An Introduction to the Heffernan-Tawn Model and its Limitations

Luke Fairley

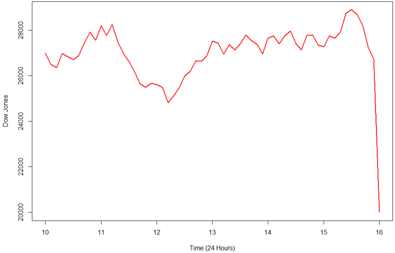
4th September 2020

Contents

- A brief overview of univariate Extreme Value Theory
- An Introduction to multivariate EVT and the complications that arise from it
- An exploration of the Heffernan-Tawn model
- Q&A

Introduction to Extremes

- Extreme Value Theory is concerned with modelling the distribution of rare events, and modelling the most extreme event that could happen in a given (long) period of time, called a return level.



Univariate Threshold Model

Let X be a random variable, and $Y = X - u \,|\, X > u$ be a random variable representing the exceedences of X above some high threshold u. Allowing for some regularity conditions, Y approximately follows a Generalised Pareto Distribution, so

$$F_Y(y;\sigma,\xi) = 1 - \left(1 + \frac{\xi y}{\sigma}\right)^{-1/\xi},$$

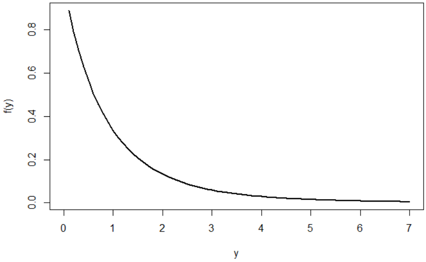
$$f_Y(y;\sigma,\xi) = \frac{1}{\sigma}\left(1 + \frac{\xi y}{\sigma}\right)^{-\frac{1}{\xi}-1}.$$

We write $Y \sim GPD(\sigma,\xi)$. Note that in the limit as $\xi \to 0$, we obtain the Exponential distribution. By convention then, if $Y \sim GPD(\sigma, \xi = 0)$, then $Y \sim \text{Exp}(1/\sigma)$.

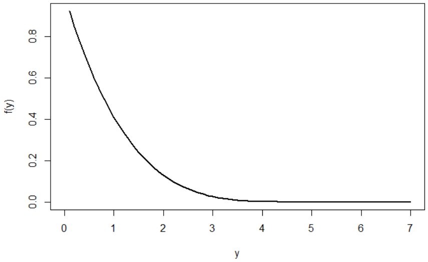
Intuition for the Threshold Model

- Positive values of ξ capture the behaviour of the distribution "trailing off", allowing for arbitrarily large excesses with low probability.
- Negative values have a "cut-off point", past which excesses are not possible. The probability density is considered 0 above $-\sigma/\xi$.

$\xi = 0.2$

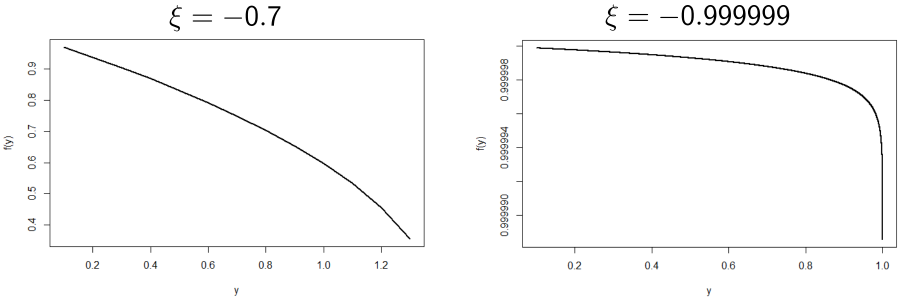


$\xi = -0.2$



Other behaviours of the model

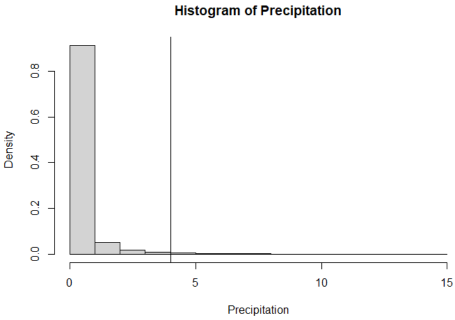
- For values of $\xi \leq -0.5$, we have slightly stranger behaviour.



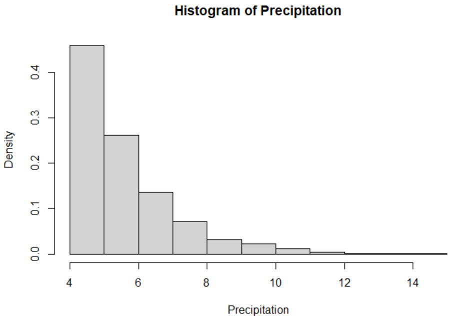
Diagnostics

- Threshold selection is generally a difficult problem.
- We choose the lowest threshold such that, when fitting for higher thresholds, estimates of the parameters are suitably "stable".
- We check the suitability of our MLE model using PP plots and QQ plots.

Example



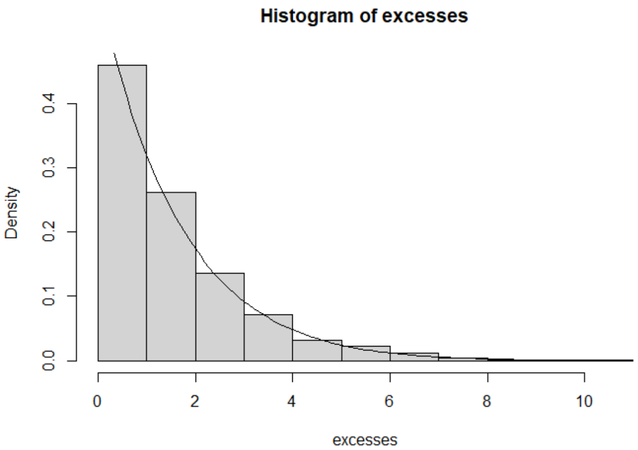
Histogram of Precipitation
Density
0.0
0.2
0.4
0.6
0.8
0
5
10
15
Precipitation



Histogram of Precipitation
Density
0.0
0.1
0.2
0.3
0.4
4
6
8
10
12
14
Precipitation

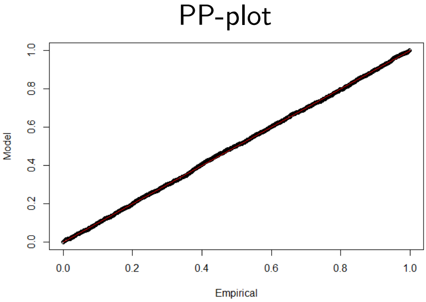
Diagnostics Plots

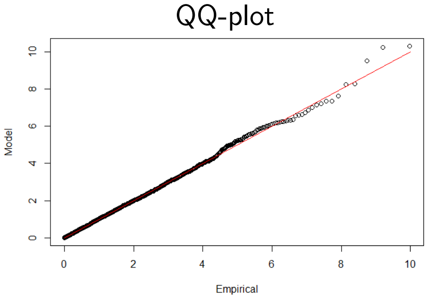
With $\hat{\sigma} \approx 1.63$ and $\hat{\xi} \approx -0.05$ we get



Diagnostics Plots

We see approximate linearity, as desired.





Semiparametric Model

Let X be a random variable such that there is a high threshold u_X such that $X - u_X \,|\, X > u_X \sim \text{GPD}(\sigma_X, \xi_X)$. Given i.i.d observations of X, we can approximate F_X by

$$\hat{F}_X(x) = \begin{cases} 1 - (1 - \tilde{F}_X(u_X))(1 + \xi_X(x - u_X)/\sigma_X)^{-1/\xi_X} & x > u_X \\ \tilde{F}_X(x) & x \leq u_X \end{cases},$$

where $\tilde{F}_X$ is the empirical distribution of the i.i.d observations.

Multivariate Extreme Value Theory

- When modelling multivariate models, we must capture both marginal probabilities and the dependence structure.
- Marginal distributions can be modelled using the univariate methods previously discussed.
- The dependency structure is far harder to capture, especially since the multivariate extreme value distributions have no closed form.

Introducing the Heffernan-Tawn Model

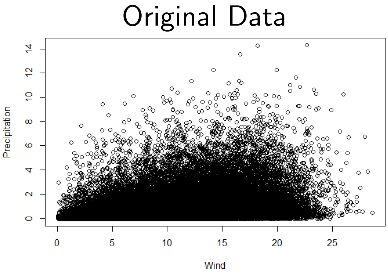
- The Heffernan-Tawn model is a conditional approach to multivariate extreme value theory, and can be used to model the random variable

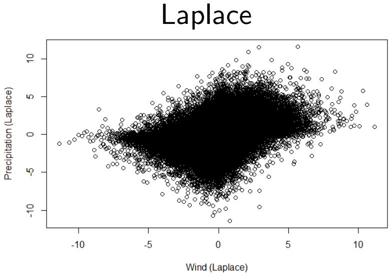
$$\boldsymbol{X} \mid X_i > u_i,$$

  where $\boldsymbol{X}$ is a multivariate random variable, X_i is one of the components of $\boldsymbol{X}$, and u_i is some high threshold for X_i.
- When we fit the model, we transform $\boldsymbol{X}$ to Laplace margins, assuming each F_{X_j} to be approximated by the semiparametric model.
- In short, the Heffernan-Tawn model seeks to model the behaviour of the whole random vector, given one of its components exceeding some high threshold.

Laplace Margins

We will demonstrate how the model is fitted using data about Wind and Precipitation at a location in the sea north-west of Scotland, from 1957 to 2018.





The Heffernan-Tawn Model

- Let (X, Y) be a random vector whose marginal distributions are Laplace, and let x be a value exceeding some high threshold u_X, distinct from the threshold used for the GPD. The Heffernan-Tawn model gives us

$$Y \mid X = x \sim \alpha x + x^{\beta} Z,$$

  where Z is some residual random variable **independent of** X, $-1 \leq \alpha \leq 1$, and $\beta < 1$.
- This generalises to a random vector of any finite dimension, but for simplicity we only consider the bivariate case throughout.

Fitting the model

We fit the model using standard inference methods, under the simplifying assumption that Z follows a $N(\mu, \sigma^2)$ distribution. This assumption is discarded after fitting the model.

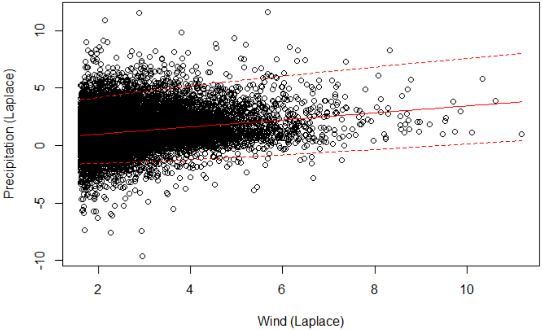


Figure: Model of Precipitation dependent on Wind above the 90% quantile.

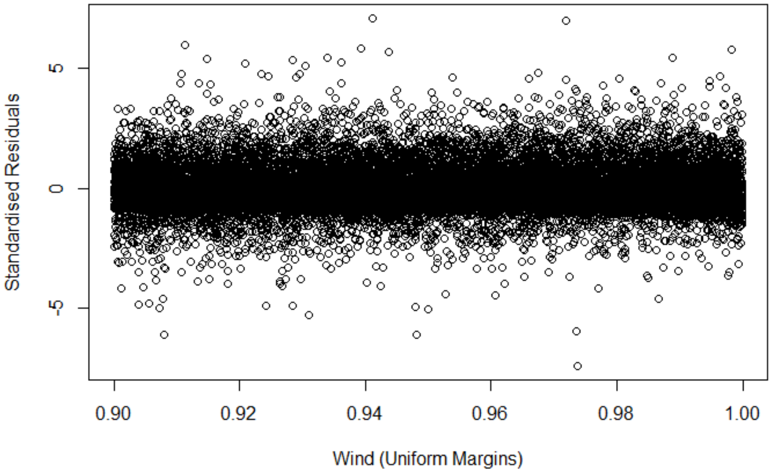
Residuals

Let $(x_1, y_1), \ldots, (x_n, y_n)$ be the data points on Laplace margins about which the model has been fitted (that is, the data points such that x is large). We define the residuals to be

$$z_i = \frac{y_i - \alpha x_i}{x_i^{\beta}}.$$

For the model to be valid, these residuals and their associated x's must be independent. We use the empirical CDF of the z_i to approximate the distribution of Z.

Diagnostic for Independence



Using the model

1. Simulate X from the Laplace distribution, conditional on exceeding a threshold greater or equal to the fitting threshold.

Using the model

1. Simulate X from the Laplace distribution, conditional on exceeding a threshold greater or equal to the fitting threshold.
2. Independently simulate Z from the empirical residual distribution.

Using the model

1. Simulate X from the Laplace distribution, conditional on exceeding a threshold greater or equal to the fitting threshold.
2. Independently simulate Z from the empirical residual distribution.
3. Set $Y = \hat{\alpha}X + X^{\hat{\beta}}Z$.

Using the model

1. Simulate X from the Laplace distribution, conditional on exceeding a threshold greater or equal to the fitting threshold.
2. Independently simulate Z from the empirical residual distribution.
3. Set $Y = \hat{\alpha}X + X^{\hat{\beta}}Z$.
4. (Optional) Transform (X, Y) back to original margins.

Using the model

1. Simulate X from the Laplace distribution, conditional on exceeding a threshold greater or equal to the fitting threshold.
2. Independently simulate Z from the empirical residual distribution.
3. Set $Y = \hat{\alpha}X + X^{\hat{\beta}}Z$.
4. (Optional) Transform (X, Y) back to original margins.
5. Repeat this many times. More points gives more accuracy, but with diminishing returns.

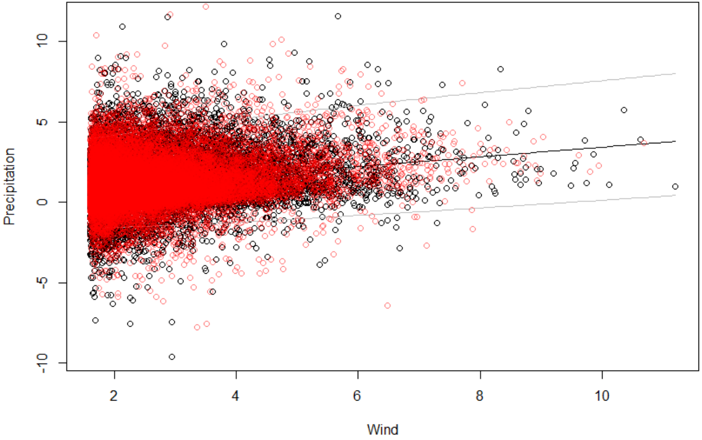
Using the model

1. Simulate X from the Laplace distribution, conditional on exceeding a threshold greater or equal to the fitting threshold.
2. Independently simulate Z from the empirical residual distribution.
3. Set $Y = \hat{\alpha}X + X^{\hat{\beta}}Z$.
4. (Optional) Transform (X, Y) back to original margins.
5. Repeat this many times. More points gives more accuracy, but with diminishing returns.

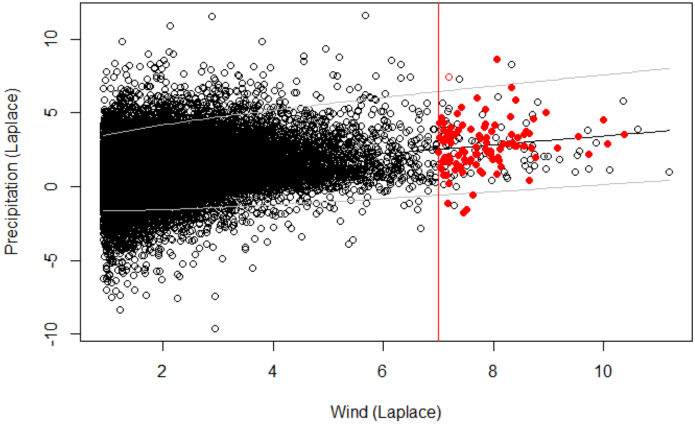
We then get an estimate of $\mathbb{P}(\boldsymbol{X} \in \mathcal{C} \mid X > v_X)$ given by

$$\hat{p} = \frac{\#\text{number of points in } \mathcal{C}}{\#\text{total number of points generated}}, \text{ with } \mathrm{Var}(\hat{p}) \approx \frac{\hat{p}(1-\hat{p})}{n}.$$

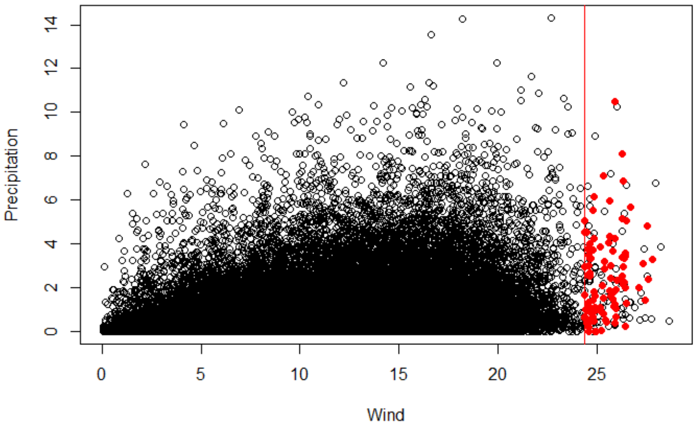
Simulated Points



Simulated Points for High Threshold



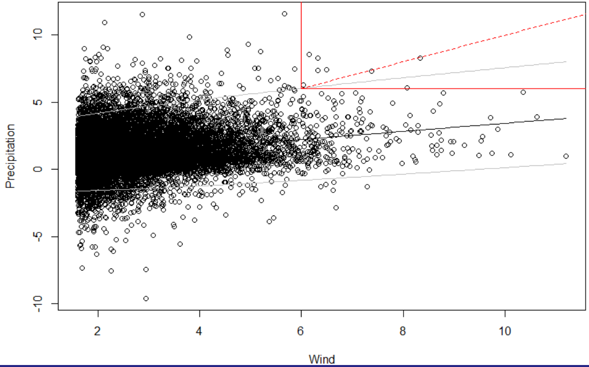
Simulated Points for High Threshold



Pedantic Sidenote

We use the decomposition

$$\mathbb{P}(\boldsymbol{X} \in \mathcal{C}) = \sum_{i=1}^{d} \mathbb{P}(\boldsymbol{X} \in \mathcal{C}_i \,|\, X_i > v_{X_i})\, \mathbb{P}(X_i > v_{X_i})$$

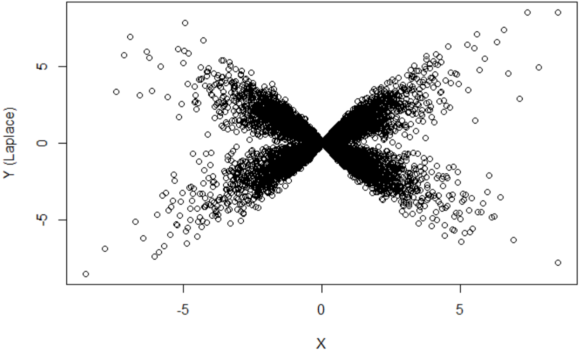


Limitations of the model

- Can show slow convergence, i.e. requires a high fitting threshold to be accurate. This is the case with Multivariate Normal data, for example.
- Assumption of Z being normal may be inadequate.
- Doesn't account for "mixed distributions".
- Limitation of the model form to $\alpha X + X^{\beta} Z$ is quite inflexible.

"Wacky" distribution

This distribution was discovered by constructing the random vector (X, Y) with $X \sim \text{Laplace}(1)$ and $Y|X = x \sim x^2 Z$ with Z standard Gaussian. This was chosen because it's similar to the model with $\alpha = 0$ and $\beta = 2$.



Potential Alterations to the model

- Explore models of the form $x\mathcal{L}_\alpha(x) + x^\beta \mathcal{L}_\beta(x) Z$, where $\mathcal{L}$ satisfies

$$\lim_{x\to\infty} \frac{\mathcal{L}(tx)}{\mathcal{L}(x)} = 1,$$

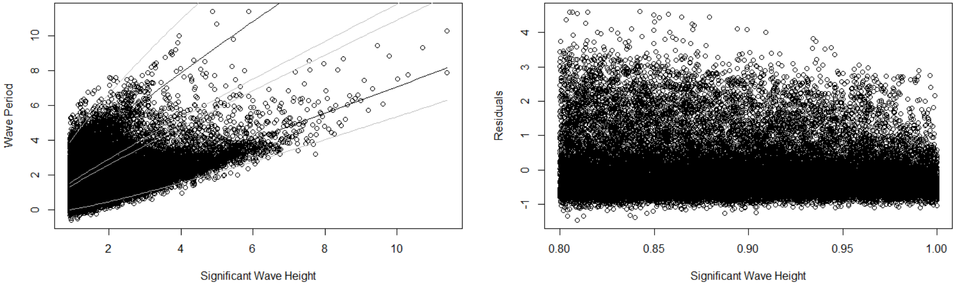
  for $t > 0$.
- Assume different distributions of Z for inference.
- Allow for mixture distributions, i.e. the model could be made up of sub-models, and when simulating a point from the model, a sub-model is chosen with some probability.

Simulation Study

This simulation study focuses on comparing the performance of the original Heffernan-Tawn model with the performance of the Heffernan-Tawn model where Z is assumed to be the mixture of two Normal distributions for inference.

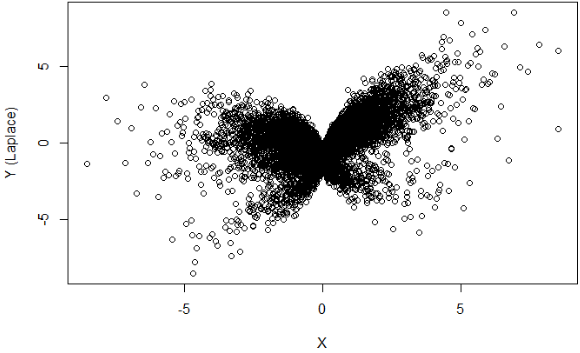
First Attempt

We first tried this on data of Significant Wave Height and Wave Period. However, we find that SWH and Residuals are not independent.



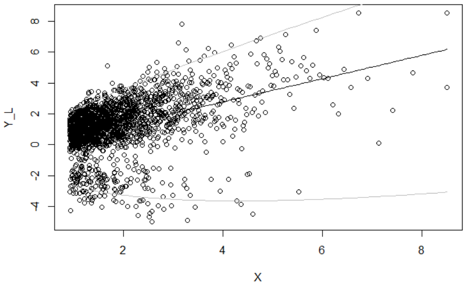
Synthetic Data

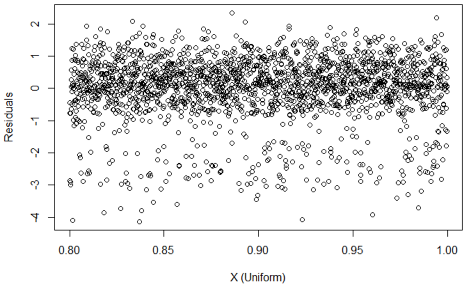
We instead compare the two models on 10000 points of synthetic data, with $X \sim \text{Laplace}(1)$ and $Y|X = x \sim 0.5x + |x|^{0.5}Z$, where Z is a mixture of $\mathcal{N}(2,1)$ and $\mathcal{N}(-2,1)$ with mixture probability $p = 0.9$. We then transform Y to Laplace margins.



Regular Fit

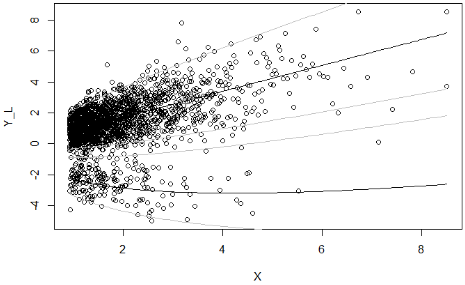
We first fit the original model to this data.

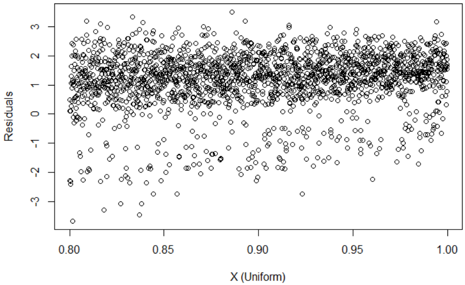




Bimodal Fit

We then fit the bimodal version.



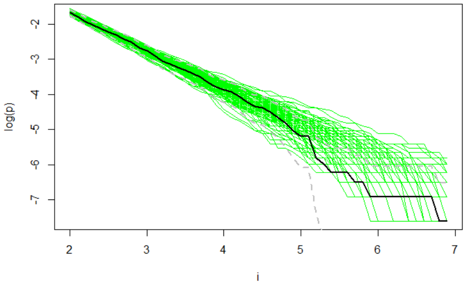


Simulation

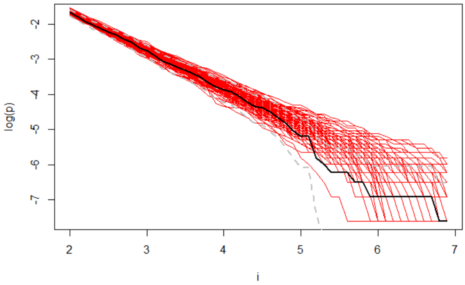
We generate 2000 points from each model, and use this to estimate probabilities $p_i = \mathbb{P}(X > i, Y > i \mid X > u_X)$, then plot lines $\{(i, \log(p_i)) : u_X \leq i \leq i_{\max}\}$, where $i_{\max}$ is the greatest i such that the data contains a point satisfying $X, Y > i$. We repeat this 100 times, and obtain confidence intervals.

All Simulations

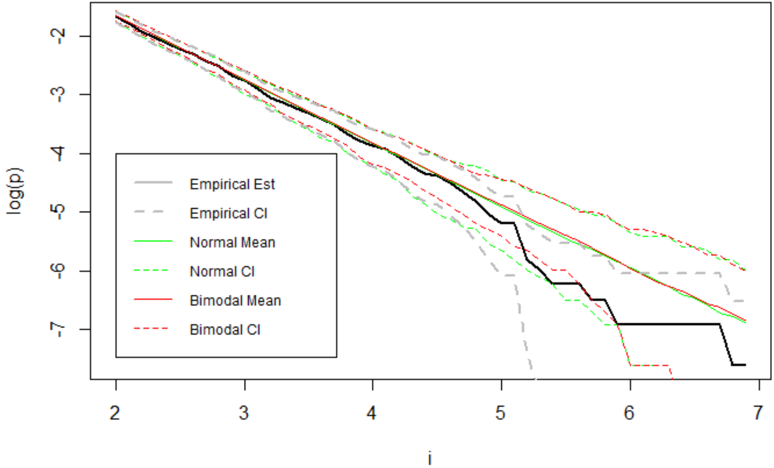
Original



Bimodal

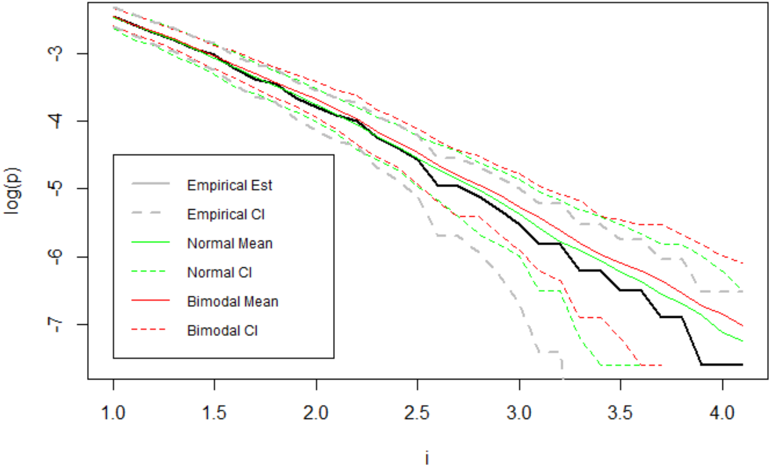


Results



log(p)
-2
-3
-4
-5
-6
-7
Empirical Est
Empirical CI
Normal Mean
Normal CI
Bimodal Mean
Bimodal CI
2
3
4
5
6
7
i

Estimating the Lower Tail



Conclusions and Extensions

- Assuming bimodality of Z does not result in a substantially better fit than assuming Z is normal.
- Bimodal Z is difficult to fit, different initial estimates often resulted in different parameter estimates.
- Other experiments showed similar results, with little improvement on the original model.
- Experiments could be carried out ad infinitum, testing different model forms, or finding distributions with which the model struggles.

References

- Stuart Coles *An Introduction to Statistical Modeling of Extreme Values* (2001)
- Janet E. Heffernan and Jonathan A. Tawn *A conditional approach for multivariate extreme values* (2004)
- T. Lugrin, A. C. Davison and J. A. Tawn *Sub-asymptotic results to motivate a new conditional multivariate extremes model* (2020)

Thanks for listening!

- Any questions?