Power-Law Estimation via MPS

Theory and Applications

Joshua D. Ingram[1, 2] Vinay Kashyap [1] Xufei Wang[3]
Xiao-Li Meng[3] Bernhard Klingenberg[2]

[1]High Energy Astrophysics Division
Center for Astrophysics | Harvard & Smithsonian

[2]Department of Data Science
New College of Florida

[3]Department of Statistics
Harvard University

RISE-CHASC Workshop, August 2022

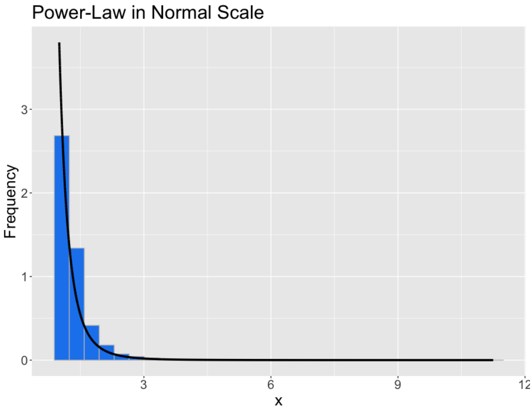
Table of Contents

Table of Contents

Power-Law Distributions in Solar Flares

Solar flare properties have been found to follow power-law distributions (Lu & Hamilton 1991, Wheatland & Litvinenko 2002, Clauset et al. 2009). These properties include total energy, peak flux, waiting times, etc.

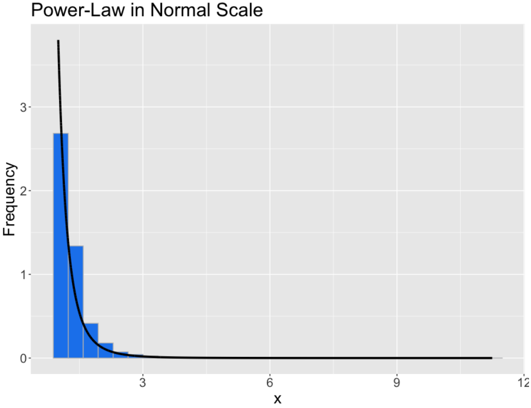
- The functional form of a power-law is given by $f(x) = cx^{-\alpha}$.



Power-Law Distributions in Solar Flares

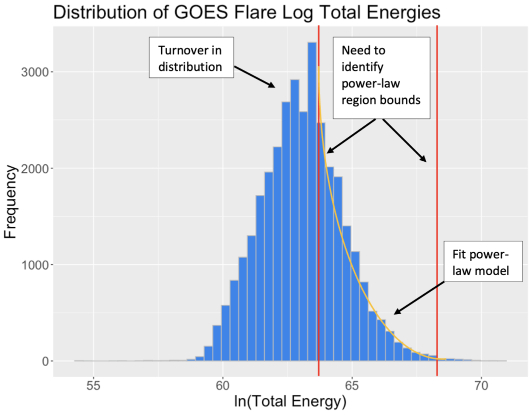
Solar flare properties have been found to follow power-law distributions (Lu & Hamilton 1991, Wheatland & Litvinenko 2002, Clauset et al. 2009). These properties include total energy, peak flux, waiting times, etc.

- The functional form of a power-law is given by $f(x) = cx^{-\alpha}$.
- In log scale, where $y = \log(x)$, it has the form $g(y) = ke^{-y(\alpha-1)}$.



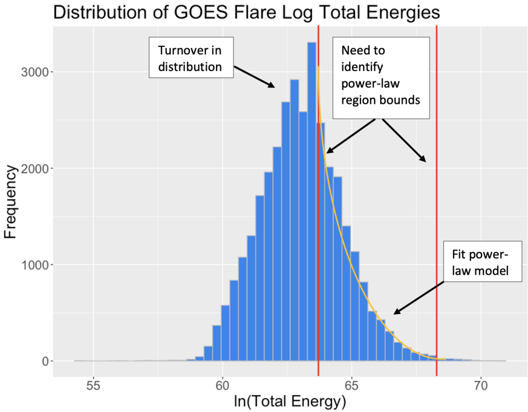
Turnover in the Distribution

- A power-law is only observed over a bounded range to the right of the mode.



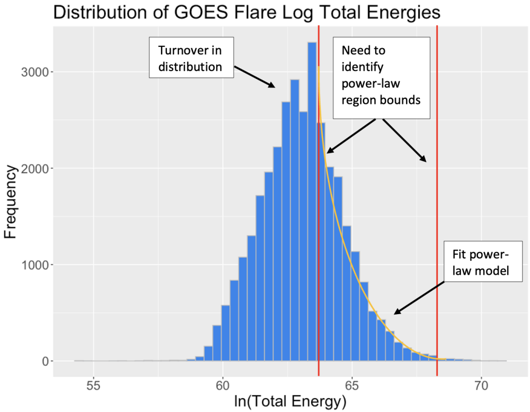
Turnover in the Distribution

- A power-law is only observed over a bounded range to the right of the mode.
- The turnover observed at lower values is the result of missing data caused by detection sensitivity.



Turnover in the Distribution

- A power-law is only observed over a bounded range to the right of the mode.
- The turnover observed at lower values is the result of missing data caused by detection sensitivity.
- The start of the power-law is typically visually determined, or by using the K-S statistic (see Clauset et al. 2009).



The Problem

How do we accurately determine the power-law region and estimate α?

Table of Contents

Maximum Product of Spacings Method

- For $X_1 < X_2 < \ldots < X_n$, fit a CDF that maximizes the product of spacings: $\prod_{i=0}^{n}\{F(X_{i+1}) - F(X_i)\}$.

Maximum Product of Spacings Method

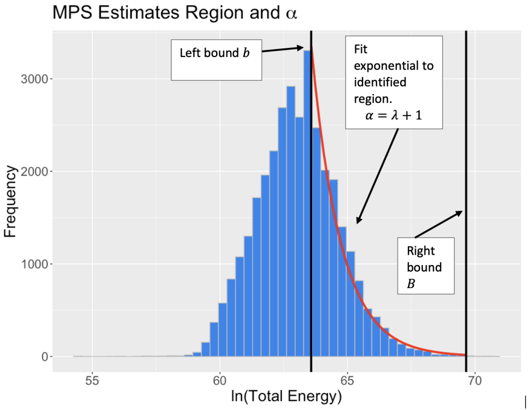
- For $X_1 < X_2 < \ldots < X_n$, fit a CDF that maximizes the product of spacings: $\prod_{i=0}^{n}\{F(X_{i+1}) - F(X_i)\}$.
- The method is very versatile. It can be parametric, non-parametric, or partially parametric.

Maximum Product of Spacings Method

- For $X_1 < X_2 < \ldots < X_n$, fit a CDF that maximizes the product of spacings: $\prod_{i=0}^{n} \{F(X_{i+1}) - F(X_i)\}$.
- The method is very versatile. It can be parametric, non-parametric, or partially parametric.
- We fit a parametric MPS model, and add a penalty term.

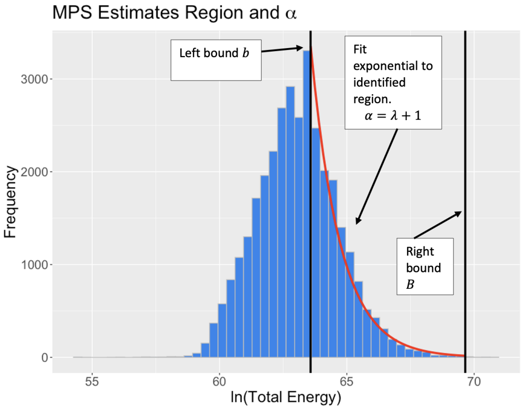
Maximum Product of Spacings Method

- Take the logarithm of the total energy. Power-law region becomes exponential.



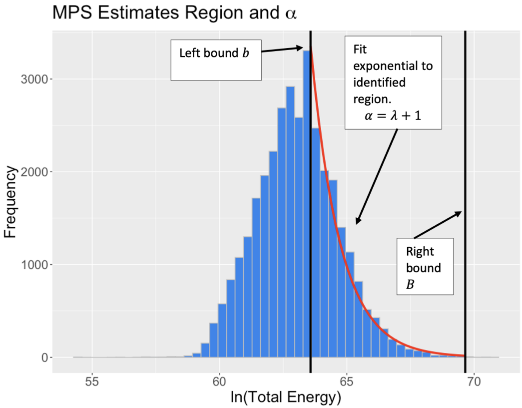
Maximum Product of Spacings Method

- Take the logarithm of the total energy. Power-law region becomes exponential.
- Within the region b and B, the data follows an exponential distribution with rate λ.



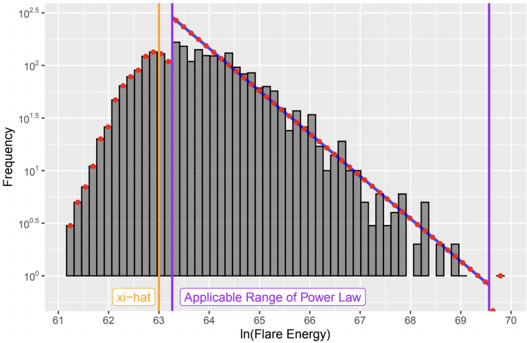
Maximum Product of Spacings Method

- Take the logarithm of the total energy. Power-law region becomes exponential.
- Within the region b and B, the data follows an exponential distribution with rate λ.
- The Maximum Product of Spacings method estimates these bounds and the rate parameter λ. $(\alpha = \lambda + 1)$



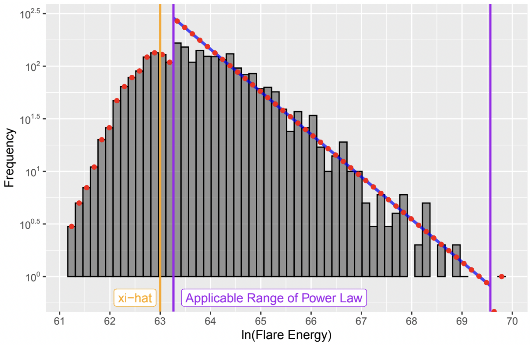
Alternative: Hybrid Parametric Multinomial Model

- Given k pre-specified energy bins, fit a multinomial model using bin probabilities.



Alternative: Hybrid Parametric Multinomial Model

- Given k pre-specified energy bins, fit a multinomial model using bin probabilities.
- Bin probabilities within the power-law region are parameterized by the power-law, outside are not.



Alternative: Hybrid Parametric Multinomial Model

- Given k pre-specified energy bins, fit a multinomial model using bin probabilities.
- Bin probabilities within the power-law region are parameterized by the power-law, outside are not.
- Maximize the likelihood function, estimate α and lower bound ξ.

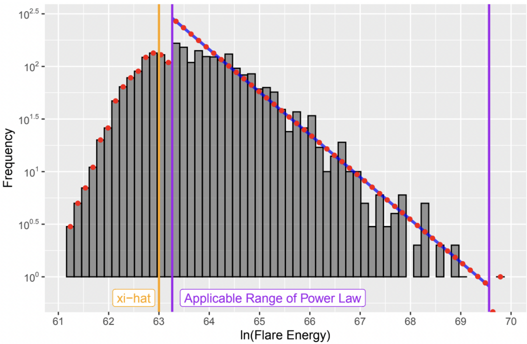
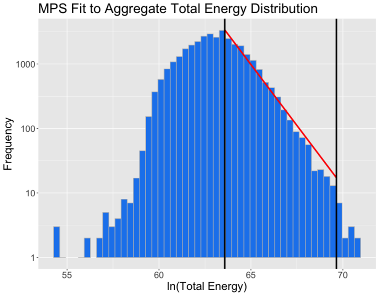


Table of Contents

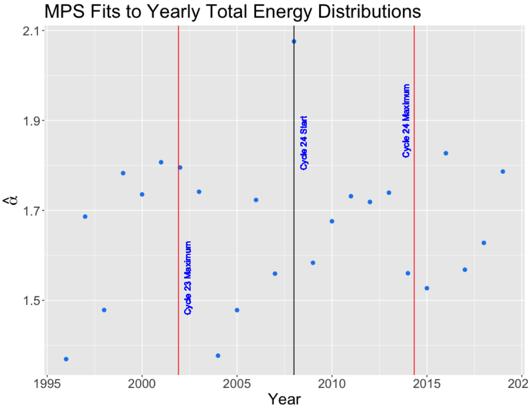
Total Energy - Aggregate and By-Cycle

Data	$\hat{\alpha}$	$\hat{b}$	$\hat{B}$	n
Aggregate	1.87	63.58	69.65	33445
Cycle 23	1.85	63.58	69.09	18457
Cycle 24	1.64	62.61	67.10	14988



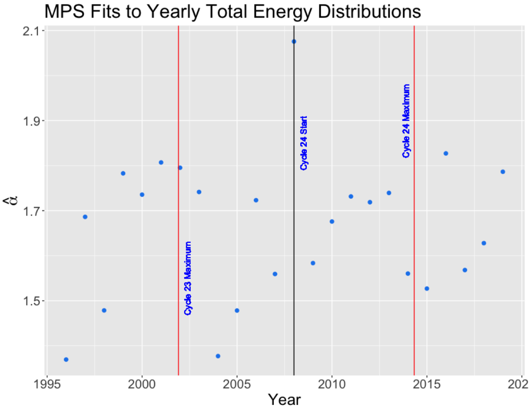
Total Energy - Yearly

- $\hat{\alpha}$ varies between 1.37 and 1.87 throughout the two solar cycles (outlier of $\hat{\alpha} = 2.07$ in 2008, where $n = 88$)



Total Energy - Yearly

- $\hat{\alpha}$ varies between 1.37 and 1.87 throughout the two solar cycles (outlier of $\hat{\alpha} = 2.07$ in 2008, where $n = 88$)
- No significant trend in $\hat{\alpha}$ within a solar cycle.



Summary

- We have further developed our model based on the MPS method to estimate the lower and upper bounds of the power-law region, and α.

Summary

- We have further developed our model based on the MPS method to estimate the lower and upper bounds of the power-law region, and α.
- For aggregate data, $\hat{\alpha} = 1.87$. Similar to previous estimates close to 1.8.

Summary

- We have further developed our model based on the MPS method to estimate the lower and upper bounds of the power-law region, and α.
- For aggregate data, $\hat{\alpha} = 1.87$. Similar to previous estimates close to 1.8.
- For yearly data, $\hat{\alpha}$ varies between 1.37 and 1.87. We identify no clear trend.

Moving Forward

- Further study performance of MPS through simulations.

Moving Forward

- Further study performance of MPS through simulations.
- Fit our power-law models to solar flare property distributions (total energy, peak flux, waiting time).

Moving Forward

- Further study performance of MPS through simulations.
- Fit our power-law models to solar flare property distributions (total energy, peak flux, waiting time).
- Compare power-law models and verify results.

References

R.C.H.Cheng and N.A.K. Amin
Estimating parameters in continuous univariate distributions with a shifted origin
Journal of the Royal Statistical Society, Series B. 45 (3): 394–403 (1983).

B. Ranneby
The Maximum Product of Spacings Method. An Estimation Method Related to the Maximum Likelihood Method
Scandinavian Journal of Statistics 11, 93-112 (1984).

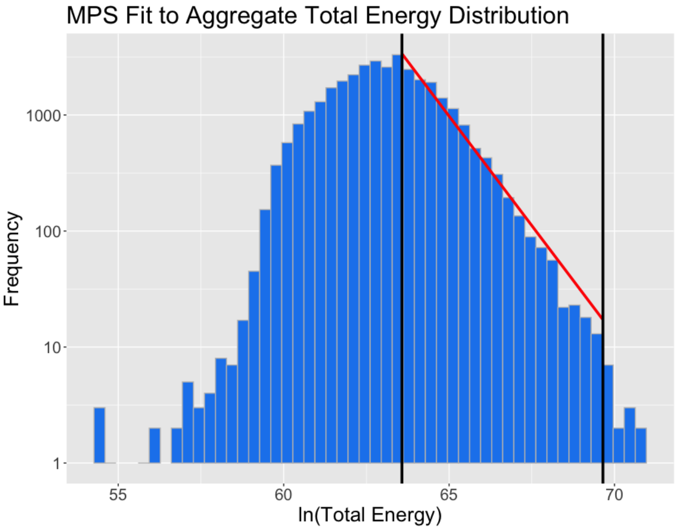
E. T. Lu and R. J. Hamilton
Avalanches and the Distribution of Solar Flares
Astrophysical Journal Letters v.380, p.L89 (1991).

References Continued...

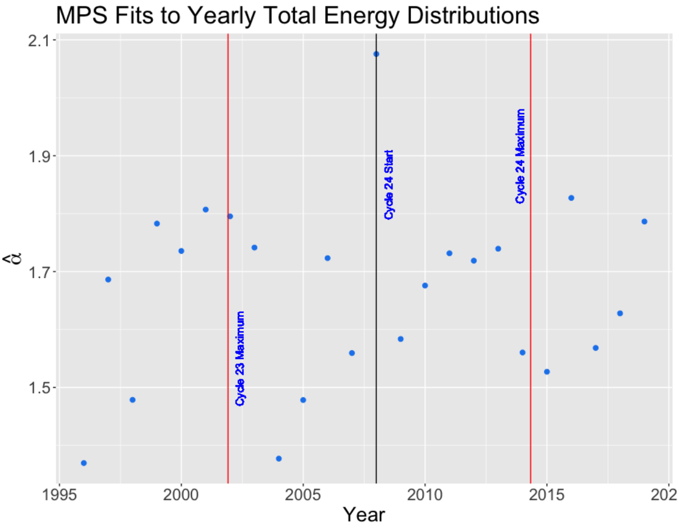
M. S. Wheatland and Y. E. Litvinenko
Understanding Solar Flare Waiting-Time Distributions
Solar Physics, v. 211, Issue 1, p. 255-274 (2002) .

A. Clauset, C. R. Shalizi, and M. E. J. Newman
Power-Law Distributions in Empirical Data
SIAM Review, vol. 51, no. 4, pp. 661-703 (2009).

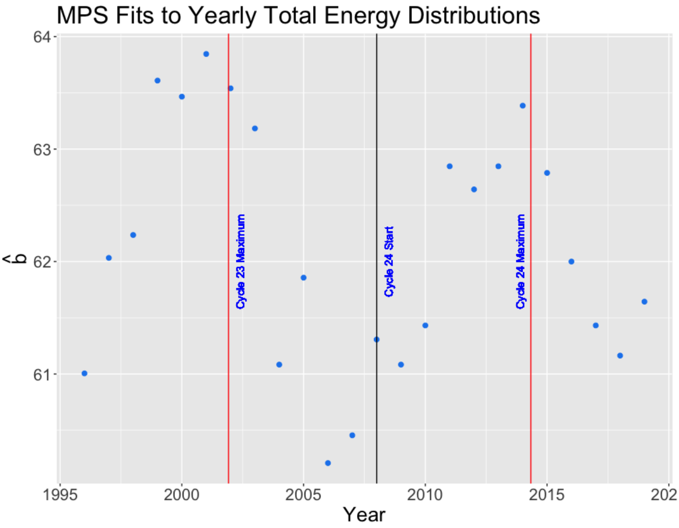
MPS Fit to Aggregate Total Energy



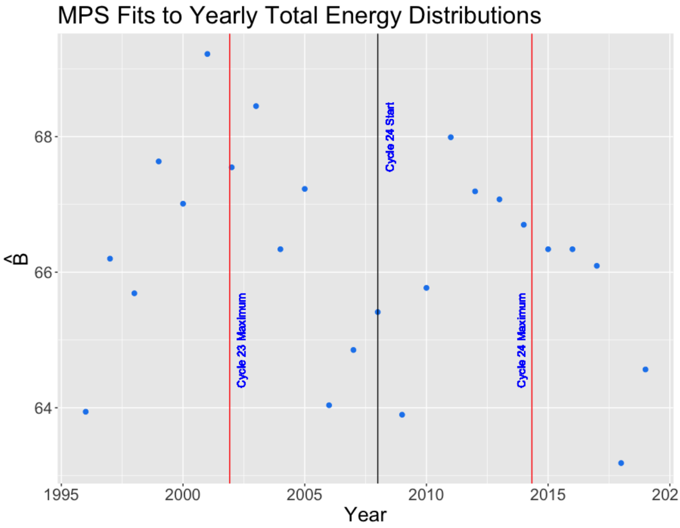
MPS α Fits to Yearly Total Energy



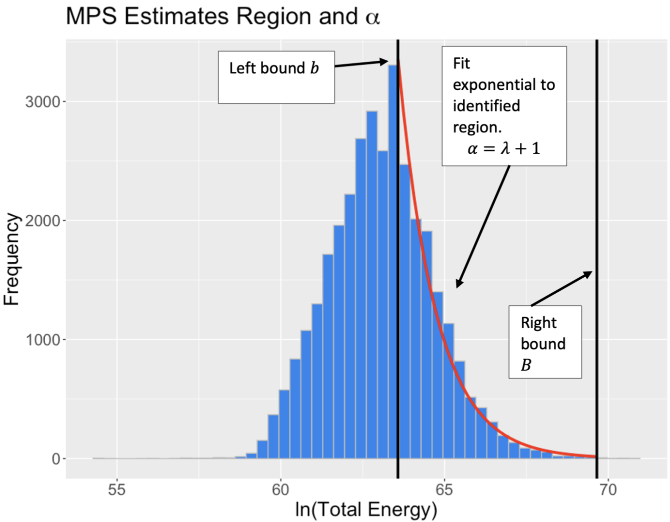
MPS Left Bound Fits to Yearly Total Energy



MPS Right Bound Fits to Yearly Total Energy



MPS Fit Method



Semi-Parametric Multinomial Model