Detection of Non-Gaussian Data in Mark5 (M5B) Files

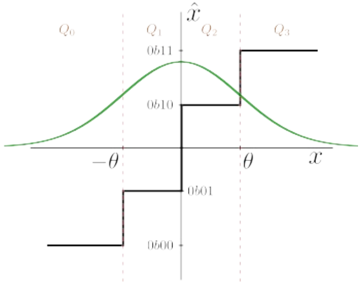
The radio astronomy data stored in M5B files are Gaussian white noise. Channel data streams are 4-level quantized to have only 2 binary digits per sample.

Problem: detect damaged fragments with non-Gaussian distribution. Do it before the correlation on GPU because of high data volumes.

Solution: estimating quantization thresholds from data may indicate non-Gaussianity

1. Data quantization uses 3 thresholds, -*θ*, *0*, +*θ*, so the only statistics are quantiles, Q_0, Q_1, Q_2, and Q_3, the quantities of data that fall into intervals (including infinities) between them.

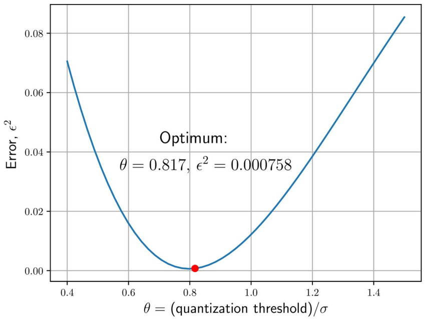
4-Level Quantization Pattern



3. The q-thresholds are not strictly specified and can be varied during operation. It is possible to find the optimal *θ* that minimizes the error between the standard Normal quantiles and those from data:

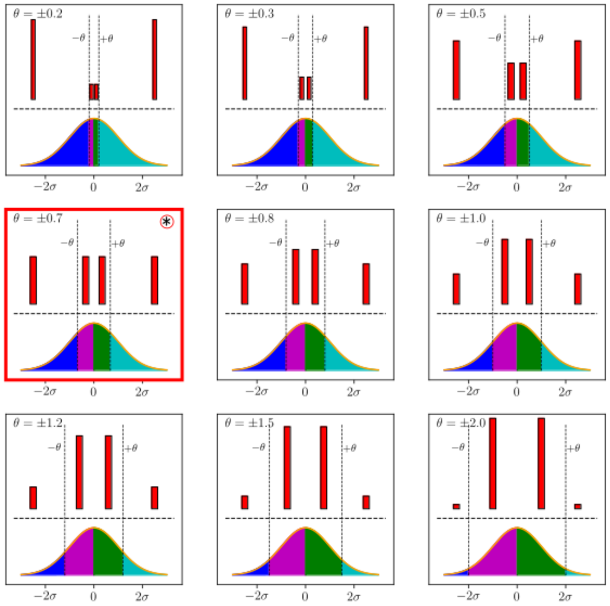
$$\varepsilon^2 = \sum_{i=0}^{3} (\hat{Q}_i - Q_i)^2$$

Error b/w M5B Data and Normal Quantiles for Tresholds [0.4 .. 1.5]



2. Below one can see the limitations on q-threshold values. The four red bars show relative values or the standard Normal probability density function.

Normal PDF Quantiles Separated by $-\infty, -\theta, 0, +\theta, +\infty$

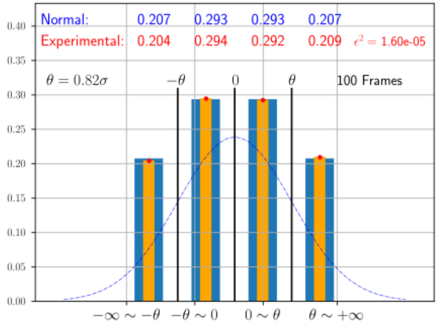


* Quantization Thresholds $\pm\theta$: MUST be $|\theta| > 0.6745\,\sigma$

4. Two examples of finding the optimal q-threshold. The bars give values of quantiles: blue – Normal, yellow – M5B data.

Good data (Normal distribution)

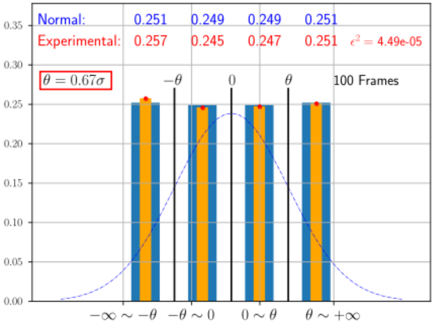
M5B Data vs Normal in Quantiles between $-\theta, 0,$ and $+\theta$



File rd1910_wz_268-1811.m5b

GARBAGE (Uniform distribution?)

M5B Data vs Normal in Quantiles between $-\theta, 0,$ and $+\theta$

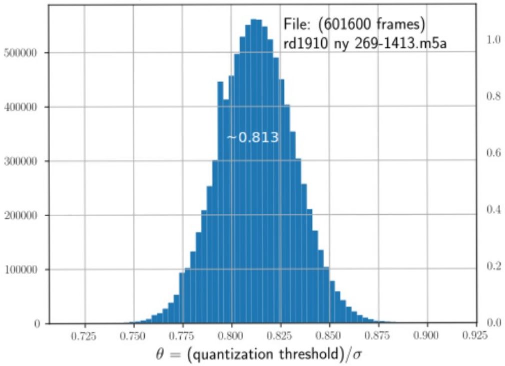


File rd1903_ft_100-0950.m5b

5. **SOLUTION**. Statistics of optimal q-threshold estimates. One threshold is found for each 2500 samples in one channel.

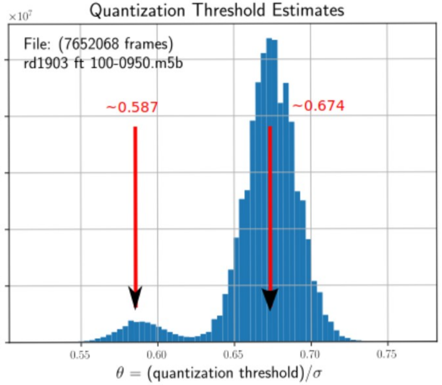
Good data (θ > 0.67)

Quantization Threshold Estimates



Bad, non-Gaussian data (θ > 0.67)

Quantization Threshold Estimates


$\times 10^7$
File: (7652068 frames)
rd1903 ft 100-0950.m5b
~0.587
~0.674
θ = (quantization threshold)/σ