Chapter 6

Typical distribution functions in geophysics, hydrology and water resources

Demetris Koutsoyiannis
Department of Water Resources and Environmental Engineering
Faculty of Civil Engineering, National Technical University of Athens, Greece

Summary

In this chapter we describe four families of distribution functions that are used in geophysical and engineering applications, including engineering hydrology and water resources technology. The first includes the normal distribution and the distributions derived from this by the logarithmic transformation. The second is the gamma family and related distributions that includes the exponential distribution, the two- and three-parameter gamma distributions, the Log-Pearson III distribution derived from the last one by the logarithmic transformation and the beta distribution that is closely related to the gamma distribution. The third is the Pareto distribution, which in the last years tends to become popular due to its long tail that seems to be in accordance with natural behaviours. The fourth family includes the extreme value distributions represented by the generalized extreme value distributions of maxima and minima, special cases of which are the Gumbel and the Weibull distributions.

5.1 Normal Distribution and related transformations

5.1.1 Normal (Gaussian) Distribution

In the preceding chapters we have discussed extensively and in detail the normal distribution and its use in statistics and in engineering applications. Specifically, the normal distribution has been introduced in section 2.8, as a consequence of the central limit theorem, along with two closely related distributions, the χ^2 and the Student (or t), which are of great importance in statistical estimates, even though they are not used for the description of geophysical variables. The normal distribution has been used in chapter 3 to theoretically derive statistical estimates. In chapter 5 we have presented in detail the use of the normal distribution for the description of geophysical variables.

In summary, the normal distribution is a symmetric, two-parameter, bell shaped distribution. The fact that a normal variable X ranges from minus infinity to infinity contrasts the fact that hydrological variables are in general non-negative. This problem has been already discussed in detail in section 5.4.1. A basic characteristic of the normal distribution is that it is closed under addition or, else, a stable distribution. Consequently, the sum (and any linear combination) of normal variables is also a normal variable. Table 6.1 provides a concise summary of the basic mathematical properties and relations associated with the normal distribution, described in detail in previous chapters.

Table 6.1 Normal (Gaussian) distribution conspectus.

Probability density function	$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \mathrm{e}^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$
Distribution function	$F_X(x) = \int_{-\infty}^{x} f_X(s)\, ds$
Range	$-\infty < x < \infty$ (continuous)
Parameters	μ: location parameter (= mean) $\sigma > 0$: scale parameter (= standard deviation)
Mean	$\mu_X = \mu$
Variance	$\sigma_X^2 = \sigma^2$
Third central moment	$\mu_X^{(3)} = 0$
Fourth central moment	$\mu_X^{(4)} = 3\sigma^4$
Coefficient of skewness	$C_{s_X} = 0$
Coefficient of kurtosis	$C_{k_X} = 3$
Mode	$x_{\mathrm{p}} = \mu$
Median	$x_{0.5} = \mu$
Second L moment	$\lambda_X^{(2)} = \frac{\sigma}{\sqrt{\pi}}$
Third L moment	$\lambda_X^{(3)} = 0$
L coefficient of variation	$\tau_X^{(2)} = \frac{\sigma}{\sqrt{\pi}\,\mu}$
L skewness	$\tau_X^{(3)} = 0$
L kurtosis	$\tau_X^{(4)} = 0.1226$

Typical calculations

The most typical calculations are the calculation of the value $u = F_X(x_u)$ of the distribution function for a given x_u, or inversely, the calculation of the u-quantile of the variable, i.e. the calculation of x_u, when the probability u is known. The fact that the integral defining the normal distribution function (Table 6.1) does not have an analytical expression, creates difficulties in the calculations. A simple solution is the use of tabulated values of the standardized normal variable $z = (x - \mu) / \sigma$, which is a normal variable with zero mean and standard deviation equal to 1 (section 2.6.1 and Table A1 in Appendix). Thus, the calculation of the u-quantile (x_u) becomes straightforward by

$$x_u = \mu + z_u \sigma \tag{6.1}$$

where z_u, corresponding to $u = F_Z(z_u)$, is taken from Table A1. Conversely, for a given x_u, z_u is calculated by (6.1) and $u = F_Z(z_u)$ is determined from Table A1.

Several numerical approximations of the normal distribution function are given in the literature, which can be utilized to avoid use of tables (Press *et al.*, 1987; Stedinger *et al.*, 1993; Koutsoyiannis, 1997), whereas most common computer applications (e.g. spreadsheets*) include ready to use functions.

Parameter estimation

As we have seen in section 3.5, both the method of moments and the maximum likelihood result in the same estimates of the parameters of normal distribution, i.e.,

$$\mu = \bar{x}\,, \quad \sigma = s_X \tag{6.2}$$

We notice that s_X in (6.2) is the biased estimate of the standard deviation. Alternatively, the unbiased estimation of standard deviation is preferred sometimes. The method of L moments can be used as an alternative (see Table 6.1) to estimate the parameters based on the mean and the second L moment.

Standard error and confidence intervals of quantiles

In section 3.4.6 we defined the standard error and the confidence intervals of the quantile estimation and we presented the corresponding equations for the normal distribution. Summarising, the point estimate of the normal distribution u-quantile is

$$\hat{x}_u = \bar{x} + z_u s_x \tag{6.3}$$

the standard error of the estimation is

$$\varepsilon_u = \frac{s_X}{\sqrt{n}}\sqrt{1 + \frac{z_u^2}{2}} \tag{6.4}$$

and the corresponding confidence limits for confidence coefficient γ are

$$\hat{x}_{u_{1,2}} \approx (\bar{x} + z_u\, s_X)\ \pm z_{(1+\gamma)/2}\frac{s_X}{\sqrt{n}}\sqrt{1 + \frac{z_u^2}{2}} = \hat{x}_u \pm z_{(1+\gamma)/2}\frac{s_X}{\sqrt{n}}\sqrt{1 + \frac{z_u^2}{2}} \tag{6.5}$$

Normal distribution probability plot

As described in section 5.3.4, the normal distribution is depicted as a straight line in a normal probability plot. This depiction is equivalent to plotting the values of the variable x (in the vertical axis) versus and the standardized normal variate z (in the horizontal axis).

5.1.2 Two-parameter log-normal distribution

The two-parameter log-normal distribution results from the normal distribution using the transformation

$$y = \ln x \leftrightarrow x = \mathrm{e}^y \tag{6.6}$$

* In Excel, these functions are NormDist, NormInv, NormSDist and NormSInv.

Thus, the variable X has a two-parameter log-normal distribution if the variable Y has normal distribution $N(\mu_Y, \sigma_Y)$. Table 6.2 summarizes the mathematical properties and relations associated with the two-parameter log-normal distribution.

Table 6.2 Two-parameter log-normal distribution conspectus.

Probability density function	$f_X(x) = \dfrac{1}{x\sqrt{2\pi}\sigma_Y} \mathrm{e}^{-\frac{1}{2}\left(\frac{\ln x - \mu_Y}{\sigma_Y}\right)^2}$
Distribution function	$F_X(x) = \int_0^x f_X(s)\, ds$
Range	$0 < x < \infty$ (continuous)
Parameters	μ_Y: scale parameter $\sigma_Y > 0$: shape parameter
Mean	$\mu_X = \mathrm{e}^{\mu_Y + \frac{\sigma_Y^2}{2}}$
Variance	$\sigma_X^2 = \mathrm{e}^{2\mu_Y + \sigma_Y^2}\left(\mathrm{e}^{\sigma_Y^2} - 1\right)$
Third central moment	$\mu_X^{(3)} = \mathrm{e}^{3\mu_Y + \frac{3\sigma_Y^2}{2}}\left(\mathrm{e}^{3\sigma_Y^2} - 3\mathrm{e}^{\sigma_Y^2} + 2\right)$
Coefficient of skewness	$C_{v_X} = \sqrt{\mathrm{e}^{\sigma_Y^2} - 1}$
Coefficient of kurtosis	$C_{s_X} = 3C_{v_X} + C_{v_X}^3$
Mode	$x_p = \mathrm{e}^{\mu_Y - \sigma_Y^2}$
Median	$x_{0.5} = \mathrm{e}^{\mu_Y}$

A direct consequence of the logarithmic transformation (6.6) is that the variable X is always positive. In addition, it results from Table 6.2 that the distribution has always positive skewness and that its mode is different from zero. Thus, the shape of the probability density function is always bell-shaped and positively skewed. These basic attributes of the log-normal distribution are compatible with observed properties of many geophysical variables, and therefore it is frequently used in geophysical applications. It can be easily shown that the product of two variables having a two-parameter log-normal distribution, has also a two-parameter log-normal distribution. This property, combined with the central limit theorem and taking into account that in many cases the variables can be considered as a product of several variables instead of a sum, has provided theoretical grounds for the frequent use of the distribution in geophysics.

Typical calculations

Typical calculations of the log-normal distribution are based on the corresponding calculations of the normal distribution. Thus, combining equations (6.1) and (6.6) we obtain

$$y_u = \mu_Y + z_u \sigma_Y \Leftrightarrow x_u = \mathrm{e}^{\mu_Y + z_u \sigma_Y} \tag{6.7}$$

where z_u is the u-quantile of the standardized normal variable.

Parameter estimation

Using the equations of Table 6.2, the method of moments results in:

$$\sigma_Y = \sqrt{\ln\left(1 + s_X^2 / \overline{x}^2\right)}, \quad \mu_Y = \ln \overline{x} - \sigma_Y^2 / 2 \tag{6.8}$$

Parameter estimation using the maximum likelihood method gives (e.g. Kite, 1988, p. 57)

$$\mu_Y = \sum_{i=1}^{n} \ln x_i / n = \overline{y}, \quad \sigma_Y = \sqrt{\sum_{i=1}^{n} (\ln x_i - \mu_Y)^2 / n} = s_Y \tag{6.9}$$

We observe that the two methods differ not only in the resulted estimates, but also in that they are based on different sample characteristics. Namely, the method of moments is based on the mean and the (biased) standard deviation of the variable X while the maximum likelihood method is based on the mean and the (biased) standard deviation of the logarithm of the variable X.

Standard error and confidence intervals of quantiles

Provided that the maximum likelihood method is used to estimate the parameters of the log-normal distribution, the point estimate of the u-quantile of y and x is then given by

$$\hat{y}_u = \ln(\hat{x}_u) = \overline{y} + z_u s_Y \Rightarrow \hat{x}_u = \mathrm{e}^{\overline{y} + z_u s_Y} \tag{6.10}$$

where z_u is the u-quantile of the standard normal distribution. The square of the standard error of the Y estimate is given by:

$$\varepsilon_Y^2 = \mathrm{Var}(\hat{Y}_u) = \mathrm{Var}(\ln \hat{X}_u) = \frac{s_Y^2}{n}\left(1 + \frac{z_u^2}{2}\right) \tag{6.11}$$

Combining these equations we obtain the following approximate relationship which gives the confidence intervals of x_u for confidence level γ

$$\hat{x}_{u_{1,2}} \approx \exp\left[(\overline{y} + z_u s_Y) \pm z_{(1+\gamma)/2} \frac{s_Y}{\sqrt{n}} \sqrt{1 + \frac{z_u^2}{2}}\right] = \hat{x}_u \exp\left[\pm z_{(1+\gamma)/2} \frac{s_Y}{\sqrt{n}} \sqrt{1 + \frac{z_u^2}{2}}\right] \tag{6.12}$$

where $z_{(1+\gamma)/2}$ is the $[(1+\gamma)/2]$-quantile of the standard normal distribution. When the parameter estimation is based in the method of moments, the standard error and the corresponding confidence intervals are different (see Kite 1988, p. 60).

Log-normal distribution probability plot

The normal distribution probability plot can be easily transformed in order for the log-normal distribution to be depicted as a straight line. Specifically, a logarithmic vertical axis has to be used. This depiction is equivalent to plotting the logarithm of the variable, $\ln x$, (in the vertical axis) versus the standard normal variate (in the horizontal axis).

Numerical example

Table 6.3 lists the observations of monthly runoff of the Evinos river basin, central-western Greece, upstream of the hydrometric gauge at Poros Reganiou, for the month of January. We wish to fit the two-parameter log-normal distribution to the data and estimate the 50-year discharge.

Table 6.3 Observed sample of January runoff volume (in hm^3) at the hydrometric station of Poros Riganiou of the Evinos river.

Hydrological year	Runoff	Hydrological year	Runoff	Hydrological year	Runoff
1970-71	102	1977-78	121	1984-85	178
1971-72	74	1978-79	317	1985-86	185
1972-73	78	1979-80	213	1986-87	101
1973-74	48	1980-81	111	1987-88	57
1974-75	31	1981-82	82	1988-89	24
1975-76	48	1982-83	61	1989-90	22
1976-77	114	1983-84	133	1990-91	51

The sample mean is

$$\bar{x} = \sum x / n = 102.4 \text{ hm}^3$$

The standard deviation (biased estimate) is

$$s_X = \left(\sum x^2 / n - \bar{x}^2\right)^{1/2} = 70.4 \text{ hm}^3$$

and the coefficient of variation

$$\hat{C}_{\mathrm{v}X} = s_X / \bar{x} = 70.4 / 102.4 = 0.69$$

The skewness coefficient (biased estimate) is

$$\hat{C}_{\mathrm{s}X} = 1.4$$

These coefficients of variation and skewness suggest a large departure from the normal distribution.

The method of moments results in

$$\sigma_Y = \sqrt{\ln(1 + s_X^2 / \bar{x}^2)} = 0.622, \quad \mu_Y = \ln \bar{x} - \sigma_Y^2 / 2 = 4.435$$

whereas the maximum likelihood estimates are

$$\mu_Y = \sum \ln x / n = 4.404, \quad \sigma_Y = \sqrt{\sum (\ln x)^2 / n - \mu_Y^2} = 0.687$$

The 50-year discharge can be estimated from $x_u = \exp(\mu_Y + z_u\,\sigma_Y)$ where $u = 1 - 1/50 = 0.98$ and $z_u = 2.054$ (Table A1). Using the parameters estimated by the method of moments we obtain $x_{0.98} = 302.7$ hm^3, while using the maximum likelihood parameter estimates we get $x_{0.98} = 335.1$. In the latter case the 95% confidence interval for that value is (based on (6.12), for $z_u = 2.054$ and $z_{(1+\gamma)/2} = 1.96$):

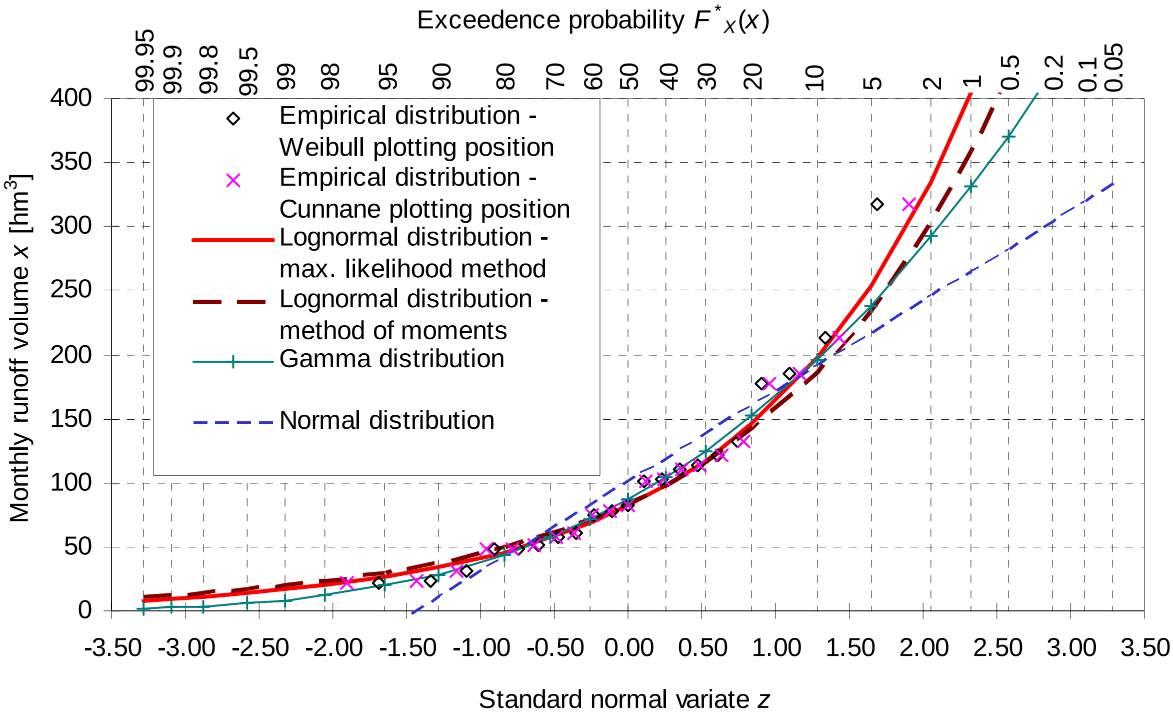

Fig. 6.1 Alternative empirical and theoretical distribution functions of the January runoff at Poros Riganiou (normal probability plot).

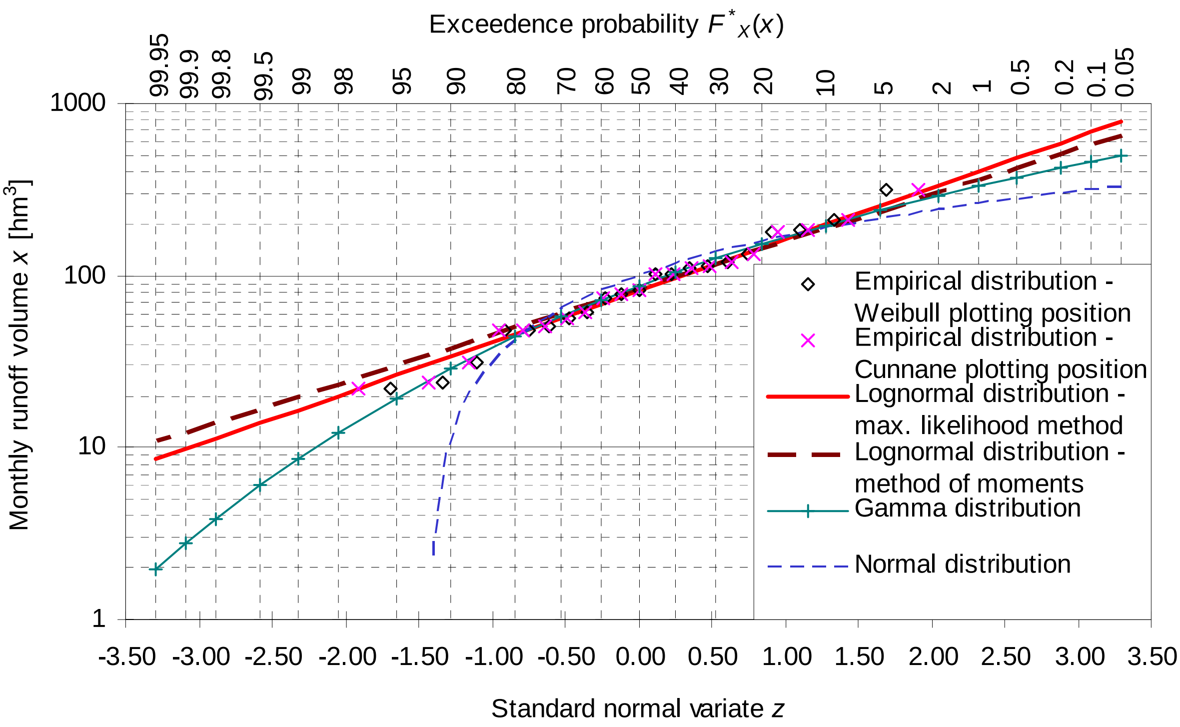

Fig. 6.2 Alternative empirical and theoretical distribution functions of the January runoff at Poros Riganiou (lognormal probability plot).

$$\hat{x}_{u_{1,2}} \approx \exp\left[4.404 + 2.054 \times 0.687 \pm 1.96 \times \frac{0.687}{\sqrt{21}} \times \sqrt{1 + \frac{2.054^2}{2}}\right]$$

$$= \exp(5.815 \pm 0.518) = \begin{cases} 562.8 \\ 199.7 \end{cases}$$

The huge width of the confidence interval reflects a poor reliability of the prediction of the 50-year January runoff. The reduction of the uncertainty would be made possible only by a substantially larger sample.

To test the appropriateness of the log-normal distribution we can use the χ^2 test (see section 5.5.1). As an empirical alternative, we depict in Fig. 6.1 and Fig. 6.2 comparisons of the empirical distribution function and the fitted log-normal theoretical distribution functions, on normal probability plot and on log-normal probability plot, respectively. For the empirical distribution we have used two plotting positions, the Weibull and the Cunnane (Table 5.8). Both log-normal distribution plots, resulted from the methods and the maximum likelihood are shown in the figures. Clearly, the maximum likelihood method results in a better fit in the region of small exceedence probabilities. For comparison we have also plotted the normal distribution, which apparently does not fit well to the data, and the Gamma distribution (see section 5.2.2).

5.1.3 Three-parameter log-normal (Galton) distribution

A combination of the normal distribution and the modified logarithmic transformation

$$y = \ln(x - \zeta) \Leftrightarrow x = \zeta + \mathrm{e}^y \tag{6.13}$$

results in the three-parameter log-normal distribution or the Galton distribution. This distribution has an additional parameter, compared to the two-parameter log-normal, the location parameter ζ, which is the lower limit of the variable. This third parameter results in a higher flexibility of the distribution fit. Specifically, if the method of moments is used to fit the distribution, the third parameter makes possible the preservation of the coefficient of skewness. Table 6.4 summarizes the basic mathematical properties and equations associated with the three-parameter log-normal distribution.

Typical calculations

The three-parameter log-normal distribution, can be handled in a similar manner with the two-parameter log-normal distribution according to the following relationship

$$y_u = \mu_Y + z_u \sigma_Y \Leftrightarrow x_u = \zeta + \mathrm{e}^{\mu_Y + z_u \sigma_Y} \tag{6.14}$$

where z_u is the u-quantile of the standard normal distribution.

Parameter estimation

Using the equations of Table 6.4 for the method of moments, and after algebraic manipulation we obtain the following relationships that estimate the parameter σ_Y.

$$\sigma_Y = \sqrt{\ln\left(1+\phi^2\right)} \tag{6.15}$$

where

$$\phi = \frac{1-\omega^{2/3}}{\omega^{1/3}}, \quad \omega = \frac{-\hat{C}_{s_X} + \sqrt{\hat{C}_{s_X}^2 + 4}}{2} \tag{6.16}$$

Table 6.4 Three-parameter log-normal distribution conspectus

Probability density function	$f_X(x) = \dfrac{1}{(x-\zeta)\sqrt{2\pi}\sigma_Y} \mathrm{e}^{-\frac{1}{2}\left(\frac{\ln(x-\zeta)-\mu_Y}{\sigma_Y}\right)^2}$
Distribution function	$F_X(x) = \int_c^x f_X(s)\,ds$
Range	$\zeta < x < \infty$ (continuous)
Parameters	ζ: location parameter μ_Y: scale parameter $\sigma_Y > 0$: shape parameter
Mean	$\mu_X = \zeta + \mathrm{e}^{\mu_Y + \frac{\sigma_Y^2}{2}}$
Variance	$\sigma_X^2 = \mathrm{e}^{2\mu_Y + \sigma_Y^2}\left(\mathrm{e}^{\sigma_Y^2} - 1\right)$
Third central moment	$\mu_X^{(3)} = \mathrm{e}^{3\mu_Y + \frac{3\sigma_Y^2}{2}}\left(\mathrm{e}^{3\sigma_Y^2} - 3\mathrm{e}^{\sigma_Y^2} + 2\right)$
Coefficient of skewness	$C_{s_X} = 3\left(\mathrm{e}^{\sigma_Y^2} - 1\right)^{1/2} + \left(\mathrm{e}^{\sigma_Y^2} - 1\right)^{3/2}$
Mode	$x_p = \zeta + \mathrm{e}^{\mu_Y - \sigma_Y^2}$
Median	$x_{0.5} = \zeta + \mathrm{e}^{\mu_Y}$

The other two parameters of the distribution can be calculated from

$$\mu_Y = \ln(s_X/\varphi) - \sigma_Y^2/2 \qquad \zeta = \bar{x} - \frac{s_X}{\varphi} \tag{6.17}$$

The maximum likelihood method is based on the following relationships (e.g. Kite, 1988, p. 74)

$$\mu_Y = \sum_{i=1}^{n} \ln\left(x_i - c\right)/n, \quad \sigma_Y^2 = \sum_{i=1}^{n}\left[\ln\left(x_i - c\right) - \mu_Y\right]^2/n \tag{6.18}$$

$$\left(\mu_Y - \sigma_Y^2\right)\sum_{i=1}^{n}\frac{1}{x_i - c} = \sum_{i=1}^{n}\frac{\ln\left(x_i - c\right)}{x_i - c} \tag{6.19}$$

that can be solved only numerically.

The estimation of confidence intervals for the three-parameter log-normal distribution is complicated. The reader can consult Kite (1988, p. 77).

5.2 The Gamma family and related distribution functions

5.2.1 Exponential distribution

A very simple yet useful distribution is the exponential. Its basic characteristics are summarized in Table 6.5.

Table 6.5 Exponential distribution conspectus

Probability density function	$f_X(x) = \dfrac{\mathrm{e}^{-\frac{x-\zeta}{\lambda}}}{\lambda}$
Distribution function	$F_X(x) = 1 - \mathrm{e}^{-\frac{x-\zeta}{\lambda}}$
Variable range	$\zeta < x < \infty$ (continuous)
Parameters	ζ: location parameter $\lambda > 0$: scale parameter
Mean	$\mu_X = \zeta + \lambda$
Variance	$\sigma_X^2 = \lambda^2$
Third central moment	$\mu_X^{(3)} = \lambda^3$
Fourth central moment	$\mu_X^{(4)} = 9\lambda^4$
Coefficient of variation	$C_{v_X} = \dfrac{\lambda}{\zeta + \lambda}$
Coefficient of skewness	$C_{s_X} = 2$
Coefficient of kurtosis	$C_{k_X} = 9$
Mode	$x_p = \zeta$
Median	$x_{0.5} = \zeta + \lambda \ln 2$
Second L moment	$\lambda_X^{(2)} = \lambda/2$
Third L moment	$\lambda_X^{(3)} = \lambda/6$
Fourth L moment	$\lambda_X^{(4)} = \lambda/12$
L coefficient of variation	$\tau_X^{(2)} = \dfrac{\lambda}{2(\lambda + \zeta)}$
L skewness	$\tau_X^{(3)} = 1/3$
L kurtosis	$\tau_X^{(4)} = 1/6$

In its simplesr form, as we have already seen in section 2.5.5, the exponential distribution has only one parameter, the location parameter λ (the second parameter ζ is 0). The probability density function of the exponential distribution is a monotonically decreasing function (it has an inverse J shape).

As we have already seen (section 2.5.5), the exponential distribution can be used to describe non-negative geophysical variables at a fine time scale (e.g. hourly or daily rainfall depths). In addition, a theorem in probability theory states that intervals between random points in time, have exponential distribution. Application of this theorem in geophysics suggests that, for instance, the time intervals between rainfall events have exponential distribution. This is verified only as a rough approximation. The starting times of rainfall events cannot be regarded as random points in time; rather, a clustering behaviour is evident, which is related to some dependence in time (Koutsoyiannis, 2006). Moreover, the duration of rainfall events and the total rainfall depth in an event have been frequently assumed to have exponential distribution. Again this is just a rough approximation (Koutsoyiannis, 2005).

5.2.2 Two-parameter Gamma distribution

The two-parameter Gamma distribution is one of the most commonly used in geophysics and engineering hydrology. Its basic characteristics are given in Table 6.6.

Table 6.6 Two-parameter Gamma distribution conspectus.

Probability density function	$f_X(x) = \dfrac{1}{\lambda^\kappa \Gamma(\kappa)} x^{\kappa-1} \mathrm{e}^{-x/\lambda}$
Distribution function	$F_X(x) = \int_0^x f_X(s)\,ds$
Range	$0 < x < \infty$ (continuous)
Parameters	$\lambda > 0$: scale parameter $\kappa > 0$: shape parameter
Mean	$\mu_X = \kappa\lambda$
Variance	$\sigma_X^2 = \kappa\lambda^2$
Third central moment	$\mu_X^{(3)} = 2\kappa\lambda^3$
Fourth central moment	$\mu_X^{(4)} = 3\kappa(\kappa+2)\lambda^4$
Coefficient of variation	$C_{v_X} = \dfrac{1}{\sqrt{\kappa}}$
Coefficient of skewness	$C_{s_X} = \dfrac{2}{\sqrt{\kappa}} = 2C_{v_X}$
Coefficient of kurtosis	$C_{k_X} = 3 + \dfrac{6}{\kappa} = 3 + 6C_{v_X}^2$
Mode	$x_\mathrm{p} = (\kappa - 1)\,\lambda$ (for $\kappa \geq 1$) $x_\mathrm{p} = 0$ (for $\kappa \leq 1$)

Similar to the two-parameter log-normal distribution, the Gamma distribution is positively skewed and is defined only for nonnegative values of the variable. These characteristics make the Gamma distribution compatible with several geophysical variables, including monthly and annual flows and precipitation depths.

The Gamma distribution has two parameters, the scale parameter λ and the shape parameter κ. For κ = 1 the distribution is identical with the exponential, which is a special case of Gamma. For $\kappa > 1$ the probability density function is bell-shaped, whereas for $\kappa < 1$ its shape becomes an inverse J, with an infinite ordinate at $x = 0$. For large κ values (above 15-30) the Gamma distribution approaches the normal.

The Gamma distribution, similar to the normal, is closed under addition, but only when the added variables are stochastically independent and have the same scale parameter. Thus, the sum of two independent variables that have Gamma distribution with common scale parameter λ, has also a Gamma distribution.

The χ^2 distribution, which has been discussed in section 2.10.4, is a special case of the Gamma distribution.

Typical calculations

Similar to the normal distribution, the integral in the Gamma distribution function does not have an analytical expression thus causing difficulties in calculations. A simple solution is to tabulate the values of the standardized variable $k = (x - \mu_X) / \sigma_X$, where μ_X and σ_X is the mean value and standard deviation of X, respectively. Such tabulations are very common in statistics books; one is provided in Table A4 in Appendix. Each column of this table corresponds to a certain value of κ (or, equivalently, to a certain skewness coefficient value $C_{sX} = 2 / \sqrt{\kappa} = 2\sigma_X / \bar{x}$).The u-quantile (x_u) is then given by

$$x_u = \mu_X + k_u \sigma_X \tag{6.20}$$

where k_u is read from tables for the specified value of $u = F_K(k_u)$. Conversely, for given x_u, the k_u value can be calculated from (6.1) and then $u = F_K(k_u)$ is taken from tables (interpolation in a column or among adjacent columns may be necessary).

Several numerical approaches can be found in literature in order to avoid the use of tables (Press *et al.*, 1987; Stedinger *et al.*, 1993; Koutsoyiannis, 1997) whereas most common computer applications (e.g. spreadsheets*) include ready to use functions.

Parameter estimation

The implementation of the method of moments results in the following simple estimates of the two Gamma distribution parameters:

$$\kappa = \frac{\bar{x}^2}{s_X^2}, \quad \lambda = \frac{s_X^2}{\bar{x}} \tag{6.21}$$

Parameter estimation based on the maximum likelihood method is more complicated. It is based in the solution of the equations (cf. e.g. Bobée and Ashkar, 1991)

$$\ln \kappa - \psi(\kappa) = \ln \bar{x} - \frac{1}{n}\sum_{i=1}^{n} \ln x_i, \quad \lambda = \frac{\bar{x}}{\kappa} \tag{6.22}$$

* In Excel, these functions are GammaDist and GammaInv.

where $\psi(\kappa) = \mathrm{d} \ln \Gamma(\kappa) / \mathrm{d}\kappa$ is the so-called Digamma function (derivative of the logarithm of Gamma function).

Standard error and confidence intervals of quantiles

A point estimate of the u-quantile of Gamma distribution is given by

$$\hat{x}_u = \bar{x} + k_u s_X \tag{6.23}$$

If the method of moments is used to estimate the parameters the square of standard error of the estimate is (Bobée and Ashkar, 1991, p. 50)

$$\varepsilon_u^2 = \frac{s_X^2}{n}\left[\left(1 + k_u C_{v_X}\right)^2 + \frac{1}{2}\left(k_u + 2C_{v_X}\frac{\partial k_u}{\partial C_{s_X}}\right)^2\left(1 + C_{v_X}\right)^2\right] \tag{6.24}$$

In a first rough approximation, the term $\partial k_u / \partial C_{sX}$ can be omitted, leading to the simplification

$$\varepsilon_u^2 = \frac{s_X^2}{n}\left[1 + 2C_{v_X} k_u + \frac{1}{2}\left(1 + 3C_{v_X}^2\right)k_u^2\right] \tag{6.25}$$

Thus, an approximation of the confidence limits for confidence coefficient γ is

$$\hat{x}_{u_{1,2}} \approx (\bar{x} + k_u s_X) \pm z_{(1+\gamma)/2}\frac{s_X}{\sqrt{n}}\sqrt{1 + 2C_{v_X} k_u + \frac{1}{2}\left(1 + 3C_{v_X}^2\right)k_u^2} \tag{6.26}$$

The maximum likelihood method results in more complicated calculations of the confidence intervals. The interested reader may consult Bobée and Ashkar (1991, p. 46).

Gamma distribution probability plot

It is not possible to construct a probability paper that depicts any Gamma distribution as straight line. It is feasible, though, to create a Gamma probability paper for a specified shape parameter κ. Clearly, this is not practical, and thus the depiction of Gamma distribution is usually done on normal probability paper or on Weibull probability paper (see below). In that case obviously the distribution is not depicted as a straight line but as a curve.

Numerical example

We wish to fit a two-parameter Gamma distribution to the sample of January runoff of the river Evinos upstream of the hydrometric station of Poros Riganiou and to determine the 50-year runoff (sample in Table 6.3).

The sample mean value is 102.4 hm^3 and the sample standard deviation is 70.4 hm^3; using the method of moments we obtain the following parameter estimates:

$$\kappa = 102.4^2 / 70.4^2 = 2.11,\ \lambda = 70.4^2 / 102.4 = 48.4 \text{ hm}^3.$$

For return period $T = 50$ or equivalently for probability of non-exceedence $F = 0.98 = u$ we determine the quantile x_u either by an appropriate computer function or from tabulated standardized quantile values (Table A4); we find $k_{0.98} = 2.70$ and

$$x_u = 102.4 + 2.70 \times 70.4 = 292.5 \text{ hm}^3$$

Likewise, we can calculate a series of quantiles, thus enabling the depiction of the fitted Gamma distribution. This has been done in Fig. 6.1 (in normal probability plot) and in Fig. 6.2 (in log-normal probability plot) in comparison with other distributions. We observe that in general the Gamma distribution fit is close to those of the log-normal distribution; in the region of small exceedence probabilities the log-normal distribution provides a better fit.

To determine the 95% confidence intervals for the 50-year discharge we use the approximate relationship (6.26), which for $z_{(1+\gamma)/2} = 1.96$, $k_u = 2.70$ and $C_{vX} = 0.69$ results in

$$\begin{aligned}\hat{x}_{u_{1,2}} &\approx 292.5 \pm 1.96 \times \frac{70.4}{\sqrt{21}} \times \sqrt{1 + 2 \times 0.69 \times 2.70 + \frac{1}{2}\left(1 + 3 \times 0.69^2\right) \times 2.70^2} \\ &\approx 292.5 \pm 110.9 = \begin{cases} 403.4 \\ 181.6 \end{cases}\end{aligned}$$

5.2.3 Three-parameter Gamma distribution (Pearson III)

The addition of a location parameter (ζ) to the two-parameter Gamma distribution, results in the three-parameter Gamma distribution or the so-called Pearson type III (Table 6.7).

Table 6.7 Pearson type III distribution conspectus.

Probability density function	$f_X(x) = \frac{1}{\lambda^\kappa \Gamma(\kappa)} (x - \zeta)^{\kappa - 1} e^{-(x-\zeta)/\lambda}$
Distribution function	$F_X(x) = \int_c^x f_X(s)\, ds$
Range	$\zeta < x < \infty$ (continuous)
Parameters	ζ: location parameter $\lambda > 0$: scale parameter $\kappa > 0$: shape parameter
Mean	$\mu_X = c + \kappa\lambda$
Variance	$\sigma_X^2 = \kappa\lambda^2$
Third central moment	$\mu_X^{(3)} = 2\kappa\lambda^3$
Fourth central moment	$\mu_X^{(4)} = 3\kappa(\kappa + 2)\lambda^4$
Coefficient of skewness	$C_{s_X} = \frac{2}{\sqrt{\kappa}}$
Coefficient of kurtosis	$C_{k_X} = 3 + \frac{6}{\kappa}$
Mode	$x_p = \zeta + (\kappa - 1)\lambda$ (for $\kappa \geq 1$) $x_p = \zeta$ (for $\kappa \leq 1$)

The location parameter ζ, which is the lower limit of the variable, enables a more flexible fit to the data. Thus, if we use the method of moments to fit the distribution, the third parameter permits the preservation of the coefficient of skewness.

The basic characteristics are similar to those of the two-parameter Gamma distribution. Typical calculations are also based in equation (6.20). In contrast, the equations used for parameter estimation differ. Thus, the method of moments results in

$$\kappa = \frac{4}{\hat{C}^2_{s_X}}, \quad \lambda = \frac{s_X}{\sqrt{\kappa}}, \quad \zeta = \bar{x} - \kappa\lambda \tag{6.27}$$

The maximum likelihood method results in more complicated equations. The interested reader may consult Bobée and Ashkar (1991, p. 59) and Kite (1988, p. 117) who also provide formulae to estimate the standard error and confidence intervals of distribution quantiles.

5.2.4 Log-Pearson III distribution

The Log-Pearson III results from the Pearson type III distribution and the transformation

$$y = \ln x \Leftrightarrow x = \mathrm{e}^y \tag{6.28}$$

Thus, the random variable X has Log-Pearson III distribution if the variable Y has Pearson III. Table 6.8 summarizes the basic mathematical relationships for the Log-Pearson III distribution.

Table 6.8 Log Pearson III distribution conspectus.

Probability density function	$f_X(x) = \dfrac{1}{x\lambda^\kappa \Gamma(\kappa)}(\ln x - \zeta)^{\kappa-1}\mathrm{e}^{-(\ln x - \zeta)/\lambda}$
Distribution function	$F_X(x) = \int_{e^\zeta}^{x} f_X(s)ds$
Range	$\mathrm{e}^\zeta < x < \infty$ (continuous)
Parameters	ζ: scale parameter $\lambda > 0$: shape parameter $\kappa > 0$: shape parameter
Mean	$\mu_X = \mathrm{e}^\zeta \left(\dfrac{1}{1-\lambda}\right)^\kappa, \quad \lambda < 1$
Variance	$\sigma_X^2 = \mathrm{e}^{2\zeta}\left[\left(\dfrac{1}{1-2\lambda}\right)^\kappa - \left(\dfrac{1}{1-\lambda}\right)^{2\kappa}\right], \quad \lambda < 1/2$
Raw moments of order r	$m_X^{(r)} = \mathrm{e}^{r\zeta}\left(\dfrac{1}{1-r\lambda}\right)^\kappa, \quad \lambda < 1/r$

The probability density function of the Log-Pearson III distribution can take several shapes like bell-, inverse-J-, U-shape and others. From Table 6.8 we can be conclude that the rth moment tends to infinity for $\lambda = 1/r$ and does not exist for greater λ. This shows that the distribution has a long tail (see section 2.5.6), which has made it a popular choice in engineering hydrology. Thus, it has been extensively used to describe flood discharges; in the USA the Log-Pearson III has been recommended by national authorities as the distribution of choice for floods.

Typical calculations

Typical calculations for the Log-Pearson III are based on those related to the Pearson III. Hence, a combination of the equations (6.20) and (6.28) gives

$$y_u = \mu_Y + k_u \sigma_Y \Leftrightarrow x_u = \mathrm{e}^{\mu_Y + k_u \sigma_Y} \tag{6.29}$$

where the standard Gamma variate k_u can be determined either from tables or numerically as described in section 5.2.2.

Parameter estimation

The parameter estimation by either the method of moments or the maximum likelihood is quite complicated (Bobée and Ashkar, 1991, p. 85· Kite, 1988, p. 138). Here we present a simpler *method of moments of logarithms*: According to this method we calculate the values $y_i = \ln x_i$ from the available sample and then we calculate the statistics of the values y_i. Finally, we apply the equations resulted from the method of moments for the variable Y, thus we have

$$\kappa = \frac{4}{\hat{C}_{sY}^2}, \quad \lambda = \frac{s_Y}{\sqrt{\kappa}}, \quad \zeta = \bar{y} - \kappa\lambda \tag{6.30}$$

As in the case of the Pearson III distribution, the estimation of the confidence intervals is pretty complicated.

Log-Pearson III probability plot

It is not possible to construct a probability paper that depicts any Log-Pearson III distribution as a straight line. Of course it is possible to make a probability paper for a specified value of the shape parameter κ but this is impractical. Thus, the depiction of the Log-Pearson III distribution is usually done on Log-normal probability paper or on Gumbel probability paper (see below). In that case the distribution is not depicted as a straight line but as a curve.

5.2.5 Two-parameter Beta distribution

The Beta distribution is an important distribution of the probability theory and has been extensively used as a conditional distribution and in Bayesian statistics. Moreover, the two-parameter Beta distribution is related to the Gamma distribution. Specifically, if X and Y are independent random variables with distributions Gamma(α, θ) and Gamma(β, θ) respectively (where Gamma(α, θ) denotes a Gamma distribution with shape parameter α and scale parameter θ), then the random variable $X / (X + Y)$ has Beta(α, β) distribution. A basic property of the Beta distribution is that the variable ranges from 0 to 1, contrary to the other distributions examined that are unbounded from above. The Beta distribution is frequently used in geophysics for doubly bounded variables, e.g. relative humidity.

The Beta distribution has two shape parameters, α and β whereas an additional scale parameter could be easily added. Depending on the parameter values, the probability density function of the Beta distribution can take a plethora of shapes. Specifically, for $\alpha = \beta = 1$ it becomes identical to the uniform distribution, while for $\alpha = 1$ and $\beta = 2$ (or $\alpha = 2$ and $\beta = 1$) it is identical to the negatively (positively) skewed triangular distribution. If $\alpha < 1$ (or $\beta < 1$) the probability density function is infinite at point $x = 0$ ($x = 1$). If $\alpha > 1$ and $\beta > 1$ the Beta probability density function is bell shaped. Table 6.9 summarizes the basic properties of the Beta distribution.

Table 6.9 Two-parameter Beta distribution conspectus.

Probability density function	$f_X(x) = \dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1}$
Distribution function	$F_X(x) = \int_0^x f_X(s)\, ds$
Variable range	$0 < x < 1$ (continuous)
Parameters	$\alpha, \beta > 0$: shape parameters
Mean	$\mu_X = \dfrac{\alpha}{\alpha+\beta}$
Variance	$\sigma_X^2 = \dfrac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$
Third raw moment	$m_X^{(3)} = \dfrac{\alpha(\alpha+1)(\alpha+2)}{(\alpha+\beta)(\alpha+\beta+1)(\alpha+\beta+2)}$
Coefficient of variation	$C_{v_X} = \sqrt{\dfrac{\beta}{\alpha(\alpha+\beta+1)}}$
Mode	$x_{\mathrm{p}} = \dfrac{\alpha-1}{\alpha+\beta-2}$ (for $\alpha, \beta > 1$)

5.3 Generalized Pareto distribution

The Pareto distribution was introduced by the Italian economist Vilfredo Pareto to describe the allocation of wealth among individuals since it seemed to describe well the fact that a larger portion of the wealth of a society is owned by a smaller percentage of the people. Its original form is expressed by the power-law equation

$$P\{X > x\} = \left(\frac{x}{\lambda}\right)^{-\frac{1}{\kappa}} \tag{6.31}$$

where λ is a (necessarily positive) minimum value of x ($x > \lambda$) and κ is a (positive) shape parameter. A generalized form, the so-called generalized Pareto distribution, in which a location parameter ζ independent of the scale parameter λ has been added, has been used in geophysics. Its basic characteristics are summarized in Table 6.10. Similar to the Log-Pearson III, the generalized Pareto distribution has a long tail. Indeed, as can be observed in Table 6.10, its third, second and first moments diverge (become infinite) for $\kappa \geq 1/3$, $\kappa \geq 1/2$ and $\kappa \geq 1$, respectively. For its long tail the distribution recently tends to replace short-tail distributions such as the Gamma distribution in modelling fine-time-scale rainfall and river discharge (Koutsoyiannis, 2004a,b, 2005). Since the analytical expression of the distribution function is very simple (Table 6.10) no tables or complicated numerical procedures are needed to handle it. Application of l'Hôpital's rule for κ = 0 results precisely in the exponential distribution, which thus can be derived as a special case of the Pareto distribution.

Table 6.10 Generalized Pareto distribution conspectus.

Probability density function	$f_X(x) = \frac{1}{\lambda}\left(1 + \kappa\frac{x-\zeta}{\lambda}\right)^{-\frac{1}{\kappa}-1}$
Distribution function	$F_X(x) = 1 - \left(1 + \kappa\frac{x-\zeta}{\lambda}\right)^{-\frac{1}{\kappa}}$
Range	For $\kappa > 0$, $\zeta \le x < \infty$ For $\kappa < 0$, $\zeta \le x < \zeta - \lambda / \kappa$ (continuous)
Parameters	ζ: location parameter $\lambda > 0$: scale parameter κ: shape parameter
Mean	$\mu_X = \zeta + \frac{\lambda}{1-\kappa}$
Variance	$\sigma_X^2 = \frac{\lambda^2}{(1-\kappa)^2(1-2\kappa)}$
Third central moment	$\mu_X^{(3)} = \frac{2\lambda^3(1+\kappa)}{(1-\kappa)^3(1-2\kappa)(1-3\kappa)}$
Skewness coefficient	$C_{s_X} = \frac{2(1+\kappa)\sqrt{1-2\kappa}}{1-3\kappa}$
Mode	$x_{\mathrm{p}} = \zeta$
Median	$x_{0.5} = \zeta + \frac{\lambda}{\kappa}\left(1 - 0.5^{-\kappa}\right)$
Second L moment	$\lambda_X^{(2)} = \frac{\lambda}{(1-\kappa)(2-\kappa)}$
Third L moment	$\lambda_X^{(3)} = \frac{\lambda(1+\kappa)}{(1-\kappa)(2-\kappa)(3-\kappa)}$
Fourth L moment	$\lambda_X^{(4)} = \frac{\lambda(1+\kappa)(2+\kappa)}{(1-\kappa)(2-\kappa)(3-\kappa)(4-\kappa)}$
L coefficient of variation	$\tau_X^{(2)} = \frac{\lambda}{[\zeta(1-\kappa)+\lambda](2-\kappa)}$
L skewness	$\tau_X^{(3)} = \frac{1+\kappa}{3-\kappa}$
L kurtosis	$\tau_X^{(4)} = \frac{(1+\kappa)(2+\kappa)}{(3-\kappa)(4-\kappa)}$

5.4 Extreme value distributions

It can be easily shown that, given a number n of independent identically distributed random variables $Y_1,\ldots,Y_n$, the largest (in the sense of a specific realization) of them (more precisely, the largest order statistic), i.e.:

$$X_n = \max(Y_1, \ldots, Y_n) \tag{6.32}$$

has probability distribution function:

$$H_n(x) = [F(x)]^n \tag{6.33}$$

where $F(x) := P\{Y_i \leq x\}$ is the common probability distribution function (referred to as the parent distribution) of each Y_i.

The evaluation of the exact distribution (6.33) requires the parent distribution to be known. For n tending to infinity, the limiting distribution $H(x) := H_\infty(x)$ becomes independent of $F(x)$. This has been utilised in several geophysical applications, thus trying to fit (justifiably or not) limiting extreme value distributions, or asymptotes, to extremes of various phenomena, and bypassing the study of the parent distribution. According to Gumbel (1958), as n tends to infinity, $H_n(x)$ converges to one of three possible asymptotes, depending on the mathematical form of $F(x)$. However, all three asymptotes can be described by a single mathematical expression, known as the generalized extreme value (GEV) distribution of maxima.

The logic behind the use of the extreme value distributions is this. Let us assume that the variable Y_i denotes the daily average discharge of a river of the day i. From (6.33), X_{365} will be then the maximum daily average discharge within a year. In practical problems of flood protection designs we are interested on the distribution of the variable X_{365} instead of that of Y_i. It is usually assumed that the distribution of X_{365} (the maximum of 365 variables) is well approximated by one of the asymptotes. Nevertheless, the strict conditions that make the theoretical extreme value distributions valid are rarely satisfied in real world processes. In the previous example the variables Y_i can neither be considered independent nor identically distributed. Moreover, the convergence to the asymptotic distribution in general is very slow, so that a good approximation may require that the maximum is taken over millions of variables (Koutsoyiannis, 2004a). For these reasons, the use of the asymptotic distributions should be done with attentiveness.

If are interested about minima, rather than maxima, i.e.:

$$X_n = \min(Y_1, \ldots, Y_n) \tag{6.34}$$

then the probability distribution function of X_n is:

$$G_n(x) = 1 - [1 - F(x)]^n \tag{6.35}$$

As n tends to infinity we obtain the generalized extreme value distribution of minima, a distribution symmetric to the generalized extreme value distribution of maxima.

These two generalized distributions and their special cases are analysed below. Nevertheless, several other distributions have been used in geophysics to describe extremes, e.g. the log-normal, the two and three-parameter Gamma and the log-Pearson III distributions.

5.4.1 Generalized extreme value distribution of maxima

The mathematical expression that comprises all three asymptotes is known as the generalized extreme value (GEV) distribution. Its basic characteristics are summarized in Table 6.11.

Table 6.11 Generalized extreme value distribution of maxima conspectus.

Probability density function	$f_X(x) = \frac{1}{\lambda}\left(1 + \kappa\,\frac{x-\zeta}{\lambda}\right)^{-1/\kappa - 1} \exp\left[-\left(1 + \kappa\,\frac{x-\zeta}{\lambda}\right)^{-1/\kappa}\right]$
Distribution function	$F_X(x) = \exp\left[-\left(1 + \kappa\,\frac{x-\zeta}{\lambda}\right)^{-1/\kappa}\right]$
Range	In general: $\kappa\, x \geq \kappa\,\zeta - \lambda$ For $\kappa > 0$ (Extreme value of maxima type II): $\zeta - \lambda / \kappa \leq x < \infty$ For $\kappa < 0$ (Extreme value of maxima type III): $-\infty < x \leq \zeta - \lambda / \kappa$
Parameters	ζ: location parameter $\lambda > 0$: scale parameter κ: shape parameter
Mean	$\mu_X = \zeta - \frac{\lambda}{\kappa}\,[1 - \Gamma(1 - \kappa)]$
Variance	$\sigma_X^2 = \left(\frac{\lambda}{\kappa}\right)^2 [\Gamma(1 - 2\,\kappa) - \Gamma^2(1 - \kappa)]$
Third central moment	$\mu_X^{(3)} = \left(\frac{\lambda}{\kappa}\right)^3 [\Gamma(1 - 3\,\kappa) - 3\,\Gamma(1 - 2\,\kappa)\,\Gamma(1 - \kappa) + 2\Gamma^3(1 - \kappa)]$
Coefficient of skewness	$C_{s_X} = \mathrm{sgn}(\kappa)\,\frac{\Gamma(1-3\,\kappa) - 3\,\Gamma(1-2\,\kappa)\,\Gamma(1-\kappa) + 2\,\Gamma^3(1-\kappa)}{[\Gamma(1-2\,\kappa) - \Gamma^2(1-\kappa)]^{3/2}}$
Second L moment	$\lambda_X^{(2)} = -\Gamma(-\kappa)\,(2^\kappa - 1)\,\lambda$
Third L moment	$\lambda_X^{(3)} = -\Gamma(-\kappa)\,[2(3^\kappa - 1) - 3(2^\kappa - 1)]\,\lambda$
Fourth L moment	$\lambda_X^{(4)} = -\Gamma(-\kappa)\,[5(4^\kappa - 1) - 10(3^\kappa - 1) + 6(2^\kappa - 1)]\,\lambda$
L Coefficient of variation	$\tau_X^{(2)} = \frac{\Gamma(1 - \kappa)\,(2^\kappa - 1)\,\lambda}{\lambda\,\Gamma(1 - \kappa) + \zeta\,\kappa - \lambda}$
L Skewness	$\tau_X^{(3)} = 2\,\frac{3^\kappa - 1}{2^\kappa - 1} - 3$
L Kurtosis	$\tau_X^{(4)} = 6 + \frac{5(4^\kappa - 1) - 10(3^\kappa - 1)}{2^\kappa - 1}$

The shape parameter κ determines the general behaviour of the GEV distribution. For $\kappa > 0$ the distribution is bounded from below, has long right tail, and is known as the type II extreme value distribution of maxima or the Fréchet distribution. For $\kappa < 0$ it is bounded from above and is known as the type III extreme value distribution of maxima; this is not of practical interest in most real world problems because a bound from above is unrealistic. The limiting case where $\kappa = 0$, derived by application of l'Hôpital's rule, corresponds to the so-called extreme value distribution of type I or the Gumbel distribution (see section 5.4.2), which is unbounded both from above and below.

Typical calculations

The simplicity of the mathematical expression of the distribution function, permits typical calculations to be made directly without the need of tables or numerical approximations. The value of the distribution function can be calculated if the variable value is known. Also, the inverse distribution function has an analytical expression, namely the u-quantile of the distribution is

$$x_u = \zeta + \frac{\lambda\left[(-\ln u)^{-\kappa} - 1\right]}{\kappa} \tag{6.36}$$

Parameter estimation

As shown in Table 6.11, both coefficients of skewness and L skewness are functions of the shape parameter κ only, which enables the estimation of κ from either of the two expressions using the samples estimates of these coefficients. However the expressions are complicated and need to be solved numerically. Instead, the following explicit equations (Koutsoyiannis, 2004b) can be used, which are approximations of the exact (but implicit) equations of Table 6.11:

$$\kappa = \frac{1}{3} - \frac{1}{0.31 + 0.91\hat{C}_{s_X} + \sqrt{(0.91\,\hat{C}_{s_X})^2 + 1.8}} \tag{6.37}$$

$$\kappa = 8c - 3c^2,\ \ c := \frac{\ln 2}{\ln 3} - \frac{2}{3 + \hat{\tau}_X^{(3)}} \tag{6.38}$$

The former corresponds to the method of moments and the resulting error is smaller than ± 0.01 for $-1 < \kappa < 1/3$ ($-2 < C_{s_X} < \infty$). The latter corresponds to the method of L moments and the resulting error is smaller than ± 0.008 for $-1 < \kappa < 1$ ($-1/3 < \tau_X^{(3)} < 1$).

Once the shape parameter is calculated, the estimation of the remaining two parameters becomes very simple. The scale parameter can be estimated by the method of moments from:

$$\lambda = c_1 s_X,\ \ c_1 = |\kappa| \,/\, \sqrt{\Gamma(1 - 2\kappa) - \Gamma^2(1 - \kappa)} \tag{6.39}$$

or by the method of L moments from:

$$\lambda = c_2\, l_X^{(2)},\ \ c_2 = \kappa/[\Gamma(1 - \kappa)(2^{\kappa} - 1)] \tag{6.40}$$

The estimate of the location parameter for both the method of moments and L moments is:

$$\zeta = \bar{x} - c_3\,\lambda,\ \ c_3 = [\Gamma(1 - \kappa) - 1]/\kappa \tag{6.41}$$

5.4.2 Extreme value distribution of maxima of type I (Gumbel)

As we have explained in the previous section, the type I or the Gumbel distribution is a special case of the generalized extreme value distribution of maxima for κ = 0. Its basic characteristics are summarized in Table 6.12, where the constant γ_E that appears in some equations is the Euler* constant.

Table 6.12 Type I or Gumbel distribution of maxima conspectus.

Probability density function	$f_X(x) = \frac{1}{\lambda}\exp\left(-\frac{x-\zeta}{\lambda}\right)\exp\left[-\exp\left(-\frac{x-\zeta}{\lambda}\right)\right]$
Distribution function	$F_X(x) = \exp\left[-\exp\left(-\frac{x-\zeta}{\lambda}\right)\right]$
Range	$-\infty < x < \infty$ (continuous)
Parameters	ζ: location parameter $\lambda > 0$: scale parameter
Mean	$\mu_X = \zeta + \gamma_E \lambda = \zeta + 0.5772\lambda$
Variance	$\sigma_X^2 = \frac{\pi^2}{6}\lambda^2 = 1.645\lambda^2$
Third central moment	$\mu_X^{(3)} = 2.404\lambda^3$
Fourth central moment	$\mu_X^{(4)} = 14.6\lambda^4$
Coefficient of skewness	$C_{s_X} = 1.1396$
Coefficient of kurtosis	$C_{k_X} = 5.4$
Mode	$x_p = \zeta$
Median	$x_{0.5} = \zeta - \lambda\ln(-\ln 0.5) = \zeta + 0.3665\lambda$
Second L moment	$\lambda_X^{(2)} = \lambda \ln2$
Third L moment	$\lambda_X^{(3)} = (2\ln3 - 3\ln2)\lambda$
Fourth L moment	$\lambda_X^{(4)} = 2(8\ln2 - 5\ln3)\lambda$
L coefficient of variation	$\tau_X^{(2)} = \frac{\ln2\,\lambda}{\zeta + \gamma_E\lambda}$
L skewness	$\tau_X^{(3)} = 2\frac{\ln3}{\ln2} - 3 \approx 0.1699$
L kurtosis	$\tau_X^{(4)} = 16 - 10\frac{\ln3}{\ln2} \approx 0.1504$

* The Euler constant is defined as the limit

$$\gamma_E := \lim_{n\to\infty}\left(1 + \frac{1}{2} + \cdots + \frac{1}{n} - \ln n\right) \approx 0.5772156649\ \ldots$$

Typical calculations

Due to the simplicity of the mathematical expression of the distribution function, typical calculations can be done explicitly without the need of tables or numerical approximations. The value of the distribution function can be calculated easily if the value of the variable is known. Moreover, the inverse distribution function has an analytical expression, namely the u-quantile of the distribution is

$$x_u = \zeta - \lambda \ln(-\ln u) \tag{6.42}$$

Parameter estimation

Since the Gumbel distribution is a special case of the GEV distribution, the parameter estimation procedures of the latter can be applied also in this case (except for the estimation of κ which by definition is zero). Specifically equations (6.39)-(6.41) for the method of moments and L moments still hold, and the constants c_i have the following values: $c_1 = \sqrt{6}/\pi = 0.78$, $c_2 = 1/\ln 2 = 1.443$ and $c_3 = \gamma_E = 0.577$.

Another method that results in similar expressions is the Gumbel method (Gumbel, 1958, p. 227). The method is based in the least square fit of the theoretical distribution function to the empirical distribution. For the empirical distribution function the Weibull plotting position must be used. The expressions of this method depend on the sample size n. The original Gumbel method is based on tabulated constants. To avoid the use of tables we give the following expressions that are good approximations of the original method:

$$\lambda = \frac{s_X}{\dfrac{1}{0.78} - \dfrac{1.57}{(n+1)^{0.65}}}, \quad \zeta = \bar{x} - \left[0.577 - \frac{0.53}{(n+2.5)^{0.74}}\right]\lambda \tag{6.43}$$

The approximation error is smaller than 0.25% for the former equation and smaller than 0.10% for the latter (for $n \geq 10$). For small exceedence probabilities, the Gumbel method results in safer predictions in comparison to the method of moments. The maximum likelihood method is more complicated; the interested reader may consult Kite (1988, p. 96).

Standard error and confidence intervals of quantiles

If the method of moments is used to estimate the parameters, then the point estimate of the u-quantile can be written in the following form that is equivalent to (6.42):

$$\hat{x}_u = \bar{x} - 0.5772\lambda - \lambda \ln(-\ln u) = \bar{x} + k_u s_X \tag{6.44}$$

where,

$$k_u = \lambda \frac{-0.5772 - \ln(-\ln u)}{s_X} = -0.45 - 0.78 \ln(-\ln u) \tag{6.45}$$

In this case it can be shown (Gumbel, 1958, p. 228· Kite, 1988, p. 103) that the square of the standard error of the estimate is

$$\varepsilon_X^2 = \mathrm{Var}(\hat{X}_u) = \frac{s_X^2}{n}\left(1 + 1.1396k_u + 1.1k_u^2\right) \tag{6.46}$$

Consequently, the confidence intervals of the u-quantile for confidence coefficient γ is approximately

$$\hat{x}_{u,2} = \left(\bar{x} + k_u s_X\right) \pm z_{(1+\gamma)/2}\frac{s_X}{\sqrt{n}}\sqrt{1 + 1.1396k_u + 1.1k_u^2} \tag{6.47}$$

Gumbel probability plot

The Gumbel distribution can be depicted as a straight line on a Gumbel probability plot. This plot can be easily constructed with horizontal probability axis $h = -\ln(-\ln F)$ (sometimes called Gumbel reduced variate) and vertical axis the variable of interest. Clearly, equation (6.42) is a straight line in this probability plot.

Numerical example

Table 6.13 lists a sample of the annual maximum daily discharge of the Evinos river upstream of the hydrometric station of Poros Reganiou. We wish to fit the Gumbel distribution of maxima and to determine the 100-year maximum discharge.

Table 6.13 Sample of annual maximum daily discharge (in m^3/s) of the river Evinos upstream of the hydrometric station of Poros Reganiou.

Hydrological year	Maximum discharge	Hydrological year	Maximum discharge	Hydrological year	Maximum discharge
1970-71	884	1977-78	365	1984-85	317
1971-72	305	1978-79	502	1985-86	374
1972-73	215	1979-80	381	1986-87	188
1973-74	378	1980-81	387	1987-88	192
1974-75	176	1981-82	525	1988-89	448
1975-76	430	1982-83	412	1989-90	70
1976-77	713	1983-84	439		

The sample average is

$$\bar{x} = \sum x / n = 385.1 \text{ m}^3\text{/s}$$

The standard deviation is

$$s_X = \sqrt{\sum x^2 / n - \bar{x}^2} = 181.5 \text{ m}^3\text{/s}$$

and the coefficient of variation is

$$\hat{C}_{\mathrm{v}_X} = s_X / \bar{x} = 181.5 / 385.1 = 0.47$$

The skewness coefficient is

$$\hat{C}_{\mathrm{s}_X} = 0.94$$

a value close to the theoretical value of the Gumbel distribution (1.14).

The method of moments results in

$$\lambda = 0.78 \times 181.5 = 141.57 \text{ m}^3/\text{s}, \zeta = 385.1 - 0.577 \times 141.57 = 303.4 \text{ m}^3/\text{s}$$

The maximum daily discharge for $T = 100$, or equivalently for $u = 1 - 1/100 = 0.99$, is

$$x_{0.99} = 303.4 - 141.57 \times \ln[-\ln(0.99)] = 955.0 \text{ m}^3/\text{s}$$

Based on (6.47), for

$$k_u = (955.0 - 385.1) / 181.5 = 3.16, z_{(1+\gamma)/2} = 1.96$$

we determine the 95% confidence intervals of the 100-year maximum daily discharge:

$$\hat{x}_{u1,2} \approx 955.0 \pm 1.96 \times \frac{181.5}{\sqrt{20}} \sqrt{1 + 1.1396 \times 3.16 + 1.1 \times 3.16^2}$$

$$\approx 955.0 \pm 313.1 = \begin{cases} 1268.1 \text{ m}^3/\text{s} \\ 641.9 \text{ m}^3/\text{s} \end{cases}$$

The Gumbel method using the equations (6.43), for $n = 20$, gives

$$\lambda = 170.36 \text{ m}^3/\text{s}, \zeta = 295.7 \text{ m}^3/\text{s}$$

and the 100-year maximum discharge estimation is

$$x_{0.99} = 295.7 - 170.36 \times \ln[-\ln(0.99)] = 1079.4 \text{ m}^3/\text{s}$$

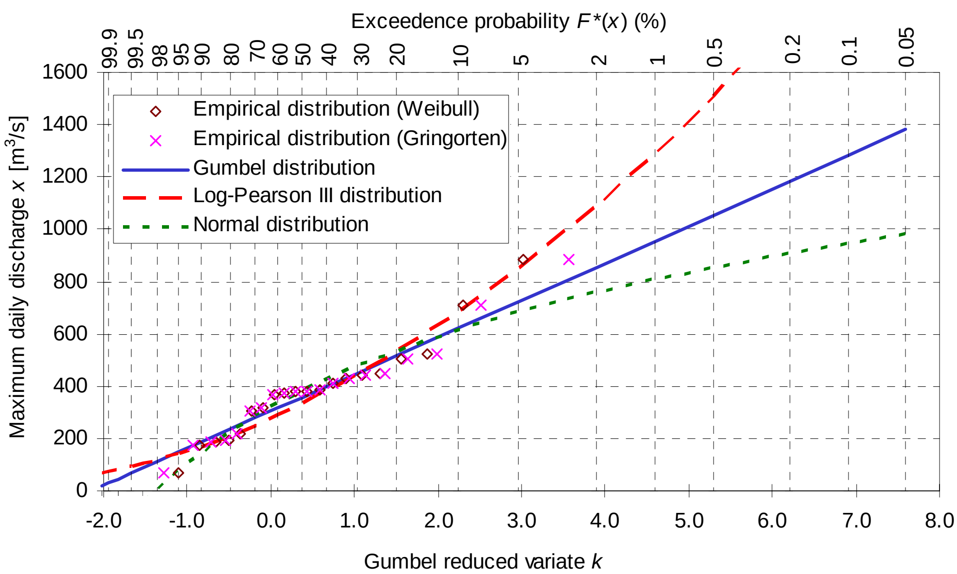

Fig. 6.3 Empirical and theoretical distribution of the daily maximum discharge of the river Evinos at station of Poros Riganiou plotted in Gumbel of maxima probability paper.

Fig. 6.3 depicts a comparison of the empirical distribution function and the theoretical Gumbel distribution of maxima on a Gumbel probability plot. For the empirical distribution function we have used the Weibull and the Gringorten plotting positions. For comparison we have also plotted the normal and log-Pearson III distributions. Clearly, the normal distribution is inappropriate (as expected) but even the Gumbel distribution does not fit well in the area of small exceedence probabilities that are of more interest, and seems to underestimate the highest discharges. The log-Person III distribution seems to be the most appropriate for the highest values of discharge. This seems to be a general problem for the Gumbel distribution. For more than have a century it has been the prevailing model for quantifying risk associated with extreme geophysical events. Newer evidence and theoretical studies (Koutsoyiannis, 2004a,b, 2005) have shown that the Gumbel distribution is quite unlikely to apply to hydrological extremes and its application may misjudge the risk, as it underestimates seriously the largest extremes. Besides, it has been shown that observed samples of typical length (like the one of this example) may display a distorted picture of the actual distribution, suggesting that the Gumbel distribution is an appropriate model for geophysical extremes while it is not. Therefore, it can be recommended to avoid the Gumbel distribution for the description of extreme rainfall and river discharge and use long-tail distributions such as the extreme value distribution of type II or log-Pearson III.

5.4.3 Generalized extreme value distribution of minima

If $H(x)$ is the generalized extreme value distribution of maxima then the distribution function $G(x) = 1 - H(-x)$ is the generalized extreme value distribution of minima. Its general characteristics are summarized in Table 6.14, where we have changed the sign convention in the parameter κ so that the distribution be unbounded from above for $\kappa > 0$ (bounded from below). This is similar to the generalized extreme value distribution of maxima where again $\kappa > 0$ corresponds to a distribution be unbounded from above. However, they are termed, respectively, type II extreme value distribution of maxima and type III extreme value distribution of minima (or the Weibull distribution). For $\kappa < 0$ the distribution of minima (similar to that of maxima) is bounded from above and is known as the type II extreme value distribution of minima; this is not of practical interest as in most real world problems a bound from above is unrealistic. The limiting case where $\kappa = 0$, derived by application of l'Hôpital's rule, corresponds to the so-called type I extreme value distribution of minima or the Gumbel distribution of minima, which is unbounded both from above and below.

Typical calculations

The mathematical expression of the generalized extreme value distribution of minima is similar to that of maxima. Thus, typical calculations can be done explicitly. The value of the distribution function can be calculated directly from the value of the variable. Also, the inverse distribution function has an analytical expression, namely the u-quantile of the distribution is given by

$$x_u = \zeta + \frac{\lambda}{\kappa}\{ [-\ln (1 - u)]^{\kappa} - 1\} \tag{6.48}$$

Table 6.14 Generalized extreme value distribution of minima conspectus.

Probability density function	$f_X(x) = \frac{1}{\lambda}\left[1+\kappa\left(\frac{x-\zeta}{\lambda}\right)\right]^{1/\kappa-1} \exp\left\{-\left[1+\kappa\left(\frac{x-\zeta}{\lambda}\right)\right]^{1/\kappa}\right\}$
Distribution function	$F_X(x) = 1 - \exp\left[-\left(1+\kappa\,\frac{x-\zeta}{\lambda}\right)^{1/\kappa}\right]$
Range	In general: $\kappa\, x \geq \kappa\,\zeta - \lambda$ For $\kappa > 0$ (Extreme value of minima type III): $\zeta - \lambda / \kappa \leq x < \infty$ For $\kappa < 0$ (Extreme value of minima type II): $-\infty < x \leq \zeta - \lambda / \kappa$
Parameters	ζ: location parameter $\lambda > 0$: scale parameter κ: shape parameter
Mean	$\mu_X = \zeta + \frac{\lambda}{\kappa}[\Gamma(1+\kappa) - 1]$
Variance	$\sigma_X^2 = \left(\frac{\lambda}{\kappa}\right)^2 [\Gamma(1+2\,\kappa) - \Gamma^2(1+\kappa)]$
Third central moment	$\mu_X^{(3)} = \left(\frac{\lambda}{\kappa}\right)^3 [\Gamma(1+3\,\kappa) - 3\,\Gamma(1+2\,\kappa)\,\Gamma(1+\kappa) + 2\Gamma^3(1+\kappa)]$
Coefficient of skewness	$C_{s_X} = \mathrm{sgn}(\kappa)\,\frac{\Gamma(1+3\,\kappa) - 3\,\Gamma(1+2\,\kappa)\,\Gamma(1+\kappa) + 2\,\Gamma^3(1+\kappa)}{[\Gamma(1+2\,\kappa) - \Gamma^2(1+\kappa)]^{3/2}}$
Second L moment	$\lambda_X^{(2)} = \Gamma(\kappa)\,(1-2^{-\kappa})\,\lambda$
Third L moment	$\lambda_X^{(3)} = \Gamma(\kappa)\,[3(1-2^{-\kappa}) - 2(1-3^{-\kappa})]\,\lambda$
Fourth L moment	$\lambda_X^{(4)} = \Gamma(\kappa)\,[5(1-4^{-\kappa}) - 10(1-3^{-\kappa}) + 6(1-2^{-\kappa})]\,\lambda$
L coefficient of variation	$\tau_X^{(2)} = \frac{\Gamma(1+\kappa)\,(1-2^{-\kappa})\,\lambda}{\lambda\,\Gamma(1+\kappa) + \zeta\,\kappa - \lambda}$
L skewness	$\tau_X^{(3)} = 3 - 2\,\frac{1-3^{-\kappa}}{1-2^{-\kappa}}$
L kurtosis	$\tau_X^{(4)} = 6 + \frac{5(1-4^{-\kappa}) - 10(1-3^{-\kappa})}{1-2^{-\kappa}}$

Parameter estimation

As shown in Table 6.14, both coefficients of skewness and L skewness are functions of the shape parameter κ only, which enables the estimation of κ from either of the two expressions using the sample estimates of these coefficients. However the expressions are complicated and need to be solved numerically. Instead, the following explicit equations (Koutsoyiannis, 2004b) can be used, which are approximations of the exact (but implicit) equations of Table 6.11:

$$\kappa = \frac{1}{0.28 - 0.9\hat{C}_{s_X} + 0.998\sqrt{(0.9\,\hat{C}_{s_X})^2 + 1.93}} - \frac{1}{3} \tag{6.49}$$

$$\kappa = 7.8c + 4.71\, c^2,\ \ c := \frac{2}{3 - \hat{\tau}_X^{(3)}} - \frac{\ln 2}{\ln 3} \tag{6.50}$$

The former corresponds to the method of moments and the resulting error is smaller than ±0.01 for $-1/3 < \kappa < 3$ ($-\infty < C_s < 20$). The latter corresponds to the method of L moments and the resulting error is even smaller.

Once the shape parameter is known, the scale parameter can be estimated by the method of moments from:

$$\lambda = c_1 s_X,\ \ c_1 = |\kappa| \,/\, \sqrt{\Gamma(1 + 2\kappa) - \Gamma^2(1 + \kappa)} \tag{6.51}$$

or by the method of L moments from:

$$\lambda = c_2\, \hat{l}_X^{(2)},\ \ c_2 = \kappa/[\Gamma(1 - \kappa)(2^\kappa - 1)] \tag{6.52}$$

The estimate of the location parameter for both the method of moments and L moments is:

$$\zeta = \bar{x} + c_3 \lambda,\ \ c_3 = [1 - \Gamma(1 + \kappa)]/\kappa \tag{6.53}$$

5.4.4 Extreme value distribution minima of type I (Gumbel)

As shown in Table 6.15, the type I distribution of minima resembles the type I distribution of maxima. The typical calculations are also similar. The inverse distribution function has an analytical expression and thus the u-quantile is given by:

$$x_u = \zeta + \lambda \ln\,[-\ln(1 - u)] \tag{6.54}$$

Since the Gumbel distribution is a special case of the GEV distribution, the parameter estimation procedures of the latter is based on equations (6.51)-(6.53) but with constants c_i as follows: $c_1 = \sqrt{6}/\pi = 0.78$, $c_2 = 1/\ln 2 = 1.443$ and $c_3 = \gamma_E = 0.577$.

We can plot the Gumbel distribution of minima on a Gumbel-of-maxima probability paper if we replace the probability of exceedence with the probability of non-exceedence. Further, we can construct a Gumbel-of-minima probability plot if we use as horizontal axis the variate $h = \ln[-\ln(1-F)]$.

5.4.5 Two-parameter Weibull distribution

If in the generalized extreme value distribution of minima we assume that the lower bound ($\zeta - \lambda/\kappa$) is zero, we obtain the special case known as the two two-parameter Weibull distribution. Its main characteristics are shown in Table 6.15, where for convenience we have performed a change of the scale parameter replacing λ/κ with α.

Typical calculations

The related calculations are simple as in all previous cases and the inverse distribution function, from which quantiles are estimated, is

$$x_u = \alpha\{\,[-\ln(1 - u)]^{\kappa}\} \tag{6.55}$$

Table 6.15 Type I or Gumbel distribution of minima conspectus.

Probability density function	$f_X(x) = \frac{1}{\lambda} \exp\left(\frac{x-\zeta}{\lambda}\right) \exp\left[-\exp\left(\frac{x-\zeta}{\lambda}\right)\right]$
Distribution function	$F_X(x) = 1 - \exp\left[-\exp\left(\frac{x-\zeta}{\lambda}\right)\right]$
Variable range	$-\infty < x < \infty$ (continuous)
Parameters	ζ: location parameter $\lambda > 0$: scale parameter
Mean	$\mu_X = \zeta - \gamma_E \lambda = \zeta - 0.5772\lambda$
Variance	$\sigma_X^2 = \frac{\pi^2}{6}\lambda^2 = 1.645\lambda^2$
Third central moment	$\mu_X^{(3)} = -2.404\lambda^3$
Fourth central moment	$\mu_X^{(4)} = 14.6\lambda^4$
Skewness coefficient	$C_{s_X} = -1.1396$
Kurtosis coefficient	$C_{k_X} = 5.4$
Mode	$x_p = \zeta$
Median	$x_{0.5} = \zeta + \lambda \ln(-\ln 0.5) = \zeta - 0.3665\lambda$
Second L moment	$\lambda_X^{(2)} = \lambda \ln 2$
Third L moment	$\lambda_X^{(3)} = -(2 \ln 3 - 3 \ln 2)\, \lambda$
Fourth L moment	$\lambda_X^{(4)} = 2(8 \ln 2 - 5 \ln 3)\, \lambda$
L coefficient of variation	$\tau_X^{(2)} = \frac{\ln 2\, \lambda}{\zeta - \gamma_E\, \lambda}$
L skewness	$\tau_X^{(3)} = -2\frac{\ln 3}{\ln 2} + 3 \approx -0.1699$
L kurtosis	$\tau_X^{(4)} = 16 - 10\frac{\ln 3}{\ln 2} \approx 0.1504$

Parameter estimation

From the expressions of Table 6.14, the estimate of κ by the method of moments can be done from:

$$\frac{\Gamma(1 + 2\kappa)}{\Gamma^2(1 + \kappa)} = \hat{C}_{v_X}^2 + 1 \tag{6.56}$$

This is implicit for κ and can be solved only numerically. An approximate solution with accuracy ±0.01 για $0 < \kappa < 3.2$ or $0 < C_{v_X} < 5$) is

$$\kappa = 2.56 \left\{\exp\{0.41 [\ln(C_v^2 + 1)]^{0.58}\} - 1\right\} \tag{6.57}$$

The L moment estimate is much simpler:

$$\kappa = \frac{-\ln(1-\tau_X^{(2)})}{\ln 2} \tag{6.58}$$

Once κ has been estimated, the scale parameter for both the method of moments and L moments is

$$\alpha = \frac{\bar{x}}{\Gamma(1+\kappa)} \tag{6.59}$$

Table 6.16 Two-parameter Weibull distribution (type III of minima) conspectus.

Probability density function	$f(x) = \frac{1}{\kappa\,\alpha}\left(\frac{x}{\alpha}\right)^{1/\kappa-1}\exp\left[-\left(\frac{x}{\alpha}\right)^{1/\kappa}\right]$
Distribution function	$F(x) = 1 - \exp\left[-\left(\frac{x}{\alpha}\right)^{1/\kappa}\right]$
Range	$0 < x < \infty$ (continuous)
Parameters	$\alpha > 0$: scale parameter $\kappa > 0$: shape parameter
Mean	$\mu_X = \alpha\Gamma(1+\kappa)$
Variance	$\sigma_X^2 = \alpha^2[\Gamma(1+2\kappa) - \Gamma^2(1+\kappa)]$
Third central moment	$\mu_X^{(3)} = \alpha^3[\Gamma(1+3\kappa) - 3\Gamma(1+2\kappa)\,\Gamma(1+\kappa) + 2\Gamma^3(1+\kappa)]$
Coefficient of variation	$C_{vX} = \frac{[\Gamma(1+2\kappa) - \Gamma^2(1+\kappa)]^{1/2}}{\Gamma(1+\kappa)}$
Coefficient of skewness	$C_{sX} = \frac{\Gamma(1+3\kappa) - 3\,\Gamma(1+2\kappa)\,\Gamma(1+\kappa) + 2\,\Gamma^3(1+\kappa)}{[\Gamma(1+2\kappa) - \Gamma^2(1+\kappa)]^{3/2}}$
Mode	$x_p = \alpha(1-\kappa)^\kappa$ (for $\kappa > 1$)
Median	$x_{0.5} = \alpha(\ln 2)^\kappa$
Second L moment	$\lambda_X^{(2)} = \Gamma(1+\kappa)\,(1-2^{-\kappa})\,\alpha$
Third L moment	$\lambda_X^{(3)} = \Gamma(1+\kappa)\,[3(1-2^{-\kappa}) - 2(1-3^{-\kappa})]\,\alpha$
Fourth L moment	$\lambda_X^{(4)} = \Gamma(1+\kappa)\,[5(1-4^{-\kappa}) - 10(1-3^{-\kappa}) + 6(1-2^{-\kappa})]\,\alpha$
L coefficient of variation	$\tau_X^{(2)} = 1 - 2^{-\kappa}$
L skewness	$\tau_X^{(3)} = 3 - 2\,\frac{1-3^{-\kappa}}{1-2^{-\kappa}}$
L kurtosis	$\tau_X^{(4)} = 6 + \frac{5(1-4^{-\kappa}) - 10(1-3^{-\kappa})}{1-2^{-\kappa}}$

We observe that the transformation $Z = \ln X$ results in

$$F_z(z) = 1 - \exp[-\mathrm{e}^{(z-\ln\alpha)/\kappa}] \tag{6.60}$$

which is a Gumbel distribution of minima with location parameter ln α and scale parameter κ. Thus, we can also use the parameter estimation methods of the Gumbel distribution applied on the logarithms of the observed sample values.

Weibull probability plot

A probability plot where the two-parameter Weibull distribution is depicted as a straight line is possible. The horizontal axis is $h = \ln[-\ln(1-F)]$ (similar to the plot of Gumbel of minima) and the vertical axis is $v = \ln x$ (logarithmic scale).

Numerical example

Table 6.17 lists a sample of annual minimum (average) daily discharge of the Evinos river upstream of the hydrometric station of Poros Reganiou. We wish to fit the Gumbel distribution of minima and the Weibull distribution and to determine the minimum 20-year discharge.

Table 6.17 Sample of annual minimum daily discharges (in m^3/s) of the river Evinos at the station of Poros Riganiou.

Hydrological year	Minimum. discharge	Hydrological year	Minimum discharge	Hydrological year	Minimum. discharge
1970-71	0.00	1977-78	2.14	1984-85	0.54
1971-72	2.19	1978-79	2.00	1985-86	0.54
1972-73	2.66	1979-80	1.93	1986-87	1.70
1973-74	2.13	1980-81	2.29	1987-88	1.70
1974-75	1.28	1981-82	2.66	1988-89	0.32
1975-76	0.56	1982-83	2.87	1989-90	1.37
1976-77	0.13	1983-84	1.88		

The sample mean is

$$\bar{x} = \sum x / n = 1.545 \text{ m}^3\text{/s}$$

The standard deviation is

$$s_X = \sqrt{\sum x^2 / n - \bar{x}^2} = 0.878 \text{ m}^3\text{/s}$$

and the coefficient of variation is

$$\hat{C}_{\mathrm{v}X} = s_X / \bar{x} = 0.878/1.545 = 0.568$$

The skewness coefficient is

$$\hat{C}_{\mathrm{s}X} = -0.40$$

The negative value of the skewness coefficient is expected for a sample of minimum discharges.

For the Gumbel distribution, the method of moments yields

$$\lambda = 0.78 \times 0.878 = 0.685 \text{ m}^3/\text{s}, \zeta = 1.545 + 0.577 \times 0.685 = 1.940 \text{ m}^3/\text{s}$$

The minimum discharge for $T = 20$ years, or equivalently for $u = 1/20 = 0.05$, is

$$x_{0.05} = 1.940 + 0.685 \times \ln[-\ln(1 - 0.05)] = -0.09 \text{ m}^3/\text{s}$$

Apparently, a negative value of discharge is meaningless; we can consider that the minimum 20-year discharge is zero.

For the two-parameter Weibull distribution, application of (6.57) for the method of moments gives

$$\kappa = 2.56 \left\{\exp\{0.41 [\ln(0.568^2 + 1)]^{0.58}\} - 1\right\} = 0.55$$

Hence

$$\alpha = \frac{1.545}{\Gamma(1 + 0.55)} = 1.740 \text{ m}^3/\text{s}$$

and the 20-year mminimum daily discharge is estimated at

$$x_{0.05} = 1.740 \{[-\ln(1-0.05)]^{0.55}\} = 0.340 \text{ m}^3/\text{s}$$

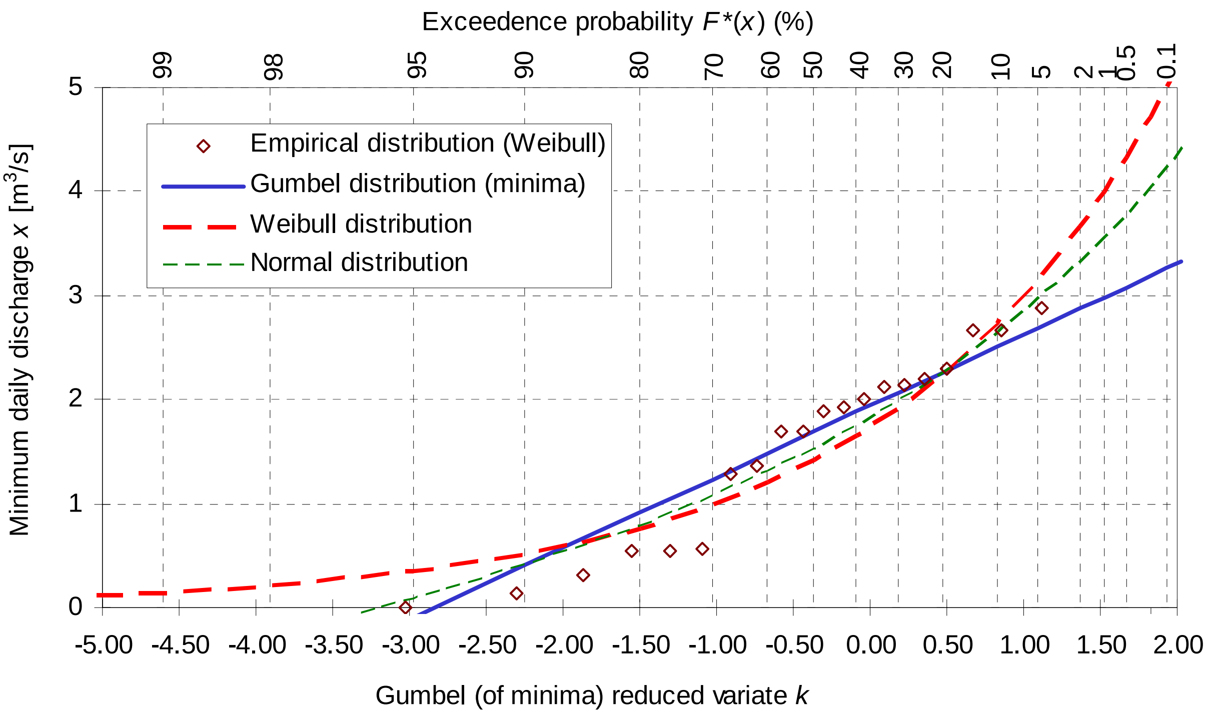

Fig. 6.4 Empirical and theoretical distribution function of the minimum daily discharge of the river Evinos at the station Poros Riganiou in Gumbel of minima probability paper.

Fig. 6.4 compares graphically the empirical distribution function with the two fitted theoretical distributions. For the empirical distribution we have used the Weibull plotting position. None the two theoretical distributions fits very well to the sample, but clearly the Gumbel distribution performs better, especially in the area of small exceedence probabilities.

The two-parameter Weibull distribution is defined for $x > 0$, which seems to be a theoretical advantage due to the consistency with nature. However, in practice it turns to be a disadvantage due to the departure of the empirical distribution for the lowest discharges. On the other hand, the Gumbel distribution of minima is theoretically inconsistent as it predicts negative values of discharge for high return periods. An *ad hoc* solution is to truncate the Gumbel distribution at zero, as we have done above. For comparison the normal distribution has been also plotted in Fig. 6.4 but we do not expect to be appropriate for this problem.

Acknowledgement I thank Simon Papalexiou for his help in translating some of my Greek texts into English and for his valuable suggestions.

References

Bobée, B., and F. Ashkar, *The Gamma Family and Derived Distributions Applied in Hydrology*, Water Resources Publications, Littleton, Colorado, 1991.

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

Kite, G. W., *Frequency and Risk Analyses in Hydrology*, Water Resources Publications, Littleton, Colorado, 1988.

Koutsoyiannis, D., *Statistical Hydrology*, Edition 4, 312 pages, National Technical University of Athens, Athens, 1997.

Koutsoyiannis, D., Statistics of extremes and estimation of extreme rainfall, 1, Theoretical investigation, *Hydrological Sciences Journal*, 49 (4), 575–590, 2004a.

Koutsoyiannis, D., Statistics of extremes and estimation of extreme rainfall, 2, Empirical investigation of long rainfall records, *Hydrological Sciences Journal*, 49 (4), 591–610, 2004b.

Koutsoyiannis, D., Uncertainty, entropy, scaling and hydrological stochastics, 1, Marginal distributional properties of hydrological processes and state scaling, *Hydrological Sciences Journal*, 50 (3), 381–404, 2005.

Koutsoyiannis, D., An entropic-stochastic representation of rainfall intermittency: The origin of clustering and persistence, *Water Resources Research*, 42 (1), W01401, 2006.

Press, W. H., B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes*, Cambridge University Press, Cambridge, 1987.

Stedinger, J. R., R. M. Vogel, and E. Foufoula-Georgiou, Frequency analysis of extreme events, Chapter 18 in *Handbook of Hydrology*, edited by D. R. Maidment, McGraw-Hill, 1993.