STATISTICAL THEORY OF EXTREME VALUES (*)

a short summary by

E.J. GUMBEL, Columbia University

---

1. *Definitions.*

To a continuous statistical variable X correspond the probability functions

$$F(x) = \text{Prob}\,(X < x); \quad P(x) = \text{Prob}\,(X \geqslant x)$$

The derivative $f(x) = F'(x)$ is called the density function. Two distributions are said to be mutually symmetrical if

$$F_1(x) = 1 - F_1(-x)$$

The *intensity function* $\mu(x) \geqslant 0$ is defined as the logarithmic derivative of the probability function. The *return period* $T(x)$ of a value greater than x is defined for observations equidistant in time or another measure by

$$T(x) = \frac{1}{1 - F(x)} > 1$$

This function increases with x.

A connected notion is the characteristic largest among n values u_n defined by

$$F(u_n) = 1 - 1/n$$

which increases with lg n. The intensity function at $x = u_n$ is written:

$$\mu(u_n) \equiv \alpha_n = nf(u_n)$$

(*) Conférence faite à la Société belge de Statistique le 18 juin 1962.

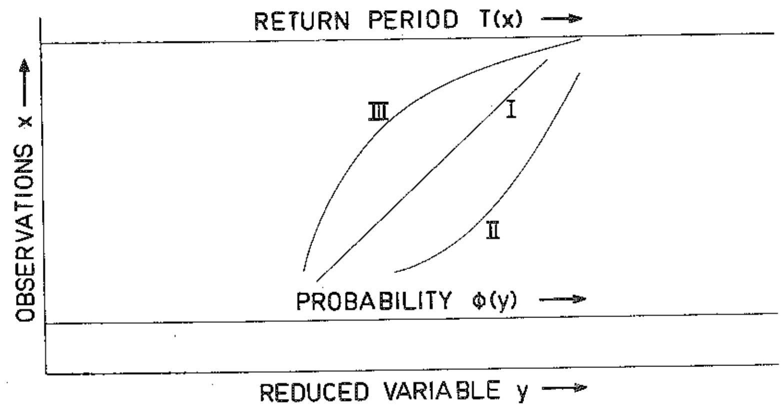
Let $\alpha\,(> 0)$ and θ be parameters, and let y be a reduced variable such that

$$y = \alpha\,(x - \theta)$$

then

$$F\,[\alpha\,(x - \theta)] = \Phi(y)$$

A probability paper consists of a scheme such as in graph 1.



Graph 1.

The mth among n observations x_m, arranged in increasing order of magnitude, is plotted at

$$F(x_m) = \frac{m}{n + 1}$$

The return period scale serves for forecasting for large values of x, the probability for the interval $0.32T$ to $3.31T$ is: 0,6827 as for $E(x) \pm \sigma$ in the normal distribution. This theorem leads to distribution free control intervals.

2. *Exact Theory.*

The probability $w\,(n, m, N, x)$ that the mth among n observations ordered in decreasing magnitude is *exceeded* x times in N future observations taken from the same population is

$$w\,(n, m, N, x) = \binom{n}{m}\binom{N}{x}\, m \bigg/ \binom{n + N}{m + x}\,(m + x) \qquad \begin{matrix} 1 \leqslant m \leqslant n \\ 0 \leqslant x \leqslant N \end{matrix}$$

The mean number of exceedances is

$$E(x) = m\,N/(n + 1)$$

The median number $\breve{x}$ in the case N = n is $\breve{x} = (m - 1)$. A forecast of the number of exceedances is more precise for the largest than for the median observation. If N = n is large, and $m = p/(N + 1)$, the distribution becomes normal. If N = n is large, and m remains fixed, the corresponding *law of rare exceedances*

$$w(m, x) = \binom{x + m - 1}{x} \left(\frac{1}{2}\right)^{m + x}$$

is similar to Poissons's law. These methods of forecasting the number of exceedances are quite general, because no knowledge of the initial distribution is required.

The exact probability functions $\Phi_n(x)$ and $\pi_n(x)$ of the largest, and the smallest, of n independent observations are

$$\Phi_n(x) = F^n(x); \qquad \pi_n(x) = P^n(x)$$

If the initial distribution is symmetrical, then the distribution of the largest value is symmetrical to the distribution of the smallest one. More generally, to any distribution of the largest value, we can construct a corresponding distribution of the smallest value (symmetry principle). For consecutive values of n, the curves $F(x)$ representing n are shifted to the right. Since the density function $\varphi_n(x)$ at $x = u_n$ is

$$\varphi_n(u_n) = \alpha_n/e$$

the curves $F^n(x)$ become more (less) concentrated if the extremal intensity α_n increases (decreases) with n. In the first (second) case the precision of a largest value increases (decreases) with the sample size n from which it is taken. The analytical properties of $\Phi_n(x)$ and $\pi_n(x)$ depend only on the properties of the initial distribution for large (and small) values of x.

For any continuous distribution (possessing the first two moments) the expected largest value $E(x_n)$ increases more slowly than $\sqrt{n/2}$ times the initial standard deviation and more slowly than $\frac{1}{2}\sqrt{n}$ for symmetrical distributions. At the characteristic largest value, the probability function becomes, even for moderate sample sizes,

$$\Phi_n(u_n) = 1/e$$

Therefore, the median largest value $\breve{x}_n$ exceeds the characteristic one. If Hopital's Rule holds for large x, in the form

$$\mu(x) \rightarrow -\frac{d \lg f(x)}{dx}$$

(distributions of the exponential type), the modal largest value $\tilde{x}_n$ converges to the characteristic largest value u_n.

For the uniform distribution

$$f(x) = 1/\theta; \quad 0 \leqslant x \leqslant \theta$$

the estimation of θ from the largest value is more precise than from the mean.

Numerous tables of the probability and density functions, the expectation, median and variance of the largest values as functions of n exist for the normal distribution (Pearson). Similar tables can easily be constructed for the exponential, log normal and gamma distributions. The following table gives some values of the characteristic largest value u_n and the extremal intensity function α_n.

	Distribution: exponential	*logistic*	*normal*
Characteristic largest value u_n	$\lg n$	$\lg n$	$\sqrt{2 \lg . 4n}$
Extremal intensity α_n	1	$\rightarrow 1$	$\rightarrow u_n$

For a known initial distribution, the exact distribution of extreme values for n observations gives a simple criterion for the rejection of outliers.

3. *Asymptotic Theory.*

The problem is: how does $\Phi_n(x)$ behave if n and therefore x increase? This was studied first, on a purely numerical basis, for the normal distribution but no analytical results of general validity were obtained. The exponential distribution is a better starting point. Different authors (Fréchet, Mises, Gnedenko, Gumbel) have used different methods to obtain the asymptotic probability function requesting different conditions on the analytic nature of the initial distribution. As shown in the following table distinctions have to be made for three classes of initial distribution, namely the exponential, the Cauchy, and the limited type. The first two are unlimited in the direction of the extreme. For the Cauchy distribution the moments of an order $l \geqslant k$ do not exist.

The usual derivation (Fisher and Tippett and Jenkinson) is based on the stability postulate which requires that the maximum of the largest value should have the same distribution as the largest value itself except for a linear transformation of the variable. There are three, and only three, asymptotic distributions of extremes shown in the tables in the form $-\lg \Phi(x)$ and $-\lg \pi(x)$.

The three asymptotic distributions of largest values.

Name	$-\lg \Phi(x)$	Variation	Conditions
exponential	$\exp[-\alpha(x-\theta)]$	$-\infty < x < +\infty$	$\mu(x) \to -\dfrac{d \lg f(x)}{dx}$
Cauchy	$\left(\dfrac{x-\varepsilon}{\theta-\varepsilon}\right)^{-k}$	$x \geqslant \varepsilon,\ k > 0$; $\theta > \varepsilon$	$(x-2)\,\mu(x) \to k$
limited	$\left(\dfrac{\omega-x}{\omega-\theta}\right)^{k}$	$x \leqslant \omega,\ k > 0$; $\omega > \theta$	$\dfrac{\mu(x)}{\omega-x} \to k$

The parameters here are chosen in such a way that ω and ε stand for the upper and lower limits and

$$\Phi(\theta) = 1/e$$

The parameter θ which corresponds to u_n is again called the characteristic largest value.

The symmetry principle leads to the three corresponding expressions for smallest values. Let $\pi(x)$ be the probability of the smallest value to exceed x.

Then

$$-\lg \pi(x) \begin{cases} = \exp[\alpha(x-\theta)] & -\infty < x < \infty \quad \text{exponential type} \\ = \left(\dfrac{\omega-x}{\omega-\theta}\right)^{-k} & \omega \geqslant x,\ \omega > \theta,\ k > 0 \quad \text{Cauchy type} \\ = \left(\dfrac{x-\varepsilon}{\theta-\varepsilon}\right)^{k} & x \geqslant \varepsilon,\ \theta > \varepsilon,\ k > 0 \quad \text{limited type} \end{cases}$$

with the same conditions on the parameters.

For the first type, the moment-generating functions for the variable $y = \alpha (x - \theta)$ are

$$C_n(t) = \Gamma(1 - t), \; C_1(t) = \Gamma(1 + t)$$

for the largest and smallest values, respectively. If an extreme value has an asymptotic distribution, then the standardized values $(x_n - \bar{x})/s$ and $(\bar{x} - x_1)/s$ have the same asymptotic distribution. Here $\bar{x}$ and s stand for the sample mean and standard deviation.

Although the asymptotic distributions of extremes were first established for independent observations it can be shown that they also hold for observations where m consecutive observations are dependent provided that m is very small compared to n. Another important theorem states that the two extremes are asymptotically independent. The same holds for the mth extremes, i.e., the second, third, etc.. value from the top and the bottom. Under certain conditions the extremes are independent of the sample quantiles. It is not yet known under what conditions the extremes are independent of the sample mean.

The theory has recently been generalized into more than one dimension. However, bivariate distributions are not determined by their margins. Therefore, we can only expect to find families of bivariate extremal distributions.

Let

$$-\lg \Phi(x) = \xi; \quad -\lg \Phi(y) = \eta; \quad -\lg \Phi(x, y) = \zeta$$

and let

$$t = \xi/\eta$$

the the general form for bivariate distributions given by Geffroy is

$$\zeta = \eta + \eta\, g(t)$$

where g is an increasing convex function which behaves asymptotically like a straight line. Only two special cases are known which can be written down explicitly as functions of the marginal distributions, namely

$$\zeta^m = \xi^m + \eta^m; \quad m \geqslant 1$$

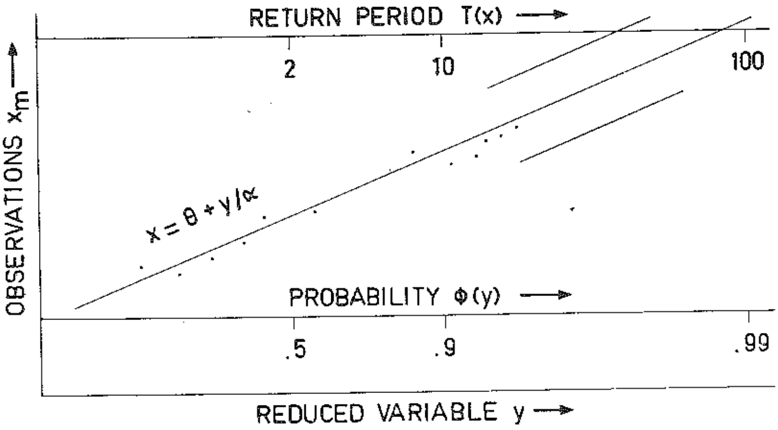
and

$$\zeta = \xi + \eta - a\,(1/\xi + 1/\zeta)^{-1}; \quad 0 \leqslant a \leqslant 1$$

The cases $m = 1$ and $a = 0$ stand for independence.

4. *Technical Applications.*

In many technical problems, it is not the mean but the extreme values which is of decisive importance. A bridge must not only withstand the mean discharge of a river, but its largest value to be expected within a given period. A building must withstand the strongest wind to be expected within a given period, represented on the return period scale. If we know the initial distribution and if the sample from which the observed extremes were taken is small, we have to use the exact theory for the analysis of the observations and the forecast. If we know only the type of initial distribution and the sample is large we have to use asymptotic theory. In most cases some reasonable assumption about the type of the initial distribution can be made. In many practical applications, however, the initial distribution is unknown and the only observations available are the extremes themselves. In aeronautics, there are many measuring devices which give only the extreme values. The extremal probability paper (see graph 2) then gives a criterion indicating which of the three types should be chosen for the analysis of observed extremes.



Graph 2.

In the first and second asymptotic distributions of largest values, there is no upper limit. Therefore, it does not make sense to speak of a maximum, but only of the maximum which is the most probable one to be reached within a given time. In the first distribution, the maxima increase as a linear function of the return period. In the second one, the logarithm of the maxima increase as a linear function of the return period, which may be identified with the number of years in hydrological or climatological problems, or with the number of traverses of about the same length in aerodynamical applications.

Many observations have confirmed the validity of the theory. The first asymptotic distribution of largest values holds for the oldest ages at death, floods, i.e., the largest annual discharges of a river, the largest rain-falls, the largest atmospheric pressures, the largest temperatures, largest snowfalls, the size of boulders in a sand pit, wind speeds, gusts, acceleration increments, effective gust velocities, maximum air speeds and similar aerodynamic phenomena. However, in certain cases, especially for the floods and the wind speeds, a good approximation is also reached by the use of the second distribution.

The third asymptotic distribution of largest values which possesses an upper limit has not yet found practical applications.

The first asymptotic distribution of smallest values has been successfully applied to the minima of the atmospheric pressures, minima of temperatures, the breaking strengths of rubber, the breakdown of voltages in capacitors. The third asymptotic distribution of smallest values holds for the annual drougths of a river, for static and dynamic breaking strengths, for fatigue failure, wave heights and the stresses on ships and airplanes. In application to breaking strengths and fatigue failure, the probability function is interpreted as a lifetable function. The use of this distribution leads to a statistical estimation of the minimum life, i.e., the number of cycles before which no fracture occurs, and of the endurance limit, the stress so small that the specimen tested may survive an infinity of cycles. The estimation of these values, which cannot be observed, is decisive for the safety of structures. This author strongly doubts the validity of the estimations of the endurance limit based on non-statistical procedures.

SHORT BIBLIOGRAPHY

N.A. BARRICELLI : Les plus grands et les plus petits maxima ou minima annuels d'une variable climatique, Arch. Math. Natur. 46 (Oslo 1943).

J.J. DRONKERS : Approximate formulæ for the statistical distribution of extreme values, Biometrika, 45 (1958) pp. 447-470.

R.A. FISHER and L.C.H. TIPPETT : Limiting forms of the frequency distribution of the largest or smallest member of a sample, Proc. Cambridge Phil. Soc. 24, Pt. 2, pp. 180-190 (April 1928).

Y.C. FUNG : Statistical Aspects of Dynamic Loads, Inst. of Aeronautical Sciences, preprint 376, (1952) New York.

J. GEFFROY : Contributions à la théorie des valeurs extrêmes. Publ. Inst. Statist. Paris, 7 (1958), pp. 37-121, Vol. 8 (1959), pp. 123-184.

B. GNEDENKO : Sur la distribution limite du terme maximum d'une série aléatoire, Ann. Math. (2) 44, pp. 423-453, 1943.

E.J. GUMBEL : La durée extrême de la vie humaine, Hermann et Cie (520) Paris, 1937, Actualités scientifiques et industrielles.

E.J. GUMBEL : Statistical theory of extreme values and some practical applications, US Commerce Dept, National Bureau of Standards, Applied Math. Series no. 33 (1954).

—— Statistical theory of floods and droughts, J. Inst. Water Engineers, London, 12 (1958), pp. 157-184.

—— Communications on the statistical theory of floods and droughts, Journ. Inst. of Water Engineers, 13, no. 1, Feb 1959.

—— Statistical theory of breaking strength and fatigue failure, Bull. Intl. Stat. Inst. 38, pt. 3 Tokyo (1961).

—— Statistics of extremes, Columbia Univ. Press, 1958, 1960, New York (375 pages).

—— Extreme values in aeronautics, with P.G. Carlson, Journ. Aeronautical Sciences, 21, pp. 389-398, New York 1954.

M. FRECHET : Sur la loi de probabilité de l'écart maximum, Ann. Soc. Polonaise Math. Cracovie 6, pp. 93-116 (1927).

A.M. FREUDENTHAL and E.J. GUMBEL : Physical and statistical aspects of fatigue, Advanc. Appl. Mech. 4 (1956), pp. 117-158.

A.F. JENKINSON : The frequency distribution of the annual maximum (or minimum) values of meteorological elements. Q.J. Royal Meteor. Soc., 87, p. 158.

R. von MISES : La distribution de la plus grande de *n* valeurs, Revue Math. l'Union Interbalkanique 1, pp. 1-20 (Athens 1936).

National Bureau of Standards : Probability tables for the analysis of extreme-value data, Appl. Math. Ser. 22, US Dept. Commerce (1953).

K. PEARSON : Tables for statisticians and biometricians. Vol. 2, London (1931).

E.S. PEARSON and H.O. HARTLEY : Biometrika tables for statisticians, Cambridge Univ. Press (1954).

M. SHINOZUKA and A.M. FREUDENTHAL : Structure safety under conditions of ultimate load failure and fatigue WADD Tech. Report 61-177 (US Dept. of Commerce) 1961.

C.J. VELZ and J.J. GANNON : Drought flow of Michigan streams, Mich. Water Resources Comm. (1960).

WEIBULL, W. : Statistical evaluation of data from fatigue and creeprupture tests: fundamental concepts and general methods, WADD Tech. Rept. 59-400, pt. I (US Dept. of Commerce) 1959.

—— A statistical representation of fatigue failures in solids, Kungl. Tekniska Hogskolans Handlinger 27 (1949).