A new adaptive density estimator for 1-D point processes: the (not so) fine details of the GRBs' redshift distribution

Zsolt Bagoly[1,2] and István Horváth[1]

[1] University of Public Service, Budapest, Hungary [2] Eötvös University, Budapest, Hungary

Introduction

We've developed a density estimator for 1-D point processes. It is based on the continous smooth function approximation, while information in the global density distribution is also applied. Using MC simulations we've determined the optimal parameters for low count distributions like the GRBs' redshift data. The results detailing the GRBs' redshift distribution also discussed briefly.

Observational data indicate that a substantial fraction of GRBs originate from high redshifts, often $z > 5$, suggesting that these events were more prevalent in the early universe. The observed variation in GRB redshifts is influenced by multiple factors, including the sensitivity of detection instruments and observational biases. GRBs at very high redshifts may be underrepresented due to technological limitations and the challenges associated with follow-up observations. Recently Ghirlanda et al. studied the evolution of GRB formation rates and luminosity functions by analyzing extensive datasets from Fermi, Swift, and CGRO. The data shows that the LGRB formation rate increases steeply with redshift up to $z \approx 3$ before declining, due to the preference for low-metallicity environments.

In this study we used a comprehensive database of GRBs with spectroscopic redshifts, primarily detected by NASA's Swift and Fermi missions. We utilize the same dataset as referenced in Horvath et al., 2022, Bagoly et al., 2023, and Horvath et al, 2024. The redshifts sourced from the Gamma-Ray Burst Online Index (GRBOX), GCN reports and J Greiner's compilation webpage. Our dataset includes observations through August 31, 2022, with 542 GRBs. For more details on the dataset's, refer to (Horvath et al, MNRAS 2024).

Model functions and estimators

We've generated three model distribution from the data. For density estimations the kernel based methods are popular. The fixed-bandwidth kernel density estimate uses a kernel with a given bandwidth to smooth data. The bandwidth controls the spread of each kernel, with larger values creating wider, smoother estimates and smaller values producing narrower, more focused ones. Adaptive bandwidth estimation allows kernel bandwidths to vary based on local data density, improving performance over fixed-bandwidth methods. The bandwidth is defined as a function of data coordinates, a popular choice is the Abramson's method (Annals of Statistics,10(4):1217â1223, 1982.) which sets it inversely proportional to the square root of the target density. This reduces smoothing in high-density areas for finer detail and increases it in low-density regions, and it is combined with edge correction for density estimation. Using the CRAN R program language we used the npreg package to generate a spline based density estimator (Spline Model). The KernSmooth function from the same package was used to generate an optimal fixed bandwith estimator ($bw = 3841/hMpc$) (Smooth Model). The spatstat.univar library's densityAdaptiveKernel method to generate an adaptive bandwidth estimator - but here we reduced the optimal smoothing size to 20% of its optimal value to generate an artificial rapidly changing model (Adaptive Model).

Contact: bagoly.zsolt@uni-nke.hu

Acknowledgements: authors thank the Hungarian programs TKP2021-NVA-16 and TKP2021-NKTA-64 for support.

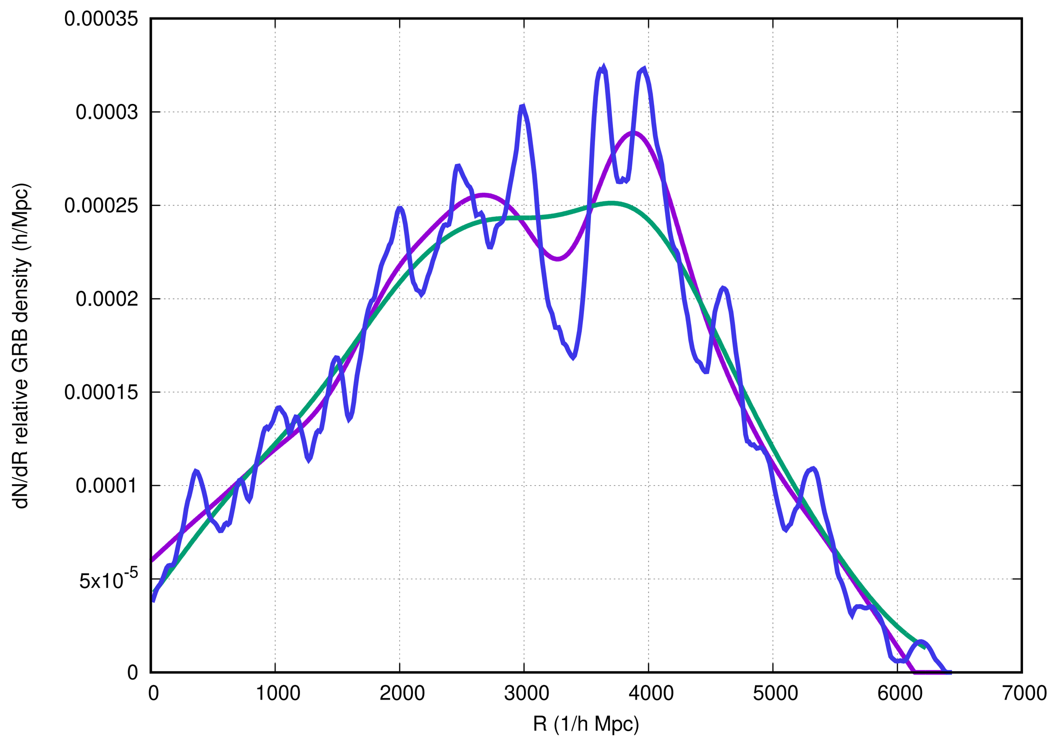

Fig. 1: The three model function used in the test, all generated from the real data: the Spline(pink), the Smooth(green) and the high-frequency Adaptive Model(blue).

The Continuous-Function Estimator

The Continuous-Function Estimator is a novel approach to estimating density functions like the two-point correlation function (Storey-Fisher and Hogg, 2021). Unlike conventional methods that rely on hard-edged separation bins, this estimator utilizes a continuous representation by projecting data onto a set of user-defined basis functions. It allows for a more flexible and accurate representation of a function, as it can adapt to the expected shape of the function. The Continuous-Function Estimator is inspired by least-squares fitting,

The Continuous-Function Estimator replaces the traditional binning with cubic splines as basis functions of $< sp_i|$. The cubic spline basis functions are selected for their ability to provide a smooth and accurate fit, as they maintain continuity and differentiability. It projects the data onto the spline basis: here we have 542 R_j distance values in the for of Dirac-deltas. Projecting these $p(r_j)$ to the $< sp_i|$ basis we obtain the $c_{ij} =< sp_i|p(r_j) >$ values and one can reconstruct the density function in the form of $p(r) = c_{ij}|sp_i >$. We used a $1\ (1/hMpc)$ grid, up to 6400 $(1/hMpc)$. The cubic splines have a compact support, the bandwidth was fixed for 96 bin, with a forresponding FWHM width of $138.4\ 1/hMpc$ for all catalogue sizes.

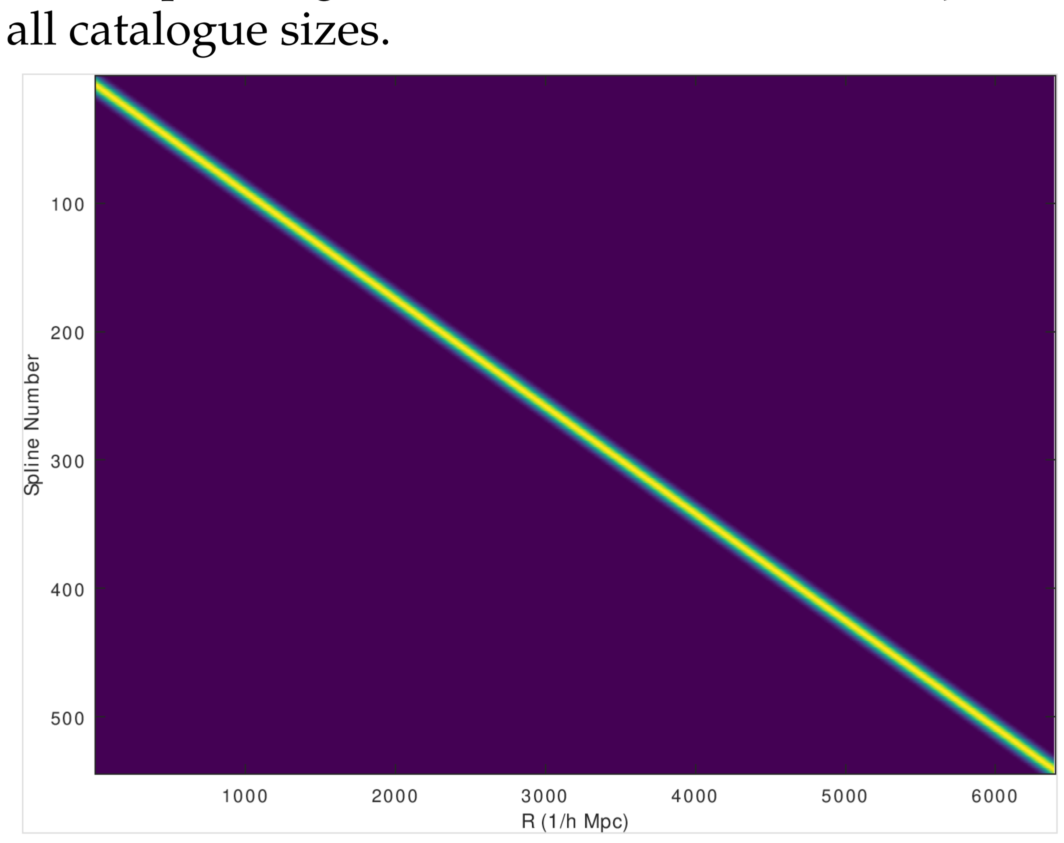

Fig. 2: The cubic splines used for the Continuous-Function Estimator. The i^{th} line in the image is the $|sp_i>$ cubic spline base function.

To generate random catalogues we've sampled the model functions 100 times using event numbers of $[128, 256, 512, 1024, 2048]$. To compare the efficiency of the Continuous-Function Estimator we've calculated the mean integrated squared error (MISE) value between the model function and the density estimation produced by the random catalogues. The catlog estimation was repeated with the R CRAN spatstat.univar library's densityAdaptiveKernel method with the optimal kernel sizes, and the corresponding MISE was calculated as well.

Results

On Figs. 3-5. the MISE values for the two methods for different catalogue sizes and different original models can be seen. One can observe that the Continuous-Function Estimator fits the data better than the densityAdaptiveKernel method above $N \approx 512$ for the Spline and Smooth Model, and it is better for densityAdaptiveKernel fitting of the Adaptive Model above $N \approx 850$.

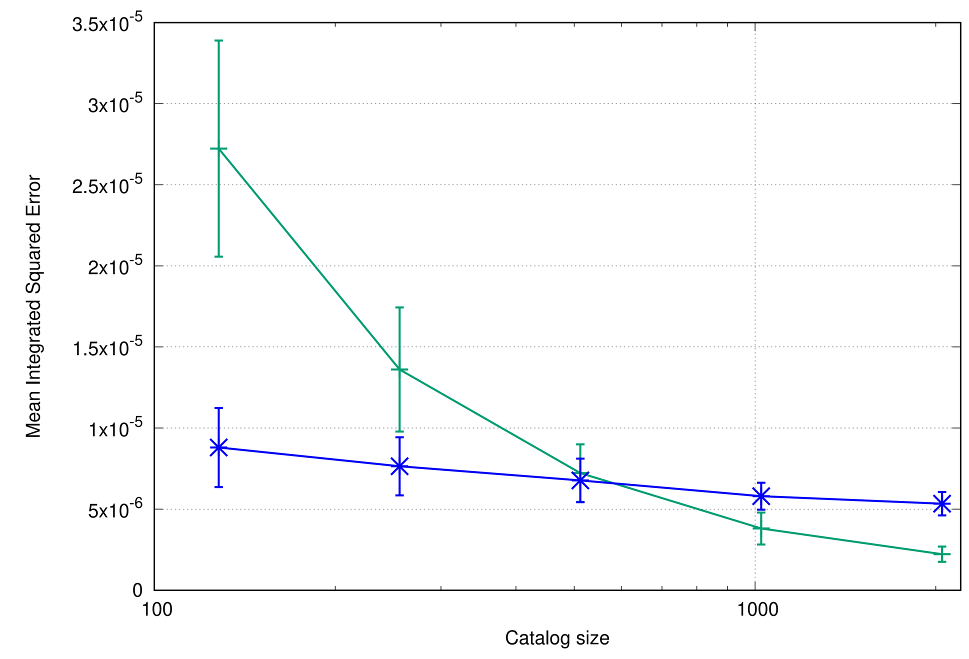

Fig. 3: The MISE values for the Continuous-Function Estimator (green) and the densityAdaptiveKernel (blue) method for the Spline model

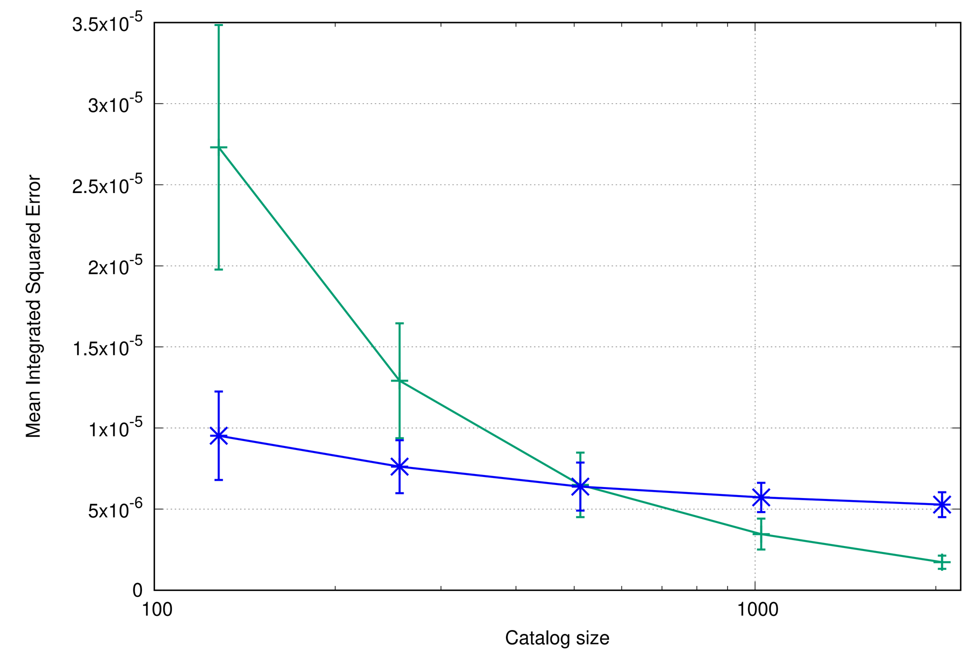

Fig. 4: The MISE values for the Continuous-Function Estimator (green) and the densityAdaptiveKernel (blue) method for the Smooth model

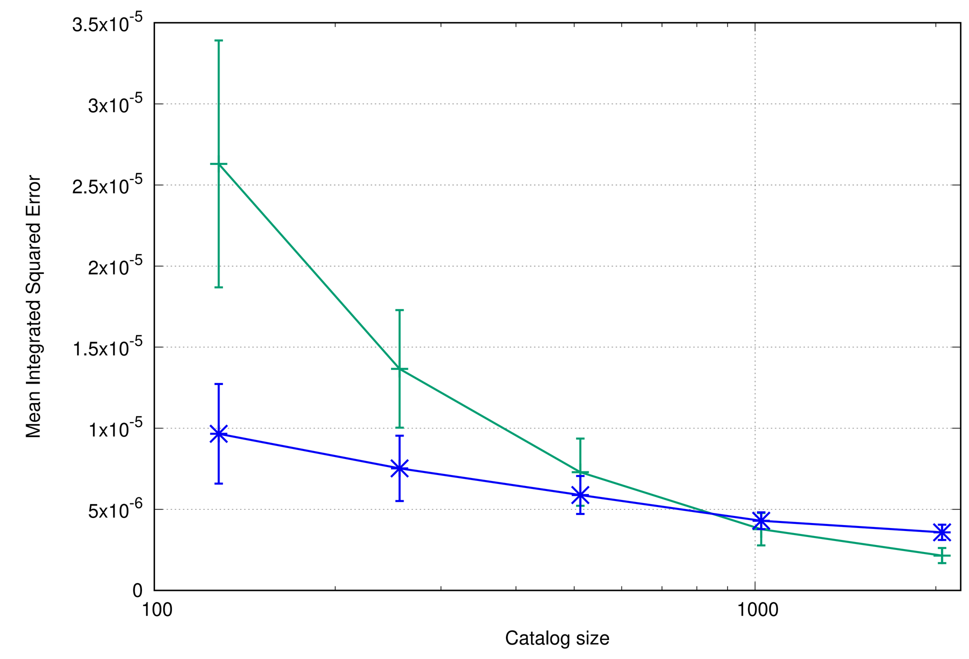

Fig. 5: The MISE values for the Continuous-Function Estimator (green) and the densityAdaptiveKernel (blue) method for the Adaptive model

The Poisson noise makes the two estimation methods quite similar in efficiency around $N = 542$, so here we can use either to smooth the GRB density. On Fig. 6 the GRB rate was plotted in scaled units of $h^3/Mpc^3/yr$ vs. the $1 + z$ redshift. The well-known maximum at $z \approx 2.6$ is imminent.

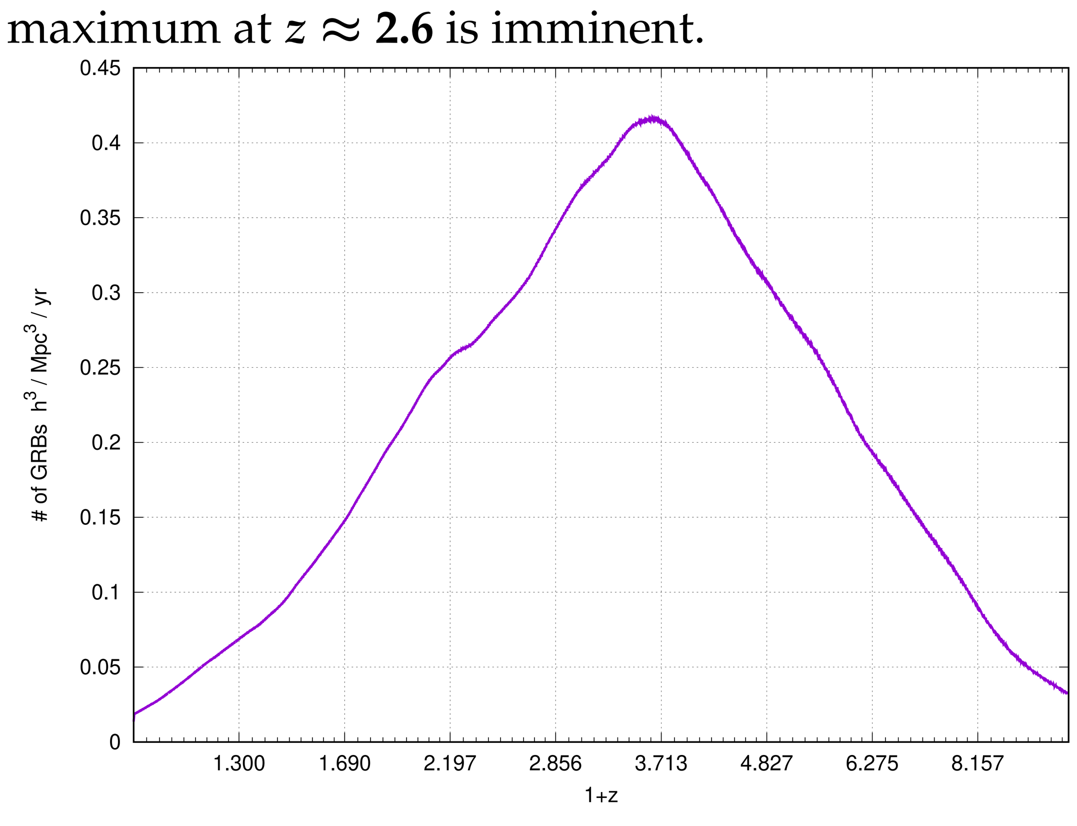

Fig. 6: The GRB rate in scaled units of $h^3/Mpc^3/yr$ as a function of the $1 + z$ redshift. The well-known maximum at $z \approx 2.6$ is visible.