Extreme value statistics: from one dimension to many

Lecture 1: one dimension

Lecture 2: many dimensions
• The challenge for extreme value statistics right now: to go from 1 or 2 dimensions to 50 or more

• The challenge for computation: To maximize fifty-dimensional likelihoods with hundreds of parameters

• The challenge for probability: To construct parametric models which makes this possible – and to understand the models (and then model validation, asymptotics, ...)

Extreme rainfall at many catchments, flooding of many dykes, storm insurance, landslide risk assessment, consistent risk estimation for financial portfolios, extreme influenza epidemics, ...
Notation

Bold symbols are d-variate vectors. For instance, $\mathbf{y} = (\gamma_1, \ldots, \gamma_d)$ and $\mathbf{0} = (0, \ldots, 0) \in \mathbb{R}^d$. Operations and relations are componentwise, with shorter vectors recycled.

For instance $a\mathbf{x} + \mathbf{b} = (a_1x_1 + b_1, \ldots, a_dx_d + b_d)$, $\mathbf{x} \leq \mathbf{y}$ if $x_j \leq y_j$ for $j = 1, \ldots, d$, and $t\mathbf{y} = (t\gamma_1, \ldots, t\gamma_d)$.
$d$-dimensional generalized extreme value (GEV) distributions

- No finite-dimensional parametrization

- Marginal distributions GEV: $G_i(x) = e^{-\left(1+\frac{x-\mu_i}{\sigma_i}\right)^{-1/\gamma_i}^+}$

- Can be expressed through “spectral representations”.

- Can be expressed in terms of point processes, see later slides
Let $\eta \in [-\infty, \infty]^d$ be the vector of lower endpoints of the marginal distributions of $G$.

The GEV distributions are the limit distributions of maxima: Let $X_1, X_2, \ldots$ be an i.i.d. sequence of $d$-dimensional vectors with cdf $F$. If for some scaling and location sequences $a_n > 0$ and $b_n$

$$\Pr(a_n^{-1}(\bigvee_{i=1}^n X_i - b_n) \vee \eta \leq x) \to_d G(x), \quad \text{as} \quad n \to \infty$$

where $G$ has non-degenerate margins, then $G$ is a GEV distribution. Conversely all GEV distributions can be obtained in this way.
The GEV distributions are the max-stable distributions: Let $X_1, X_2, \ldots$ be an i.i.d. sequence with nondegenerate marginal cdf-s $G_i$. If for some scaling and location sequences $\alpha_n > 0$ and $\beta_n$

$$\Pr(\alpha_n^{-1} \left( \bigvee_{i=1}^{n} X_i - \beta_n \right) \leq x) = G(x), \quad \text{for } n \to 1, 2, \ldots$$

then $G$ is a GEV distribution. Conversely all GEV distributions can be obtained in this way.
Think of $X_1, X_2, \ldots, X_n$ as points in $\mathbb{R}^d$. Subtract $a_n$ and divide by $b_n$ to get a point process with points

$$\left\{\frac{X_i - a_n}{b_n} \vee \eta; \ i = 1, \ldots, n\right\}.$$

It converges as $n \to \infty$ iff $X \in D(G)$, to a limit point process

$$\{W_i \times R_i^{-\gamma}\}$$

where (in sequel think of $\gamma > 0$) $W_i$ are i.i.d. copies of some ("any") positive stochastic $d$-dimensional vector and the $R_i$ are points of a unit rate Poisson process on $R_+$. 
All d-dimensional GEV distributions may be obtained as

\[ G(x) = Pr(\max_i \{W_i \times R_i^{-Y} \} \leq x) \]

- Often the vector \( W = W^{(d)} \) is thought of as obtained by sampling from a process \( W(s) \) with \( s \in R^d \) (as in figure on previous slide)
- Often \( G \) is defined on a standardized scale, say \( \gamma = 1 \), and model then is transformed to the real scale
- **Smith process**: \( W \) is a non-random Gaussian density function
- **Brown-Resnick process**: \( W(s) = \exp\{X(s) - \sigma^2 / 2\} \), with \( X \) a Gaussian process with variance \( \sigma^2 \)
- **Schlater model**
- ...
The block maxima method

• **As in one dimension:** Get observations $X_1, \ldots, X_n$ of block maxima; assume the observations are i.i.d and follow a GEV distribution; use $X_1, \ldots, X_n$ to estimate the parameters of the GEV distribution

• **Likelihood estimation:** Distribution functions often are possible to calculate, but differentiation with respect to the $d$ component variables often lead to combinatorial explosion of number of terms (because often $G(x) = \exp\{-\ell(x)\}$)

• **Composite likelihood:** Use sum of likelihoods of pairs, or triplets, or ...

• **Stephenson-Tawn likelihoods:** in between block maxima and PoT

• A. C. Davison, S. A. Padoan and M. Ribatet “Statistical Modeling of Spatial Extremes”, *Statistical Science* 2012

• R. Huser, A. C. Davison, M. G. Genton “Likelihood estimators for multivariate extremes”, *Extremes* 2016
• Copula modelling (i.e. modeling after transforming marginal distributions to uniformity) is often used
• Popular copulas: logistic = Gumbel; asymmetric logistic; inverted logistic; Gaussian; Husler-Reiss; bilogistic; Dirichlet; ...
pointwise 25-year return levels for rainfall (mm) obtained from latent variable and max-stable models.  
*Top and bottom rows*: lower and upper bounds of 95% pointwise credible/confidence intervals. *Middle row*: predictive pointwise posterior mean and pointwise estimates.  
*Left column*: latent variable model.  
*Middle column*: Another latent variable model.  
*Right column*: Extremal t copula model.

Copied from Davison, Padoan & Ribatet paper
Multivariate peaks over thresholds modelling and likelihood inference

Or:

News from the exciting exploration of GP-territory

Joint with Anna Kiriliouk, Johan Segers, Maud Thomas, Jennifer Wadsworth
The philosophy is simple

- Extreme episodes are often quite different from ordinary everyday behavior, and ordinary behavior then has little to say about extremes, so that only other extreme events give useful information about future extreme events.

- Operationally: choose thresholds $u_1, \ldots, u_d$, say that an extreme episode occurs if at least one of the components of the observation $Y = (Y_1, \ldots, Y_d)$ exceeds its threshold, only model times of occurrence and undershoots and overshoots, $X = Y - u$, of extreme episodes.

- Theory: times of occurrence follows a Poisson process, $X$ is a generalized Pareto (GP) distribution.
Generalized Pareto (GP) distributions

\[ H(x) = \frac{1}{-\log G(0)} \log \frac{G(x)}{G(x \wedge 0)}, \]

where \( G \) is a GEV cdf with \( G(0) > 0 \).
The GP distributions are the limit distributions threshold excesses: Let \( X \sim F \). If there exist continuous threshold and scaling functions \( u_t \) and \( s_t > 0 \), with \( F(u_t) < 1 \) and \( F(u_t) \to 1 \) as \( t \to \infty \), such that

\[
Pr\left( s_t^{-1} (X - u_t) \leq x \mid X \neq u_t \right) \to_d H(x), \quad \text{as } n \to \infty,
\]

where \( H \) has non-degenerate margins, then \( H \) is a GP distribution. Conversely all GP distributions can be obtained in this way.

- maxima of \( F \) are in “the domain of attraction of a GEV cdf” if and only if threshold excesses of \( F \) are in “the threshold domain of attraction” of a GP cdf \( H \)
The GP distributions are the threshold-stable distribution: Let $X \sim H$, with $H$ nondegenerate. If there exist continuous threshold and scaling functions $u_t, s_t \geq 0$ and with $F(u_t) < 1$ and $F(u_t) \to 1$ as $t \to \infty$, such that

$$Pr\left(s_t^{-1}(X - u_t) \leq x \mid X > u_t\right) = H(x), \quad \text{for all } t \geq 1,$$

then $H$ is a GP distribution. Conversely all GP distributions has this property.
Properties of the class of GP distributions

- Closed under
  - Scaling
  - Increase of level
  - Taking limits
  - Mixing
  - Going to conditional margins

- Not closed under
  - Location changes
  - Going to (unconditional) margins

- Conditional 1-d marginals GP: \( H_j^+(x) = 1 - \left(1 + \frac{\gamma_j}{\sigma_j} x\right)^{-1/\gamma_j} \)

\( (= 1 - e^{-x/\sigma_j} \text{ if } \gamma_j = 0) \)
Multivariate PoT modelling

Should ideally contain three components:

1) A model for the behavior of extreme episodes in the physical world

2) Understanding of how conditioning on threshold exceedance changes the distribution obtained in 1)

3) The possibility to incorporate trends in the models (→ likelihood inference)

Inference should be for the original observations, and not after (approximate) transformation of margins of observations to some standard form, say standard Pareto or standard GP
Think of $X_1, X_2, ..., X_n$ as points in $R^d$. Subtract $a_n$ and divide by $b_n$ to get a point process with points

$$\left\{ \frac{X_i - a_n}{b_n} \vee \eta; \ i = 1, ... n \right\}.$$

It converges as $n \rightarrow \infty$ iff $X \in D(G)$, to a limit point process

$$\{ W_i \times R_i^{-y} \}$$

where (in sequel think of $y > 0$) $W_i$ are i.i.d. copies of some ("any") positive stochastic $d$-dimensional vector and the $R_i$ are points of a unit rate Poisson process on $R_+$.
A typical point (positive shape)

\[ R \times T^{-\gamma} \]

- \( R \) an “arbitrary” \( d \)-dimensional random vector, subject to \( E(R_j) < \infty \), and \( R_j \geq 0 \) if \( \gamma_j > 0 \)
- \( T \) improper random variable, has density \( = 1 \) with respect to Lebesgue measure on \( R_+ \)
- \( R \) and \( T \) are mutually independent

- If \( \gamma_j > 0 \) then \( T^{-\gamma_j} \) has density \( x^{\frac{-1}{\gamma_j}-1} \) with respect to Lebesgue measure on \( R_+ \)
- If \( \gamma_j = 0 \) then \( R_j \times T^{-\gamma_j} \) is defined to mean \( R_j - \log T \) and \( \log T \) has density \( e^{-x} \) with respect to Lebesgue measure on \( R \)
Conditioning on threshold exceedance

\( F \) d.f. of \( \mathbf{R} \) in typical point \( \mathbf{R} \times T^{-\gamma} \).

- \( X_u = X - u \),
- \( \Pr(X_u \leq x|X_u \neq 0) = \frac{\Pr(X_u \leq x) - \Pr(X_u \leq x \wedge 0)}{\Pr(X_u \neq 0)} \)
- Set \( X = \mathbf{W} \times R^{-\gamma} \) and \( u = \frac{\sigma}{\gamma} \) and condition (formally) on \( T \)

\[
H(x) = \frac{\int_0^\infty \{F(t^\gamma (x + \frac{\sigma}{\gamma})) - F(t^\gamma (x \wedge 0 + \frac{\sigma}{\gamma}))\}dt}{\int_0^\infty \bar{F}(t^\gamma \frac{\sigma}{\gamma})dt}
\]

(R) is a GP – and all GP-s can be obtained this way
Three representations

(R) \[
H_R(x) = \frac{\int_{0}^{\infty} \left\{ F_R \left( t \gamma(x + \frac{\sigma}{\gamma}) \right) dt - F_R \left( t \gamma(x \wedge 0 + \frac{\sigma}{\gamma}) \right) \right\} dt}{\int_{0}^{\infty} F_R(t \gamma \sigma) dt}
\]
is a GP – and all GP-s can be obtained this way.

(R_s) \[
H_S(x) = \frac{\int_{0}^{\infty} \left\{ F_S \left( \frac{1}{\gamma} \log(x + \frac{\sigma}{\gamma}) + \log t \right) dt - F_S \left( \frac{1}{\gamma} \log(x \wedge 0 + \frac{\sigma}{\gamma}) + \log t \right) \right\} dt}{\int_{0}^{\infty} F_S(\log t) dt}
\]
is a GP – and all GP-s can be obtained this way.

(Rₖ) \[
H_{\tilde{S}}(x) = \int_{0}^{1} F_{\tilde{S}} \left( \frac{1}{\gamma} \log \left( x + \frac{\sigma}{\gamma} \right) + \log t \right) dt
\]
\[
= \int_{0}^{\infty} F_{\tilde{S}} \left( \frac{1}{\gamma} \log \left( x + \frac{\sigma}{\gamma} \right) - t \right) e^{-t} dt
\]
for \( \tilde{S} = S - \max S_j \) – and all GP-s can be obtained this way.

Ferreira de Haan (2014) representation.
Densities

- Explicit formulas for densities for all three representations
- Formula for $\hat{R}_{\bar{s}}$ density simplest !!
- Censored likelihood contributions computable
Estimation

Likelihood for (R):

$$\prod_{i=1}^{N} \frac{\int_{0}^{\infty} t^{\sum_{j=1}^{d} \gamma j f\left(t \gamma \left(x_{i}+\gamma \right)\right)}dt}{\int_{0}^{\infty} F\left(t \gamma \right)dt}$$

- One-dimensional integrals, so relatively easy to compute if $f, \bar{F}$ are easy to compute. In some cases integrals can be computed analytically.

- Censoring makes computation harder

- Parenthesis: Coefficient of asymptotic dependence:

$$\chi_d := \lim_{q \to 1} \frac{\Pr(F_1(X_1) > q, ..., F_d(X_d) > q)}{1 - q}.$$

GEV and GP modelling aimed at asymptotic dependence, i.e. $\chi > 0$ for some subset of components. Substantial literature (Heffernan, Tawn, Resnick, Wadsworth, ...) on modelling under asymptotic independence, when $\chi = 0$
Parametrization and identifiability

• Understanding is on the (R) scale
• Choice between (R), (Rₜ), and (Rₜ̃) a choice of parametrization
• $\sigma$ and $\gamma$ parameters of conditional margins
• Replacing $R$ by $Z^\gamma R$ doesn’t change $H_r$. Similar properties of $H_s$ and $H_{ₜ̃}$
• (Rₜ) and (Rₜ̃) continuous at $\gamma_i = 0$. Not so for (R)

• Much left to explore
Probabilities and conditional probabilities

• Probabilities of general events given by “same” formula as $H_r$.
• Conditional densities same type of formulas.
• Conditional probabilities same type of formulas (prediction).
• If $\gamma_1 = \cdots = \gamma_d =: \gamma$ then weighted sums of components, conditioned to be positive, are also GP.
If $\gamma_1 = \cdots = \gamma_d := \gamma$ then weighted sums of components, conditioned to be positive, are also GP
Simulation

1. Direct for \( (R_S) \)
2. Change of measure + rejection sampling for \( (R) \) and \( (R_S) \)
3. Change of measure + MCMC for \( (R) \) and \( (R_S) \)
4. Approximate method for \( (R) \) and \( (R_S) \)
Before Gudrun

After Gudrun

Windstorm Gudrun, January 2005

SEK 2.8 billion loss (LF)

55% of total loss 1982-2005

Remember: analysis based on data 1982-1993: 1% chance biggest damage next 15-year period will be more than SEK 2.5 billion

???

Bivariate logistic PoT analysis based on data 1982-2005: 10% chance biggest damage next 15-year period will be more than SEK 7 billion (this time we didn’t miss damage to forest – but did we miss something else?)

Brodin & Rootzén, Insurance 2009
Landslides

- Landslide caused by one day with very extreme rainfall, or by two consecutive days with extreme rainfall. Simplest model:

\[ X_1, X_2, \text{ independent positive variables,} \]

\[ \gamma_1 = \gamma_2 = \gamma \]

\[ \frac{R}{T^\gamma} = \left( \frac{X_1 \lor X_2}{T^\gamma}, \frac{X_1 + X_2}{T^\gamma} \right) = \left( \frac{X_1 \lor X_2}{T^\gamma}, \frac{X_1 \lor X_2 + X_1 \land X_2}{T^\gamma} \right) \]

\[ = (\text{rainiest day, rainest day + rainest adjacent day}) \]
- Daily accumulated precipitation data from Abisko between January 1st 1913 to December 31st 2014
- Construct a three-dimensional dataset whose components represent daily, two-day, and three-day extreme rainfall amounts.
- **Structured components model.**
Data: weekly negative returns of four largest UK banks, October 2007 - April 2016 (444 observations)

Asymptotic dependence seems reasonable; $\chi_{1.4}(q)$ looks constant above $q \approx 0.83$
Modelling strategy

1. standardize the data to common GP margins using the rank transformation
2. fit most complicated dependence model within each class to the standardized data
3. select as the dependence model class that which produces the closest fit to the data, in the sense of largest maximized log likelihood (or AIC)
4. use likelihood ratio tests to test for simplification of models within the selected dependence class, and select a final dependence model
5. fit GP margins simultaneously with this dependence model (i.e. fit full MGPD)
6. test for simplifications in the marginal parameterization
Above strategy leads to Gumbel model with a single dependence parameter(!)

- Tripletwise and quadruple $\chi$ all look fine too
Likelihood ratio tests support common shape ($\hat{\gamma} = 0.46$) but separate scale parameters (standard errors reduced by $\approx 30\%$ by use of joint fit).

GPD fit to $\sum_{j=1}^{d} (X_j - u_j) | \sum_{j=1}^{d} (X_j - u_j) > 0$.
Prediction of influenza epidemics in France

- $y_1 =$#cases week 3, ..., $y_8 =$#cases week 11
- $\psi = 0$ (from data), $R_i \sim N(m_i, s_i^2), m_1 = 0$
- **Model**: conditional distribution of $R - \sigma \log T$ given that $R_1 > \sigma \log T$
- Total # cases GP with $\gamma = 0, \sigma = \sigma_1 + \cdots \sigma_8$
- LR-test of parabolic form for the $m_i$, equality of the $s_i^2$
Challenges

- Model construction
- Parametrization
- Time series
- Prediction
- Asymptotics