Methods for Univariate Extreme Value Inference

Zhang Ruiyang

Supervisors: Conor Murphy, Lídia André

Aug 2022
Introduction

Figure: Fort Collins, Colorado, USA
Introduction

Figure: Fort Collins Flood, July 1997
Outline

1. Peaks Over Threshold
2. Mixture Model
3. Simulation Study
Peaks Over Threshold

Figure: Fort Collins Daily Maximum Precipitation Data with Threshold, 1900-1999
Density function of **generalised Pareto distribution** (GPD) is

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

where \( x > 0 \).

**Figure:** GPD Density Functions with Different Shape Parameters
**Method of Murphy et al.**

**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.
Method of Murphy et al.

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

(b) Mean Distance Plot with Minimum in Red
Mixture Model

- Model the extreme by one distribution (e.g. GPD)
- Model the non-extreme by another distribution (e.g. Gaussian)
- Average the two distributions by imposing a (continuous / discrete) weight.
Dynamically Weight Model (DWM)

Based on Frigessi, Haug, and Rue (2002).

Dynamically Weight Model has the density function

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

where

\[ g(x) := \frac{1}{\sigma} \left( 1 + \frac{\xi x}{\sigma} \right)^{-1 - 1/\xi}, \]

\[ f(x) := \beta \lambda^\beta x^{\beta - 1} \exp[-(\lambda x)^\beta], \]

\[ p(x) := \frac{1}{2} + \frac{1}{\pi} \arctan \left( \frac{x - \mu}{\tau} \right). \]

\( g \) is GPD density, \( f \) is Weibull density, and \( p \) is Cauchy CDF. \( Z \) is a normalising constant.
Dynamically Weight Model (DWM)

(a) Weight Function with \( \mu = 0 \) and \( \tau = 2 \)

(b) Dynamically Weighted Mixture Model
Hybrid Pareto Model Distribution (HPD)

Based on Carreau and Bengio (2009).

Hybrid Pareto Model Distribution 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 \( f_{\mu, \sigma_N} \) is the normal density, \( g_{\xi, \sigma} \) is the GPD density, and \( \gamma \) is the normalising constant. \( u \) is the threshold.
Hybrid Pareto Model Distribution (HPD)

Based on Carreau and Bengio (2009).

Hybrid Pareto Model Distribution 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 \( f_{\mu, \sigma_N} \) is the normal density, \( g_{\xi, \sigma} \) is the GPD density, and \( \gamma \) is the normalising constant. \( u \) is the threshold.

To enforce continuity of the distribution, two constraints are imposed at the threshold \( u \):

1. continuity at \( u \), \( f(u) = g(0) \)

2. continuity of derivative at \( u \), \( f'(u) = g'(0) \)
Figure: Hybrid Pareto Distribution with $\mu = 0$, $\sigma = 1$, $\xi = 0.4$
Simulation Study - Setup

Models: (1) Method of Murphy *et al.* (2) HPD (3) DWM

Sample Distributions: (1) GPD scale = 1, shape = 0.5, threshold = 1 (2) Standard Normal (3) Beta (2,5)

Sample Size = 1000

Repetition = 100.

Comparisons: (1) $10^3$, $10^4$, $10^5$, $10^6$ Return Levels. (2) Root Mean Squared Errors
Simulation Study - GPD Samples

(a) Estimated Return Levels

(b) Root Mean Squared Errors
Simulation Study - Standard Normal Samples

(a) Estimated Return Levels  
(b) Root Mean Squared Errors
Simulation Study - Beta(2,5) Samples

(a) Estimated Return Levels

(b) Root Mean Squared Errors
Conclusion

Summary

- Method of Murphy et al. provides most accurate return level estimations and lowest RMSEs.
- HPD provides relatively accurate return level estimations too.
Conclusion

Summary
- Method of Murphy \textit{et al.} provides most accurate return level estimations and lowest RMSEs.
- HPD provides relatively accurate return level estimations too.

Extensions
- Simulation studies for an even more varied sample distributions.
- Incorporate confidence intervals for estimations by Murphy \textit{et al.}
Questions?