Bayesian Blocks

Sofie Bruun and Ida Stoustrup

March 7, 2018

1 Common binning

In histograms, it is common to use uniform bins. This is often not optimal as the amount of data and the variation in the data is often not constant. This means that uniform binning will often result in too many bins in regions with little, non-varying data and too few bins in well-described regions where the mean changes rapidly.
Aside from this, it is common to choose the number of bins arbitrarily, which increases the sensitivity to bias and does not guarantee reasonable results. Arbitrary, uniform binning may also mean that the placement of the bins splits a signal and thereby decreases the signal strength.
All this means that fits performed on data that is binned arbitrarily and uniformly may suffer greatly in both visual and statistical quality.

2 Bayesian Blocks

The idea of Bayesian blocks is to approximate the data as part-wise constant and to find the optimal amount of constant segments (called blocks) and the placement of these. This bears obvious resemblance to what is done in a histogram and so the method is often applied to optimize histograms.
Bayesian blocks are non-uniform, so when applied to histograms the binwidth varies. Each block, or bin for histograms, is separated by change points at the edges. They start at the first data point and end at the last one with no gaps, although bins may be empty. One could alternatively adjust the bin widths manually to achieve some of the same advantages, but bayesian blocks introduces an objective way of achieving optimal binning in multiple respects.
In Bayesian blocks a fitness function that depends on the location and number of change points is optimized to find the correct number and placement of them. The total fitness function is the sum of the fitness function of each block. The choice of fitness function will depend on the type of data and could for unbinned event data be the log-likelihood of a Poisson distribution[1]. This fitness function has the form:

$$f(B_i) = ln(L_i(\lambda)) = N_i ln(\lambda) - \lambda T_i \tag{1}$$

where $f(B_i)$ is the fitness function of block i, $L_i(\lambda)$ is the likelihood of block i for a given value of λ, λ is the constant value of the block (it doesn't have a subscript since it's the variable), N_i is the number of events in the block and T_i is the width of the block. Inserting the value of λ that maximizes this fitness function gives:

$$f(B_i) = N_i(ln(N_i) - ln(T_i)) \tag{2}$$

The fitness function can then be maximized by varying the number of events in each block and the width of each block. The total fitness $F_{total} = \sum_{i=0}^{K} f(B_i)$ (for K blocks) is calculated $n-1$ times for each of n iterations, where the number of iterations is equal to the number of blocks (n), so the computation time is $\mathcal{O}(N^2)$ rather than $\mathcal{O}(2^N)$.
For a histogram to show a useful nonparametric estimate of the pdf of the data, the number of events in each bin should be high, which means that $N_{bins} \ll N$. To obtain this one could include a prior that favors fewer bins. A choice for such a prior could be the geometric function:

$$P(N_b) = P_0 \gamma^{N_{bins}} \tag{3}$$

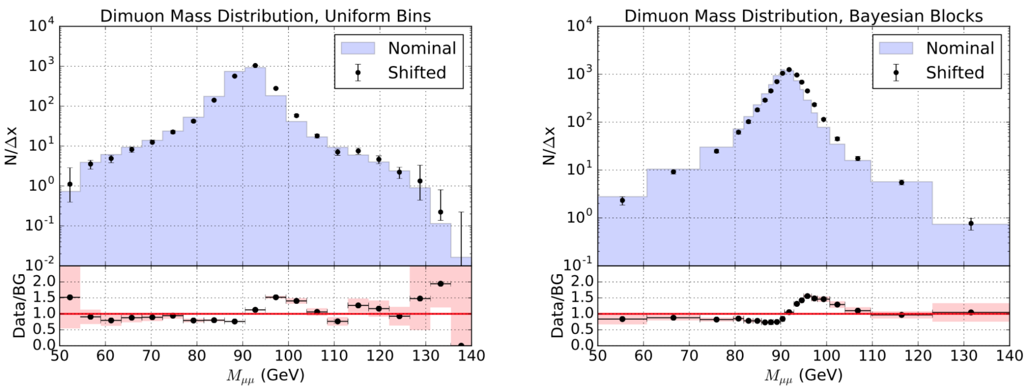
where P_0 is a normalization constant and γ is a constant. Alternatively to setting the value of γ, the prior can be determined by choosing a false positive rate for change points and simulating data. Through these simulations the correct gamma for the given false positive rate can be achieved. Too few change points will mean that important changes will be overlooked. Too many will mean that we will see more changes that are just based on fluctuations.
An advantage of Bayesian Blocks, it that the algorithm can center the peak of a signal in a bin, so that it will

[1] The derivation of this can be found in [2] and is applied to Bayesian Blocks in [3], p. 31.

be as clear as possible compared to the background.
Bayesian blocks are better if you are interested in peaks than tails, as tails have few entries and therefore do not change much and have few change points.
For the same reason, Bayesian blocks work better the more data you have, as the changes can be too small to make change points in small data sets.

3 Applications

An example of an application of the Bayesian blocks is given in [1] and shown here in figure 1. The advantage of the Bayesian Blocks is clear: The peak, which contains lots of rapidly changing data, has many bins so that it does not unnecessarily lose resolution. The tails of the distribution that has little, near-constant data have much broader bins, which means that the average value of these regions becomes much more certain than in the case of the uniform bins. The bottom of the figure shows the residuals from a comparison between data and a simulation where the calibration of a measured quantity was shifted. This produces a characteristic S-shape which becomes much clearer and easy to detect with the Bayesian blocks, because of the superior resolution of the peak.
Another example is in figure 2, where a hybrid binning method is used to isolate a signal into few bins. It works by calculating the change points first of the background and then of the signal, and then combining the two, such that the change points from the signal dominates in the region where the signal is present.



(a) Fixed-width binning.

(b) Bayesian Block binning.

Figure 1: Comparison of Bayesian blocks (right) and uniform bins (left) for simulated Drell-Yan distributions. Figure borrowed from [1].

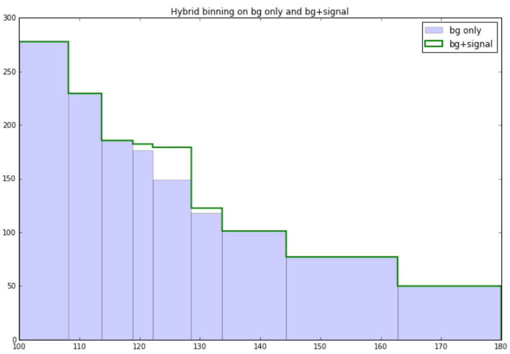


Figure 2: Hybrid binning applied to isolate the events of a signal in almost only one bin. On the x-axis is some measured quantity and on the y-axis, events divided by bin width. Figure borrowed from [1].

References

[1] S. B. Brian Pollack and M. Schmitt, *Bayesian Blocks in High Energy Physics: Better Binning made easy!*, (2017).

[2] W. Cash, *Parameter Estimation in Astronomy Through Application of the Likelihood Ratio*, (1979).

[3] J. D. Scargle, *Studies in Astronomical Time Series Analysis: VI. Bayesian Block Representations*, (2012).