AN INTRODUCTION TO EXTREME ORDER STATISTICS AND ACTUARIAL APPLICATIONS

2004 ERM Symposium, Chicago
April 26, 2004
Sessions CS 1E, 2E, 3E:
Extreme Value Forum

H. N. Nagaraja
(hnn@stat.ohio-state.edu)
Ohio State University, Columbus

Hour 1: ORDER STATISTICS - BASIC THEORY & ACTUARIAL EXAMPLES

Hour 2: EXTREME VALUES - BASIC MODELS

Hour 3: EXTREME VALUES - INFERENCE AND APPLICATIONS

Hour 1: ORDER STATISTICS - BASIC THEORY & ACTUARIAL EXAMPLES

1. Exact Distributions of Order Statistics
   - Uniform Parent
   - Exponential
   - Normal
   - Gumbel
2. Order Statistics Related Quantities of Actuarial Interest
   - Last Survivor Policy
   - Multiple Life Annuities
   - Percentiles and VaR (Value at Risk)
   - Tail Probabilities
   - Conditional Tail Expectation (CTE)
   - Exceedances and Mean Excess Function
3. Large Sample Distribution of Order Statistics
   - Central Order Statistics (Central Percentiles)
   - Extreme Order Statistics
   - Intermediate Order Statistics

4. Sum of top order statistics & sample CTE
5. Data Examples and Model Fitting
   - Nationwide Data
   - Danish Fire-loss Data
   - Missouri Flood level Data
   - Venice Sea Level Data

Hour 2: EXTREME VALUES - BASIC MODELS

1. Limit Distributions For the Maximum
   - Fréchet
   - Weibull
   - Gumbel (or the extreme-value)
2. Generalized Extreme Value (GEV) Distribution
3. Limit Distributions for the Other Upper Extremes
4. Limit Distributions for the Minimum
5. Exceedances - Generalized Pareto Distribution (GPD)
6. Examples

Hour 3: EXTREME VALUES - INFERENCE AND APPLICATIONS

1. GEV Distribution- Maximum Likelihood Estimates
2. Blocked Maxima Approach
3. Inference based on top k extremes
   - Pickands' Estimator
   - Hills Estimator
   - Estimating High Quantiles
4. GPD and Estimates Based on Exceedances
5. General Diagnostic Plots
   - QQ Plot
   - Empirical Mean Excess Function Plot
6. Examples
7. General Remarks
8. Software Resources
9. References

Hour 1

1 Exact Distribution of An Order Statistic

If the random variables $X_1, ..., X_n$ are arranged in order of magnitude and then written as

$$X_{(1)} \leq \cdots \leq X_{(n)},$$

we call $X_{(i)}$ the ith order statistic ($i = 1, ..., n$). We assume X_i to be statistically independent and identically distributed - a *random sample* from a continuous population with cumulative distribution function (cdf) $F(x)$ and probability density function (pdf) $f(x)$.

CDF and PDF of X_r

$$\begin{aligned} F_{(r)}(x) &= \Pr\{X_{(r)} \leq x\} \\ &= \sum_{i=r}^{n} \binom{n}{i} F^i(x)[1 - F(x)]^{n-i} \end{aligned}$$

$$\begin{aligned} \text{Minimum}: \ F_{(1)}(x) &= 1 - [1 - F(x)]^n \\ \text{Maximum}: \ F_{(n)}(x) &= [F(x)]^n \end{aligned}$$

$$f_{(r)}(x) = \frac{n!}{(r-1)!(n-r)!} F^{r-1}(x)[1 - F(x)]^{n-r} f(x).$$

Standard Uniform Parent – PDFs

$$f(x) = 1, 0 \le x \le 1.$$

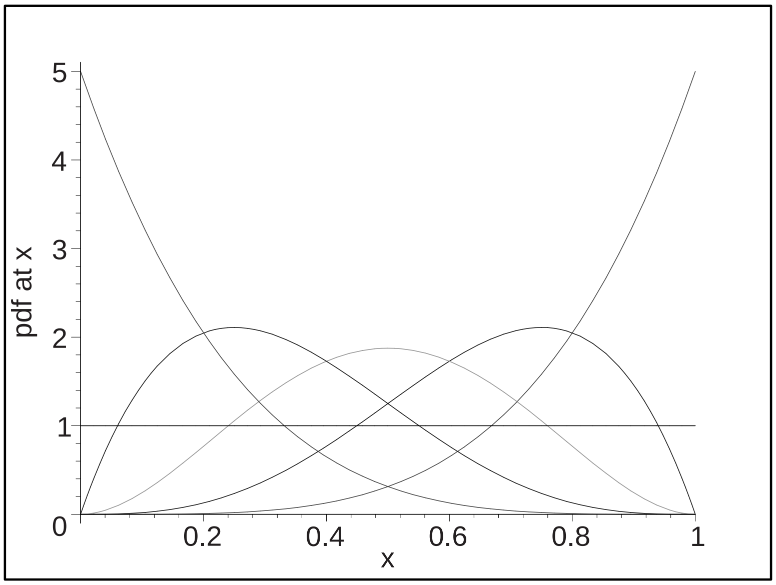


Figure 1: **The probability density function (pdf) of the standard uniform population and the pdfs of the first, second, third, fourth and largest order statistics from a random sample of size 5.**

Standard Uniform Parent – CDFs

$$F(x) = x, 0 \leq x \leq 1.$$

$$E(X_{(r)}) = \frac{r}{n+1}; \quad Var(X_{(r)}) = \frac{r(n-r+1)}{(n+1)^2}.$$

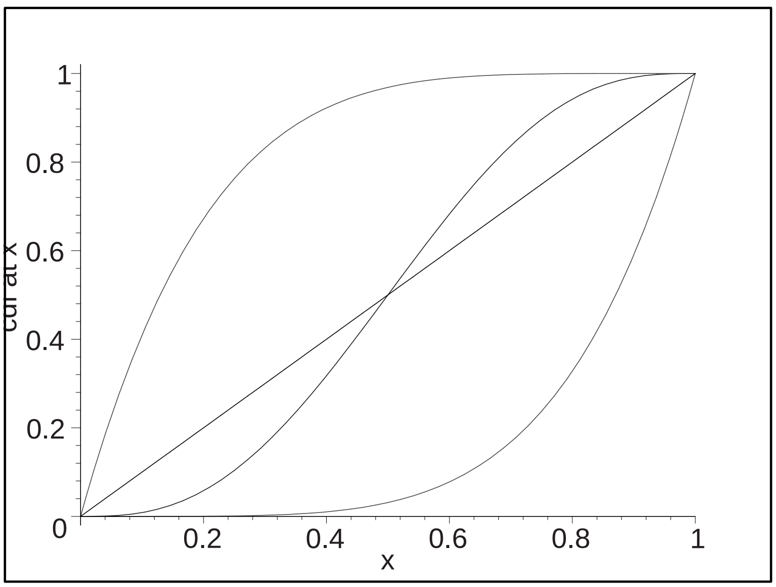


Figure 2: **The cumulative distribution function (cdf) of the standard uniform population and the cdfs of the first (minimum), third (median), and the largest (maximum) order statistics from a random sample of size 5.**

Standard Normal Parent – PDFs

$$f(x;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma}\exp\{-\frac{1}{2\sigma^2}(x-\mu)^2\}, -\infty < x < \infty.$$

Here μ is the mean and σ is the standard deviation.
Standard Normal: $\mu = 0; \sigma = 1$.

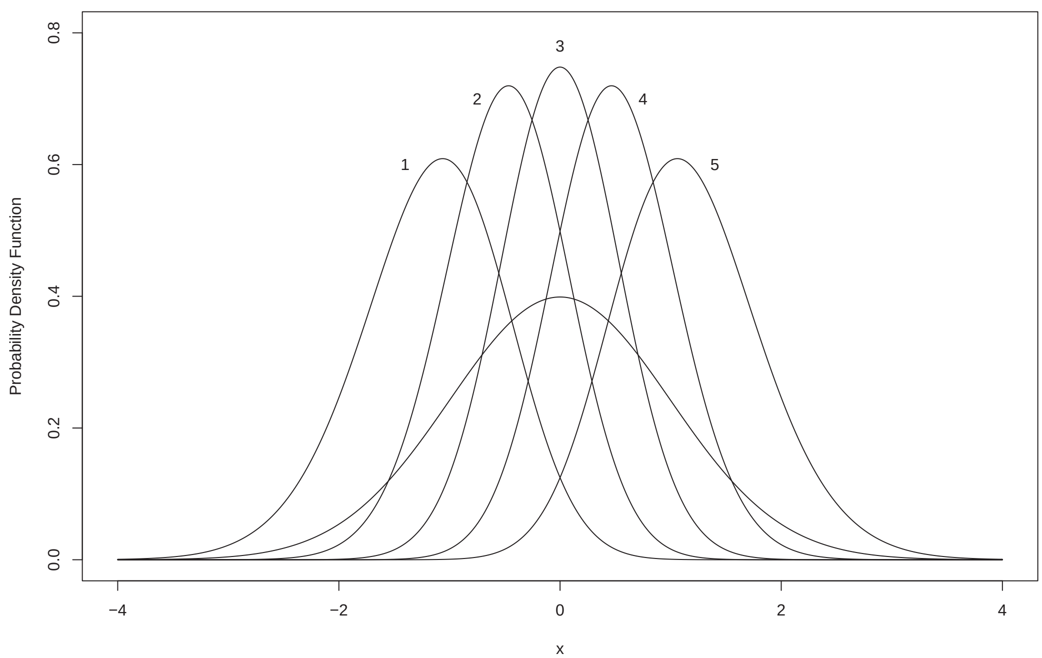


Figure 3: **The pdf of the standard normal population and the pdfs of the first, second, third, fourth and the largest order statistics from a random sample of size 5.**

Standard Normal Parent – PDFs of the Maxima

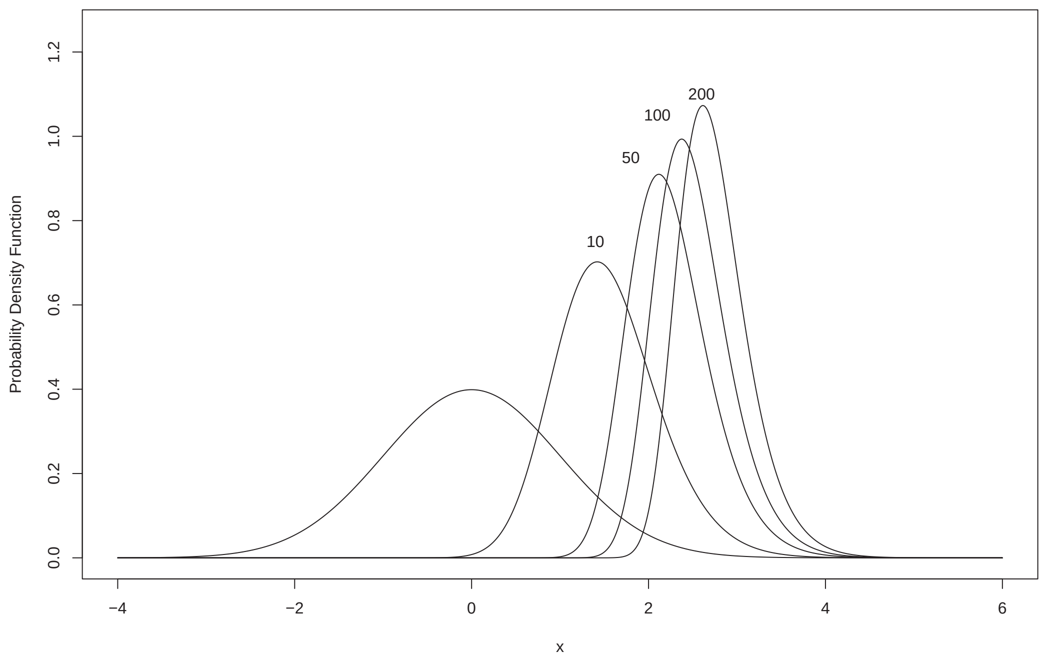


Figure 4: **The pdf of the standard normal parent and the pdfs of the maximum from random samples of size** $n = 10, 50, 100, 200$**.**

Standard Exponential Parent – PDFs

$$f(x;\lambda) = \lambda \exp\{-\lambda x\}, 0 \leq x < \infty.$$

Here mean = std. dev. = $1/\lambda$; Std. Exponential: $\lambda = 1$.

Moments for the Max:

$E(X_{(n)}) \approx \log(n) - 0.5772$; $Var(X_{(n)}) \approx \pi^2/6 \approx 1.64$.

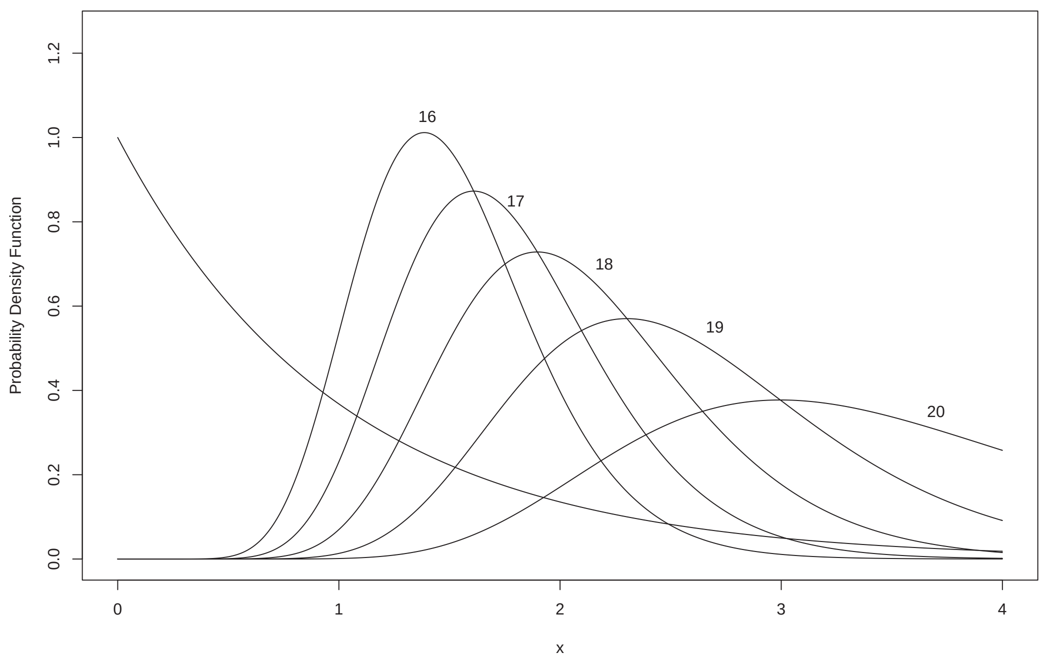


Figure 5: **The pdfs of the standard exponential parent and those of the top 5 order statistics from a random sample of size 20.**

Standard Gumbel Parent – PDFs

$$F(x;a,b) = \exp\{-\exp[-(x-\mu)/\sigma]\}, -\infty < x < \infty.$$

Interesting Fact: $X_{(n)}$ is Gumbel with μ and σ/n.
Standard Gumbel: $\mu = 0, \sigma = 1$;
$E(X) = -\gamma(\text{Euler's constant}) \approx -0.5772$;
Variance $= \pi^2/6$.

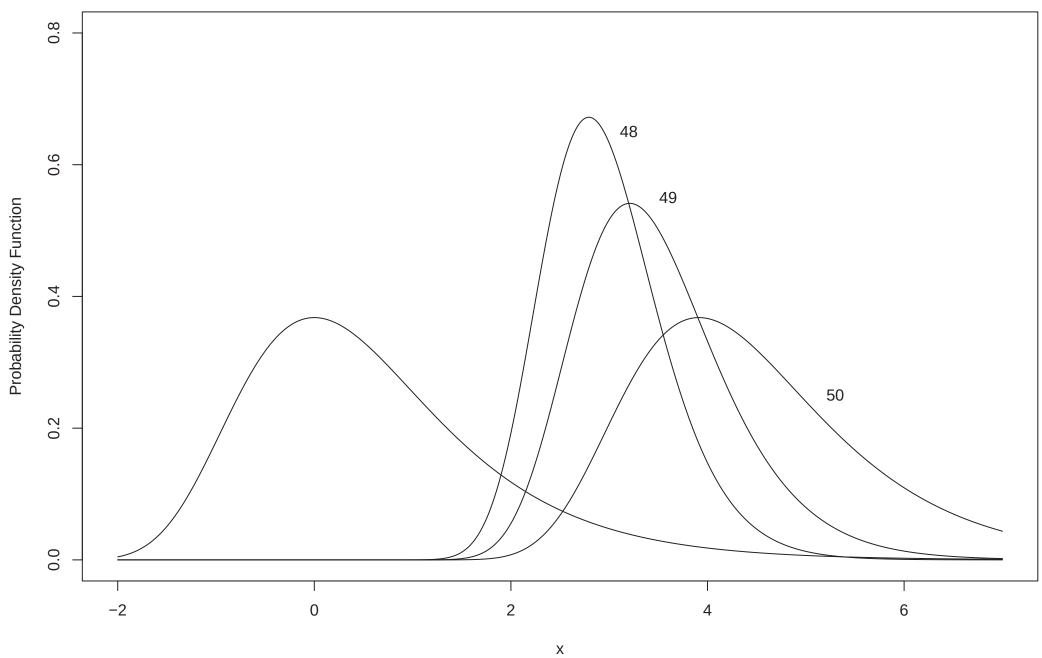


Figure 6: **The pdfs of the standard Gumbel parent and those of the top 3 order statistics from a random sample of size $n = 50$.**

2 Order Statistics Related Quantities of Actuarial Interest

Last Survivor Policy and Multiple Life Annuities

(a) *Last Survivor Policy:* Distribution of the Maximum; of interest will be the mean, standard deviation and percentiles of the distribution.

(b) *Multiple Life Annuities:* Distribution of the rth order statistic from a sample of size n; say the third to die out of five.

Percentiles and VaR (Value at Risk)

Upper pth percentile of the Distribution of X. Here X could be the random variable corresponding to profits and losses (P&L) or returns on a financial instrument over a certain time horizon. Let

$$x_p = F^{-1}(1 - p),$$

where F^{-1} is the *inverse cdf* or the *quantile* function.

If $p = 0.25$, or $p = .05$ say – central order statistics, $p = 0.01$ or $p = 0.001$ – extreme order statistics. VaR = *Value-at-Risk* is generally an extreme percentile.

Exceedances and Tail Probabilities

Exceedance is an event that an observation from the population of X exceeds a given threshold c. Probability exceeding or below a certain threshold c; probability of ruin.

$$\Pr(X > c).$$

When c is an extreme threshold, we need EVT.

Conditional Tail Expectation (CTE)

Suppose the top $100p\%$ of the population are selected and one is interested in the average of the selected group.

$$E(X|X > x_p) = \mu_p = \frac{1}{p}\int_{x_p}^{\infty} xf(x)dx.$$

Mean Excess Function

Suppose c is the threshold and the interest is the average excess above it.

$$\begin{aligned} E(X - c|X > c) = m(c) &= \frac{1}{1 - F(c)}\int_c^{\infty}(x - c)f(x)dx \\ &= \frac{1}{1 - F(c)}\int_c^{\infty}[1 - F(x)]dx. \end{aligned}$$

3 Large Sample Distribution of Order Statistics

- Central Order Statistics (Central Percentiles)
- Extreme Order Statistics
- Intermediate Order Statistics

The asymptotic theory of order statistics is concerned with the distribution of $X_{r:n}$, suitably standardized, as n approaches ∞. If $r/n \approx p$, fundamentally different results are obtained according as

(a) $0 < p < 1$ (central or quantile case)
(b) r or $n - r$ is held fixed (extreme case)
(c) $p = 0$ or 1, with $r = r(n)$ (intermediate case).

In Cases (a) and (c) the limit distribution is generally normal. In case (b), it is one of the three *extreme-value* distributions, or a member of the family of *Generalized Extreme Value* (GEV) distributions.

Central Case

Result 1. *Let* $0 < p < 1$, *and assume* $r \approx np$, *and* $0 < f(F^{-1}(p)) < \infty$. *Then the asymptotic distribution of*

$$n^{\frac{1}{2}}(X_{r:n} - F^{-1}(p))$$

is normal with zero mean and variance

$$\frac{p(1-p)}{[f(F^{-1}(p))]^2}.$$

- With $0 < p_1 < p_2 < 1$,

  $$\left(n^{\frac{1}{2}}(X_{r_1:n} - F^{-1}(p_1)), n^{\frac{1}{2}}(X_{r_2:n} - F^{-1}(p_2))\right)$$

  is bivariate normal with correlation

  $$\sqrt{\frac{p_1(1-p_2)}{p_2(1-p_1)}}.$$

- These limiting means, variances and correlations can be used to approximate the moments of central order statistics.

- One can use this result to estimate and find confidence intervals for central percentiles.

- One can find the large sample distribution of, e.g., the sample interquartile range (IQR).

Intermediate Case

Result 2. *Suppose the upper-tail of F satisfies some smoothness conditions (von Mises conditions). As $r \to \infty$ and $r/n \to 0$, with $p_n = r/n$,*

$$nf(x_{p_n})r^{-\frac{1}{2}}\left(X_{n-r+1:n} - z_{p_n}\right)$$

is asymptotically standard normal.

- x_p is the upper pth quantile.
- There are 3 types of von Mises conditions - tied to EVT.
- Here and in the quantile case, one can obtain local estimates of the pdf f using neighboring order statistics.

Upper and Lower Extremes

None of the limit distributions is normal. There are 3 families of distributions that generate the limit distributions for the upper extremes, and for the lower extremes.

- A nonsymmetric parent can have different types of limit distributions for the max. and min.

- The limit distribution for the kth maximum is related to, but different from that of the maximum.

Example. Consider a random sample from a standard exponential population. That is,

$$\Pr(X \leq x) = 1 - \exp(-x), \quad x \geq 0.$$

Then, for large n $X_{(n)} - \log(n) \approx$ Gumbel with location parameter 0 and scale 1. That is,

$$\Pr(X_{(n)} - \log(n) \leq x) \approx \exp[-\exp(-x)], \text{ for all real } x.$$

In contrast, the sample minimum is *exactly* Exponential with mean $1/n$ without any location shift. That is,

$$\Pr(X_{(1)} \leq x) = 1 - \exp(-nx), \quad x \geq 0.$$

Other Upper Extremes

$$\Pr(X_{(n-k)} - \log(n) \leq x) \approx \exp[-\exp(-x)] \sum_{j=0}^{k} \frac{\exp(-jx)}{j!}.$$

Other Lower Extremes

$$\Pr(nX_{(k)} \leq x) \approx 1 - \exp(-x) \sum_{j=0}^{k} \frac{(x)^j}{j!}, \quad x \geq 0.$$

4 Sum of top order statistics & sample CTE

Suppose we are interested in estimating the CTE $\mu_p = E(X|X > x_p)$ where x_p is the upper pth quantile and p is not very small (say 5%) and let $k/n \approx p$. Define the sample CTE as

$$D_k = \frac{1}{k} \sum_{j=n-k+1}^{n} X_{(j)}.$$

Result 3. *For large n,*

$$\sqrt{k}(D_k - \mu_p) \approx N(0, \sigma_p^2 + p(x_p - \mu_p)^2).$$

where $\sigma_p^2 = Var(X|X > x_p)$.

- On the right, x_p, μ_p and σ_p^2 can be estimated from the sample using the sample quantile $X_{(n-k)}$, D_k and S_D^2, the sample variance of the top k order statistics.
- The CTE and Mean Excess Function are related.

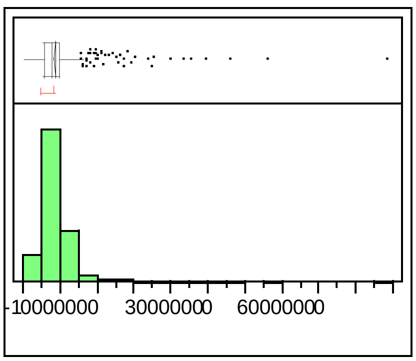
$$E(X|X > c) = c + m(c).$$

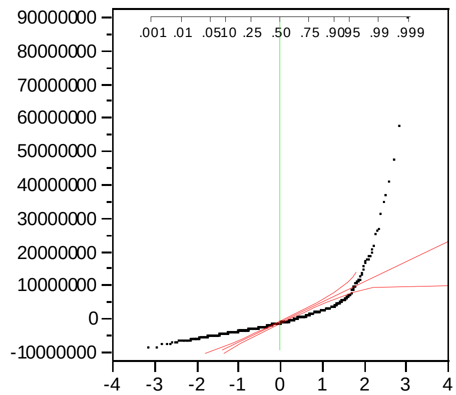
5 Data Examples and Model Fitting

Capital Requirements Data

Data Source: Steve Craighead, Nationwide Insurance, Columbus

Values are negatives of the cap-requirements





Normal Quantile Plot

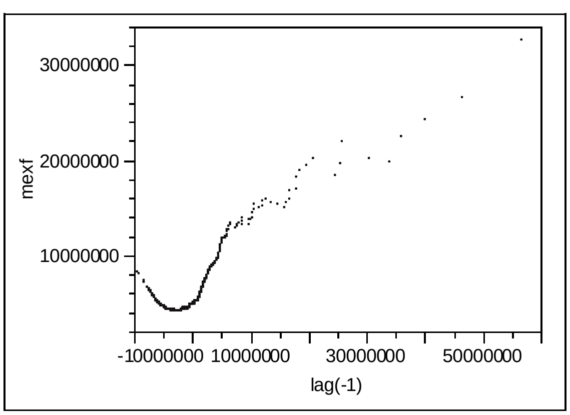
Quantiles

100.0%	maximum	89490993
99.5%		36103098
97.5%		11807434
90.0%		2782728.4
75.0%	quartile	207245.65
50.0%	median	-2.094e+6
25.0%	quartile	-3.8605e6
10.0%		-5.1367e6
2.5%		-6.5963e6
0.5%		-8.0968e6
0.0%	minimum	-9.133e+6

Moments

Mean	-1.0354e6
Std Dev	5983789.7
Std Err Mean	189318.73
upper 95% Mean	-663845.7
lower 95% Mean	-1.4069e6
N	999

Bivariate Fit of mexf By (-capreq)_lag1

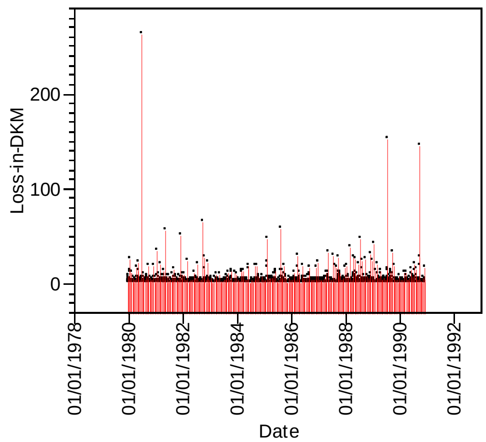


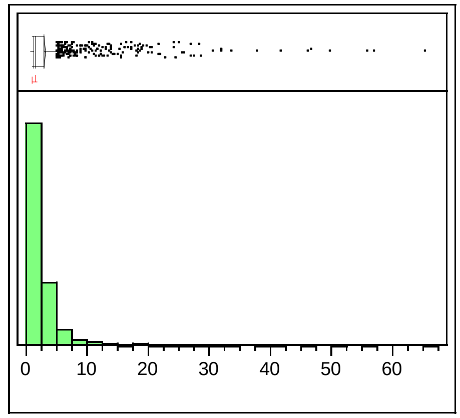
Danish Fire Insurance Claims Data

Data Source: http://www.math.ethz.ch/~mcneil/

(Alexander McNeil – Swiss Federal Institute of Technology)

Original Source: Rytgaard (ASTIN Bulletin)





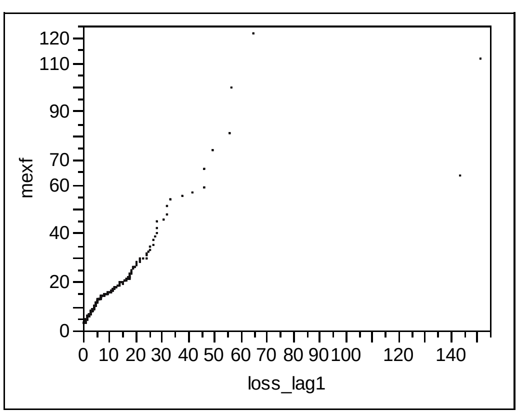
Quantiles (Note: Top 3 values were excl.)

100.0%	maximum	65.707
99.5%		32.402
97.5%		15.921
90.0%		5.506
75.0%	quartile	2.964
50.0%	median	1.775
25.0%	quartile	1.321
10.0%		1.113
2.5%		1.020
0.5%		1.000
0.0%	minimum	1.000

Moments (Top 3 excl.)

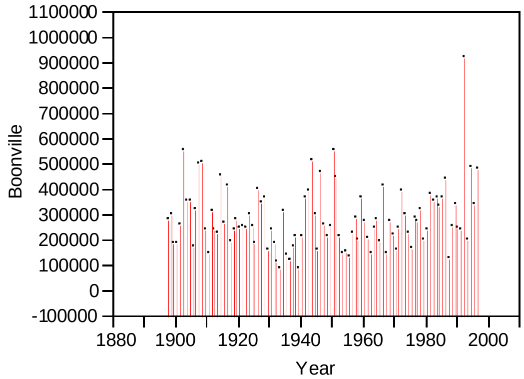
Mean	3.1308527
Std Dev	4.6580174
Std Err Mean	0.1001319
upper 95% Mean	3.3272175
lower 95% Mean	2.9344879
N	2164

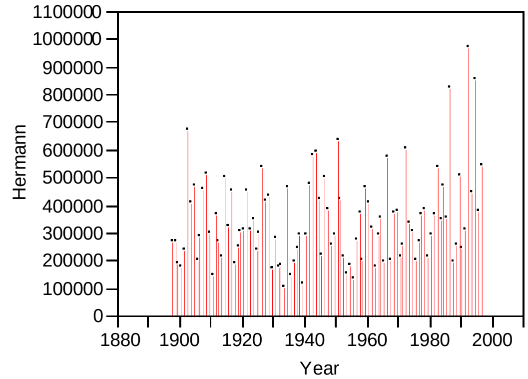
Bivariate Fit of mexf By loss_lag1

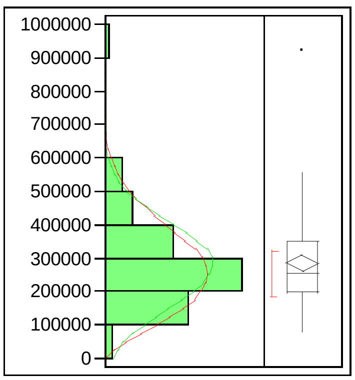


Missouri River Annual Maximum Flow Data for the Towns of Boonville and Hermann

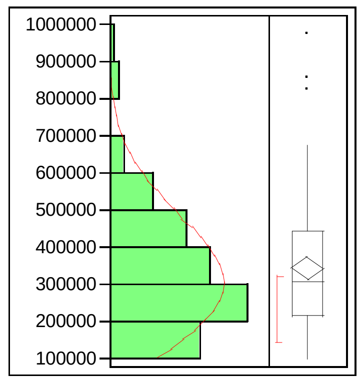
Data Source: US Army Corp of Engineers



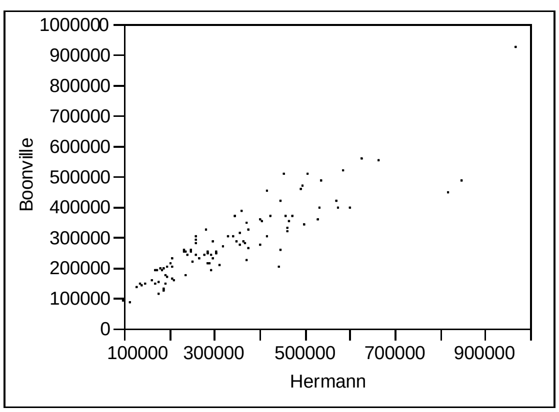




Extreme-Value and Normal fits for Boonville



Extreme Value Fit for Hermann

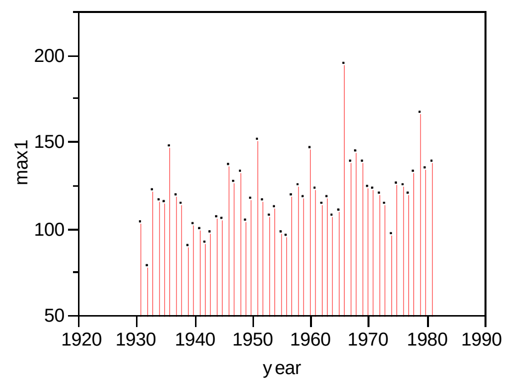


Pearson Correlation

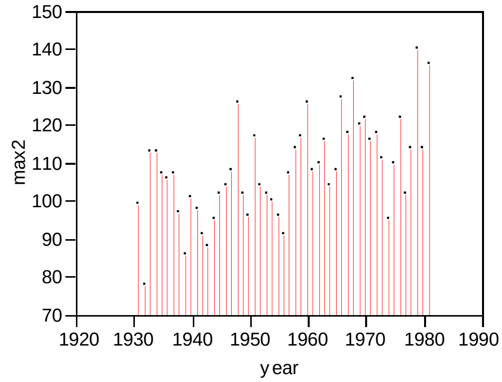
	Boonville	Hermann
Boonville	1.0000	0.8936
Hermann	0.8936	1.0000

Venice Annual Extreme Sea level Readings (1931-1981)

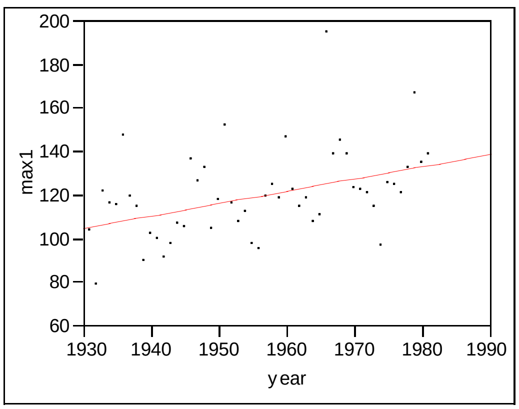
Data Source: Reiss and Thomas (1997); XTREMES software.



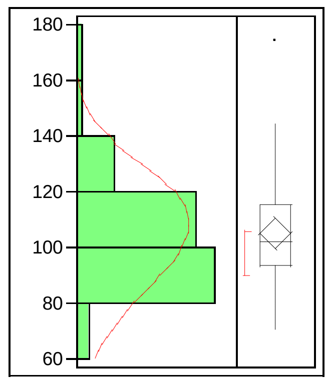
Highest Annual level



Second Best Annual level



max1 = -990.0706 + 0.5673303 year



Extreme Value Fit for mod.max1

Pairwise Pearson Correlations

	max1	max2	max3	max4	max5	max6
max1	1.0000	0.8103	0.7746	0.6724	0.6741	0.6685
max2	0.8103	1.0000	0.9455	0.9145	0.8872	0.8626
max3	0.7746	0.9455	1.0000	0.9524	0.9208	0.8901
max4	0.6724	0.9145	0.9524	1.0000	0.9662	0.9253
max5	0.6741	0.8872	0.9208	0.9662	1.0000	0.9682
max6	0.6685	0.8626	0.8901	0.9253	0.9682	1.0000

Hour 2: EXTREME VALUES - BASIC MODELS

1 Sample Maximum

For an arbitrary parent distribution, $X_{n:n}$, the largest in a random sample of n from a population with cdf $F(x)$, even after suitable standardization, may not possess a limiting distribution.

Result 3. *If $F(x)$ is such that $(X_{(n)} - a_n)/b_n$ has a limit distribution for large n, then the limiting cdf must be one of just three types:*

$$\begin{aligned}
\textit{(Fréchet)}\quad G_1(x;\alpha) &= 0 \qquad\qquad\qquad x \leq 0, \alpha > 0, \\
&= \exp(-x^{-\alpha}) \qquad x > 0; \\[1em]
\textit{(Weibull)}\quad G_2(x;\alpha) &= \exp\left[-(-x)^{\alpha}\right] \;\; x \leq 0, \alpha > 0, \\
&= 1 \qquad\qquad\qquad x > 0; \\[1em]
\textit{(Gumbel)}\quad G_3(x) &= \exp\left(-e^{-x}\right) \;\; -\infty < x < \infty.
\end{aligned} \tag{1}$$

Notes

- For a given parent cdf F there can be only one type of G.
- Norming constants a_n and b_n need to be estimated from the data and formulas depend on the type of G and parent cdf F. Convenient choices are given below in Result 4.
- For the sample minimum, the limit distributions are similar and have one-to-one relationship with the family in (1).
- necessary and sufficient conditions for each of the 3 possibilities are known and are very technical. They depend on the right-tail thickness of F.
- For many insurance applications, the right-tail is Pareto like resulting in limiting Fréchet distribution.
- Simpler sufficient conditions are available for continuous distributions.
- There are parent cdfs for which no (nondegenerate) limit distribution exists for the max.

Formulas for Norming Constants For the Maximum

Result 4. *The location shift* a_n *and scale shift* b_n *are*

(i) $a_n = 0$ *and* $b_n = x_{1/n}$ *if* $G = G_1$

(ii)$a_n = F^{-1}(1)$ *and* $b_n = F^{-1}(1) - x_{1/n}$ *if* $G = G_2$

(iii) $a_n = x_{1/n}$ *and* $b_n = E(X - a_n | X > a_n) \approx [nf(a_n)]^{-1}$ *if* $G = G_3$

Domains For Common Distributions

Fréchet: F has infinite upper limit.

Cauchy; Pareto; Burr; Stable with index < 2; Loggamma.

Weibull: F Bounded Above.

Uniform; Power law (upper end); Beta.

Gumbel: Bounded or unbounded F.

Exponential; Weibull; Gamma; Normal; Benktander-type I and type II.

An Example: Pareto Distribution

Suppose the cdf is

$$F(x) = 1 - x^{-\alpha}, \quad x \geq 1, \alpha > 0.$$

Then

$$\Pr(X_{(n)} \leq n^{1/\alpha} x) \approx \exp\{-x^{-\alpha}\}$$

for all $x > 0$.

Note: The α of the limit distribution of the maximum is the α of the parent distribution representing the tail thickness.

Another Example

Suppose the cdf is

$$F(x) = 1 - (\log(x))^{-1}, \quad x \geq e.$$

Then you cannot normalize $X_{(n)}$ so that we get a nondegenerate limit distribution.

2 Generalized Extreme-Value Distribution

The G_k in (1) are members of the family of *generalized extreme-value* (GEV) distributions

$$G_\xi(y;\mu,\sigma) = \begin{cases} \exp\left[-(1+\xi\frac{y-\mu}{\sigma})^{-\frac{1}{\xi}}\right], & 1+\xi\frac{y-\mu}{\sigma} > 0, \\ & \xi \neq 0, \\ \exp\left(-e^{-[(y-\mu)/\sigma]}\right), & -\infty < y < \infty, \\ & \xi = 0. \end{cases} \quad (2)$$

ξ is the *shape*, μ is the location and σ is the scale parameter.

Connections

$G_1(y;\alpha)$ when $\xi > 0$, $\sigma = 1$, $\mu\xi = 1$ and $\alpha = 1/\xi$,

$G_2(y;\alpha)$ when $\xi < 0$, $\sigma = 1$, and $\alpha = -1/\xi$

$G_3(y)$ when $\xi = 0$, $\sigma = 1$ and $\mu = 0$.

Moments of GEV

Take $\mu = 0$, $\sigma = 1$.

When $\xi \neq 0$,

$$\text{Mean} = \frac{1}{\xi}\left[\Gamma(1-\xi) - 1\right], \xi < 1;$$

$$\text{Variance} = \frac{1}{\xi^2}\left[\Gamma(1-2\xi) - \Gamma^2(1-\xi)\right], \xi < 1/2;$$

$$\text{Mode} = \frac{1}{\xi}\left[(1+\xi)^{-\xi} - 1\right].$$

When $\xi = 0$, we have the Gumbel distribution and

$$\text{Mean} = \gamma = 0.5772.. = \text{Euler's constant};$$

$$\text{Variance} = \pi^2/6; \quad \text{Mode} = 0.$$

Quantile Function of GEV

Take $\mu = 0$, $\sigma = 1$.

When $\xi \neq 0$,

$$G_\xi^{-1}(q) = \frac{1}{\xi}\left\{[-\log(q)]^{-\xi} - 1\right\}$$

For the Gumbel cdf, $G_0^{-1}(q) = -\log[-\log(q)]$.

3 Limit Distributions of the top Extremes

kth Order Statistic

Result 5. *If $F(x)$ is such that $(X_{(n)} - a_n)/b_n$ has a limiting cdf G, then, for a fixed k, the limiting cdf of $(X_{(n-k+1)} - a_n)/b_n$ is of the form:*

$$G^{(k)}(x) = G(x) \sum_{j=0}^{k-1} \frac{[-\log G(x)]^j}{j!}.$$

Notes

- The same G and the same set of norming constants work.
- The form of $G^{(k)}(x)$ has a Poisson partial sum form.

Pareto Example Contd.

Suppose the cdf is

$$F(x) = 1 - x^{-\alpha}, \quad x \geq 1, \alpha > 0.$$

Then

$$\Pr(X_{(n-k+1)} \leq n^{1/\alpha} x) \approx \exp\{-x^{-\alpha}\} \sum_{j=0}^{k-1} \frac{x^{-j\alpha}}{j!}$$

for all $x > 0$.

Limiting Joint Distribution of the top kth Order Statistics

Result 6. *If $(X_{(n)} - a_n)/b_n$ has limiting cdf G and g is the pdf of G then the k-dimensional vector*

$$\left(\frac{X_{(n)} - a_n}{b_n}, \ldots, \frac{X_{(n-k+1)} - a_n}{b_n}\right) \tag{3}$$

has a limit distribution with joint pdf

$$g_{(1,\ldots,k)}(w_1, \ldots, w_k) = G(w_k) \prod_{i=1}^{k} \frac{g(w_i)}{G(w_i)}, \quad w_1 > \cdots > w_k. \tag{4}$$

Pareto Example Contd.

Suppose the cdf is

$$F(x) = 1 - x^{-\alpha}, \quad x \geq 1, \alpha > 0.$$

Then

$$\left(X_{(n)}/n^{1/\alpha}, \ldots, X_{(n-k+1)}/n^{1/\alpha}\right)$$

has limiting joint pdf

$$g_{(1,\ldots,k)}(w_1, \ldots, w_k) = \exp\{-w_k^{-\alpha}\} \prod_{i=1}^{k} \frac{\alpha}{w_i^{\alpha+1}},$$

for all $w_1 > \cdots > w_k > 0$.

Notes

- Result 6 is useful for making inference when only the top k observations are available.
- The joint density in (4) will involve the location, scale and shape parameters associated with the Generalized Extreme Value Distribution of interest.
- Methods of Maximum Likelihood or Moments for the Estimation of the parameters will involve this joint density.

Discrete Parent Populations

While the results hold, no familiar discrete distribution is in the domain of maximal attraction (or minimal attraction).

4 Limiting Distribution of the Sample Minimum

Result similar to Result 3 holds for the asymptotic distribution of the standardized minimum, $(X_{(1)} - a_n^*)/b_n^*$.

Result 7. *The three possible limiting cdfs for the minimum are as follows:*

$$
\begin{aligned}
G_1^*(x;\alpha) &= 1 - \exp\left[-(-x)^{-\alpha}\right] \quad x \le 0, \alpha > 0, \\
&= 1 \qquad\qquad\qquad x > 0; \\
G_2^*(x;\alpha) &= 0 \qquad\qquad\qquad x \le 0, \alpha > 0, \\
&= 1 - \exp(-x^{\alpha}) \qquad x > 0; \\
G_3^*(x) &= 1 - \exp\left(-e^{x}\right) \qquad -\infty < x < \infty.
\end{aligned} \tag{5}
$$

Clearly, $G_i^*(x) = 1 - G_i(-x)$, $i = 1, 2, 3$. where G_i is given by (1).

Notes

- Here $G_2^*(x;\alpha)$ is the real Weibull distribution!
- Norming constants can be obtained using Result 4 with appropriate modifications.
- Instead of dealing with the sample minimum from X one can see it as the negative of the maximum from

$-X$. Thus need to know only the upper extremes well.

Pareto Example Contd.

Suppose the cdf is

$$F(x) = 1 - x^{-\alpha}, \quad x \geq 1, \alpha > 0.$$

Then

$$P(X_{(1)} > x) = [1 - F(x)]^n = x^{-n\alpha}, x \geq 1.$$

So

$$P\{n(X_{(1)} - 1) > y\} = [1 + \frac{y}{n}]^{-n\alpha}, y \geq 0.$$

This approaches $\exp(-y^\alpha)$ as n increases. So the limiting cdf is $G_2^*(x; \alpha)$.

Joint Limiting Distribution of Max and Min.

Any "lower" extreme $X_{(r)}$ is asymptotically independent of any "upper" extreme $X_{(n+1-k)}$. This result is very useful for finding the limiting distributions of statistics such as the range and the midrange.

5 Distributions of Exceedances and the Generalized Pareto Distribution (GPD)

When F is unbounded to the right, there is an interesting connection between the limit distribution for $X_{(n)}$ for large n and the limit behavior as $t \to \infty$ of standardized *excess life* $(X - a(t))/b(t)$, conditioned on the event $\{X > t\}$.

Balkema and de Haan (1974), and Pickands (1975) characterize the family of limiting distributions for the excess life.

Result 8. *For a nondegenerate cdf $H(x)$ if*

$$\lim_{t \to \infty} \Pr\left\{ \frac{X - a(t)}{b(t)} > x \middle| X > t \right\} = 1 - H(x)$$

then $H(x)$ is a Pareto-type cdf having the form

$$H_\xi(x) = \begin{cases} 1 - (1 + \xi x)^{-\frac{1}{\xi}}, & x \geq 0, \quad \text{if } \xi > 0 \\ \\ 1 - e^{-x}, & x \geq 0, \quad \text{if } \xi = 0 \end{cases} \tag{6}$$

if and only if

$$\Pr\left(\frac{X_{(n)} - a_n}{b_n} \leq x \right) \to G_\xi(x)$$

where G_ξ is the GEV cdf given in (2) with the same shape parameter $\xi \geq 0$.

Moments of GPD

When $\xi \neq 0$,

$$\text{Mean} = \frac{1}{1-\xi}, \xi < 1;$$

$$\text{Variance} = \left[(1-2\xi)(1-\xi)^2\right]^{-1}, \xi < 1/2;$$

$$\text{Mode} = \frac{1}{\xi}\left[(1+\xi)^{-\xi} - 1\right].$$

$$\text{Mean Excess Function} = E(X - t | X > t) = \frac{1+\xi t}{1-\xi}, \xi < 1.$$

When $\xi = 0$, we have the standard exponential distribution and

Mean = Standard Deviation = 1 = Mean Excess Function.

Quantile Function of GPD

When $\xi \neq 0$,

$$H_\xi^{-1}(q) = \frac{1}{\xi}\left\{(1-q)^{-\xi} - 1\right\}.$$

For the Exponential cdf, $H_0^{-1}(q) = -\log(1-q)$.

Note

In addition to the shape parameter ξ the GPD has one more parameter, namely the scale parameter σ.

Hour 3: INFERENCE FOR EXTREME VALUE MODELS

1 Generalized Extreme Value Distribution

Let Y be a random variable having a *generalized extreme-value* (GEV) distribution with shape parameter ξ, location parameter μ and scale parameter σ. The cdf is (see (2))

$$G_\xi(y;\mu,\sigma) = \begin{cases} \exp\left[-(1+\xi\frac{y-\mu}{\sigma})^{-\frac{1}{\xi}}\right], & 1+\xi\frac{y-\mu}{\sigma}>0, \\ & \xi\neq 0, \\ \exp\left(-e^{-[(y-\mu)/\sigma]}\right), & -\infty<y<\infty, \\ & \xi=0. \end{cases}$$

The pdf is

$$g_\xi(y;\mu,\sigma) = \begin{cases} \frac{1}{\sigma}G_\xi(y;\mu,\sigma)(1+\xi\frac{y-\mu}{\sigma})^{-\frac{1}{\xi}}, & \\ \quad \xi(y-\mu)>-\sigma, & \xi\neq 0, \\ \frac{1}{\sigma}\exp\left(-e^{-[(y-\mu)/\sigma]}\right)e^{-[(y-\mu)/\sigma]}, & \\ \quad -\infty<y<\infty, & \xi=0. \end{cases} \tag{7}$$

1.1 Maximum Likelihood Method

- For a given data $(Y_1, \ldots, Y_m)$ whose joint likelihood L (i.e., probability density function) is known but for an unknown parameter θ, we choose the parameter value such that the likelihood L is maximized.

- It is generally chosen by looking at the first derivatives with respect to θ (if θ represents a multiparameter) of the logarithm of the likelihood ($\log L$) and equating to 0 and solving for θ. That is,

$$\frac{\partial \log L}{\partial \theta} = 0$$

  is solved and the solutions are checked to see which one corresponds to the largest possible value for the likelihood L.

- This general recipe produces estimates called Maximum Likelihood Estimates (MLE) (say $\hat{\theta}$) that have good large-sample properties under "regularity conditions".

- One condition is that the range of the data is free of the unknown parameter we are trying to estimate (GEV with $\xi \neq 0$ is a "nonregular situation".)

- In particular, if m is the sample size and m is large, generally,

$$\sqrt{m}(\hat{\theta} - \theta) \approx N(0, \sigma_\theta^2)$$

where σ_θ^2 is related to the average Fisher information in the sample.

- This variance σ_θ^2 can be estimated by the reciprocal of

$$-\frac{1}{m}\left.\frac{\partial^2 \log L}{\partial \theta^2}\right|_{\theta=\hat{\theta}}$$

Example. If we have a random sample of size m from an exponential distribution with pdf

$$f(y;\theta) = \frac{1}{\theta}\exp\{-y/\theta\}, y \geq 0,$$

the log-likelihood is given by

$$\log L(\theta) = -m\log(\theta) - \frac{1}{\theta}\sum_{i=1}^{m} y_i.$$

Differentiating this with respect to θ and equating to 0 we get only one solution namely,

$$\hat{\theta} = \overline{y},$$

the sample mean. Var$(Y) = \theta^2$ and this is the reciprocal of the Fisher information per observation. We know (central limit theorem) that

$$\sqrt{m}(\hat{\theta} - \theta) \approx N(0, \theta^2).$$

2 Blocked Maxima Approach

Suppose we have a LARGE sample of size n from a distribution with cdf F whose sample maximum is in the *domain of maximal attraction*. We are interested in the large percentiles and upper tails of F.
Suppose we can divide the total sample into m nonoverlapping blocks each of size r where r itself is big (say over 100). Note that $n = m \cdot r$. Let Y_1 be the largest in the first block,..., Y_m be the largest of the mth block. Then $Y_1, \ldots, Y_m$ behaves like a random sample from the GEV distribution with pdf $g(y)$ given by (7).

Note: If we know that F is in the domain of attraction of Gumbel, then $\xi = 0$ and hence we will have one less parameter and we will satisfy the regularity conditions for nice properties for the MLE to hold. We will consider it first.

Gumbel Distribution

From (7) we can write the joint pdf of $Y_1, \ldots, Y_m$. The log-likelihood, $\log L(\mu, \sigma)$ will be

$$-m \log \sigma - \sum_{i=1}^{m} \left(\frac{y_i - \mu}{\sigma} \right) - \sum_{i=1}^{m} \exp \left\{ - \left(\frac{y_i - \mu}{\sigma} \right) \right\}.$$

We differentiate this with respect to μ and σ and obtain two equations. Solving iteratively using numerical techniques, we obtain the MLE $(\hat{\mu}, \hat{\sigma})$. If m is large, the vector is approximately normally distributed with mean (μ, σ) and covariance matrix

$$\frac{1}{m}\frac{6\sigma^2}{\pi^2}\begin{pmatrix} \frac{\pi^2}{6} + (1-\gamma)^2 & (1-\gamma) \\ (1-\gamma) & 1 \end{pmatrix}$$

where γ is the Euler's constant 0.5772... This can be used to provide confidence intervals for the parameter estimates.

GEV Distribution

The log-likelihood for the GEV parameters is given by

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

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.

Smith (1985)

- When $\xi > -0.5$, MLEs have the usual properties.
- $-1 < \xi \leq -0.5$: MLEs are generally obtainable but do not have the standard asymptotic properties.
- $\xi < -1$: MLEs are unlikely to be obtainable from numerical techniques- likelihood is too bumpy.
- If $\xi \leq -0.5$, we have a short bounded upper tail for F - not encountered; Fréchet has $\xi > 0$.
- $(\hat{\xi}, \hat{\mu}, \hat{\sigma})$ is asymptotically normal with mean (ξ, μ, σ) and the covariance matrix can be approximated using the inverse of the observed Fisher information matrix.

Large Quantile Estimation

Let z_p be the upper pth quantile of the GEV. Then its estimate is given by

$$\hat{z_p} = \begin{cases} \hat{\mu} - \frac{\hat{\sigma}}{\hat{\xi}}(1 - y_p^{\hat{\xi}}), & \text{if } \hat{\xi} \neq 0 \\ \\ \hat{\mu} - \hat{\sigma}\log(y_p), & \text{if } \xi = 0 \ \ \text{(Gumbel)}, \end{cases}$$

where $y_p = -\log(1-p)$ is the upper pth quantile of the standard exponential population. Confidence interval for this percentile estimate can be computed using the covariance matrix of the MLEs of the individual parameters and linear approximations.

What we need are the estimates of the large quantiles of F. Suppose

$F(x_{p_0}) = 1 - p_0$ or $F^r(x_{p_0}) = \Pr(X_{(r)} \le x_{p_0}) = (1-p_0)^r$.

Then with $p = 1 - (1 - p_0)^r$ we have

$$\hat{x}_{p_0} = \hat{z}_p$$

given above. For example, suppose we need to estimate 99.5th ($= 1 - p_0$) percentile of the parent population and we have used blocks of size 100. Then $p = 1 - (0.995)^{100} \approx 0.39$ and we use this percentile estimate $\hat{z}_{.39}$.

Tail Probability Estimation

$$\begin{aligned}
\Pr(X > c) &= 1 - F(c) \\
&= 1 - [F^r(c)]^{1/r} \\
&\approx 1 - [G_\xi(c;\mu,\sigma)]^{1/r}.
\end{aligned}$$

Use the estimates of ξ, μ, and σ in the above formula to estimate the tail probability for the population.

3 Inference Using Top k Order Statistics

Suppose n is large, but not large enough to make blocks of subsamples of substantial size. One can think of using the top k order statistics. If k is small when compared to n, one can use the joint pdf given in Result 6, namely,

$$g(y_k) \prod_{i=1}^{k} \frac{g(y_i)}{G(y_i)}, \quad y_1 > \cdots > y_k$$

where g and G correspond to the GEV distribution. Thus, the likelihood will be

$$\exp\left\{-\left[1+\xi(\frac{y_i-\mu}{\sigma})\right]^{-\frac{1}{\xi}}\right\}\prod_{i=1}^{k}\frac{1}{\sigma}\left[1+\xi(\frac{y_i-\mu}{\sigma})\right]^{-\frac{1}{\xi}-1},$$

where $\xi(y_i-\mu) > -\sigma, i = 1,\ldots,k$. Using this likelihood, one can obtain the MLEs of ξ, μ, σ. These estimates can be used to estimate the high percentiles of F by using $p = 1-(1-p_0)^n$ and the $\hat{z}_p$ given earlier.

Remarks

- If we have data of top k order statistics from several blocks, say m, then we use the product of these likelihoods and determine the MLEs using the combined likelihood.

- How large k can we take? A tricky question. Larger k means decreased standard error of the estimators, but increased bias as we move away from G towards F. Typically one increases k and looks at a plot of the estimates of the parameters; when they stabilize, one stops.

Quick Estimators of ξ

(a) Hill's (1975) Estimator

- Assume $\xi > 0$ (Fréchet type). This means upper limit of F is infinite.

$$\hat{\xi}_H = \frac{1}{k} \sum_{i=n-k+1}^{n} \log(X_{(i)}) - \log(X_{(n-k)}).$$

- This is the mean excess function for the $\log(X)$ values! $\log(X)$ values are defined for $X > 0$ and this is the case for the upper extremes from large samples.
- The Hill estimator is also asymptotically normal.
- This is the MLE of ξ assuming $\mu = 0$ and $\sigma = 1$. The MLE of $\alpha = 1/\hat{\xi}$.
- Choice of k is determined by the examination of the plot of $(k, \hat{\xi}_H(k))$.

(b) Pickands' Estimator

- Works for any ξ real.

$$\hat{\xi}_P = \frac{1}{\log 2} \log \left\{ \frac{X_{(n-k+1)} - X_{(n-2k+1)}}{X_{(n-2k+1)} - X_{(n-4k+1)}} \right\}.$$

- Here k is large but k/n is small.
- Choice of k is determined by the examination of the plot of $(k, \hat{\xi}_P(k))$.

Remark

There are other estimators that are refinements of Hill's and Pickands'. They have explicit forms and hence computation is easy, but there are no clear preferences.

4 Generalized Pareto Distribution – Estimates Based on Exceedances

Let t be the threshold and we use only the X_i that exceed t. Then, for t large, the exceedances $Y^* = X - t$ have cdfs that can be approximated by a Pareto-type cdf having the form

$$H_\xi(y;\sigma) = \begin{cases} 1-(1+\xi\frac{y}{\sigma})^{-\frac{1}{\xi}}, & y>0; \xi y > -\sigma, \quad \text{if} \quad \xi \neq 0 \\ \\ 1-e^{-y/\sigma}, & y \geq 0, \quad \text{if} \quad \xi = 0 \end{cases} \tag{8}$$

If there are k exceedances above t, say $y_1^*, \ldots, y_k^*$, the MLEs of ξ and σ are obtained by using the log-likelihood

$$-k\log(\sigma) - (1+\frac{1}{\xi})\sum_{i=1}^{k}\log(1+\xi\frac{y_i^*}{k})$$

and for the case $\xi = 0$ we have the MLE based on the exponential distribution.

- The variance covariance matrix of $(\hat{\xi}, \hat{\sigma})$ is approximated by using the inverse of the sample Fisher information matrix.

- Let z_p be the upper pth percentile of $H_\xi(y;\sigma)$. Then an estimate of the upper pth percentile of F is given by

$$\hat{x_p} = t + \frac{\hat{\sigma}}{\hat{\xi}} \left\{ \left(\frac{\Pr\{X > t\}}{p} \right)^{\hat{\xi}} - 1 \right\}$$

where we assume $\Pr\{X > t\} > p$ and $\Pr\{X > t\}$ is estimated by the sample proportion exceeding t.

- Choice of the threshold t - how close the exceedances are modeled by the Generalized Pareto distribution.
- Such data are called Peaks Over Threshold (POT) data.

5 General Diagnostic Plots

5.1 QQ plots

Plots of Sample Order Statistics against the Quantiles of the fitted/hypothesized distribution. Generally $\big(X_{(i)}, F^{-1}(\frac{i}{n+1})\big)$ for $i = 1, \ldots, n$, is plotted. Sometimes $(i - \frac{1}{2})/n$ is used in place of $\frac{i}{n+1}$. Linear trend indicates good fit.

5.2 Mean Excess Function Plots

Plots of $(k, e_n(k))$, $k = 1, 2, ...$ that tell us how close the data agrees with the assumed distribution. Here $e_n(k)$ is the sample mean excess function

$$e_n(k) = \frac{1}{k} \sum_{i=n-k+1}^{n} X_{(i)} - X_{(n-k)} = D_k - X_{(n-k)}.$$

5.3 Plots for Choosing the top k Order Statistics

Plots of $(k, \hat{\theta}(k))$ where $\hat{\theta}(k)$ is the estimate of the parameter θ based on top k order statistics; it is accompanied by standard errors of the estimate. Used to examine the trade-off characteristics between bias and variance, and choose k for making inference.

6 General Remarks

1. The distribution of the maximum of a Poisson number of IID excesses over a high threshold is a GEV. (Crucial for stop-loss treaties.)
2. Relaxing Model Assumptions- Adjusting for Trend.
3. Stationarity with mild dependence structures.
4. Non-identical but independent, with no dominating distribution.
5. Linear processes (ARMA etc.)

7 Examples

Extreme Sea Levels in Venice Data

Largest annual sea levels from 1931-1981.

a) Maxima of blocks of $n = 365$ days.

b) Scatterplot and Adjusting for trend.

c) Fitting GEV.

d) Estimating large percentile for 2005.

Reference: Ferreiras article in Reiss and Thomas (1997); XTREMES software.

8 Computational/Software Resources

- XTREMES.
  A free-standing software that comes with Reiss and Thomas' book (academic edition- with limited data capacity). Professional edition is also available. Website:
  http://www.xtremes.math.uni-siegen.de/xtremes/

- Alexander McNeil's Website:
  http://www.math.ethz.ch/ mcneil/software.html
  Has free software attachment that works with S-PLUS.

- http://www.maths.bris.ac.uik/ masgc/ismev/summary.html
  Associated with Coles' book.

9 References

1. Arnold B. C., Balakrishnan N., and Nagaraja, H. N. (1992). *A First Course in Order Statistics.* John Wiley, New York.
   An elementary introduction to Order Statistics.

2. Balkema, A. A. and de Haan, L. (1974). Residual lifetime at great age. *Annals of Probability* **2**, 792-804.

3. Cebrián, A. C., Denuit, M., and Lambert, P. (2003). Generalized Pareto fit to the Society of Actuaries' Large Claims Database. *North American Actuarial Journal* **7**, No. 3, 18-36.
   Shows how the GPD model outperforms parametric counter parts for this data set.

4. Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values.* Springer Verlag, London.
   A truly enjoyable applied introduction to extremes; an excellent starting point. Has S-Plus codes and help to run them. Has two data examples from finance.

5. Craighead, S. (2004). Personal communication - Capital Requirements Data.

6. David H. A. and Nagaraja, H. N. (2003). *Order Statistics, Third Edition.* John Wiley, New York.
An encyclopedic reference - a dense treatment.

7. Embrechts P., Klüppelberg C., and Mikosch, T. (1997). *Modelling Extremal Events for Insurance and Finance.* Springer Verlag, Berlin.
A compendium of theory and applications - requires a strong mathematical background; has examples from Insurance and Finance.

8. Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution. *Annals of Statistics* **3**, 1163-1174.

9. McNeil, A. J. Webpage: http://www.math.ethz.ch/ mcneil/
An excellent resource for software, datasets, and papers.

10. McNeil, A. J. (1996). Estimating the tails of loss severity distributions using extreme value theory. (his webpage)

11. Nagaraja, H. N., Choudhary, P.K., and Matalas, N. (2002). Number of records in a bivariate sample with application to Missouri river flood data. *Methodol-*

ogy and Computing in Applied Probability **4**, 377-392.

12. Pickands, J. (1975). Statistical inference using extreme order statistics. *Annals of Statistics* **3**, 119-131.

13. Reiss R.-D. and Thomas, M. (1997). *Statistical Analysis of Extreme Values.* Birkhäuser Verlag, Basel, Switzerland.
    XTREMES software comes with the book.

14. SAS JMP software, Version 5.1. (2003). SAS Institute, Cary NC.

15. Smith, R. L. (1985). Maximum likelihood estimation in a class of non-regular cases. *Biometrika* **72**, 67-90.