Extremes

- An extreme value is an unusually large – or small – magnitude.
- Extreme value analysis (EVA) has as objective to quantify the stochastic behavior of a process at such unusual levels.
- EVA usually requires estimation of the probability of events that are more extreme than any that have already been observed.
- Typical question: What is the maximum value expected over a given time period (1 day – 100 year) and given a specified degree of confidence (typically 95% or 99%)?
Largest 10 annual sea-levels in Venice

source of the data: ismev (R-package)
Largest 10 annual sea-levels in Venice

(510 data values)
Before Extreme Value Theory two main approaches were used:

- In economics and finance, value-at-risk (VaR), Gauss distribution; see for instance http://en.wikipedia.org/wiki/Value_at_risk

Fig. 1. Plots of the GEV (generalized extreme value) probability density function with $\mu = 0$, $\sigma = 1$, $\xi = -0.2$ (Weibull type), $\xi = 0$ (Gumbel), and $\xi = 0.2$ (Fréchet).

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

(3)

Here \(\mu\) is termed a location, \(\sigma > 0\) a scale, and \(\xi\) a shape parameter. If the random variable \(X\) has a GEV distribution (Eq. 3), then the standardized variable \((X - \mu)/\sigma\) has a distribution that does not depend on either \(\mu\) or \(\sigma\), only on \(\xi\).

The shape of the GEV distribution assumes three possible types

(i) \(\xi = 0\), a light-tailed (or Gumbel) distribution;

(ii) \(\xi > 0\), a heavy-tailed (or Fréchet) distribution;

(iii) \(\xi < 0\), a bounded (or Weibull) distribution.
i.i.d and $N \to \infty$

- $X_i$ is a sequence ($i = 1, 2, \cdots, N$) or other collection of random variables independent and identically distributed (i.i.d.). The sequence is \textbf{i.i.d.} if each $X_i$ has the same probability distribution as the others and all are mutually independent. \textbf{If i.i.d.} then it is stationary and independent. \textbf{If i.i.d.} then it is \textbf{ergodic}.

- \textbf{in short:} $X_i$ is a time series produced by some (random) device.

- \textbf{Reality:} Not stationary, not independent, not ergodic, one sample, finite length.
Figure 1: One-thousand random samples each of size 1,000 simulated from a normal distribution with mean zero and unit standard deviation. The histograms shown here are for the means (left) and maxima (right) for each of these samples. The solid lines show the best fit normal pdf and the dashed lines show the best fit GEV (using maximum-likelihood estimation).

General limitations with the GEV paradigm and data.

The GEV model and other he will discuss later are developed using asymptotic arguments (N ->\( \infty \)) and care is needed in treating them as exact results for finite samples. It was shown that usually the convergence to the GEV distribution is extremely slow.

The unknown parameters of a model, \( \mu, \theta, \alpha \), are inferred on the basis of historical data. We can argument if this enable extrapolation to unobserved levels.

Other data, equally representative of the true process being studied, would have led to different estimates. A model must be complemented with measures of the uncertainty due to sampling variability, usually by calculating a confidence interval.

No more credible alternative has been proposed to date.
Figure 2: Daily rainfall values recorded in Venezuela.
A note on confidence intervals

If samples of the same size are drawn repeatedly from a population, and a confidence interval is calculated from each sample, then 95% of these intervals should contain the population mean.

Picturing 50 realizations of a 95%-confidence interval.

Usually replacing population parameter by approximations calculated using the observed sample. Data (measurement) errors and model errors (parametric or nonparametric, gaussian etc) are not included. Neither estimation error associated to the estimation method: Maximum Likelihood, Bayesian, Moments etc. Uncertainty is largely underestimated.

Independence and classical EVA approach

Block maxima

- In the classical extreme value analysis, data are blocked into sequences of observations of length R, generating a series of block maxima, $m_1, \ldots, m_j$. Often the blocks correspond to a time period of length one year to which the GEV distribution can be fitted using some technique.

The log-likelihood for the GEV parameters is given by

$$-m \log \sigma - (1 + \frac{1}{\xi}) \sum_{i=1}^{m} \log \left[ 1 + \xi \left( \frac{y_i - \mu}{\sigma} \right) \right] - \sum_{i=1}^{m} \left[ 1 + \xi \left( \frac{y_i - \mu}{\sigma} \right) \right]^{-1/\xi},$$

provided $\xi(y_i - \mu) > -\sigma$ for $i = 1, \ldots, m$. Since the range depends on the unknown parameters, the MLEs may not have the usual nice properties.

FIG. 1. Definition of maxima: A time series $(x_i), i=1, \ldots, N$, of, e.g., daily data is separated into segments of length $R=365$ days. The maximum values $m_j (\bigcirc)$ in each segment, e.g., annual maxima, define another time series $(m_j), j=1, \ldots, N/R$. 

Threshold Models

Generalized Pareto Distribution (GPD)

Exceedances over thresholds \((X-U)\). In this context there is an analog of the generalized extreme value distribution: the generalized Pareto distribution. The basic idea is to pick a high threshold \(U\) and study all the exceedances of \(U\). The amount by which the threshold is exceeded follows a generalized Pareto Distribution (GPD):

\[ G(x) = \exp \left\{ - \left[ 1 + \xi \left( \frac{x - \mu}{\sigma} \right) \right]^{-1/\xi} \right\} \]

for some \(\mu, \sigma > 0\) and \(\xi\). Then, for large enough \(u\), the distribution function of \((X-u)\), conditional on \(X > u\), is approximately

\[ H(y) = 1 - \left( 1 + \frac{\xi y}{\hat{\sigma}} \right)^{-1/\xi} \]

defined on \(\{y : y > 0 \text{ and } (1 + \xi y/\hat{\sigma}) > 0\}\), where

\[ \hat{\sigma} = \sigma + \xi(u - \mu). \]

Text from: An Introduction to Statistical Modeling of Extreme Values by Stuart Coles (Springer, 2001)
Picture from: Estimation and attribution of changes in extreme weather and climate events: opening lecture to the IPCC Workshop in Beijing, by David Stephenson.
Threshold selection

• Several procedures for threshold selection are given in the literature. One method is based on the expected value for the excess of a threshold. It can be shown that:

\[ E(X-u \mid X>u) \] is a linear function of \( u \).

• “Is tempting to conclude that there is no stability until \( u=60 \), after which there is approximate linearity”

• From: Coles 2001 “An Introduction to Statistical Modeling of Extreme values”

![FIGURE 4.1. Mean residual life plot for daily rainfall data.](image)
Poisson-GDP-Peaks over Threshold model

(a) The number, $N$, of exceedances of the level $u$ in any one year has a Poisson distribution with mean $\lambda$; 
(b) Conditionally on $N \geq 1$, the excess values $Y_1, \ldots, Y_N$ are i.i.d. from the GPD.

We call this the Poisson-GPD model.

Of course, there is nothing special here about one year as the unit of time — we could just as well use any other time unit — but for environmental processes in particular, a year is often the most convenient reference time period.

$$\lambda = \left(1 + (u - \mu)/\psi \right)^{-\xi}$$

The Poisson-GPD model is closely related to the Peaks Over Threshold (POT) model originally developed by hydrologists. In cases with high serial correlation the threshold exceedances do not occur singly but in clusters, and in that case, the method is most directly applied to the peak values within each cluster.

From: Coles 2001: An Introduction to Statistical Modeling of Extreme Values.

[...] recalling that the Poisson distribution arises as an approximation to the binomial for rare events, it is reasonable to assume that the sequence of times the event occurs is governed by a Poisson process, say with rate parameter $\lambda > 0$. In this case, the number of occurrences in a time interval of length $T$, $N_T$ say, has a Poisson distribution with mean $\lambda T$; that is,

$$\Pr\{N_T = k\} = \left(\frac{\lambda T e^{-\lambda T}}{k!}\right), \quad k = 0, 1, \ldots$$

From: Coles 2001: An Introduction to Statistical Modeling of Extreme Values.

[...] there are physical phenomena for which the Poisson process is a poor model: processes where there is a natural spacing, such as the location of trees in a forest; or processes that have natural clustering, such as the occurrence times of rainstorms.
Figure 4. The estimated (a) 10 years, (b) 100 years and (c) 1000 years floods for the ML, PML and PWM estimators using block size (BS) 12 months, 6 months and 3 months, and thresholds (u) 25, 30, 40, and 50 m³/s.
Other methods

• Why are there so many models for extremes?

• As the real nature of natural data is unknown, and because there is always some differences between model’s results and data, we can always suspect the basic hypothesis of a specific model and try to find another that supposedly fits better to the data. It is with no surprise that a variety of models exists, each model having its own set of assumptions.
Non-stationary process

\[ Z_t \sim \text{GEV}(\mu(t), \sigma(t)) \]

\[ \mu(t) = \beta_0 + \beta_1 t \]

\[ \mu(t) = \beta_0 + \beta_1 t + \beta_2 t^2 \]

\[ Z_t \sim \text{GEV}(\mu(t), \sigma(t), \xi(t)) \]

\[ \ell(\mu, \sigma, \xi) = -\sum_{t=1}^{\infty} \left\{ \log \sigma(t) + (1 + 1/\xi(t)) \log \left[ 1 + \xi(t) \left( \frac{z_t - \mu(t)}{\sigma(t)} \right) \right] \right. \\
+ \left. \left[ 1 + \xi(t) \left( \frac{z_t - \mu(t)}{\sigma(t)} \right) \right]^{-1/\xi(t)} \right\} \]

**TABLE 6.1.** Maximized log-likelihoods and parameter estimates, with standard errors in parentheses, of various models for \( \mu \) in GEV model for minima applied to race time data of Example 1.4

<table>
<thead>
<tr>
<th>Model</th>
<th>Log-likelihood</th>
<th>( \beta )</th>
<th>( \hat{\sigma} )</th>
<th>( \hat{\xi} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-54.5</td>
<td>239.3</td>
<td>3.63</td>
<td>-0.469</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.9)</td>
<td>(0.64)</td>
<td>(0.141)</td>
</tr>
<tr>
<td>Linear</td>
<td>-51.8</td>
<td>(242.9, -0.311)</td>
<td>2.72</td>
<td>-0.201</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(1.4, 0.101)</td>
<td>(0.49)</td>
<td>(0.172)</td>
</tr>
<tr>
<td>Quadratic</td>
<td>-48.4</td>
<td>(247.0, -1.395, 0.049)</td>
<td>2.28</td>
<td>-0.182</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(2.3, 0.420, 0.018)</td>
<td>(0.45)</td>
<td>(0.232)</td>
</tr>
</tbody>
</table>

**FIGURE 6.3.** Fitted estimates for \( \mu \) in GEV model for minima applied to fastest annual race times for women's 1500 meters event.

**FIGURE 6.4.** Fitted estimates for exponential \( \mu \) in GEV model for minima applied to fastest annual race times for women's 1500 meters event.
Figure 4. (a) Plots of (a) the time series of Boeotios Kephisos annual runoff, (b) a hundred-year part of the Nilometer time series; (c) the complete Nilometer time series.
Changes in location, scale, and shape

Fig. 5.1 from PhD thesis of Tahl Kestin:

a) Shift in mean
b) Increase in variance
c) Change in shape
Our strategy

It is to show overconfidence that to presume the present trend will continue and that it is impossible other natural or man-caused modifications will not induce trend variability. Even Newton’s Physics proved wrong.

- Time Evolution of Extreme values:
- Scenarios
- Numerical simulations
- Translate results (which?) from the numerical world to the real one.

Preliminary results

Stationary, block maxima = 1y, error bar = 95% confi.int.

Stationary, block maxima = 1y, error bar = 95% confi.int.

En effet, la variance du paramètre de forme est donné par
\[ \sigma^2 = \frac{(\xi+1)^2}{\xi^2} \]. Alors si \( \xi = 0.2 \), on obtient \( \sigma^2 = \frac{0.64}{0.1} \). Les bornes de l'intervalle de confiance sont donc données par l'équation suivante :
\[ \xi \pm \sqrt{\frac{0.64}{n}} \].

Même si la valeur moyenne est proche de la valeur exacte nous avons besoin de \( n = 560 \) données pour réduire la largeur de l'intervalle de confiance qui pourrait permettre une détection d'une variation de \( \pm 0.1 \).