Investigating the effect of aggregation on extremal dependence

Dylan Bahia

September 6, 2019

What is an extreme value?

What is an extreme value?

What is an extreme value?

- An extreme value is one which occurs with a very low probability.

What is an extreme value?

- An extreme value is one which occurs with a very low probability.
- There a two ways in which such a value can be classified

What is an extreme value?

- An extreme value is one which occurs with a very low probability.
- There a two ways in which such a value can be classified
- The classification used depends on the problem

What is an extreme value?

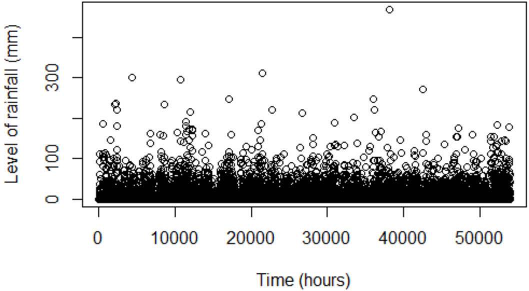

Figure: Level of rainfall at fixed location over time

What is an extreme value?

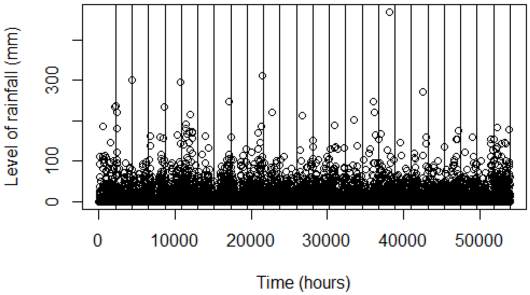

Figure: Level of rainfall at fixed location over time

What is an extreme value?

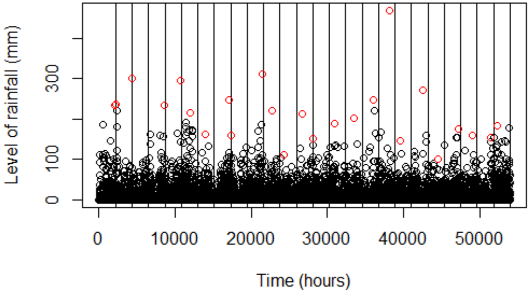

Figure: Level of rainfall at fixed location over time, with block maxima in red

What is an extreme value?

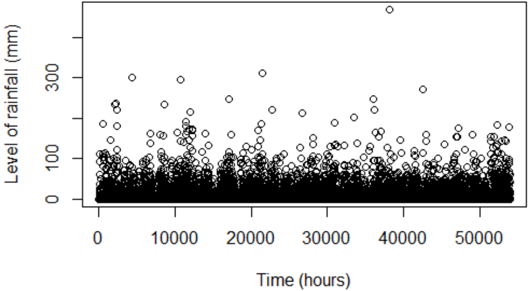

Figure: Level of rainfall at fixed location over time

What is an extreme value?

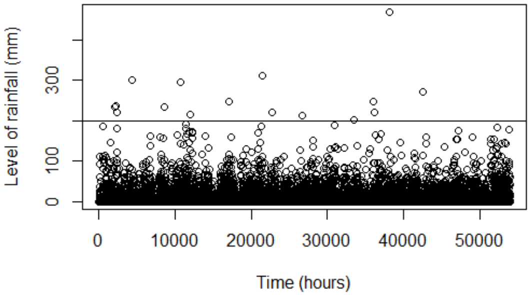

Figure: Level of rainfall at fixed location over time

What is an extreme value?

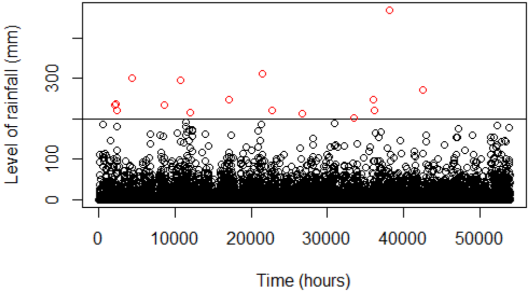

Figure: Level of rainfall at fixed location over time, with values exceeding threshold in red

Block maxima vs Threshold

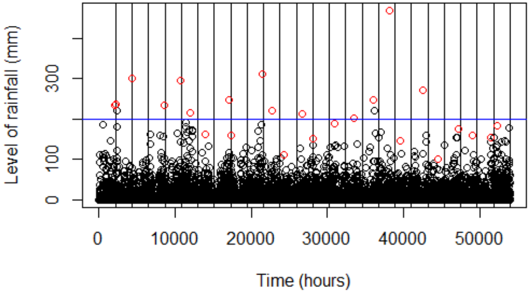

Figure: Level of rainfall at fixed location over time, with block maxima in red and threshold line in blue

Distribution of block maxima

Distribution of block maxima

- Let X_i denote the i^{th} observation of a random variable X

Distribution of block maxima

- Let X_i denote the i^{th} observation of a random variable X
- Let $M_n = max\{X_1, ..., X_n\}$.

Distribution of block maxima

- Let X_i denote the i^{th} observation of a random variable X
- Let $M_n = max\{X_1, ..., X_n\}$.
- Suppose there exist a sequence of constants $a_n > 0$ and b_n such that

$$P\left(\frac{(M_n - b_n)}{a_n} \leq z\right) \rightarrow G(z) \text{ as } n \rightarrow \infty,$$

where G is a non-degenerate distribution function.

Distribution of block maxima

- Let X_i denote the i^{th} observation of a random variable X
- Let $M_n = max\{X_1, ..., X_n\}$.
- Suppose there exist a sequence of constants $a_n > 0$ and b_n such that

$$P\left(\frac{(M_n - b_n)}{a_n} \leq z\right) \to G(z) \text{ as } n \to \infty,$$

where G is a non-degenerate distribution function.

- Then G is a Generalised Extreme Value (GEV) distribution, where

$$G(z) = exp\left(-\left(1 + \xi\left(\frac{(z-\mu)}{\sigma}\right)\right)^{-\frac{1}{\xi}}\right)$$

Distribution of block maxima

- Location parameter: $-\infty < \mu < \infty$

Distribution of block maxima

- Location parameter: $-\infty < \mu < \infty$
- Scale parameter: $\sigma > 0$

Distribution of block maxima

- Location parameter: $-\infty < \mu < \infty$
- Scale parameter: $\sigma > 0$
- Shape parameter: $-\infty < \xi < \infty$

Distribution of block maxima

- Location parameter: $-\infty < \mu < \infty$
- Scale parameter: $\sigma > 0$
- Shape parameter: $-\infty < \xi < \infty$
- Support: $\{z : 1 + \frac{\xi(z-\mu)}{\sigma} > 0\}$

Effect of ξ

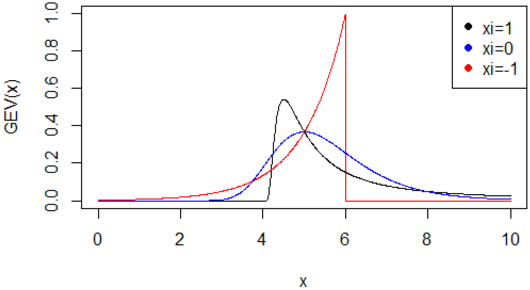

Figure: GEV dist. with $\mu = 5, \sigma = 1$ and varying value of ξ

Distribution of values exceeding threshold

Distribution of values exceeding threshold

- If the block maxima are GEV distributed, then for large enough u, the distribution function of $(X - u)$, conditional on $X > u$, is approximately

$$\begin{aligned} H(y) &= P(X < u + y \mid X > u) \\ &= 1 - \left(1 + \frac{\xi y}{\widetilde{\sigma}}\right)^{-\frac{1}{\xi}} \end{aligned}$$

where $\widetilde{\sigma} = \sigma + \xi(u - \mu)$

Distribution of values exceeding threshold

- If the block maxima are GEV distributed, then for large enough u, the distribution function of $(X - u)$, conditional on $X > u$, is approximately

$$
\begin{aligned}
H(y) &= P(X < u + y \mid X > u) \\
&= 1 - \left(1 + \frac{\xi y}{\widetilde{\sigma}}\right)^{-\frac{1}{\xi}}
\end{aligned}
$$

where $\widetilde{\sigma} = \sigma + \xi(u - \mu)$

- This called the Generalised Pareto Distribution (GPD)

Distribution of values exceeding threshold

- If the block maxima are GEV distributed, then for large enough u, the distribution function of $(X - u)$, conditional on $X > u$, is approximately

$$\begin{aligned} H(y) &= P(X < u + y \mid X > u) \\ &= 1 - \left(1 + \frac{\xi y}{\widetilde{\sigma}}\right)^{-\frac{1}{\xi}} \end{aligned}$$

  where $\widetilde{\sigma} = \sigma + \xi(u - \mu)$
- This called the Generalised Pareto Distribution (GPD)
- It is has support $y \geq u$ if $\xi \geq 0$ & $u \leq y \leq u - \frac{\sigma}{\xi}$ if $\xi < 0$

Choosing a threshold

Choosing a threshold

- There are two factors to consider when choosing a threshold:

Choosing a threshold

- There are two factors to consider when choosing a threshold:
  - If the threshold is too low, the asymptotic properties of the model will no longer hold, thus leading to bias.

Choosing a threshold

- There are two factors to consider when choosing a threshold:
  - If the threshold is too low, the asymptotic properties of the model will no longer hold, thus leading to bias.
  - If the threshold is too high, there won't be enough data to estimate the model accurately, leading to high variance.

Choosing a threshold

- There are two factors to consider when choosing a threshold:
  - If the threshold is too low, the asymptotic properties of the model will no longer hold, thus leading to bias.
  - If the threshold is too high, there won't be enough data to estimate the model accurately, leading to high variance.
- The standard practice is to choose the threshold to be as low as possible, whilst making sure the model provides a reasonable approximation.

Return Levels

- A series of independent observations X_i, $i = 1, ..., d$, can be blocked into sequences of length n (the length is often 1 year), generating a series of block maxima $M_{n,1}, ...M_{n,m}$, where m is the number of maxima.

Return Levels

- A series of independent observations X_i, $i = 1, ..., d$, can be blocked into sequences of length n (the length is often 1 year), generating a series of block maxima $M_{n,1}, ...M_{n,m}$, where m is the number of maxima.
- A GEV distribution can be fitted to these maxima, and extreme quantiles can be estimated by inverting the distribution function:

$$z_p = \begin{cases} \mu - \frac{\sigma}{\xi}(1 - (-log(1-p))^{-\xi}) \text{ for } \xi \neq 0 \\ \mu - \sigma \log(-\log(1-p)) \text{ for } \xi = 0 \end{cases}$$

where $G(z_p) = 1 - p$.

Return Levels

- A series of independent observations X_i, $i = 1, ..., d$, can be blocked into sequences of length n (the length is often 1 year), generating a series of block maxima $M_{n,1}, ...M_{n,m}$, where m is the number of maxima.
- A GEV distribution can be fitted to these maxima, and extreme quantiles can be estimated by inverting the distribution function:

$$z_p = \begin{cases} \mu - \frac{\sigma}{\xi}(1 - (-log(1-p))^{-\xi}) \text{ for } \xi \neq 0 \\ \mu - \sigma \log(-\log(1-p)) \text{ for } \xi = 0 \end{cases}$$

  where $G(z_p) = 1 - p$.
- z_p is the return level associated with the return period $1/p$

Return Levels

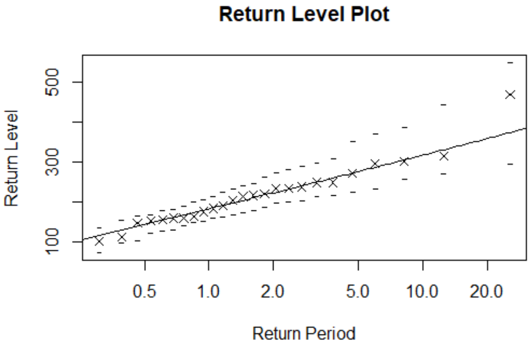

Figure: Expected return level after given period of years

Extremal Dependence Measures

Extremal Dependence Measures

- It is very useful to look at how likely extremes of different variables are to occur together

Extremal Dependence Measures

- It is very useful to look at how likely extremes of different variables are to occur together
- This is known as asymptotic dependence

Extremal Dependence Measures

- It is very useful to look at how likely extremes of different variables are to occur together
- This is known as asymptotic dependence
- The are various summary measures for characterising this dependence

Extremal Dependence Measures

- It is very useful to look at how likely extremes of different variables are to occur together
- This is known as asymptotic dependence
- The are various summary measures for characterising this dependence
- I have looked into two of them; χ and η

Extremal Dependence Measures - χ

Extremal Dependence Measures - χ

- χ is known as the upper tail index

Extremal Dependence Measures - χ

- χ is known as the upper tail index
- It is a measure of asymptotic dependence

Extremal Dependence Measures - χ

- χ is known as the upper tail index
- It is a measure of asymptotic dependence
- Given two random variables X and Y, they can be transformed to uniform margins, producing respective variables U and V

Extremal Dependence Measures - χ

- χ is known as the upper tail index
- It is a measure of asymptotic dependence
- Given two random variables X and Y, they can be transformed to uniform margins, producing respective variables U and V
- $\chi = lim_{u \to 1} P(V > u \mid U > u)$

Extremal Dependence Measures - χ

- χ is known as the upper tail index
- It is a measure of asymptotic dependence
- Given two random variables X and Y, they can be transformed to uniform margins, producing respective variables U and V
- $\chi = lim_{u \to 1} P(V > u \mid U > u)$
- χ takes values between 0 and 1

Extremal Dependence Measures - χ

- χ is known as the upper tail index
- It is a measure of asymptotic dependence
- Given two random variables X and Y, they can be transformed to uniform margins, producing respective variables U and V
- $\chi = lim_{u \to 1} P(V > u \mid U > u)$
- χ takes values between 0 and 1
- 0 corresponds to asymptotic independence

Extremal Dependence Measures - χ

- χ is known as the upper tail index
- It is a measure of asymptotic dependence
- Given two random variables X and Y, they can be transformed to uniform margins, producing respective variables U and V
- $\chi = lim_{u \to 1} P(V > u \mid U > u)$
- χ takes values between 0 and 1
- 0 corresponds to asymptotic independence
- 1 corresponds to perfect asymptotic dependence

Extremal Dependence Measures - χ

- χ is known as the upper tail index
- It is a measure of asymptotic dependence
- Given two random variables X and Y, they can be transformed to uniform margins, producing respective variables U and V
- $\chi = lim_{u \to 1} P(V > u \mid U > u)$
- χ takes values between 0 and 1
- 0 corresponds to asymptotic independence
- 1 corresponds to perfect asymptotic dependence
- $\chi > 0$ implies asymptotic dependence

Extremal Dependence Measures - χ

- χ is known as the upper tail index
- It is a measure of asymptotic dependence
- Given two random variables X and Y, they can be transformed to uniform margins, producing respective variables U and V
- $\chi = lim_{u \to 1} P(V > u \mid U > u)$
- χ takes values between 0 and 1
- 0 corresponds to asymptotic independence
- 1 corresponds to perfect asymptotic dependence
- $\chi > 0$ implies asymptotic dependence
- As χ increases, asymptotic dependence increases

Empirical calculation of χ

Empirical calculation of χ

- Let X and Y be set of observations of random variables of length n and m respectively

Empirical calculation of χ

- Let X and Y be set of observations of random variables of length n and m respectively
- Let u be the threshold

Empirical calculation of χ

- Let X and Y be set of observations of random variables of length n and m respectively
- Let u be the threshold
- Let $U = \frac{rank(X)}{(n+1)}$
- Let $V = \frac{rank(Y)}{(n+1)}$

Empirical calculation of χ

- Let X and Y be set of observations of random variables of length n and m respectively
- Let u be the threshold
- Let $U = \frac{rank(X)}{(n+1)}$
- Let $V = \frac{rank(Y)}{(n+1)}$
- Then $\widehat{\chi} = \frac{\Sigma(U>u|V>u)}{\Sigma(V>u)}$

Simulations

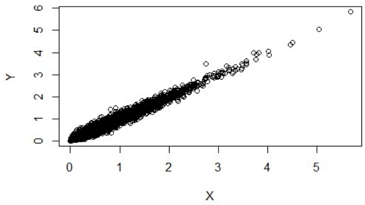

Figure: Simulation of random variables X and Y such that $\widehat{\chi} = 0.9$

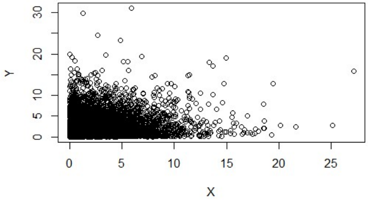

Figure: Simulation of random variables X and Y such that $\widehat{\chi} = 0.1$

Extremal Dependence Measures - η

Extremal Dependence Measures - η

- η is known as the coefficient of tail dependence

Extremal Dependence Measures - η

- η is known as the coefficient of tail dependence
- It is a measure of asymptotic independence

Extremal Dependence Measures - η

- η is known as the coefficient of tail dependence
- It is a measure of asymptotic independence
- Let X and Y be random variables on exponential margins

Extremal Dependence Measures - η

- η is known as the coefficient of tail dependence
- It is a measure of asymptotic independence
- Let X and Y be random variables on exponential margins
- $P(X > x, Y > x) \sim L(x) exp(-\frac{x}{\eta})$
  where $L(x)$ is a slowly varying function

Extremal Dependence Measures - η

- η is known as the coefficient of tail dependence
- It is a measure of asymptotic independence
- Let X and Y be random variables on exponential margins
- $P(X > x, Y > x) \sim L(x)exp(-\frac{x}{\eta})$
  where $L(x)$ is a slowly varying function
- A slowly varying function satisfies

$$\frac{L(cx)}{L(x)} \sim 1 \textbf{ as } x \to \infty$$

  for some constant $c > 0$

Extremal Dependence Measures - η

- η is known as the coefficient of tail dependence
- It is a measure of asymptotic independence
- Let X and Y be random variables on exponential margins
- $P(X > x, Y > x) \sim L(x) exp(-\frac{x}{\eta})$
  where $L(x)$ is a slowly varying function
- A slowly varying function satisfies
  $$\frac{L(cx)}{L(x)} \sim 1 \textbf{ as } x \to \infty$$
  for some constant $c > 0$
- It takes values between 0.5 and 1

Extremal Dependence Measures - η

- η is known as the coefficient of tail dependence
- It is a measure of asymptotic independence
- Let X and Y be random variables on exponential margins
- $P(X > x, Y > x) \sim L(x) exp(-\frac{x}{\eta})$
  where $L(x)$ is a slowly varying function
- A slowly varying function satisfies

  $$\frac{L(cx)}{L(x)} \sim 1 \textbf{ as } x \to \infty$$

  for some constant $c > 0$
- It takes values between 0.5 and 1
- 1 corresponds to asymptotic dependence

Extremal Dependence Measures - η

- η is known as the coefficient of tail dependence
- It is a measure of asymptotic independence
- Let X and Y be random variables on exponential margins
- $P(X > x, Y > x) \sim L(x)exp(-\frac{x}{\eta})$
  where $L(x)$ is a slowly varying function
- A slowly varying function satisfies
  $$\frac{L(cx)}{L(x)} \sim 1 \textbf{ as } x \to \infty$$
  for some constant $c > 0$
- It takes values between 0.5 and 1
- 1 corresponds to asymptotic dependence
- $eta < 1$ implies asymptotic independence
- As η decreases, asymptotic independence increases

Application

Application

- Gridded hourly precipitation data taken over the North West of England

Application

- Gridded hourly precipitation data taken over the North West of England
- Taken over December, January and February from 1990 to 2014

Application

- Gridded hourly precipitation data taken over the North West of England
- Taken over December, January and February from 1990 to 2014
- Data comes from climate simulations

Applications

- I calculated empirical estimates for pairwise values of χ and η between locations

Applications

- I calculated empirical estimates for pairwise values of χ and η between locations
- The purpose is to look at how the dependence structure changes between aggregation levels

Changes in χ

Changes in χ

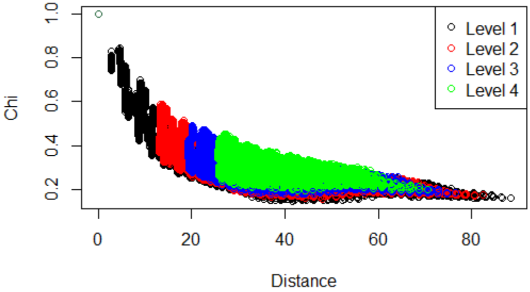

Figure: Values of χ with respect to distance at 4 different aggregation levels

Changes in χ

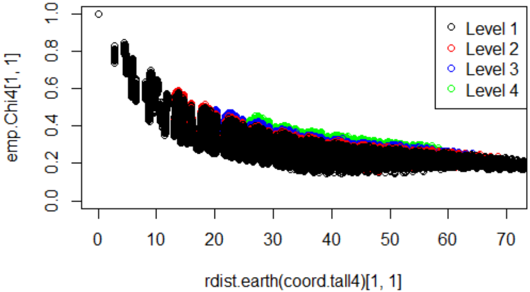

Figure: Values of χ with respect to distance at 4 different aggregation levels

Changes in η

Changes in η

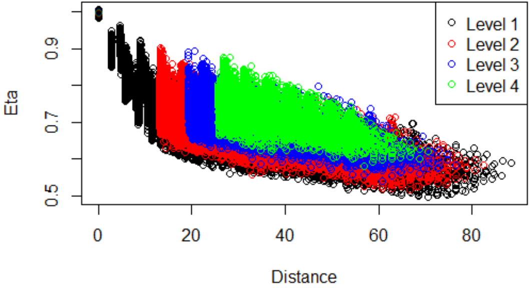

Figure: Values of η with respect to distance at 4 different aggregation levels

Changes in η

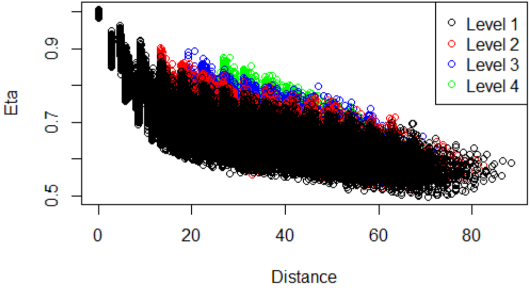

Figure: Values of η with respect to distance at 4 different aggregation levels