An Introduction to Extremes

RSS / ESSG Short Course

Stuart Coles
University of Bristol

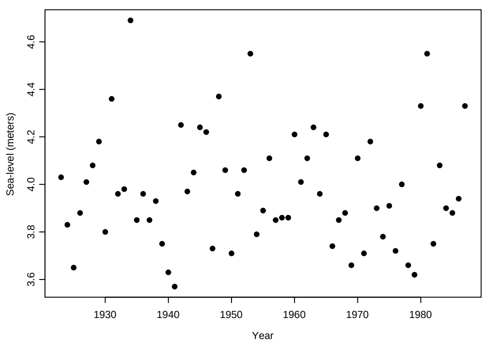
6, Nov, 2001

`stuart.coles@bristol.ac.uk`
`http://www.statistics.bristol.ac.uk/~masgc`

Applications of Extremes

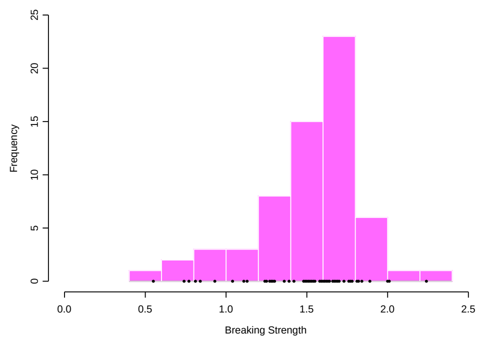
1. **Environmental**: Sea-levels, wind speeds, pollutant concentrations,. . .
2. **Reliability**: Breaking strength, corrosion level, pit depth, . . .
3. **Financial**: Insurance risk, portfolio risk, value-at-risk,. . .
4. **Miscellaneous**: Management strategy, sports data, . . .

Sea Levels



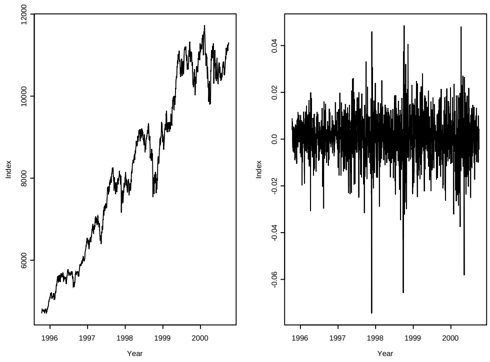
Objective: Design coastal defence to afford protection against extreme sea-levels.

Breaking Strengths



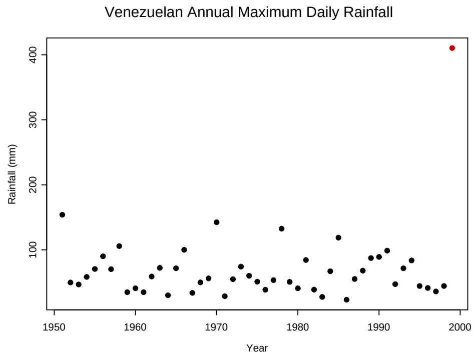
Objective: Predict minimum possible breaking strength.

Price Index Series



Objective: Estimate risk of large change in daily price.

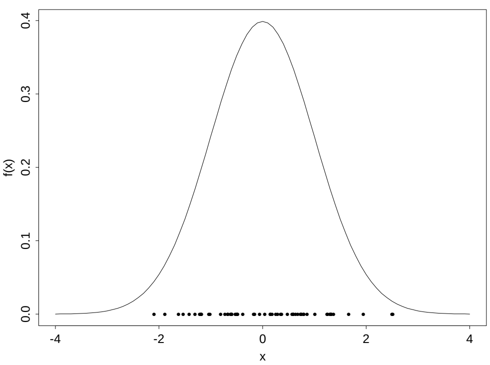
Venezuelan Annual Maximum Rainfall



Objective: Assess likelihood of future catastrophic rainfall levels.

Basics of extreme value theory

Simplest case: independent variables $X_1, \ldots X_n \sim F$. Require accurate inferences on tail of F.



The principal issues in considering this problem are as follows . . .

Issues in building models for extremes

- By definition, extremes are rare;
- Estimates are often required beyond X_{max}, the largest observed data value;
- Standard density estimation techniques fit well where the data have greatest density, but can be severely biased in estimating tail probabilities.

Absence of physical or empirical basis for extrapolations leads to the **extreme value paradigm:**

Asymptotic arguments should be used to generate suitable families of models for extremes.

Model Building

Let $X_1, X_2, \ldots, X_n$ is a sequence of iid random variables with distribution function F, and define

$$M_n = \max\{X_1, X_2, \ldots, X_n\}.$$

Then the distribution function of M_n is found as:

$$\begin{aligned}\Pr\{M_n \le x\} &= \Pr\{X_1 \le x, \ldots, X_n \le x\} \\ &= \Pr\{X_1 \le x\} \ldots \Pr\{X_n \le x\} \\ &= \{F(x)\}^n.\end{aligned}$$

But: F is unknown.

Hence, approximate the distribution by limit distributions as $n \to \infty$.

What distributions can arise?

The Generalized Extreme Value Distribution

Careful analysis implies that (subject to regularity) the approximate distribution of $F^n(x)$ for large n falls within the family

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

defined on $\{x : 1 + \xi(x-\mu)/\sigma > 0\}$.

The parameters μ and σ are location and scale parameters; ξ is a shape parameter determining the rate of decay in the tail.

The Gumbel Distribution

The special case of the GEV distribution in which $\xi = 0$, referred to as the **Gumbel** distribution, is of special interest. It is obtained by letting $\xi \to 0$ in the GEV family, leading to the distribution function

$$G(x) = \exp\left[- \exp\left\{ - \left(\frac{x-\mu}{\sigma} \right) \right\} \right],$$

defined on $-\infty < x < \infty$.

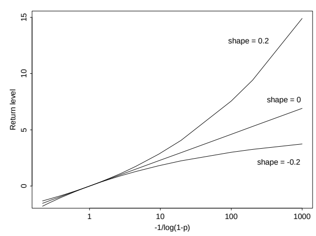
Quantiles

In terms of quantiles:

$$x_p = \mu - \frac{\sigma}{\xi} \left[1 - \{-\log(1-p)\}^{-\xi} \right],$$

where $G(x_p) = 1 - p$.

In extreme value terminology, x_p is the *return level* associated with the *return period* $1/p$



Plotting on this scale ensures:

1. Effects of extrapolation are clearly highlighted;
2. Plots are linear when $\xi = 0$.

Inference

Given observed 'annual maxima' $X_1, X_2, \ldots, X_k$, we need to make inferences on the GEV parameters (μ, σ, ξ) . Possibilities include:

- Graphical techniques;
- Moment-based estimators;
- Maximum Likelihood.

Reasons for preferring Maximum Likelihood include:

1. Simple approximations for standard error and confidence interval calculations;
2. Large sample optimality properties;
3. Provides framework for model building and extension.

Modelling Procedure

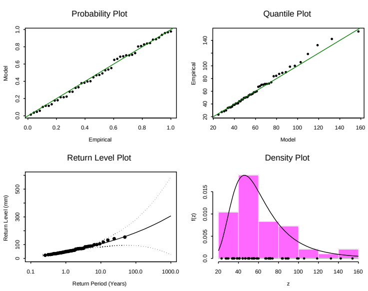
1. Specification of log-likelihood function:

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

2. Numerical maximization of log-likelihood.
3. Calculation of standard errors from inverse of observed information matrix (also obtained numerically).
4. Diagnostic checks: probability plots, quantile plots, return level plots.
5. Calculation of confidence intervals for return levels.

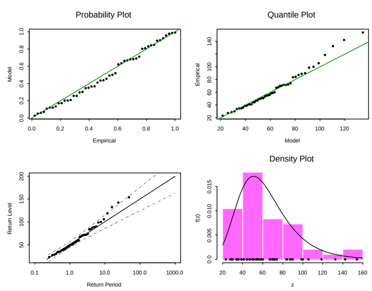
Example: Venezuelan Rainfall

```
gev.fit(rain)

$nllh:
[1] 224.0537

$mle:
[1] 49.0496173 19.9318432  0.1662012

$se:
[1] 3.365197 2.662967 0.139700
```



Example: Venezuelan Rainfall (Gumbel Model)

```
gum.fit(rain)

$nllh:
[1] 224.8877

$mle:
[1] 50.90696 21.54820

$se:
[1] 3.261058 2.545455
```



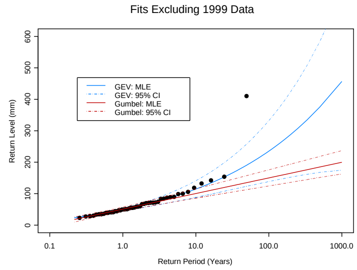
GEV or Gumbel?

Is GEV model preferable to Gumbel?

1. Estimate of ξ in GEV model is 0.166 with standard error of 0.140. Estimate is little more than one standard error from zero, suggesting data are consistent with $\xi = 0$.
2. Consider deviance statistic
$$D = 2\{224.89 - 224.05\} = 1.68.$$
The bigger the value of D, the stronger the evidence for $\xi \neq 0$. More formally, the strength of evidence for $\xi \neq 0$ is approximated by the exceedance probability of D on a χ_1^2 distribution:
```
1-pchisq(1.68,1)
0.1949245
```
3. Despite 1. and 2. the diagnostic plots of the Gumbel model are poorer than those of the GEV.

The bigger picture

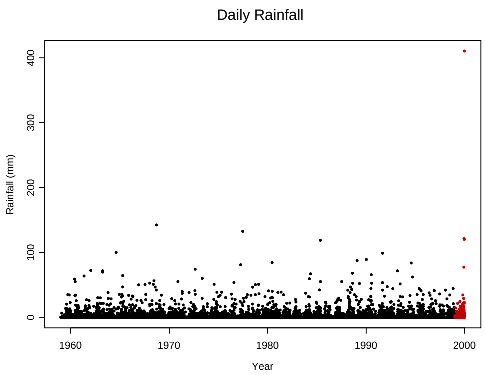
In actual fact, the Gumbel model proved to be **hopelessly inadequate** when the 1999 event occurred.



Conclusion: Even if formal tests support model reduction from GEV to Gumbel, the GEV model should be preferred, as its conservatism affords increased protection.

Threshold Models

Modelling only annual maxima is wasteful if other data are also available. For example, the Venezuelan data:



Modelling threshold exceedances

Threshold approach leads to approximate families for

$$\Pr(X < u + y \mid X > u) \quad y > 0$$

for large values of the threshold u. Similar arguments to those leading to GEV generate the Generalized Pareto family:

$$H(y) = 1 - \left(1 + \frac{\xi y}{\tilde{\sigma}}\right)^{-1/\xi}$$

defined on $\{y : \; y > 0 \text{ and } (1 + \xi y/\tilde{\sigma}) > 0\}$.

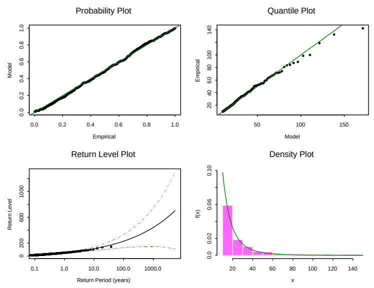
Threshold Modelling Strategy

1. Determine suitable threshold (various diagnostics available – look for linearity of mean residual life plot);
2. Maximize likelihood of GPD model for threshold exceedances;
3. Calculate standard errors etc.;
4. Validate model fit;
5. Use model to estimate return levels.

$$z_N = u + \frac{\sigma}{\xi}\left[(N n_y \lambda_u)^{\xi} - 1\right],$$

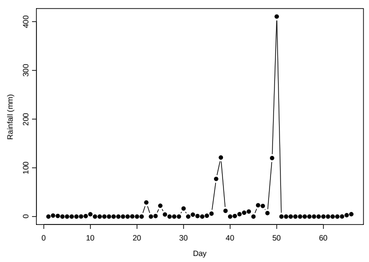
Example: Venezuelan data

```
gpd.fit(rain.day)

$nllh:
[1] 1884.81

$mle:
[1] 10.2016905  0.2607349

$se:
[1] 0.75118031 0.06041446
```



Other Issues 1: Temporal Dependence

In practice data usually display **temporal dependence**.



However:

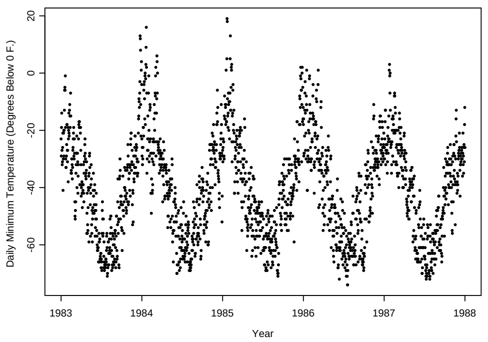
> **Standard limit results are not invalidated provided dependence is 'short range'.**

In particular:

1. GEV can be estimated for annual maximum without any alteration to procedure;
2. GPD can also be estimated for distribution of threshold exceedances, though allowance should be made for fact that observations are not independent. (Various techniques available: simplest is to identify clusters of extremes and fit GPD to cluster maxima only).

Other Issues 2: Non-Stationarity

Data are often non-stationary: trends, seasonality etc.

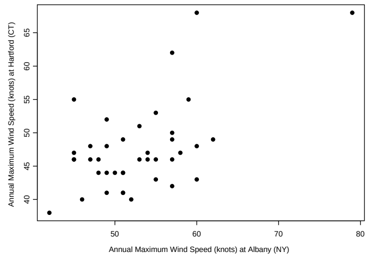


Usual approach:

Model non-stationarity as time-variation in extreme value parameters. Estimate model through maximum likelihood.

Other Issues 3: Multivariate Extremes

Extremes of different processes may be statistically related.



A joint analysis of extremes of several processes, using multivariate analogs of the standard extreme value models, enables:

- Improved precision through information transfer;
- Probability estimates of combined extreme events.

Other Issues 4: Bayesian Inference

Modern computational techniques (MCMC) enable ‘straightforward’ Bayesian inferences for extreme value models as an alternative to Maximum Likelihood. Advantages include:

- Facility to incorporate genuine prior knowledge or information;
- Inferences that have a simplified interpretation and that do not rely on asymptotic approximations of the likelihood function;
- More coherent approach to inference.

In particular, predictive versions of return level plots can be produced that account for parameter uncertainty.