Statistical Analysis of

EXTREMES in GEOPHYSICS

Zwiers FW and Kharin VV. 1998. Changes in the extremes of the climate simulated by CCC GCM2 under CO_2 doubling. *J. Climate* 11:2200–2222.

http://www.ral.ucar.edu/staff/ericg/readinggroup.html

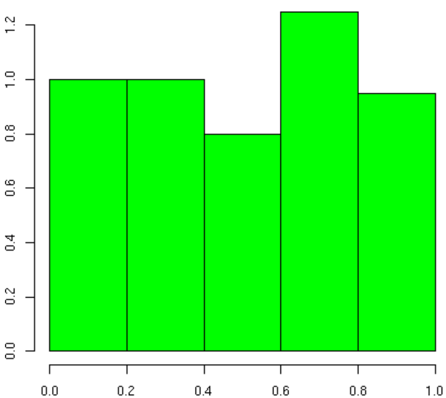
Outline

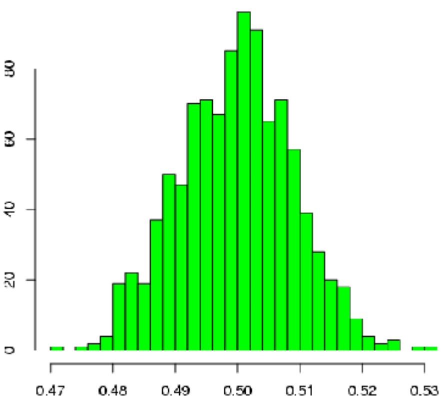
- Some background on Extreme Value Statistics
  - Extremal Types Theorem
  - Max-stability
  - Peaks Over Threshold Approach
  - Quantiles/Return values
  - Estimation methods
- Discuss paper

Statistical Analysis of Averages

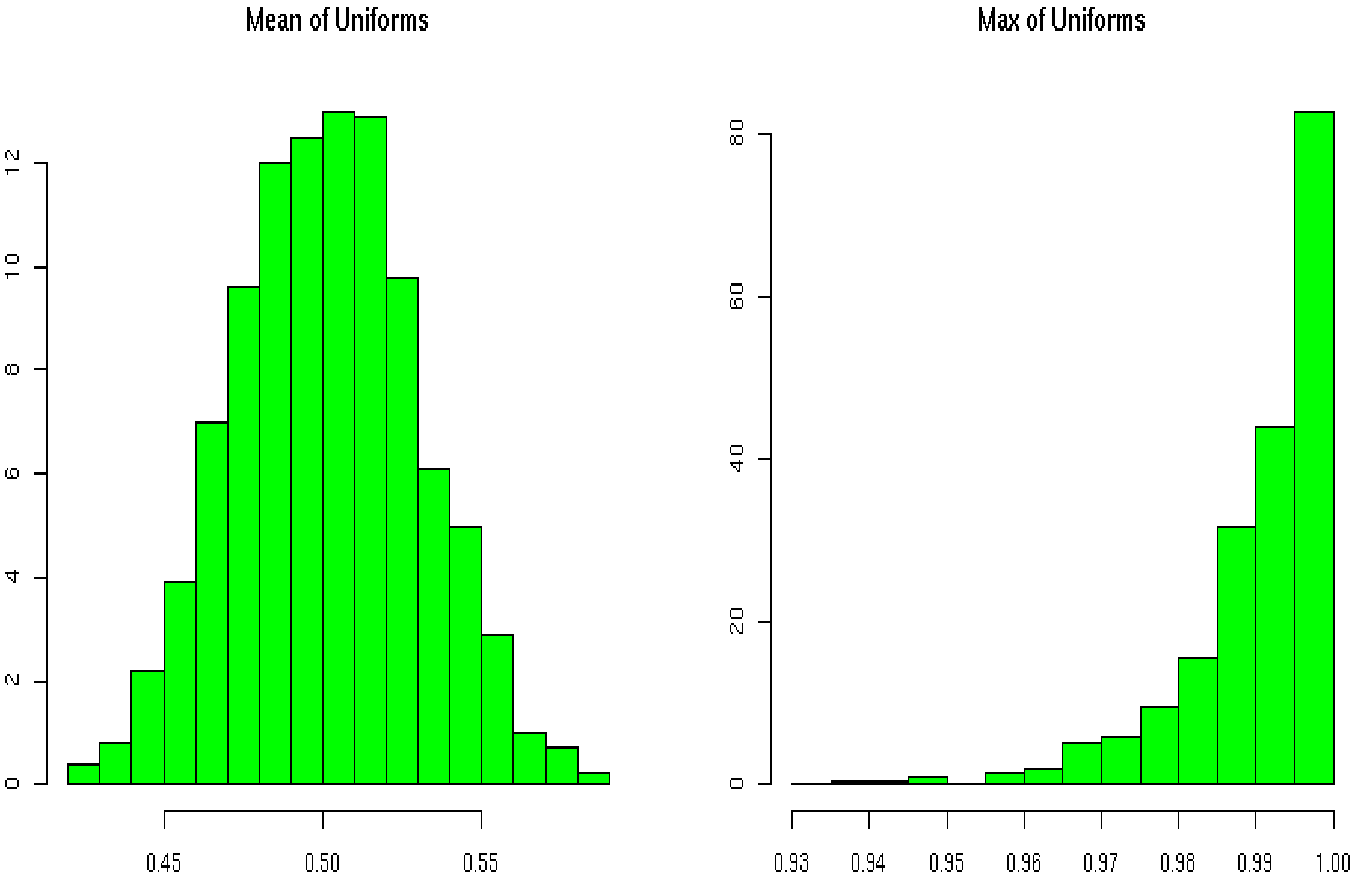
Central Limit Theorem: $X_1, \ldots, X_n$ random sample from any distribution.

$$\frac{X_1 + X_2 + \cdots X_n - n\mu}{\sigma\sqrt{n}} \longrightarrow Z \sim \mathrm{N}(0,1) \text{ as } n \longrightarrow \infty$$





What if interest is in *maximums* instead of *means*?



Statistical Analysis of Extremes

Extremal Types Theorem: $X_1, \ldots, X_n$ random sample from any distribution.

$$\mathrm{Pr}\{\frac{\max\{X_1, \ldots, X_n\} - b_n}{a_n} \leq z\} \longrightarrow G(z) \text{ as } n \longrightarrow \infty$$

where $G(z)$ is one of three types of distributions.

I. (Gumbel) $G(z) = \exp\{-\exp\left[-\left(\frac{z-b}{a}\right)\right]\}$, $-\infty < z < \infty$.

II. (Fréchet) $G(z) = \exp\{-\left(\frac{z-b}{a}\right)^{-\alpha}\}$, $z > b$ and 0 otherwise.

III. (Weibull) $\exp\{-\left[-\left(\frac{z-b}{a}\right)^{\alpha}\right]\}$, $z < b$ and 1 otherwise.

(where $a > 0$, $\alpha > 0$ and b are parameters).

Statistical Analysis of Extremes

Extremal Types Theorem

The above three distributions can be combined into a single family of distributions.

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

G is called the generalized extreme value distribution (GEV).

Three parameters: location (μ), scale (σ) and shape (ξ).

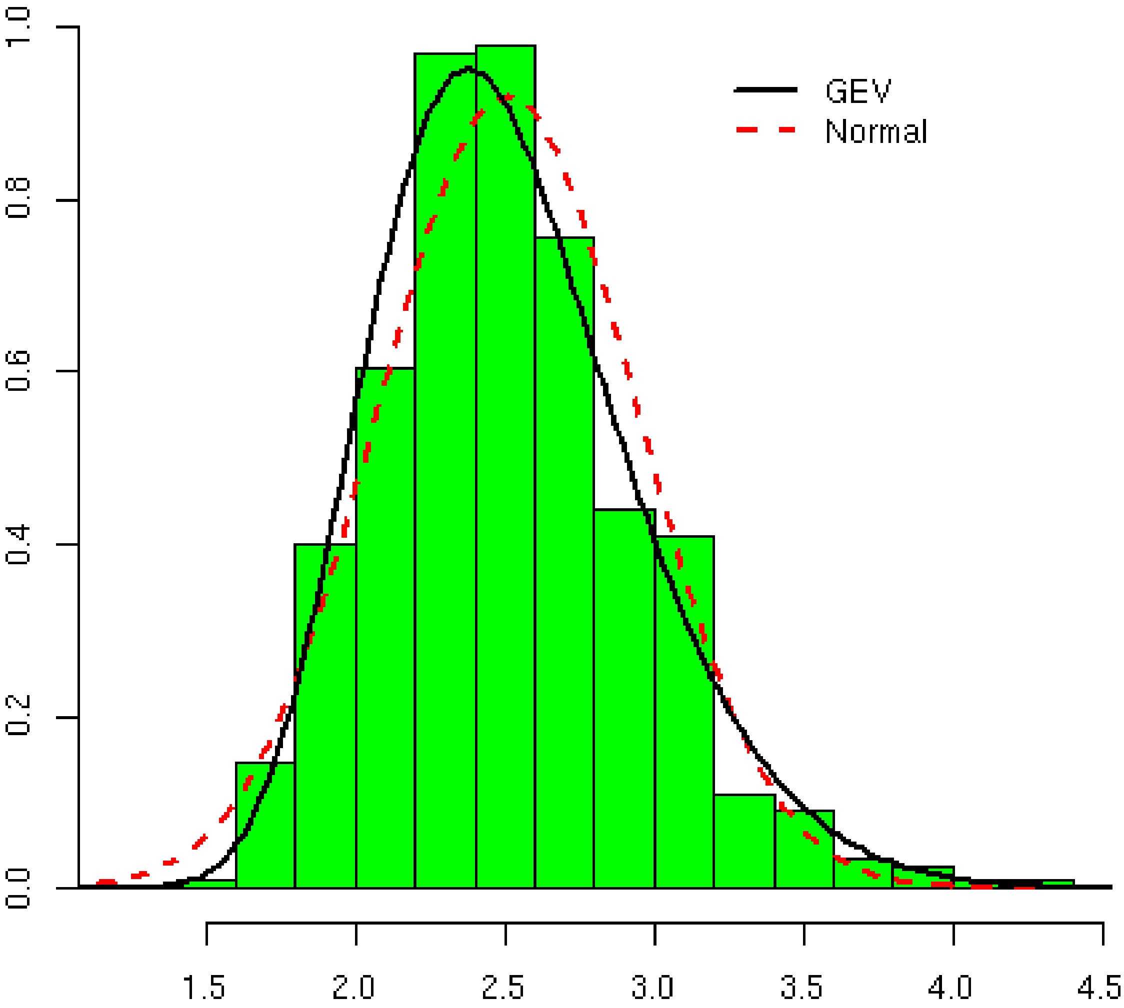
This paper uses ξ for location, α for scale and κ for shape.

Statistical Analysis of Extremes

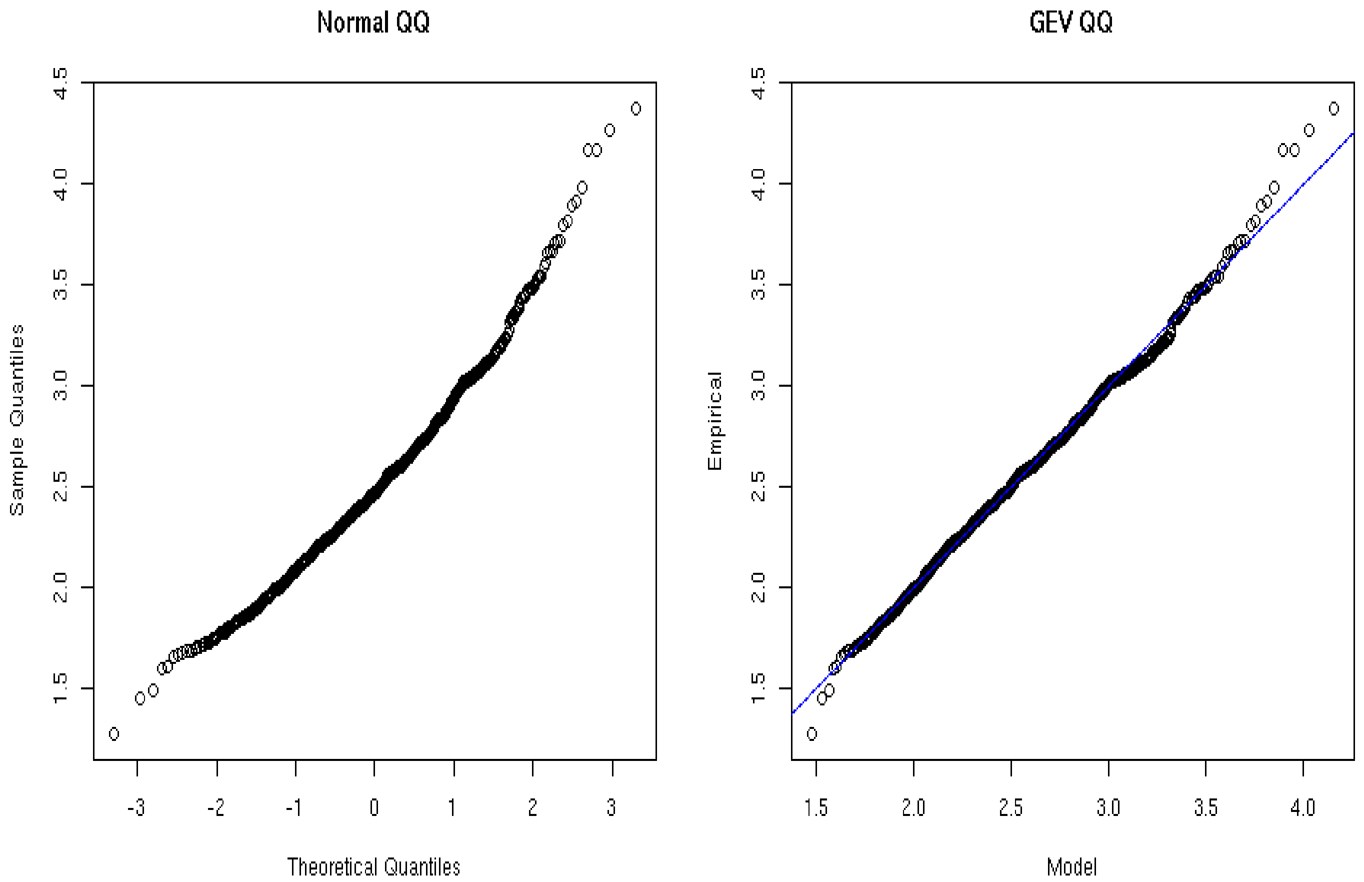
Note that the GEV is only defined when $1+\xi\left(\frac{z-\mu}{\sigma}\right) > 0$. Also,

- $\xi \longrightarrow 0$ corresponds to Gumbel.
- $\xi > 0$ corresponds to Fréchet (lower end-point at $\mu - \frac{\sigma}{\xi}$).
- $\xi < 0$ corresponds to Weibull (upper end-point at $\mu - \frac{\sigma}{\xi}$).

Statistical Analysis of Extremes



Statistical Analysis of Extremes



Max-Stability:

Definition: A distribution G is said to be max-stable if, for every $n = 2, 3, \ldots$, there are constants $\alpha_n > 0$ and β_n such that

$$G^n(\alpha_n z + \beta_n) = G(z).$$

Theorem: A distribution is max-stable if, and only if, it is a generalized extreme value distribution.

Generalized Pareto Distribution (GPD)

Exceedance Over Threshold Model

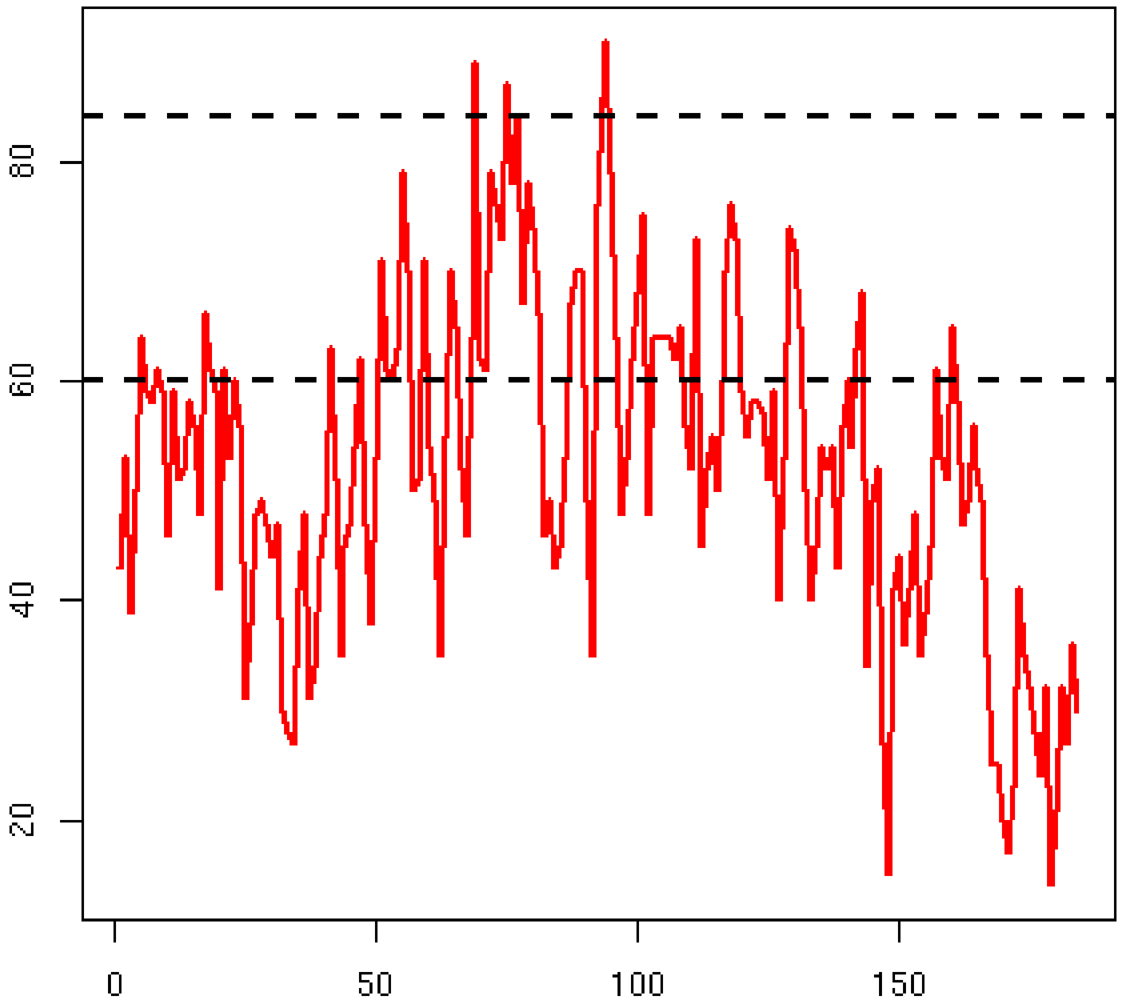
For X random (with cdf F) and a (large) threshold u

$$\Pr\{X > x | X > u\} = \frac{1 - F(x)}{1 - F(u)}$$

Then for $x > u$ (u *large*), the GPD is given by

$$\frac{1 - F(x)}{1 - F(u)} \approx [1 + \frac{\xi}{\sigma}(x - u)]^{-1/\xi}$$

Extreme Value Distributions: GPD



Fitting the GPD to data

Method

Maximum Likelihood is easy to evaluate and maximize (also can use L-moments, PWM, Bayesian).

Threshold Selection: Variance vs. Bias

Trade-off between a low enough u to have enough data (low variance), but high enough for the limit model to be a reasonable approximation (low bias).

Confidence Intervals

The parameter distributions are generally skewed. So, the best method for finding confidence intervals (or sets) are based on the likelihood value (or surface) and a χ^2 critical value.

Quantiles and Return Levels

For any extreme value distribution, interest is typically in quantiles.

Quantile Function

The p-quantile of a distribution function, F, is $x_p = F^{-1}(p)$. (i.e., the probability of exceeding x_p is $1 - p$)

For a GPD fit to *daily* exceedances (365 days), the 100-yr. *return level* is the

$$1 - \frac{1}{365 * 100} \text{ quantile}$$

Quantiles and Return Levels

GPD Quantile function

$$x_p = \begin{cases} u + \frac{\sigma}{\xi}[(\zeta_u \cdot \frac{1}{p})^{\xi} - 1] & \xi \neq 0 \\ \\ u + \sigma \log(\zeta_u \cdot \frac{1}{p}) & \xi = 0 \end{cases}$$

where $\zeta_u = \Pr(X > u) = 1 - F(u)$.

Estimation methods

- Maximum Likelihood Estimation (MLE)
- Method of L Moments
- Bayesian estimation

MLE

Assuming $Z_1, \ldots, Z_m$ are iid random variables that follow the GEV distribution the log-likelihood is given by the following.

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

L Moments

Probability Weighted Moments (PWM)

$$M_{p,r,s} = E\left[X^p\{F(X)\}^r\{1 - F(X)\}^s\right]$$

L-moments are based on the special cases $\alpha_r = M_{1,0,r}$ and $\beta_r = M_{1,r,0}$. Specifically, let $x(u)$ be the quantile function for a distribution, then:

$$\alpha_r = \int_0^1 x(u)(1-u)^r du$$

$$\beta_r = \int_0^1 x(u)u^r du$$

Compare to ordinary moments: $E(X^r) = \int_0^1 \{x(u)\}^r du$.

L-moments

Much more to it, but the moments derived in the paper come from:

- $\lambda_1 = \alpha_0 = \beta_0$,
- $\lambda_2 = \alpha_0 - 2\alpha_1 = 2\beta_1 - \beta_0$ and
- $\lambda_3 = \alpha_0 - 6\alpha_1 + 6\alpha_2 = 6\beta_2 - 6\beta_1 + \beta_0$

More generally

$$\lambda_r = \int_0^1 x(u) \sum_{k=0}^{r-1} \frac{(-1)^{r-k-1}(+k-1)!}{(k!)^2(r-k-1)!} du$$

Alternatively

- For $n = 1$, $X_{1:1}$ estimates location. If distribution is shifted to larger values, then $X_{1:1}$ is expected to be larger. (Hence, $\lambda_1 = E(X_{1:1})$)
- For $n = 2$, $X_{2:2} - X_{1:1}$ estimates scale (dispersion). If dist'n is tightly bunched, small value. (Hence, $\lambda_2 = \frac{1}{2}E(X_{2:2} - X_{1:2})$)
- For $n = 3$, $X_{3:3} - 2X_{2:3} + X_{1:3}$ measures skewness. (i.e., $X_{3:3} - X_{2:3} \approx X_{2:3} - X_{1:3}$). (Hence, $\lambda_3 = \frac{1}{3}E(X_{3:3} - 2X_{2:3} + X_{1:3})$)

And in general,

$$\lambda_r = r^{-1} \sum_{j=0}^{r-1} (-1)^j \frac{(r-1)!}{j!(r-j-1)!} E(X_{r-j:r})$$

For more on L-Moments

Hosking JRM and Wallis JR. 1997. *Regional Frequency Analysis: An Approach Based on L-Moments*. Cambridge University Press.

Some References

Applied Introductory references to extreme value statistical analysis

Coles S. 2001. *An Introduction to Statistical Modeling of Extreme Values*. Springer.

Gilleland E and Katz, RW. 2005. Tutorial to Extremes Toolkit.
http://www.assessment.ucar.edu/toolkit

Katz RW, Parlange MB, and Naveau P. 2002. Statistics of extremes in hydrology. *Advances in Water Resources*, **25**:1287–1304.

Smith RL. 2002. *Statistics of extremes with applications in environment, insurance and finance*.
http://www.stat.unc.edu/postscript/rs/semstatrls.pdf

Some References

More theoretical ...

Leadbetter MR, Lindgren G, Rootzen H. 1983. *Extremes and Related Properties of Random Sequences and Processes*, Springer, New York.

Some Possible Topics for Discussion

- Spatial (temporal) Data. Spatial Analysis?
- Statistical Significance.
- L-moments vs. MLE (vs. Bayesian).
- Parameter uncertainty (last paragraph 2203).
- Interpretations.
- Validation.

Other Things

- Next session
- Format
- Papers/topics
- Run?