Extremal Dependence: A Few Ideas for Where to Start

Robert L. Wolpert

Department of Statistical Science
Nicholas School of the Environment
Duke University
Draft 3

Mon, 2011 Oct 17

Four Half-baked Ideas

Our subgroup might consider:

1. Building testbed multivariate extreme examples for which inference is tractable, and considering their theoretical properties
2. Thinking harder about the “inbetween” range for dependent but extremally independent random variables (e.g. correlated Gaussians)
3. Exploring various measures of dependence in the literature
4. Thinking harder about threshold models for multivariate extremes

Half-baked Idea I: A Few Testbed Examples

If $\{X_i\} \sim \mathrm{Ca}(\gamma_i)$ have independent centered Cauchy distributions

$$f(x \mid \gamma) = \frac{\gamma/\pi}{\gamma^2 + x^2}$$

and if $a \in \mathbb{R}^n$ then

$$Y := \sum a_i X_i \sim \mathrm{Ca}\left(\sum |a_i| \gamma_i\right)$$

and, for large u, $X \sim \mathrm{Ca}(\gamma)$ satisfies

$$\mathrm{P}[X > u] = \frac{1}{\pi} \arctan\left(\frac{\gamma}{u}\right) \approx \frac{\gamma}{\pi\ u},$$

while the Unit Fréchet Y has

$$\mathrm{P}[Y > u] = 1 - e^{-1/u} \qquad \approx \frac{1}{u},$$

so $\frac{\pi}{\gamma} X$ has Unit Fréchet tails.

Linear Combinations of Cauchys

If $\{\zeta_i\} \overset{\text{iid}}{\sim} \mathsf{Ca}(\gamma = \pi)$ and $a, b \in S_n$, then

$$X = \sum a_i \zeta_i \qquad \text{and} \qquad Y = \sum b_i \zeta_i$$

have unit Fréchet tails and $u \gg 1 \Rightarrow$

$$\begin{aligned} \chi &\equiv \lim \mathsf{P}[Y > u \mid X > u] \\ &\approx \sum \mathsf{P}[b_i\zeta_i > u \mid a_i\zeta_i > u]\ \mathsf{P}[a_i\zeta_i > u \mid X > u] \\ &\approx \sum \left(1 \wedge \frac{b_i}{a_i}\right) \frac{a_i}{\sum a_j} = \sum (a_i \wedge b_i) \end{aligned}$$

Extremally Dependent if any $(a_i \wedge b_i) > 0$.

Linear Combinations of Cauchys

If $\{\zeta_i\} \overset{\text{iid}}{\sim} \mathsf{Ca}\,(\gamma = \pi)$ and $a, b \in S_n$, then

$$X = \sum a_i \zeta_i \qquad \text{and} \qquad Y = \sum b_i \zeta_i$$

have unit Fréchet tails and $u \gg 1 \Rightarrow$

$$\begin{aligned} \chi &\equiv \lim \mathsf{P}[Y > u \mid X > u] \\ &\approx \sum \mathsf{P}[b_i\zeta_i > u \mid a_i\zeta_i > u]\;\mathsf{P}[a_i\zeta_i > u \mid X > u] \\ &\approx \sum \left(1 \wedge \frac{b_i}{a_i}\right)\;\frac{a_i}{\sum a_j} = \sum (a_i \wedge b_i) \end{aligned}$$

Extremally Dependent if any $(a_i \wedge b_i) > 0$.

Cauchys (cont)

$$X = \sum a_i \zeta_i \qquad Y = \sum b_i \zeta_i \qquad a, b \in S_n, \quad \{\zeta_i\} \overset{\text{iid}}{\sim} \mathsf{Ca}(\pi)$$

$$\mathsf{P}[X + Y > u] \approx \sum (a_i + b_i)/u = 2/u$$

$$\mathsf{P}\left[\frac{(X,Y)}{X+Y} = \frac{(a_i, b_i)}{a_i + b_i}\right] \approx \mathsf{P}[\ \zeta_i \text{ is huge } \mid X + Y \text{ is huge }]$$

$$\approx \frac{(a_i + b_i)/u}{2/u} = (a_i + b_i)/2$$

SO, can achieve *any* discrete Spectral Measure w/mean $(\frac{1}{2}, \frac{1}{2})$:

$$H(d\sigma) = \sum p_i\ \delta_{\sigma_i}(d\sigma), \qquad \sigma_i = (x_i, 1 - x_i) \in S_1$$

on the one-simplex S_1 by linear combinations of Cauchys, with

$$a_i = 2x_i p_i \qquad\qquad p_i = (a_i + b_i)/2$$

$$b_i = 2x_i(1 - p_i) \qquad\qquad x_i = a_i/(a_i + b_i)$$

Cauchys (cont)

$$X = \sum a_i \zeta_i \qquad Y = \sum b_i \zeta_i \qquad a, b \in S_n, \quad \{\zeta_i\} \overset{\text{iid}}{\sim} \text{Ca}(\pi)$$

$$\text{P}[X + Y > u] \approx \sum (a_i + b_i)/u = 2/u$$

$$\text{P}\left[\frac{(X, Y)}{X + Y} = \frac{(a_i, b_i)}{a_i + b_i}\right] \approx \text{P}[\ \zeta_i \text{ is huge } \mid X + Y \text{ is huge }]$$

$$\approx \frac{(a_i + b_i)/u}{2/u} = (a_i + b_i)/2$$

SO, can achieve *any* discrete Spectral Measure w/mean $(\frac{1}{2}, \frac{1}{2})$:

$$H(d\sigma) = \sum p_i \ \delta_{\sigma_i}(d\sigma), \qquad \sigma_i = (x_i, 1 - x_i) \in S_1$$

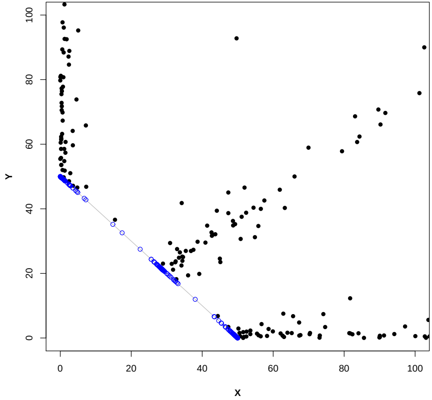
on the one-simplex S_1 by linear combinations of Cauchys, with

$$a_i = 2x_i p_i \qquad p_i = (a_i + b_i)/2$$

$$b_i = 2x_i(1 - p_i) \qquad x_i = a_i/(a_i + b_i)$$

Bivariate Spectrum

Example with $n = 3$:



Half-baked Ideas I: A few extensions

- Multivariate case is no harder:

$$X_i = \sum_j A_{ij}\zeta_j, \qquad \sum_j A_{ij} = 1$$

- Infinite collections (*e.g.*, AR(1) time series) are okay too
- Arbitrary ("spatial") dependence okay
- The general α-stable $\mathrm{St_A}(\alpha, \beta, \gamma, \delta)$ isn't much harder than the Cauchy

It's not much more interesting, either.

Half-baked Ideas I: A few extensions

- Multivariate case is no harder:

$$X_i = \sum_j A_{ij}\zeta_j, \qquad \sum_j A_{ij} = 1$$

- Infinite collections (*e.g.*, AR(1) time series) are okay too
- Arbitrary ("spatial") dependence okay
- The general α-stable $\mathrm{St_A}(\alpha, \beta, \gamma, \delta)$ isn't much harder than the Cauchy

It's not much more interesting, either.

Half-baked Ideas I: A few extensions

- Multivariate case is no harder:

$$X_i = \sum_j A_{ij} \zeta_j, \qquad \sum_j A_{ij} = 1$$

- Infinite collections (*e.g.*, AR(1) time series) are okay too
- Arbitrary ("spatial") dependence okay
- The general α-stable $\mathrm{St_A}(\alpha, \beta, \gamma, \delta)$ isn't much harder than the Cauchy

It's not much more interesting, either.

Half-baked Ideas I: A few extensions

- Multivariate case is no harder:

$$X_i = \sum_j A_{ij}\zeta_j, \qquad \sum_j A_{ij} = 1$$

- Infinite collections (*e.g.*, AR(1) time series) are okay too
- Arbitrary ("spatial") dependence okay
- The general α-stable $\mathsf{St}_\mathsf{A}(\alpha, \beta, \gamma, \delta)$ isn't much harder than the Cauchy

It's not much more interesting, either.

Half-baked Ideas I: A few extensions

- Multivariate case is no harder:

$$X_i = \sum_j A_{ij}\zeta_j, \qquad \sum_j A_{ij} = 1$$

- Infinite collections (*e.g.*, AR(1) time series) are okay too
- Arbitrary ("spatial") dependence okay
- The general α-stable $\mathsf{St_A}(\alpha, \beta, \gamma, \delta)$ isn't much harder than the Cauchy

It's not much more interesting, either.

Half-baked Ideas II: The Inbetween Range

- Precipitation extremes for nearby locations x, y likely exhibit extremal dependence, because their biggest storms coincide.
- As the distance between locations x, y increases, this gradually ceases to be true.
- Most well-studied models are dichotomous: All locations are tail-dependent, or all are tail-independent.
- Perhaps we can overcome this with *finite-range covariance structures* (like *spherical covariance*) in models of Smith (1990, *unp.*), Schlather (2002 *Extremes*), etc.
- Or a convex threshold like $x^p + y^p \geq u^p$ for $0 < p < 1$ or even $p_u \to 0$, excluding areas near the axes.

Half-baked Ideas II: The Inbetween Range

- Precipitation extremes for nearby locations x, y likely exhibit extremal dependence, because their biggest storms coincide.
- As the distance between locations x, y increases, this gradually ceases to be true.
- Most well-studied models are dichotomous: All locations are tail-dependent, or all are tail-independent.
- Perhaps we can overcome this with *finite-range covariance structures* (like *spherical covariance*) in models of Smith (1990, *unp.*), Schlather (2002 *Extremes*), etc.
- Or a convex threshold like $x^p + y^p \geq u^p$ for $0 < p < 1$ or even $p_u \to 0$, excluding areas near the axes.

Half-baked Ideas II: The Inbetween Range

- Precipitation extremes for nearby locations x, y likely exhibit extremal dependence, because their biggest storms coincide.
- As the distance between locations x, y increases, this gradually ceases to be true.
- Most well-studied models are dichotomous: All locations are tail-dependent, or all are tail-independent.
- Perhaps we can overcome this with *finite-range covariance structures* (like *spherical covariance*) in models of Smith (1990, *unp.*), Schlather (2002 *Extremes*), etc.
- Or a convex threshold like $x^p + y^p \geq u^p$ for $0 < p < 1$ or even $p_u \to 0$, excluding areas near the axes.

Half-baked Ideas II: The Inbetween Range

- Precipitation extremes for nearby locations x, y likely exhibit extremal dependence, because their biggest storms coincide.
- As the distance between locations x, y increases, this gradually ceases to be true.
- Most well-studied models are dichotomous: All locations are tail-dependent, or all are tail-independent.
- Perhaps we can overcome this with *finite-range covariance structures* (like *spherical covariance*) in models of Smith (1990, *unp.*), Schlather (2002 *Extremes*), etc.
- Or a convex threshold like $x^p + y^p \geq u^p$ for $0 < p < 1$ or even $p_u \to 0$, excluding areas near the axes.

Half-baked Ideas II: The Inbetween Range

- Precipitation extremes for nearby locations x, y likely exhibit extremal dependence, because their biggest storms coincide.
- As the distance between locations x, y increases, this gradually ceases to be true.
- Most well-studied models are dichotomous: All locations are tail-dependent, or all are tail-independent.
- Perhaps we can overcome this with *finite-range covariance structures* (like *spherical covariance*) in models of Smith (1990, *unp.*), Schlather (2002 *Extremes*), etc.
- Or a convex threshold like $x^p + y^p \geq u^p$ for $0 < p < 1$ or even $p_u \to 0$, excluding areas near the axes.

Half-baked Ideas III: Measures of Dependence

We might read through papers like Coles, Hefferman, & Tawn (1999, *Extremes*): Dependence measures for extreme value analysis to explore the measures of tail dependence (χ, $\bar{\chi}$, etc.) they recommend, and see how they would behave in

- toy examples, like the Cauchy models above;
- abstract models, like those of
  - Smith (1990, *unp.*),
  - Schlather (2002 *Extremes*),
  - Kabluchko/Schlather/de Haan (2009 *AoP*)
- realistic ones with real data (weather, finance, volcanos)—

with a view of going beyond existing measures to capture important features.

Half-baked Ideas III: Measures of Dependence

We might read through papers like Coles, Hefferman, & Tawn (1999, *Extremes*): Dependence measures for extreme value analysis to explore the measures of tail dependence (χ, $\bar{\chi}$, etc.) they recommend, and see how they would behave in

- toy examples, like the Cauchy models above;
- abstract models, like those of
  - Smith (1990, *unp.*),
  - Schlather (2002 *Extremes*),
  - Kabluchko/Schlather/de Haan (2009 *AoP*)
- realistic ones with real data (weather, finance, volcanos)—

with a view of going beyond existing measures to capture important features.

Half-baked Ideas IV: Threshold Models

For univariate extremes there is a well-understood duality between Block Maximum methods (for example, reducing a longitudinal data set to its sequences of annual maxima) and Threshold Exceedance methods (for example, extracting from a longitudinal data set the subset of observations exceeding some large value u).

Multivariate ("Max Stable") methods are well-established for Block Maxima–

Not so much for Threshold Exceedance data.

We have some preliminary ideas... but are open to your

Suggestions?

Half-baked Ideas IV: Threshold Models

For univariate extremes there is a well-understood duality between Block Maximum methods (for example, reducing a longitudinal data set to its sequences of annual maxima) and Threshold Exceedance methods (for example, extracting from a longitudinal data set the subset of observations exceeding some large value u).

Multivariate ("Max Stable") methods are well-established for Block Maxima–

Not so much for Threshold Exceedance data.

We have some preliminary ideas... but are open to your

Suggestions?

Half-baked Ideas IV: Threshold Models

For univariate extremes there is a well-understood duality between Block Maximum methods (for example, reducing a longitudinal data set to its sequences of annual maxima) and Threshold Exceedance methods (for example, extracting from a longitudinal data set the subset of observations exceeding some large value u).

Multivariate ("Max Stable") methods are well-established for Block Maxima–

Not so much for Threshold Exceedance data.

We have some preliminary ideas... but are open to your

Suggestions?

Half-baked Ideas IV: Threshold Models

For univariate extremes there is a well-understood duality between Block Maximum methods (for example, reducing a longitudinal data set to its sequences of annual maxima) and Threshold Exceedance methods (for example, extracting from a longitudinal data set the subset of observations exceeding some large value u).

Multivariate ("Max Stable") methods are well-established for Block Maxima–

Not so much for Threshold Exceedance data.

We have some preliminary ideas... but are open to your

Suggestions?

Half-baked Ideas IV: Threshold Models

For univariate extremes there is a well-understood duality between Block Maximum methods (for example, reducing a longitudinal data set to its sequences of annual maxima) and Threshold Exceedance methods (for example, extracting from a longitudinal data set the subset of observations exceeding some large value u).

Multivariate ("Max Stable") methods are well-established for Block Maxima–

Not so much for Threshold Exceedance data.

We have some preliminary ideas... but are open to your

Suggestions?