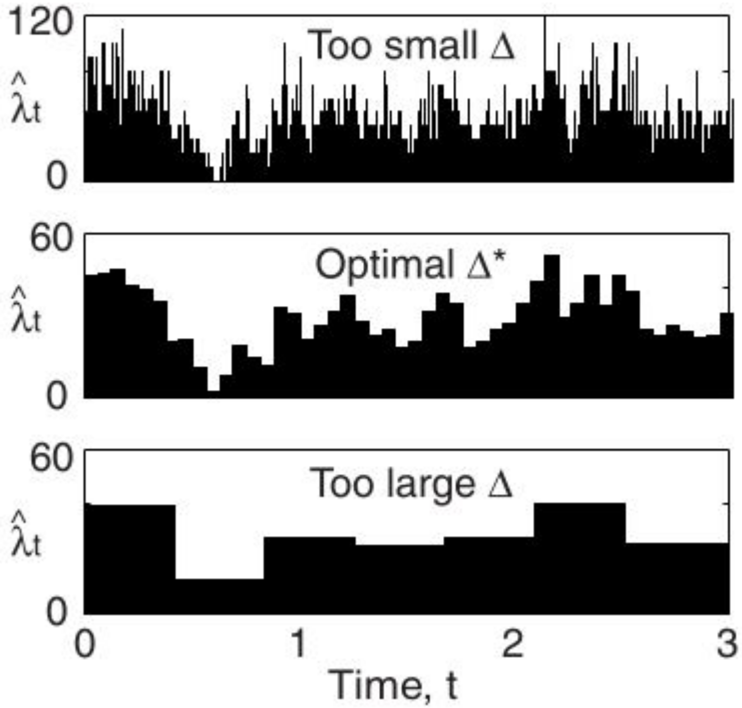
Optimal binning of time histograms using variable bin size

IFAE Summer Fellowship 2022

Gerard Marcet Carbonell - Pizza seminar - September 21st 2022

Why do we care about bin size

- Loss of information
- Introduction of bias



Previous work on the subject

A Method for Selecting the Bin Size of a Time Histogram

Hideaki Shimazaki
shimazaki@ton.scphys.kyoto-u.ac.jp
Shigeru Shinomoto
shinomoto@scphys.kyoto-u.ac.jp
Department of Physics, Kyoto University, Kyoto 606-8502, Japan

Defining a cost function - MISE approach

$$\mathrm{MISE} \equiv \frac{1}{T}\int_0^T E\,(\hat{\lambda}_t - \lambda_t)^2\,dt,$$

$$\hat{\theta} = \frac{k}{\Delta}$$

Defining a cost function - MISE approach

$$C_n(\Delta) \equiv \mathrm{MISE} - \frac{1}{T}\int_0^T (\lambda_t - \langle\theta\rangle)^2\, dt$$

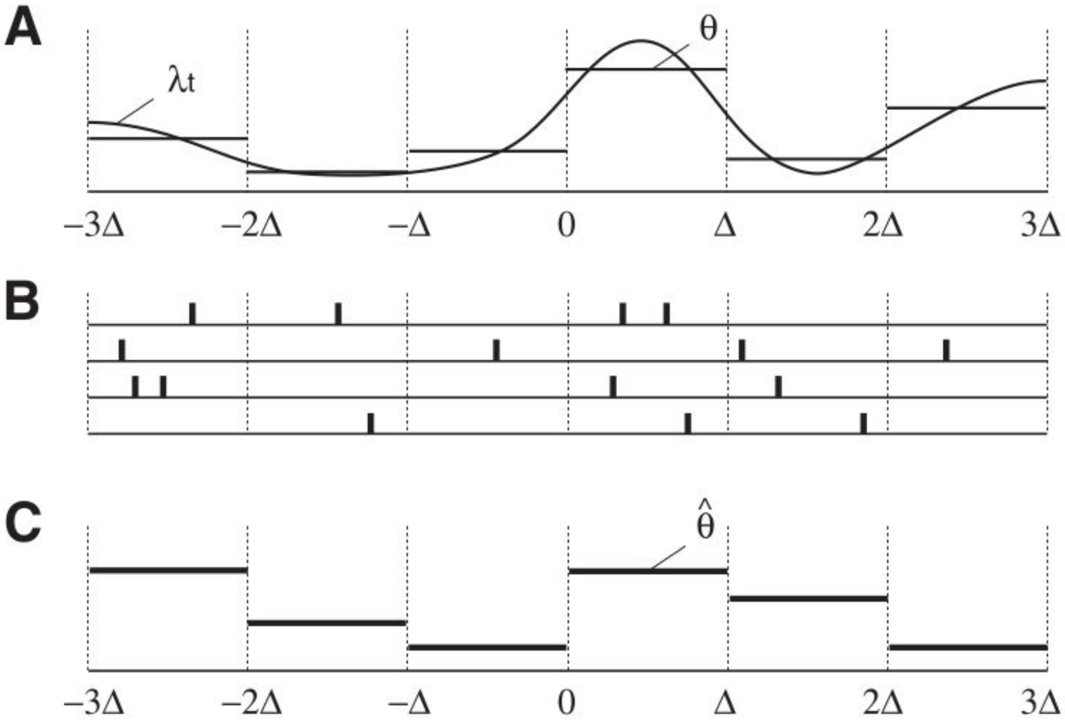
$$= \langle E(\hat{\theta} - \theta)^2\rangle - \langle(\theta - \langle\theta\rangle)^2\rangle.$$

$$C_n(\Delta) = \frac{2\bar{k} - v}{(\quad \Delta)^2}.$$

$$\theta = \frac{1}{\Delta}\int_0^\Delta \lambda_t\, dt.$$

$$E\hat{\theta} = \theta$$

Defining a cost function - MISE approach



Defining a cost function with variable bins

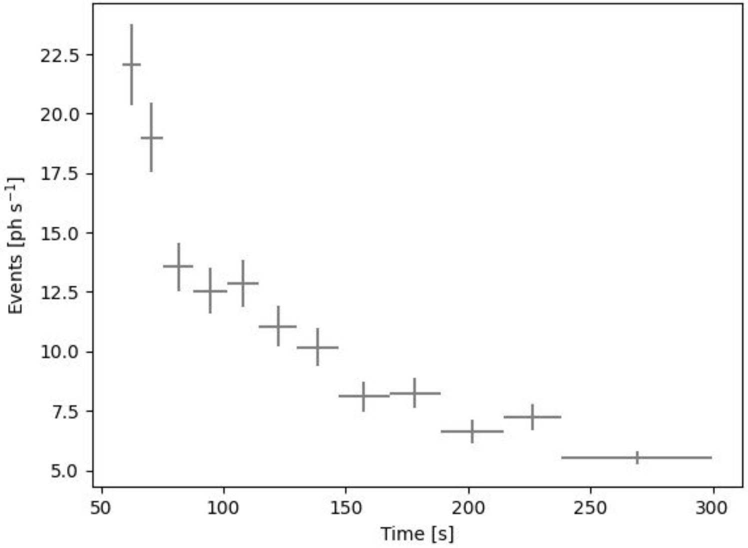
$$C(\text{lightcurve}) = \frac{2}{T}\sum_{i=1}^{N}\frac{k_i}{\Delta_i} - \sum_{i=1}^{N}\frac{\Delta_i}{T}\left(\frac{k_i}{\Delta_i} - \frac{K}{T}\right)^2$$

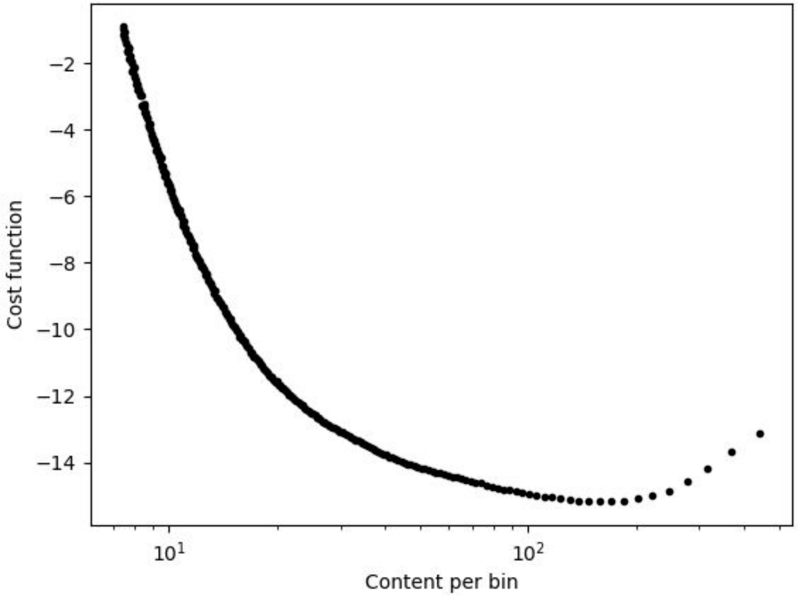
$$C_n(\Delta) = \frac{2\bar{k} - v}{(\quad\Delta)^2}.$$

← Shimazaki and Shinomoto's cost function

Application of the method: content constant binning

170 events/bin

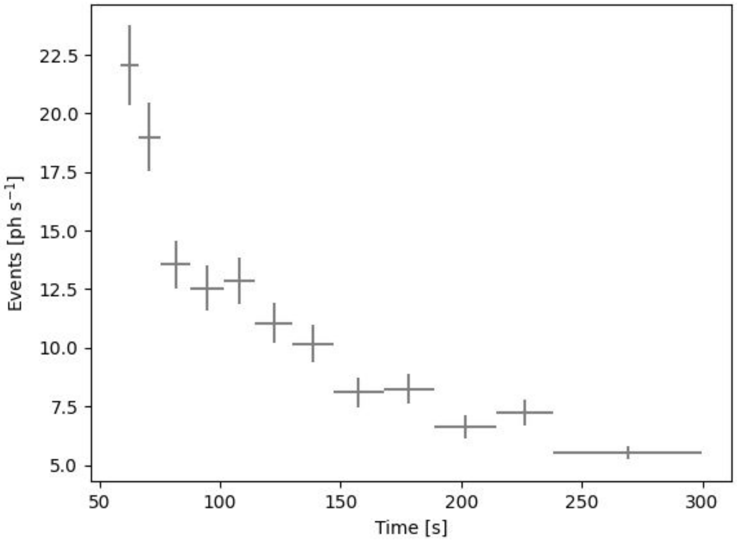




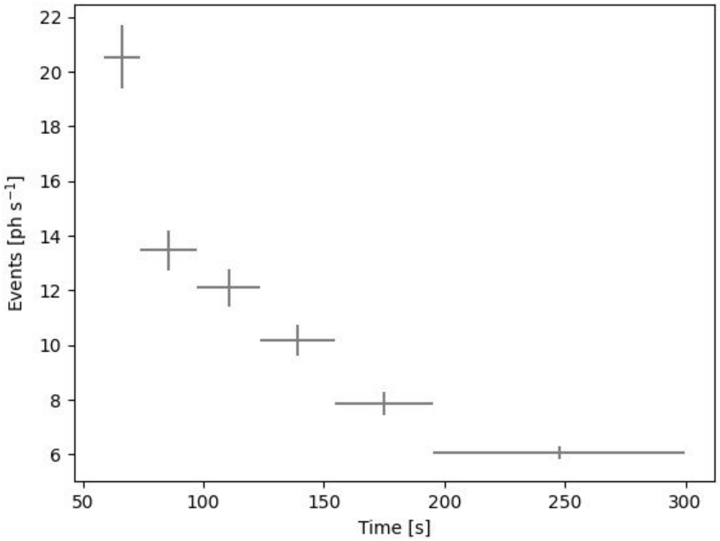
Dataset: MAGIC Collaboration. Teraelectronvolt emission from the γ-ray burst GRB 190114C. *Nature* **575,** 455–458 (2019).

Application of the method: content constant binning

Optimal content per bin
(170 events/bin)



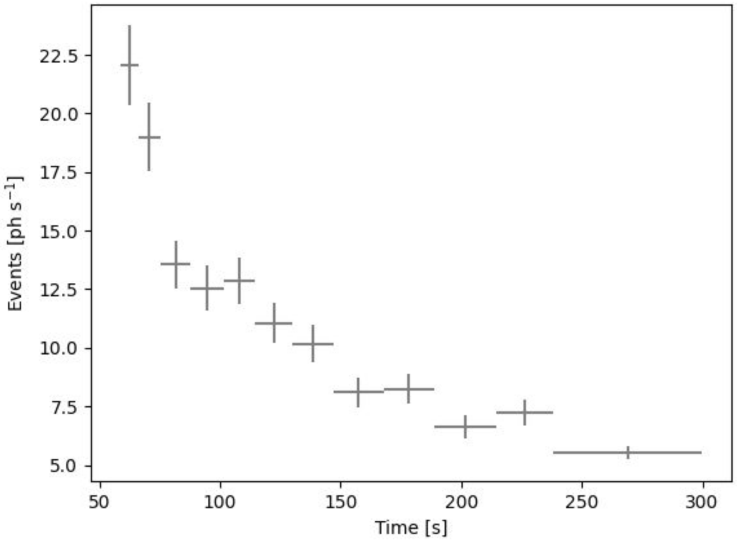
Too many events per bin
(317 events/bin)



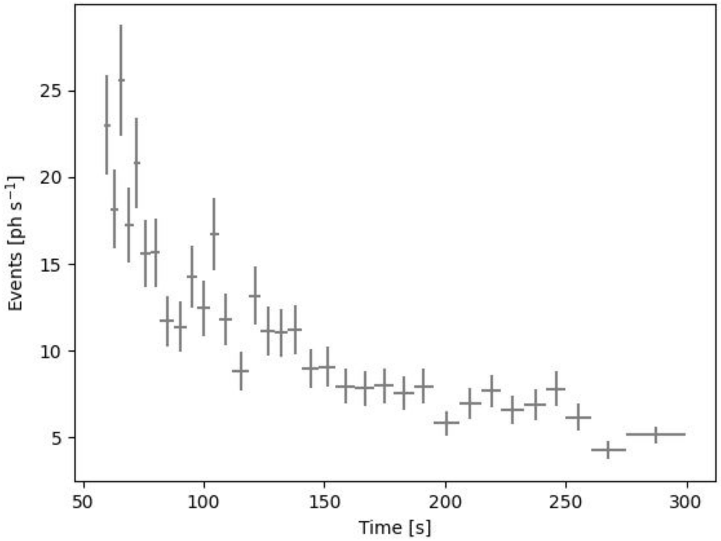
Dataset: MAGIC Collaboration. Teraelectronvolt emission from the γ-ray burst GRB 190114C. *Nature* **575,** 455–458 (2019).

Application of the method: content constant binning

Optimal content per bin
(170 events/bin)



Too few events per bin
(63 events/bin)



Dataset: MAGIC Collaboration. Teraelectronvolt emission from the γ-ray burst GRB 190114C. *Nature* **575,** 455–458 (2019).

LST4 Camera integration