Bayesian Blocks

Better binning made easy

Ida Stoustrup and Sofie H. Bruun
Niels Bohr Institute

Table of Contents

Histograms

Motivations

- Non-parametric estimate of a PDF
- Illustrating data distributions
- Data size reduction

Common Binning

Uniform bins

- The amount and variation of data often changes rapidly throughout a dataset.
- Bins in areas with lots of changing data will be too big and lower precision.
- Bins in areas with little data will have large uncertainties

Number of bins is subjective

- Risk of bias
- No quality/reasonableness guarantied
- Requires human input

Bayesian Blocks

The algorithm

- A fitness function is optimized by varying the number of blocks K and the block width T_i of each block.

$$F_{total} = \sum_{i=0}^{K} f(B_i)$$

- Given unbinned data the fitness function could be the log-likelihood of a Poisson distribution: [3] [2]

$$f(B_i) = ln(L_i(\lambda)) = N_i ln(\lambda) - \lambda T_i$$

Bayesian Blocks

- Iterates for each block, storing the best values left-to-right.
- This means the computational time is $\mathcal{O}(N^2)$ rather than $\mathcal{O}(2^N)$.

Prior

- Prior should prefer few bins, since often $N_{bins} \ll N$. Could be:

$$P(N_b) = P_0 \gamma^{N_{bins}} \tag{1}$$

- Setting γ adjusts the false positive rate for change points.

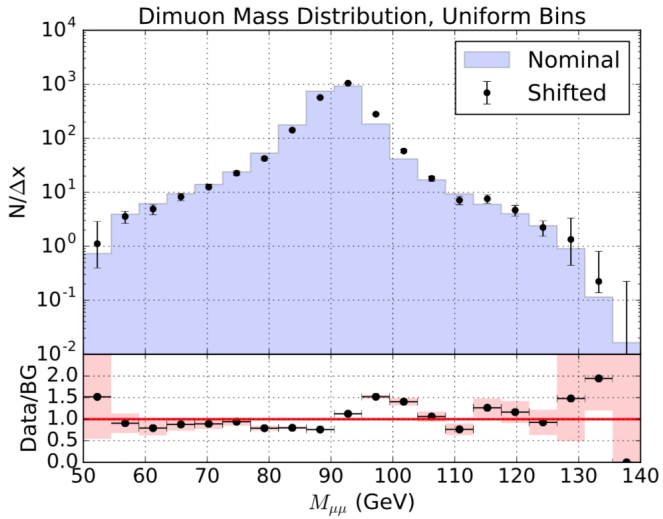
Advantages

- Objective
- Visual quality
- Statistical quality - better fits
- Adjustable false positive rate for change points
- Easy to automatically make reasonable histograms e.g. for compressing data

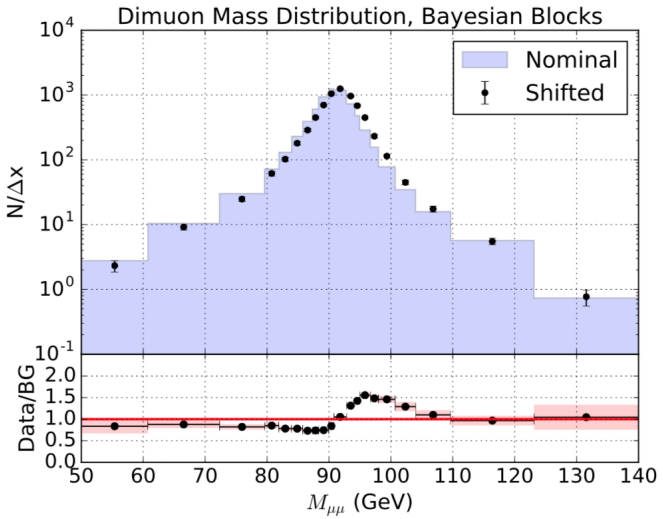
Challenges

- Not always enough change points when there are few data points.
- Not suited for examining tails as these will have coarse binning.
- When the binning is not done correctly for a signal on a background, a hybrid method can be used, but it requires separated data/simulations of the signal and the background.

Applications I



(a) Fixed-width binning.

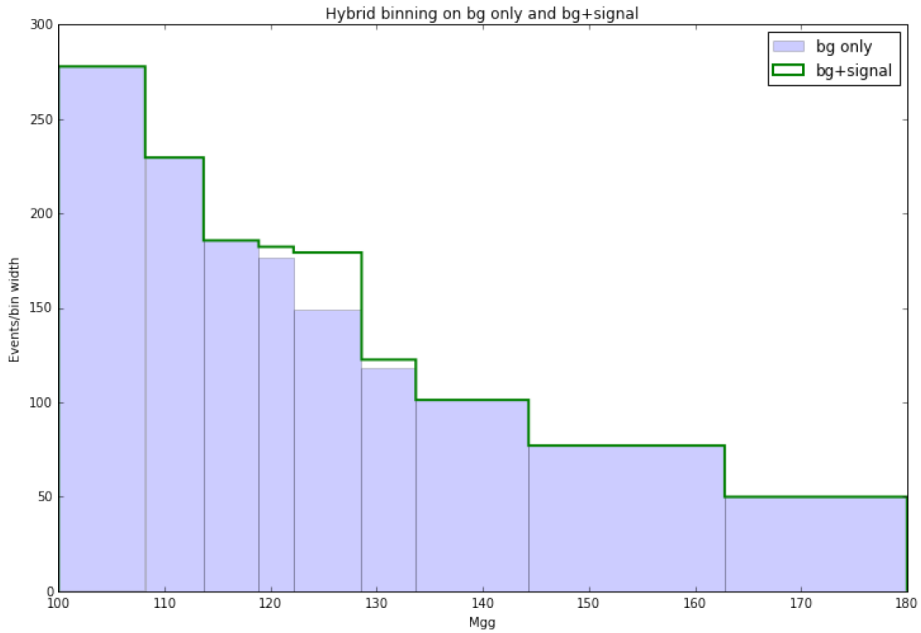


(b) Bayesian Block binning.

[1]

- Bayesian blocks give much higher resolution around the peak
- Bayesian Blocks make large bins in areas with little data, so that the uncertainty of these areas is smaller.

Applications II



[1]

- A hybrid method collects the influence of a signal into few bins
- This makes the signal easier to discern from the background.

Conclusion

- Bayesian Blocks provide a method for objective binning in histograms.
- They can improve the visual and statistical quality of histograms in many cases, but have some limitations.

Referencer I

[1] Brian Pollack, Saptaparna Bhattacharya, and Michael Schmitt.
Bayesian Blocks in High Energy Physics: Better Binning made easy!
2017.

[2] Webster Cash.
Parameter Estimation in Astronomy Through Application of the Likelihood Ratio.
1979.

[3] Jeffrey D. Scargle.
Studies in Astronomical Time Series Analysis: VI. Bayesian Block Representations.
2012.