Motivation

2020 in review . . .

- Apart from Covid-19, in 2020 there were
  - record-breaking wildfires in Australia still burning from 2019,
  - major flash floods in Indonesia and Afghanistan,
  - huge flooding of the Yangtze River in May,
  - a record number of named North Atlantic hurricanes,
  - many severe convective storms (thunderstorms with tornadoes, floods and hail) in the US,
  - major cyclones in the Philippines, Bangladesh and India,
  - huge wildfires in California — outside the usual wildfire season!

- These events were once unusual, but more and worse will come due to the increased energy in the climate system from global heating.

Extrapolation

- For disaster planning, public health, construction (and insurance) we need to extrapolate to
  - the tails of distributions, beyond previous events
  - new conditions in a warmer world.

Motivations for modelling extremes

- Estimation of changes in extremes, for better forecasting.
- Risk assessment at a single important site.
- Risk estimation for particular events:
  - What is the risk of crop failure due to drought over a large region?
  - What might the total insurance payout be in case of a major windstorm, or flooding of a major city?
- Attribution of events to possible causes: to what extent is a heatwave caused by climate change?

- These involve:
  - accurate space/time interpolation or extrapolation;
  - accurate marginal and/or joint modelling of extreme events.
- Risk estimation typically involves bold extrapolation:
  - e.g., prediction of ‘ten-thousand year event’ from 80 years of data.
- Basic problem: the events are rare, their probabilities are small, and there may be little (or no) directly relevant data.
Vargas tragedy

- Following two weeks of intermittent rainfall, torrential rainfall on 14–16 December 1999 spawned landslides throughout the upper watersheds of the Cerro Grande River in the Venezuelan state of Vargas.
- Mud floods, debris flows and flood surges destroyed much of Tanaguarena and other coastal tourist towns. Perhaps 30,000 people died.
- Similar events, fortunately with less loss of life, had previously occurred nearby.

Tanaguarena

Available rainfall data

- Data from the airport at Maiquetia, before the event:
Available rainfall data

- Data from the airport at Maiquetia, including the event:

![Graph showing daily rainfall (mm) from 1950 to 2000.]

Zurich and the river Sihl

![Map of Zurich and the river Sihl.]

- ![Images of Zurich and the river Sihl in different conditions.]
Here there are hourly rainfall totals (mm) on a grid for 2013–2018 from radar measurements (Météoswiss).

Spatial averages and spatial maxima of hourly total rainfall and the two largest events:
Desiderata

Why specialised models?

- Basic problem is **extrapolation** to rarer events.
- Geostatistics is mostly based on multivariate normal distributions, inappropriate for modelling distribution tails.
- Extrapolation from a fit to the entire distribution can be misleading:
  - there may be regime change in the tails,
  - different fits to the bulk may give very different tail estimates—in particular, the light tails of the Gaussian density can grossly underestimate probabilities of rare events,
  - this is worse for multivariate data, for which Gaussian models predict independence of very rare events (‘the formula that killed Wall Street’).
- Standard copulas can deal with transformations to marginal distributions, but not with joint dependence.

---

Gaussian tails and probabilities

- Gaussian (black), Cauchy (red) and $t_{20}$ (blue) densities matched to have probabilities 0.05 for $|Y| > 1.96$
- Ratios of Cauchy/Gaussian and $t_{20}$/Gaussian probabilities for $|Y| > y$:

<table>
<thead>
<tr>
<th>$y$</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cauchy ratio</td>
<td>1.08</td>
<td>12</td>
<td>387</td>
<td>34247</td>
<td>$8.3 \times 10^6$</td>
<td>$5.5 \times 10^9$</td>
</tr>
<tr>
<td>$t_{20}$ ratio</td>
<td>1.01</td>
<td>1.7</td>
<td>6.1</td>
<td>58</td>
<td>1589</td>
<td>$1.3 \times 10^6$</td>
</tr>
</tbody>
</table>
Joint Gaussian tails for rare events

- Conditional probabilities $P(Z_2 > z \mid Z_1 > z)$ as a function of $z$, for
  - bivariate Gaussian data with correlation 0.9 (left)
  - bivariate extreme-value data with extremal coefficient 1.23 (right)

Extremal paradigm

- We need a basis for extrapolation outside the sample, possibly based on a small (extreme) subset of the data.
- The uncertainty will inevitably be large, and must be taken into account.
- Standard models and methods are too limiting, because
  - they cannot accommodate heavy tails
  - their joint tail properties are too inflexible for wide use, and risk underestimating probabilities for joint events.
- Hence the extremal paradigm:
  - Fit asymptotically-justified models which extrapolate 'appropriately'.
Extremal Models

Poisson process
- Random point pattern $\mathcal{P}$ in a state space $\mathcal{E}$ defined by properties of counts
  \[ N(A) = |\{x : x \in \mathcal{P} \cap A\}|, \quad A \subset \mathcal{E} : \]
  - $N(A_1), \ldots, N(A_k)$ independent for disjoint $A_1, \ldots, A_k$,
  - $N(A) \sim \text{Pois}(\mu(A))$,

  where the measure $\mu$ is non-atomic (diffuse), and often has an intensity $\mu$.
- \textbf{Mapping theorem}: if $g : \mathcal{E} \to \mathcal{E}^*$ does not create atoms, then $\mathcal{P}^* = g(\mathcal{P})$ is also a Poisson process.
- Restriction of process $\mathcal{P}$ to $\mathcal{E}' \subset \mathcal{E}$ is also Poisson.

Poisson process convergence
- Let $X_1, \ldots, X_n \overset{iid}{\sim} F$ and for $b_n \in \mathbb{R}$ and $a_n > 0$ define point processes
  \[ \mathcal{P}_n = \left\{ \frac{X_j - b_n}{a_n} : j = 1, \ldots, \right\}, \quad \mathcal{E} = \mathbb{R}. \]
- Then the rescaled maximum
  \[ M_n = \frac{\max(X_1, \ldots, X_n) - b_n}{a_n} \]

  has a non-degenerate limiting distribution iff $\mathcal{P}_n$ converges to a Poisson process with mean measure
  \[ \Lambda(y) \equiv \Lambda\{(y, \infty)\} = \left(1 + \frac{y - \eta}{\tau}\right)^{-1/\xi}, \quad y \in \mathbb{R}, \]

  where $u_+ = \max(u, 0)$ for real $u$.

Extremal types theorem
- $M_n \leq y$ iff $\mathcal{P}_n \cap (y, \infty) = \emptyset$, so
  \[ P(M_n \leq y) \to \exp\{-\Lambda(y)\}, \quad n \to \infty. \]
- The \textbf{Extremal Types Theorem} states that the only possible non-degenerate limiting distribution for $M_n$ is\textbf{ generalized extreme-value (GEV)}:
  \[ G(y) = \begin{cases} \exp\left\{ - \left(1 + \frac{y - \eta}{\tau}\right)^{-1/\xi}\right\}, & \xi \neq 0, \\ \exp\left\{ - \exp\left(-\frac{y - \eta}{\tau}\right)\right\}, & \xi = 0, \end{cases} \]

  where
  - $\eta$ is a real location parameter,
  - $\tau$ is a positive scale parameter,
  - $\xi$ is a real shape parameter, though usually $|\xi| < \frac{1}{2}$ in applications.
- 'Universal' law, explains the use of the GEV for modelling maxima, analogous to use of Gaussian distribution for averages.
- Minima are modelled by transformation $y \mapsto -y$ or by fitting $1 - G(-y)$. 
GEV and ‘three types’

- $\xi > 0$ giving the heavy-tailed Fréchet, Type II, bounded below (red);
- $\xi = 0$ giving the light-tailed Gumbel, Type I, with support on $\mathbb{R}$ (black);
- $\xi < 0$ giving the short-tailed (reverse) Weibull, Type III, bounded above (blue).

The usual Weibull distribution gives a model for minima.

Max-stability

- The underlying idea is max-stability: ‘the maximum of 100 consecutive years of data equals the maximum of the ten decadal maxima’.

This implies that any limiting distribution must satisfy

$$G^T(b_T + a_T y) = G(y), \quad T > 0,$$

where $G^T$ is the shifted and scaled distribution.

Max-stability gives a mathematical basis for extrapolation, though it assumes no regime changes outside the observed data range.
Extrapolation

To extrapolate to $T$-year maxima (below, with $T = 50$ in red) from a GEV fitted to annual maxima, we use

$$G_T^T(y; \eta, \tau, \xi) = G(\tau_T, \tau_T, \xi),$$

where $\eta_T = \eta + \tau(T^\xi - 1)/\xi$ and $\tau_T = \tau T^\xi$.

Point process approximation

Take $X_1, \ldots, X_n \stackrel{iid}{\sim} F$ for which the GEV limit holds and form binomial processes

$$P_n = \left\{ \left( \frac{j}{n + 1}, \frac{X_i - b_j}{a_j} \right) : j = 1, \ldots, n \right\}.$$

As $n \to \infty$, $P_n$ converges to a Poisson process $P$ on $[0, 1] \times \mathbb{R}$ with measure

$$\mu([t_1, t_2] \times (x, \infty)) = (t_2 - t_1) \Lambda(x), \quad 0 \leq t_1 < t_2 \leq 1, x \in \mathbb{R}.$$
Threshold exceedances

□ Exceedances of a threshold \( u \) occur at the times of a homogeneous Poisson process of rate \( \Lambda(u) \), and their sizes are independent with the **generalized Pareto distribution (GPD)**

\[
H(x) = \begin{cases} 
1 - (1 + \xi x/\sigma)^{-1/\xi} & \xi \neq 0, \\
1 - \exp(-x/\sigma) & \xi = 0,
\end{cases} \quad x > 0,
\]

where \( \xi \in \mathbb{R} \) and \( \sigma = \tau + \xi(u - \eta) > 0 \).

□ Use of this approximation is called **peaks over threshold (POT) analysis**.

□ Analogous to the max-stability of the GEV, the GPD is **threshold-stable**:

\[
\frac{1 - H(x + x'; \xi, \sigma)}{1 - H(x'; \xi, \sigma)} = 1 - H(x; \xi, \sigma + \xi x') \quad x, x' > 0,
\]

so it is the natural model for exceedances over high thresholds—and under low ones, by replacing \( H(x) \) with \( 1 - H(-x) \).

□ Extrapolation to higher levels is analogous to the GEV.

---

**Generalized Pareto distribution**

□ A flexible distribution whose density can take a variety of shapes.

□ Left: exponential density (\( \xi = 0 \), black), heavy-tailed density (\( \xi = 0.5 \), red) and light-tailed density (\( \xi = -0.2 \), blue, with upper terminal shown); all have \( \sigma = 1 \).

□ Right: densities with negative shape parameter and upper terminal \( x_F = 1 \), with \( \xi = -1 \) (black), \( \xi = -2 \) (red), \( \xi = -0.5 \) (blue) and \( \xi = -0.8 \) (green).
Shape parameter $\xi$

- $\xi$ has characteristic ranges for different types of data:
  - for rainfall, typically $\xi \approx 0.1$,
  - for extreme hot or cold temperatures, typically $\xi \approx -0.2$,
  - for wind speeds, typically $\xi < 0$,
  - for athletics data, typically $\xi < 0$,
  - for negative financial returns, typically $\xi > 0$, maybe even $\xi > 1$ for very risky assets.

- $\xi$ is difficult to estimate, and its uncertainty dominates extrapolation to higher values, so
  - it is useful to combine data from different (but compatible!) sources to reduce uncertainty, or
  - to use Bayesian methods, if good prior information is available.

- The $r$th moment of the GEV exists only if $\xi < 1/r$, so the mean exists only if $\xi < 1$, the variance only if $\xi < 1/2$, etc.

- In applications (particularly in finance) some moments may not exist.

- MLE does not have its usual properties if $\xi \leq -1/2$.

Statistical implementation

- The GEV and GPD are limit models, but are used as approximations for maxima for finite block size $m$ and for finite threshold $u$, so there is a trade-off:
  - taking $m$ too small/$u$ too low gives more data but estimation may be biased,
  - increasing $m$ or $u$ reduces bias, but may give too little data for useful assessment of uncertainty.

- The GEV and GPD are asymptotic models used for sub-asymptotic data, so there is always a bias, and we must check their fit carefully, using standard plots.

- Methods for automatic choice of $m$ or $u$ have been proposed, but don’t generally perform well, so informal methods based on stability are often used.

- Quantile regression is sometimes used to choose $u$ in big datasets.

- Sensitivity analysis is crucial: our the conclusions should not depend heavily on the block size or the threshold.

- Prefer fitting by likelihood/Bayes methods, which are flexible and general.
Quantiles and return levels

Assume we use the GEV to model annual maxima, \( Y \).

Take \( 0 < p < 1 \) and define the GEV quantile \( y_p \) by \( G(y_p) = p \), giving

\[
y_p = \eta + \tau \left( -\log p \right)^{\frac{\xi}{\tau}} - 1.
\]

We call \( y_p \) the return level associated with the return period \( 1/(1 - p) \), so \( y_{0.95} \) is the 20-year return level, \( y_{0.99} \) is the 100-year return level, etc.

Is this useful in a non-stationary setting?

Measures of risk

In finance the \( p \) quantile is called the value at risk, \( \text{VaR}_p \), (be careful whether \( Y \) represents a profit or a loss!).

Another common financial risk measure is the expected shortfall,

\[
\text{ES}_p = E(Y - y_p \mid Y > y_p),
\]

computed for \( p = 0.95, 0.99, \ldots \).

In engineering contexts one may be interested in the probability that \( Y \) falls into a risk set \( \mathcal{A} \)

\[
P(Y \in \mathcal{A}),
\]

where typically \( \mathcal{A} = [v, \infty) \) or \( (-\infty, v] \) when \( Y \) is scalar.

All estimates of risk measures are likely to have very asymmetric distributions (they are high quantiles or small probabilities), so standard symmetric confidence intervals

\[
\text{Est} \pm 2 \text{SE}
\]

should be avoided in favour of likelihood-based or Bayes intervals.
Estimation or prediction?

- Return levels, values at risk, expected shortfalls and probabilities would be known if we knew the underlying data generating mechanism—they are parameters to be estimated.
- Very often we are interested in future events, e.g., the largest flood $Y_T$ to be seen in the next $T$ years, which is a random variable—even if we knew the data generating mechanism exactly, we should consider $Y_T$ as random until the $T$ years have passed.
- This suggests that we should focus on prediction of future events, not on their probabilities.
- In a Bayesian context, this is (in principle) straightforward, we compute the posterior predictive density of $Y_T$ conditional on the observed data $Y = y$, i.e.,

$$f(y_T | y) = \frac{\int f(y_T | y, \theta) f(y | \theta) \pi(\theta) \, d\theta}{\int f(y | \theta) \pi(\theta) \, d\theta},$$

where $\pi(\theta)$ is the prior density for parameters $\theta$. We could also compute summaries, such as quantiles of $f(y_T | y)$ or its mean $E(Y_T | Y = y)$.
- In a frequentist setting, we may estimate properties of the $T$-year maximum, such as its median or its expectation (if finite).

Examples

Annual maxima for Vargas

![Annual maxima for Vargas](image)
Gumbel QQ plot

Gumbel QQplots for annual maxima, with (left) and without (right) maximum for 1999, and GEV simulation envelopes based on $R = 10^4$ replicate samples.

Effect of uncertainty
Vargas GEV fit

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Location $\eta$</td>
<td>47.15$^{+3.77}_{-3.77}$</td>
<td>47.87$^{+3.73}_{-3.73}$</td>
</tr>
<tr>
<td>Scale $\tau$</td>
<td>20.55$^{+1.29}_{-1.29}$</td>
<td>19.52$^{+2.92}_{-2.92}$</td>
</tr>
<tr>
<td>Shape $\xi$</td>
<td>0.36$^{+0.15}_{-0.15}$</td>
<td>0.14$^{+0.16}_{-0.16}$</td>
</tr>
</tbody>
</table>

Parameter estimates and standard errors with and without the 1999 maximum.

- Note:
  - the sizes of the standard errors relative to the estimates;
  - the large change in $\hat{\xi}$ due to dropping the final maximum;
  - the Gumbel distribution ($\xi = 0$) is well inside a 95% confidence interval for $\xi$ if 1999 is dropped, but not otherwise.

- The largest observation has a huge effect on inferences, particularly for $\xi$.

Predictive densities

Predictive densities for annual daily maximum (black) and 39-year daily maximum (red), based on data without 1999, with the 1999 daily maximum shown by the vertical line:
Gumbel fits to the monthly maxima would be poor:
- there can be zero maxima (March, April?);
- mixtures — some values lie well off the line for the rest;

Peaks over threshold (POT) analysis, $u = 40\text{mm}$
Exploratory techniques

- The mean of the GPD satisfies

\[ E(X - u \mid X > u) = \frac{\sigma_u}{1 - \xi} = \frac{\tau + \xi u}{1 - \xi}, \quad \xi < 1. \]

so if the GPD is applicable above some threshold \( v \), a plot of

\[ \frac{\sum_{j=1}^{n} (x_j - u)I(x_j > u)}{\sum_{j=1}^{n} I(x_j > u)} \text{ vs } u \]

should be a straight line of gradient \( \xi/(1 - \xi) \) when \( u > v \).

- Likewise, if the point process model is appropriate for data above some threshold \( u \), plots of the ML estimates of \( \eta, \tau \) and \( \xi \) based on data above those thresholds should become constant, above \( u \).

- The main issue with these plots (and many others for extremal data) is that the region of interest usually has too few observations, and hence too wide confidence intervals, to draw firm conclusions.

Example: Venezuela rainfall
Mean residual life plot (top left) and parameter stability plots for fits of point process model:
Extremogram

The extremogram for a stationary series \( \{X_t\} \) estimates

\[
\pi_h(u) = P(X_{t+h} > u \mid X_t > u), \quad h = 1, 2, \ldots.
\]

If there is no serial dependence, we should see \( \pi_h(u) = P(X_t > u) \) for all \( h \) (blue in picture, upper 95% point is red).

- This is (almost) the ACF for the time series \( I(X_t > u) \);
- estimated by (almost) the corresponding correlogram;
- beware poor sampling properties—is there an annual cycle for \( X_t > 20 \text{mm} \)?

Comments

- Modelling the threshold exceedances would involve fitting the GPD for some chosen threshold, and would allow more detailed modelling of individual rare events.
- Using annual maxima avoids having to model the seasonality, which could stem from
  - variation in the numbers of large rainfall days, or
  - variation in the sizes of large rainfall amounts, or
  - variation in both due to a mixture of underlying rainfall processes?
- This dataset ends after the largest event, so there is a stopping rule. Ignoring this will bias estimates of risk upwards.
- There is (slight) autocorrelation in large values, leading to (slight) clustering of extremes.
- The basic Poisson process approximation can be extended to allow for local clustering, by including the extremal index \( \theta \in (0, 1] \), which allows clusters of extremes of mean size \( 1/\theta \) and otherwise arbitrary configuration.

Trends in severe US thunderstorms

- Severe thunderstorms can have devastating impacts, leading to major losses of property and life.
- More likely if there are simultaneously high values of
  - convective available potential energy (CAPE) and storm relative helicity (SRH),
  - and hence of \( \text{PROD} = \sqrt{\text{CAPE} \times \text{SRH}} \).
- Fit the GEV to monthly maxima (allows for seasonality), with parameters
  \[
  \eta_{d,t} = \eta_d + \eta'_d \text{ENSO}_t, \quad \tau_d, \quad \xi, \quad d = 1, \ldots, D,
  \]
  depending linearly on El Niño-Southern Oscillation (ENSO) and time.
- Repeat similar analyses for \( D \) grid points and 12 months, so must allow for multiple testing using Benjamini–Hochberg procedure.
- Koch et al. (Journal of Climate, to appear).
Changes in PROD


GEV fit at a random gridpoint

Figure 1: In-sample fit of the GEV: QQ plots for PROD (left), CAPE (centre) and SRH (right) May maxima. The shaded regions indicate the 95% confidence bounds.
Effect of ENSO on SRH in February

Figure 2: Values and significance of the ENSO coefficient $\hat{\eta}_d$ for SRH maxima in February. Large and small circles indicate significance (after accounting for multiple testing using the BH procedure) at any level not lower than 5% and 20%, respectively. The units of $\hat{\eta}_d$ are $m^2 s^{-2} C^{-1}$.

Comments

- This analysis does not allow for spatial dependence, except for the multiple testing step (the BH procedure is OK under a particular type of dependence between the test statistics . . . ).
- The time series at each grid point are treated as independent, but a more efficient inference would allow for spatial dependence.
- Would require models for multivariate extremes (next . . . ).
Multivariate Extremes

Multivariate extremes
- Many climatic/environmental extremal problems are intrinsically multivariate:
  - overwhelming of sea defences by high tides and strong winds;
  - flooding at many locations of a river system;
  - heatwaves have successive very hot days over a wide spatial area.
- More statistically, uncertainty may be reduced by combining information from several sources.
- In one dimension it’s obvious what is ‘extreme’, but we must now consider:
  - what is ‘extreme’ in two or more dimensions?
  - how can we summarize and model extremal dependence of two or more variables?
- One approach is to reduce a multivariate problem to a scalar structure variable.

Structure variable
- A structure variable $S = s(X_1, \ldots, X_D)$ is a univariate function:
  - for example, insurance loss
    \[
    S = \sum_{d=1}^{D} a_d(X_d),
    \]
    where the functions $a_d(\cdot)$ express losses due to risks $X_d$.
- Then we have a scalar time series $S_1, \ldots, S_n$ to which previous ideas apply, using block maxima or threshold exceedances.
- **Advantages**: simple analysis, ignores dependence between $X_1, \ldots, X_D$.
- **Disadvantages**:
  - if a new structure variable is introduced, a new analysis is needed—which may disagree with original;
  - missing values of $X_d$ not allowed;
  - don’t learn which combinations of $X_1, \ldots, X_D$ yield extreme events.

Extremes for two variables
- For simplicity, consider bivariate case $(X, Y)$ with the same marginal distributions.
- Given a high threshold $u$, we might consider any of the following scenarios as extreme:
  - at least one of $X$ and $Y$ exceeds $u$, i.e., $\max(X, Y) > u$;
  - both $X$ and $Y$ exceed $u$, i.e., $\min(X, Y) > u$;
  - a function $s(X, Y)$ exceeds $u$, e.g., $X + Y > u$, though $s(\cdot)$ could also measure distance from some multivariate centre for the data; or
  - given that $X > u$, we consider the distribution of $Y$, where $Y$ is called a concomitant of $X$; the extremal set is $X > u$.
- There are other possibilities, but these already make life complicated enough.
- The grey regions on the next slide are considered under these four scenarios.
Extremes in two dimensions

Models for multivariate extremes

- In extension of the univariate case, we ask:
  
  If non-degenerate limiting distributions exist for maxima of rescaled \((X_1, \ldots, X_D)\), what forms can they have?

- The rescaled margins must have limiting GEV distributions, so we can consider the component-wise transformations
  
  \[
  \mathcal{P}_n = \left\{ \frac{1 + \xi X_j - b_n}{a_n}^{1/\xi} : j = 1, \ldots, D \right\} \subset \mathcal{E} = \mathbb{R}_+^D = \{0\},
  \]

  where \(X_1, \ldots, X_n, a_n > 0, b_n\) and \(\xi\) are all \(D \times 1\) vectors, and \(\mathcal{P}_n\) converges to a Poisson process \(\mathcal{P}\) on \(\mathcal{E}\) with mean measure

  \[
  \lim_{n \to \infty} n\mathbb{P}\left\{ \frac{1 + \xi X_j - b_n}{a_n}^{1/\xi} \in \cdot \right\} = \mu(\cdot).
  \]

- We replace \((1 + \xi)^{1/\xi}\) by \(\exp(\cdot)\) when \(\xi = 0\).
Poisson process limit

Let \( z = (z_1, \ldots, z_D) \in \mathcal{E} \) and let
\[
\mathcal{A}_z = \mathcal{E} - [0, z_1] \times \cdots \times [0, z_D].
\]
The maximum of \( \mathcal{P}_n \) lies below \( z \) iff \( \mathcal{P}_n \cap \mathcal{A}_z = \emptyset \), and this has limiting probability
\[
P(Z \leq z) = P(\mathcal{P} \cap \mathcal{A}_z = \emptyset) = \exp\left\{-\mu(\mathcal{A}_z)\right\}, \quad z \in \mathcal{E},
\]
where \( Z \) denotes the componentwise maximum of the points in \( \mathcal{P} \).

Equivalently we define the exponent function
\[
V(z_1, \ldots, z_D) = \mu(\mathcal{A}_z),
\]
and can show that
- the marginal unit Fréchet distributions of the \( Z_d \) yield \( V(z, \infty, \ldots, \infty) = 1/z \) for any permutation of the arguments;
- the function \( V \) is homogeneous of order \(-1\), i.e.,
\[
V(tz_1, \ldots, tz_D) = t^{-1}V(z_1, \ldots, z_D), \quad z_1, \ldots, z_D > 0, t > 0,
\]
which implies max-stability of the distribution of \( Z \).

Limit distribution of componentwise maxima

**Theorem 1** If \( X_1, X_2, \ldots \) are independent copies of a \( D \)-dimensional random variable whose componentwise maxima can be linearly renormalised to converge as \( n \to \infty \) to a random variable \( Z = (Z_1, \ldots, Z_D) \) that has a non-degenerate distribution with unit Fréchet margins, then
\[
P(Z_1 \leq z_1, \ldots, Z_D \leq z_D) = \exp\left[\left.-D\right\{\max\limits_{d=1}^{D}(W_d/z_d)\right\}\right], \quad z_1, \ldots, z_D > 0, \quad (1)
\]
where the angular variable \( W = (W_1, \ldots, W_D) \) lies in the \((D-1)\)-dimensional simplex, i.e.,
\[
W \in \mathcal{S}_{D-1} = \{(w_1, \ldots, w_D) : w_d \geq 0, \sum_{d} w_d = 1\}
\]
and satisfies the marginal mean constraints
\[
E(W_d) = 1/D, \quad d = 1, \ldots, D.
\]
The angular distribution \( \nu \) of \( W \) is otherwise arbitrary, and the exponent function is
\[
V(z_1, \ldots, z_D) = D\max\limits_{d=1}^{D}(W_d/z_d).
\]
Extremal functions

If we write \( P = \{ Q_j : j = 1, 2, \ldots \} \) using extremal functions \( Q_j \), then
\[
Q_j = R_j W_j, \quad R_j > 0, W_j \in S_{D-1},
\]
where
- the pseudo-radii \( R_j \) are points of a Poisson process on \((0, \infty)\) with intensity \( D/r^2 \) independent of
- the pseudo-angles \( W_j \) iid \( \sim \nu \),
and the extremal functions form a Poisson process on \( E \) with intensity
\[
\mu(dq) \equiv \mu(dr, dw) = \frac{dr}{r^2} \times \nu(dw).
\]

The \( Q_j \) represent individual extreme events (storms, heatwaves, \ldots). We can simulate the \( Q_j \) by starting with the largest \( R_j \) and working downwards. The same decomposition applies in more generality, with the \( W \) functions lying in a suitable function space.

Hülsler–Reiss simulations

Simulated Poisson processes for the Hülsler–Reiss model with \( \lambda = 0.5 \). In each case 10000 points with the largest pseudo-radii have been simulated; the limits appear curved because of the log axes. The intersections of the dotted lines show the componentwise maxima: on the right both arise from a single event, whereas on the left they arise from two separate events.
Hüsler–Reiss distribution

- The natural analogue of the normal distribution in extremal contexts.
- The bivariate version has a scalar parameter $\lambda > 0$ and
  \[ V(z_1, z_2) = \frac{1}{z_1} \Phi \left( \frac{\lambda}{2} + \frac{1}{\lambda} \log \left( \frac{z_2}{z_1} \right) \right) + \frac{1}{z_2} \Phi \left( \frac{\lambda}{2} + \frac{1}{\lambda} \log \left( \frac{z_1}{z_2} \right) \right), \quad z_1, z_2 > 0, \]
  where $\Phi$ denotes the standard normal cumulative distribution function.
- Its limits are total independence and total dependence,
  \[ V(z_1, z_2) \to \begin{cases} 
 1/z_1 + 1/z_2, & \lambda \to \infty, \\
 1/\min(z_1, z_2), & \lambda \to 0.
  \end{cases} \]
  For this model,
  \[ W^{-1} = 1 + \exp(\lambda e + \lambda^2 I/2), \]
  where $I = \pm 1$ with equal probabilities independently of $e \sim \mathcal{N}(0, 1)$.

Parametric models

- It is tricky to formulate parametric models that satisfy the mean constraints in higher dimensions, but numerous models exist for bivariate data.
- The logistic model for general $D$ has a single parameter $\alpha \in (0, 1]$ and
  \[ V(z_1, \ldots, z_D) = \left( \sum_{j=1}^{D} z_j^{-1/\alpha} \right)^\alpha, \quad z_1, \ldots, z_D > 0. \]
  Independence and total dependence arise as limits as $\alpha \uparrow 1$ and $\alpha \downarrow 0$ respectively.
- A limitation of the Hüsler–Reiss and logistic models is their symmetry, but they can be extended in various useful ways.

Asymmetric parametric models

Asymmetric alternatives to the logistic and Hüsler–Reiss models include:
- the bilogistic model
  \[ \tilde{\nu}(w) = \frac{1}{2}(1 - \alpha)(1 - w)^{-1} w^{-2} (1 - u) u^{1 - \alpha} (\alpha(1 - u) + \beta u)^{-1}, \quad 0 < w < 1, \]
  where $0 < \alpha, \beta < 1$, and $u = u(w, \alpha, \beta)$ satisfies
  \[ (1 - \alpha)(1 - w)(1 - u)^{\beta} - (1 - \beta) w u^{\alpha} = 0; \]
- and the Dirichlet model
  \[ \tilde{\nu}(w) = \frac{\alpha\beta \Gamma(\alpha + \beta + 1)(\alpha w)^{\alpha-1} (\beta(1 - w))^{\beta-1}}{2\Gamma(\alpha)\Gamma(\beta)(\alpha w + \beta(1 - w))^{\alpha+\beta+1}}, \quad 0 < w < 1, \]
  for parameters $\alpha, \beta > 0$.
- The function \texttt{fbvevd} (see also \texttt{dbvevd}) in the R library \texttt{evd} enables MLE for numerous parametric bivariate models for maxima.
Logistic and Dirichlet densities

Left: logistic densities $\hat{\nu}(w)$ with $\alpha = 0.1$ (black), 0.3 (red), 0.5 (blue), 0.9 (green).

Right: Dirichlet densities $\hat{\nu}(w)$ with parameters $(\alpha, \beta) = (0.5, 0.5)$ (black), $(0.5, 1)$ (red), $(0.5, 2)$ (blue) and $(2, 3)$ (green).

Events from subsets of $W$

- If $V$ is differentiable, there can be densities on the $D$-dimensional simplex $S_{D-1}$ and on each of its sub-faces, defined by setting subsets of the $w_d$ to zero.
- Hence $\nu$ can have $2^D - 1$ components in general — a complicated object!
- These correspond to particular combinations of extremes and can be viewed as components of a mixture distribution.
- If $D = 3$, for example, there are three singleton components, three pair components, and one triple component, so perhaps the limiting rare events are
  
  $W_1$ only, $W_2, W_3$ together, or $W_1, W_2, W_3$ together,

  corresponding to

  $Q_1 > 0, Q_2 = Q_3 = 0$, $Q_1 = 0, Q_2, Q_3 > 0$, or $Q_1, Q_2, Q_3 > 0$,

  with other combinations impossible.
- In applications we never see $Q_d = 0$, so we have to declare $Q_d \equiv 0$ when $Q_d < \epsilon$ for some small positive $\epsilon$. 

Asymptotic dependence and independence

- All max-stable models are asymptotically dependent (AD), i.e.,
  \[ \chi(u) = P \{ X_2 > F_2^{-1}(u) \mid X_1 > F_1^{-1}(u) \} \to \chi > 0, \quad u \to 1, \]
  or exactly independent if \( \chi = 0 \).

- In many applications \( \chi(u) \to 0 \) as \( u \to 1 \), i.e., the variables are asymptotically independent (AI), and we then use
  \[ \overline{\chi}(u) = \frac{2 \log P \{ F_2(X_2) > u \} - \log P \{ F_2(X_2) > u, F_1(X_1) > u \}}{\log P \{ F_2(X_2) > u \}} - 1 \to \chi, \quad u \to 1, \]
  to measure the level of AI. The scaling is chosen so that if
  - \( X \) and \( Y \) are independent, \( \chi = 0 \);
  - if \( X \) and \( Y \) are perfectly dependent, \( \chi \equiv 1 \);
  - if \( X \) and \( Y \) are AD, \( \chi = 1 \);
  - \(-1 < \chi(u) \leq 1\), and \( \chi \) increases with increasing dependence.

- Both \( \chi(u) \) and \( \overline{\chi}(u) \) can be estimated non-parametrically from data.

- Can construct AI models from AD ones by inversion.

- Models that encompass both AD and AI exist (Heffernan & Tawn, 2004).

Extremal coefficient

- A simple summary of dependence is the extremal coefficient,
  \[ \theta = V(1, \ldots, 1), \]
  which satisfies \( \theta = 1 \) for perfectly dependent data, and \( \theta = D \) for independent data, and is (loosely) interpreted as the ‘number of independent maxima’ contributing to \( Z \), because
  \[
  P(\max(Z_1, \ldots, Z_D) \leq z) = P(\bar{Z_1} \leq z, \ldots, \bar{Z_D} \leq z) = \exp\{-V(z, \ldots, z)\} = \exp\{-V(1, \ldots, 1)/z\} = \{\exp(-1/z)\}^{V(1, \ldots, 1)}, \quad z > 0,
  \]
  the distribution of the maximum of \( \theta = V(1, \ldots, 1) \) independent Fréchet variables.

- When \( D = 2 \),
  \[ \chi = \lim_{z \to \infty} P(Z_1 > z \mid Z_2 > z) = 2 - \theta, \quad 1 \leq \theta \leq 2, \]
  and \( 2 - \theta \) is sometimes called the extremal correlation.
Extremal coefficients for snow depth

Extremal coefficient computed relative to Koppigen, Adelboden, Davos and Maloja (white points), kriged to the whole of Switzerland using a linear trend on absolute altitude difference (Blanchet & Davison, 2011).

Concurrence probability

- The **concurrence probability** is the probability \( p_D \) that all \( D \) components of a multivariate maximum \( Z \) arise from a single extreme event, or equivalently that the limiting Poisson process contains a point \( Q \) such that \( Q = Z \) with probability one.

- This event occurs if \( Q = z \) and \( A_z \) is void, or equivalently if the maximum of the other points of the Poisson process is below \( z \), so
  
  \[
  p_D = \int_E \mu(z) \exp\{-\mu(A_z)\} \, dz = E_W \left[ E_W^* \left\{ \max_d \left( W_d / W_d^* \right) \right\}^{-1} \right],
  \]

  where \( W, W^* \) iid \( \nu \).

- Total dependence yields \( p_D = 1 \), whereas independence yields \( p_D = 0 \).

- \( p_D \) has a nice interpretation, but its computation typically involves numerical integration.

- In spatial applications the probability that the same \( Q \) leads to extremes at locations \( s \) and \( s' \) is \( p_D \equiv p_2(s, s') \), and
  
  \[
  I(s) = \int p_2(s, s') \, ds'
  \]

  measures the mean area of the events leading to extremes at \( s \).
Concurrence for high/low temperature extremes

Figure 8. Maps of the extremal concurrence probability for the Worland station (triangle) for the Fall (top) and Spring (bottom) seasons. The left panels show results for the time period 1911–1950 while the middle panels to 1951–2010. The rightmost panels plot the pointwise p-values relative to the hypothesis test of concurrence probabilities difference between the two time periods—see Section 4.5.

From Dombry, Ribatet and Stoev (2018) slide 71

Concurrence for high/low temperature extremes

Figure 9. Top: Spatial distribution of the expected extremal concurrence cell areas anomalies, that is, the pointwise difference between $E(|C(s)|)$ for period 1951–2010 and period 1911–1950. Bottom: Spatial distribution of the extremal concurrence cell area variations, that is, the pointwise ratio between $\text{var}(|C(s)|)$ for period 1951–2010 and 1911–1950. Each column corresponds to one season. From left to right: Fall, Winter, Spring, and Summer.

From Dombry, Ribatet and Stoev (2018) slide 72
Summary

- After transformation of the margins to a standard form, the joint distributions have a nonparametric structure subject to mean constraints.
- As in the scalar case, max-stability constrains the possible limiting distributions for multivariate maxima.
- There are many parametric models for bivariate data, but fewer for $D > 2$.
- There are close links between the maximum and point process representations.
- Dependence measures exist:
  - $\chi$ and $\chi'$ measure asymptotic dependence (AD) and asymptotic independence (AI);
  - the extremal coefficient is a scalar summary of dependence, with $\theta = 1$ for fully dependent data and $\theta = D$ for independent data;
  - the concurrence probability is another scalar summary, with a simple interpretation;

Modelling Multivariate Extremes

Inference for multivariate maxima

- Inference will consist of the following steps:
  - fitting of marginal GEV distributions and transformation to standard Fréchet;
  - choice/estimation of dependence model $\nu$;
  - model checking;
  - computation of probabilities for events of interest.
- Ideally all the estimation is performed at once, by fitting marginal and dependence models together (infeasible in complex cases).
- When $D = 2$, the joint density for the maxima $(Y_1, Y_2)$ can be written as

$$f(y_1, y_2) = \frac{\partial z_1 \partial z_2}{\partial y_1 \partial y_2} \times \left\{ \frac{\partial V(z_1, z_2)}{\partial z_1} \frac{\partial V(z_1, z_2)}{\partial z_2} - \frac{\partial^2 V(z_1, z_2)}{\partial z_1 \partial z_2} \right\} \times \exp\{-V(z_1, z_2)\},$$

where the first two (Jacobian) terms depend on the marginal parameters and the remainder depend both on those parameters and those of $V$ (or equivalently $\nu$).
- For larger $D$ the number of terms with derivatives of $V$ increases very rapidly, but the structure of the density is the same.
Stephenson–Tawn estimation

- The joint density $f(y_1, y_2)$ of the maxima has two terms,
  \[
  \frac{\partial V(z_1, z_2)}{\partial z_1} \frac{\partial V(z_1, z_2)}{\partial z_2} - \frac{\partial^2 V(z_1, z_2)}{\partial z_1 \partial z_2},
  \]
  corresponding to the joint maximum occurring either
  - in the same event, $-\partial^2 V(z_1, z_2)/\partial z_1 \partial z_2$, or
  - in two different events, $\partial V(z_1, z_2)/\partial z_1 \times \partial V(z_1, z_2)/\partial z_2$,
  and corresponding to the two partitions of the set \{1, 2\}, either as 12 or as 1 | 2.
- So if we have information on the timing of events (same storm, different storms?) then we need only include the relevant term in the likelihood.
- Likewise, if $D = 3$, the corresponding partitions are
  \[123, \ 1 \mid 23, \ 2 \mid 13, \ 3 \mid 12, \ 1 \mid 2 \mid 3,\]
  each of which gives a different likelihood contribution, which must all be included if the timing of the maxima is unknown.

Poisson process approximation

- To apply the Poisson process approximation we must apply a marginal transformation of the original data into $\mathcal{E}$, and choose an ‘extreme’ region $A$.
- For marginal transformation, we take thresholds $u_1, \ldots, u_D$, usually corresponding to the same quantile (e.g., 0.95) of each margin. Suppose that $n_u$ observations exceed these thresholds, and let $\hat{p}_u = n_u/n$ be the estimated exceedance probability.
- We fit GPDs above these thresholds, giving fitted distribution
  \[
  \hat{F}_d(x) = \begin{cases} 
  \frac{\#\{j : x_{j,d} \leq x\}}{n}, & x \leq u_d, \\
  1 - \hat{p}_u \left\{1 + \frac{\xi_d(x - u_d)}{\hat{\sigma}_d}\right\}^{-1/\xi_d}, & x > u_d,
  \end{cases} \quad d = 1, \ldots, D,
  \]
  based on the fitted GP parameters $(\hat{\sigma}_d, \hat{\xi}_d)$ for the $D$ margins.
- The (component-wise) transformed variables
  \[q_j = -1 / \log \hat{F}(x_j), \quad j = 1, \ldots, n,\]
  lie in $\mathcal{E}$ and have approximate unit Fréchet marginal distributions.
- The corresponding pseudo-radii and pseudo-angles are
  \[r_j = \|q_j\|_1 > 0, \quad w_j = q_j/r_j \in S_{D-1}, \quad j = 1, \ldots, n.\]
Nonparametric estimation

Homogeneity of $V$ gives

$$
\mu(A) = t\mu(tA), \quad A \subset \mathcal{E}, t > 0, \quad \text{where} \quad tA = \{tx : x \in A\}.
$$

Thus we can compute the probability $P(X \notin A) = \exp\{-\mu(A)\}$ using

$$
\mu(A) = t\mu(tA) = t \times D \int_{tA} \frac{dV}{r^2} \nu(dw), \quad t > 0,
$$

where the second term on the right is the measure of the set $tA$.

Hope that the theory applies near the origin, take $t$ small, and let

$$
\hat{\mu}(A) = t \times \#\{j : q_j \in tA\}.
$$

Likelihood for events

Base extremal modelling on those individual events $q$ falling into extreme set $A$:

- allows more detailed modelling and may include more data,
- if $\mu(A)$ is readily computed, likelihood is

$$
\exp\{-\mu(A)\} \times \prod_{q \in A} \hat{\mu}(q), \quad \hat{\mu}(q) = -\frac{\partial^D V(z_1, \ldots, z_D)}{\partial z_1 \cdots \partial z_D},
$$

- but components of some $q$ may be non-extreme, so use a **censored likelihood**.
Brown–Resnick likelihood

If \( z_d > u \) for \( d = 1, \ldots, C \) and \( z_d < u \) for \( d \in C' = \{ C + 1, \ldots, D \} \), and \( C = \{ 2, \ldots, C \} \), the censored likelihood contribution has form

\[
\frac{1}{z_1 \cdots z_C} \times \phi_{C-1}(\log \tilde{z}_C; \tilde{\Omega}_{C,C}) \times \Phi_{D-C}(\tilde{\mu}_{C'|C}; \tilde{\Omega}_{C'|C})
\]

where \( \phi_k \) and \( \Phi_k \) denote the \( k \)-dimensional normal density and distribution functions, \( \Omega \) is defined in terms of the variogram \( \gamma \), and

\[
\begin{align*}
\log \tilde{z}_d &= \log z_d - \log z_1 + \Omega_{d,1}/2, \quad d = 2, \ldots, C, \\
\tilde{\Omega}_{c,d} &= \frac{1}{2} \{ \Omega_{c,1} + \Omega_{1,d} - \Omega_{c,d} \}, \quad c, d \in \{ 2, \ldots, D \}, \\
\mu_{C'|C} &= (\log u - \log z_1 + \frac{1}{2} \Omega_{1,C'}) - \tilde{\Omega}_{C'|C} \tilde{\Omega}_{C,C}^{-1} \log z_C, \\
\tilde{\Omega}_{C'|C} &= \tilde{\Omega}_{C'|C} - \tilde{\Omega}_{C'|C} \tilde{\Omega}_{C,C}^{-1} \tilde{\Omega}_{C,C}.
\end{align*}
\]

Gradient score needed for higher \( D \):
- Differentiate with respect to data, so normalising constants not needed;
- Use weight function to downweight effects of observations near thresholds.

Similar computations are possible for extremal-\( t \) processes.

Saudi Arabian rainfall

Jeddah liable to intense (but rare!) strong convective rainstorms, leading to flash floods, extensive damage and deaths.

15-minute radar data available at 750 grid cells over 17 years, so daily annual maxima are space-rich but time-poor.

(Davison et al. 2019)
Saudi Arabian rainfall

- Censor annual maxima < 3mm.
- Use local likelihood estimates of location and scale parameters, with $\xi \approx 0.14$ constant.
- Transform maxima to standard Fréchet scale, and fit spatial models using censored pairwise local likelihood

$$\ell(d) = \sum_{i=1}^{17} \sum_{d' < d} w_{d',d} I(z_{i,d'} > u'_d, z_{i,d} > u_d) \log \left\{ \frac{\exp(-V)(V_1V_2 - V_{12})}{p(u'_d, u_d)} \right\}.$$ 

- Information criteria suggest reasonable fit of isotopic Brown–Resnick model with variogram $(h/\lambda)^{\alpha}$, with range $\lambda \approx 13$km range and shape $\alpha \approx 0.7$. 

---

**Saudi Arabian rainfall: Estimates**

---

slide 82

slide 83
Saudi Arabian rainfall: Model fit

Empirical and fitted extremal coefficients \( \theta_D \in [1, D] \) for locations around Jeddah.

<table>
<thead>
<tr>
<th>Region ( D = {s_1, \ldots, s_D} )</th>
<th>( D )</th>
<th>Empirical ( \hat{\theta}_D )</th>
<th>Smith Schl. B.-R. Ext.-t</th>
</tr>
</thead>
<tbody>
<tr>
<td>([39, 40]^\circ\ E \times [21, 22]^\circ\ N)</td>
<td>14</td>
<td>4.17 ([1.90, 6.44])</td>
<td>3.37 3.16 4.41 3.44</td>
</tr>
<tr>
<td>([39, 41]^\circ\ E \times [21, 23]^\circ\ N)</td>
<td>62</td>
<td>14.25 ([6.50, 22.00])</td>
<td>9.73 6.06 11.27 8.71</td>
</tr>
<tr>
<td>([39, 42]^\circ\ E \times [21, 24]^\circ\ N)</td>
<td>142</td>
<td>20.90 ([9.54, 32.26])</td>
<td>19.00 9.01 20.10 15.96</td>
</tr>
</tbody>
</table>


Saudi Arabian rainfall: Risk estimation

- Use simulation of individual events to compute probabilities that annual maximum averaged over 14 grid cells \( S \) around Jeddah/Makkah exceeds \( v \) mm/day, i.e.,

\[
p(v) = P\left\{ |S|^{-1} \sum_{s \in S} Z(s) > v \right\},
\]

obtaining

\[
p(50) = 0.072, \quad p(71.1) = 0.019, \quad p(100) = 0.0048,
\]

with respective return periods around 14, 54 and 208 years.

- Daily rainfall total on 25 November 2009 was 71.1 mm/day, leading to 122 deaths.

Summary

- Structure variable approach sometimes useful, focused on a scalar of interest.
- Multivariate models can be fitted using either joint maxima or point process approaches.
- Possible in fairly high dimensions, using pairwise likelihoods or gradient score methods.
- Assume that the asymptotic representations are adequate for finite samples/thresholds, so must be checked carefully (e.g., by varying the threshold).
- Multivariate AI models also exist.
- Simultaneity can be incorporated into the maximum model, if times of events are available.
- Nonparametric estimators also exist.
Discussion

- Extreme-value statistics
  - is a well-developed (and growing!) domain of statistics dealing with extrapolation for rare(r) events
  - relies on point process theory and regular variation (suppressed here)
  - relies on asymptotically-justified models that are never exactly correct
  - can be fitted (with some effort) to high-dimensional problems/processes
  - has many applications to complex problems in climate science

References (mostly review papers)

- de Fondeville & Davison (2021+) *arXiv*.
- Koch, Koh, Davison, Lepore & Tippett (2021) *Journal of Climate*.