Statistical Theory of Extremes
Homepage: http://www.gathacognition.com/book/gcb14
http://dx.doi.org/10.21523/gcb1

Part 2
Statistics for Univariate Extremes

Chapter 5

Statistical Analysis of the Gumbel Model

J. Tiago de Oliveira
Academia das Ciências de Lisboa (Lisbon Academy of Sciences), Lisbon, Portugal.

Abstract

It is assumed that the samples of maxima come from a Gumbel distribution. Gumbel characteristics are studied: moment generating and characteristic functions, mean value, variance, skewness, kurtosis, quantiles and mode. Return periods and return levels are considered. Statistical inference for location/scale parameters: Maximum Likelihood Estimators (MLE), Moments Estimators (ME), Best Linear Unbiased Estimators (BLUE), Best Linear Invariant Estimators (BLIE), confidence interval, tests of hypotheses, point and interval prediction, tolerance intervals and overpassing probability of some threshold. A worked example on maximal yearly precipitations is presented.

Published Online

23 June 2017

Keywords

Statistical inference;
MLE;
ME;
BLUE;
BLIE;
Test of hypotheses;
Tolerance intervals.

Editor(s)

J.C. Tiago de Oliveira

5.1 Introduction

In what follows we will assume that the samples of maxima or minima we are dealing with have a Gumbel distribution for maxima $\Lambda((x-\lambda)/\delta)$, sometimes denoted $\Lambda(x|\lambda,\delta)$, or for minima $1-\Lambda(-(x-\lambda)/\delta)$. We will deal only with maxima, minima being dealt with, as said by the relation $\min(x_i) = -\max(-x_i)$. Recall that the Gumbel distribution plays a central role in extreme value theory, being sometimes called *the* extreme value distribution; it is the limit of both Fréchet and Weibull distributions according

Originally published in 'Statistical Analysis of Extremes', 1997, 2016
http://dx.doi.org/10.21523/gcb1.1710

to $\theta \downarrow 0$ or $\theta \uparrow 0$ in the von Mises-Jenkinson formula and those distributions can be transformed to a Gumbel distribution by simple (but not always convenient) logarithmic transformations.

5.2 Characteristics of the Gumbel distribution

As the distribution $\Lambda((x-\lambda)/\delta)$ is not symmetrical, the natural location parameter need not be the mean; also the convenient scale parameter will not be the standard deviation. Define the *standardized* or *reduced* random variable as

$Z = X - \lambda/\delta$ or $X = \lambda + \delta Z$.

The graph of the probability density $\frac{1}{\delta} e^{-(x-\lambda)/\delta} \exp(-e^{(x-\lambda)/\delta})$ for $\lambda = 0$ and $\delta = 1$ is given in Figure 5.1.

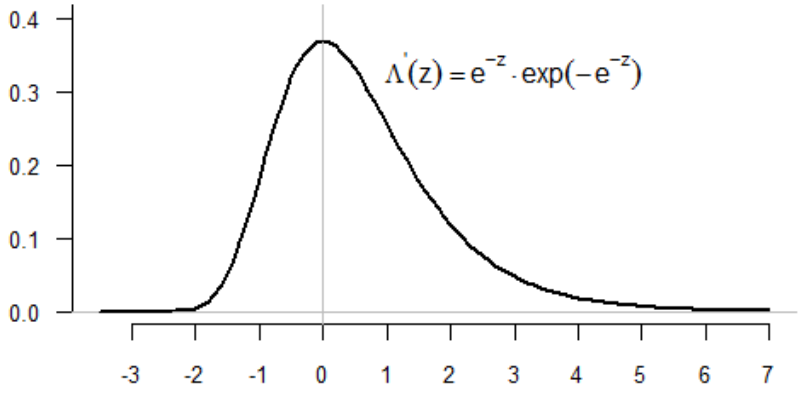

Figure 5.1 Reduced Gumbel density (for maxima)

The moment generating function of Z is $m_Z(\theta) = \Gamma(1-\theta)$ (for $\theta < 1$) and that of X is $m_X(\theta) = e^{\lambda\theta}\ m_Z(\delta\theta) = e^{\lambda\theta}\Gamma(1-\delta\theta)(\delta\theta < 1)$; the corresponding characteristic functions are $\Phi_Z(t) = \Gamma(1 - i\,t)$ and $\Phi_X(t) = e^{i\lambda t}\,\Gamma(1 - i\,\delta t)$.

All the moments exist as follows from the infinite differentiability of $\Gamma(1-it)$. Let us recall that when the mean value μ exists, the central moments (i.e., the moments about μ) have the expressions, if they exist,

$\mu_i = M((X-\mu)^i)$

with the variance $\sigma^2 = \mu_2$ and the frequently used linearly invariant measures of skewness and kurtosis,

$$\beta_1 = \frac{\mu_3}{\sigma^3}, \qquad \beta_2 = \mu_4/\sigma^4.$$

In our case we should consider the random variables Z and $X = \lambda + \delta Z$ with distribution functions $\Lambda(z)$ and $\Lambda((x-\lambda)/\delta)$, and the moments and coefficients should be noted $\mu_{Z,i}$ and $\mu_{X,i}$, the values of β_1 and β_2 being independent of λ and δ. We shall, for simplicity, drop the index Z always and we have

$\mu_i = \mu_{Z,i} = M(Z-\mu)^i$

$\mu_{X,i} = M(Z-\mu_X)^i = \delta^i \mu_i$

with $\mu_X = M(X) = \lambda + \delta M(Z) = \lambda + \delta\mu = \lambda + \delta\mu_Z$.

Note also that an analogous relation is valid for quantiles:

if $\Lambda(\chi_p) = \Lambda(\chi_{Z,p}) = p$ we have $\chi_{X,p} = \lambda + \delta\,\chi_{Z,p} = \lambda + \delta\,\chi_p$.

For the mode μ' , as for the median $\tilde{\mu}$ which is the .5-quantile, we have also

$\mu'_x = \lambda + \delta\,\mu'_Z = \lambda + \delta\mu'$

and $\tilde{\mu}_x = \lambda + \delta\,\tilde{\mu}_Z = \lambda + \delta\tilde{\mu}$

We now have the following values for the coefficients associated with Z, see Gumbel (1958):

$\mu_1 = \gamma = .57722\ldots =$ Euler's constant;

$\mu' = 0 = \chi_{1/e}$;

$\tilde{\mu} = -\log\log 2 = .36651\ldots$;

$\chi_p = -\log\log(p^{-1})$;

$\mu_2 = \sigma^2 = \pi^2/6 = 1.64493\ldots$;

$\sigma = \sqrt{\pi^2/6} = 1.28255\ldots$;

$\mu_3 = 2.40411\ldots$;

$\mu_4 = 3\pi^4/20 = 14.61136\ldots;$

$\beta_1 = 1.13955;$

$\beta_2 = 5.40000.$

The return period, for the overpassing the level (or threshold) a, is, as said before,

$T(a) = (1 - \Lambda((a-\lambda)/\delta))^{-1};$

on average $T(a)$ trials will be made until the maximum overpasses a, after the last overpassing. Note that the probability of overpassing a before or at the m^{th} trial (year, for instance, if we are thinking of dams) is $1-(1-p)^m$ where $p = 1 - \Lambda((a-\lambda)/\delta)$. Thus is can be written as $1-(1-\frac{1}{T(a)})^m$.

The probability that the overpassing will occur before or at the return period is

$1-(1-\frac{1}{T(a)})^{[T(a)]} = 1 - e^{-1} = .63212$ for large $T(a)$.

The maximum likelihood estimator of $T(a)$ is $T(a) = N_k/k$, where N_k is the time to the $k-$th occurrence of the overpassing. This is not very convenient and it is better to estimate $T((a-\lambda)/\delta)$ by $T((a-\hat{\lambda})/\hat{\delta})$, $(\hat{\lambda}, \hat{\delta})$ being the maximum likelihood estimators of (λ, δ) (to be dealt with in the next section), or with other estimators of (λ, δ).

A simple and useful conservative approximation for $T(a)$, given by Fuller (see Tiago de Oliveira (1972)) is

$$T(a) \approx \tilde{T}(a) = \exp\{\frac{a-\lambda}{\delta}\}$$

with the bounds $\tilde{T}(a) \leq T(a) \leq \tilde{T}(a)(1 - 1/2\,\tilde{T}(a))^{-1}$.

Usually, when $T(a) = 100$, engineers call a "the largest flood in 100 years" or "the 100-years flood"; the expressions in quotes are misnomers because the largest flood in 100 years is a random variable. In a sense, *on average*, we expect to have to wait 100 years to observe a maximum yearly flood greater than or equal to a.

From $\Lambda((a-\lambda)/\delta) = 1 - \frac{1}{T((a-\lambda)/\delta)}$ we see that "the 100-year flood" is the $.99$ quantile.

If a random variable has a Gumbel distribution with parameters $\lambda, \delta > 0$ the interval $[\lambda + b\,\delta, \lambda + a\,\delta]$ has the probability $\Lambda(a) - (b)$ and its length is $\delta(a - b)$. The shortest interval with a fixed probability β is given by the equations (obtained by the Lagrange multipliers method)

$$\Lambda(a) - \Lambda(b) = \beta = 1 - \alpha$$

$$\Lambda'(a) - \Lambda'(b) = 0$$

or $\Lambda(a) - \Lambda(b) = \beta = 1 - \alpha$

$$a + e^{-a} = b + e^{-b}.$$

A table for some relevant values is given below.

Table 5.1

α	β	b	A	Λ (b)	Λ (a)	a-b
1.00	0.00	0.000000	0.000000	1/e	1/e	0.000000
0.50	0.50	-0.651256	0.831143	0.146908	0.646908	1.482399
0.20	0.80	-1.125023	1.787968	0.045946	0.845946	2.912991
0.10	0.90	-1.369180	2.479152	0.019602	0.919602	3.848332
0.05	0.95	-1.561344	3.161461	0.008521	0.958521	4.722805
0.01	0.99	-1.893530	4.740459	0.001305	0.991305	6.633989

As could by the expected from the skewness of the distribution the shorted interval is not symmetrical; for instance the shorted intervals containing 90%, 95% and 99% of the population are roughly $[\lambda - 1.4\,\delta, \lambda + 2.5\,\delta]$, $[\lambda - 1.6\,\delta, \lambda + 3.2\,\delta]$ and $[\lambda - 1.9\,\delta, \lambda + 4.7\,\delta]$. The interval with $a = 2$, $b = 5$, i.e.,$[\lambda - 2\,\delta, \lambda + 5\,\delta]$, contains more than 99% of the population.

5.3 Point Estimation

The earliest methods of point estimation are the method of moments and the method of maximum likelihood. They were evidently used for Gumbel distribution. Thus these procedures will be given first, followed by some other treatments involving best linear unbiased estimators and best linear invariant estimators.

5.3.1. Maximum Likelihood Estimators (MLE)

The logarithm of the likelihood of a i.i.d. sample $(x_1, x_2, \dots, x_n)$ of n observations with the distribution function $\Lambda((x-\lambda)/\delta)$ is

$$\log L\,(\lambda, \delta | x_i) = -n\,\log\delta - \sum_1^n e^{-(x_i-\lambda)/\delta} - \sum_1^n (x_i-\lambda)/\delta.$$

Setting the partial derivatives of $\log L$ with respect to λ and δ equal to zero, we get, with easy algebra, the maximum likelihood equations:

$$\hat{\delta} = \bar{x} - \sum_1^n x_i\, e^{-x_i/\hat{\delta}} \Big/ \sum_1^n e^{-x_i/\hat{\delta}}$$

$$\hat{\lambda} = -\hat{\delta}\log\,(\sum_1^n e^{-x_i/\hat{\delta}}\,/\,n).$$

The second equation is very easy to deal with once the first one is solved. But this one, involving only $\hat{\delta}$, is intractable without computers. One could use an iterative procedure, starting with the moment estimators for λ and δ to follow. However, Kimball developed a simple procedure to get approximate solutions. This is described by him in the sections of Gumbel (1958) which he wrote, summarizing some of his papers. Panchang and Aggarwal (1962) in the "Poona Report" were among the first to use the iterative procedure indicated above, with the use of computers.

Maximum likelihood estimators are asymptotically efficient and unbiased with the variance-covariance matrix given by the inverse of

$$-n\begin{bmatrix} M(\frac{\partial^2 \log g}{\partial\,\lambda^2}) & M(\frac{\partial^2 \log g}{\partial\lambda\;\partial\delta}) \\ M(\frac{\partial^2 \log g}{\partial\lambda\;\partial\delta}) & M(\frac{\partial^2 \log g}{\partial\,\delta^2}) \end{bmatrix}$$

with $g = g(x|\lambda,\delta) = \frac{1}{\delta}\,\Lambda'(\frac{x-\lambda}{\delta})$, the density of Gumbel distribution with parameters $(\lambda\,,\delta)$. It is known that:

The estimators $(\hat{\lambda}\,,\hat{\delta})$ *are asymptotically binormal with mean values* $(\lambda\,,\delta)$, *the asymptotic variance-covariance matrix is*

$$\frac{1}{n}\begin{bmatrix} (1+\frac{6(1-\gamma)^2}{\pi^2})\,\delta^2 & \frac{6(1-\gamma)}{\pi^2}\delta^2 \\ \frac{6(1-\gamma)}{\pi^2}\,\delta^2 & \frac{6}{\pi^2}\delta^2 \end{bmatrix}$$

and the asymptotic correlation coefficient is

$$\rho(\hat{\lambda},\hat{\delta}) = (1+\frac{\pi^2}{6(1-\gamma)^2})^{-1/2} = .31307.$$

5.3.2. Moments Estimators (ME)

From the relations

$$M(X) = \lambda + \delta\gamma$$

$$V(X) = \frac{\pi^2}{6}\,\delta^2$$

we get as estimators by the method of moments

$$\delta^* = \frac{\sqrt{6}\ s}{\pi} = .77970\ s$$

$$\lambda^* = \bar{x} - \frac{\sqrt{6}\ \gamma}{\pi}\ s = \bar{x} - .45501\ s$$

where $\bar{x}$ and s^2 are the average $(\sum_1^n x_i/n)$ and the empirical variance $(\sum_1^n (x_i - \bar{x})^2/n)$.

(λ^*, δ^*) is asymptotically binormal but its efficiency is very low (48%); see Tiago de Oliveira (1963). The usefulness of this method, now that computers are used, is to give a simple first estimate as a seed for iteration to solve the maximum likelihood estimator equations.

It should be noted that Gumbel (1958) and Posner (1965) developed methods close to the method of moments (called "simplified" and "modified"), using the plotting positions and a fitted linear regression.

Mann (1968), by simulation and also using Lieblein's (1954) simulation, has shown that, for estimating quantiles, the Posner (1965) method gives poor estimators for either small or large samples, although the Gumbel (1958) method is relatively good for small samples. The BLUE methods - see below - are preferable.

5.3.3. Best Linear Unbiased Estimators (BLUE)

Lieblein (1954) was the first to recognize that as Gumbel plotting used order statistics, information was wasted if estimators did not use order statistics. He then considered estimating the linear combination

$$\theta = \theta\,(a\,,b) = a\,\lambda + b\,\delta,$$

with, as special cases, $\lambda = \theta\,(1{,}0)\,, \delta = \theta\,(0{,}1)$ and the percentile

$$\theta_p = \theta\big(1, \Lambda^{-1}(p)\big) = \lambda + \delta\,\Lambda^{-1}(p)$$

where $\Lambda^{-1}(p) = -\log(-\log p)$. He proposed to estimate θ by θ^*, a linear combination of order statistics in which the weights are determined to yield minimum variance unbiased estimators. More specifically, θ^* based on double censored data is given by

$$\theta^* = \sum_{i=r+1}^{N-s} w_i\, x_i{}'$$

where, obviously, $w_i = a\,a_i + b\,b_i$. Note that for $s = r = 0$ we have the complete sample.

The values of a_i and b_i for which $M(\theta^*) = \theta$ and $V(\theta^*)$is minimized, are functions of n, r, s and may be determined by the generalized least squares procedure of Loyd (1962), based on the generalized Gauss-Markov theorem. As the distribution of order statistics $Z_i^{'}$ is parameter-free, the mean values, variances, and covariances depend only on the standardized distribution function and the sample size. For example

$$M\big(Z_i^{'\,k}\big) = \frac{n\,!}{(i-1)!(n-i)!}\ \int_{-\infty}^{\infty} z^k \Lambda^{i-1}(z)\,[1-\Lambda(z)]^{n-1}\Lambda'(z)\,d\,z\,.$$

Those moments were given numerically by Lieblein (1953).

We start with the fact, observed earlier, because $X_i^{'}, = \lambda + \delta\,Z_i^{'}$, that

$$M\big(Z_i^{'\,k}\big) = \lambda + \delta\,\mu_i, C\big(X_i^{'}, X_j^{'}\big) = \delta^2\sigma'_{ij}$$

where $\mu'_i = M(Z_i^{'})$ and σ'_{ij} is the (i,j) the element of the variance-covariance matrix of $Z = (Z_1^{'}, Z_2^{'}, \dots, Z_n^{'})^T$. The generalized least squares theorem then states that if X is the vector of observations,

$Z = (Z_1', Z_2', \ldots, Z_n')^T$

having $M(X) = A\,\theta$ with $\theta = [\lambda, \delta]^T$, and $C(X) = C(\lambda + \delta z) = [\sigma'_{ij}]\,\delta^2 = V\,\delta^2$ where $A = \begin{bmatrix} 1 & 1 & & 1 \\ \mu'_1 & \mu'_2 & \ldots & \mu'_n \end{bmatrix}^T$, both A and V being known and $\det(V) \neq 0$, the least squares estimator of the vector $\theta = [\lambda, \delta]^T$ is the value that minimizes

$(X - A\,\theta)^T\ V^{-1}(X - A\,\theta)$.

The minimum solution is the one that satisfies the so-called normal equations,

$A^T V^{-1} A\,\theta^* = A^T V^{-1} X$ and so

$\theta^* = (A^T V^{-1} A)^{-1}\ A^T V^{-1} X$

and the variance-covariance matrix of θ^* being

$C(\theta^*) = (A^T V^{-1} A)^{-1}\,\delta^2$

The estimator θ^* is unbiased and the θ_i^* have minimum variance in the class of unbiased linear statistics.

Calculation of the mean values variances and covariances of the order statistics Z_i' and the weights a_i, b_i are tedious and require computers even for small values of n. They are now tabulated by Mann (1967) (Appendix C) to $N = 2(1)25, r = 0, s = 0(1)N - 2$, after some previous work by Lieblein and Zelen (1956).

5.3.4. Best Linear Invariant Estimators (BLIE)

The unbiased condition of the previous section can be relaxed by defining loss as the mean squared error divided by δ^2 and by considering weighted sums of $Z_i', i = 1,2,\ldots,r \leq n$, as estimators of (λ, δ) and of the quantiles with mean squared error, invariant under translations and multiplications. The estimators among those with smallest mean squared error were called the best linear invariant estimators. The BLUE of Lieblein and Zelen (1956) are also invariant. Mann (1968) has shown that those estimators of (λ, δ), and of the quantiles, are linear functions of the BLUE of λ and δ. The mean squared errors of the BLIE are uniformly less than those of the corresponding BLUE, Mann (1968). She says that when $r << n$, the ratios of the mean losses of the

BLUE relative to those of the corresponding ones of the BLIE are large; for $r = 2, n = 2(1)25$ the mean losses of the BLUE of λ and δ increase with n from $.659$ to 8.24 and $.712$ to $.980$, respectively. The corresponding ranges of losses when the BLIE are used are $.657$ to 4.49 and $.416$ to $.495$.

Furthermore, for $m = 2$, the losses of the BLIE of the quantiles and δ are evidently less than or equal to those of the corresponding MLE, which are also invariant. Tables of weights for obtaining the BLIE and the losses of the estimators are available in Mann (1967), (Appendix C).

Asymptotically efficient linear estimators of location and dispersion parameters were obtained and shown to be asymptotically normal by Chernoff *et al.* (1967). They are invariant and quite efficient with respect to the BLUE, even for fairly small n, Mann (1968). Johns and Lieberman (1966) used this procedure to determine the weights in the asymptotically efficient linear estimators of λ and δ, but there exist (or existed) computer difficulties in comparing the losses of those methods as moment calculations (mean values, variances and covariances) of order statistics involve large rounding errors, requiring multiple-precision computing techniques. Some comparison, based on limited computation, appears in Mann (1968).

5.3.5. Further methods of linear point estimation

Other approximating methods exist to generate the weights for linear estimators of (λ, δ).

We will enumerate some of them, giving sufficient references; in general they do not need, as does Johns and Lieberman's (1966) procedure, the tedious and error-prone computation of the first and second moments of order statistics.

A method of splitting the sample at random in blocks of $4,5$ or 6, computing the averages of the blocks, estimating the parameters (λ, δ) from those averages with convenient coefficients, and averaging the estimates, was developed by Lieblein and Zelen (1956) and also studied in Tiago de Oliveira (1972). It was shown there that the asymptotic efficiency for the estimator $\chi^*_{x,p} = \lambda^* + \chi_p \delta^*$ of the quantile χ_p has the expression

$$\text{eff}(\chi^*_{x,p}) = \frac{1 + .60793(.42278 + \chi_p)^2}{6(.19117 + .06274\,\chi_p + .13196\,\chi_{p2})}$$

having as $\chi_p \to \pm\infty$ the limit value 0.77, with $\text{eff}(\chi^*_{X,1/e}) = .97$ for the mode and $\text{eff}(\chi^*_{X,p}) > .77$ for $\chi_p > -.99802$ whose probability is $\Lambda(-.98) = .066$. See in Figure 5.2 the graph of the efficiency.

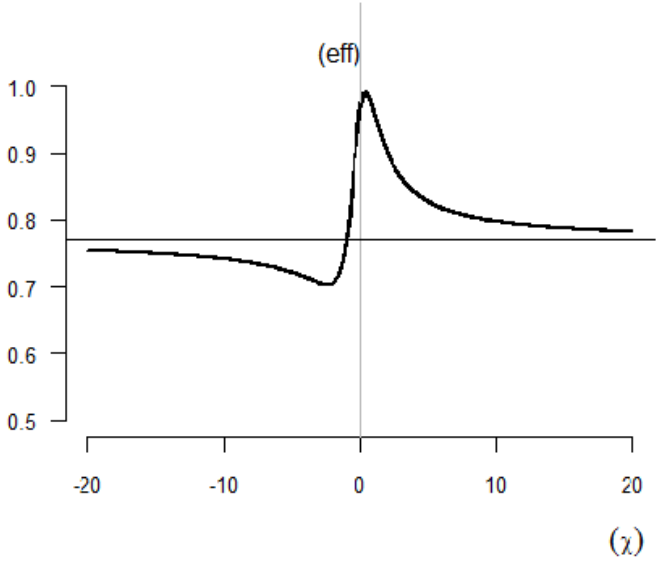

Figure 5.2 Asymptotic efficiency of Lieblein-Zelen method

Let us also refer to a method due to Downton (1966), developed for Gumbel minima distribution, now translated to the maxima data.

Denoting by $\bar{x} = \sum_1^n x_i/n$ and $\bar{w} = \frac{2}{n(n-1)} \sum_1^n i\, x_i'$, where x_i' are the usual order statistics, we know that and $\bar{x}$ and $\bar{w}$ have mean values $\lambda + \gamma\,\delta$ and $\lambda + (\gamma + \frac{n-1}{n+1}\log 2)\,\delta$. The unbiased estimators

$$\lambda^* = \bar{x} - \gamma\,\delta^*$$

$$\delta^* = \frac{n+1}{(n-1)\log 2}(\bar{w} - \bar{x})$$

(note that $\bar{w} > \bar{x}$) have the asymptotic efficiency for the estimation of the p-quantile

$$\text{eff}(\chi_{X,p}) = \frac{1 + .60793(.42278 + \chi_p)^2}{1.112825 + .459414\,\chi_p^2 + .804621\,\chi_p^2}$$

with a graph analogous to the previous one.

See also Tiago de Oliveira (1972) for details. In this paper it was also searched the asymptotically best sample quantiles π and $\pi'(0 < \pi < \pi' < 1)$ such that the variance of the estimators $\lambda^* = a\, x'_{\pi} + (1 - a)x'_{\pi'}, \delta^* = b\,(x'_{\pi'} - x'_{\pi})$, was minimum. The correct results (the ones in the paper had a computation error) are $\pi = .07,\ \pi' = .76, a = -.43065,\ b = .44032$ and the asymptotic efficiency is $\text{eff}\,(\lambda^*, \delta^*) = 40.8\%$. For more details see in the Chapter 9 the section "Estimation using two or three quantiles".

Blom (1958) and in Saharan and Greenberg (1962) developed the nearly best unbiased linear estimators (NBULE) making use of the quantiles $\Lambda^{-1}(i/(n+1))$ with relation to the order statistics X'_i, which Hassanein (1964) applied to the Gumbel distribution. The method can be extended to any distribution having only location and dispersion parameters, directly or by transformation.

Lieblein (1954) proposed a technique using a finite number (in general 3) of empirical quantiles. Hassanein (1964) increased the number of quantiles to be used for the estimation of λ and δ. With the sub-sample $(x'_{a_1}, x'_{a_2}, \dots, x'_{a_k})$ where $1 \le a_1 < a_2 < \cdots < a_k \le n$, for $k = 2(1)10$, and with the results of Ogawa (1951), he determined the quantiles to be used which maximize the joint efficiency of the estimators.

Recently Kubat and Epstein (1980) sought the sets of 2 (or 3) order statistics x'_r and $x'_{r'}$ with $(1 \le r = [n\,\pi] < r' = [n\pi'] \le n)$ and coefficients a and b, such that (λ^*, δ^*) with $\lambda^* = a\, x'_r + (1 - a)x'_{r'}$, $\delta^* = b(x'_{r'} - x'_r)$ were asymptotically unbiased, asymptotically binormal, and with the best efficiency that could be attained to estimate a quantile χ_p with $\pi < p < \pi'$.

Let us refer in greater detail to the method of block partitioning, which seems simple to execute and does not depend on knowledge of the coefficients of linear combination of the observations, although it needs, once and for all, the computation of efficiency and the best choice of splitting points; see Kubat (1982) for other modifications. The method can be extended to Fréchet and Weibull distributions, as will be seen later.

Take $r < s$ such that $1 < r < s < n$. The averages are

$\bar{x}_1 = \frac{1}{r}\sum_1^r x_i^{'} < \bar{x}_2 = \frac{1}{s-r}\sum_{r+1}^s x_i^{'} < \bar{x}_3 = \frac{1}{n-s}\sum_{s+1}^n x_i^{'}$, with probability 1.

It can be shown that $(\bar{x}_1, \bar{x}_2, \bar{x}_3)$ is asymptotically trinormal and we can seek coefficients a_j and $b_j (j = 1,2,3)$ such that the random pair

$$\lambda^* = \lambda^*(x_i) = \sum_1^3 a_i \ \bar{x}_i$$

$$\delta^* = \delta^*(x_i) = \sum_1^3 b_i \ \bar{x}_i$$

which is always asymptotically binormal has minimum variance; the condition of invariance (i.e., such that $\lambda^*(\alpha + \beta\, x_i) = \alpha + \beta\, \lambda^*(x_i)$, $\delta^*(\alpha + \beta\, x_i) = \beta\, \delta^*(x_i), (\beta > 0)$ implies $\sum_1^3 a_i = 1$)and $\sum_1^3 b_i = 0$. Then we can seek the best r and s in terms of efficiency and thus compute the a_i and b_i.

5.4 Further comments

As is well known, the method of maximum likelihood is asymptotically the best and its use is to be recommended.

The other methods (from the moments and the various linear estimators) are not asymptotically efficient and can be considered either as quick methods, if we use only a calculator (and tables if needed), or as a seed for the iterative solution of the $\hat{\delta}$ equation. If we do not use tables of coefficients, the most practical methods are those of moments — as said with efficiency 48% -or the linear methods with 2 quantiles of Tiago de Oliveira (1972) and of Kubat and Epstein (1980) or of Downton (1966).

5.5 Confidence interval estimation

In principle we should always use the maximum likelihood estimator, either for the p-quantile also with a confidence interval, or for the parameter point.

The maximum likelihood estimator $\hat{\lambda} + \chi_p\, \hat{\delta}$ is asymptotically normal with mean value $\lambda + \chi_p\, \delta$ and variance $(1 + \frac{6}{\pi^2}\left(1 - \gamma + \chi_p\right)^2)^{1/2}\ \frac{\delta^2}{n}$.

As

$$\sqrt{n}\,\frac{\left(\hat{\lambda}+\chi_p\,\hat{\delta}\right)-\left(\lambda+\chi_p\,\delta\right)}{\left(1+\frac{6}{\pi^2}(1-\gamma+\chi_p)^2\right)^{1/2}\,\delta}$$

is asymptotically standard normal and, as $\hat{\delta}\xrightarrow{P}\delta$, by the δ-method we know that:

$\sqrt{n}\,\frac{\left(\hat{\lambda}+\chi_p\,\hat{\delta}\right)-\left(\lambda+\chi_p\,\delta\right)}{\left(1+\frac{6}{\pi^2}(1-\gamma+\chi_p)^2\right)^{1/2}\,\hat{\delta}}$ *is asymptotically standard normal, which gives a simple confidence interval for* $\lambda+\chi_p\,\delta$ *independent of the parameter* δ as did not happen with the first asymptotic relation.

We could also compute the bias of estimator of the p-quantile $M(\hat{\lambda}-\lambda)+M(\hat{\delta}-\delta)\,\chi_p$ as well as its variance $V(\hat{\lambda}+\chi_p\,\hat{\delta})=V(\hat{\lambda})+2\,C(\hat{\lambda},\hat{\delta})\chi_p+V(\hat{\delta})\,\chi_p^2$. It seems that the bias converges to zero as $O(n^{-1})$ when the variance converges to zero as $O(n^{-1})$.

As a result of the asymptotic binormality of $(\hat{\lambda},\hat{\delta})$, the asymptotic confidence interval with significance level α for (λ,δ) is given by the curve in the (λ,δ) plane

$$(\hat{\lambda}-\lambda)^2-2(1-\gamma)\left(\hat{\lambda}-\lambda\right)\left(\hat{\delta}-\delta\right)+\left[\tfrac{\pi^2}{6}+(1-\gamma)^2\right]\left(\hat{\delta}-\delta\right)^2\le-\tfrac{2}{n}\log\alpha\cdot\delta^2\,,$$

that is,

$$(\hat{\lambda}-\lambda)^2-0.84556\left(\hat{\lambda}-\lambda\right)\left(\hat{\delta}-\delta\right)+1.82367\left(\hat{\delta}-\delta\right)^2\le-\frac{2}{n}\log\alpha\cdot\delta^2;$$

recall that the quadratic form of a binormal distribution has the exponential distribution; see Cramér (1946).

Also from $\hat{\delta}\xrightarrow{P}\delta$ and by the δ-method we can have asymptotically a part of the ellipsis (as $\delta>0$)

$$(\hat{\lambda}-\lambda)^2-2(1-\gamma)\left(\hat{\lambda}-\lambda\right)\left(\hat{\delta}-\delta\right)+\left[\tfrac{\pi^2}{6}+(1-\gamma)^2\right]\left(\hat{\delta}-\delta\right)^2\le-\tfrac{2}{n}\log\alpha\cdot\hat{\delta}^2\,.$$

The other estimators given in "Point Estimation" can be used to obtain an interval estimation or to test hypotheses about the parameters. *Most of them*

are asymptotically normal with known asymptotic variances and covariances, so that one can compute asymptotic confidence intervals; in some cases we may resort to Monte Carlo simulation. For more details see Johns and Lieberman (1966) and Herbach (1970). The reliability function $R(x) = \exp\{-\exp\{-v/\delta\}\}$, where $v = \lambda - x$, is an increasing function of v/δ and a lower confidence limit on v/δ can be used to determine a lower confidence limit on $R(x)$, as $\text{Prob}\{L < v/\delta\} = \text{Prob}\{\exp\{-\exp(-L)\} < \exp\{-\exp(-v/\delta)\} = R(x) = \beta$. They proposed to use their asymptotically efficient linear estimators of v and δ based on the first r order statistics in a sample of n. Let $Y_i^{'} = X_i^{'}/a$. Then for estimates of v, δ we use

$$Y_a = \sum_{i=1}^{r} a_i \; Y_i^{'} \; , Y_b = \sum_{i=1}^{r} b_i \; Y_i^{'} \; .$$

Let $V_a = (Y_a - v)/\delta \, , V_b = Y_a/\delta$. The joint distribution of (V_a, V_b) is parameter-free. Next define a function $L(t)$ with the property that for fixed $\beta \, , 0 < \beta < 1$, assuming that $\text{Prob}\{V_b > 0\} = 1$, we have

$$\text{Prob}\,(L(t) < t\,V_b - V_a) = \beta \qquad \text{for all } t.$$

Then

$$\text{Prob}\,\{L(Y_a \,/\, Y_b) < v/\delta\} = \text{Prob}\{(Y_a \,/\, Y_b) < L^{-1}\,(v/\delta)\} =$$

$$= \text{Prob}\,\{[\frac{v}{\delta} < L^{-1}\,(\frac{v}{\delta})V_b - V_a]\}$$

$$= \text{Prob}\{L\,[L^{-1}\,(\frac{v}{\delta})] < L^{-1}\,(\frac{v}{\delta})V_b - V_a\} = \beta.$$

Thus $L(Y_a \,/\, Y_b)$ being a lower confidence limit for v/δ with confidence coefficient β , correspondingly $\exp\{-\exp\{-L(Y_a \,/\, Y_b)\}\}$ is a lower confidence limit for $R(x)$, also with confidence coefficient β.

Finally they show that, asymptotically, L and $\exp\{-\exp\{-L\}\}$ are efficient bounds for v/δ and $R(x)$. For $n = 10, 15, 20, 30, 50, 100$ and with four values of r for each n , they computed Monte Carlo simulations of the distribution of $tV_b - V_a$ for *fixed* values of $t = Y_a/Y_b$. Thus they generated tables from which one can obtain the bounds for $R(x)$ for specific x for $\beta = .5, .75, .90, .95$ and $.99$. To get a lower bound on $R^{-1}(w)$ we find in the table a value of $\exp\{-\exp\{-L(Y_a/Y_b)\}\}$ corresponding to w which is $Y_a - t\,Y_b$

where t is the table value of Y_a/Y_b associated with $\exp\{-\exp\{-L(Y_a/Y_b)\}\}$ for the appropriate confidence level.

5.6 Tests of Hypotheses

Hypothesis testing on parameters can be handled, as usual, by relating the acceptance region to the confidence region.

Consequently, the asymptotic test of $(\lambda, \delta) = (\lambda_0, \delta_0)$ vs. $(\lambda, \delta) \neq (\lambda_0, \delta_0)$ with significance level α, is given by the acceptance region : accept (λ_0, δ_0) if this point falls in the interior of the ellipsis given for the confidence region, and reject otherwise.

Concretely, we will use as the acceptance region, for the significance level α,

$$(\hat{\lambda} - \lambda_0)^2 - 2(1-\gamma)(\hat{\lambda} - \lambda_0)(\hat{\delta} - \delta_0) + (\frac{\pi^2}{6} + (1-\gamma)^2)(\hat{\delta} - \delta_0)^2 \leq -\frac{2}{n} \log \alpha \cdot \delta_0^2.$$

Here we do not need to substitute δ_0^2 by $\hat{\delta}^2$ in the RHS because in the test δ is supposed known.

Evidently we could obtain, by the classical method, the locally most powerful and locally most powerful unbiased tests if one of the parameters, or a relation between them, is fixed; but we would obtain analogous difficulties and, finally, resort to tests based on maximum likelihood estimation with optimal asymptotic behaviour. They can be let as exercises.

Evidently we could also use other estimators such as the ones described previously, but their efficiencies would in general be low.

5.7 Point and Interval Prediction

Let us begin by considering the point prediction of the maximum of (the next) m observations from a sample of n. The likelihood of the sample $(x_1, \ldots, x_n)$ and the (future) maximum $W = w$ is evidently

$$L(\lambda, \delta | x_i, w) = \frac{1}{\delta^n} e^{-n(\bar{x}-\lambda)/\delta} \exp(\sum_1^n e^{-(x_i-\lambda)/\delta}) \cdot \frac{m}{\delta} e^{-(w-\lambda)/\delta} e^{-m(w-\lambda)/\delta}.$$

We could seek, following Tiago de Oliveira (1966) and (1968), the quasi-linearly invariant predictor $p(x_1, \ldots, x_n)$ (e.g., the predictor such that

$p(\alpha + \beta\, x_i) = \alpha + \beta\, p(x_i), \beta > 0)$ such that $M(w - p(x_i))^2$ is minimum. The result is not manageable, even using computers, as can be seen from the expression

$$p(x_i) = \frac{\int_0^{+\infty} \xi^{n-1} \frac{e^{-n\,\bar{x}\,\xi}}{(\sum_1^n e^{-\xi\, x_i})^n} (\gamma + \log m - \log \sum_1^n e^{-\xi\, x_i} + \frac{\Gamma'(n)}{\Gamma(n)}) d\,\xi}{\int_0^{+\infty} \xi^n \frac{e^{-n\,\bar{x}\,\xi}}{(\sum_1^n e^{-\xi\, x_i})^n} d\,\xi}.$$

As the $M(W) = \lambda + (\gamma + \log m)\,\delta$ and $(\hat{\lambda}, \hat{\delta})$ is asymptotically optimal it is then natural to use, as point predictor, the function $p(x_i) = \hat{\lambda} + k\,\hat{\delta}$ which is quasi-linearly invariant because $\hat{\lambda}(\alpha + \beta\, x_i) = \alpha + \beta\, \hat{\lambda}(x_i)$ and $\hat{\delta}\,(\alpha + \beta\, x_i) = \beta\, \delta(x_i)$.

Using the mean-square error as a measure of distance between the prediction and the (to be) observed value of $W = w$, we have

$$MSE(p) = M\left(W - \left(\hat{\lambda} + k\,\hat{\delta}\right)^2\right)$$

$$= M\big(W - M(W)\big)^2 + M(\hat{\lambda} + k\,\hat{\delta} - M(W))^2 =$$

$$= \frac{\pi^2}{6}\delta^2 + M(\hat{\lambda} + k\,\hat{\delta}) - (\lambda + (\gamma + \log m)\delta)^2.$$

The mean-square error of the point predictor $p(x_i) = \hat{\lambda} + k\,\hat{\delta}$ is asymptotically a minimum for $k = \gamma + log\ m$ and its variance is that of the quantile $\chi = \gamma + log\ m$, i.e., asymptotically.

$$\frac{\pi^2}{6}\delta^2 + [1 + \frac{6}{\pi^2}(\gamma + \log m)^2]\frac{\delta^2}{n}.$$

The determination of prediction regions, also dealt with in the same papers, leads to the same difficulties in computing integrals. Thus we will once more use the maximum likelihood estimators.

Let us now seek, then, one-sided and two-sided prediction intervals based on $(\hat{\lambda}, \hat{\delta})$, for a prediction level ω, in general close to 1.

The most important one-sided prediction interval for the prediction level ω, for W is $W \le \hat{\lambda} + c\,\hat{\delta}$ with c such that

$$Prob\{W \le \hat{\lambda} + c\,\hat{\delta}\} = \omega$$

which as

$\text{Prob}\{W \leq y\} = \Lambda^m((y-\lambda)/\delta) = \Lambda((y-\lambda)/\delta) - \log\ m$

is reduced to

$$M(\Lambda(\frac{\hat{\lambda} + c\,\hat{\delta} - \lambda}{\delta} - \log\ m)) = \omega.$$

Evidently the first approximation to c is from the relation $\Lambda(c - \log m) = \omega$ or $c = \chi_\omega + \log m$. Consider now the approximation $c = \chi_\omega + \log m + \frac{\varphi}{\sqrt{n}} + O\,(n^{-1/2})$. Let us put $\sqrt{n}\,\frac{\hat{\lambda}-\lambda}{\delta} = A_n, \sqrt{n}\,(\frac{\hat{\delta}}{\delta} - 1) = B_n$. Then we get

$$M(\Lambda(\frac{\hat{\lambda} + c\,\hat{\delta} - \lambda}{\delta} - \log m)) = M(\Lambda(c - \log m + \frac{A_n + c\,B_n}{\sqrt{n}})) = \omega.$$

Developing c through $\chi_\omega + \log m + \frac{\varphi}{\sqrt{n}} + O\,(n^{-1/2})$ we get the equation, by easy algebra,

$$\varphi + M(A_n) + (\chi_\omega + \log \omega)M(B_n) + O\big(n^{-1/2}\big) = 0,$$

and as $M(A_n) \to 0$ and $M(B_n) \to 0$ we get $\varphi = 0$. Thus $c = \chi_\omega + \log m + O(n^{-1/2})$, or $c = \chi_\omega + \log m$ with a linear error of the order of $O(n^{-1/2})$. The prediction level differs from ω in terms of $(O(n^{-1/2}))^2 = O(n^{-1})$ which shows that $c = \chi_\omega + \log m$ gives a good approximation to the quantile with an error in the prediction level of the order $O(n^{-1})$. Consequently:

The one-sided prediction interval with prediction level ω, to the order of $O(n^{-1})$ is, $\hat{\lambda} + (\chi_\omega + log\ m)\,\hat{\delta}$ with $\Lambda(\chi_\omega) = \omega$.

A subsequent summand Ψ/n of order n^{-1} could be obtained in the same way supposing now that $\sqrt{n}\,M(A_n) \to 0$ and $\sqrt{n}\,M(B_n) \to 0$; we have then $\Psi = \frac{1+\log\omega}{2}(1 + \frac{6}{\pi^2}(1 - \gamma + \chi_\omega + \log m)^2)$ and so the second order approximation to c, with a linear error of the order of $O(n^{-1})$, is

$$c = \chi_\omega + \log m + \frac{1 + \log \omega}{2n}\ (1 + \frac{6}{\pi^2}(1 - \lambda + \chi_\omega + \log m)^2)$$

and the prediction level differs from ω by a quantity of $O(n^{-3/2})$.

The other one-sided interval can be dealt with in the same way.

Let us sketch two-sided prediction intervals with prediction level ω and asymptotically shortest average length. If $a' \leq b'$ are quantities such that

$$\text{Prob}\left(\hat{\lambda} + a'\,\hat{\delta} \leq X \leq \hat{\lambda} + b'\,\hat{\delta}\right) = w$$

and $b' - a'$ is minimum, they are given by the equations

$$a' + e^{-a'} = b' + e^{-b'}.$$

It is then easy to show that if $a \leq b$ and are such that

$$\text{Prob}\left(\hat{\lambda} + a\,\hat{\delta} \leq W \leq \hat{\lambda} + b\,\hat{\delta}\right) = \omega$$

we shall have $a = a' + \log m, b = b' + \log m$, apart from terms of order n^{-1} with the length $(b - a)\,\delta = (b' - a')\delta$ which is asymptotically shortest on average.

Other estimators of (λ, δ) can be considered as well to form predictors.

5.8 Miscellaneous Results

5.8.1. Discrimination

Consider that we have two populations whose distribution functions are $\Lambda((x - \lambda)/\delta)$ and $\Lambda((x' - \lambda')/\delta)$, where the scale parameter is the same.

Consider that we want to decide if some observation y belongs to the first or the second population (i.e., has the location parameter λ or λ'). Supposing λ, λ' and δ to be known, we are in a situation analogous to usual Neyman-Pearson theory for testing hypothesis. Denoting by A and $A' = A^c$ the acceptance regions for λ and λ', the misclassification errors of the test are

$$\text{Prob}\{A|\lambda'\} = \int_A d\,\Lambda\left(\frac{y-\lambda'}{\delta}\right)$$

and $$\text{Prob}\{A'|\,\lambda\} = \int_{A'} d\,\Lambda\left(\frac{y-\lambda}{\delta}\right) = 1 - \int_A d\,\Lambda\left(\frac{y-\lambda}{\delta}\right).$$

To be fair we stipulate these misclassification errors to be equal, i.e.,

$$\int_A \left[d\,\Lambda\left(\frac{y-\lambda}{\delta}\right) + d\,\Lambda\left(\frac{y-\lambda'}{\delta}\right)\right] = 1$$

and minimize the (equal) misclassification errors $\int_A d\,\Lambda\left(\frac{y-\lambda'}{\delta}\right)$.

This test, by the Neyman-Pearson theorem, has the acceptance region for λ,

$$\frac{\frac{1}{\delta}\,\Lambda'\left(\frac{y-\lambda'}{\delta}\right)}{\frac{1}{\delta}\,\Lambda'\left(\frac{y-\lambda}{\delta}\right)+\frac{1}{\delta}\,\Lambda'\left(\frac{y-\lambda'}{\delta}\right)} \le k$$

or

$$e^{-y/\delta}\left(e^{\lambda/\delta}-e^{\lambda'/\delta}\right) \le c\,(\lambda,\lambda',\delta).$$

Suppose that $\lambda > \lambda'$. Then the acceptance region A is given by $y > c\,(\lambda,\lambda',\delta)$, where c is defined by

$$\Lambda\left(\frac{c-\lambda}{\delta}\right)+\Lambda\left(\frac{c-\lambda'}{\delta}\right)=1$$

and the misclassification error is given by

$$\int_A d\,\Lambda\left(\frac{y-\lambda'}{\delta}\right)=1-\Lambda\left(\frac{c-\lambda'}{\delta}\right);$$

if $\lambda < \lambda'$ the changes are obvious.

As λ, λ' and δ are not known it is natural to estimate them and substitute in the previous relations.

For a sample of $(x_1,\ldots,x_n)$ of the first population and of $(x_1',\ldots,x_n')$ of the second one, the maximum likelihood estimators are

$$\hat{\lambda}=-\hat{\delta}\log\frac{\sum_1^n e^{-x_i/\hat{\delta}}}{n},\quad \lambda'=-\hat{\delta}\log\frac{\sum_1^{n'} e^{-x'_i/\hat{\delta}}}{n'}$$

and

$$\hat{\delta}=\frac{n}{n+n'}\left(\bar{x}-\frac{\sum_1^n x_i\,e^{-x_i/\delta}}{\sum_1^n e^{-x_i/\delta}}\right)+\frac{n'}{n+n'}\left(\bar{x}'-\frac{\sum_1^{n'} x'_i\,e^{-x'_i/\hat{\delta}}}{\sum_1^{n'} e^{-x'_i/\hat{\delta}}}\right).$$

Thus we substitute λ, λ' and δ by their estimators in the fair test above, taking for A the discrimination region corresponding to the largest estimated location parameter.

For more details and the analysis of the asymptotic behaviour (consistency) see Tiago de Oliveira (1973).

5.8.2. Tolerance intervals

The concept of tolerance intervals was introduced by Wilks (1941) as a procedure for the prediction of the values of a population. In Themido (1985), parameter-free tolerance intervals for the Gumbel distribution are obtained and a comparison, for large samples, is made with the distribution-free tolerance intervals proposed by Wilks, a comparison based on the respective expected lengths.

Let $(x_1,\ldots,x_n)$ be an i.i.d. sample of one-dimensional continuous random variables, and $L_i = L_i\,(x_1,\ldots,x_n)$, $i = 1,2$, be statistics such that $L_1\,(x_i) < L_2\,(x_i)$. The random interval $[L_1, L_2]$ is a (β,ω) -tolerance interval if

$$\text{Prob}\,\{\,\text{Prob}[L_1 \le X \le L_2] \ge \beta\} = \omega,$$

ω being the tolerance level of the interval; if, instead of this relation, we have

$$M(\text{Prob}[L_1 \le X \le L_2]) = \alpha,$$

the interval $[L_1, L_2]$ is an α-average tolerance interval.

In the first case for given β and $\omega(0 < \beta, \omega < 1)$, we seek tolerance limits L_1 and L_2, based on the sample, such that $[L_1, L_2]$ contains, with probability ω at least $100\,\beta\,\%$ of the populations or such that we have a confidence of $100\,\omega\,\%$ that the probability that a future observation will lie between L_1 and L_2 is at least β ; for the second case we must find L_1 and L_2 such that the average of mean coverage of the interval is $\alpha(0 < \alpha < 1)$.

If we have $L_1 = -\infty$ or $L_2 = +\infty$ we obtain one-sided tolerance intervals.

Regarding the distribution function of the X_i we should consider two situations:

- the non-parametric case where the functional form of the distribution of the X_i is unknown, apart from the fact that it is (absolutely) continuous;
- the parametric case where the functional form of the distribution of the X_i is known and only the values of a finite number of parameters of the distribution are unknown.

For the determination of tolerance intervals for one-dimensional distributions in the non-parametric case, an exact solution, using order statistics, was given by Wilks (1941).

In the parametric case it is natural to expect that tolerance limits better than those of Wilks can be obtained.

For the Gumbel distribution $\Lambda(x|\lambda,\delta)=\Lambda((x-\lambda)/\delta), -\infty<\lambda<+\infty, 0<\delta<+\infty$, let us use as tolerance limits the maximum likelihood estimators of the quantiles of probabilities $P_i=\Lambda(k_i), i=1,2, L_i=L_i(x_1,\dots,x_n)=\hat{\lambda}+k_i\,\hat{\delta}$, with k_1 and k_2 $(k_1<k_2)$ real coefficients independent of n, a choice made because $\hat{\lambda}(x_i)$ and $\hat{\delta}(x_i)$ are optimum estimators of λ and δ in terms of asymptotic efficiency. Recall that $\hat{\lambda}+k\,\hat{\delta}$ are the unique quasi-linearly invariant functions of $\hat{\lambda}$ and $\hat{\delta}$.

The (random) proportion of the Gumbel distribution covered by the tolerance interval $[L_1, L_2]$ is

$$P_{1,2}=\Lambda\left(\frac{\hat{\lambda}(x_i)}{\delta}+k_2\frac{\hat{\delta}(x_i)}{\delta}\right)-\Lambda\left(\frac{\hat{\lambda}(x_i)}{\delta}+k_1\frac{\hat{\delta}(x_i)}{\delta}\right)=$$

$$=\Lambda(\hat{\lambda}(Z_i)+k_2\,\hat{\delta}(Z_i))-\Lambda(\hat{\lambda}(Z_i)+k_1\hat{\delta}(Z_i)),$$

where $(z_1,\dots,z_n)$ is the corresponding random sample from the reduced Gumbel distribution. This follows from the invariance of maximum likelihood estimators and from the form of the quasi-linearly invariant tolerance limits considered. Thus, the probability $\widehat{P}_{1,2}$ of $[L_1, L_2]$ is parameter-free and the (β,ω)-tolerance interval can be written as $\text{Prob}\{\widehat{P}_{1,2}\geq\beta\}=\omega$. Using the δ-method Tiago de Oliveira (1982) we can show that the asymptotic distribution of $\widehat{P}_{1,2}$ is normal, as usual, with mean value $\Lambda(k_2)-\Lambda(k_1)$ and variance $O(n^{-1})$. Even if we take $k_1(n)=k_i+a_i\sqrt{n}+O(n^{-1/2})$, the a_i can be computed to give a better approximation; details of an analogous approach can be found in Chapter 8.

For each pair $(\beta,\omega), 0<\beta,\omega<1$, there is an infinity of pairs (k_1,k_2) such that the tolerance limits L_1 and L_2 define $a(\beta,\omega)$-tolerance interval for the Gumbel distribution. We can determine the minimum length parameter-free (β,ω)-tolerance interval.

Also, in terms of the α-average tolerance interval, $0 < \alpha < 1$, we can seek the one with minimum length.

The parameter-free (β, ω)-tolerance limits and the parameter-free α-average tolerance limits, being estimators of quantiles for the Gumbel distribution converge in probability, as $n \to +\infty$, to the quantiles which are the end-points of the shortest quantile intervals with probability β and α respectively. Tolerance intervals thus correspond to the intervals above but with random end-points.

We can also deal with one-sided parameter-free (β, ω) -tolerance interval; in this case we have only one coefficient, k_1 or k_2 for lower or upper intervals respectively.

In the comparison for large samples, between distribution-free and parameter-free two-sided tolerance interval for the Gumbel distribution, with respect to the mean length criterion, the parameter-free tolerance interval using the information about the distribution are, as expected, better than distribution-free tolerance interval.

For the proofs, a tabulation of (k_1, k_2), and other details, see Themido (1985).

5.8.3. Multisample analysis

Consider two samples $(x_1, \ldots, x_n)$ and $(x_1', \ldots, x_n')$, for which we assume that they have a Gumbel distribution with parameters λ and λ' and the same parameter δ. We want to test if $\lambda = \lambda'$. This is the first step of Multisample Analysis for Extremes, analogous to Analysis of Variance, but practically unsolved for extremes, and the only case that will be dealt with here.

Analogously to what happened to discrimination we have

$$\hat{\lambda} = -\hat{\delta} \log \frac{\sum_1^n e^{-x_i/\hat{\delta}}}{n}, \qquad \hat{\lambda}' = -\hat{\delta} \log \frac{\sum_1^{n'} e^{-x'_i/\hat{\delta}}}{n'}$$

and

$$\hat{\delta} = \frac{n}{n+n'} \left(\bar{x} - \frac{\sum_1^n x_i e^{-x_i/\hat{\delta}}}{\sum_1^n e^{-x_i/\hat{\delta}}}\right) + \frac{n'}{n+n'} \left(\bar{x}' - \frac{\sum_1^{n'} x'_i e^{-x'_i/\hat{\delta}}}{\sum_1^{n'} e^{-x'_i/\hat{\delta}}}\right)$$

if λ is supposed different from λ' and

$$\hat{\hat{\lambda}} = -\hat{\hat{\delta}}\ \log\frac{\sum_1^n e^{-x_i/\hat{\hat{\delta}}} + \sum_1^{n'} e^{-x'_i/\hat{\hat{\delta}}}}{n+n'}$$

$$\hat{\hat{\delta}} = \frac{n\,\bar{x} + n'\,\bar{x}'}{n+n'} - \frac{\sum_1^n x_i\, e^{-x_i/\hat{\hat{\delta}}} + \sum_1^n x'_i\, e^{-x'_i/\hat{\hat{\delta}}}}{\sum_1^n e^{-x_i/\hat{\hat{\delta}}} + \sum_1^{n'} e^{-x'_i/\hat{\hat{\delta}}}}$$

if we suppose $\lambda' = \lambda$.

The likelihood ratio criterion leads to the test statistic

$$-2\log\frac{\max_{\lambda,\delta}[L_n(x|\lambda,\delta)L_{n'}(x'|\lambda,\delta)]}{\max_{\lambda,\lambda',\delta}[L_n(x|\lambda,\delta)L_{n'}(x'|\lambda,\delta)]} =$$

$$= 2\{(n+n')\log\frac{\hat{\hat{\delta}}}{\hat{\delta}} + \frac{n(\bar{x}-\hat{\hat{\lambda}})+n'(\bar{x}'-\hat{\hat{\lambda}})}{\hat{\hat{\delta}}} - \frac{n(\bar{x}-\hat{\lambda})+n'(\bar{x}'-\hat{\lambda}')}{\hat{\delta}}\}$$

which is asymptotically a χ^2 with one degree of freedom in the hypothesis tested.

5.8.4. Overpassing probability of a level

Let a be a level and suppose that X has a Gumbel distribution $\Lambda((x-\lambda)/\delta)$. Then $P(a) = P(a|\lambda,\delta) = \text{Prob}\{X > a\} = 1 - \Lambda((a-\lambda)/\delta)$. We will take as estimator of $P(a|\lambda,\delta)$, using the estimators $(\hat{\lambda}, \hat{\delta})$ of (λ,δ),

$$\hat{P}(a) = P\left(a\middle|\hat{\lambda}, \hat{\delta}\right)$$

and by the use δ-method we have that $\sqrt{n}\,(\hat{P}(a) - P(a|\lambda,\delta))$ has the asymptotically normal distribution of the linearized expression.

$$\sqrt{n}\,[(\hat{\lambda}-\lambda)\frac{\partial P(a)}{\partial\lambda} + (\hat{\delta}-\delta)\frac{\partial P(a)}{\partial\delta}] =$$

$$= \sqrt{n}(1-P(a))\,(-\log(1-P(a))\,[\frac{\hat{\lambda}-\lambda}{\delta} + \frac{a-\lambda}{\delta}\ \frac{\hat{\delta}-\delta}{\delta}].$$

Its asymptotic mean value is zero and the variance is

$$V(a|\lambda,\delta) = (1-P(a))^2\,(\log(1-P(a)))^2\,\{1 + \frac{6}{\pi^2}(1-\gamma+\frac{a-\lambda}{\delta})^2\}.$$

Also $\sqrt{n}\,\frac{\hat{P}(a)-P(a)}{\sqrt{V(a|\lambda,\delta)}}$ is asymptotically standard normal even with $V(a|\lambda,\delta)$ substituted by $\hat{V}(a) = V(a|\hat{\lambda},\hat{\delta})$.

Then, λ_α denoting as usual the solution of $N(x) = 1 - \alpha$, we have

$\text{Prob}\{\sqrt{n}\ \frac{\hat{P}(a) - P(a)}{\sqrt{\hat{V}(a)}} \geq -\lambda_\alpha\} \to 1 - \alpha$ as $n \to \infty$ and thus:

The one-sided asymptotic confidence interval for $P(a)$ with significance level α is

$$0 \leq P(a) \leq \hat{P}(a) + \frac{\lambda_\alpha}{\sqrt{n}}\sqrt{\hat{V}(a)}.$$

5.8.5. A worked example

Sneyers (1977) gives the following $n = 35$ maximal yearly precipitations from 1938 to 1972 at Uccle, in 1 mm units, for the durations of 24 h, 60 min, 10 min and 1 min.

Experience says Sneyers - and the plotting confirms it - shows that for small durations the maximal precipitation follows a Gumbel distribution but for large durations it follows a Fréchet distribution.

The ML method applied to 1 min duration data gives $\hat{\lambda} = 1.709286$ and $\hat{\delta} = .778273$, close to 1.694 and $.7777$ or 1.691 and $.7484$ obtained by Sneyers by other methods.

From the values of $(\hat{\lambda}, \hat{\delta})$ we can solve other decision problems such as the ones discussed before.

To test the goodness of fit of a *known* (or *fixed*) distribution $F(x)$ we can use the Kolmogoroff-Smirnov test statistic

$$KS_n = \sup_x |S_n(x) - F(x)| = \max_1^n \{\max\,(|F(x_i') - \tfrac{i}{n}|, |F(x_{i+1}') - \tfrac{i}{n}|)\},$$

where $S_n(x) = S_n(x\,|x_i)$ denotes the sample distribution function ($x_1' \leq x_2' \leq \cdots \leq x_n'$), the order statistics of the i.i.d. sample and is taken $F(x_{n+1}') = 1$.

We accept $F(x)$, at the significance levels $.05$ and $.01$, according to $\sqrt{n}\,KS_n \leq 1.36$ or $\sqrt{n}\ KS_n \leq 1.63$, rejecting otherwise. When we assume *model* $F(x|\theta)$, we estimate θ by $\hat{\theta}$ (say) and instead of the unknown (parameterized) $F(x|\theta)$ we compute $\hat{K}S_n = \sup_x |\,S_n(x) - F(x|\hat{\theta})|$ and act in

the same way as before. Although the procedure is not completely justified, it can be expected to be sufficiently approximate when n is large.

Table 5.2

Year	24 h	1 min	10 min	60 min
1938	33.8	2.5	6.5	14.0
1939	27.7	1.0	8.5	12.8
1940	60.0	0.5	5.0	12.9
1941	24.0	0.9	8.4	11.9
1942	72.3	1.5	13.2	20.6
1943	50.7	4.4	11.9	29.1
1944	18.7	1.0	3.8	6.2
1945	41.2	3.0	13.0	21.1
1946	26.6	3.3	11.1	11.2
1947	27.2	2.0	13.0	18.0
1948	23.8	1.8	6.5	15.6
1949	19.8	1.0	5.7	8.7
1950	34.3	2.0	13.3	23.8
1951	28.2	4.0	12.2	12.2
1952	51.1	2.0	8.4	29.0
1953	37.5	1.0	5.0	9.9
1954	34.3	2.0	6.9	12.5
1955	22.2	1.6	6.2	9.6
1956	35.6	3.0	8.5	18.8
1957	34.2	1.6	9.8	12.0
1958	24.3	2.0	5.5	12.0
1959	20.3	1.2	9.8	11.6
1960	48.0	2.0	9.5	15.3
1961	32.4	1.5	11.5	19.2
1962	59.6	2.9	12.7	42.8
1963	60.4	3.7	9.0	13.0
1964	27.0	2.7	13.0	15.7
1965	45.8	2.0	12.2	15.4
1966	39.8	2.9	9.5	14.3
1967	21.6	3.0	11.9	13.1
1968	19.7	2.1	8.3	14.9
1969	54.4	2.3	15.3	25.8
1970	29.1	2.2	13.8	17.1
1971	41.6	1.6	7.0	21.2
1972	26.0	2.8	8.7	16.3

In our case, assuming the Gumbel model, with the values of $(\hat{\lambda}, \hat{\delta})$ above, we have

$$\hat{K}S_n = \sup_x |S_n(x) - \Lambda((x - \hat{\lambda})/\hat{\delta})| = .0102432$$

and as $n = 35$ we have $\sqrt{n}\ \hat{K}S_n = .06059$, accepting thus the Gumbel model.

References

Blom, G., 1958. *Statistical Estimation and Transformed Beta-Variables*, Wiley, New York.

Chernoff, H., Gastwirth, L., and Johns, M. V., 1967. Asymptotic distribution of linear combinations of functions of order statistics with applications to estimation. *Ann. Math. Statist*., 38, 52-71.

Cramér, H., 1946. *Mathematical Methods of Statistics*, Princeton University Press, New Jersey.

Downton, F., 1966. Linear estimates of parameters in the extreme value distribution. *Technometrics*, 8, 3-17.

Gumbel, E. J., 1958. *Statistics of Extremes*, Columbia University Press, New York.

Hassanein, K. M., 1964. *Estimation of the parameters of the extreme value distribution by order statistics*, University of North Carolina, Chapell Hill.

Herbach, L., 1970. An exact asymptotically efficient confidence bound for reliability in the case of Gumbel population. *Technometrics*, 12, 700-701.

Herbach, L., 1970. Introduction, Gumbel model. in *Statistical Extremes and Applications*, Tiago de Oliveira ed., 49-80, D. Reidel, Dordrecht.

Johns Jr, M. V., and Lieberman, G. J., 1966. An exact asymptotically efficient confidence bound for reliability in the case of Weibull distribution. *Technometrics*, 8, 135-175.

Kubat, P., and Epstein, B., 1980. Estimation of quantiles of location scale distribution based on 2 or 3 order statistics. *Technometrics*, 22, 575-581.

Kubat, P., 1982. Simple large sample estimators of scale and location parameters based on blocks of order statistics. *Trab. Estad. Inv. Oper*., 33, 86-118.

Lieblein, J., 1953. On the exact evaluation of the variance and covariance of order statistics in samples from the extreme value distribution. *Ann. Math. Statist.*, 24, 282-287.

Lieblein, J., 1954. A new method of analysing extreme value data. *Nat. Advisory Comm. for Aviation, Techn*.

Lieblein, J., and Zelen, M., 1956. Statistical investigation of the fatigue life of deep groove ball bearings. *J. Res. Nat. Bur. Stand*., 57, 273-316.

Loyd, E. H., 1962. Generalized least squares theorem. in *Contributions to Order Statistics*, Ahmed E. Sarhan and Bernard G. Greenberg eds., 20-27, Wiley.

Mann, N. R., 1967a. Results on the location and scale parameter estimation with applications to extreme-value distribution. *Aerospace Res. Labs.*, Rep 67-0023, Wright-Patterson AFB, Ohio.

Mann, N. R., 1967b. Tables for obtaining the best linear invariant estimates of the parameters of the Weibull distributions. *Technometrics*, 9, 629-645.

Mann, N., 1968. Point and interval estimation procedures for the two-parameter Weibull and extreme-value distributions. *Technometrics*, 10, 231-256.

Ogawa, J., 1951. Contributions to the theory of systematic statistics. *Osaka Math. J*., 3, 175-213.

Panchang, G. M., and Aggarwal, V. P., 1962. Peak flow estimation by the method a maximum likelihood. *Techn. Mem*. HLO-2, Central Water and Power Research Statics, Poona, India.

Posner, E. C., 1965. The applications of extreme value theory to error free communication. *Technometrics*, 7, 517-519.

Sarhan, A. E., and Greenberg, B.G., (eds.) 1962. *Contributions to Ordered Statistics*, Wiley, New York.

Sneyers, R., 1977. L'intensite maximale des precipitations en Belgique, Inst. Royal Meteor. Belgique, Ser. B, n° 86.

Themido, T., 1985. Intervalos de tolerância equivariantes para a distribuição de Gumbel. Universidade de Lisboa, Tese de Mestrado.

Tiago de Oliveira, J., 1963. Decision results for the parameters of the extreme value distribution (Gumbel) based on the mean and the standard deviation. *Trab. Estad. Inv. Oper*., 14, 61-81, Madrid.

Tiago de Oliveira, J., 1966. Quasi-linearly invariant prediction. *Ann. Math. Statist*., 37, 1684-1687.

Tiago de Oliveira, J., 1968. Efficiency evaluation for quasi-linear invariant predictors. *Rev. Belge Statist. Rech. Oper.*, 9, 1-9.

Tiago de Oliveira, J., 1972. Statistics for Gumbel and Fréchet distributions. in *Structural Safety and Reliability*, A. M. Freudenthal ed., 91-105.

Tiago de Oliveira, J., 1973. Quasi-linear discrimination. in *Discriminant Analysis and Applications*, T. Cacoullos ed., 291-309, Academic Press, New York.

Tiago de Oliveira, J., 1982a. A definition of estimator efficiency in k-parameter case. *Inst.Statist. Math.*, 34 A, 411-421.

Tiago de Oliveira, J., 1982b. Efficient estimation for quantiles of Weibull distributions. *Beige Statist. Rech. Oper.*, 22, 3-10.

Wilks, S. S., 1941. Determination of a sample sizes for setting tolerance limits. *Ann. Math. Statist.*, 12, 91-96.

*****