NCAR

Statistical Extreme Value Theory (EVT) Part I

Eric Gilleland

Research Applications Laboratory

21 July 2014, Trieste, Italy

National Center for Atmospheric Research

What is extreme?

NCAR

- The upper quartile of temperatures during the year?
- The maximum temperature during the year?
- The highest stream flow recording during the day?
- The lowest stream flow recordings accumulated over a three-day period for the year?
- Winning the lottery?

What is extreme?

- When a variable exceeds some high threshold?
- When the year's maximum of a variable is very different from the usual maximum value?
- When an unusual event takes place, regardless of whether or not it is catastrophic?
- Only when an event causes catastrophes?

What is extreme?

- The United States wins the World Cup for Soccer?
- Any team wins the World Cup?
- The Denver Broncos win the Super Bowl?

What is extreme?

NCAR

Suppose an “event” of interest, E, has a probability, p, of occurring and a probability 1 – p of not occurring.

Then, the number of events, N, that occur in *n* trials follows a binomial distribution. That is, the probability of having *k* “successes” in *n* trials is governed by the probability distribution:

$$\Pr\{N = k\} = \binom{n}{k} p^k (1 - p)^{n-k}$$

What is extreme?

$$\Pr\{\mathrm{N}=k\}=\binom{n}{k}p^{k}(1-p)^{n-k}$$

Now, suppose p is very small. That is, E has a very low probability of occurring.

Simon Denis Poisson introduced the probability distribution, named after him, obtained as the limit of the binomial distribution when p tends to zero at a rate that is fast enough so that the expected number of events is constant.

That is, as n goes to infinity, the product *n*p remains fixed.

What is extreme?

That is, for p "small," N has an approximate Poisson distribution with intensity parameter *nλ*.

$$\Pr\{\mathrm{N}=0\}=e^{-n\lambda},\ \Pr\{N>0\}=1-e^{-n\lambda}$$

What is extreme?

NCAR

We're looking for events with a very low probability of occurrence (e.g., if $n \geq 20$ and $p \leq 0.05$ or $n \geq 100$ and $np \leq 10$).

Does not mean that the event is impactful!

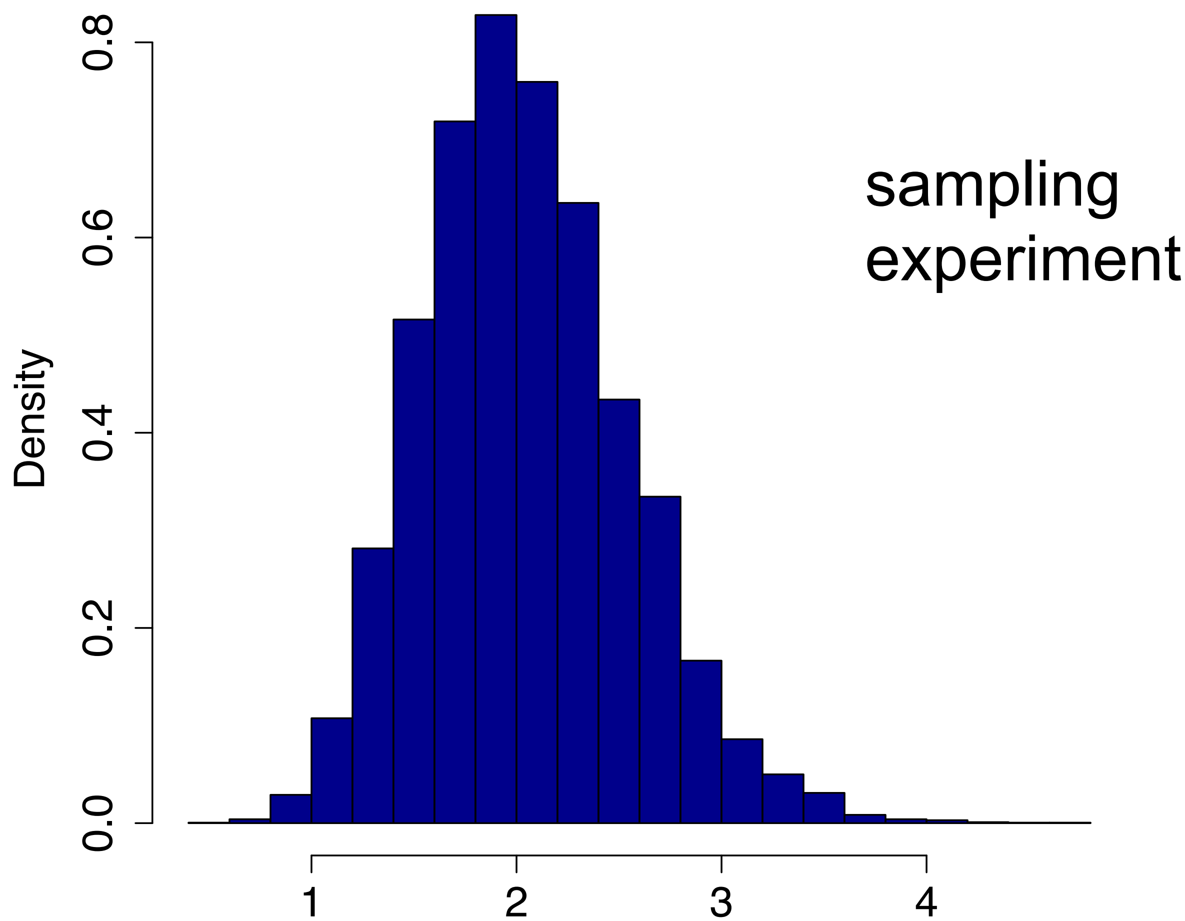
Does not mean that an impactful event must be governed by EVT!

Inference about the maximum

- If I have a sample of size 10, how many maxima do I have?
- If I have a sample of size 1000?
- Is it possible, in the future, to see a value larger than what I've seen before?
- How can I infer about such a value?
- What form of distribution arises for the maximum?

Extreme Value Theory Background

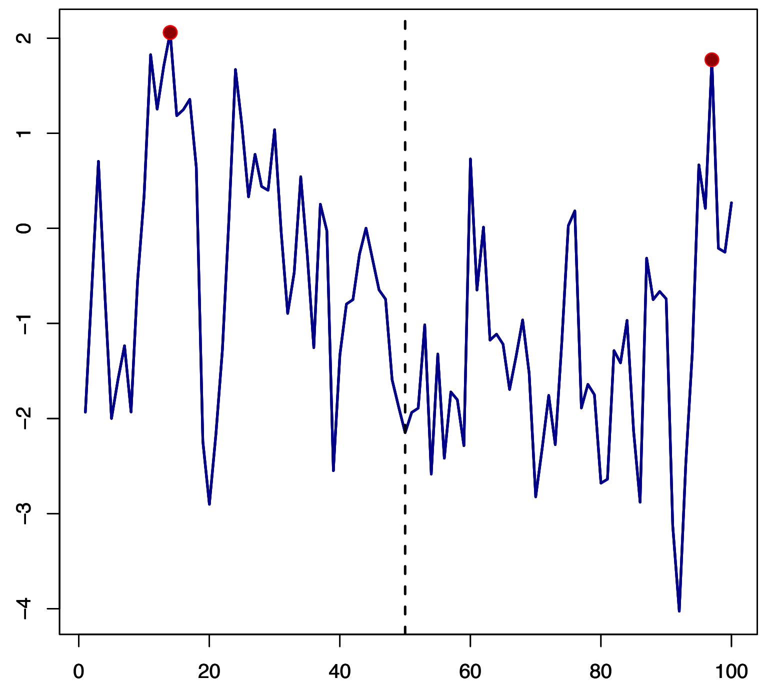
Maxima of samples of size 30
from standard normal distribution



Extreme Value Theory Background: Max Stability

$\max\{ x_1, \ldots, x_{2n} \} =$

$\max\{ \max\{ x_1, \ldots, x_n \}, \max\{ x_{n+1}, \ldots, x_{2n} \}\}.$



Extreme Value Theory Background: Max Stability

In other words, the cumulative distribution function (cdf), say G, must satisfy

$$G^2(x) = G(ax + b)$$

where $a > 0$ and b are constants.

Extreme Value Theory Background: Extremal Types Theorem

Let $X_1,\ldots,X_n$ be independent and identically distributed, and define $M_n = \max\{X_1,\ldots,X_n\}$.

Suppose there exist constants $a_n > 0$ and b_n such that

$$\Pr\{ (M_n - b_n) / a_n \leq x \} \longrightarrow G(x) \text{ as } n \longrightarrow \infty,$$

where G is a non-degenerate cdf.

Extreme Value Theory Background: Extremal Types Theorem

Then, G must be a generalized extreme value (GEV) cdf. That is,

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

defined where the term inside the [ ] and σ are positive.

Extreme Value Theory Background: Extremal Types Theorem

Note that this is of the same form as the Poisson distribution!

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

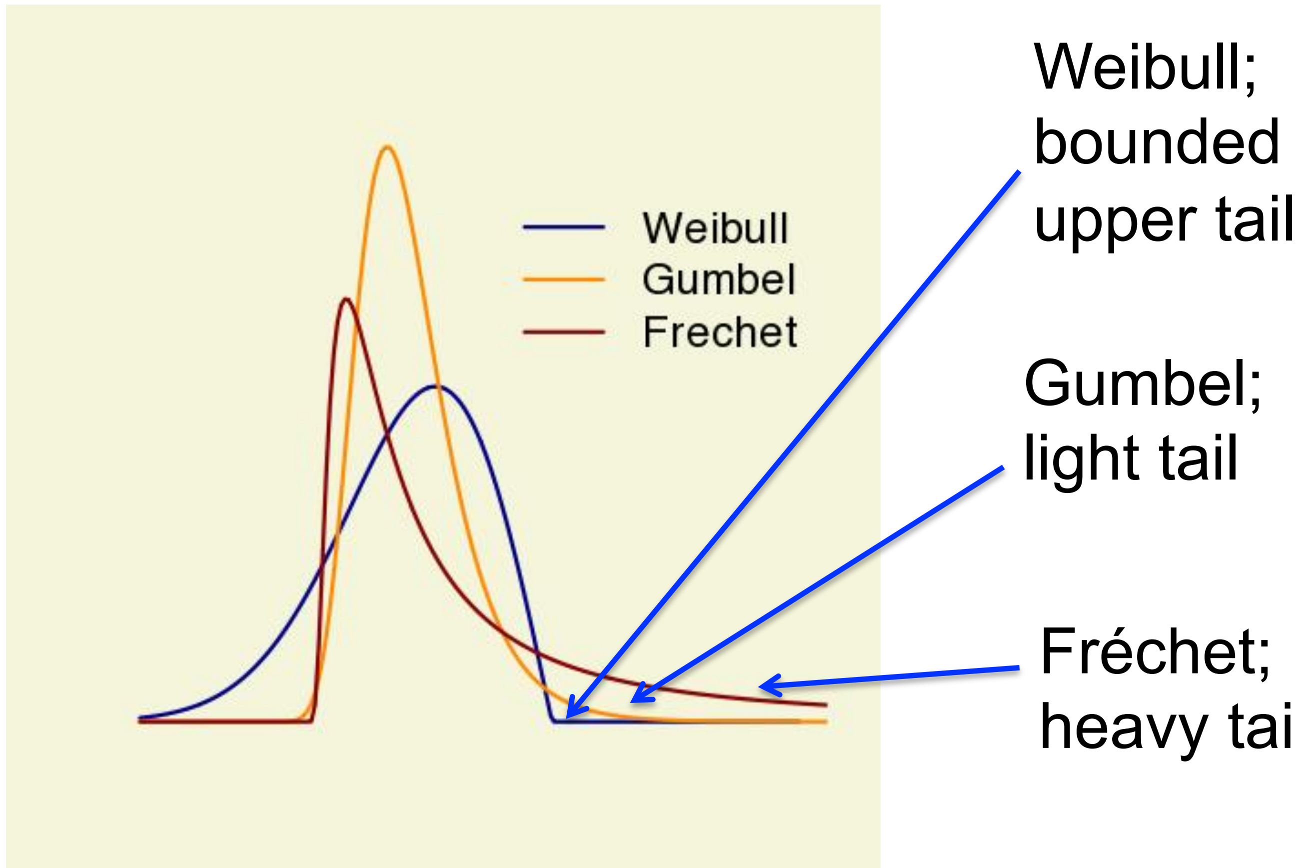
Extreme Value Theory Background: Extremal Types Theorem

Three parameters:

- μ is the location parameter
- σ > 0 the scale parameter
- ξ is the shape parameter

Extreme Value Theory Background: Extremal Types Theorem

Three types of Extreme Value df's



Extreme Value Theory Background: Extremal Types Theorem

Paula Radcliffe, 11.6 mph world marathon record London Marathon, 13 April 2003

Predicted Speed Limits

Thoroughbreds (Kentucky Derby)	**≈ 38 mph**
Greyhounds (English Derby)	≈ 38 mph
Men (100 m distance)	≈ 24 mph
Women (100 m distance)	≈ 22 mph
Women (marathon distance)	≈ 12 mph
Women (marathon distance using a different model)	≈ 11.45 mph

Denny, M.W., 2008, *J. Experim. Biol*., **211**:3836–3849.

Extreme Value Theory Background: Extremal Types Theorem

NCAR

Weibull type:
temperature, wind speed, sea level
negative shape parameter
bounded upper tail at:

$$\mu - \frac{\sigma}{\xi}$$

Extreme Value Theory Background: Extremal Types Theorem

Gumbel type:

"Domain of attraction" for many common distributions (e.g., normal, exponential, gamma)

limit as shape parameter approaches zero.

"light" upper tail

Extreme Value Theory Background: Extremal Types Theorem

NCAR

Fréchet type:

precipitation, stream flow, economic damage

positive shape parameter

“heavy” upper tail

infinite k-th order moment if $k \geq 1 / \xi$
(e.g., infinite variance if $\xi \geq ½$)

Extreme Value Analysis

- Fit directly to *block* maxima, with relatively long blocks
  - annual maximum of daily precipitation amount
  - highest temperature over a given year
  - annual peak stream flow
- Advantages
  - Do not necessarily need to explicitly model annual and diurnal cycles
  - Do not necessarily need to explicitly model temporal dependence

Extreme Value Analysis

NCAR

Parameter estimation

- Maximum Likelihood Estimation (MLE)
- L-moments (other moment-based estimators)
- Bayesian estimation
- various fast estimators (e.g., Hill estimator for shape parameter)

Extreme Value Analysis

MLE

Given observed block maxima

$Z_1 = z_1, \ldots, Z_m = z_m$,

minimize the negative log-likelihood

$(-\ln L(z_1, \ldots, z_m; \mu, \sigma, \xi))$

of observing the sample with respect to the three parameters.

Extreme Value Analysis

NCAR

MLE

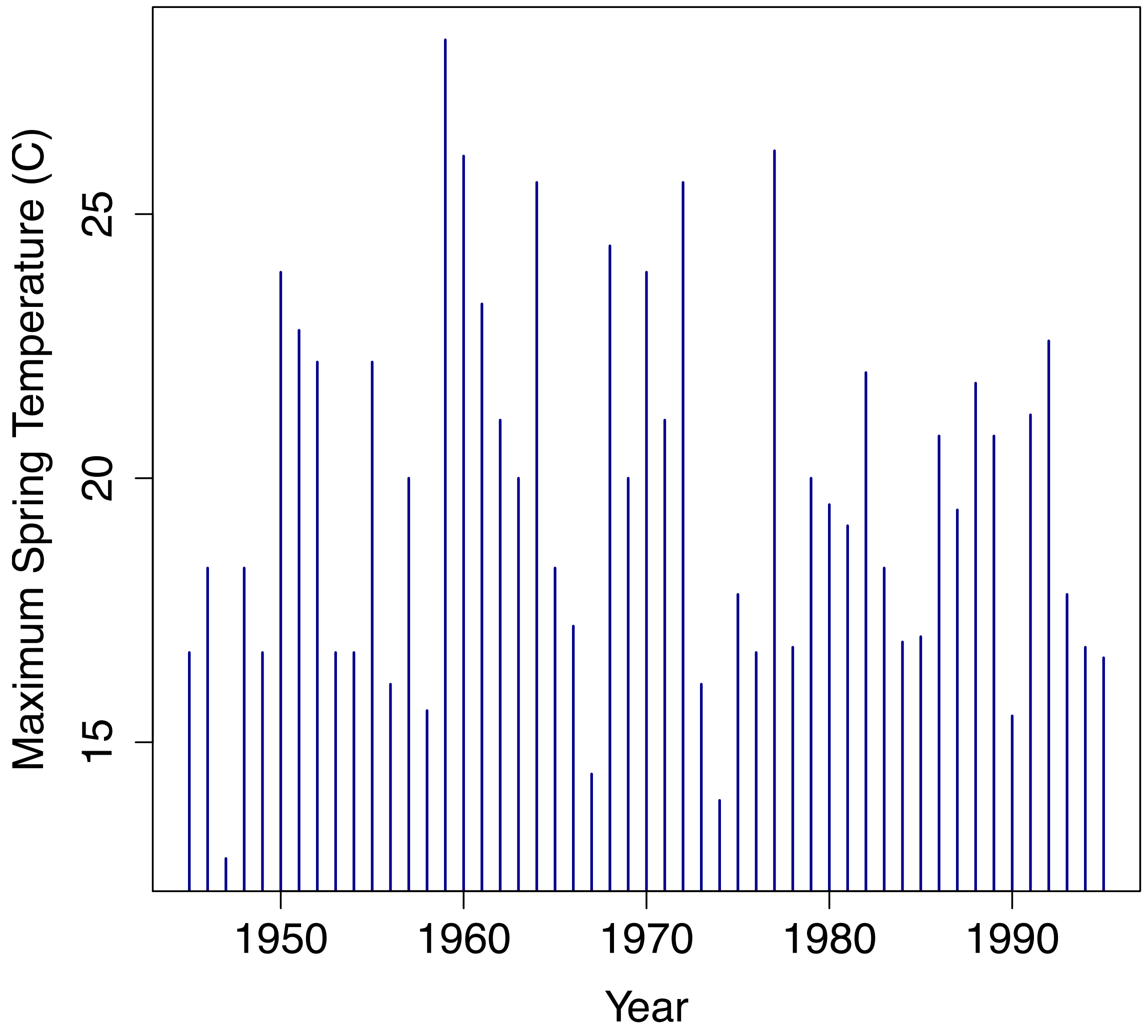
Allows for employing the likelihood-ratio test to test one model against another (nested) model.

Model 1: $-\ln L(z_1, \ldots, z_m; \mu, \sigma, \xi = 0)$

Model 2: $-\ln L(z_1, \ldots, z_m; \mu, \sigma, \xi)$

If $\xi = 0$, then $V = 2 * (\text{Model 2} - \text{Model 1})$ has approximate χ^2 distribution with 1 degree of freedom for large m.

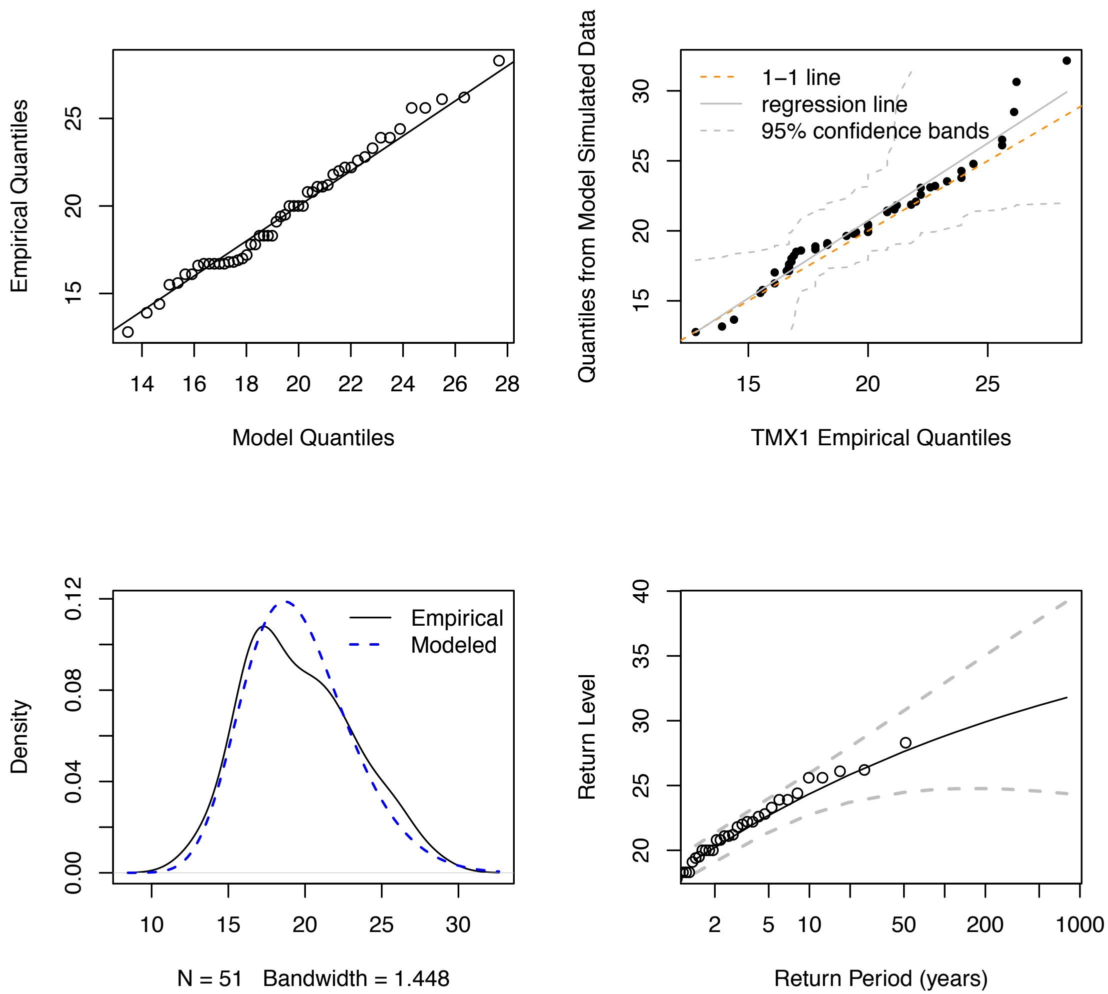
Extreme Value Analysis



Sept, Iles, Québec

Extreme Value Analysis

fevd(x = TMX1, data = SEPTsp)



Sept, Iles, Québec

Extreme Value Analysis

NCAR

	95% lower CI	Estimate	95% upper CI
μ	17.22	18.20	19.18
σ	2.42	3.13	3.84
ξ	-0.37	-0.14	0.09
100-year return level	24.72 °C	28.81 °C	32.90 °C

Sept, Iles, Québec

Extreme Value Theory: Return Levels

NCAR

Assume stationarity (i.e. unchanging climate)

Return period / Return Level

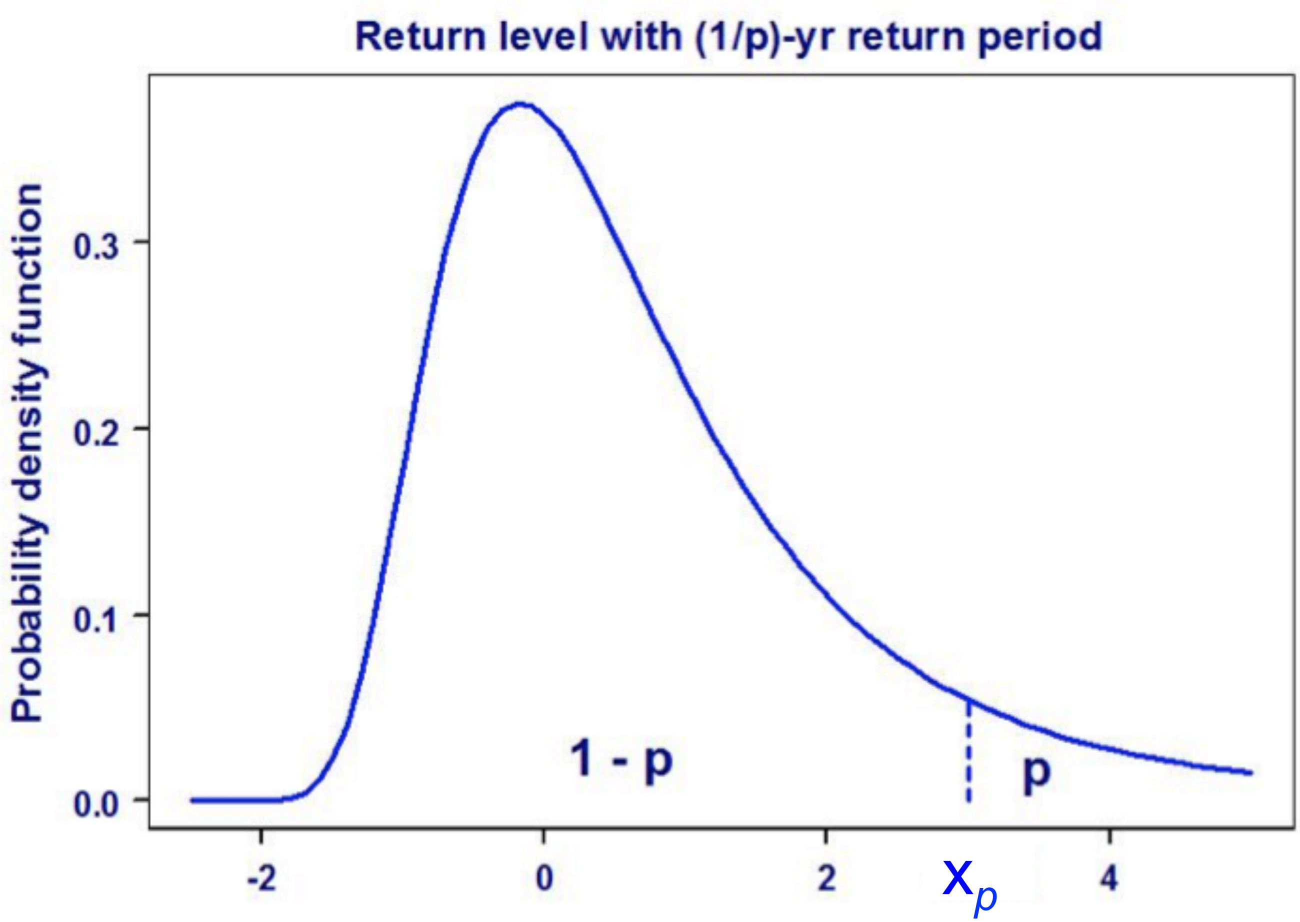
Seek x_p such that $G(x_p) = 1 – p$, where $1 / p$ is the return period. That is,

$$x_p = G^{-1}(1 – p; \mu, \sigma, \xi), 0 < p < 1$$

Easily found for the GEV cdf.

Example, $p = 0.01$ corresponds to 100-year return period (assuming annual blocks).

Extreme Value Theory: Return Levels



Conclusion of Part I: Questions?