A Conditional Approach to Modeling Multivariate Extremes

By Heffernan & Tawn

Presented by Whitney Huang

Department of Statistics
Purdue University

April 30, 2014
Outline

Motivation & Background

Conditional Extreme Value Models
  Model
  Inference

Air quality application
A central aim of multivariate extremes is trying to describe the structure of tail dependence.
Background: componentwise maxima approach

- Consider \( \mathbf{X} = (X_1, \cdots, X_d) \) be a \( d \)-dimensional random vector with distribution function \( F \), \( \mathbf{X}_i = (X_{i1}, \cdots, X_{id}) \), \( 1 \leq i \leq n \) be random sample from \( F \).
- Define \( M_{nj} = \max_{i=1,\ldots,n} X_{ij} \), if there exist sequences \( a_{nj} > 0, \ b_{nj} \in \mathbb{R}, j = 1, \cdots, d \) such that
\[
\mathbb{P}\left( \frac{M_{n1} - b_{n1}}{a_{n1}} \leq x_1, \cdots, \frac{M_{nd} - b_{nd}}{a_{nd}} \leq x_d \right) \to G(\mathbf{x})
\]
weakly as \( n \to \infty \), and marginal distributions \( G_1, \cdots, G_d \) of \( G \) are non-degenerate, we say that \( F \) is in the domain of attraction of \( G \).
- Such \( G \) is called an multivariate extreme value distribution which is the limiting distribution function of the vector component-wise maxima of \( (X_1, \cdots, X_n) \).
Background: componentwise maxima approach

- Consider $\mathbf{X} = (X_1, \cdots, X_d)$ be a $d$-dimensional random vector with distribution function $F$, $\mathbf{X}_i = (X_{i1}, \cdots, X_{id})$, $1 \leq i \leq n$ be random sample from $F$.

- Define $M_{nj} = \max_{i=1,\ldots,n} X_{ij}$, if there exist sequences $a_{nj} > 0$, $b_{nj} \in \mathbb{R}$, $j = 1, \cdots, d$ such that

$$
\mathbb{P} \left( \frac{M_{n1} - b_{n1}}{a_{n1}} \leq x_1, \cdots, \frac{M_{nd} - b_{nd}}{a_{nd}} \leq x_d \right) \to G(\mathbf{x})
$$

weakly as $n \to \infty$, and marginal distributions $G_1, \cdots, G_d$ of $G$ are non-degenerate, we say that $F$ is in the domain of attraction of $G$.

- Such $G$ is called an multivariate extreme value distribution which is the limiting distribution function of the vector component-wise maxima of $(X_1, \cdots, X_n)$. 
Background: componentwise maxima approach

- Consider $\mathbf{X} = (X_1, \cdots, X_d)$ be a $d$-dimensional random vector with distribution function $F$, $\mathbf{X}_i = (X_{i1}, \cdots, X_{id})$, $1 \leq i \leq n$ be random sample from $F$.

- Define $M_{nj} = \max_{i=1,\cdots,n} X_{ij}$, if there exist sequences $a_{nj} > 0$, $b_{nj} \in \mathbb{R}$, $j = 1, \cdots, d$ such that

$$
\mathbb{P}\left(\frac{M_{n1} - b_{n1}}{a_{n1}} \leq x_1, \cdots, \frac{M_{nd} - b_{nd}}{a_{nd}} \leq x_d\right) \rightarrow G(\mathbf{x})
$$

weakly as $n \rightarrow \infty$, and marginal distributions $G_1, \cdots, G_d$ of $G$ are non-degenerate, we say that $F$ is in the domain of attraction of $G$.

- Such $G$ is called a multivariate extreme value distribution which is the limiting distribution function of the vector component-wise maxima of $(\mathbf{X}_1, \cdots, \mathbf{X}_n)$.
Componentwise maxima approach

- Extensive literature along this direction—max-stable models
  - Model asymptotic independent variables as exactly independent (Ledford & Tawn, 1996)
  - Only allows investigation of joint tail
Componentwise maxima approach

- Extensive literature along this direction—\textit{max-stable} models
- Model \textit{asymptotic independent} variables as \textit{exactly independent} (Ledford & Tawn, 1996)
- Only allows investigation of joint tail
Componentwise maxima approach

- Extensive literature along this direction—max-stable models

  - Model asymptotic independent variables as exactly independent (Ledford & Tawn, 1996)

  - Only allows investigation of joint tail
Motivating Problem: air quality data

Problem of interest: to study the extremal dependence structure among air pollutants and to estimate critical (extreme) functionals
Motivation & Background

Conditional Extreme Value Models
  Model
  Inference

Air quality application
Basic idea

1. condition on one variable being extreme
2. examine behavior of remaining variables conditional on having an extreme component
3. Rationale: it is less likely that component-wise maxima occurs together e.g. NO and O₃ (in winter)
Characterizing the tail of a multivariate distribution

Set-up

- \( \mathbf{X} = (X_1, \cdots, X_d); \mathbf{X}_{-i} = (X_1, \cdots, X_{i-1}, X_{i+1}, \cdots, X_d) \)
- \( n \) i.i.d. observations of \( \mathbf{X} \) from unknown distribution \( F \)

To characterize the tail it requires

- \( \mathbb{P}(X_i > u_i) \) needs a univariate threshold
- \( X_i | X_i > u_i \) needs a marginal model
- \( \mathbf{X}_{-i} | X_i > u_i \) needs a dependence model
Marginal Model

1. **GPD model for threshold exceedances, empirical distribution for “non-extremes”**

\[
\begin{align*}
\mathbb{P}(X_i > x + u_i | X_i > u_i) &= \left(1 + \xi_i \frac{x}{\beta_i}\right)^{-\frac{1}{\xi_i}} \quad \text{for } x > 0 \\
\mathbb{P}(X_i \leq u_i) &= \hat{F}_{X_i}(x) \quad \text{for } x \leq u_i
\end{align*}
\]

2. Apply probability integral transform

\[Y_i = -\log\left[-\log\left(\hat{F}_{X_i}(x)\right)\right]\]

to obtain approximately Gumbel distribution i.e.

\[\mathbb{P}(Y_i \leq y) \approx \exp\left(e^{-y}\right)\]

3. Henceforth assume marginal distributions are exactly Gumbel and concentrate on dependence among \(Y_1, \cdots, Y_d\)
Marginal Model

1. GPD model for threshold exceedances, empirical distribution for “non-extremes”

\[
\begin{align*}
\mathbb{P}(X_i > x + u_i | X_i > u_i) &= \left(1 + \frac{\xi_i x}{\beta_i}\right)^{-\frac{1}{\xi_i}} \quad \text{for } x > 0 \\
\mathbb{P}(X_i \leq u_i) &= \hat{F}_{X_i}(x) \quad \text{for } x \leq u_i
\end{align*}
\]

2. Apply probability integral transform

\[ Y_i = -\log \left[ -\log \left( \hat{F}_{X_i}(x) \right) \right] \] to obtain approximately Gumbel distribution i.e. \( \mathbb{P}(Y_i \leq y) \approx \exp(e^{-y}) \)

3. Henceforth assume marginal distributions are exactly Gumbel and concentrate on dependence among \( Y_1, \cdots, Y_d \)
Marginal Model

1. **GPD** model for threshold exceedances, empirical distribution for “non-extremes”

\[
\begin{align*}
\mathbb{P}(X_i > x + u_i | X_i > u_i) &= \left(1 + \xi_i \frac{x}{\beta_i}\right)^{-\frac{1}{\xi_i}} \quad \text{for } x > 0 \\
\mathbb{P}(X_i \leq u_i) &= \hat{F}_{X_i}(x) \quad \text{for } x \leq u_i
\end{align*}
\]

2. Apply probability integral transform

\[
Y_i = -\log \left[ -\log \left( \hat{F}_{X_i}(x) \right) \right]
\]


\text{to obtain approximately Gumbel distribution i.e. } \mathbb{P}(Y_i \leq y) \approx \exp \left( e^{-y} \right)

3. Henceforth assume marginal distributions are exactly Gumbel and concentrate on dependence among \(Y_1, \cdots, Y_d\)
Existing techniques

Most existing extreme value methods with Gumbel margins reduce to

\[ P(Y \in t + A) \approx e^{-\frac{t}{\eta_Y}} P(Y \in A) \]

for some \( \eta_Y \in (0, 1] \)

**Ledford-Tawn classification**

- \( \eta_Y = 1 \): asymptotic dependence
- \( \frac{1}{d} < \eta_Y < 1 \): positive extremal dependence
- \( 0 < \eta_Y < \frac{1}{d} \): negative extremal dependence
- \( \eta_Y = \frac{1}{d} \): near extremal independence

**Disadvantage:** doesn't work for extreme sets that are not simultaneously extreme in all components
Define $\mathbf{Y}_{-i} = (Y_1, \cdots, Y_{i-1}, Y_{i+1}, \cdots, Y_d)$. For d-dimensional $\mathbf{Y}$, consider for each $i = 1, \cdots, d$

$$\mathbb{P}(\mathbf{Y}_{-i} \leq \mathbf{y}_{-i} | Y_i = y_i)$$

as $y_i \to \infty$
Asymptotic assumptions

Assume there exist vector normalizing functions

- $a_i(y_i) : \mathbb{R} \mapsto \mathbb{R}^{d-1}$
- $b_i(y_i) : \mathbb{R} \mapsto \mathbb{R}^{d-1}$

such that for any fixed $z_{-i}$ and any sequence $y_i \to \infty$

$$\lim_{y_i \to \infty} \mathbb{P} \left( \frac{Y_{-i} - a_i(y_i)}{b_i(y_i)} \leq z_{-i} \mid Y_i = y_i \right) = G_{|i}(z_{-i})$$

where $G_{|i}$ has non-degenerate margins
Normalizing functions

The authors examined a wide range of multivariate extremal dependence models. Here are some of the key functions:

<table>
<thead>
<tr>
<th>Extremal dependence</th>
<th>$\eta$</th>
<th>$a(y)$</th>
<th>$b(y)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perfect pos. dependence</td>
<td>1</td>
<td>$y$</td>
<td>1</td>
</tr>
<tr>
<td>Bivariate EVD</td>
<td>1</td>
<td>$y$</td>
<td>1</td>
</tr>
<tr>
<td>Bivariate Normal ($\rho &gt; 0$)</td>
<td>$\frac{1+\rho}{2}$</td>
<td>$\rho^2 y$</td>
<td>$y^{\frac{1}{2}}$</td>
</tr>
<tr>
<td>Inverted logistic ($\alpha \in (0, 1]$)</td>
<td>$2^{-\alpha}$</td>
<td>0</td>
<td>$y^{1-\alpha}$</td>
</tr>
<tr>
<td>Independent</td>
<td>$\frac{1}{2}$</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Bivariate Normal ($\rho &lt; 0$)</td>
<td>$\frac{1+\rho}{2}$</td>
<td>$- \log(\rho^2 y)$</td>
<td>$y^{-\frac{1}{2}}$</td>
</tr>
<tr>
<td>Perfect neg. dependence</td>
<td>0</td>
<td>$- \log(y)$</td>
<td>1</td>
</tr>
</tbody>
</table>
They found

- \( a_i(y) \) and \( b_i(y) \) fall in the parametric family:

\[
\begin{align*}
a_i(y) &= a_i y + 1_{\{a_i=0,b_i<0\}} (c_i - d_i \log y) \\
b_i(y) &= y^{b_i}
\end{align*}
\]

- There is no general form for \( G_i \) or for its marginal distributions \( G_{j|i} \).
Model assumptions

Asymptotic structure assumed to hold exactly above high threshold

\[
P\left( \frac{Y_i - a_i(y_i)}{b_i(y_i)} \leq z_i | Y_i = y_i \right) = G_i(z_i) \quad \text{for } y_i > u_i
\]

Then

\[
Z_i = \frac{Y_i - a_i(y_i)}{b_i(y_i)}, \quad \text{for } y_i > u_i
\]

are assumed to follow \( G_i \) and be independent of \( Y_i \)

Remarks:

- parametric forms for \( a_i \) and \( b_i \)
- nonparametric model for \( G_i \)
Estimation of marginal parameters

- Let $\psi = (\beta = (\beta_1, \cdots, \beta_d), \xi = (\xi_1, \cdots, \xi_d))$ denotes parameters for individual GPD’s
- Maximize

$$
\log L(\psi) = \sum_{i=1}^{d} \sum_{k=1}^{n_{uX_i}} \log \hat{f}_{X_i}(x_{i|i,k})
$$

where $u_{X_i}$ is the number of threshold exceedances in $i^{th}$ component and $\hat{f}_{X_i}(x_{i|i,k})$ is the GP density evaluated at the $k^{th}$ exceedance.
Single conditional: estimation of \( \theta_i = (a_i(y_i), b_i(y_i)) \)

- **The problem**: don’t know the distribution of \( Z_i \)
- **The solution**: assume \( Z_i \) has two finite marginal moments. This leads to
  - \( \mu_i(y) = a_i(y) + \mu_i b_i(y) \)
  - \( \sigma_i(y) = \sigma_i b_i(y) \)
- **Estimating equation**

\[
Q_i = - \sum_{j \neq i} \sum_{k=1}^{n_{uy_i}} \left[ \log \sigma_{j|i}(y_{i|i,k}) + \frac{1}{2} \left\{ \frac{y_{j|i,k} - \mu_{j|i}(y_{i|i,k})}{\sigma_{j|i}(y_{i|i,k})} \right\}^2 \right]
\]
All conditionals

To estimate all $a_i(y_i)$ and $b_i(y_i)$ jointly maximize

$$Q = \sum_{i=1}^{d} Q_i$$

**Remarks:**

- Assume independence between conditional distributions
- Analogous with pseudolikelihood estimation (Besag 1975)
Uncertainty

Uncertainty comes from

- semiparametric marginal models
- parametric normalization functions of conditional dependence structure
- nonparametric models for $Z_i$

$\Rightarrow$ Bootstrap
Outline

Motivation & Background

Conditional Extreme Value Models
  Model
  Inference

Air quality application
Air quality monitoring application

- **Data**: daily values of five air pollutants ($O_3$, $NO_2$, $NO$, $SO_2$, $PM_{10}$) in Leeds, U.K., during 1994-1998
- **Two seasons**: winter (NDJF) and early summer (AMJJ)
- **Problem of interest**: to study the extremal dependence among air pollutants and to estimate critical functionals
Transforming data to known margins

Transform variable $X$ to have standard Gumbel marginal distributions
Air quality data–extrapolation

$PM_{10}$ and $NO$: 

![Data, original margins](image1)
![Data, Gumbel margins](image2)
![Monte Carlo sample, Gumbel](image3)
![Monte Carlo sample, original](image4)
Conditional approach: Summary

- accommodates any functional of multivariate extreme events
- handles asymptotic dependence and asymptotic independence including negative dependence
- not just model bivariate extremes
Besag, J.
Statistical analysis of non-lattice data
*The statistician*, 179–195, 1975

Heffernan J. E. & Tawn J. A.
A conditional approach for multivariate extreme values
(with Discussion)

Ledford, A. W., & Tawn, J. A.
Statistics for near independence in multivariate extreme values
Asymptotic (in)dependent

Definition
A bivariate random variable \((X_1, X_2)\) is called asymptotically independent if

\[
\lambda = \lim_{x \to x_F} \mathbb{P}(X_2 > x | X_1 > x) = 0
\]

where \((X_1, X_2)\) are identically distributed with \(x_F = \sup_{x \in \mathbb{R}} \{x : \mathbb{P}(X_1 \leq x) < 1\}\). If \(\lambda > 0\), then \((X_1, X_2)\) is called asymptotic dependent.