | 1.2 | 1.5 | 1.9 |
| 90% CI | 0–1 | | 0–2 | | 1–2 | 0.9–3 | 1–2 |
| detected | 0.8 | | 1.5 | | placed | CPT | flat |

Figure 153: Bi-modal dataset A19.

changepoints, placed correctly and bounding the inter-mode transition. There are five flats, one per mode that covers the corresponding flat, one in the transition, and two for the tails. One changepoint matches an endpoint of a level section.

Even more clearly separated, sample A19 (Figure 153) has two identical normal draws separated by 21.6 standard deviations. Like the chirp, the interval peak's height equals the separation and the length of the flats to each side is the size of the draws. These draws are tighter than in A18 and the interval spacing is smoother, supporting flats. The results are close to A18, with a peak and two flats, all significant, repeatably appearing in the low-pass spacing. A peak and two flats also appear in the interval spacing, although the acceptance rate is lower, with the peak found significant in a third of the runs but with most of the flats passing. Both peaks and flats agree, with the matching rate limited by the number of features in the interval spacing. The five level sections correspond to the two modes, the transition between, and the two tails. They cover almost all of the significant flats and match the one or two changepoints that bound the transition region.

A20 (Figure 154) is a symmetric bi-modal distribution, starting from the A19 setup and adding a very small beta draw to each mode to add some values in the gap between them. There are 1600 points in the total draw, to allocate 32 points to the beta variates. As with A19, the low-pass spacing contains one peak and two flats, all significant and all stable over the trials. The interval spacing has one significant peak and three flats, the two that mark the modes significant and a third in the transition, which because of the beta variates is wide enough — has enough points — to pass the length requirement. Both features agree. The level section test is very noisy but fairly stable, finding thirteen features with many clustering around the transition gap and in the tails. Such splitting reduces the number that overlap significant flats. There are five changepoints, with a large confidence interval. Two bound the transition region and are placed correctly. Two also match section endpoints,



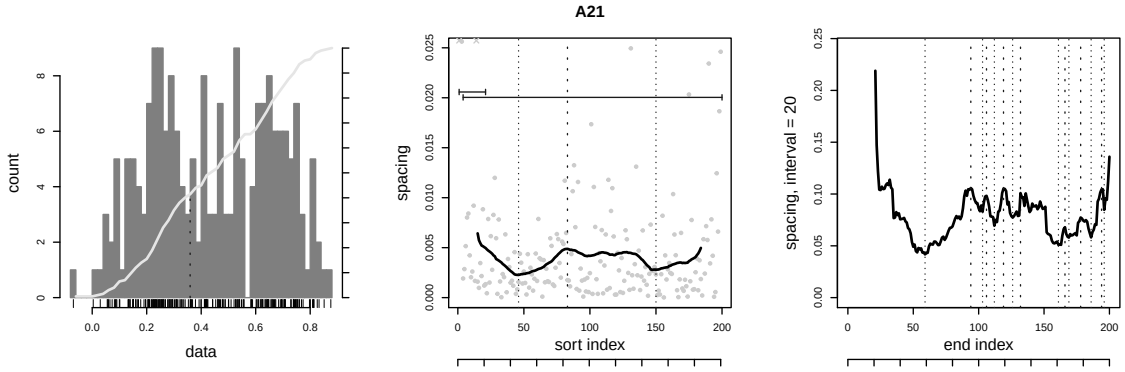

| A20 (bi) | peaks | | flats | | CPT | level | |
|---|---|---|---|---|---|---|---|
| | LP | Diw | LP | Diw | | | |
| mean count | 1.0 | 1.0 | 2.0 | 2.0 | 4.6 | 12.8 | |
| 90% CI | 1–1 | 1–1 | 2–2 | 2–2 | 2–7 | 11–14 | |
| detected | 1.0 | 1.0 | 2.0 | 3.0 | | | |
| matching | 0 | | 2.0 | | 2.2 | 0.8 | 1.8 |
| 90% CI | | | 2–2 | | 1–4 | 0–2 | 1–2 |
| detected | 0 | | 2.0 | | placed | CPT | flat |

Figure 154: Bi-modal dataset A20.

which is a low rate given the large number of endpoints available for matching.

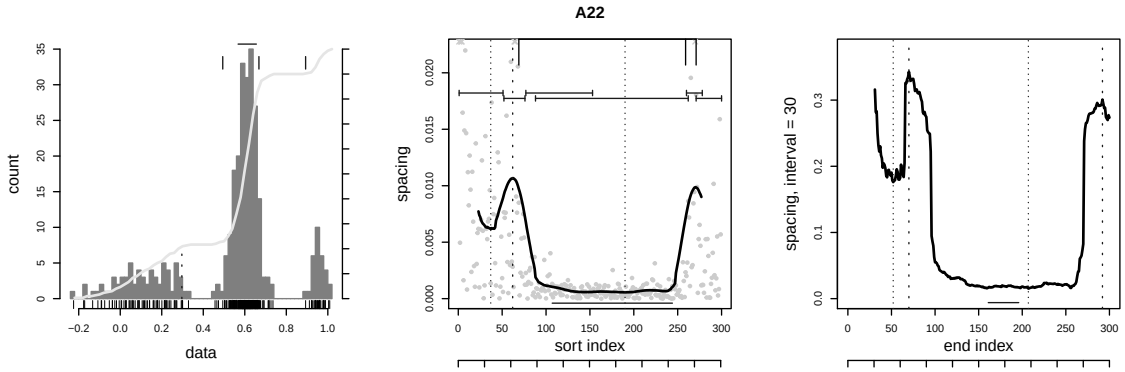
Sample A21 (Figure 155) is a tri-modal distribution with unequal draws and different normal parameters. The first two draws are close but the second is tighter, while the second and third draws are easier to distinguish. There are two peaks in the low-pass spacing, although only one is significant, and that in half the runs. The interval spacing is rougher as it responds to non-uniformities in the draw without much smoothing. It finds six peaks, and tests accept one as significant. Both spacings contain different levels, with higher average values in the middle, and these correspond to the standard deviations of the three draws. The transition at the edges is too gentle to raise a large peak, however. The transition between the first two draws raises a small bump at index 83, which is the significant peak in either spacing, but there is no local maximum at the other side, around index 140. The different levels are too rough to support flats, especially in the interval spacing, and the tails also round off the first and third regions, so we cannot compare their average value to see the offset. In short, the spacing does not split this sample. There is one changepoint that does not align to an endpoint of the three level sections. Neither of these features can define the three modes.

Three unequal normal draws make up the A22 sample (Figure 156). There are 300 points in the total draw, so that the smallest variate has 30 points. The three modes have a large gap between them, and distinguishing them should be easy. However, as we have seen in other setups, such as W14 and A13, having a small draw at the edge of the data means that it gets ignored because the filter kernel partially overlaps it. That happens here. The low-pass spacing finds the peak between the first two variates in all trials, and the tests accept it as significant. The interval spacing finds both peaks, but tests only pass one. The detected peaks match but the significant peaks only align half the time, which means significance is split 50-50 between the interval spacing features. There is a flat in both methods that marks the middle variate. Three changepoints border either transition



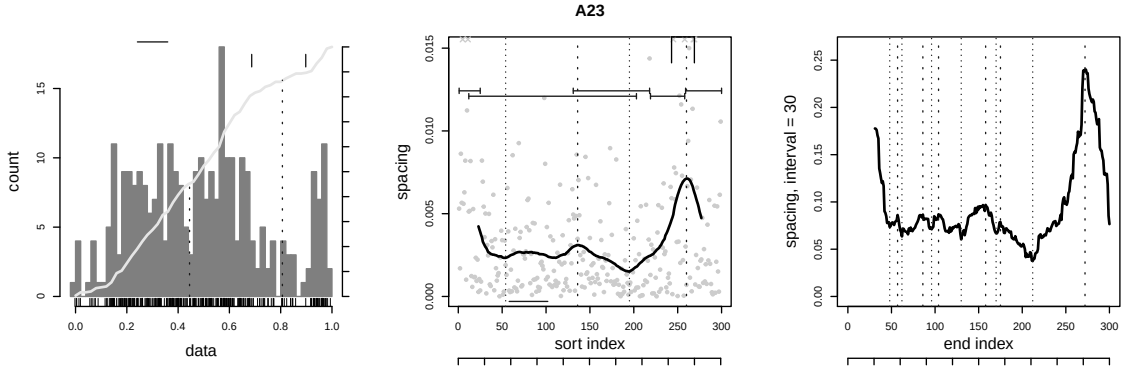

| A21 (tri) | peaks | | flats | | CPT | level | |
|---|---|---|---|---|---|---|---|
| | LP | Diw | LP | Diw | | | |
| mean count | 0.5 | 0.9 | 0.1 | 0 | 0.8 | 2.7 | |
| 90% CI | 0–1 | 0–3.1 | 0–1 | | 0–2 | 2–4 | |
| detected | 1.8 | 5.8 | 0.1 | 0 | | | |
| matching | 0.1 | | 0 | | 0 | 0.2 | 0.1 |
| 90% CI | 0–1 | | | | | 0–1 | 0–1 |
| detected | 1.6 | | 0 | | placed | CPT | flat |

Figure 155: Tri-modal dataset A21.




| A22 (tri) | peaks | | flats | | CPT | level | |
|---|---|---|---|---|---|---|---|
| | LP | Diw | LP | Diw | | | |
| mean count | 1.0 | 0.8 | 1.0 | 0.8 | 2.8 | 7.0 | |
| 90% CI | 1–1 | 0–2 | 1–1 | 0–2 | 2–4 | 6–8 | |
| detected | 1.0 | 1.9 | 1.0 | 0.8 | | | |
| matching | 0.6 | | 0 | | 1.3 | 1.5 | 1.0 |
| 90% CI | 0–1 | | | | 1–2 | 0–3 | 1–1 |
| detected | 1.0 | | 0 | | placed | CPT | flat |

Figure 156: Tri-modal dataset A22.



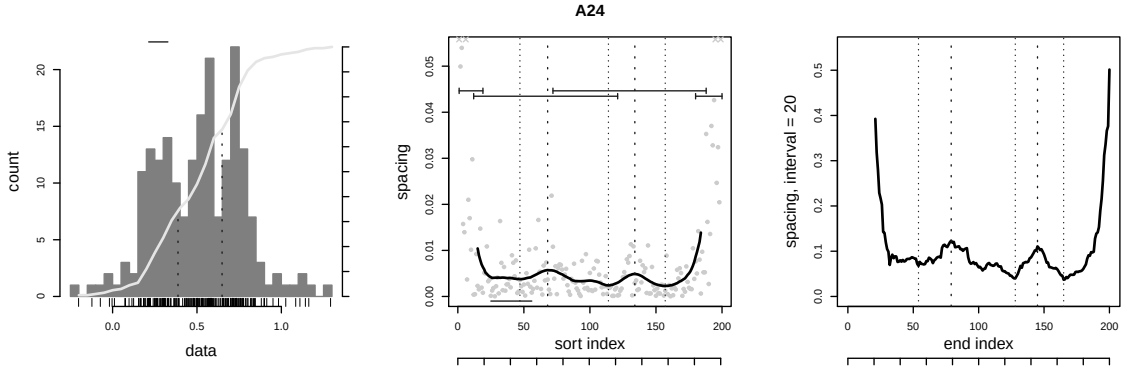

| A23 (tri) | peaks | | flats | | CPT | level | |
|---|---|---|---|---|---|---|---|
| | LP | Diw | LP | Diw | | | |
| mean count | 1.0 | 1.3 | 0.9 | 0 | 1.9 | 4.4 | |
| 90% CI | 0–1 | 0–3 | 0–2 | | 1–3 | 3–6 | |
| detected | 2.1 | 5.0 | 1.7 | 0 | | | |
| matching | 0.7 | | 0 | | 0.7 | 1.1 | 0.8 |
| 90% CI | 0–1 | | | | 0–2 | 0–2 | 0–2 |
| detected | 1.7 | | 0 | | placed | CPT | flat |

Figure 157: Tri-modal dataset A23.

region, with one falling between a peak and flat. In the figure level sections do not distinguish the first and third draw from the tails; the seven sections identify the three modes with some splits and the two transitions. The large number of endpoints helps the matching to changepoints, so that one or two align per trial. Level sections overlap all significant flats.

A23 (Figure 157) is a tri-modal distribution formed from two identical normal draws separated by 2.86 standard deviations, and a small, narrow normal variate. There are 300 points in total drawn in each run. The gap between the second and third draws is small enough that the tail of the second extends past the third. This is enough to form a clear peak in the low-pass spacing between them, with a second, weaker peak appearing in the transition between the first two draws. The tests accept one of the two as significant. The interval spacing contains five peaks, one considered significant. Both detected and significant peaks match, the second partially. There are two flats in the low-pass spacing, one significant; in the figure it lies in the first mode, because the other transition creates a slope across the second. The interval spacing is too rough to support any. Two changepoints border the transition between the second and third modes. There are four or five level sections that split up the modes, with a moderate spread. One changepoint on average matches a section endpoint. Most significant flats are covered by a level section.

The A24 sample (Figure 158) is a tri-modal distribution, with three narrow normal variates atop a broad background, similar to the claws of W10 and W12. The three modes are separated by 3.16 standard deviations and are centered atop the background. Both transitions are found as peaks in the low-pass spacing, although tests reject both. The interval spacing also contains two peaks, one which is occasionally found significant. There is good but not perfect matching of the detected features. The peaks are small, so smaller filter widths or intervals would make them more prominent, although that would not change the acceptance rate and few would be significant. The low-pass spacing contains




| A24 (tri) | peaks | | flats | | CPT | level | |
|---|---|---|---|---|---|---|---|
| | LP | Diw | LP | Diw | | | |
| mean count | 0 | 0.2 | 0.5 | 0 | 0.9 | 3.6 | |
| 90% CI | | 0–1 | 0–1 | | 0–2 | 3–5 | |
| detected | 1.7 | 2.2 | 0.9 | 0 | | | |
| matching | 0 | | 0 | | 0 | 0.6 | 0.4 |
| 90% CI | | | | | | 0–2 | 0–1 |
| detected | 1.3 | | 0 | | placed | CPT | flat |

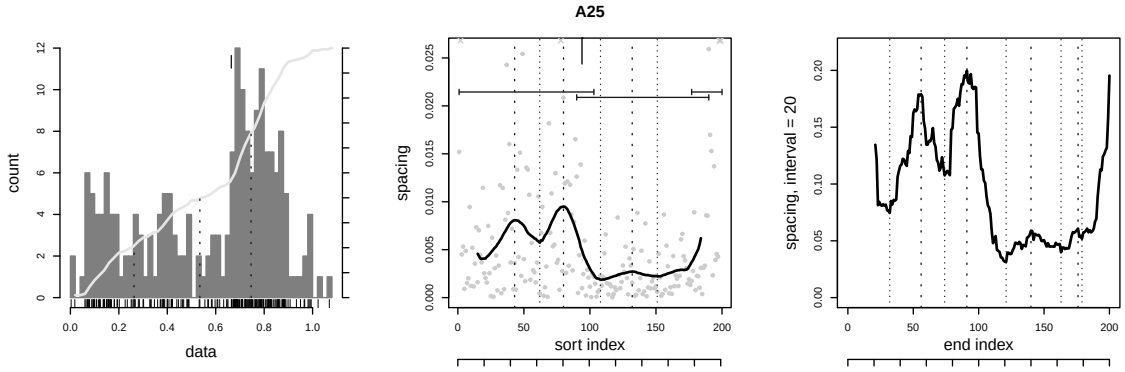
Figure 158: Tri-modal dataset A24.

one flat, significant in half the trials; because the three draws have 30 points each and this is the smallest length of a flat, the tests may ignore some flats. Yet there are three or four level sections, which, after subtracting two for either tail, means that the sections cover more than one mode. They do cover most significant flats. There is one changepoint on average that usually aligns with a section endpoint.

A25 (Figure 159) is a tri-modal distribution built from three unequal normal variates. Two are small, narrow draws separated by 2.91 standard deviations, the third is larger and wider and sits to the side. The figure shows two moderate peaks at the transitions between the draws, plus a minor peak within the broad draw. The depicted level sections do not distinguish the two small draws, but do separate them from the larger. A changepoint marks the transition to the broad variate. The low-pass spacing contains two peaks, one of which tests significant. The interval spacing has five, but also passes one. These significant peaks tend to align. There is a low-pass flat in half the runs that is always found significant, but none exist in the interval spacing. The spacing has two changepoints, the existence of a flat determining the correct placement rate. One of them lines up with an endpoint of one of the four level sections that are found, so one section must break at an inter-mode transition. A section covers any flat.

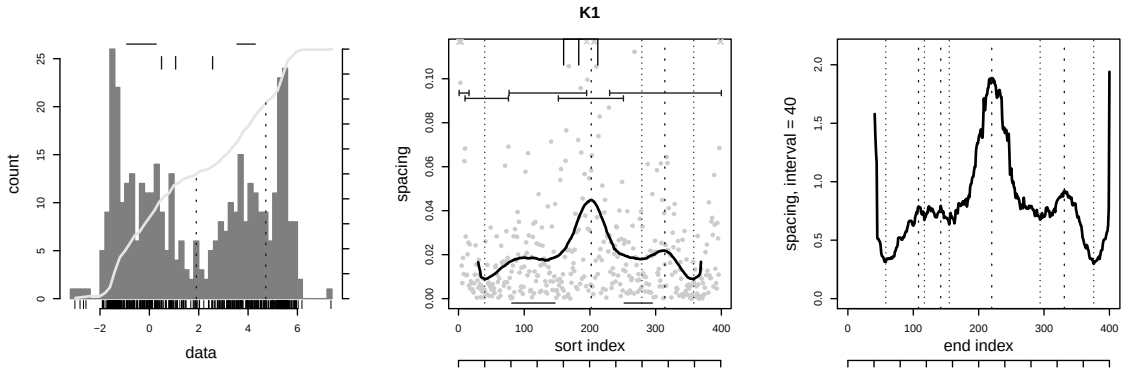
### K Samples

The analysis of the K1 sample (Figure 160) resolves some of the four draws. In the figure the central gap produces a large spacing, while the side draws are closer to a step in the variation which might either produce a local maximum or blend smoothly, as seen to the right and left sides, respectively. Flats are found in the two large draws, but neither spacing reaches a stable level in the smaller. The level sections do pick up the four variates, plus the central gap and two tails, while



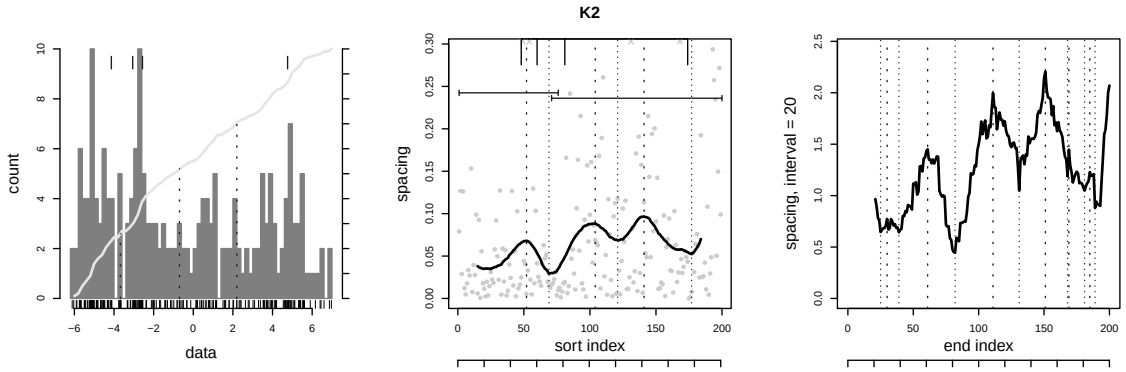

| A25 (tri) | peaks | | flats | | CPT | level | |
|---|---|---|---|---|---|---|---|
| | LP | Diw | LP | Diw | | | |
| mean count | 1.1 | 1.1 | 0.6 | 0 | 1.9 | 4.1 | |
| 90% CI | 1–2 | 0–4 | 0–1 | | 1–3 | 3–6 | |
| detected | 1.9 | 4.5 | 0.6 | 0 | | | |
| matching | 0.6 | | 0 | | 0.6 | 0.8 | 0.6 |
| 90% CI | 0–1 | | | | 0–2 | 0–1 | 0–1 |
| detected | 1.6 | | 0 | | placed | CPT | flat |

Figure 159: Tri-modal dataset A25.




| K1 (quad) | peaks | | flats | | CPT | level | |
|---|---|---|---|---|---|---|---|
| | LP | Diw | LP | Diw | | | |
| mean count | 1.7 | 1.2 | 0.3 | 0 | 3.2 | 5.9 | |
| 90% CI | 1–3 | 0–3 | 0–1 | | 1–5 | 5–7 | |
| detected | 2.2 | 3.7 | 2.0 | 0 | | | |
| matching | 0.3 | | 0 | | 1.7 | 0.6 | 0.2 |
| 90% CI | 0–1 | | | | 0–3 | 0–2 | 0–1 |
| detected | 0.9 | | 0 | | placed | CPT | flat |

Figure 160: Quad-modal dataset K1.



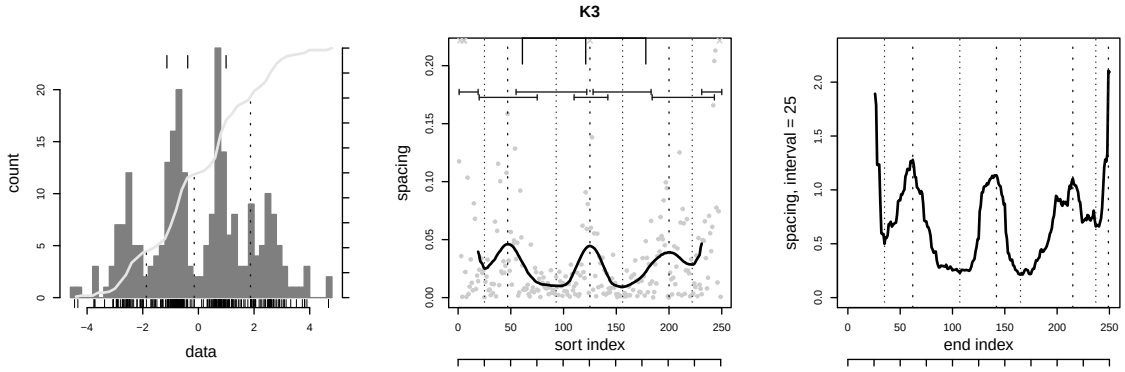

| K2 (quad) | peaks | | flats | | CPT | level | |
|---|---|---|---|---|---|---|---|
| | LP | Diw | LP | Diw | | | |
| mean count | 1.7 | 1.0 | 0 | 0 | 2.4 | 3.1 | |
| 90% CI | 1–3 | 0–3 | | | 1–5 | 2–5 | |
| detected | 2.8 | 5.1 | 0 | 0 | | | |
| matching | 0.5 | | 0 | | 0 | 0.7 | 0 |
| 90% CI | 0–2 | | | | | 0–2 | |
| detected | 2.3 | | 0 | | placed | CPT | flat |

Figure 161: Quad-modal dataset K2.

changepoints bound the gap. In the analysis the interval spacing has four peaks, with tests passing one as significant. The low-pass spacing finds two, most accepted. The two methods do not match well, as only one of the detected peaks and sometimes a significant one aligns. There are two flats in the low-pass spacing, but only one occasionally passes testing. The six level sections are stable, covering most but not all the significant flats. There are three changepoints, two flanking the gap and correctly placed between the maximum and a flat. Their position matches a section endpoint only once in two runs, however.

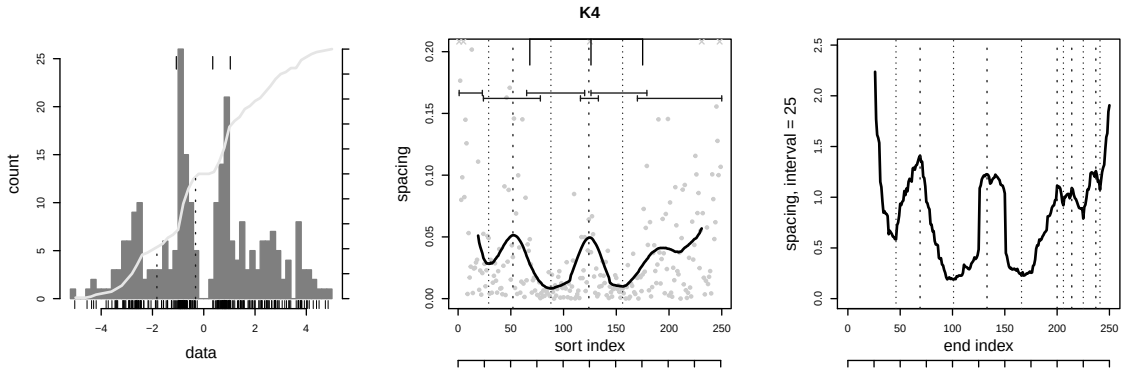
Figure 161 shows that peaks resolve the draws in the K2 sample, and the changepoints partially, but the level sections and flats do not. The analysis is less clear. The low-pass spacing finds the three peaks between each variate, and tests pass two of them as significant. There are five interval peaks, but only one is accepted. Detected peaks match, but the rate falls for significant peaks to one in half the trials. The loss in matching can either be mis-positioning, perhaps from the rough interval spacing or because different peaks pass as significant. There are two or three changepoints with a moderately wide confidence interval, and one matches up with an endpoint of a level section. The three sections cannot be said to identify the four variates, since the two tails often create a section.

K3 (Figure 162) has larger peaks in each gap between the four draws that are not always considered significant, but which do set a changepoint and split the level sections. There are one or two low-pass peaks, all considered significant, and three or four interval peaks of which one is significant. The two methods match somewhat, for half the significant peaks and three quarters of the detected. There are three changepoints that correspond to each peak in the figure, with a wide confidence interval. The correct placement rate is limited by the appearance of low-pass flats, which rarely occur but are significant when they do. There are six level sections that cover the variates and tails, and perhaps the gaps. They cover any flat, and an endpoint matches half the changepoints.



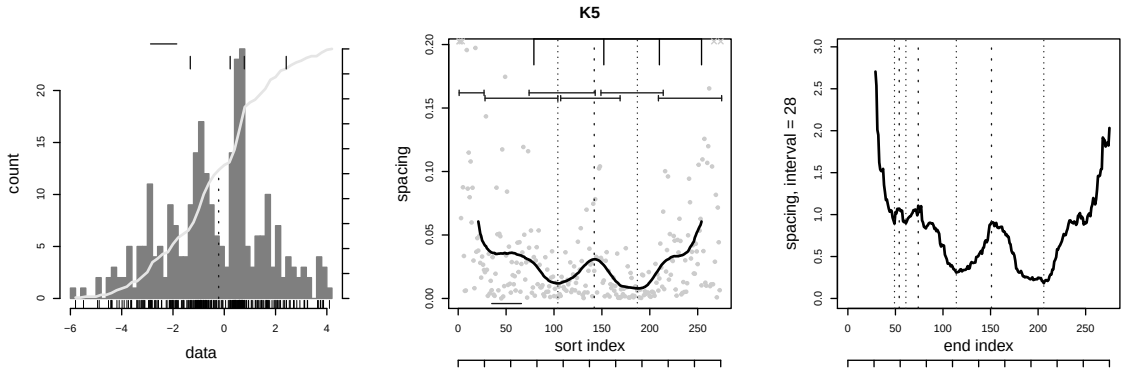

| K3 (quad) | peaks | | flats | | CPT | level | |
|---|---|---|---|---|---|---|---|
| | LP | Diw | LP | Diw | | | |
| mean count | 1.5 | 1.2 | 0.2 | 0 | 3.0 | 6.0 | |
| 90% CI | 1–2 | 0–3 | 0–1 | | 1–5 | 5–7 | |
| detected | 1.6 | 3.4 | 0.2 | 0 | | | |
| matching | 0.6 | | 0 | | 0.1 | 1.7 | 0.2 |
| 90% CI | 0–2 | | | | 0–1 | 0–3 | 0–1 |
| detected | 1.2 | | 0 | | placed | CPT | flat |

Figure 162: Quad-modal dataset K3.




| K4 (quad) | peaks | | flats | | CPT | level | |
|---|---|---|---|---|---|---|---|
| | LP | Diw | LP | Diw | | | |
| mean count | 2.4 | 1.3 | 0.1 | 0 | 2.9 | 6.0 | |
| 90% CI | 1–3 | 0–3 | | | 1–4 | 5–7 | |
| detected | 2.5 | 4.1 | 0.1 | | | | |
| matching | 0.7 | | 0 | | 0.1 | 2.1 | 0.1 |
| 90% CI | 0–2 | | | | | 1–3 | |
| detected | 1.8 | | 0 | | placed | CPT | flat |

Figure 163: Quad-modal dataset K4.



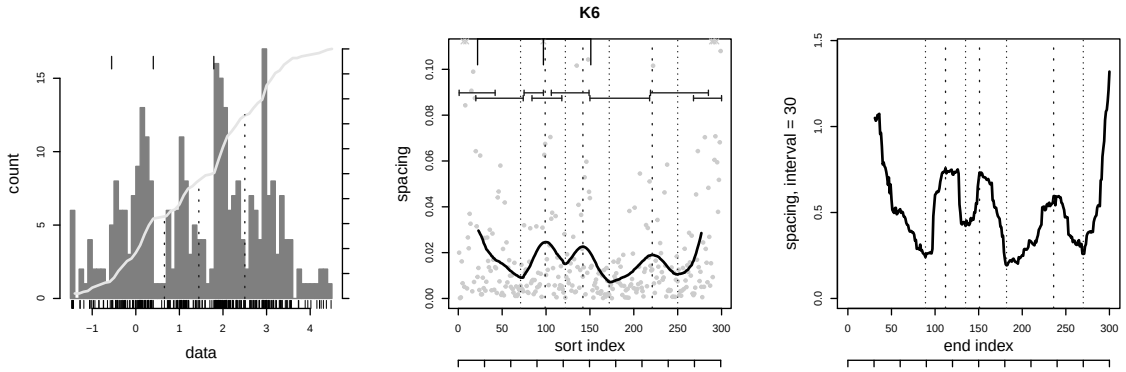

| K5 (quad) | peaks | | flats | | CPT | level | |
|---|---|---|---|---|---|---|---|
| | LP | Diw | LP | Diw | | | |
| mean count | 1.1 | 0.9 | 0.1 | 0 | 2.6 | 5.7 | |
| 90% CI | 0–2 | 0–2 | 0–1 | | 0.9–4 | 5–7 | |
| detected | 1.7 | 3.4 | 0.4 | 0 | | | |
| matching | 0.3 | | 0 | | | 1.4 | 0.1 |
| 90% CI | 0–1 | | | | | 0–3 | 0–1 |
| detected | 1.3 | | 0 | | placed | CPT | flat |

Figure 164: Quad-modal dataset K5.

The spacing clearly increases in the main central transition between the draws of sample K4 (Figure 163), but the side draws seem marginal. No flats appear, as the modes are not large enough to produce a stable spacing. Changepoints do pick out the anti-modes and level sections span each variate. These qualitative descriptions are stable over the runs. There are two or three peaks in the low-pass spacing, all judged significant. The interval spacing picks up an extra detected peak, but tests only accept the central one, and in a third of the runs one of the side transitions. Half the significant peaks align between the two methods, while three quarters of the detected peaks do. There are no flats in either spacing. The spacing contains six level sections and three changepoints, and the two that match endpoints indicate both checks respond to the same increase in spacing.

The spacing in Sample K5 (Figure 164) has a peak at the gap between the central draws, but not necessarily within each pair to the sides. Like K1 there is a change in the variance within the pair, but the filters transition smoothly without ringing at the edges. Flats may mark one of the modes, and the level sections do better at covering each draw. The changepoints do seem to separate all modes. The low-pass spacing has one or usually two peaks, one of which tests significant. There are twice as many peaks in the interval spacing, but still only one passes. There is poor alignment of the significant peaks between the two methods, while most of the detected peaks do match. A low-pass flat appears in half the runs but is usually not found significant. There are six level sections and three changepoints, one or two of which align with endpoints.

K6 (Figure 165) simulates a discrete draw, using tight normal variates at integer means. The standard deviation for the variates is not small enough to create clear gaps between the variates. The distribution proves separable. There are three low-pass peaks in the gaps separating the modes, and tests accept two as significant. The interval spacing has a fourth peak, but tests reduce the four to one or two significant. Matching between the methods is marginal, with one significant peak at the same



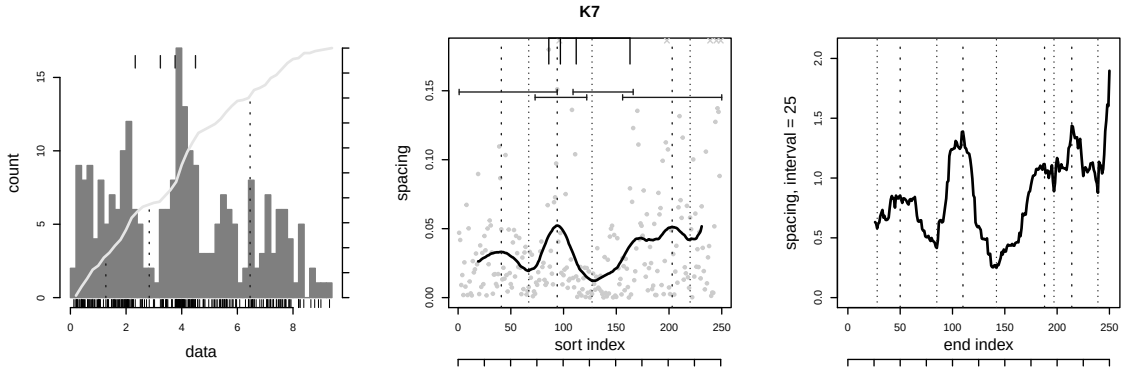

| K6 (quad) | peaks | | flats | | CPT | level | |
|---|---|---|---|---|---|---|---|
| | LP | Diw | LP | Diw | | | |
| mean count | 2.1 | 1.5 | 0.1 | | 4.0 | 6.6 | |
| 90% CI | 1–3 | 0–4 | 0–1 | | 2–7 | 4–8 | |
| detected | 3.0 | 4.0 | 0.1 | | | | |
| matching | 0.6 | | 0 | | 0.1 | 1.9 | 0.1 |
| 90% CI | 0–2 | | | | 0–1 | 1—4 | 0–1 |
| detected | 1.9 | | 0 | | placed | CPT | flat |

Figure 165: Quad-modal dataset K6.

location in half the trials and in two out of three detected peaks. This implies that the two methods are identifying different peaks as significant or that the local maximum is not stable, suggested by the rough tops of the interval peaks in the figure. Neither spacing contains flats. There are six or seven level sections spanning the four draws, the two tails, and possibly a peak. Four changepoints divide the draws, although the confidence interval in the count is wide. Half of them align with a section endpoint. The rate may be raised because there so many endpoints, making coming close to one easier.

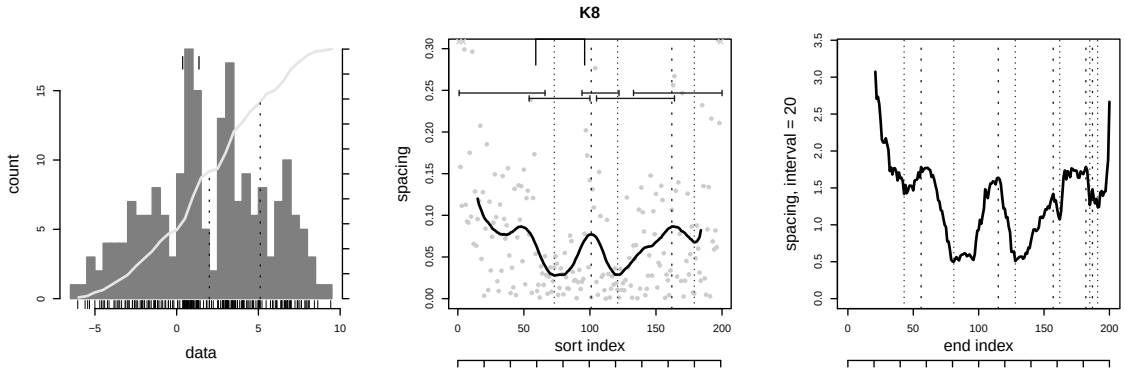
In sample K7 (Figure 166) the spacing can separate the four draws in principle, but testing does not accept most of the features. There are two or three peaks in the low-pass spacing corresponding to the gaps between draws, with one found significant. The interval spacing has four peaks, with one passing. Again we see a disconnect between the matching of detected and significant peaks, suggesting that the position of the largest maximum is not stable. In the figure the main interval peak has ringing which would explain the drift. The low-pass spacing has a flat in half the runs, and tests accept it as significant; the interval spacing is rough and has no flats. There are four level sections that in the figure do not distinguish the close, left two variates, leaving the fourth section to span the main gap. A section covers any significant flat. There are two or three changepoints that bound the central and right gaps, but do not separate the left variates. One changepoint aligns with a section endpoint. In half the trials a changepoint lies in the gap between a peak and level section; this is limited by the number of flats that exist. We can only partially resolve this example.

Figure 167 shows that the left and right draws of sample K8 merge into the starting and ending tails, making it harder to identify a peak in the spacing. Changepoints bound the second mode, with a moderate confidence interval. The level sections do correspond to the four modes, but neither filter settles enough to support flats. Quantitatively we find two or three peaks in the low-pass spacing,



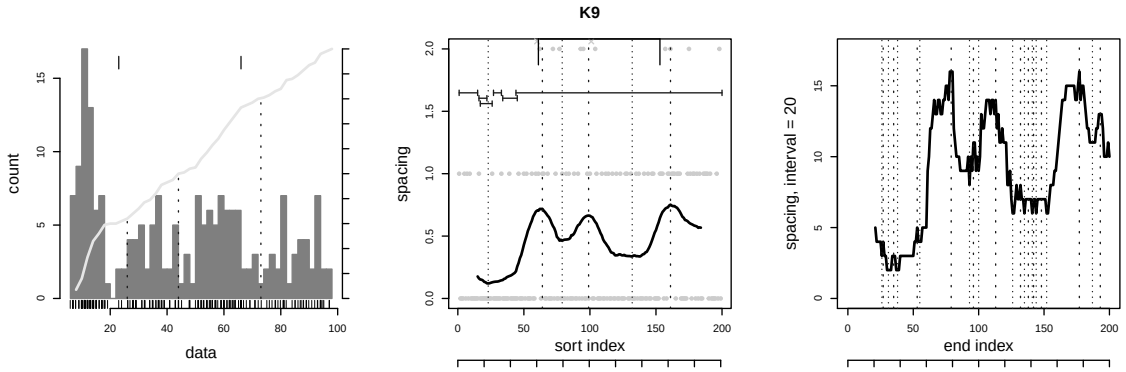

| K7 (quad) | peaks | | flats | | CPT | level | |
|---|---|---|---|---|---|---|---|
| | LP | Diw | LP | Diw | | | |
| mean count | 0.8 | 0.9 | 0.3 | 0 | 2.5 | 3.9 | |
| 90% CI | 0–2 | 0–2 | 0–1 | 0 | 1–4 | 3–5 | |
| detected | 2.5 | 4.1 | 0.4 | 0 | | | |
| matching | 0.2 | | 0 | | 0.5 | 1.0 | 0.3 |
| 90% CI | 0–1 | | | | 0–2 | 0–2 | 0–1 |
| detected | 1.8 | | 0 | | placed | CPT | flat |

Figure 166: Quad-modal dataset K7.




| K8 (quad) | peaks | | flats | | CPT | level | |
|---|---|---|---|---|---|---|---|
| | LP | Diw | LP | Diw | | | |
| mean count | 1.4 | 1.4 | 0.1 | 0 | 2.1 | 4.4 | |
| 90% CI | 0–3 | 0–5 | | 0 | 1–4 | 3–5 | |
| detected | 2.4 | 4.4 | 0.1 | 0 | | | |
| matching | 0.6 | | 0 | | 0.1 | 1.1 | 0.1 |
| 90% CI | 0–2 | | | | | 0–2 | |
| detected | 2.1 | | 0 | | placed | CPT | flat |

Figure 167: Quad-modal dataset K8.



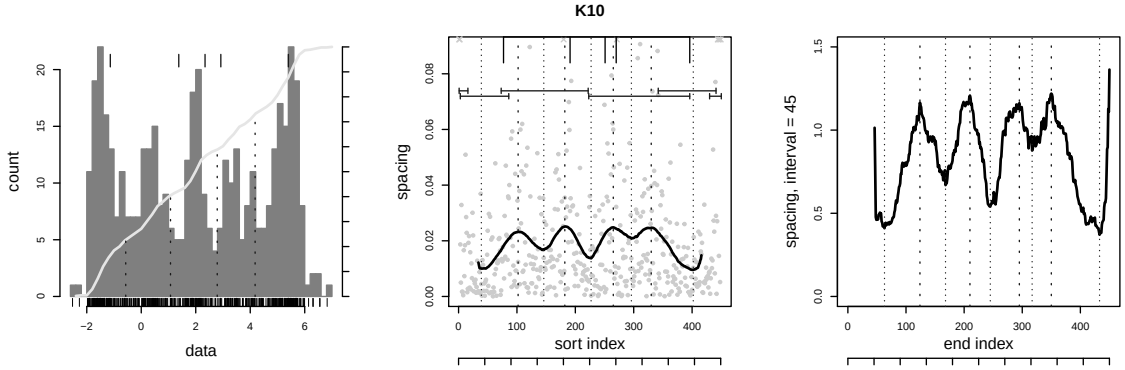

| K9 (quad) | peaks | | flats | | CPT | level | |
|---|---|---|---|---|---|---|---|
| | LP | Diw | LP | Diw | | | |
| mean count | 1.6 | 2.4 | 0 | 0 | 2.3 | 7.9 | |
| 90% CI | 1–3 | 0–8 | | | 1–4 | 5–12 | |
| detected | 2.8 | 9.7 | 0 | 0 | | | |
| matching | 0.4 | | 0 | | 0 | 1.1 | 0 |
| 90% CI | 0–1 | | | | | 0–3 | |
| detected | 2.1 | | 0 | | placed | CPT | flat |

Figure 168: Quad-modal discrete dataset K9.

with tests rejecting one of them as insignificant. The interval spacing is rougher, especially in the left and right draws. It has four or five peaks, one or two accepted as significant. The position of the detected peaks match up between the two methods, but only one significant peak does in half the runs. There are two changepoints, missing one of the inter-modal transitions; the lack of flats limits their correct placement. There are four or five level sections, one endpoint lining up with a changepoint.

K9 (Figure 168) is a discrete distribution that depends on filtering to smooth the integral jumps in the spacing. We do find three peaks in the low-pass spacing, half of which the tests accept as significant. The discrete data make the interval spacing very rough, creating a large number of peaks, ten, that reduce to two or three significant, although the 90% confidence interval on the average counts is very wide. With so many potential matches, alignment between detected peaks is good, with three quarters lining up, but is poor between significant peaks. There are two changepoints on average that correspond to the gaps between draws. This is better than we've seen with other discrete data, where the algorithms can trigger at each step in value; even the confidence interval is reasonable. The level section algorithm, in contrast, generates eight sections with a wide confidence interval. The sections in the figure extend beyond the modes and cannot be used.

The peaks and changepoints in the K10 example (Figure 169) mark or bound the transitions between the five draws. The low-pass spacing captures three of the four gaps, with all considered significant. The interval spacing finds four peaks, with only one or two passing tests. Matching between the methods is moderate, however, compared to other K samples. The low-pass spacing usually contains one flat which tests often accept, yielding a significant feature in 40% of the runs. One of the five level sections covers any significant flat. The confidence interval is wide, and despite the high number the endpoints the alignment with changepoints is poor. In the figure the level sections




| K10 (penta) | peaks | | flats | | CPT | level | |
|---|---|---|---|---|---|---|---|
| | LP | Diw | LP | Diw | | | |
| mean count | 2.9 | 1.3 | 0.4 | 0 | 3.4 | 4.8 | |
| 90% CI | 2–4 | 0–3 | 0–1 | | 1–5 | 2.9–8 | |
| detected | 2.9 | 4.2 | 0.7 | 0 | | | |
| matching | 0.3 | | 0 | | 0.7 | 0.5 | 0.4 |
| 90% CI | 0–1 | | | | 0–2 | 0–2 | 0–1 |
| detected | 0.9 | | 0 | | placed | CPT | flat |

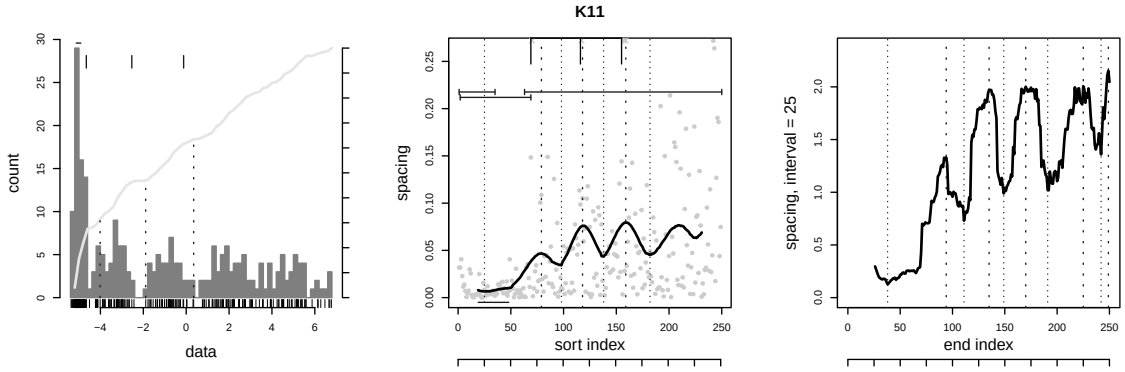
Figure 169: Penta-modal dataset K10.

do not identify each variate, and two sections mark the tails. There are three or four changepoints that partially identify the inter-mode transitions. The placement rate is higher than the number of flats, so changepoints must lie to either side of the flat.

Sample K11 (Figure 170) generates three or four peaks in the low-pass spacing and five in the interval, with the rightmost transition between draws consistently appearing. Testing judges two low-pass peaks and one interval as significant. Alignment between the two methods is moderate, holding for a third of the significant peaks and half the detected. The low-pass spacing contains one flat, significant, in the large tight draw. There are four level sections, one covering the flat. With up to two in the tails they cannot resolve all draws, as seen in the figure. There are three changepoints marking anti-modes, placed correctly; one matches the section endpoint at the end of the large draw. Overall the spacing separates four of the five draws with peaks and changepoints, although only two meet the significance tests, and the flats and level sections span one mode. It only partly resolves the modality.

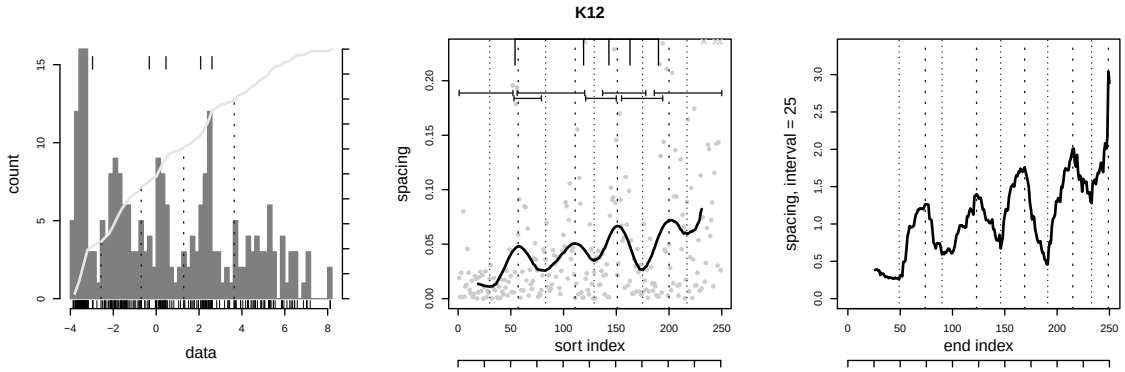
The transition between the fourth and fifth variates in sample K12 (Figure 171) is smaller than in K11, and this appears in the low-pass spacing results. There is one fewer detected peak and one less significant peak, and the flat goes away. The interval spacing still has five peaks, one significant. There are also fewer changepoints, still marking the transitions between draws, either in a gap or bounding it. The level sections don't change, even if the flat in the first draw no longer exists; in the figure the sections run from gap to gap. There are moderate spreads in the number of changepoints and level sections.

K13 (Figure 172) simulates a discrete distribution, using tight normal variates with integral means in place of draws that generate integer values. It adds a fifth normal draw to the K6 setup and tweaks



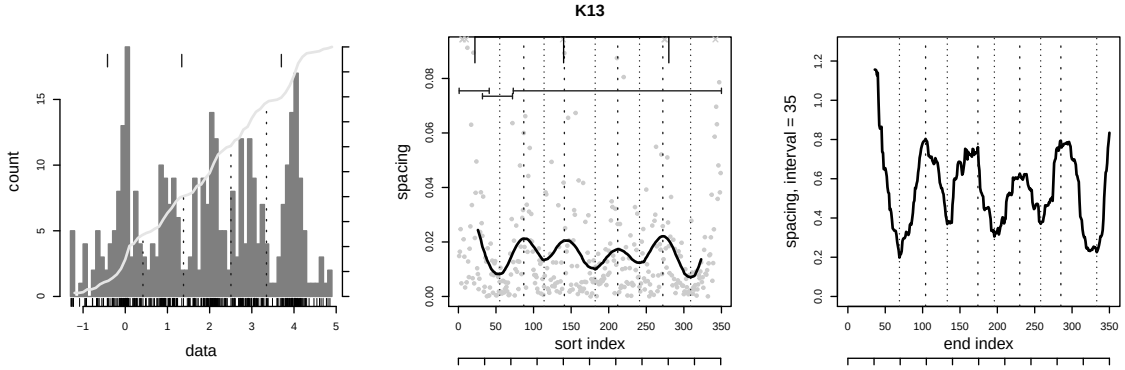

| K11 (penta) | peaks | | flats | | CPT | level | |
|---|---|---|---|---|---|---|---|
| | LP | Diw | LP | Diw | | | |
| mean count | 1.9 | 1.0 | 0.9 | 0 | 2.8 | 4.1 | |
| 90% CI | 1–3 | 0–3 | 0–1 | | 1–5 | 3–6 | |
| detected | 3.5 | 4.7 | 0.9 | 0 | | | |
| matching | 0.3 | | 0 | | 2.1 | 1.3 | 0.9 |
| 90% CI | 0–1 | | | | 0–4 | 0–2 | 0–1 |
| detected | 1.9 | | 0 | | placed | CPT | flat |

Figure 170: Penta-modal dataset K11.




| K12 (penta) | peaks | | flats | | CPT | level | |
|---|---|---|---|---|---|---|---|
| | LP | Diw | LP | Diw | | | |
| mean count | 0.7 | 0.8 | 0.2 | 0 | 2.2 | 4.2 | |
| 90% CI | 0–2 | 0–2 | 0–1 | | 1–4 | 3–6 | |
| detected | 2.8 | 5.0 | 0.2 | 0 | | | |
| matching | 0.1 | | 0 | | 0.1 | 1.1 | 0.1 |
| 90% CI | 0–1 | | | | 0–1 | 0–2 | 0–1 |
| detected | 1.7 | | 0 | | placed | CPT | flat |

Figure 171: Penta-modal dataset K12.



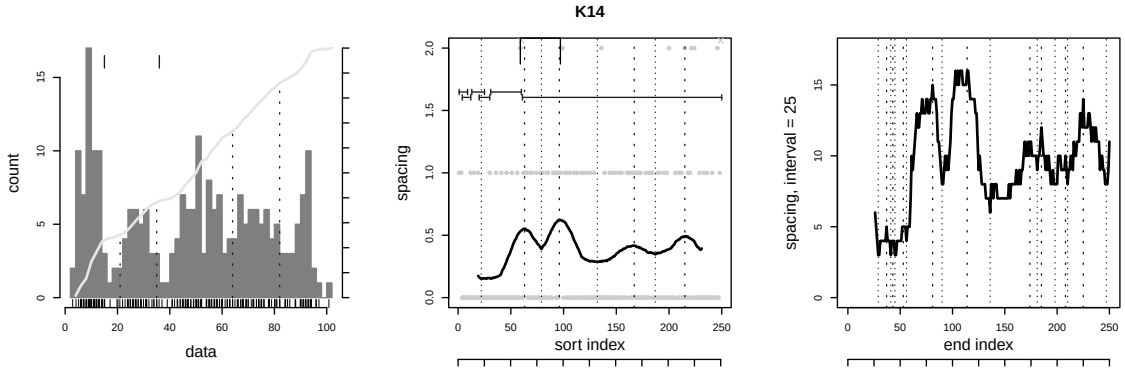

| K13 (penta) | peaks | | flats | | CPT | level | |
|---|---|---|---|---|---|---|---|
| | LP | Diw | LP | Diw | | | |
| mean count | 3.0 | 1.6 | 0 | 0 | 4.4 | 5.2 | |
| 90% CI | 2–4 | 0–4 | | | 2–7 | 3–9 | |
| detected | 3.7 | 4.5 | 0 | 0 | | | |
| matching | 0.6 | | 0 | | 0 | 1.4 | 0 |
| 90% CI | 0–2 | | | | | 0–3 | |
| detected | 1.7 | | 0 | | placed | CPT | flat |

Figure 172: Penta-modal dataset K13.

the standard deviations for variety. A sixth uniform variate provides an additional background that fills in the gaps between the normal draws. The spacing finds the inter-mode transitions, but the spread in values around each mean, together with the background and smoothing by the filter, prevent flats. K13 is an extension of K6 and the results are similar. The low-pass spacing adds one peak, giving four at the transitions. Tests judge three significant. The interval spacing adds a peak in half the trials, but the significant count is unchanged. Matching of the peak positions does not change. The interval peaks have a rough top making the location of the local maximum noisy. The algorithms add changepoints which match the transitions or tails. There are five level sections sharing endpoints with one or two changepoints. They partially span the modes, since two may appear in the left and right tails. In the figure one section spans four of the draws.

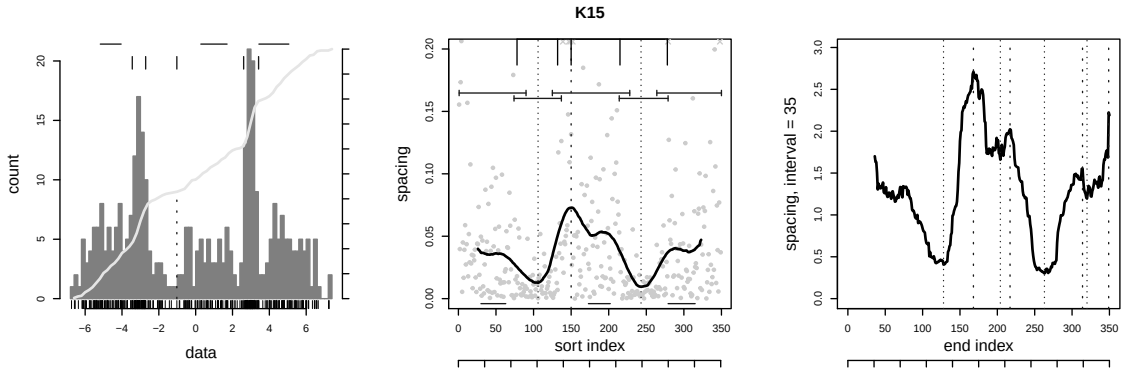
K14 (Figure 173) is a discrete distribution built from five binomial variates. It is an extension of the K9 setup, and the comments from that example apply. Two draws apparently cannot be resolved, as the average counts are very similar to K9, including the low-pass and interval peaks detected and significant, the changepoints, and level sections. Even the confidence intervals do not change.

The example total draw for sample K15 in Figure 174 shows three separate modes, with a smooth transition in the first and third between the broad background and tight extra variates. This transition creates a change in the average local spacing but without a gap there is no or a minimal increase at the edge and the low-pass filtered spacing does not raise a peak. The interval spacing can succeed in this case. Flats correspond to the three large draws, but the tight extras are not large enough to create a stable floor. The interval spacing here does better at not rounding the signal, but the draws are too small to meet the minimum length. Changepoints do bracket the tight draws and can mark the transitions to either side of the central mode. The level sections span each draw separately, identifying the two tight ones. On average there are two or three peaks in the low-pass spacing, all




| K14 (penta) | peaks | | flats | | CPT | level | |
|---|---|---|---|---|---|---|---|
| | LP | Diw | LP | Diw | | | |
| mean count | 1.4 | 2.0 | 0.1 | 0 | 2.1 | 7.6 | |
| 90% CI | 1–2 | 0–8 | 0–1 | | 1–4 | 4.9–12 | |
| detected | 2.8 | 10.2 | 0.1 | 0 | | | |
| matching | 0.1 | | 0 | | 0.1 | 0.9 | 0.1 |
| 90% CI | 0–1 | | | | | 0–3 | 0–1 |
| detected | 1.4 | | 0 | | placed | CPT | flat |

Figure 173: Penta-modal discrete dataset K14.




| K15 (penta) | peaks | | flats | | CPT | level | |
|---|---|---|---|---|---|---|---|
| | LP | Diw | LP | Diw | | | |
| mean count | 2.4 | 2.5 | 0.4 | 0 | 4.2 | 6.2 | |
| 90% CI | 1–4 | 1–4 | 0–1 | | 3–5 | 5–8 | |
| detected | 2.4 | 4.3 | 0.4 | 0 | | | |
| matching | 0.7 | | 0 | | 0.7 | 2.1 | 0.3 |
| 90% CI | 0–2 | | | | 0–3 | 1–3 | 0–1 |
| detected | 1.0 | | 0 | | placed | CPT | flat |

Figure 174: Penta-modal dataset K15.

significant, so the spacing can distinguish one of the tight variates in half the runs. The interval spacing has four peaks, with two rejected by the tests. Matching between the two methods is poor. A low-pass flat appears in half the trials, and passes testing. There are six level sections to span the five draws and possibly two tails. They mostly cover the flat if it appears, but not always. There are four changepoints, with half matching endpoints of the level sections. This rate may be inflated by the large number of endpoints making alignment to one easier.

### Detail 38 Summary of Literature Datasets

#### Sample Count

To summarize the features found in the literature samples [Detail 35], [Detail 36], [Detail 37] and their significance, we combine the results by the number of modes in the draw. 'abi' stands for an asymmetric bi-modal sample, and the chirp is described in W14. The total samples per class are

| | uni | bi | abi | tri | quad | penta | chirp |
|---|---|---|---|---|---|---|---|
| number samples | 16 | 12 | 18 | 14 | 11 | 9 | 1 |

Some comments about each sample occur repeatedly and provide a qualitative assessment of using spacing to analyze modality. The interval spacing is rough, which generates more peaks and fewer flats. Tuning the detector parameters for each case may be inevitable. The roughness may degrade the accuracy of the peak's location and matching to an anti-mode. Tests reject many of the detected peaks, moreso for the interval than low-pass spacing. Tests do accept most flats, however. Since there are fewer flats than peaks, a more conservative test for the peaks seems appropriate. The changepoint algorithms and level section test often trigger at edge of each tail, where the spacing inevitably increases strongly. The changepoints generally do not line up with level section endpoints. But almost always some level section covers any significant flat completely.

#### Feature Occurrence

We begin by looking at the number of features found as the number of modes increases.

The low-pass and interval spacing have an interesting difference between the raw and test results. The detection rate of peaks is much higher in the interval spacing than the low-pass, but the number of significant peaks in the two methods is largely the same. The low-pass spacing contains more peaks than expected in the uni-modal and bi-modal samples, but misses one inter-mode transition between three or more variates. Said differently, it finds the expected uni-modal and bi-modal transitions and begins to miss one in the tri-modal case. Without a model of which transitions give measurable increases in the spacing, we cannot say if the samples should be resolved, or if variations in the draws prevent that, or if the detectors are finding all features that they can. We can only say the tri-modal and more complicated samples generate marginally detectable peaks in the transitions. The interval spacing generally has two more peaks than the low-pass, with the smallest difference for the uni-modal samples and largest for the tri-modal. It over-estimates the number of modes before testing. We find fewer significant peaks than modes, except for the uni-modal and bi-modal samples, and the counts in the two spacings are much more similar, so testing is removing many of the extra peaks in the interval spacing. There are no significant low-pass peaks for uni-modal samples, and only a few in the interval spacing, so the false positive rate is low. Although the number of significant peaks increases with the complexity, it lags the actual number of modes. Significant peaks conservatively estimate the modality of data.

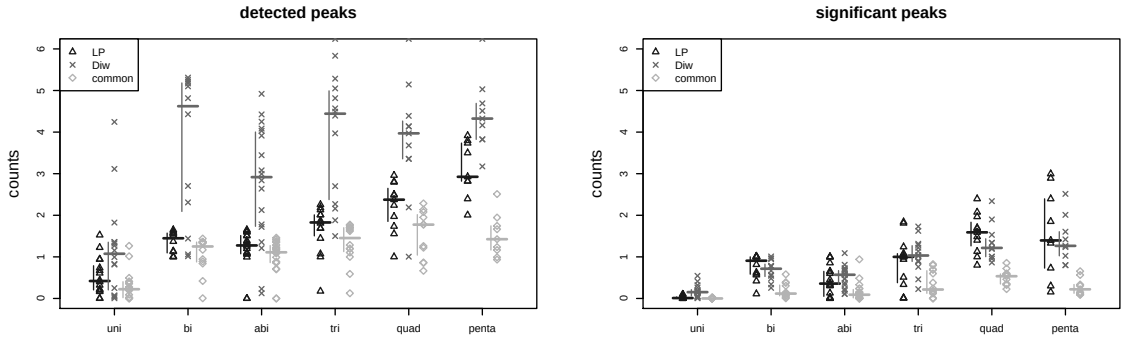



Figure 175: Counts of peaks averaged over all runs and grouped by complexity of sample. Bars present median and inter-quartile range.

These trends are visible in Figure 175 when comparing all samples grouped by their complexity. The plots are a form of boxplot, showing individual points for each sample grouped by the number of modes. A horizontal bar marks the median (also in Table 41) and vertical line the inter-quartile range. The interval spacing detected peak counts are higher than the low-pass. There are many more detected interval peaks with a much larger inter-quartile range for the bi- and tri-modal samples, but the counts do not show a trend. The two methods have similar counts of significant peaks, with the interval peaks now showing an increasing trend, but with too small a slope.

Table 41: Median Local Peaks Count

| | | uni | bi | abi | tri | quad | penta | chirp |
|---|---|---|---|---|---|---|---|---|
| detected | LP | 0.42 | 1.45 | 1.27 | 1.83 | 2.38 | 2.92 | 2.00 |
| | Diw | 1.07 | 4.62 | 2.92 | 4.44 | 3.97 | 4.33 | 3.00 |
| significant | LP | 0.01 | 0.91 | 0.36 | 1.00 | 1.59 | 1.40 | 1.15 |
| | Diw | 0.15 | 0.72 | 0.57 | 1.03 | 1.22 | 1.26 | 1.23 |

The graphs also include a common category that will be used when discussing the consistency of the tests.

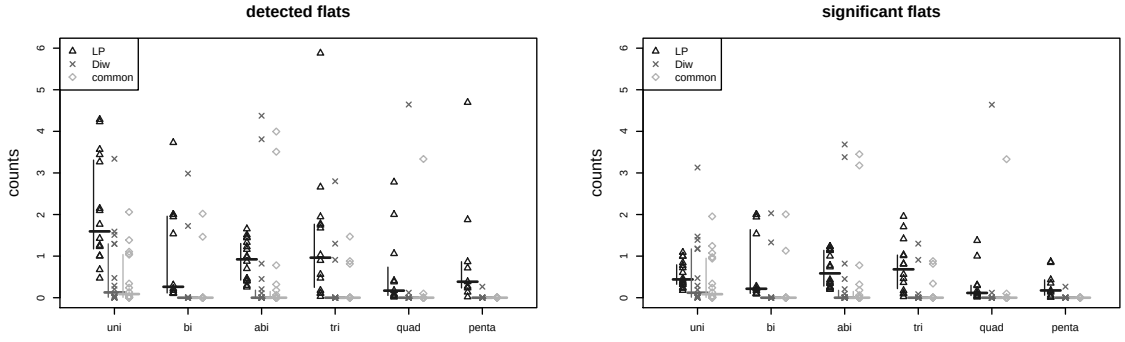



Figure 176: Counts of flats averaged over all runs and grouped by complexity of sample.

The analysis of flats in the interval spacing is complicated because the roughness of the signal

prevents the detection of flats in most samples. The median counts are zero except for the uni-modal samples, and there only rarely. But the counts in Figure 176 show large outliers that, if we were to present the average number instead of the median, would raise the values by 0.5, independent of the class (Table 42). The mean significant counts would be very close. The low-pass spacing has more flats, enough to use median counts. We detect fewer flats as more modes are added, but the bi-modal samples have unusually few, and the counts stabilize in the quad- and penta-modal examples. The inter-quartile range shrinks with the complexity class. The trend may be a trade-off in the filtering: more smoothing reduces the signal's ripple and improves detection, but also broadens the edge of a mode which shrinks its size. This, coupled with smaller variates because we try to keep the total sample size the same, can produce fewer flats.

Table 42: Local Flats Count

| | | uni | bi | abi | tri | quad | penta | chirp |
|---|---|---|---|---|---|---|---|---|
| median | | | | | | | | |
| detected | LP | 1.59 | 0.26 | 0.92 | 0.96 | 0.17 | 0.39 | 2.00 |
| | Diw | 0.13 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 4.00 |
| significant | LP | 0.44 | 0.21 | 0.58 | 0.68 | 0.12 | 0.17 | 2.00 |
| | Diw | 0.09 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.35 |
| average | | | | | | | | |
| detected | LP | 2.07 | 1.03 | 0.89 | 1.36 | 0.64 | 1.02 | 2.00 |
| | Diw | 0.73 | 0.39 | 0.54 | 0.36 | 0.43 | 0.03 | 4.00 |
| significant | LP | 0.56 | 0.73 | 0.68 | 0.75 | 0.32 | 0.32 | 2.00 |
| | Diw | 0.59 | 0.28 | 0.48 | 0.16 | 0.43 | 0.03 | 0.35 |

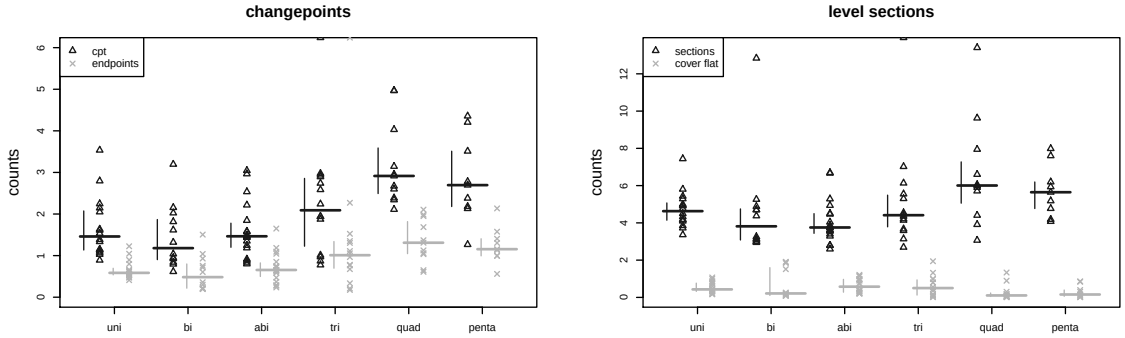



Figure 177: Changepoint and level section counts averaged over all runs and grouped by complexity of sample.

We see the number of changepoints increasing with the modality (Figure 177, left, and Table 43), lagging by one. The penta-modal is an exception, intermediate between the tri- and quad-modal counts. The uni-modal samples have 1.46, and although they could be detecting shoulders and other composite variates in some examples, this suggests that the algorithms also respond to the tails. If this forms a background count then it is not clear if the bi-modal changepoints correspond to a feature. We will return to this question when discussing the selectivity of the feature's placement. What we do not see is multiple changepoints per mode, so they do not bound a mode or transition on both sides, only on one. The inter-quartile range is independent of the modality class, outside the asymmetric bi-modal samples which have the smallest range.

Table 43: Median Changepoint Count

| | uni | bi | abi | tri | quad | penta | chirp |
|---|---|---|---|---|---|---|---|
| detected | 1.46 | 1.18 | 1.46 | 2.09 | 2.92 | 2.69 | 4.21 |

The level section test finds more sections than modes (Figure 177, right, and Table 44). The number of level sections does no better than the other features at reflecting the modality of the sample. The counts of the two bi-modal classes are consistent. The uni-modal samples have many sections, more than the tri-modal, showing that the test is inherently sensitive to variations in the individual draws and is noisy. It cannot mark changes in modality with any degree of confidence. The bi-, tri-, and quad-modal classes have samples that produce unusually many level sections, enough to distort an average count. The medians presented in the graph and table are centered and have small inter-quartile ranges, and represent the overall performance well.

Table 44: Median Level Section Count

| | uni | bi | abi | tri | quad | penta | chirp |
|---|---|---|---|---|---|---|---|
| detected | 4.63 | 3.81 | 3.75 | 4.41 | 6.00 | 5.64 | 10.95 |

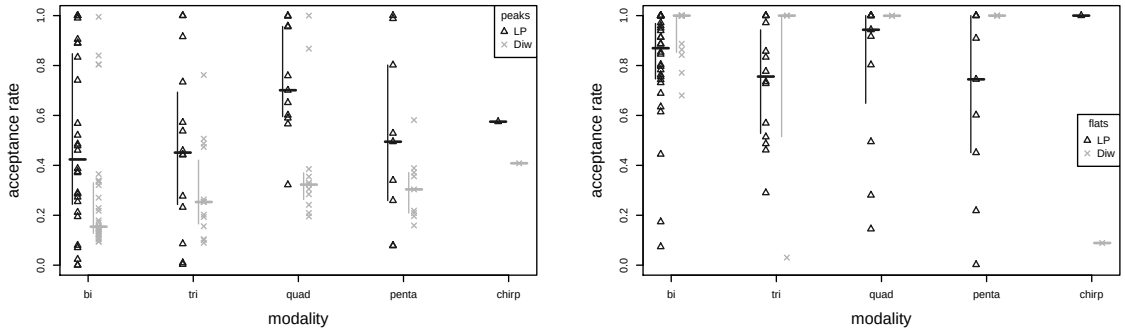



Figure 178: Passing rate of detected features at the 0.01 level for peaks (left) and flats (right) in each spacing.

Define the acceptance rate as the fraction of detected features $N_{det}$ that pass any test and are accepted as significant $N_{sig}$,

$$\text{acceptance} = \frac{N_{sig}}{N_{det}} \tag{34}$$

At the 0.01 level, Figure 178 shows the rate for peaks and flats in either spacing. Half the detected peaks in the low-pass spacing survive testing, or three-quarters in the quad-mode samples. The inter-quartile range is large, however, about half the possible range from 0.0 to 1.0, and for the bi- and quad-modal classes the median does not center within it. Acceptance is lower for peaks in the interval spacing, showing a slight improvement as the sample's complexity grows, rising from less than a fifth to a third of the detected peaks. The inter-quartile range is smaller, implying the tests are more stable. The acceptance rate for flats is much higher, above three-quarters in the low-pass spacing and essentially perfect for the interval flats. The conservative nature of the tests is countered by the strict requirements for a flat; the first would tend to drive the acceptance rate down, but the rarity of the feature would improve it.

**Feature Stability**

The tables in [Detail 37] contain 90% confidence intervals for the number of features, taken over

all 200 trials of each literature sample. The size of the ranges measures the stability of the features over draw features. A range of 2–2 with size 0 means every trial has two features, while one of 2–5 with size 3, taken from the level section count for X3, shows moderate variation. It is the size and not the location of the interval that reflects the stability. Whether the average or median is centered in the interval reflects a bias in the test that we do not consider. In Figure 179 we plot how many samples generate confidence intervals of each size for significant features, passed by any test.

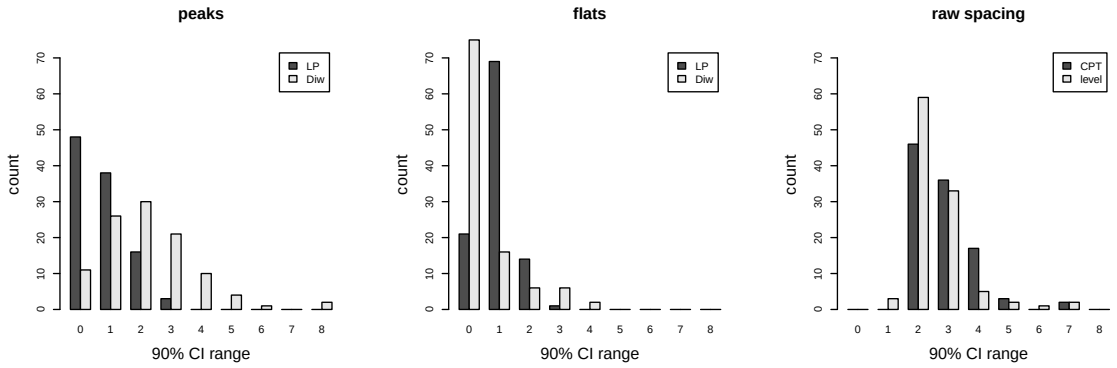



Figure 179: Histogram of widths of 90% critical intervals for peaks (left), flats (middle), and raw spacing features (right).

There are a few differences in the histograms. Peaks in the low-pass spacing have smaller confidence intervals than they do in the interval spacing, both in terms of the average size and how quickly the histogram drops off. Flats distinguish these two aspects, however. Interval flats have more ranges of zero size, but this is because they appear in so few samples. They do extend to larger sizes, suggesting that if we had more flats, for example by relaxing the ripple specfication, the flats would not appear to be very stable. The confidence intervals for the features in the raw spacing, changepoints and level sections, are surprisingly similar, since it seemed from the individual samples that the level sections would vary more. The variability of these two features is noticeably higher than the peaks and flats. The histogram does not include a single outlier, the changepoints in sample P1.

**Feature Placement Accuracy**

Some common threads appear in the discussions in [Detail 36] about the matching of features to modes and gaps in the ideal density of each sample. The low-pass peaks match gaps and avoid modes, especially when judged significant. The interval peaks and changepoints scatter between modes and anti-modes. Their position is not as cleanly defined. In several samples the interval peaks match the mode to the left of a gap, and not the gap itself. Testing helps to reduce the dispersal but does not eliminate it. The features generally follow any symmetry present in the sample's setup, although in marginal setups the variability of each draw can make this harder to see. Gaps do not seem to be identified consistently by all the tests, especially when the separation between variates makes them hard to distinguish. Flats do mostly lie within modes, although because we keep the total draw size roughly constant, more flats extend over mode and gap as the number of variates increases and the width of a mode approaches the minimum flat size. The features in this case naturally extend over the halfway point between mode and gap. There are very few flats in the interval spacing. The level sections are long and extend over more than one gap or mode, a consequence of them filling the entire sample, leaving no point uncovered. Any short sections mostly correspond to the tails of the total sample, and there is an asymmetry between the left and right tails caused by the algorithm using the end of the prior section as a basis for the start of the next.

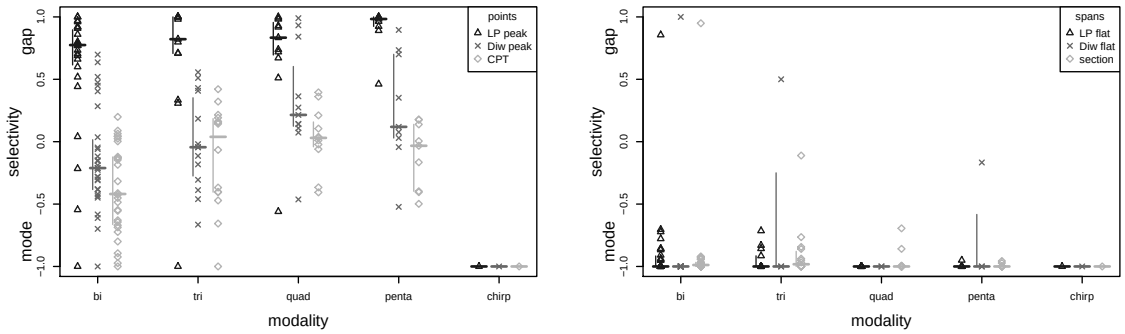



Figure 180: Selectivity of point features (left) and spans (right).

Figure 180 plots the selectivity (20) for features significant at the 0.01 level, combining the bi-modal and asymmetric bi-modal sets into one. There is a clear difference between the methods. Peaks in the low-pass spacing align well with the gaps, improving slightly with the number of modes although the inter-quartile range does not change. The interval spacing peaks are closer to balanced, and for the bi-modal and tri-modal samples actually align more often to modes. The inter-quartile range is large and the median lies within the lower half. It appears in the graph that the selectivity has two clusters within each class, with a small subset of samples having a better selectivity. For this subset the interval peaks locate the anti-mode, but in general the spacing is rough and the placement of the peaks noisy. The accuracy is poor.

Flats are consistently tied to the modes, in either spacing.

Figure 180 also contains the selectivity of the changepoints and level sections. The former resemble the interval peaks. Of course we expect the algorithms to trigger in the transition between modes, not at them or the gaps, so their selectivity should be balanced. They lie closer to a mode in almost all the bi-modal samples. The inter-quartile ranges for the tri- and penta-modal samples are not symmetric about the median, leaning towards modes. Selectivity for level sections involves only those features that overlap a single mode or gap, and excludes those that lie in the tail. The selectivity of these remaining short sections resembles the low-pass flats and may reflect the overlap of any flat by a section.

Another way to check the accuracy of the placement of point features is to measure how far away they are from the gap, as indices in the data (Figure 181). The histograms are gathered over all 200 trials of each sample for peaks significant at the 0.01 level, ignoring those that are closer to modes than gaps. The placement is good, within 5 points. This is independent of the distance we used to define matching, but it justifies that value. Peaks in the interval spacing are a bit more loosely placed, needing to integrate to a separation of 11 to include 95% of all trials, while the low-pass peaks reach that at a distance of 6. The tails for the histograms are extremely long, with a maximum separation of 70 for the low-pass peaks and 151 for the interval.

[Detail 36] contains information about two other aspects of the interval features (Figure 182). Flats and level sections may extend beyond a single mode; these are not included in the selectivity rate. The features are less confined as the sample complexity increases. 10% of the significant low-pass flats are long in the bi-modal samples, and there is a similar cluster in the tri-modal examples. However, this cluster includes less than half the samples, and the remainder pull the median fraction

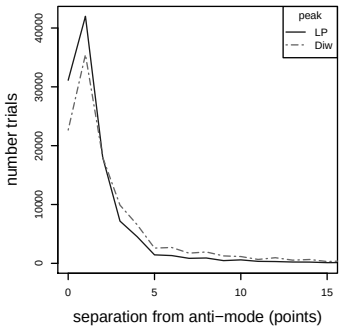



Figure 181: Histogram of separation between peaks and anti-modes over all trials and samples.

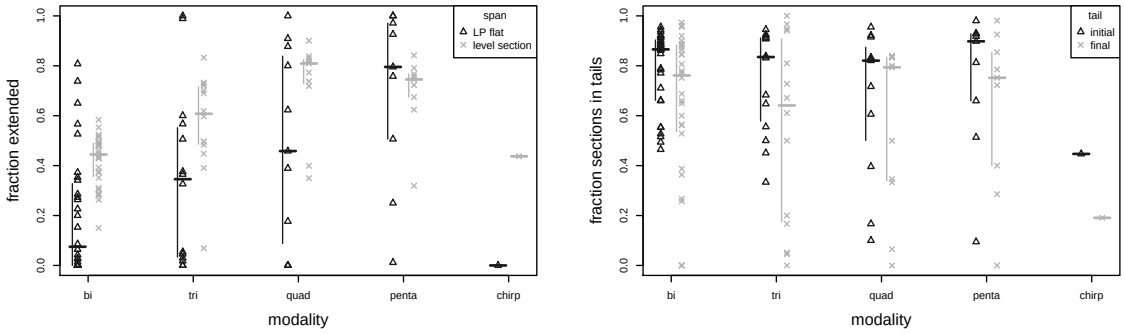



Figure 182: Fraction of features that extend beyond a mode (left) and of level sections in the tails (right).

to 35%. The quad-modal samples approach a 50% rate and have a much larger inter-quartile range, and the rate rises to three-quarters in the penta-modal examples. The level section rates are higher, above 50%. Although the selectivity implies that flats and sections do identify modes, the extension rate argues against this. The higher values for flats may reflect an increased difficulty in distinguishing draws, as seen in the lower acceptance rate of the penta-modal samples. It may also be a consequence of the minimum flat length and smaller variates, so that significant flats by their nature will tend to be long. Level sections on the other hand simply do not consistently identify modes. As a result of their count and length, they overlap and more 20% of each sample is covered by two level sections. This is independent of the complexity, but the variance is high.

The second aspect is the number of short intervals that are located in the tails, before and after the outer modes. More than 80% of the level sections inside the first mode actually mark the spacing change in the initial tail. The fraction is lower, between two-thirds and three-quarters, for the trailing tail, because of the asymmetry in the growth of level sections by the detection algorithm. Such sections are not included in the selectivity because they do not identify a mode.

**Feature Consistency**

Figures 175 and 176 include the common count of the same features that are present in both the

low-pass and interval spacing. Matching was defined arbitrarily in [Detail 37] as peaks lying within 5 points of each other, or flats that overlap by 70% of their length. Detected peaks align well, close to the maximum possible, which is against the fewest features in either spacing. They do less well as the sample complexity increases. Significant peaks do not match, with a count almost independent of the number of modes. Flat alignment is good as long as they appear in the interval spacing. This is seen not in the median bars in the graph, but in the individual points.

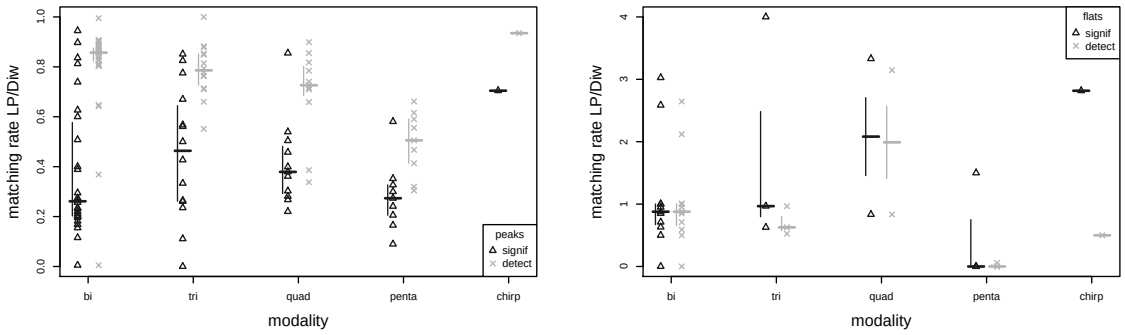



Figure 183: Matching of peaks (left) and flats (right) between the low-pass and interval spacing.

To remove the dependence in the raw counts on the number of features, define the matching rate as the fraction of the number of matching features $N_{match}$ to the smaller of the feature counts, $N_{LP}$ or $N_{Diw}$.

$$\text{matching} = \frac{N_{match}}{\min(N_{LP}, N_{Diw})} \tag{35}$$

For peaks the fraction is less than one, but with flats it can be greater because we can have multiple overlapping flats in the same match. Figure 183 presents the matching rate for peaks and flats, both in the detected features and those significant at the 0.01 level. Since there are more peaks in the interval spacing than the low-pass (Figure 175), the matching rate is a fraction of the low-pass count. Half to a quarter of the significant low-pass peaks have a counterpart in the interval spacing, with a rate that generally drops as more modes are added, although the bi-modal samples have a wide spread and a low median rate. The variability in the rate also decreases with the sample's complexity. Three-quarters to half of the detected peaks match, and show a consistent decline with complexity. Significant flats match in the bi- and tri-modal samples. The high rate in the quad-modal samples comes from multiple short interval flats per low-pass, and matching in the penta-modal samples is poor. There is little difference in the rate of detected and significant flats because their acceptance rate is high.

As continually said in [Detail 37], level sections completely overlap almost all significant flats, shown in Figure 177. The count appears to be less than one, but is limited by the number of flats. The matching rate is greater than 90% for all samples. Despite the large number of level sections, their endpoints do not align to changepoints. 40% of the changepoints match an endpoint, consistently across the complexity classes.