Methods for Univariate Extreme Value Inference

Ruiyang Zhang[1]

ruiyang.zhang.20@ucl.ac.uk https://shusheng3927.github.io/website/

STOR-i Lancaster University

1. Motivation

- In many areas, like seismology and hydrology, we are interested in modelling and performing inferences for rare events. A framework to better study these rare extreme events is required.
- A common approach to model extreme data is to select a threshold and study data beyond that threshold. The quality of the model fit by this approach, however, relies heavily on the selection of threshold.
- Another approach is to use a mixture model to model both the extreme and non-extreme events. Such mixture models, like threshold based methods, shed light on the threshold and provide a way to perform inferences about extreme events as well.
- We are interested to investigate whether incorporating non-extreme data in the model provides better inferences.

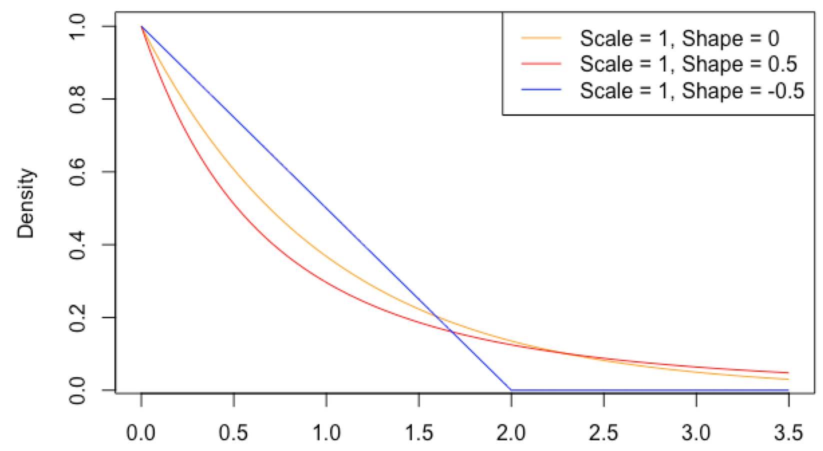
2. Peaks-Over-Threshold (POT)

- For independent samples $X_i \sim F$ for $i = 1, 2, \cdots, n$, if we denote any X_i by X, we have the **conditional excess distribution function** F_u for a threshold u, defined by $F_u(y) := \mathbb{P}(X \leq u + y \mid X \geq u)$ for $y > 0$.
- F_u describes the behaviour of exceedance data after a threshold. For an appropriate threshold, F_u can be modelled by a **generalised Pareto distribution** (GPD).
- The quality of GPD model strongly depends on the threshold, yet it is hard to select a good one.

GPD has density function

$$g_{\xi,\sigma}(x) = \begin{cases} \frac{1}{\sigma}\left(1 + \frac{\xi x}{\sigma}\right)^{-1-1/\xi} & \text{for } \xi \neq 0 \\ \frac{1}{\sigma}\, e^{-x/\sigma} & \text{for } \xi = 0, \end{cases}$$

where $x > 0$.



GPD Density Functions with Different Shape Parameters

3. Method of Murphy *et al.*

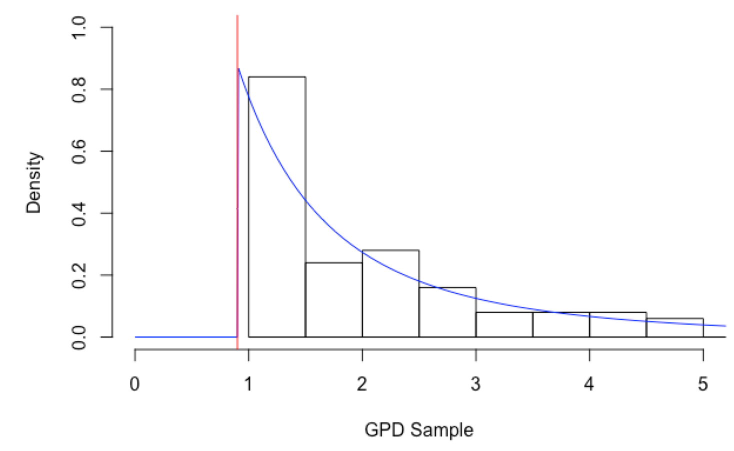
Exceedances of an appropriate threshold follow a GPD. We can compare a set of proposed thresholds, taking into account the bias-variance trade-off, using the following algorithm.

Input: data, proposal thresholds, # of bootstraps $= k$, # of quantile levels $= m$.

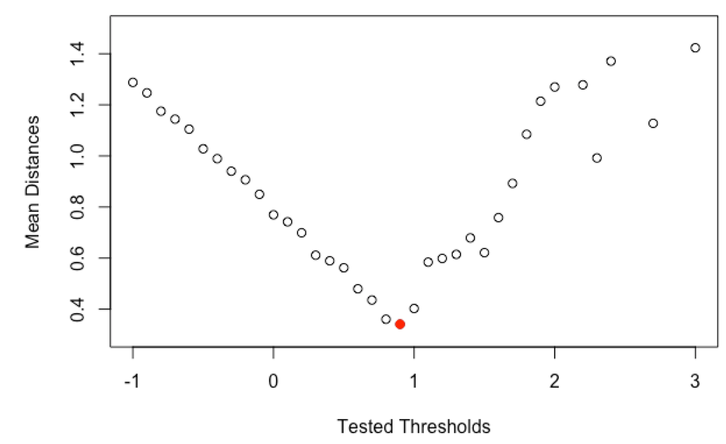
For each of the proposal thresholds:

1. Estimate the GPD parameters using MLE.
2. Measure average distance at m equally spaced quantile levels between estimated GPD and data exceedance.
3. Bootstrap the exceedances over the proposed threshold and repeat Step 2 for k times.
4. Compute mean of the k average distances.

Output: proposal threshold with least mean average distance.



(a) Method of Murphy *et al.* with GPD Samples



(b) Mean Distance Plot with Minimum in Red

4. Dynamically Weighted Mixture (DWM)

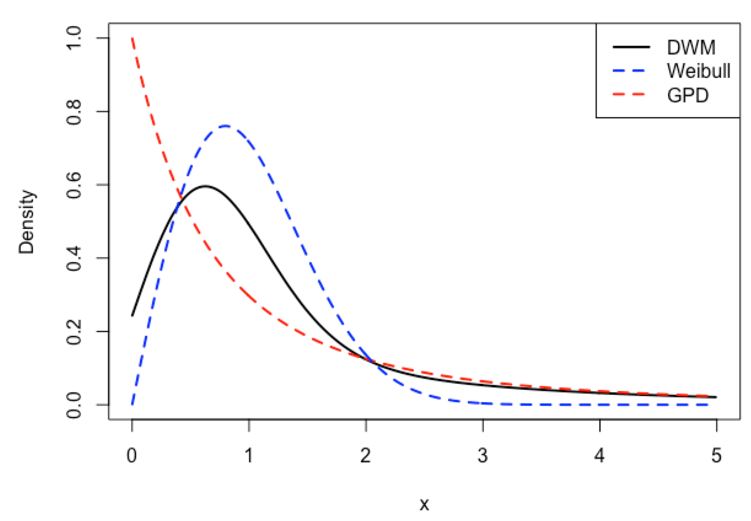
- The **Dynamically Weighted Mixture** (DWM) model is a combination of a Weibull distribution for the non-extreme and a GPD for the extreme observations.
- The mixing occurs by using a continuous Cauchy weight function (could be stepwise too).
- The weight will tend more towards the GPD for large values, and more towards the Weibull for small values.
- Although we use a Weibull for bulk here, we can use any other light-tailed distribution (say Gaussian) to model non-extreme observations.

DWM has density function

$$l(x) = \frac{(1 - p(x))f(x) + p(x)g(x)}{Z}$$

where

$$g(x) := \frac{1}{\sigma}\left(1 + \frac{\xi x}{\sigma}\right)^{-1-1/\xi} \qquad f(x) := \beta\lambda^{\beta}x^{\beta-1}\exp[-(\lambda x)^{\beta}] \qquad p(x) := \frac{1}{2} + \frac{1}{\pi}\arctan\left(\frac{x-\mu}{\tau}\right).$$



Dynamically Weighted Mixture Model Density

5. Hybrid Pareto Distribution (HPD)

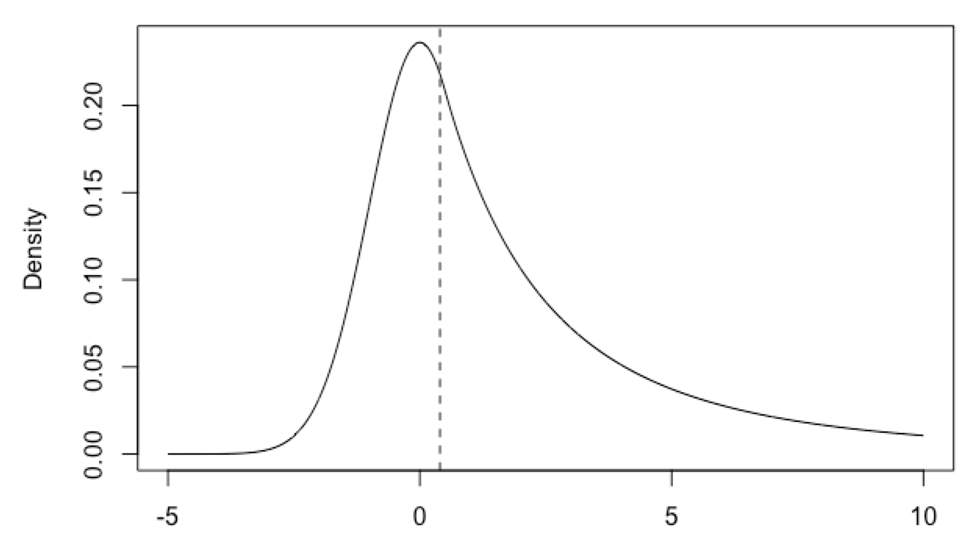
- The **Hybrid Pareto Distribution** (HPD) is a model that uses a Gaussian distribution to model non-extreme observations and a GPD to model extreme observations.
- The two distributions are stitched at a junction point, with two constraints imposed to enforce continuity of this junction.
- For junction point u, we need $f(u) = g(0)$ and $f'(u) = g'(0)$.
- The mixture is discrete and the junction serves as a threshold.

HPD has the density function

$$h(y) = \begin{cases} \frac{1}{\gamma} f_{\mu,\sigma_N}(y) & y \leq u \\ \frac{1}{\gamma} g_{\xi,\sigma}(y-u) & y > u, \end{cases}$$

where

$$f_{\mu,\sigma_N}(y) := \frac{1}{\sqrt{2\pi\sigma_N^2}}\exp\left[-\frac{(y-\mu)^2}{2\sigma_N^2}\right] \qquad g_{\xi,\beta}(y-u) := \begin{cases} \frac{1}{\sigma}\left(1 + \frac{\xi}{\sigma}(y-u)\right)^{-1/\xi-1} & \xi \neq 0, \\ \frac{1}{\sigma}\exp\left(-\frac{y-u}{\sigma}\right) & \xi = 0. \end{cases}$$



Hybrid Pareto Distribution with $\mu = 0$, $\sigma = 1$, $\xi = 0.4$

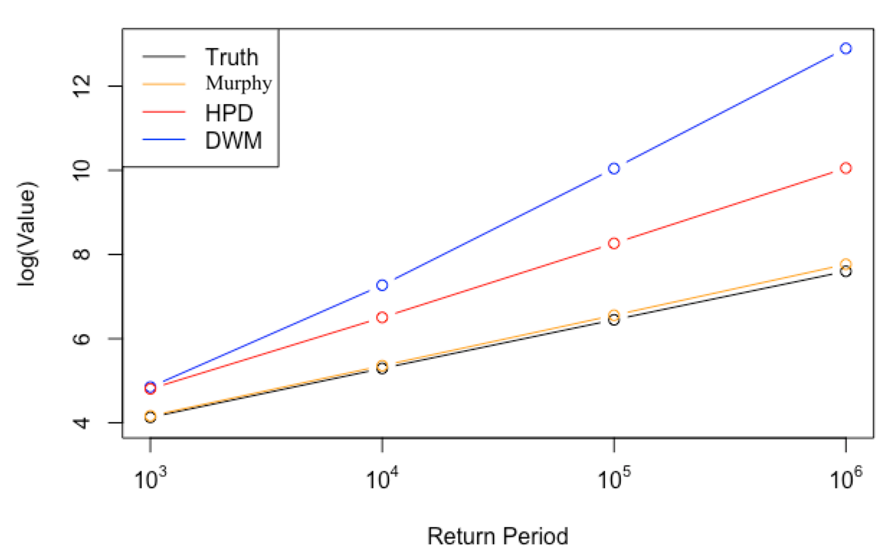
6. Simulation Study

Setup:

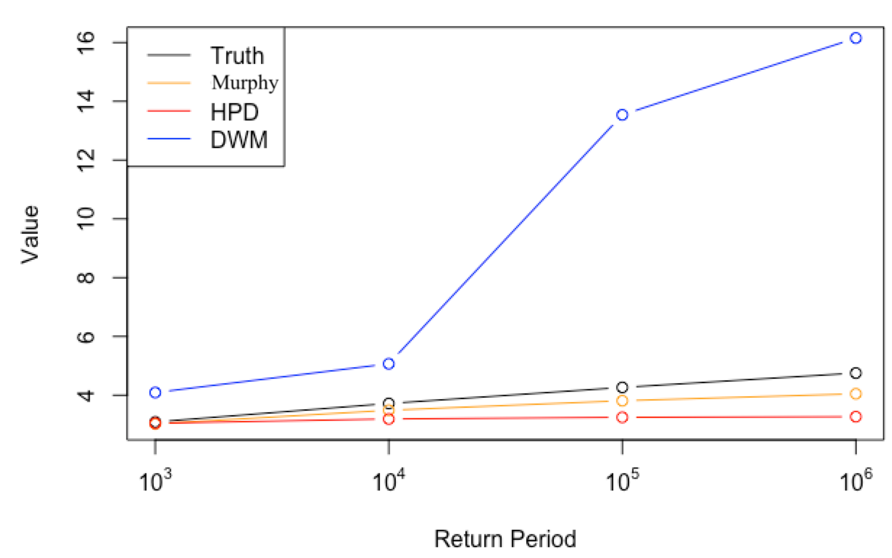
- Models: (1) Method of Murphy *et al.* (**M**) (2) HPD (**H**) (3) DWM (**D**)
- Sample Distributions: (1) GPD scale = 1, shape = 0.5, location = 1 (2) Standard Normal (3) Beta(2,5)
- Sample Size = 1000, Repetition = 100.

Findings:

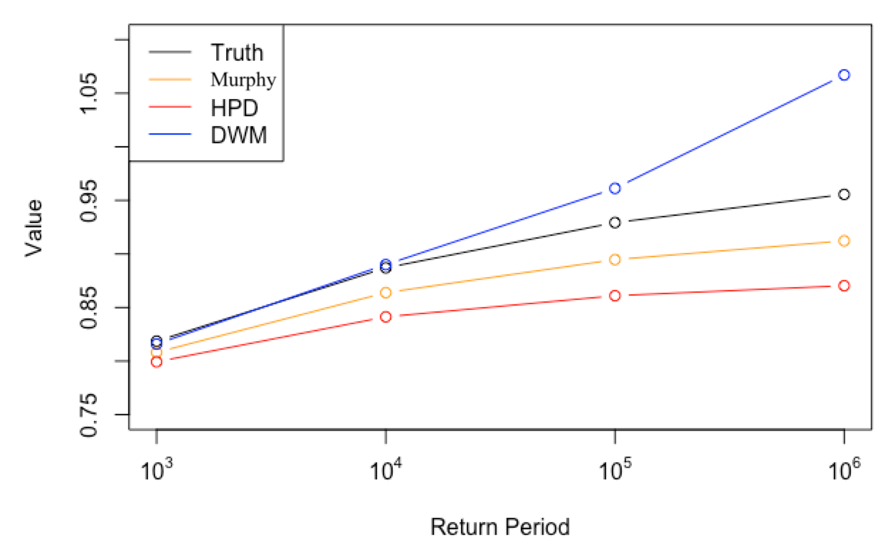
- Overall, **M** does very well at estimating return levels and has low RMSE.
- **H** does a decent job at estimating return levels and has relatively low RMSE. Also, though not shown here, **H** is the most computationally efficient method among the three.



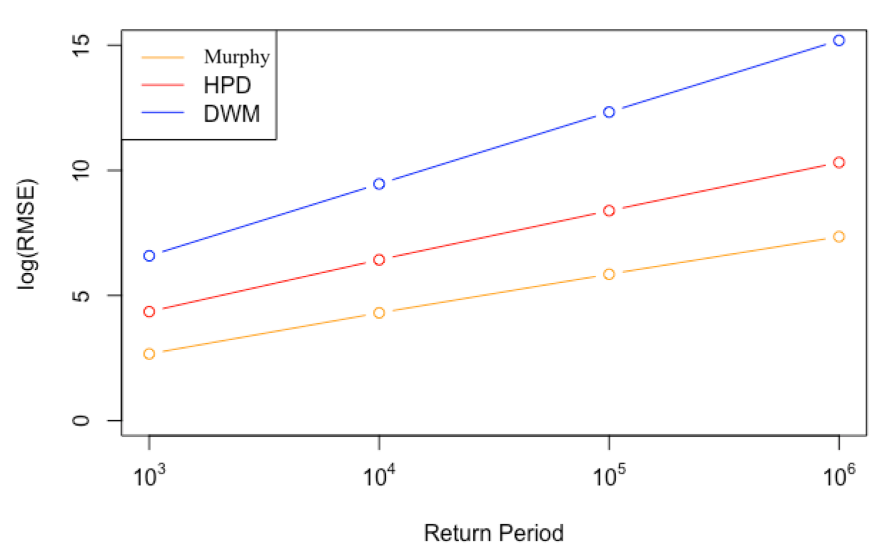
(a) Estimated Return Levels - GPD



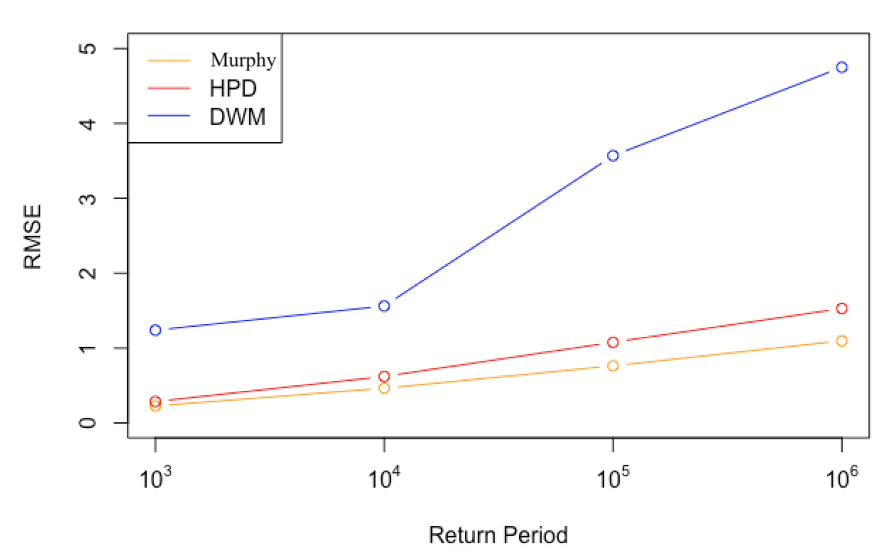
(b) Estimated Return Levels - Normal



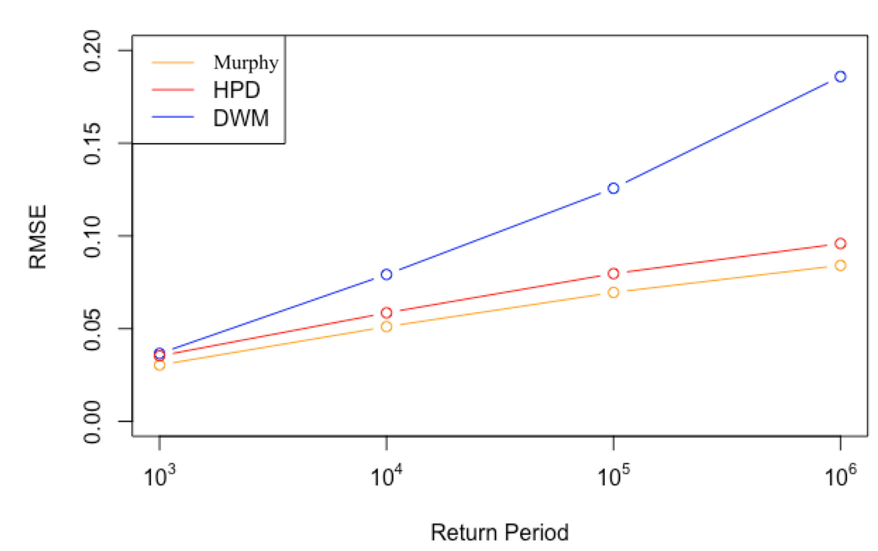
(c) Estimated Return Levels - Beta



(d) Root Mean Squared Error - GPD



(e) Root Mean Squared Error - Normal



(f) Root Mean Squared Error - Beta

Supervisors: Conor Murphy[2] and Lídia André[2]

[1] University College London [2] STOR-i CDT, Lancaster University

UKRI Engineering and Physical Sciences Research Council