Outline

1. Introduction and Motivation
   - A Motivating Example
   - Thresholding Data

2. Data
   - Events
   - Vulnerabilities

3. Methodology
   - Approaches to Thresholding

4. Results
Between 1994-1998: Volcano eruption in Rabaul, Cyclone Justin in the Milne Bay (SE from map selection), and El Niño-induced drought
In the univariate setting thresholding is straightforward...

..the separation of data into regular-valued and extreme-valued portions.
In the univariate setting thresholding is straightforward...

..the separation of data into regular-valued and extreme-valued portions.
Taking multivariate $q$, say, we want to return the set $\mathcal{T}$ such that

$$\mathcal{T} = \{t | F(T > t) > c\}$$  \hspace{1cm} (1)$$

Censor the data:

$$\mathcal{T} \supset \mathcal{T}_* = \{t | t_i > c, \forall i\}$$  \hspace{1cm} (2)$$

And the output is: $F$ for $i = 1, 2$ is $F(T \leq t_*) = F_1 + F_2 - F_1 F_2$

and $F_1 = Pr(T \leq t_*)$; $F_2 = F_1 = Pr(T \leq t | T > t_*)$
In the Multivariate setting this is to fit some contour that partitions multivariate data into

- Regular valued
- Extreme valued
Pop vs. PGA

Density Plot

Lupton, Abayomi, Lacer

MEV Thresholding
Global Natural Disaster Risk Hotspots

Worldwide data has been gridded to $1\frac{1}{2}^\circ$ boxes for 8 predictor variables.

- GDP
- Population
- Peak Ground Acceleration (PGA)
- Floods
- Cyclones
- Drought
- Volcanoes
- Landslides
Global Natural Disaster Risk Hotspots

Worldwide data has been gridded to $1\frac{1}{2}^\circ$ boxes for 8 predictor variables.

- GDP
- Population
- Peak Ground Acceleration (PGA)
- Floods
- Cyclones
- Drought
- Volcanoes
- Landslides
Incidence Maps, gridded to 1.5° lat-lon, 8 variables

- Floods
- Volcano
- Drought
- Earthquake
- GNP: 1990 Gross National Product in US dollars
- Population: Gridded population count (estimate) 1995
.9 ptile of Flood counts
Volcanos

‘.9’ ptile of Volcano incidence
Droughts: Classifying a drought.

Example of a drought event defined by monthly precipitation being below a threshold of 75% of the long-term median value for at least 3 consecutive months. In this case, the duration of the event was 6 months.
Droughts

50 pct Weighted Anomaly Standardized Precipitation (WASP)
75 pct Weighted Anomaly Standardized Precipitation (WASP)
Drought declaration vs. Drought classification
Peak Ground Acceleration
Population Density
Income

GNP
Select Bivariate Plots

- PGA vs. Floods

- PGA vs. Volcanoes

![PGA vs. Floods Plot](image)
Select Bivariate Plots

- PGA vs. Volcanoes

![Graph showing PGA vs. Volcanoes]
We proceed as follows:

- Select a thresholding level
- Fit an extreme-valued parametric model to the data’s tail
- Measure distance between the parametric model and an empirical distribution function
We proceed as follows:

- Select a thresholding level
- Fit an extreme-valued parametric model to the data’s tail
- Measure distance between the parametric model and an empirical distribution function
We proceed as follows:

- Select a thresholding level
- Fit an extreme-valued parametric model to the data’s tail
- Measure distance between the parametric model and an empirical distribution function
We proceed as follows:

- Select a thresholding level
- Fit an extreme-valued parametric model to the data’s tail
- Measure distance between the parametric model and an empirical distribution function
Asymmetric Logistic Distribution (Tawn 1990):

\[ F_{\Theta}(x_1, \ldots, x_d) = \exp \left[ -\sum_{b \in B} \left[ \sum_{j \in b} \left( \frac{\theta_{j,b}}{y_j} \right)^{1/\alpha_b} \right]^{\alpha_b} \right] \]

- \( j \in \{1, \ldots, d\} \), and \( y_j \) is the transformed data
- \( B = \text{PowerSet}\{1, \ldots, d\} \setminus \emptyset \). Hence, \(|B| = 2^d - 1\)
- Say, \( b = \{2, 4, 7\} \), then the inner sum is over \( j = 2, 4, 7 \)
- \( \alpha_b \in (0, 1] \forall b \in B \setminus B_1 \) are dependence parameters
- \( \theta_{j,b} \) are asymmetry parameters, with the constraint: \( \sum_{b \in B(j)} \theta_{j,b} = 1 \) for \( j = 1, \ldots, d \) to force univariate margins to be of the correct form. Here, \( B(j) = \{b \in B : j \in b\} \).
- \(|B(j)| = 2^{d-1}\).
Asymmetric Logistic Distribution (Tawn 1990):

\[ F_\Theta(x_1, \ldots, x_d) = \exp \left[ - \sum_{b \in B} \left( \sum_{j \in b} \left( \frac{\theta_{j,b}}{y_j} \right)^{1/\alpha_b} \right)^{\alpha_b} \right] \]

- \( j \in \{1, \ldots, d\} \), and \( y_j \) is the transformed data
- \( B = \text{PowerSet}\{1, \ldots, d\} \setminus \emptyset \). Hence, \(|B| = 2^d - 1\)
- Say, \( b = \{2, 4, 7\} \), then the inner sum is over \( j = 2, 4, 7 \)
- \( \alpha_b \in (0, 1] \ \forall \ b \in B \setminus B_1 \) are dependence parameters
- \( \theta_{j,b} \) are asymmetry parameters, with the constraint:
  \[ \sum_{b \in B(j)} \theta_{j,b} = 1 \] for \( j = 1, \ldots, d \) to force univariate margins to be of the correct form. Here, \( B(j) = \{ b \in B : j \in b \} \).
- \(|B(j)| = 2^{d-1}\).
Asymmetric Logistic Distribution (Tawn 1990):

\[ F_\Theta(x_1, \ldots, x_d) = \exp \left[ - \sum_{b \in B} \left( \sum_{j \in b} \left( \frac{\theta_{j,b}}{y_j} \right)^{1/\alpha_b} \right)^{\alpha_b} \right] \]

- \( j \in \{1, \ldots, d\} \), and \( y_j \) is the transformed data
- \( B = \text{PowerSet}\{1, \ldots, d\} \setminus \emptyset \). Hence, \(|B| = 2^d - 1\)
- Say, \( b = \{2, 4, 7\} \), then the inner sum is over \( j = 2, 4, 7 \)
- \( \alpha_b \in (0, 1] \ \forall \ b \in B \setminus B_1 \) are dependence parameters
- \( \theta_{j,b} \) are asymmetry parameters, with the constraint:
  \( \sum_{b \in B(j)} \theta_{j,b} = 1 \) for \( j = 1, \ldots, d \) to force univariate margins to be of the correct form. Here, \( B(j) = \{b \in B : j \in b\} \).
- \(|B(j)| = 2^{d-1}\).
Asymmetric Logistic Distribution (Tawn 1990):

$$F_\Theta(x_1, \ldots, x_d) = \exp \left[ -\sum_{b \in B} \left[ \sum_{j \in b} \left( \frac{\theta_{j,b}}{y_j} \right)^{1/\alpha_b} \right]^{\alpha_b} \right]$$

- $j \in \{1, \ldots, d\}$, and $y_j$ is the transformed data
- $B = \text{PowerSet}\{1, \ldots, d\} \setminus \emptyset$. Hence, $|B| = 2^d - 1$
- Say, $b = \{2, 4, 7\}$, then the inner sum is over $j = 2, 4, 7$
- $\alpha_b \in (0, 1] \forall b \in B \setminus B_1$ are dependence parameters
- $\theta_{j,b}$ are asymmetry parameters, with the constraint:
  $$\sum_{b \in B(j)} \theta_{j,b} = 1 \text{ for } j = 1, \ldots, d \text{ to force univariate margins to be of the correct form. Here, } B(j) = \{b \in B : j \in b\}.$$
- $|B(j)| = 2^{d-1}$. 

Asymmetric Model
Asymmetric Logistic Distribution (Tawn 1990):

\[ F_\Theta(x_1, \ldots, x_d) = \exp \left[ - \sum_{b \in B} \left( \sum_{j \in b} \left( \frac{\theta_{j,b}}{y_j} \right)^{1/\alpha_b} \right)^{\alpha_b} \right] \]

- \( j \in \{1, \ldots, d\} \), and \( y_j \) is the transformed data
- \( B = \text{PowerSet}\{1, \ldots, d\} \setminus \emptyset \). Hence, \( |B| = 2^d - 1 \)
- Say, \( b = \{2, 4, 7\} \), then the inner sum is over \( j = 2, 4, 7 \)
- \( \alpha_b \in (0, 1] \ \forall \ b \in B \setminus B_1 \) are dependence parameters
- \( \theta_{j,b} \) are asymmetry parameters, with the constraint: \( \sum_{b \in B(j)} \theta_{j,b} = 1 \) for \( j = 1, \ldots, d \) to force univariate margins to be of the correct form. Here, \( B(j) = \{b \in B : j \in b\} \).
- \( |B(j)| = 2^{d-1} \).
Asymmetric Logistic Distribution (Tawn 1990):

\[ F_{\Theta}(x_1, \ldots, x_d) = \exp \left[ -\sum_{b \in B} \left[ \sum_{j \in b} \left( \frac{\theta_{j,b}}{y_j} \right)^{1/\alpha_b} \right]^{\alpha_b} \right] \]

- \( j \in \{1, \ldots, d\} \), and \( y_j \) is the transformed data
- \( B = \text{PowerSet}\{1, \ldots, d\} \setminus \emptyset \). Hence, \( |B| = 2^d - 1 \)
- Say, \( b = \{2, 4, 7\} \), then the inner sum is over \( j = 2, 4, 7 \)
- \( \alpha_b \in (0, 1] \ \forall \ b \in B \setminus B_1 \) are dependence parameters
- \( \theta_{j,b} \) are asymmetry parameters, with the constraint:
  \( \sum_{b \in B(j)} \theta_{j,b} = 1 \) for \( j = 1, \ldots, d \) to force univariate margins to be of the correct form. Here, \( B(j) = \{b \in B : j \in b\} \).
- \( |B(j)| = 2^{d-1} \).
Asymmetric Logistic Distribution (Tawn 1990):

$$F_{\Theta}(x_1, \ldots, x_d) = \exp \left[ -\sum_{b \in B} \left( \sum_{j \in b} \left( \frac{\theta_{j,b}}{y_j} \right)^{1/\alpha_b} \right)^{\alpha_b} \right]$$

- $j \in \{1, \ldots, d\}$, and $y_j$ is the transformed data
- $B = \text{PowerSet}\{1, \ldots, d\} \setminus \emptyset$. Hence, $|B| = 2^d - 1$
- Say, $b = \{2, 4, 7\}$, then the inner sum is over $j = 2, 4, 7$
- $\alpha_b \in (0, 1] \ \forall \ b \in B \setminus B_1$ are dependence parameters
- $\theta_{j,b}$ are asymmetry parameters, with the constraint:
  \[ \sum_{b \in B(j)} \theta_{j,b} = 1 \] for $j = 1, \ldots, d$ to force univariate margins to be of the correct form. Here, $B(j) = \{ b \in B : j \in b \}$.
- $|B(j)| = 2^{d-1}$. 

Lupton, Abayomi, Lacer  MEV Thresholding
To derive the pdf, we make use of the positive stable (PS) distribution and its Laplace transform (Stephenson 2009):

- \[ \int_0^\infty h_1(s) \exp(-st) ds = \exp(-t^\alpha) \]
- Take \( S_b \sim \text{PS}(\alpha_b) \) \( \forall \ b \in B \setminus B_1 \), and \( S = \{S_b \mid b \in B \setminus B_1\} \).
- Then we have for \( j = 1, \ldots, d \):

\[
\Pr(X_j < x_j \mid S = s) = \exp \left[ - \sum_{b \in B(j)} s_b \left( \frac{\theta_{j,b}}{y_j} \right)^{1/\alpha_b} \right]
\]

while \( X_1, \ldots, X_d \) are conditionally independent given \( S = s \).

Thus, each marginal asymmetric logistic pdf can be given by:

\[
f_j(x_j|s) = \sigma_j^{-1} y_j^{-x_j} \left[ \sum_{b \in B(j)} (z_{j,b}/\alpha_b) \right] \exp \left( - \sum_{b \in B(j)} z_{j,b} \right)
\]

where \( z_{i,b} = s_b(\theta_{i,b}/y_i)^{1/\alpha_b} \)
To derive the pdf, we make use of the positive stable (PS) distribution and its Laplace transform (Stephenson 2009):

- $\int_{0}^{\infty} h_1(s)\exp(-st)ds = \exp(-t^\alpha)$
- Take $S_b \sim PS(\alpha_b) \quad \forall \ b \in B \setminus B_1$, and $S = \{S_b \mid b \in B \setminus B_1\}$.
- Then we have for $j = 1, \ldots, d$:

$$\Pr(X_j < x_j \mid S = s) = \exp \left[ - \sum_{b \in B(j)} s_b \left( \frac{\theta_{j,b}}{y_j} \right)^{1/\alpha_b} \right]$$

while $X_1, \ldots, X_d$ are conditionally independent given $S = s$.

Thus, each marginal asymmetric logistic pdf can be given by:

$$f_j(x_j | s) = \sigma_j^{-1} y_j^{-x_i_j} \left[ \sum_{b \in B(j)} (z_{j,b}/\alpha_b) \right] \exp \left( - \sum_{b \in B(j)} z_{j,b} \right)$$

where $z_{i,b} = s_b(\theta_{i,b}/y_j)^{1/\alpha_b}$.
Conditional Representation

To derive the pdf, we make use of the positive stable (PS) distribution and its Laplace transform (Stephenson 2009):

- \[ \int_0^\infty h_1(s)\exp(-st)ds = \exp(-t^\alpha) \]
- Take \( S_b \sim \text{PS}(\alpha_b) \ \forall \ b \in B \setminus B_1 \), and \( S = \{ S_b \mid b \in B \setminus B_1 \} \).
- Then we have for \( j = 1, \ldots, d \):

\[
\Pr(X_j < x_j \mid S = s) = \exp \left[ - \sum_{b \in B(j)} s_b \left( \frac{\theta_{j,b}}{y_j} \right)^{1/\alpha_b} \right]
\]

while \( X_1, \ldots, X_d \) are conditionally independent given \( S = s \).

Thus, each marginal asymmetric logistic pdf can be given by:

\[
f_j(x_j|s) = \sigma_j^{-1} y_j^{-x_i} \left[ \sum_{b \in B(j)} (Z_{j,b}/\alpha_b) \right] \exp \left( - \sum_{b \in B(j)} z_{j,b} \right)
\]

where \( z_{i,b} = s_b(\theta_{i,b}/y_i)^{1/\alpha_b} \).
To derive the pdf, we make use of the positive stable (PS) distribution and its Laplace transform (Stephenson 2009):

\[ \int_0^\infty h_1(s) \exp(-st) ds = \exp(-t^\alpha) \]

Take \( S_b \sim \text{PS}(\alpha_b) \) \( \forall b \in B \setminus B_1 \), and \( S = \{ S_b \mid b \in B \setminus B_1 \} \).

Then we have for \( j = 1, \ldots, d \):

\[
\Pr(X_j < x_j \mid S = s) = \exp \left[ - \sum_{b \in B(j)} s_b \left( \frac{\theta_j,b}{y_j} \right)^{1/\alpha_b} \right]
\]

while \( X_1, \ldots, X_d \) are conditionally independent given \( S = s \)

Thus, each marginal asymmetric logistic pdf can be given by:

\[
f_j(x_j \mid s) = \sigma_j^{-1} y_j^{-x_i_j} \left[ \sum_{b \in B(j)} (Z_j,b/\alpha_b) \right] \exp \left[ - \sum_{b \in B(j)} z_j,b \right]
\]

where \( z_{j,b} = s_b(\theta_{j,b}/y_j)^{1/\alpha_b} \)
Parameter Estimation

- We begin by estimating the marginal parameters $(\mu_j, \sigma_j, \text{ and } \xi_j)$ from univariate data and keep them fixed throughout.

- Simplifying assumptions: we consider high-dimensional (5 and more) asymmetry parameters to be trivial; also, we assume a non-informative prior.

- To obtain estimates for $\alpha$ and $\theta$, we use Metropolis-Hastings within Gibbs to calculate conditional posterior means.
We begin by estimating the marginal parameters \((\mu_j, \sigma_j, \text{ and } \xi_j)\) from univariate data and keep them fixed throughout.

Simplifying assumptions: we consider high-dimensional (5 and more) asymmetry parameters to be trivial; also, we assume a non-informative prior.

To obtain estimates for \(\alpha\) and \(\theta\), we use Metropolis-Hastings within Gibbs to calculate conditional posterior means.
We begin by estimating the marginal parameters \((\mu_j, \sigma_j, \text{ and } \xi_j)\) from univariate data and keep them fixed throughout.

Simplifying assumptions: we consider high-dimensional (5 and more) asymmetry parameters to be trivial; also, we assume a non-informative prior.

To obtain estimates for \(\alpha\) and \(\theta\), we use Metropolis-Hastings within Gibbs to calculate conditional posterior means.
To select the best threshold, we minimize distances between our parametric fit $F_{\hat{\theta}}$ and the empirical distribution function $\hat{F}_n$ – which is given by:

$$\hat{F}_n(t_1, \ldots, t_d) = \frac{1}{nk} \sum_{j=1}^{d} \sum_{i=1}^{n} 1\{x_{ij} < t_j\}$$
To select the best threshold, we minimize distances between our parametric fit $F_{\hat{\theta}}$ and the empirical distribution function $\hat{F}_n$ – which is given by:

$$\hat{F}_n(t_1, \ldots, t_d) = \frac{1}{nk} \sum_{j=1}^{d} \sum_{i=1}^{n} 1\{x_{ij} < t_j\}$$
To select the best threshold, we minimize distances between our parametric fit $F_{\hat{\theta}}$ and the empirical distribution function $\hat{F}_n$ – which is given by:

$$\hat{F}_n(t_1, \ldots, t_d) = \frac{1}{nk} \sum_{j=1}^{d} \sum_{i=1}^{n} 1\{x_{ij} < t_j\}$$
Pickands suggesting minimizing KS distance

\[ d_k = \sup_q |\hat{F}_n(t) - \hat{F}_\theta(t)| \]

with \( k = 1, 2, \ldots [n/4] \)
Joe suggests computing measure of association and setting cutoff to maximize tail dependence

\[
\max_k \tau_{1-k/n} = \max \tau(t | t > C_k) = \max_k 4E[C_\theta(t | t > C_k)] - 1
\]

[Joe 1992]
Generalization of Joe Type

Maximum likelihood over minimum distance:

$$\max_{\theta} \min_{k} d_{\theta}(q, C_{k,\theta})$$

$$= \max_{\theta} \min_{k} E[\ln(\frac{dG_{\theta}(q)}{dG_{\theta}(C_{k})})]$$
Kendall’s Tau on tails

<table>
<thead>
<tr>
<th>$\tau_{1-k/n}$</th>
<th>$\tau_{.9}$</th>
<th>$\tau_{.95}$</th>
<th>$\tau_{.99}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pop-Pga</td>
<td>.072</td>
<td>.186</td>
<td>.472</td>
</tr>
<tr>
<td>GNP-Flood</td>
<td>.113</td>
<td>.270</td>
<td>.326</td>
</tr>
<tr>
<td>GNP-Drought</td>
<td>.208</td>
<td>.290</td>
<td>.168</td>
</tr>
</tbody>
</table>
70-percentile
75-percentile
80-percentile
95-percentile

Lupton, Abayomi, Lacer

MEV Thresholding
99-percentile
We fit a flexible model to high-dimensional data. This framework allows for the identification of multivariate extremes via either $\mathcal{L}^1$ or Pickands distance, Kullback-Liebler or Expected Entropic Distance.
Summary

- The method (on data ending in 2003) identified several, *post hoc*, locations → Haiti.
- Compare thresholded ‘hotspots’ with disaster record from 2003-2010.