Statistical Estimation Methods in Hydrological Engineering

P.H.A.J.M. van Gelder
TU Delft, The Netherlands
p.vangelder@ct.tudelft.nl, http://surf.to/vangelder

Introduction

In designing civil engineering structures use is made of probabilistic calculation methods. Stress and load parameters are described by statistical distribution functions. The parameters of these distribution functions can be estimated by various methods. An extensive comparison of these different estimation methods is given in this paper. The main point of interest is the behaviour of each method for predicting p-quantiles (the value which is exceeded by the random variable with probability p), where p«1. The estimation of extreme quantiles corresponding to a small probability of exceedance is commonly required in the risk analysis of hydraulic structures. Such extreme quantiles may represent design values of environmental loads (wind, waves, snow, earthquake), river discharges, and flood levels specified by design codes and regulations (TAW, 1990). In this paper the performance of the parameter estimation methods with respect to its small sample behaviour is analyzed with Monte Carlo simulations, added with mathematical proofs.

In civil engineering practice many parameter estimation methods for probability distribution functions are in circulation. Well known methods are for example:
- the method of moments (Johan Bernoulli, 1667-1748),
- the method of maximum likelihood (Daniel Bernoulli, 1700-1782),
- the method of least squares (on the original or on the linearized data), (Gauss, 1777-1855),
- the method of Bayesian estimation (Bayes, 1763),
- the method of minimum cross entropy (Shannon, 1949),
- the method of probability weighted moments (Greenwood et al., 1979),
- the method of L-moments (Hosking, 1990).

Textbooks, such as Benjamin and Cornell (1970), Berger (1980), treat the traditional methods in detail. The methods will be briefly reviewed in this paper.

Many attempts (for instance, Goda and Kobune, (1990), Burcharth and Liu, (1994), Yamaguchi, (1997), and Van Gelder and Vrijling, (1997a)), have been made to find out which estimation method is preferable for the parameter estimation of a particular probability distribution in order to obtain a reliable estimate of the p-quantiles. In this paper, we will in particularly investigate the performance of the parameter estimation method with respect to three different criteria; (i) based on the relative bias and root mean squared error (RMSE), (ii) based on the over- and underdesign.

It is desirable that the quantile estimate be unbiased, that is, its expected value should be equal to the true value. It is also desirable that an unbiased estimate be efficient, i.e., its variance should be as small as possible. The problem of unbiased and efficient estimation of extreme quantiles from small samples is commonly encountered in the civil engineering practice. For example, annual flood discharge

data may be available for past 50 to 100 years and on that basis one may have to estimate a design flood level corresponding to a 1,000 to 10,000 years return period (Van Gelder et al., 1995).

The first step in quantile estimation involves fitting an analytical probability distribution to represent adequately the sample observations. To achieve this, the distribution type should be judged from data and then parameters of the selected distribution should be estimated. Since the bias and efficiency of quantile estimates are sensitive to the distribution type, the development of simple and robust criteria for fitting a representative distribution to small samples of observations has been an active area of research. In this paper three different methods for the selection of the distribution type will be reviewed, extended and tested. The first method is based on Bayesian statistics, the second one on linear regression, and the third one on L-moments. Certain linear combinations of expectations of order statistics, also referred to as L-moments by Hosking (1990), have been shown to be very useful in statistical parameter estimation. Being a linear combination of data, they are less influenced by outliers, and the bias of their small sample estimates remains fairly small. A measure of kurtosis derived from L-moments, referred to as L-kurtosis, was suggested as a useful indicator of distribution shape (Hosking, 1992).

Hosking (1997) proposed a simple but effective approach to fit 3-parameter distributions. The approach involves the computation of three L-moments from a given sample. By matching the three L-moments, a set of 3-parameter distributions can be fitted to the sample data. In this paper, a distribution type selection which is based on the the 4^{th} L-moment is suggested to be the most representative distribution, which should be used for quantile estimation. In essence, the L-kurtosis, which is related to the 4^{th} L-moment, can be interpreted as a measure of resemblance between two distributions having common values of the first three L-moments.

The concept of probabilistic distance or discrimination between two distributions is discussed in great detail in modern information theory (Kullback 1959, Jumarie 1990). Mathematically sound measure of probabilistic distance, namely, the divergence, has been used to establish resemblance between two distributions or conversely to select the closest possible posterior distribution given an assumed prior distribution. The divergence is a comprehensive measure of probabilistic distance, since it involves the computation of departure of a distribution from the reference parent distribution over an entire range of the random variable. Apart from the performance of estimation methods based on bias and RMSE, and the performance based on under- and overdesign, is suggested in this paper.

Furthermore, this paper will focus on evaluating the robustness of the L-kurtosis measure in the distribution selection and extreme quantile estimation from small samples. The robustness is evaluated against the benchmark estimates obtained from the information theoretic measure, namely, the divergence. For this purpose, a series of Monte Carlo simulation experiments were designed in which probability distributions were fitted to the sample observations based on L-kurtosis and divergence based criteria, and the accuracies of quantile estimates were compared. The simulation study revealed that the L-kurtosis measure is fairly effective in quantile estimation.

Finally, this paper shows some analytical considerations concerning statistical estimation methods and probability distribution functions. The paper ends with a discussion.

Classical estimation methods

To make statements about the population on the basis of a sample, it is important to understand in what way the sample relates to the population. In most cases the following assumptions will be made:
1. Every sample observation x is the outcome of a random variable X which has an identical distribution (either discrete or continuous) for every member of the population;
2. The random variables X_1, X_2, ..., X_n corresponding to the different members of the sample are independent.

These two assumptions (abbreviated to i.i.d. (independent identically distributed)) formalize what is meant by the statement of drawing a random sample from a population.

We have now reduced the problem to one which is mathematically very simple to state: we have i.i.d. observations x_1, x_2, ..., x_n of a random variable X with probability function (in the discrete case) or probability density function (in the continuous case) f, and we want to estimate some aspect of this population distribution (for instance the mean or the variance).

It is helpful here to stress the notation that we are using in this paper: small case letters x_i denote actual sample values. But each x_i is the realisation of a random variable X_i, denoted by capitals. Thus $X=(X_1, X_2, ..., X_n)$ denotes a random sample, whose particular value is $x=(x_1, x_2, ..., x_n)$. The distinction helps us to distinguish between a random quantity and the outcome this quantity actually realises.

A statistic, T(X), is any function of the data (note that T(X) denotes that this is a random quantity which varies from sample to sample; T(x) will denote the value for a specific sample x). If a statistic is used for the purpose of estimating a parameter then it is called an estimator and the realised value T(x) is called an estimate. The basis of our approach will be to use T(x) as the estimate of θ, but to look at the sampling properties of the estimator T(x) to judge the accuracy of the estimate. Since any function of the sample data is a potential estimator, how should we determine whether an estimator is good or not? There are, in fact, many such criteria: we will focus on the two most widely used:
- Though we cannot hope to estimate a parameter perfectly, we might hope that 'on average' the estimation procedure gives the correct result.
- Estimators are to be preferred if they have small variability; in particular, we may require the variability to diminish as we take samples of a larger size.

These concepts are formalized as follows.
The estimator T(X) is unbiased for θ if :

$$E(T(X)) = \theta \qquad (0.1)$$

Otherwise, $B(T) = E(T(X)) - \theta$ is the bias of T.
If $B(T) \to 0$ as the sample size $n \to \infty$, then T is said to be asymptotically unbiased for θ.
The mean-squared error of an estimator is defined by:

$$MSE(T) = E((T(X) - \theta)^2) \qquad (0.2)$$

Note that $MSE(T)=var(T)+B^2(T)$. Indeed $MSE(T) = E(T^2(X)-2\theta T(X)+\theta^2) = E(T^2(X)) - 2\theta E(T(X)) + \theta^2 = E(T^2(X)) - 2\theta(B(T)+\theta) + \theta^2 = E(T^2(X)) - 2\theta B(T) - \theta^2$

and $\text{var}(T) = E(T^2(X)) - E^2(T(X)) = E(T^2(X)) - (B(T) + \theta)^2 = E(T^2(X)) - B^2(T) - 2\theta B(T) - \theta^2$. This proves the equality.

The root mean-squared error of an estimator is defined as:

$$\text{RMSE} = \sqrt{\text{MSE}} \tag{0.3}$$

An estimator T is said to be mean-squared consistent for θ if $\text{MSE}(T) \to 0$ as the sample size $n \to \infty$.

Ideally, estimators are both unbiased and consistent. We also prefer estimators to have as small a variance as possible. In particular, given two estimators T_1 and T_2 both being unbiased for θ, then T_1 is said to be more efficient than T_2 if :

$$\text{var}(T_1(X)) < \text{var}(T_2(X)) \tag{0.4}$$

If estimators are not unbiased it is not so straightforward to determine efficiency: we often have to make a choice between estimators that have low bias but high mean squared error and estimators that have high bias but low mean squared error.

Having established some possible criteria by which to judge estimators, we now turn to general procedures for constructing estimators. In this Section we will concentrate on three classical estimation methods, and derive their properties for a certain number of distribution functions. Furtheron in the paper an extremely powerful procedure based on L-Moments estimations, as well as the Bayesian method will be presented.

Method of Moments (MOM)

It is difficult to trace back who introduced the MOM, but Johan Bernoulli (1667-1748) was one of the first who used the method in his work. With the MOM, the moments of a distribution function in terms of its parameters are set equal to the moments of the observed sample. Analytical expressions can be derived quite easily (Appendix), but the estimators can be biased and not efficient. The moment estimators however, can be very well used as a starting estimation in an iteration process.

The central moments of a distribution are given by:

$$\begin{aligned} &\mu_r = E(X - \mu)^r = \int (x - \mu)^r f_X(x) dx \\ &Variance: \sigma^2 = \mu_2 \\ &Skewness: \sqrt{\beta_1} = \frac{\mu_3}{\mu_2^{3/2}} \\ &Kurtosis: \beta_2 = \frac{\mu_4}{\mu_2^2} \end{aligned} \tag{0.5}$$

The sample moments are given by:

$$\begin{aligned} &\bar{x} = n^{-1}\Sigma x_i \\ &m_r = n^{-1}\Sigma (x_i - \bar{x})^r \end{aligned} \tag{0.6}$$

The sample mean $\bar{x}$ is a natural estimator for μ. The higher sample moments m_r are reasonable estimators of the μ_r , but they are not unbiased. Unbiased estimators are

often used. In particular σ^2, μ_3 and the fourth cumulant $\kappa_4=\mu_4 - 3\mu_2^2$ are unbiasedly estimated by:

$$s^2 = (n-1)^{-1}\Sigma(x_i - \bar{x})^2$$

$$m_3^* = \frac{n^2}{(n-1)(n-2)} m_3 \qquad (0.7)$$

$$k_4^* = \frac{n^2}{(n-2)(n-3)}\left\{\frac{n+1}{n-1} m_4 - 3m_2^2\right\}$$

The sample standard deviation $s = \sqrt{s^2}$ is an estimator of σ but is not unbiased. The sample estimators of CV (Coefficient of Variation), skewness and kurtosis are respectively:

$$\hat{C}_v = s / \bar{x}$$

$$g = m_3^* / s^3 \qquad (0.8)$$

$$k = k_4^* / s^4 + 3$$

Finding theoretical moments as a function of θ is not easy for all probability distributions. The method is difficult to generalize to more complex situations (dependent data, covariates, non-identically distributed data). Sample covariances may be used to estimate parameters that determine dependence. For some distributions (such as Cauchy), moments may not exist. In the Appendix moments are given for most familiar PDFs.

Method of Maximum Likelihood (MML)

Also with the MML it is difficult to say who discovered the method, although Daniel Bernoulli (1700-1782) was one of the first who reported about it (Kendall, 1961). The likelihood function gives the relative likelihood of the obtained observations, as a function of the parameters θ:

$$L(\theta,x)=\Pi f(x_i , \theta) \qquad (0.9)$$

With this method one chooses that value of θ for which the likelihood function is maximized. In the Appendix an overview is given of the likelihood functions of some familiar distribution functions. The ML-method gives asymptotically unbiased parameter estimations and of all the unbiased estimators it has the smallest mean squared error. The variances approach asymptotically to:

$$Var(\Theta)= - E(\partial^2 \log L(\theta,x)/\partial\, \theta^2) \qquad (0.10)$$

Furthermore these estimators are invariant, consistent and sufficient. For the definitions we refer to Hald (1952). Analytical expressions for the parameter estimators are sometimes difficult to derive. In those cases, numerical optimization routines have to be used to determine the maximum of the likelihood function, which can also be quite difficult since the optimum of the likelihood function can be extremely flat for large sample sizes. Optimization of the likelihood function may also be hampered by the presence of local maxima. Furthermore:
- MML is (usually) straightforward to implement,
- Maximum likelihood estimators (MLEs) may not exist, and when they do, they may not be unique or give a bias error (Koch, 1991),

- MLE may give inadmissible results (Lundgren, 1988),
- The likelihood function can be used for much more than just finding the MLE: values close to the MLE are more plausible than those further away, for example. This argument can be used to obtain an interval $[p_L,p_U]$ which comprises a plausible range of values for θ,
- ML is adaptable to more complex modeling situations, because the MLE satisfies a very convenient invariance property: If $q=h(\theta)$ where h is a bijective function then $q_{ML}=h(\theta_{ML})$. For example, if $q=1/\theta$, then $q_{ML}=1/\theta_{ML}$. So, having found the MLE for one parameterization, the MLEs for other parameterizations are immediate,
- The maximum likelihood estimator is unbiased, fully efficient (in that it achieves the Cramér-Rao bound under regularity conditions), and normally distributed; all of them in asymptotical sense. Regularity conditions are not fulfilled if the range of the random variable X depends on unknown parameters as is the case for many distributions in the present paper.

The MML is extremely useful since it is often quite straightforward to evaluate from the MLE and the observed information. Nonetheless it is an approximation, and should only be trusted for large values of n (though the quality of the approximation will vary from model to model).

If the available sample sizes are large, there seems little doubt that the maximum-likelihood estimator is a good choice. It should be emphasized, however, that the properties above are asymptotic (large n), and better estimators may be available when sample sizes are small.

Method of Least Squares (MLS)

Least Squares were introduced by Gauss (1777-1855). Given the observations $x=(x_1, x_2, ..., x_n)$ and $y=(y_1, y_2, ..., y_n)$, a regression model can be fitted. For the general case:

$$E(Y|x) = \alpha + \beta x \qquad (0.11)$$

with σ^2 the assumed constant variance of Y around its regression line, the parameter estimates are:

$$\begin{aligned} \alpha^* &= m_Y - \beta^* \, m_X \\ \beta^* &= s_{XY} / s_X^2 \\ \sigma^{2*} &= n/(n-2)\,(1-r_{XY}^2)s_Y^2 \end{aligned} \qquad (0.12)$$

in which m_X, m_Y, s_X, s_Y, s_{XY} and r_{XY} are defined as:

$$\begin{aligned} m_X &= 1/n \, \Sigma_{i=1..n} x_i \\ m_Y &= 1/n \, \Sigma_{i=1..n} y_i \\ s_X^2 &= 1/n \, \Sigma_{i=1..n} (x_i - m_X)^2 \\ s_Y^2 &= 1/n \, \Sigma_{i=1..n} (y_i - m_Y)^2 \\ s_{XY} &= 1/n \, \Sigma_{i=1..n} (x_i - m_X)(y_i - m_Y) \\ r_{XY} &= s_{XY}/s_X s_Y \end{aligned} \qquad (0.13)$$

The estimators α^* and β^* are linear functions of the Y_i 's and they are unbiased. Their variances are:

$\sigma_A^2 = \sigma^2/n\ (1+m_X^2/s_X^2)$
$\sigma_B^2 = \sigma^2 / n s_X^2$ (0.14)
in which A and B are respectively, α^* and β^* treated as random variables.

With the above regression techniques, the MLS can be defined. Assume that we have n observations in sorted order given by $x_{1:n}$, $x_{2:n}$, ..., $x_{n:n}$. Define the plotting position $p_i = i/(n+1)$ of the i-th observation. We want to estimate the optimal value θ of the distribution function $F(x|\theta)$ in least squares sense. The following options are available:

1) Least squares error: $\Sigma_{i=1}^{n} [p_i - F(x_{i:n}|\theta)]^2$ is to be minimized over θ (sometimes also referred to as non-linear regression (Demetracopoulis, 1994)).

Assume that we can linearize the probability distribution F(x) under consideration (with scale parameter B and location parameter A). So we can find a function g s.t.:

$$g(F(x)) = (x-A)/B \quad (0.15)$$

For instance, for the Exponential distribution g is given by $g_E(\zeta) = -\ln(1-\zeta)$ and for the Gumbel distribution, we have $g_G(\zeta) = -\ln(-\ln(\zeta))$ (see Appendix for a complete overview of linearization functions).

2) Least squares linearized error $\Sigma_{i=1}^{n} [g(p_i) - g(F(x_{i:n}|\theta))]^2$ to be minimized over θ.

3) Weighted least squares error $\Sigma_{i=1}^{n} [w_i(p_i) - w_i(F(x_{i:n}|\theta))]^2$ to be minimized over θ, where for instance w_i can weigh the extreme observations to be better fitted by the distribution function (for instance $w_i = 1/(1-p_i)^4$).

4) Least squares to the ordered observations themselves: $\Sigma_{i=1}^{n} [F^{-1}(p_i) - x_{i:n}]^2$ as suggested by Moharram et al. (1993).

In Van Gelder (1996b), several LS-methods are applied to wave data at the location of Pozzallo in Southern Italy. It can be shown that the parameter estimation of the scale of various 2-parameter distributions with LS always leads to larger p-quantiles than with the method of moments. Furthermore, note that under the assumption that the errors in a regression model are independent and normally distributed with zero mean and a fixed standard deviation, the ML-estimators of the error-distribution are exactly the same as the LS-estimators (Gauss, 1777-1855). Problems with LS-estimators are described by McCuen et al. (1990), in which they showed that logarithmic transformations may lead to a biased model.

Method of L-Moments

Hosking, (1990), introduced the L-moments. They have become popular tools for solving various problems related to parameter estimation, distribution identification, and regionalization. It can be shown that L-moments are linear functions of probability weighted moments (PWM's) and hence for certain applications, such as the estimation of distribution parameters, serve identical purposes (Hosking, 1986). In other situations, however, L-moments have significant advantages over PWM's, notably their ability to summarize a statistical distribution in a more meaningful way. Since L-moment estimators are linear functions of the ordered data values, they are

virtually unbiased and have relatively small sampling variance. L-moment ratio estimators also have small bias and variance, especially in comparison with the classical coefficients of skewness and kurtosis. Moreover, estimators of L-moments are relatively insensitive to outliers. These often-heard arguments in favor of estimation of distribution parameters by L-moments (or PWM's) should, nevertheless, not be accepted blindly. *In for instance a wave height frequency analysis, the interest is the estimation of a given quantile*, not in the L-moments themselves. Although the latter may have desirable sampling properties, the same does not necessarily apply to a function of them, such as a quantile estimator. In fact, several simulation studies have demonstrated that for some distributions, *other estimation methods may be superior in terms of mean square errors of quantile estimators* (Hosking and Wallis, 1987; Rosbjerg et al., 1992). As compared with for example the classical method of moments, the robustness vis-à-vis sample outliers is clearly a characteristic of L-moment estimators. However, estimators can be "too robust" in the sense that large (or small) sample values reflecting important information on the tail of the parent distribution are given too little weight in the estimation. Hosking (1990) assessed that L-moments weigh each element of a sample according to its relative importance.

In this section first the theory of L-Moments will be briefly described, followed with an overview of papers with applications of L-moments. The literature review has shown that the theory of L-moments have mostly been applied in a regionalized setting combining data from more than one site. However, in univariate settings the method of L-moments has not been investigated so much. Therefore furtheron in this paper a Monte Carlo experiment is designed in a univariate setting in order to compare the L-moments method with the classical parameter estimation methods (MOM, MML, and MLS). The performance of these methods will also be analyzed w.r.t. inhomogeneous data.

Finally, L-moments are in fact nothing else than summary statistics for probability distributions and data samples. They are analogous to ordinary moments -- they provide measures of location, dispersion, skewness, kurtosis, and other aspects of the shape of probability distributions or data samples -- but are computed from linear combinations of the ordered data values (hence the prefix L). Hosking and Wallis (1997) give an excellent overview on the whole theory of L-Moments.

L-Moments for data samples

Probability weighted moments, defined by Greenwood et al. (1979), are precursors of L-moments. Sample probability weighted moments, computed from data values $x_{1:n}$, $x_{2:n}$, ... $x_{n:n}$, arranged in increasing order, are given by:

$$b_0 = n^{-1}\sum_{j=1}^{n} x_{j:n}$$
$$b_r = n^{-1}\sum_{j=r+1}^{n} \frac{(j-1)(j-2)...(j-r)}{(n-1)(n-2)...(n-r)} x_{j:n} \qquad (0.16)$$

L-moments are certain linear combinations of probability weighted moments that have simple interpretations as measures of the location, dispersion and shape of the data sample. A sample of size 2 contains two observations in ascending order $x_{1:2}$ and $x_{2:2}$. The difference between the two observations $x_{2:2} - x_{1:2}$ is a measure of the scale of the distribution. A sample of size 3 contains three observations in ascending order

$x_{1:3}$, $x_{2:3}$ and $x_{3:3.}$ The difference between the two observations $x_{2:3} - x_{1:3}$ and the difference between the two observations $x_{3:3} - x_{2:3}$ can be subtracted from eachother to have a measure of the skewness of the distribution. This leads to: $x_{3:3} - x_{2:3}$ - ($x_{2:3} - x_{1:3}$) = $x_{3:3} - 2x_{2:3} + x_{1:3}$. A sample of size 4 contains four observations in ascending order $x_{1:4}$, $x_{2:4}$, $x_{3:4}$ and $x_{4:4.}$ A measure for the kurtosis of the distribution is given by: $x_{4:4} - x_{1:4}$ - 3 ($x_{3:4} - x_{2:4}$). In short: the above linear combinations of the elements of the ordered sample contain information about the location, scale, skewness and kurtosis of the distribution from which the sample was drawn. A natural way to generalize the above approach to samples of size n, is to take all possible sub-samples of size 2 and to average the differences ($x_{2:2} - x_{1:2}$)/2:

$$l_2 = \frac{1}{2}\binom{n}{2}^{-1} \sum\sum_{i>j} (x_{i:n} - x_{j:n}) \tag{0.17}$$

Furthermore, the skewness and kurtosis are related with:

$$l_3 = \frac{1}{3}\binom{n}{3}^{-1} \sum\sum\sum_{i>j>k} (x_{i:n} - 2x_{j:n} + x_{k:n}) \tag{0.18}$$

$$l_4 = \frac{1}{4}\binom{n}{4}^{-1} \sum\sum\sum\sum_{i>j>k>l} (x_{i:n} - 3x_{j:n} + 3x_{k:n} - x_{l:n}) \tag{0.19}$$

Hosking (1990) showed that the first few L-moments follow from PWMs via:

$$\begin{aligned} l_1 &= b_0 \\ l_2 &= 2b_1 - b_0 \\ l_3 &= 6b_2 - 6b_1 + b_0 \\ l_4 &= 20b_3 - 30b_2 + 12b_1 - b_0 \end{aligned} \tag{0.20}$$

The coefficients in Eqn. (0.20) are those of the shifted Legendre polynomials. The first L-moment is the sample mean, a measure of location. The second L-moment is (a multiple of) Gini's mean difference statistic (Johnson et al., 1994), a measure of the dispersion of the data values about their mean. By dividing the higher-order L-moments by the dispersion measure, we obtain the L-moment ratios,

$$t_r = l_r/l_2 \tag{0.21}$$

These are dimensionless quantities, independent of the units of measurement of the data; t_3 is a measure of skewness and t_4 is a measure of kurtosis -- these are respectively the L-skewness and L-kurtosis. They take values between -1 and +1 (exception: some even-order L-moment ratios computed from very small samples can be less than -1). The L-moment analogue of the coefficient of variation (standard deviation divided by the mean), is the L-CV, defined by:

$$t=l_2/l_1 \tag{0.22}$$

It takes values between 0 and 1 (if $X \geq 0$).

L-Moments for probability distributions

For a probability distribution with cumulative distribution function F(x), probability weighted moments are defined by:

$$\beta_r = \int x\{F(x)\}^r dF(x), \qquad r = 0,1,2,... \tag{0.23}$$

L-moments are defined in terms of probability weighted moments, analogously to the sample L-moments:

$$\begin{aligned} \lambda_1 &= \beta_0 \\ \lambda_2 &= 2\beta_1 - \beta_0 \\ \lambda_3 &= 6\beta_2 - 6\beta_1 + \beta_0 \\ \lambda_4 &= 20\beta_3 - 30\beta_2 + 12\beta_1 - \beta_0 \end{aligned} \tag{0.24}$$

L-moment ratios are defined by:

$$\tau_r = \lambda_r / \lambda_2 \tag{0.25}$$

The L-moment analogue of the coefficient of variation, is the L-CV, defined by:

$$\tau = \lambda_2 / \lambda_1 \tag{0.26}$$

Examples (for a complete overview, see the Appendix):
Uniform (rectangular) distribution on (0,1):

$$\lambda_1 = 1/2, \quad \lambda_2 = 1/6, \quad \tau_3 = 0, \quad \tau_4 = 0. \tag{0.27}$$

Normal distribution with mean 0 and variance 1:

$$\lambda_1 = 0, \quad \lambda_2 = 1/\sqrt{\pi}, \quad \tau_3 = 0, \quad \tau_4 \approx 0.123. \tag{0.28}$$

The theory of L-moments has been applied in numerous papers. The following work is worth to mention: Rao and Hamed (1997), Duan et al. (1998), Ben-Zvi and Azmon (1997), Van Gelder and Neykov (1998), Demuth and Kuells (1997), Pearson et al. (1991), Ruprecht and Karafilis (1994), Anctil et al. (1998), Lin and Vogel (1993) and Gingras and Adamowski (1994).

Relation of L-Moments with order statistics

Consider a sample consisting of *n* observations $\{x_1, x_2, \ldots, x_n\}$ randomly drawn from a statistical population. If the sample values are rearranged in a non-decreasing order of magnitude, $x_{1:n} \le x_{2:n} \le \ldots \le x_{n:n}$, then the *r*-th member ($x_{r:n}$) of this new sequence is called the *r*-th *order statistic* of the sample (Harter, 1969). When all the sample values come from a common parent population with cumulative distribution function $F(x)$, the probability distribution (CDF) of the *r*-th *order statistic*, i.e., Prob[$X_{r:n} \le x$], means that at least *r* observations in a sample of *n* do not exceed a fixed value, *x*.

A sample randomly drawn from a distribution is analogous to a Bernouilli experiment in which the success is defined by the sampled value being less than the threshold, *x*. Naturally, the probability of success in such an experiment is given as $p = F(x)$, and the number of successes, a random variable, follows the binomial distribution. Based on this argument, the CDF of the *r*-th order statistic, $F_{(r)}(x)$, can be mathematically expressed as

$$F_{(r)}(x) = \sum_{k=r}^{n} \binom{n}{k} F^{k}(x)[1-F(x)]^{(n-k)} \tag{0.29}$$

The incomplete Beta function $I_x(a,b)$ (Kendall and Stuart 1977) is defined via the Beta function B(a,b) as:

$$I_x(a,b) = \frac{1}{B(a,b)} \int_0^x t^{a-1}(1-t)^{b-1} dt$$

in which: (0.30)

$$B(a,b) = \int_0^1 t^{a-1}(1-t)^{b-1} dt = \frac{(a-1)!(b-1)!}{(a+b-1)!} \quad \text{if } a,b > 0$$

So, the expression (0.29) can be written in terms of an incomplete Beta function as:

$$F_{(r)}(x) = r\binom{n}{r} \int_0^{F(x)} u^{r-1}(1-u)^{n-r} du \quad = \quad I_{F(x)}(r,n\text{-}r\text{+}1) \tag{0.31}$$

Indeed; note that $\frac{1}{B(r,n-r+1)} = \frac{n!}{(r-1)!(n-r)!} = r\binom{n}{r}$.

The probability density function of $X_{r:n}$ is given by the first derivative of Eqn. (0.31):

$$f_{(r)}(x) = r\binom{n}{r} F^{r-1}(x)[1-F(x)]^{n-r} f(x) \tag{0.32}$$

Now, the expected value of *r*-th order statistics can be obtained as

$$E[X_{r:n}] = \int_{-\infty}^{\infty} x f_{(r)}(x) dx \tag{0.33}$$

Substituting from eqn.(0.32) into (0.33) and introducing a transformation, $u = F(x)$ or $x = F^{-1}(u)$, $0 \le u \le 1$, leads to

$$E[X_{r:n}] = r\binom{n}{r} \int_0^1 x(u) u^{r-1}(1-u)^{n-r} du \tag{0.34}$$

Note that $x(u)$ denotes the quantile function of a random variable. The expectation of the maximum and minimum of a sample of size n can be easily obtained from eqn.(0.34) by setting $r = n$ and $r = 1$, respectively.

$$E[X_{n:n}] = n\int_0^1 x(u) u^{n-1} du, \quad \text{and} \quad E[X_{1:n}] = n\int_0^1 x(u)(1-u)^{n-1} du \tag{0.35}$$

The probability weighted moment (PWM) of a random variable was formally defined by Greenwood et al. (1979) as

$$M_{i,j,k} = E[X^i u^j (1-u)^k] = \int_0^1 x(u)^i u^j (1-u)^k du \tag{0.36}$$

The following two forms of PWM are particularly simple and useful:

Type 1: $\alpha_k = M_{1,0,k} = \int_0^1 x(u)(1-u)^k du \quad (k = 0, 1, \ldots n)$ (0.37)

and

Type 2: $\beta_k = M_{1,k,0} = \int_0^1 x(u) u^k du \quad (k = 0, 1, \ldots n)$ (0.38)

Comparing eqns.(0.37) and (0.38), it can be seen that α_k and β_k, respectively, are related to the expectations of the minimum and maximum in a sample of size *k*

$$\alpha_k = \frac{1}{k+1} E[X_{1:k+1}] \quad , \qquad \beta_k = \frac{1}{k+1} E[X_{k+1:k+1}] \qquad (k \geq 1) \tag{0.39}$$

In essence, PWM's are the normalized expectations of maximum/minimum of k random observations; the normalization is done by the sample size (k) itself. From Eqn. (0.39), we notice that $E(X_{n:n}) = n\beta_{n-1}$ and from Eqn. (0.23) we have $\beta_{n-1} = \int xF^{n-1}(x)f(x)dx$. So $E(X_{n:n}) = \int xnf(x)F^{n-1}(x)dx$. On the other hand, using Eqn. (0.16), we have $b_{n-1} = n^{-1}\sum_{j=n}^{n} x_j = n^{-1}x_{n:n}$. From this it indeed follows that b_{n-1} is an unbiased estimator of β_{n-1}. Landwehr et al. (1979) gave a proof that b_r is an unbiased estimator of β_r for other values of r.

The expression for β_r in Eqn. (0.23) can numerically calculated by using a plotting-position formula as follows:

$$\beta_r = \int x\{F(x)\}^r dF(x) \approx \sum_{j=1}^{n} x_{j:n}\left(\frac{j}{n+1}\right)^r \frac{1}{n} = n^{-1}\sum_{i=1}^{n} l_i x_{i:n} \tag{0.40}$$

Notice that the expression looks almost the same as Eqn. (0.16) by writing:

$$b_r = n^{-1}\sum_{j=r+1}^{n} \frac{(j-1)(j-2)...(j-r)}{(n-1)(n-2)...(n-r)} x_{j:n} = n^{-1}\sum_{i=r+1}^{n} k_i x_{i:n} \tag{0.41}$$

The terms l_i and k_i in Eqn. (0.40) and (0.41) are compared in Figure 1 and we notice indeed a very close similarity. Reiss (1989) derived more approximate distributions of order statistics and Durrans (1992a) derived distributions of fractional order statistics.

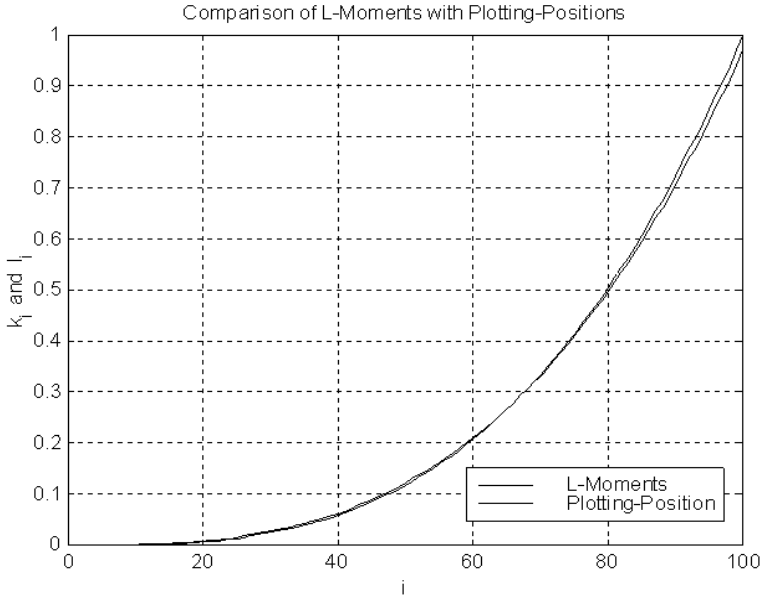


Figure 1: Comparison of Eqn. (0.40) and (0.41).

Summarizing: L-moments are certain linear combinations of probability weighted moments that are analogous to ordinary moments in a sense that they also provide measures of location, dispersion, skewness, kurtosis, and other aspects of the shape of probability distributions or data samples. An r^{th} order L-moment is mathematically defined as:

$$\lambda_r = \sum_{k=1}^{r} p^*_{r-1,k-1}\beta_{k-1} \tag{0.42}$$

where $p^*_{r,k}$ represents the coefficients of shifted Legendre polynomials (Hosking 1990). The following normalized form of higher order L-moments is convenient to work with:

$$\tau_r = \frac{\lambda_r}{\lambda_2} \quad , \quad r = 3,4,\ldots \text{ and } |\tau_r| < 1 \tag{0.43}$$

The normalized fourth order L-moment, τ_4, is referred to as the L-kurtosis of a distribution. Hosking and Wallis (1997) showed that L-moments are very efficient in estimating parameters of a wide range of distributions from small samples. The required computation is fairly limited as compared with other traditional techniques, such as maximum likelihood and least squares. In the Appendix the L-moment formulae are given for a selection of PDFs. Apart from the well-known moments diagrams, also L-Moment diagrams exist in which L-skewness and L-kurtosis are plotted against eachother. However, the L-Moment diagrams do not form a complete class; that is to say points in the diagram may correspond to more than one probability distribution. This is in contrast to the ordinary moment diagram and also to the δ_1-δ_2 diagram of Halphen distributions (Bobee et al., 1993).

The Bayesian method

An introduction

As seen in Section 2.2.3, a consistent method to model parameter uncertainty is given by the Bayesian approach (see also Box and Tiao, 1973). As we will see later, also model uncertainty can be modeled by the Bayesian approach. As an introduction to the Bayesian approach, the following Gumbel model will be analyzed. The Gumbel likelihood model with location λ and scale δ is given by:

$$l(x|\lambda,\delta) = (1/\delta)^n \exp(-(\Sigma x_i-\lambda)/\delta) \exp(-\Sigma\exp(-(x_i-\lambda)/\delta)) \tag{0.44}$$

Where $x=(x_1, x_2, ..., x_n)$. If, for instance, we assume a normal distribution for the location parameter $p(\lambda_G) = N(\lambda_G|\mu_G,\sigma_G)$ then, the posterior distribution of λ becomes:

$$p(\lambda|x)= C\, N(\lambda|\mu_G,\sigma_G)\ (1/\delta)^n\ \exp(-(\Sigma x_i-\lambda)/\delta)\exp(-\Sigma\exp(-(x_i-\lambda)/\delta)) \tag{0.45}$$

in which C a normalisation constant. Figure 2 shows the prior and posterior of the λ-parameter when the following values are given: $\lambda_G = 20$, $\delta_G = 20$, $(\mu_\lambda,\sigma_\lambda) = (17, 2)$, n=150.

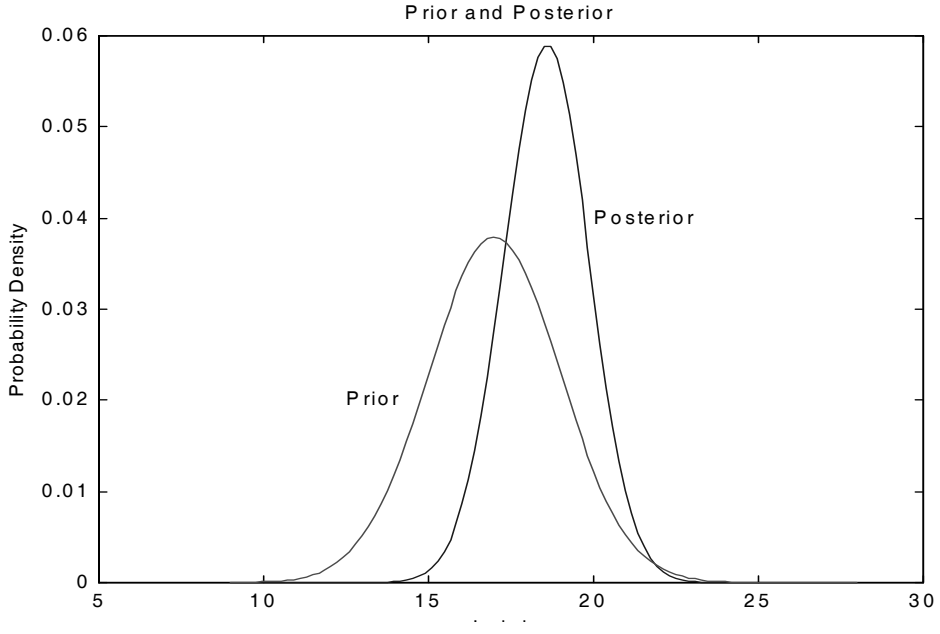


Figure 2: Prior and posterior of the λ-parameter of the Gumbel likelihood model

Notice that the updating process of 150 observations has led to a more peaked and shifted distribution function of the λ-parameter.

The way in which an engineer applies his information about a parameter Θ depends on the objectives in analyzing the data. If he is involved for instance in calculating the frequency of wave heights H, then the inferences he makes on H should reflect the uncertainty in Θ. In the Bayesian framework we are interested in the so-called predictive probability density function:

$$f_H(h)=\int_\Theta f_H(h|\Theta)\, f(\Theta|h_1, h_2, ..., h_n, I)d\Theta \qquad (0.46)$$

where $f_H(h|\Theta)$ is the probabilistic model of wave heights, conditional on the parameters Θ, $f_H(h)$ is the predictive density of the wave heights (now parameter free), and $f(\Theta|h_1, h_2, ..., h_n, I)$ is the posterior density of Θ when both the prior information I and the observations $h_1, h_2, ..., h_n$ are given. In popular words: "the uncertainty in the Θ parameters has been integrated out".

The predictive distribution can be interpreted as being the density $f_H(h|\Theta)$ weighted by $f(\Theta|h_1, h_2, ..., h_n, I)$. Inferences made by combining new information are achieved by updating the distributions of the uncertain parameters through Bayes' theorem and then by calculating the updated predictive function $f_H(h)$.

If we want to "summarize" the posterior distribution of Θ by one parameter we can use the Bayes estimator $\Theta^* = E(f(\Theta))$ or $\Theta^* = \max(f(\Theta))$ (associated with a quadratic loss function and 0-1 loss function respectively). *In making inferences on wave heights it is important to use the predictive function for h*, as opposed to the probabilistic model for h with the Bayes estimator for the parameter set Θ, i.e. $f(q|\Theta^*)$. This is because using point estimators for uncertain parameters underestimates the variance in wave heights.

This is in short the Bayesian way of thinking (see also Kuczera, 1994). A note on the way of implementation is valuable. If we have to calculate an integral of the form $\int f(\theta)g(\theta)d\theta$ we can apply Riemann integration which says:

$$\int f(\theta)g(\theta)d\theta \approx \Sigma\, f(\theta_i)g(\theta_i)\Delta\theta_i \qquad (0.47)$$

where θ_i is a suitable discretisation of θ with discretisation interval $\Delta\theta_i$.
If g is a probability distribution function, we can also apply Monte Carlo simulation. Draw θ_i (i=1..n) from g and we have:

$$\int f(\theta)g(\theta)d\theta \approx \lim_{n\to\infty} 1/n\Sigma_{i=1..n} f(\theta_i) \qquad (0.48)$$

More advanced implementation techniques are available like Markov Chain Monte Carlo methods amongst others the Gibbs sampling method, Metropolis-Hastings method, and more (see Carlin and Louis, 1996).

Obtaining conjugate priors

Provided they are not in direct conflict with our prior beliefs, and provided such a family can be found, the simplicity obtained by using a conjugate prior is very large. But in which situations can a conjugate family be obtained?

It appears that the only case where conjugates can be easily obtained is for data models within the exponential family (Bernardo and Smith, 1994). That is,

$$f(x|\theta) = h(x)g(\theta)e^{t(x)c(\theta)} \tag{0.49}$$

for functions h,g,t and c such that

$$\int f(x|\theta)dx = g(\theta)\int h(x)e^{t(x)c(\theta)}dx = 1 \tag{0.50}$$

This might seem restrictive, but in fact Eqn. (0.49) includes the exponential distribution, the Poisson distribution, the one-parameter Gamma distribution, the Binomial distribution, the Normal distribution (with known variance), and the extreme value distributions. For instance, the Gumbel distribution can be written as:

$$\begin{aligned} f(x\,|\,\alpha,\xi) &= \alpha\ \exp\{-(x-\xi)\alpha\}\exp[-\exp\{-(x-\xi)\alpha\}] = \\ &= \alpha\exp(-\alpha x)\exp(\xi\alpha)\exp[-\exp(-\alpha x)\exp(\xi\alpha)] = h(x)g(\xi)\exp(t(x)c(\xi)) \end{aligned}$$

in which

$$\begin{aligned} h(x) &= \alpha\exp(-\alpha x) \\ g(\xi) &= \exp(\xi\alpha) \\ t(x) &= \exp(-\alpha x) \\ c(\xi) &= -\exp(\alpha\xi) \end{aligned} \tag{0.51}$$

α is considered as a constant; ξ is the conjugate parameter. With a prior of θ, f(θ), we can write:

$$\begin{aligned} f(\theta|x) &\propto f(\theta)l(x|\theta) = f(\theta)\prod_{i=1}^{n} h(x_i)g(\theta)^n e^{\sum_{i=1}^{n} t(x_i)c(\theta)} \propto \\ &\propto f(\theta)g(\theta)^n e^{\sum_{i=1}^{n} t(x_i)c(\theta)} \end{aligned} \tag{0.52}$$

Thus if we choose:

$$f(\theta) \propto g(\theta)^d e^{bc(\theta)} \tag{0.53}$$

In case of the Gumbel distribution, $f(\xi) \propto \exp(d\alpha\xi)\exp(-b\exp(\alpha\xi))$, we obtain:

$$f(\theta|x) \propto g(\theta)^{n+d}\exp\{c(\theta)\left[\sum_{i=1}^{n} t(x_i) + b\right]\} = g(\theta)^{\tilde{d}}e^{\tilde{b}c(\theta)} \tag{0.54}$$

giving a posterior in the same family as the prior, but with modified parameters $\tilde{d} = n + d, \quad \tilde{b} = \sum_{i=1}^{n} t(x_i) + b.$ Indeed, for the Gumbel likelihood model and conjugate prior with parameters d and b, we obtain:

$$f(\xi|x) = e^{d\xi\alpha} e^{-be^{\xi\alpha}} \prod \alpha e^{-\alpha(x_i-\xi)} e^{-e^{-\alpha(x_i-\xi)}} = \alpha^n e^{d\xi\alpha - \Sigma\alpha(x_i-\xi)} e^{-be^{\xi\alpha} - \Sigma e^{-\alpha(x_i-\xi)}} =$$

$$= \alpha^n e^{d\xi\alpha - \alpha\Sigma x_i + n\alpha\xi} e^{-be^{\xi\alpha}} e^{-e^{\alpha\xi}\Sigma e^{-\alpha x_i}} = \alpha^n e^{-\alpha\Sigma x_i} e^{d\xi\alpha + n\xi\alpha} e^{-e^{\alpha\xi}(b+\Sigma e^{-\alpha x_i})} =$$

$$\propto e^{\xi\alpha(n+d)} e^{-e^{\alpha\xi}(b+\Sigma e^{-\alpha x_i})}$$

(0.55)

The use of conjugate priors should be seen for what it is: a convenient mathematical device. However, expression of one's prior beliefs as a parametric distribution is always an approximation. In many situations the richness of the conjugate family is great enough for a conjugate prior to be found which is sufficiently close to one's beliefs (see also Pannullo et al., 1993).

The conjugate priors for the scale parameters of other PDFs can be easily found with the techniques from this section, and they are summarized in the Appendix.

The predictive distribution

So far, we have focused on parameter estimation. That is, we have specified a probability model to describe the random process which has generated a set of data, and have shown how the Bayesian framework combines sample information and prior information to give parameter estimates in the form of a posterior distribution. Commonly the purpose of formulating a statistical model is to make predictions about future values of the process. This is handled much more elegantly in Bayesian statistics than in the corresponding classical theory. The essential point is that in making predictions about future values on the basis of an estimated model there are two sources of uncertainty (Van Gelder, 2000):

- Statistical uncertainty in the parameter values which have been estimated on the basis of past data; and
- Inherent uncertainty due to the fact that any future value is itself a random event.

In classical statistics it is usual to fit a model to the past data, and then make predictions of future values on the assumption that this model is correct, the so-called estimative approach. That is, only the second source of uncertainty is included in the analysis, leading to estimates which are believed to be more precise than they really are. There is no completely satisfactory way around this problem in the classical framework since parameters are not thought of as being random.

Within Bayesian inference it is straightforward to allow for both sources of uncertainty by simply averaging over the uncertainty in the parameter estimates (the information of which is completely contained in the posterior distribution).

So, suppose we have past observations of a variable with density function (or likelihood) $f(x|\theta)$ and we wish to make inferences about the distribution of a future value y from this process. With a prior distribution $f(\theta)$, Bayes' theorem leads to a posterior distribution $f(\theta|x)$. Then the predictive density function of y given the data x is:

$$f(y|x) = \int f(y|\theta) f(\theta|x) d\theta \qquad (0.56)$$

Thus, the predictive density is the integral of the likelihood (of a single observation) times the posterior. Notice that this definition is simply constructed from the usual laws of probability manipulation, and the definition itself has a straightforward interpretation itself in terms of probabilities.

The corresponding approach in classical statistics would be, for example, to obtain the maximum-likelihood estimate θ^* of θ and to base inference on the distribution $f(y|\theta^*)$, the estimative distribution.

To emphasize again, this makes no allowance for the variability incurred as a result of estimating θ, and so gives a false sense of precision (the predictive density $f(y|x)$ is more variable by averaging across the posterior distribution for θ).
For data models within the exponential family, we obtain from (0.49):

$$f(y|x) = \int f(y|\theta) f(\theta|x) d\theta = \int h(y) g(\theta) e^{t(y)c(\theta)} g(\theta)^{\tilde{d}} e^{\tilde{b}c(\theta)} d\theta = h(y) \int g(\theta)^{\tilde{d}+1} e^{c(\theta)(t(y)+\tilde{b})} d\theta$$

Note:

$$\int \theta^{\nu-1} e^{-\mu\theta} d\theta = \frac{\Gamma(\upsilon)}{\mu^{\nu}} \qquad (0.57)$$

The latter expression can be used to solve the integral in case of simple h, g, c and t functions. In case of the Gumbel likelihood model, we can derive (by applying the substitution $u=e^{\alpha\theta}$):

$$f(y|x) \propto h(y) \int u^{\tilde{d}+1} e^{-u(e^{-\alpha y}+\tilde{b})} \frac{du}{\alpha u} = h(y) \frac{1}{\alpha} \int u^{\tilde{d}} e^{-u(e^{-\alpha y}+\tilde{b})} du = e^{-\alpha y} \frac{\Gamma(\tilde{d}+1)}{(e^{-\alpha y}+\tilde{b})^{\tilde{d}+1}} \qquad (0.58)$$

The normalising constant follows from $\int f(y|x)dy = 1$. The posterior functions of the location and shape parameters of the exponential family (except for the normal distribution) cannot be expressed in explicit form. Numerical integration has to be performed. Uniform prior distributions can be used for the scale and location parameters. For a large selection of models, the following information is given in the Appendix: likelihood model, non-informative prior (and corresponding posterior), the conjugate prior, the conjugate posterior, and the conjugate posterior predictive.

Some comments on the Empirical Bayes Method and the Bayes Linear Estimation

The empirical Bayes method is a way of using sample information to assist in specifying the prior distribution. As such, the procedure is not strictly Bayesian, since in a proper Bayesian procedure the prior distribution must be formulated independently of the data. However, the technique is now widely used.

It is clearly very difficult to formulate one's prior information very accurately, and it may be possible only to specify means, variances and covariances with any real faith. However, the posterior distribution will depend on a complete specification of the prior. The Bayes linear estimator of a parameter is an estimator whose value depends only on means and covariances and so does not require a fuller prior specification.

Nonparametric methods

The basic idea of nonparametric density estimation is to relax the parametric assumptions about the data, typically replacing these assumptions with ones about the smoothness of the density. The most common and familiar nonparametric estimator is the histogram. Here the assumption is that the density is fairly smooth (as determined by the bin widths) and an estimate is made by binning the data and displaying the proportion of points in each bin (producing a necessarily non-differentiable, but still useful estimate).

The kernel density estimator is related to the histogram, but produces a smooth (differentiable) estimate of the density. It has been studied widely since its introduction in Rosenblatt (1956). Given i.i.d. data $x_1,...,x_n$ drawn from the unknown density α, the standard kernel estimator (SKE) is the single bandwidth estimator:

$$\hat{\alpha}(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - x_i}{h}\right) \tag{0.59}$$

The bandwidth h determines the amount of smoothing produced by the estimator. See the recent books by Silverman 1986, Scott 1992 and the bibliographies contained therein, for a good introduction to kernel estimators. Much work has been done on selecting the optimal bandwidth h under different assumptions on α or different optimality criteria.

An obvious problem with this kind of estimator for a finite data set is that it uses a single bandwidth (or smoothing parameter) throughout the entire support of the density. For densities with long tails, or modes with different variances, this can be a problem.

The group at Utah Water Research Laboratory, under the guidance of Prof. Upmanu Lall has been working on developing and applying nonparametric estimation techniques to a wide range of surface and groundwater hydrologic problems including time-series forecasting. Research by Upmanu Lall and his co-workers has focused on developing nonparametric statistical methods for the estimation of probabilities of rare floods that are more appropriate in such situations (Moon and Lall, 1994, Moon et al., 1993, and Lall et al., 1993).

For densities with long tails, or modes with different variances, there are modifications to the standard kernel estimator (0.59) proposed by Marchette (1995), and Marchette et al. (1994) which uses a small number of bandwidths rather than a single one as in (0.59), which allows local tuning of the density. As motivation for this estimator, consider a density which consists of a mixture of two normals with different variances which are very far apart. For concreteness, let:

$$\alpha(x) = p\varphi(x, -\mu, \sigma_1^2) + (1 - p)\varphi(x, \mu, \sigma_2^2) \tag{0.60}$$

Suppose we wished to use the standard (single bandwidth) kernel estimator if possible. It seems reasonable that near the left-hand mode one would wish to use a bandwidth appropriate to that normal, and similarly on the right. So a reasonable approach might be to filter the data into two distinct data sets, one from the right component and one from the left, and estimate these two components separately.

One way to do this (approximately) is to define

$$\begin{aligned} \rho_1(x) &= \chi_{\{x>0\}}(x) \\ \rho_2(x) &= \chi_{\{x\le 0\}}(x) \end{aligned} \tag{0.61}$$

and define our estimator to be:

$$\hat{\alpha}(x) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{\rho_1(x_i)}{h_1}K\left(\frac{x-x_i}{h_1}\right)+\frac{\rho_2(x_i)}{h_2}K\left(\frac{x-x_i}{h_2}\right)\right) \tag{0.62}$$

This allows us to use the bandwidths appropriate to the different components in the different regions where they are supported. Equation (0.61) is not quite right, however, for as we move the two components closer together, the overlapping region becomes more and more significant. What we really want to do is use the posteriors for each component as our ρ functions. This is the motivation of the filtered kernel estimator as proposed by Marchette (1995).

With the above example in mind, suppose we wish to have a small number of bandwidths where each bandwidth is associated with a region of the support of the density. To each bandwidth we associate a function which "filters" the data, as in (0.61). Basically, the filter will define the extent to which each local bandwidth is to be used for any particular data point. We can then construct a kernel estimator which is a combination of the kernel estimators constructed using each bandwidth, with the data filtered by the filtering functions.

Another method of nonparametric estimations is the following. Suppose we have independent random variables X_1 ; X_2 ; ... , all of them with distribution function F . Suppose that F is in the domain of attraction of some extreme value distribution G_γ. In other words: suppose 1-F is regularly varying at infinity, i.e.,

$$\lim_{t\to\infty}\frac{1-F(tx)}{1-F(t)} = x^{-1/\gamma} \tag{0.63}$$

for x>0, where γ is a positive parameter.
Let $X_{1:n}$ $X_{2:n}$ .... $X_{n:n}$ be the n-th order statistics. For some m < n define:

$$\hat{\gamma}_p = (\log 2)^{-1}\log\frac{X_{n-m:n}-X_{n-2m+1:n}}{X_{n-2m+1:n}-X_{n-4m+1:n}} \tag{0.64}$$

This is Pickands' estimator for γ (Pickands, 1975).
Another estimator for γ is the moment estimator (Dekkers and De Haan, 1989):

$$\hat{\gamma}_M = M_n^{(1)} + 1 - \frac{1}{2}\left\{1-\frac{(M_n^{(1)})^2}{M_n^{(2)}}\right\}^{-1} \tag{0.65}$$

in which:

$$M_n^{(i)} = \frac{1}{k}\sum_{i=0}^{k-1}(\log X_{n-i:n} - \log X_{n-k:n})^i \tag{0.66}$$

The estimators are consistent for any γ. Under appropriate smoothness conditions on F and a further bound on the rate of increase of k(n) the estimators for γ are asymptotically normal after normalization (see e.g. Dekkers and de Haan, 1989).

Performance of the statistical estimation methods

In the previous six sections an overview of statistical estimation methods has been given. The large number of distributions and estimation methods proposed in the literature may cause confusion which method and/or distribution to use (Bobee et al., 1993b). The World Meteorological Organization (1989) and Cunnane (1987-1988) published a few reports which compare current methodologies, and recommend a number of statistical distributions and estimation procedures. These reports are already ten years old and the last decade has been extremely busy with new results on this topic. In the present section we report the latest developments, show advantages and disadvantages of the various methods, and propose new ideas for a possible comparison strategy.

Variability of estimators

As written earlier in the paper we prefer estimators which have low variability. But the question arises as to whether there is any limit to the accuracy that an estimator can achieve (Beard, 1994). Intuitively, we would expect that there is, since the variation from one sample to another means that any estimator is bound to have some degree of variation. However, is it possible to bound the accuracy which an estimator can achieve? In fact, provided we restrict ourselves to unbiased estimators there is a remarkable theorem which provides a bound on the variance of any (unbiased) estimator (Barnett, 1973).
The quantity :

$$I(p)=-1/nE(\partial^2 \log L(p,x)/\partial p^2) \qquad (0.67)$$

is extremely important in statistical theory. It is termed the Expected Information (and related with the Fisher matrix of Eqn. (2.2) up to a factor n^{-1} (Fisher (1934)). Since second derivatives measure curvature, I(p) is a measure of the expected curvature of the likelihood function at the true parameter value. The importance of the expected information in the context of minimum variance estimation is provided by the following result, which is known as the Cramér-Rao theorem (1946):

If T(X) is an unbiased estimator of p then

$$\mathrm{Var}(T(X)) > I^{-1}(p) \qquad (0.68)$$

Thus, within the class of unbiased estimators, no estimator can have a variance which is smaller than the reciprocal of the expected information. So, as required, we obtain a bound on the maximum precision that can be attained within the class of unbiased estimators. Note that the bound only applies to unbiased estimators. It is always possible to obtain an estimator with lower variance by giving up on the property of unbiasedness. (For example, the estimator T(X)=100 has zero variance for any problem).

The result is important for both theoretical and practical reasons; it formulates in a precise way the limits that can be achieved in statistical inference due to random variation in the population. Though the Cramér-Rao bound puts a lower limit on the variance of unbiased estimators, the bound may not be achievable.

So far we have used the likelihood function only to determine an estimate of an unknown parameter θ. We have noted several times, however, that values of θ with relatively high likelihood are more plausible than those with low likelihood, and that it should be possible to exploit this knowledge to construct a credible range for θ. This is achieved formally by looking at the sampling properties of the likelihood function: quantifying the variation in $L(\theta)$ from sample to sample.

In most situations exact calculation is very difficult and we will be forced to use (asymptotic) approximations that assume large sample sizes. In the next Sub-section we present some methods which are useful in deriving the sampling properties of the likelihood function. It will be shown that these methods can be used to derive the sampling distributions of the ML and LM estimators of the exponential distribution.

Methods for deriving the sampling distribution of estimators

Mood et al. (1974) present methods for deriving the sample distribution of estimators. In some cases it is possible to establish the exact sampling distribution of an estimator and use this as the basis for confidence interval estimation. More generally this is not possible but approximations can be obtained. The following relations can be used in deriving the exact sampling distributions:

If X and Y are random variables with PDF's f and g respectively. Let

$$Z=X+Y,\ U=X-Y,\ V=XY \text{ and } W=X/Y$$

then the PDF's of Z, U, V and W are, respectively, given by:

$$\begin{aligned} f_Z(z) &= \int f(x)g(z-x)\,dx \\ f_U(u) &= \int f(u+y)g(y)\,dy \\ f_V(v) &= \int f(x)g(v/x)\,|x|^{-1}\,dx \\ f_W(w) &= \int f(xw)g(x)\,|x|\,dx \end{aligned} \tag{0.69}$$

The proof of these formulae is given in Mood et al. (1974). In textbooks on statistics the following relation is proven:

$$E(XY) = E(X)E(Y) \tag{0.70}$$

$$Var(\sum_{i=1}^{n} a_i X_i) = \sum_{i=1}^{n} a_i^2 Var(X_i) + 2\sum_{i=1}^{n}\sum_{j=i+1}^{n} a_i a_j Cov(X_i, X_j)$$

Furthermore it is possible to derive the following property for the product of random variables:

$$Var(V) = Var(X)Var(Y) + E^2(X)Var(Y) + E^2(Y)Var(X) \quad (0.71)$$

If exact calculations are not possible, the following approximation rules can be used (using Taylor's formula):

$$g(X) = g(m_X) + (X - m_X)\frac{dg(x)}{dx}_{|x=m_X} + \frac{(X - m_X)^2}{2}\frac{d^2 g(x)}{dx^2}_{|x=m_X} + \ldots \quad (0.72)$$

From this, we can derive:

$$E(g(X)) \approx g(E(X)) \quad (0.73)$$

$$Var(g(X)) \approx Var(X)\,[g'(m_X)]^2 \quad (0.74)$$

If the coefficient of variation of X is less than c, the error involved in these approximations is less than c^2. In particular, the following useful approximations can be used:

$$\begin{aligned} E(\sqrt{X}) &\approx \sqrt{E(X)}, \qquad Var(\sqrt{X}) \approx \frac{Var(X)}{4E(X)} \\ E(X^{-1}) &\approx \frac{1}{E(X)}, \qquad Var(X^{-1}) \approx \frac{Var(X)}{E^4(X)} \end{aligned} \quad (0.75)$$

From Eqn. (0.75) the approximation $CV\left(\frac{1}{X}\right) = \frac{\sigma\left(\frac{1}{X}\right)}{E\left(\frac{1}{X}\right)} \approx \frac{\frac{\sigma(X)}{E^2(X)}}{\frac{1}{E(X)}} = \frac{\sigma(X)}{E(X)} = CV(X)$ follows directly. This approximation was shown to be very useful in Sec. 2.6.

The second equation in (0.75) is useful in deriving the sampling distribution of the standard devation. It is well known that $Var(s^2) = 2\sigma^4/N$ in which σ is the population standard deviation of the normal distributed X_i and all X_i's are i.i.d. Therefore:

$$\sigma(s) = \sqrt{\frac{\frac{2\sigma^4}{N}}{4\sigma^2}} = \frac{\sigma}{\sqrt{2N}} \quad (0.76)$$

Exact calculations for the estimations of the 2-parameter Exponential distribution will be given in the next sub-section.

Exact sampling distributions for the Exponential distribution

The techniques from the previous section will shown to be succesful in deriving the exact sampling distributions for the parameter estimations of the MML and MLM of the exponential distribution.

The expressions of the MML and MLM for the location and scale parameter of the exponential distribution are given in the Appendix. Suppose that $X_1, X_2, \ldots, X_n \sim$

Exp(ξ,α) i.i.d. This is the usual exponential model parameterized in such a way that ξ+α is the population mean. The sampling distribution of $m_X = \frac{1}{n}\sum_{i=1}^{n} X_i$ is the sum of n exponential distributions divided by n. According to Johnson et al. (1994), a summation of n independent exponential distributions, each with scale parameter α is gamma distributed with parameters (n, α), and a constant k times a gamma distribution with parameters (α,β) is also gamma distributed with parameters (α,β/k). Therefore m_X is ξ + gamma distributed with parameters (n,αn).

Furthermore $P(\min(X_1, X_2, ..., X_n) < x) = 1 - P(\min(X_1, X_2, ..., X_n) > x) =$
$= 1 - P^n(X_i > x) = 1 - (1 - P(X_i < x))^n$. So the minimum of n exponential distributions, each with scale parameter α, is also exponentially distributed with scale parameter α/n. Finally, Sukhatme (1937) showed that $2n/\alpha(m_X - \min((X_1, X_2, ..., X_n))$ has a Chi-square distribution with 2(n-1) degrees of freedom.

With the above considerations and with Eqn. (0.70) we can derive that the following properties for the sampling distribution of the MLEs:

$$\begin{aligned} &E(\alpha_{ML}) = \alpha(1-1/n) \\ &Var(\alpha_{ML}) = \alpha^2(n^{-1} - n^{-2}) \\ &E(\xi_{ML}) = \xi + \alpha/n \\ &Var(\xi_{ML}) = \alpha^2/n^2 \end{aligned} \tag{0.77}$$

In order to derive the sampling distribution for the MLM, we use the following properties of the order statistics of the standard exponential distribution (Johnson et al., 1994):

$$\begin{aligned} &E(X_{(i)}) = \sum_{j=1}^{i} \frac{1}{n-j+1} \\ &Var(X_{(i)}) = \sum_{j=1}^{i} \frac{1}{(n-j+1)^2} \\ &Cov(X_{(i)}, X_{(k)}) = \sum_{j=1}^{i} \frac{1}{(n-j+1)^2} = Var(X_{(i)}) \end{aligned} \tag{0.78}$$

The following summations have been derived:

$$\begin{aligned} &\sum_{j=1}^{n}\sum_{k=1}^{j} \frac{1}{n-k+1} = n \\ &\sum_{j=2}^{n}\sum_{k=1}^{j} \frac{j-1}{n-k+1} = c(n) \\ &\sum_{j=2}^{n} \frac{2j-1-n}{n(n-1)} = 1 + \frac{1}{n} \end{aligned} \tag{0.79}$$

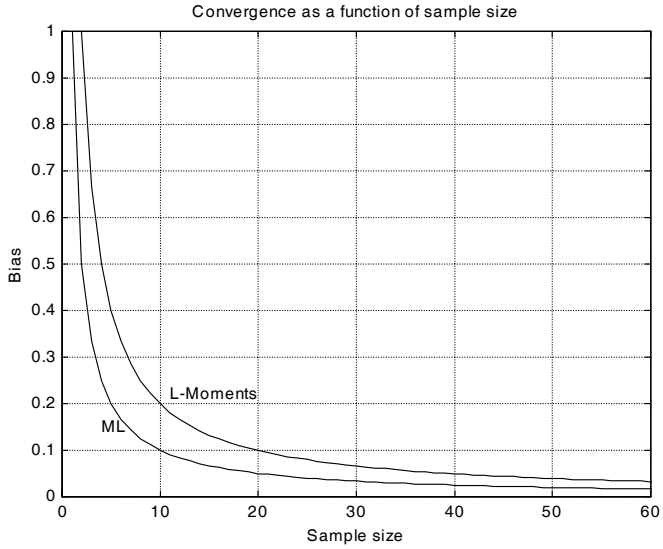
The function c(n) can be derived to $c(n) = \Psi(n)\frac{(n-1)n}{2} - \sum_{j=1}^{n-1}(n-j)\Psi(j)$, in which Ψ is the Psi-function $\Psi(x) = \frac{\Gamma'(x)}{\Gamma(x)}$, having the property $\sum_{i=k}^{n}\frac{1}{i} = \Psi(n) - \Psi(k)$ (Abromowitz and Stegun, 1965). The function c(n) behaves as n^2 for large values of n. We write $c(n)=O(n^2)$.

For n=2 the L-Moments estimator of α reads $\alpha_{LM} = x_{2:2} - x_{1:2}$. The difference $x_{2:2} - x_{1:2}$ has an exponential distribution with parameter α. Consequently $E(\alpha_{LM}) = \alpha$ (unbiased) and $Var(\alpha_{LM}) = \alpha^2$ for n=2. The Eqn. (0.79) and the asymptotic covariance matrix of the sample L-moments of Hosking (1986) are necessary to derive the following sampling properties of MLM:

$$
\begin{aligned}
&E(\alpha_{LM}) = \alpha(4n^{-1}(n-1)^{-1}c(n)-2) \\
&Var(\alpha_{LM}) = 4\alpha^2 / 3n + O(n^{-2}) \\
&E(\xi_{LM}) = \xi + \alpha(4n^{-1}(n-1)^{-1}c(n)-3) \\
&Var(\xi_{LM}) = \alpha^2 /3n + O(n^{-2})
\end{aligned}
\qquad (0.80)
$$

Note that the variance of α_{LM} is slightly larger than the variance of α_{ML}. The variance of ξ_{LM} decreases with $O(n^{-1})$ instead of $O(n^{-2})$ in case of the maximum likelihood method. Furthermore, the expressions $3- 4n^{-1}(n-1)^{-1}c(n)$ and n^{-1} can be plotted in Figure 3 for a graphical comparison of the biases:

Figure 3: Expressions $3 - 4n^{-1}(n-1)^{-1}c(n)$ and n^{-1}



Notice that the ML-method has a slightly better performance for the parameter estimation of the exponential distribution than the L-Moments method; its bias and RMSE is lower than the MLM-equivalents.

Performance based on relative bias and RMSE of the estimators

With Monte Carlo simulation studies, datasets can be generated from a beforehand known probability distribution function (and known p-quantile). Different parameter estimation methods can be applied on these datasets and compared with respect to their estimates of the p-quantiles. The estimation method with the smallest bias and/or variance is then considered to be the best method for that particular distribution function. In order to familiarize the reader with the concept, in the following we start with a simulation analysis of six estimation methods for the scale parameter of a one-parameter exponential distribution $F(x)=1-e^{-\lambda x}$. In Figure 4 we see the performances of the six methods for one simulation of 20 values the exponential distribution with

scale parameter λ=1.5. The different methods cause a very wide range of frequency lines. The 100 year event (or the 0.99 quantile) of the theoretical distribution is 3.07m.

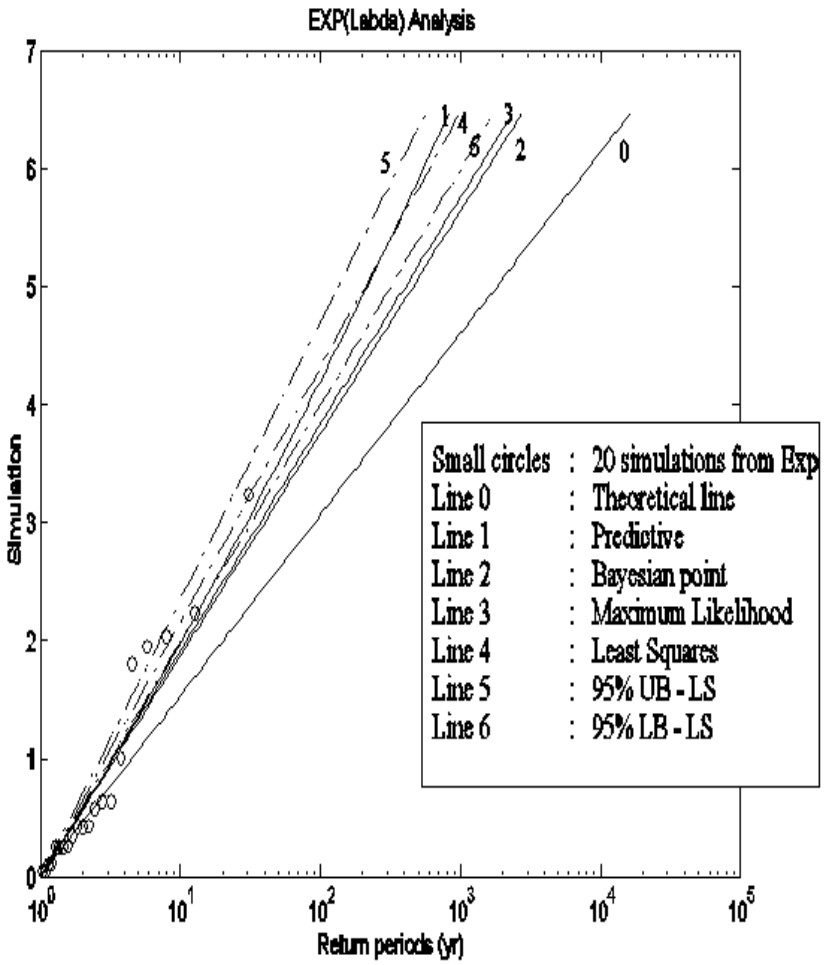


Figure 4: Performance of six methods

The above simulation analysis is repeated 200 times and from each simulation we store per estimation method the results of the 1/100 year prediction. This excersise is also repeated for other sample sizes. Apart from n=20 values we look at n=3, 6, 9,..., 60 values. In the next Figure 5 we have plotted the mean and standard deviations of each estimation method.

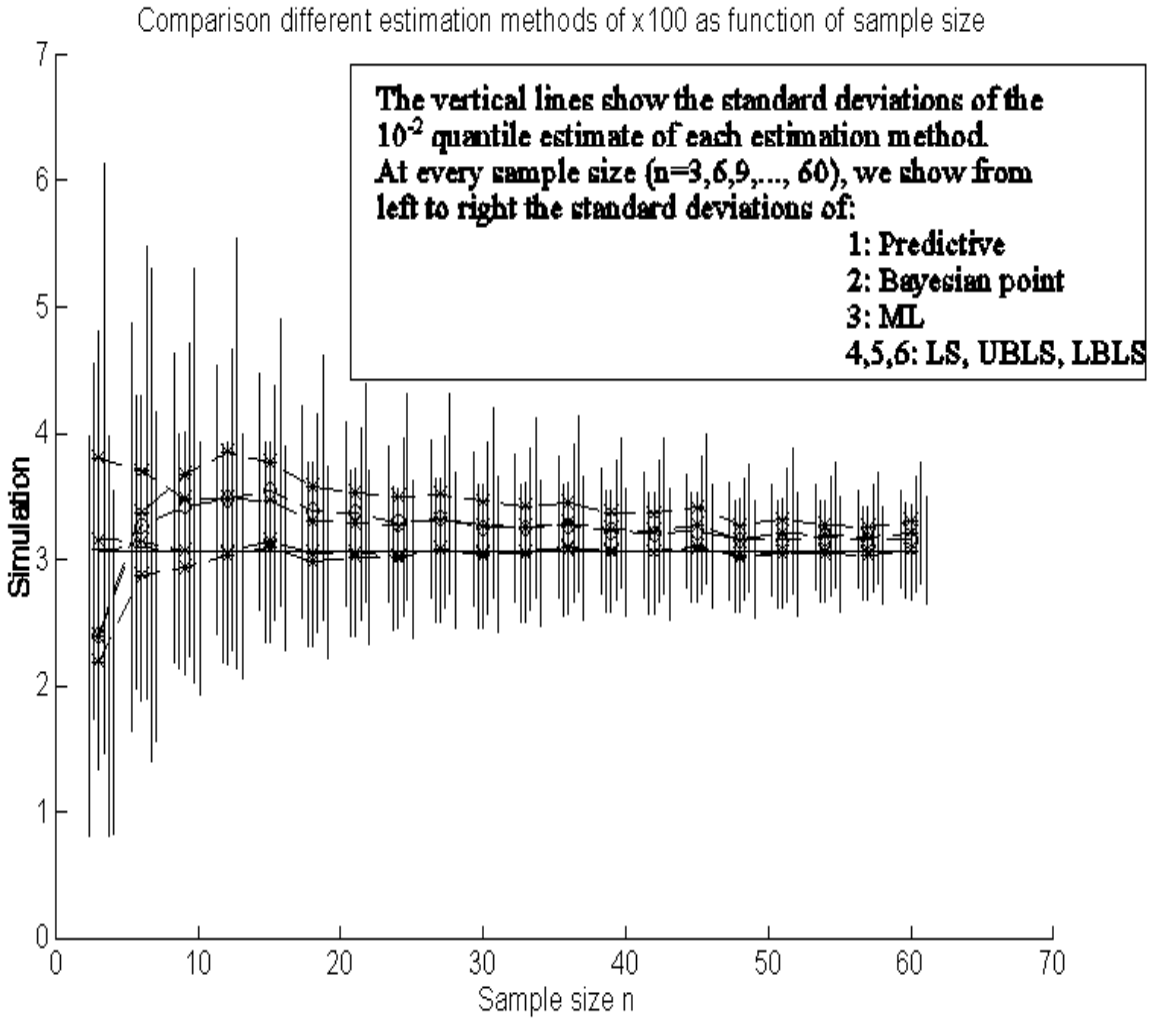


Figure 5: Performance of six estimation methods as function of sample size (source: Van Gelder and Vrijling (1997a))

In general it is impossible to say which estimation method is the most appropriate method for a particular model and dataset. This depends on the size of the sample, the type of the distribution, the choice of the parameters of the distribution, the inhomogeneity that is embedded in the data, and of course the choice of the criterion. However, a Monte Carlo simulation is the suitable method to examine the performance for a certain choice of the above mentioned dependencies.

Lots of simulation work has been performed to judge the performance of the estimation methods based on the relative bias and RMSE of the distribution parameters. An overview of this work is given in the following Table 1:

Table 1: Literature review (only first authors are shown)

	MOM	LM (or PWM)	ML	LS	Bayes	ME
GUM	Takara, 1989 Yamaguchi, 97 Carter, 1983 Landwehr, 79	Takara, 1989 Yamaguchi, 97 Landwehr, 79 Guo, 1991	Takara, 1989 Yamaguchi, 97 Carter, 1983 Corsini, 1995	Takara, 1989 Yamaguchi, 97 Carter, 1983	Coles, 1996	Takara, 1989

WEIB	Yamaguchi, 97 Abernethy, 83	Yamaguchi, 97	Yamaguchi, 97 Smith, 1987	Yamaguchi, 97	Smith, 1987	Singh, 1990
GPA	Moharran, 93 Hosking 1987 Castillo 1997	Moharran, 93 Hosking 1987 Castillo 1997	Moharran, 93 Hosking 1987 Castillo 1997	Moharram, 93	Coles, 1996	Singh, 1997
GEV	Takara, 1989 Yamaguchi, 97 Sank., 1999	Takara, 1989 Yamaguchi, 97 Lu, 1992b Wang. 1998	Takara, 1989 Yamaguchi, 97 Hosking, 1985	Yamaguchi, 97	Fill, 1998 Coles, 1996 Preumont, 88	Singh, 1992 Jowitt, 1979 Lind, 1991
LN	Stedinger,1980 Hoshi, 1986 Yamaguchi, 97 Goda, 1992	Takara,1990 Sank., 1999 Yamaguchi, 97 Takeuchi 1988	Stedinger,1980 Takara, 1990 Yamaguchi, 97 Takeuchi 1988	Takara, 1990 Lechner, 1991	Corbyn 1988 Lye, 1988	Singh, 1987
GAM	Hoshi, 1986 Bobee, 1991	Rasmussen, 94 Wu, 1991 Durrans, 92b	Hoshi, 1986 Hu, 1987 Stacy, 1965	Ashkar, 1998	vNoortwijk, 99 Ribeiro 1993	Singh 1985

Maximum likelihood estimation of the generalized Pareto distribution (GPA) has previously been considered in the literature, but Hosking et al. (1987) show that unless the sample size is 500 or more, estimators derived by the method of moments or the method of probability-weighted moments are more reliable. They also use simulations to assess the accuracy of confidence intervals for the parameters and quantiles of the generalized Pareto distribution.

Various estimation methods for the three-parameter case of Generalized Pareto distribution are given in Moharram et al. (1993), including an alternative method based on least squares (LS). Modified formulae for computing estimators are provided. The performances of these methods are compared by using Monte Carlo simulation. It is found that the LS method has generally a lower root mean squared error (RMSE) than that obtained using other methods. The LS method also performs best in terms of BIAS when the shape parameter is greater than zero, while the probability weighted moments (PWM) method performs best when the shape parameter less than zero.

In Castillo and Hadi (1997), it is shown that when the shape parameter of the GPA is greater than 1, the maximum likelihood estimates do not exist, and when the shape parameter is between 1/2 and 1, they may have problems. Furthermore, for shape parameters less than or equal to -1/2, second and higher moments do not exist, and hence both the method-of-moments (MOM) and the probability-weighted moments (PWM) estimates do not exist. Another and perhaps more serious problem with the MOM and PWM methods is that they can produce nonsensible estimates (i.e., estimates inconsistent with the observed data). In Castillo and Hadi (1997), a simulation study is carried out to evaluate the performance of the parameter estimation methods and to compare them with other methods suggested in the literature. The simulation results indicate that no method is uniformly best for all the parameter values.

In Fill and Stedinger (1995), it was shown that for realistic generalized extreme value (GEV) distributions and short records, a simple index-flood quantile estimator performs better than two-parameter (2P) GEV quantile estimators with probability weighted moment (PWM) estimation using a regional shape parameter and at-site mean and L-coefficient of variation (L-CV), and full three-parameter at-site GEV/PWM quantile estimators. However, as regional heterogeneity or record lengths increase, the 2P-estimator quickly dominates. Fill and Stedinger (1995) generalizes the index flood procedure by employing regression with physiographic information to refine a normalized T-year flood estimator. A linear empirical Bayes

estimator uses the normalized quantile regression estimator to define a prior distribution which is employed with the normalized 2P-quantile estimator. Monte Carlo simulations indicate that this empirical Bayes estimator does essentially as well as or better than the simpler normalized quantile regression estimator at sites with short records, and performs as well as or better than the 2P-estimator at sites with longer records or smaller L-CV.

In Wang (1998), approximate goodness-of-fit tests of fitted generalized extreme value (GEV) distributions using LH moments are formulated on the basis of comparison of sample LH kurtosis estimates and theoretical LH kurtosis values of the fitted distributions. Their tests are different from those that have been derived for testing the GEV distributions of which parameter values are known a priori. The tests are intended to answer the following questions: Does a fitted GEV distribution describe adequately a given data series? If not, can the GEV distribution function describe adequately the larger events in that data series for use for high quantile estimation? If so, what degree of emphasis on the larger events is needed in order that the GEV distribution becomes acceptable? The use of the GEV distribution in conjunction with the LH moment estimation method and the formulated tests should alleviate the need for finding the "correct" distribution. The tests are evaluated by Monte Carlo simulations using generated samples of both GEV and Wakeby distributions.

Takeuchi and Tsuchiya (1988) derive PWM solutions for Normal and 3-parameter Lognormal distributions. Their paper presents their relative accuracy in comparison with other parameter estimation procedures such as Moment, Maximum-likelihood, Quantile and Sextile methods through Monte Carlo simulation experiments. Simulation results revealed that PWM estimates of quantiles are unbiased for the Normal distribution and less biased than those of the Moment method for Lognormal distribution with a large coefficient of skewness. It was also revealed that the RMSE of PWM estimates of quantiles is as small as that of the Moment method for the Normal distribution but larger for the Lognormal distribution.

In Lechner (1991) three common estimators for the parameters of the lognormal distribution are evaluated. Correction factors which eliminate essentially all the bias, and formulas for the standard deviations of the estimators, are presented. It is reported that the Persson-Rootzen estimators are about as good as the maximum-likelihood estimators, without the penalty of requiring iterative optimization. Also, the estimators resulting from (least squares) fitting a line to the plot of log lifetimes on normal (Gaussian) probability paper are reasonably good. Formulas are given for obtaining these latter estimators without actually plotting the points. Lechner (1991) simulated 5k to 30k samples (more samples for smaller N for each case) and calculated the following: the means, standard deviations, and third moments of each estimator; correlations between the two members of each pair; comparisons between the estimators; and simple corrections to improve the performance of the estimators.

In Corbyn (1988), methods are developed for the determination of the posterior distribution of the first moment of the lognormal distribution with exponential and other prior distributions. Bayesian methods of statistical inference are compared with the more generally used method of inference based on confidence limits. The general problem of Bayesian estimation of the mean of a correlated random variable is discussed.

Sankarasubramanian et al. (1999) deals with fitting of regression equations for the sampling properties, variance of L-standard deviation, and bias and variance of L-skewness, based on Monte-Carlo simulation results, for generalised Normal

(Lognormal-3) and Pearson-3 distributions. These fitted equations will be useful in formulating goodness-of-fit test statistics in regional frequency analysis. The second part of their paper presents a comparison of the sampling properties between L-moments and conventional product moments for generalised Normal, generalised Extreme Value, generalised Pareto and Pearson-3 distributions, in a relative form. The comparison reveals that the bias in L-skewness is found to be insignificant up to a skewness of about 1.0, even for small samples. In case of higher skewness, for a reasonable sample size of 30, L-skewness is found to be nearly unbiased. However, the conventional skewness is found to be significantly biased, even for a low skewness of 0.5 and a reasonable sample size of 30. The overall performance evaluation in terms of "Relative-RMSE in third moment ratio" reveals that conventional moments are preferable at lower skewness, particularly for smaller samples, while L-moments are preferable at higher skewness, for all sample sizes.

Corsini et al. (1995) analyze the Maximum likelihood (ML) algorithms and Cramer-Rao (CR) bounds for the location and scale parameters of the Gumbel distribution are discussed. First they consider the case in which the scale parameter is known, obtaining the estimator of the location parameter by solving the likelihood equation and then evaluating its performance. They also consider the case where both the location parameter and the scale parameter are unknown and need to be estimated simultaneously from the reference samples. For this case, performance is analyzed by means of Monte-Carlo simulation and compared with the asymptotic CR bound.

Also results of a Monte Carlo study are presented in Guo and Cunnane (1991) comparing different simulation procedures and assessing the value of historical floods for at-site flood frequency analysis on the assumption of a Gumbel distribution.

In Wu et al. (1991), a new procedure, the method of lower-bound (MLB), is proposed for determining the design quantile Xp. Their basic concept is first to determine an estimate of the location parameter using the probability weighted moment (PWM) method, and then to transform the variable X from the original (X) space to a new (Y) space. The variable Y is considered to have a two parameter gamma distribution. In Y-space, the two parameters are estimated by the PWM or an autocovariance method, then transformed from the Y-space to the X-space. Results from the Monte Carlo experiments show that the MLB estimates are less biased than comparable moment estimates and maximum likelihood estimates, and more efficient than those of PWM for design quantiles x_p.

In Hu (1987a), the determination of confidence intervals for design floods using the Pearson Type III distribution with a known skewness is analyzed. Tables for the confidence factors based on moment and curve-fitting estimates were developed by Monte Carlo simulation technique and were used to construct the confidence intervals for frequency curves. The performance of methods using tables presented herein, the method of B values based on the curve-fitting method are evaluated.

In Naghavi and Yu (1996), it is shown that the quantile prediction accuracy of the log-Pearson type III (LP3) distribution depends largely on the accuracy of the parameter-estimation method used. The performance of a parameter-estimation method, on the other hand, depends on both the individual population chosen from the LP3 family and the sample size. In this study Monte Carlo experiments were conducted to evaluate four parameter-estimation methods that are frequently used in hydrological analysis. The four methods tested are the method of indirect moments (MMI), the method of mixed moments (MIX), the method of direct moments (MMD), and a modification of MMI using optimization techniques (MMO). A quantile ratio index (QRI) was devised to identify the limits (sample size and LP3 population

subset) within which each of these methods will perform best. This study suggested that when QRI is less than or equivalent to 1.14, MMI or MMO should be used for sample size N less than equivalent to 30, MIX for 30 less than N less than 100, and any of the four methods for N greater than equivalent to 100. When QRI greater than 1.14, MMO is recommended for N less than equivalent to 30, MIX for 30 less than N less than 100, and MIX, MMO, or MMI for N greater than equivalent to 100.

In Pilon and Adamowski (1993), maximum likelihood and censored sample theory are applied for flood frequency analysis purposes to the log Pearson Type III (LP3) distribution. The logarithmic likelihood functions are developed and solved in terms of fully specified floods, historical information, and parameters to be estimated. The asymptotic standard error of estimate of the T-year flood is obtained using the general equation for the variance of estimate of a function. The variances and covariances of the parameters are obtained through inversion of Fisher's information matrix. Monte Carlo studies to verify the accuracy of the derived asymptotic expression for the standard errors of the 10, 50, 100, and 500 year floods, indicate that these are accurate for both Type I and Type II censored samples, while the bias is less than 2.5%. Subsequently, the Type II censored data were subjected to a random, multiplicative error. Results indicate that historical information contributes greatly to the accuracy of estimation of the quantities even when the error of its measurement becomes excessive. (Type I: The threshold is fixed and the number of censored values is a random variable. Type II: The number of censored values is fixed and the threshold is a random variable).

In Lye et al. (1988), the three-parameter lognormal distribution is studied with Bayesian estimates of the parameters and of the T-year quantile and their posterior variances of the estimates are obtained by using Lindley's Bayesian approximation procedure. These estimates are compared to estimates obtained by the method of maximum likelihood. In all cases the posterior variances of the Bayes estimates of the T-year flood events are less than the corresponding variances of their maximum likelihood counterparts.

In Lye et al. (1993), ML estimators were compared with Bayesian estimators for the reliability functions of the extreme value distributions.

In Yamaguchi (1997), Monte Carlo simulations have been performed to determine a preferable parameter estimation method for each of eight distribution functions. It was also shown that a jackknife method is beneficial to correct the bias and RMSE irrespective of the estimation method used.

In Takara and Stedinger (1994) the use of two-parameter distributions is recommended from the viewpoint of quantile estimation accuracy for datasets having sample skewness greater than 0.38 and less than 1.8. They found that the quantile lower bound estimators are likely to provide more accurate quantile estimators than other procedures.

Singh and Guo (1997) and Singh and Singh (1985, 1987) showed that the Method of Entropy yielded parameter estimates for the Generalized Pareto, the Gamma and the lognormal distributions which were comparable or better within certain ranges of sample size and coefficient of variation in comparison with MOM, PWM and ML.

Performance based on over- and underdesign

In Van Gelder (1996b) it was suggested to measure the performance of statistical estimation methods with respect to over- and underdesign. In fact there is a strong relation with the performance based on relative bias and RMSE from the previous section. The relative bias and RMSE can give a first indication how much the under- or overdesign is, however, in case of very skewed distributions of the quantile, this indication might give a false impression of the amount of under- or overdesign. Under- or overestimation of the p-quantiles have an important meaning in civil engineering practice as well. Underestimation may give rise to unsafe structures whereas overestimation may lead to conservatism or too expensive structures. Therefore it is very useful to study the probabilities of under- and overdesign of a certain estimation method.
As a typical result from Van Gelder (1996b), in this section, we will in particularly concentrate on the under- and overestimation of the p-quantile of an Exponential and Gumbel distribution with a ML- and LS-parameter estimation method. Different sample sizes are considered (n=10, 30 and 100) for the same quantile of interest x_{100} such that $P(x>x_{100} \mid \text{Data}) = 1/100$. Typical results are shown in the next Tables 3 and 4.

Data from Gumbel	n	ML	LS
Fitted by Gumbel	10	0.59	0.34
	30	0.56	0.36
	100	0.53	0.38
Fitted by Exponential	10	0.18	0.19
	30	0.01	0.12
	100	0.00	0.05

Table 3: Probabilities of underdesign p_u

Data from Exponential	n	ML	LS
Fitted by Gumbel	10	0.85	0.50
	30	0.96	0.55
	100	0.99	0.67
Fitted by Exponential	10	0.59	0.37
	30	0.54	0.37
	100	0.52	0.37

Table 4: Probabilities of underdesign p_u

The probabilities of overdesign follow from the relation $p_o=1-p_u$. From the Tables 3 and 4, it follows that the least squares method usually gives lower probabilities of underdesign than the maximum likelihood method. That's why a least squares method is so popular under engineers. If we define assymetric loss-functions, in which we can model the risk aversion of a designer towards underdesign, we can determine optimal

choices for distribution type and estimation method. For example if we penalize underdesign with a factor 4 more than overdesign we get the following Table 5.

n	Gumbel		Exponential	
	f*	EM*	f*	EM*
10	Exp	ML	Exp	LS
30	G	LS	Exp	LS
100	G	LS	Exp	ML

Table 5: Optimal choices for distribution type f* and estimation method EM*

From this table, we indeed notice a preference for the least squares method, except for large sample sizes from an exponential distribution where a ML-method is prefered and for small sample sizes from a Gumbel distribution which are better modeled by an exponential distribution with a ML-method for risk-averse engineers.

Discussion

In this section, an overview and references were given of parameter estimation techniques that are well known in civil engineering practice. With Monte Carlo simulation studies, these estimation techniques can easily be compared. Table 1 gave an overview of all the simulation work that has been performed for the pairs (f,EM): distribution and estimation method. With the references given in this table it is in principle possible to determine the optimal (w.r.t. minimum bias and RMSE) estimation method given a certain distribution function. However, the optimal choice for a pair (f,EM) can change very quickly if the main assumption of i.i.d. data is violated, or when the criteria for the optimal pair is changed (performance measured in terms of under- and overdesign), as was shown in this paper. Under- and overdesign are important measures for the engineer. Based on simulations from an Exponential distribution and some mathematical proofs an ordening in risk aversion of the different estimation techniques can be made. The Maximum Likelihood, Bayesian point estimation (mean of posterior distributions) and Method of Moments parameter estimation techniques give a relatively higher proportion of underdesign than the Bayesian predictive (integration over the posterior distribution) and Least Squares techniques. Asymetric loss criteria can be used to model the risk aversion of the engineer in mathematical terms. The optimal choice of the probability distribution function and the parameter estimation method can then be determined by minimizing the assymetric loss. In Van Gelder (1996b), this idea has been worked out for more types of loss functions and includes parameter and model uncertainty. In Sec. 5.5 and 5.6 we come back to this issue.

Most of the in the paper given probability models have been implemented in computer programs. Kuczera, (1995) made a program called FLIKE in which the Bayesian analysis of GEV, LN and GPA models are included. Perreault et al. (1994) and Perron et al. (1994) developed the very powerful and user-friendly AJUSTE software.

Discussion

The choice of statistical estimation methods for probability distribution functions is one of the most challenging problems within civil engineering, and one that is filled with many controversies. However, it is a topic with great practical importance and it needs to be dealt with (see also Lambert et al. (1994), Haimes et al. (1994), and Seiler and Alvarez (1996)). Attempts to develop new methods have been extremely abundant (see also Bardsley (1994), Capehart et al. (1998), and Chow and Watt (1990)). This is the reason for the large number of references in this paper. This paper focused on the most important statistical estimation methods that are in circulation under civil engineers. The question which estimation method can best be used is impossible to answer. This depends on too many factors. The definition of what is considered best is one of those factors. However, Table 1 gives an overview of the most important journal papers which investigate the performance of a certain pair (f, EM) w.r.t. minimum bias and RMSE of a quantile. Depending on various conditions, sometimes a classification in the performance of the estimation methods can be made. Furthermore, methods for deriving sampling distributions have been described in this paper and have been applied to the Exponential distribution. The problem of weight factor estimation has been investigated with various methods. Bayes factors appear to perform very well and weight factors based on L-Kurtosis show to be in fairly close agreement with those obtained from the minimum divergence criterion.

It is recommended in this paper (as well as in Mendel and Chick (1993)) to use theoretical considerations as much as possible in the distribution selection. Chick et al. (1995, 1996) proposed a physics-based approach to determine the PDFs of extreme river discharges. In their papers, a new model for predicting the frequency of extreme river levels is proposed which encapsulates physical knowledge about river dynamics, including formulae which describe river discharge. The model accounts for the river dynamics at a given location by modeling both how water gets into the river (via upstream tributaries) and how water leaves (discharge modeled by Chézy's equation). Although the simplified physical model makes several rough approximations (using memoryless properties and Chézy's equation for approximating discharge), insights were gained in the effects of Chezy's equation parameters on the shape of the curves relating the river level and flood return frequency can be shown with Chick's approach. These shapes do not always conform to the curves found for traditional models. In particular, the relation is not necessarily linear on log paper, as with the Exponential model. It was shown that an increase in the power parameter of Chezy's equation led to a non-linear relation on log paper. As the power increased, the slope of the curve relating flood volume and the frequency of extreme floods decreased. This may be true for more complicated systems as well. It was concluded by Chick et al. (1995, 1996) that flood protection designs based on drawing straight lines on log paper would be conservative for extremely rare floods.

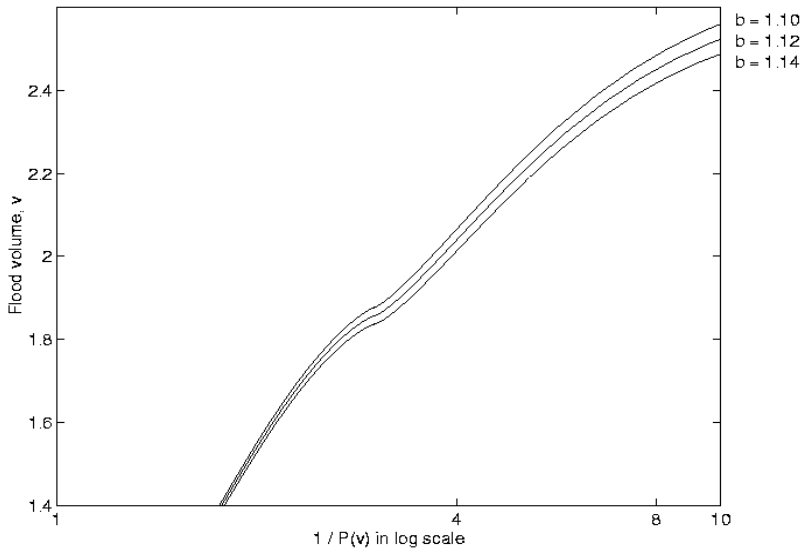


Figure 6: Effect of changes in the power parameter of Chezy's equation on the flood frequency curve (source: Chick et al., 1996)

References

Abernethy, R.B., 1983. Weibull Analysis Handbook, Pratt & Whitney Aircraft, November 1983.

Abramowitz, M., and Stegun, I., 1965. Handbook of Mathematical Functions, Dover Publications Inc., New York.

Anctil, F., Martel, N., and Van Diem Hoang. 1998. Analyse régionale des crues journalières de la province de Québec. Canadian Journal of Civil Engineering, Pages 360-369 Volume 25, Number 2, April 1998 ISSN 1208-6029.

Ashkar, F., and Ouarda, T.B.M.J., 1998. Approximate confidence intervals for quantiles of gamma and generalized gamma distributions. Journal-of-Hydrologic-Engineering. v 3 n 1 Jan 1998, p 43-51.

Bardsley W.E., 1994. Against objective statistical-analysis of hydrological extremes , Journal of hydrology, 162: (3-4) 429-431 NOV 1994.

Barnett, V., 1973. Comparitive Statistical Inference, John Wiley and Sons, Inc.

Bayes, T., 1763. An essay towards solving a probleme in the doctrine of chances, Reprinted from Philos. Trans. Roy. Soc. London. Vol. 53: 370-418.

Beard, L.R., 1994. Anatomy of the best estimate. Journal of Hydraulic Engineering, Vol. 120, No. 6, June, 1994.

Benjamin, J.R., and Cornell, C.A., 1970. Probability, Statistics and Decision for Civil Engineers, McGraw-Hill, Inc.

Ben-Zvi, A., and Azmon, B., 1997. Joint use of L-moment diagram and goodness-of-fit test: A case study of diverse series, Journal of Hydrology. v 198 n 1-4 Nov 1997. p 245-259.

Berger, J.O., 1980. Statistical Decision Theory, Foundations, Concepts, and Methods, Springer-Verlag.

Bernardo, J.M., and Smith, A.F.M., 1994. Bayesian Theory, John Wiley & Sons, Inc., Wiley Series in Probability and Mathematical Statistics.

Bernoulli Family, 1999. The MacTutor History of Mathematics archive, http://www-history.mcs.st-and.ac.uk/history/

Bobée, B., and F. Ashkar, 1991. The Gamma Family and Derived Distributions Applied in Hydrology, 203 pp., Water ResourcesPublication, Littleton, CO.

Bobée, B., Ashkar, F., and Perreault, L., 1993a. Two kinds of moment ratio diagrams and their applications in hydrology, Stochastic Hydrol. Hydraul., 7, 41-65.

Bobée, B., Cavadias, G., Ashkar, F., Bernier, J., and Rasmussen, P.F., 1993b. Towards a systematic approach to comparing distributions used in flood frequency analysis, J. Hydrol., 142, 121-136.

Box, G.E.P., and Tiao, G.C., 1973. Bayesian Inference in Statistical Analysis, Addison-Wesley.

Burcharth, H.F., and Liu, Z., 1994. On the extreme wave height analysis, In Port and Harbour Research Institute, Ministry of Transport, editor, International Conference on Hydro-Technical Engineering for Port and Harbor Construction (Hydro-Port), Yokosuka, Japan, 1994, pages 123-142, Yokosuka: Coastal Development Institute of Technology.

Capehart, B.L., Mahoney, J.F., and Sivazlian, B.D., 1998. Technological risk assessment: The statistical evaluation of low frequency events.Publ by Inderscience Enterprises Ltd, Geneva, Switz p 749-755. Technology Management 1: Proceedings of the First International Conference on Technology Management. Special Publication of the International Journal of Technology Management. Miami, FL, USA

Carlin, B.P., and Louis, T.A., 1996. Bayes and Empirical Bayes Methods for Data Analysis, Chapman & Hall.

Carter, D.J.T., and Challenor, P.G., 1983. Methods of Fitting the Fisher-Tippett Type 1 Extreme Value Distribution, Ocean Engineering, Vol.10, No. 3, pp.191-199.

Castillo, E., and Hadi, A.S., 1997. Fitting the generalized Pareto distribution to data, Journal-of-the-American-Statistical-Association. DEC 1997; 92 (440) : 1609-1620.

Chick, S., Shortle, J., Van Gelder, P., and Mendel, M., 1995. A Physics-Based Approach to Predicting the Frequency of Extreme River Levels, Engineering Probabilistic Design and Maintenance for Flood Protection, pp.89-107. Discussion: pp.109-138.

Chick, S., Shortle, J., Van Gelder, P., and Mendel, M., 1996. A Model for the Frequency of Extreme River Levels Based on River Dynamics, Structural Safety, Vol. 18, Nr. 4, pp.261-276.

Chow, K.C.A., and Watt, W.E., 1990. A knowledge-based expert system for flood frequency analysis, Can. J. Civ. Eng., 17, 597-609.

Coles, S.G., and Powell, E.A., 1996. Bayesian methods in extreme value modelling: A review and new developments, International Statistical Review, 64: (1) 119-136 APR 1996.

Coles, S.G., and Tawn, J.A., 1996. A Bayesian analysis of extreme rainfall data, applied statistics-journal of the royal statistical society series C, 45: (4) 463-478 1996.

Corbyn, J.A., 1988. Statistical analysis of samples from lognormal distribution by Bayesian methods with minerals industry applications.Transactions-of-the-Institution-of-Mining-and-Metallurgy,-Section-A, v 97 Jul 1988, p 118-124.

Corsini, G., Gini, F., Greco, M.V., and Verrazzani, L., 1995. Cramer-Rao bounds and estimation of the parameters of the Gumbel distribution, IEEE-Transactions-on-Aerospace-and-Electronic-Systems. V 31 n 3 Jul 1995, p 1202-1204.

Cramer, H., 1946. Mathematical methods of statistics, Princeton University Press.

Cunnane, C., 1987. Review of statistical methods for flood frequency estimation, in Hydrologic Frequency Modeling, edited by V.P. Singh, pp. 49-95, D. Reidel, Dordrecht.

Cunnane, C., 1988. Methods and merits of regional flood frequency analysis, J. Hydrol., 100, p. 269-290.

Dekkers, A.L.M., and De Haan, L., 1989. On the estimation of the extreme-value index and large quantile estimation. Annals of Statistics 17, 1795-1832.

Demetracopoulos, A.C., 1994. nonlinear-regression applied to hydrologic data, journal of irrigation and drainage engineering-ASCE, 120: (3) 652-659 May-June 1994.

Demuth, S., and Kuells, C., 1997. Probability analysis and regional aspects of droughts in southern Germany. Symp 1: Sustainability of Water Resources under Increasing Uncertainty IAHS Publication (International Association of Hydrological Sciences). n 240.

Duan, J., Selker, J., and Grant, G.E., 1998. Evaluation of probability density functions in precipitation models for the Pacific Northwest, Journal of the American Water Resources Association. v 34 n 3 Jun 1998. p 617-627.

Durrans, S.R., 1992a. Distributions of fractional order statistics in hydrology, Water Resour. Res., 28(6), 1649-1655.

Durrans, S.R., 1992b. Parameter estimation for the Pearson type 3 distribution using order statistics, J. Hydrol., 133, 215-232.

Fill, H.D., and Stedinger, J.R., 1995. Homogeneity tests based upon Gumbel distribution and a critical appraisal of Dalrymple's test, Journal of Hydrology. v 166 p 81-105.

Fill, H.D., and Stedinger, J.R., 1998. Using regional regression within index flood procedures and an empirical Bayesian estimator, Journal-of-Hydrology. SEP 1998; 210 (1-4) : 128-145.

Fisher, R.A., 1934. Statistical methods for research workers, Publisher Edinburgh, Biological monographs and manuals. no. 5.

Gauss, C.F., 1901. Sechs Beweise des Fundamentaltheorems über quadratische Reste (Six proofs of the fundamental theorems on quadratic residuals; in German), Publisher Leipzig : Engelmann, HRSG von E. Netto. 1901, 111 pages, Series Ostwald's Klassiker der exakten Wissenschaften Nr. 122.

Gingras, D., and Adamowski, K., 1994. Performance of L-moments and nonparametric flood frequency analysis. Canadian Journal of Civil Engineering. v 21 n 5 Oct 1994. p 856-862.

Goda, Y., and Kobune, K., 1990. Distribution Function Fitting for Storm Wave Data, Proceedings of the International Conference on Coastal Engineering, Delft, The Netherlands.

Goda, Y., Hawkes, P.J., Mansard, E., Martin, M.J., Mathiesen, M., Peltier, E., Thompson, E.F., and Van Vledder, G., 1993. Intercomparison of extremal wave analysis methods using numerically simulated data, Ocean wave measurement and analysis, pp. 963-977.

Greenwood, J.A., Landwehr, J.M., Matalas, N.C., and Wallis, J.R., 1979. Probability weighted moments: Definition and relation to parameters of several distributions expressable in inverse form. Water Resources Research, 15(5), 1049-1054.

Guo, S.L., and Cunnane, C., 1991. Evaluation of the usefulness of historical and palaeological floods in quantile estimation, J. Hydrol., 129, 245-262.

Haimes, Y.Y., Barry, T., and Lambert, J.H., 1994. When and how can you specify a probability distribution when you don't know much, Risk-Based Decision Making in Water Resources VI. Y.Y. Haimes, D.A. Moser, and E.Z. Stakhiv, Eds., New York: American Society of Civil Engineers, pp. 19-25.

Hald, A., 1952. Statistical Theory with Engineering Applications, John Wiley and Sons, Inc., New York.

Harter, H.L., 1969. Order Statistics and Their Use in Testing and Estimation. Volume 2, Aerospace Research Laboratories, U.S. Air Force, Washington DC, USA.

Hoshi, K., and Leeyavanija, U., 1986. A new approach to parameter estimations of gamma-type distributions, Journal of Hydroscience and Hydraulic Engineering, JSCE, 4 (2), 79-95.

Hosking, J.R.M., 1986. The theory of probability weighted moments, Research Rep. RC 12210, 160 pp., IBM Research Division, Yorktown Heights, NY.

Hosking, J.R.M., 1990. L-moments: Analysis and estimation of distributions using linear combinations of order statistics. J. R. Stat. Soc. Ser. B., vol. 52, 105-124.

Hosking, J.R.M., 1992. Moments or L moments? An example comparing two measures of distribution shape. The American Statistician, 46(3), 186-189.

Hosking, J.R.M., 1997. Fortran Routines for Use with the Method of L-Moments. IBM Research Report, RC20525, Yorktown Heights, NY.

Hosking, J.R.M., and Wallis, J.R., 1987. Parameter and quantile estimation for the generalized Pareto distribution, Technometrics, 29(3), 339-349.

Hosking, J.R.M., and Wallis, J.R., 1997. Regional Frequency Analysis: An Approach based on L-Moments. Cambridge University Press, Cambridge, UK.

Hosking, J.R.M., Wallis, J.R., and Wood, E.F., 1985. Estimation of the Generalized Extreme Value Distribution by the Method of Probability Weighted Moments, Technometrics, August 1985, Vol. 27, No.3, pp.251-261.

Hu, S., 1987a. Determination of confidence intervals for design floods. J-Hydrol. v 96 n 1-4 Dec 15 1987, p 201-213.

Hu, S., 1987b. Problems with outlier test methods in flood frequency analysis. J-Hydrol. v 96 n 1-4 Dec 15 1987, p 375-383.

Johnson, N.L., Kotz, S., and Balakrishnan, N., 1994. Continuous univariate distributions. Vol. 1 Publisher New York : Wiley, 1994 ISBN 0-471-58495-9, 756 pages.

Jowitt, P.W., 1979. The extreme-value type-1 distribution and the principle of maximum entropy, Journal of Hydrology, 42, pp.23-28.

Jumarie, G., 1990. Relative Information: Theories and Applications. Springer-Verlag, Berlin, Germany.

Kendall, M.G., 1961. Daniel Bernoulli on maximum likelihood, Biometrika 48, 1-18.

Koch, S.P., 1991. Bias error in maximum likelihood estimation, J. Hydrol., 122, p. 289-300.

Kuczera, G., 1994. Comprehensive Bayesian at-site flood frequency inference, Proceedings of the Water Down Under 1994 Conference. Part 3 (of 3), IE Aust, Crows Nest, NSW, Aust. p 205-210.

Kullback, S., 1959. Information Theory and Statistics. John Wiley & Sons Inc., New York, USA.

Lall, U., Moon, Y., and Bosworth, K., 1993. Kernel Flood Frequency Estimators: Bandwidth Selection and Kernel Choice. Water Resources Research , April 1993, pp. 1003-1016.

Lambert, J.H., Matalas, N.C., Ling, C.W., Haimes, Y.Y., and Li, D., 1994. Selection of Probability Distributions in Characterizing Risk of Extreme Events, Risk Analysis, Vol.14, No.5.

Landwehr, J.M., Matalas, N.C., and Wallis, J.R., 1979. Probability weighted moments compared with some traditional techniques in estimating Gumbel parameters and Quantiles, Water Resources Research, Vol. 15., No.5, October 1979, pp.1055-1064.

Lechner, J.A., 1991. Estimators for type-II censored (log)normal samples. IEEE-Trans-Reliab. v 40 n 5 Dec 1991, p 547-552.

Lin, B., and Vogel, J.L., 1993. Comparison of L-moments with method of moments Proceedings of the Symposium on Engineering Hydrology Proc Symp Eng Hydrol. Publ by ASCE, New York, NY, USA. 1993. p 443-448.

Lind, N.C., and Hong, H.P., 1991. Entropy estimation of hydrological extremes, Stochastic Hydrol. Hydraul., 5, 77-87.

Lu, L.-H., and Stedinger, J.R., 1992b. Variance of two- and three-parameter GEV/PWM quantile estimators: formulae, confidence intervals, and a comparison, J. Hydrol., 138, 247-267.

Lundgren, H., 1988. Variational criteria in spectrum analysis and image reconstruction, Signal Processing IV, Elsevier, B157pp. 1573-1576.

Lye, L.M., Hapuarachchi, K.P., and Ryan, S., 1993. Bayes estimation of the extreme-value reliability function, IEEE Transactions on Reliability, Vol. 42, No. 4, 1993 December.

Lye, L.M., Sinha, S.K., and Booy, C., 1988. Bayesian analysis of the tau-year events for flood data fitted by a three-prameter lognormal distribution. Civ-Eng-Syst. v 5 n 2 Jun 1988, p 81-86.

Marchette, D.J., 1995. An Investigation of Misspecification of Terms in Mixture Densities, Technical Report, Computational Statistics at George Mason University

Marchette, D.J., Priebe, C.E., Rogers G.W., and Solka, J.L., 1994. The Filtered Kernel Estimator, Center for Computational Statistics, GMU, Tech Rpt No 104, October 1994.

McCuen, R.H., Leahy, R.B., and Johnson, P.A., 1990. Problems with logarithmic transformations in regression , journal of hydraulic engineering-ASCE, 116: (3) 414-428 MAR 1990.

Mendel, M.B., and Chick, S.E., 1993. The Geometry and Calculus of Engineering Probability, UC-Berkeley, Lecture notes, December 1993.

Moharram, S.H., Gosain, A.K., and Kapoor, P.N., 1993. A comparative study for the estimators of the generalized Pareto distribution, J. Hydrol., 150, 169-185.

Mood, A.M., Graybill, F.A., and Boes, D.C., 1974. Introduction to the theory of statistics. 3rd ed. Publisher New York : McGraw-Hill, 1974 Series in probability and statistics.

Moon, Y., and Lall, U., 1994. A Kernel Quantile Function Estimator For Flood Frequency Analysis, in Extreme Values: Floods and Droughts, ed. K. Hipel, Kluwer.

Moon, Y., Lall, U., and Bosworth, K., 1993. A Comparison of Tail Probability Estimators for Flood Frequency Analysis. Journal of Hydrology , V. 151, 1993, pp. 343-363.

Naghavi, B., and Yu, F.-X., 1996. Selection of parameter-estimation method for LP3 distribution. Journal-of-Irrigation-and-Drainage-Engineering. v 122 n 1 Jan-Feb 1996, p 24-30.

Pannullo, J.E., Li, D., and Haimes, Y.Y., 1993. Posterior analysis in assessing risk of extreme events: a conjugate family approach, Proceedings-of-the-IEEE-International-Conference-on-Systems,-Man-and-Cybernetics. v 1 1993, Publ by IEEE, IEEE Service Center, Piscataway, NJ, USA,CH3242-5. p 477-482.

Pearson, C.P., McKerchar, A.I., and Woods, R.A., 1991. Regional flood frequency analysis of Western Australian data using L-moments. International Hydrology and Water Resources Symposium 1991 Part 2. v 2 n 91 pt 22. p 631-632.

Perreault, L., Bobee, B., and Perron, H., 1994. Ajuste II: hydrological frequency analysis software part 1: theoretical aspects, International-Conference-on-Hydraulic-Engineering-Software,-Hydrosoft,-Proceedings-2. 1994, Computational Mechanics Publ, Southampton, Engl. p 243-250

Perron, H., Bobee, B., Perreault, L., and Roy. R., 1994. Ajuste II: hydrological frequency analysis software part 2: software aspects and applications, International-Conference-on-Hydraulic-Engineering-Software,-Hydrosoft,-Proceedings-2. 1994, Computational Mechanics Publ, Southampton, Engl. p 251-258.

Pickands, J., 1975. Statistical inference using extreme order statistics, The Annals of Statistics, 3(1), 119-131.

Pilon, P.J., and Adamowski, K., 1993. Asymptotic variance of flood quantile in log Pearson Type III distribution with historical information, J. Hydrol., 143, 481-503.

Preumont, A., 1988. Bayesian analysis of the extreme value of a time history. Reliab-Eng-Syst-Saf. v 20 n 3 1988, p 165-172.

Rao, A.R., and Hamed, K.H., 1997. Regional frequency analysis of Wabash River flood data by L-moments, Journal of Hydrologic Engineering. v 2 n 4 Oct 1997. p 169-179.

Rasmussen, P.F., Bobée, B., and Bernier, J., 1994. Parameter estimation for the Pearson type 3 distribution using order statistics (Durrans, 1992)---Comment, J. Hydrol., 153, 417-424.

Reiss, R.-D., 1989. Approximate Distributions of Order Statistics, With Applications to Nonparametric Statistics, Spinger Verlag Series in Statistics.

Ribeiro-Correa, J., and Rousselle, J., 1993. A hierarchical and empirical Bayes approach for the regional Pearson type III distribution, Water Resour. Res., 29(2), p. 435-444.

Rosbjerg, D., Madsen, H., and Rasmussen, P.F., 1992. Prediction in partial duration series with generalized Pareto-distributed exceedances, Water Resour. Res., 28(11), p. 3001-3010.

Rosenblatt, M., 1956, Remarks on some Nonparametric Estimates of a Density Function, Ann. Math. Statist., 27, p. 832-835.

Ruprecht, J.K., and Karafilis, D.W., 1994. Regional flood frequency - caution needed Proceedings of the Water Down Under 1994 Conference. Part 3 (of 3). Adelaide, Aust. 19941121-19941125.

Sankarasubramanian, A., and Srinivasan, K., 1999. Investigation and comparison of sampling properties of L-moments and conventional moments. Journal-of-Hydrology. May 10 1999; 218 (1-2) : 13-34.

Scott, D.W., 1992. Multivariate Density Estimation, New York: John Wiley.

Seiler, F.A., and Alvarez, J.L., 1996. On the selection of distributions for stochastic variables, Risk Analysis, Vol. 16, No. 1, pp. 5-18.

Shannon, C.E., 1949. The Mathematical Theory of Communication. The University of Illinois Press, Urbana, IL, USA.

Silverman, B.W., 1986. Density Estimation for Statistics and Data Analysis, New York: Chapman and Hall.

Singh V.P., 1997. The use of entropy in hydrology and water resources, hydrological processes, 11: (6) 587-626 MAY 1997.

Singh V.P., and Guo H., 1997. Parameter estimation for 2-parameter generalized pareto distribution by Pome, Stochastic hydrology and hydraulics, 11: (3) 211-227 JUN 1997.

Singh, V.P., and Fiorentino, M., 1992. A Historical perspective of entropy applications in water resources, In Singh, V.P., and Fiorentino, M. (eds.), Entropy and Energy Dissipation in Water Resources, pp.21-61, Kluwer Academic Publishers.

Singh, V.P., and Singh, K., 1985. Derivation of the Gamma distribution by using the principle of maximum entropy (POME). Water Resources Bulletin, Vol. 21, No. 6, pp.941-952.

Singh, V.P., and Singh, K., 1987. Parameter estimation for TPLN distribution for flood frequency analysis. Water Resources Bulletin, vol. 23, No. 6, pp.1185-1992.

Singh, V.P., Cruise, J.F., and Ma, M., 1990. A Comparative-evaluation of the estimators of the 3-parameter lognormal-distribution by Monte-Carlo Simulation, Computational statistics & data analysis, 10: (1) 71-85 AUG 1990.

Smith, R.L., and Naylor, J.C., 1987. A comparison of Maximum Likelihood and Bayesian Estimators for the Three-parameter Weibull Distribution, Appl. Statist. 36, No. 3, pp. 358-369.

Smith, R.L., and Weissman, I., 1987. Large deviations of tail estimators based on the Pareto approximation. J. Appl. Prob. Vol. 24, pp. 619-630.

Stacy, E.W., and Mihram, G.A., 1965. Parameter estimation for a generalized gamma distribution, Technometrics, Vol. 7, No. 3, pp. 349-358, August 1965.

Stedinger, J.R., 1980. Fitting log normal distributions to hydrologic data, Water Resources Research, 16 (3), 481-490.

Sukhatme, P.V., 1937. Tests of significance for samples of the Chi-squared population with two degrees of freedom, Ann. Eugen., Vol. 8, pp. 52-56.

Takara, K., and Takasao, T., 1990. Comparison for parameter estimation methods for hydrologic frequency analysis models, Proc. Hydraulic Engineering, JSCE, 34, 7-12 (in Japanese).

Takara, K., Takasao, T., and Shimizu, A., 1989. Comparison of parameter estimation methods for extreme value distributions. Annuals, DPRI, Kyoto University, 32 B-2, 455-469 (in Japanese).

Takara, K.T., and Stedinger, J.R., 1994. Recent Japanese contributions to frequency analysis and quantile lower bound estimators, In: K.W. Hipel (ed.), Stochastic and Statistical methods in hydrology and environmental engineering, Vol. 1, p. 217-234, Kluwer Academic Publishers.

Takeuchi, K., and Tsuchiya, K., 1988. On relative accuracy of PWM estimates of normal and 3-parameter lognormal distributions, Proc. of JSCE, Vol.393/ll-9, May,pp. 103-112.

TAW: Technical Advisory Committee on Water Defences, 1990. Probabilistic Design of Flood Defences. Report 141, Centre for Civil Engineering Research and Codes, The Netherlands.

Van Gelder, P.H.A.J.M., 1999b. Risks and safety of flood protection structures in the Netherlands, Proceedings of the Participation of Young Scientists in the Forum Engelberg 1999 on Risk and Safety of Technical Systems - in View of Profound Changes, pp.55-60, p.93.

Van Gelder, P.H.A.J.M., and Neykov, N.M., 1998. Regional frequency analysis of extreme water levels along the Dutch coast using L-moments: A preliminary study, In: Stochastic models of hydrological processes and their applications to problems of environmental preservation, pp.14-20.

Van Gelder, P.H.A.J.M., and Vrijling, J.K., 1997a. A comparative study of different parameter estimation methods for statistical distribution functions in civil engineering applications, Structural Safety and Reliability, Vol. 1, pp.665-668.

Van Gelder, P.H.A.J.M., Pandey, M.D., and Vrijling, J.K., 1999b. The use of L-Kurtosis in the Estimation of Extreme Floods, 9th Annual Conference Risk Analysis: Facing the New Millennium, Rotterdam, The Netherlands, October 10 - 13, 1999.

Van Gelder, P.H.A.J.M., 2000. Statistical methods for the risk-based design of civil structures, Publisher TU Delft, Descr. 248 p, Series Communications on Hydraulic and Geotechnical Engineering ISSN:0169-6548 00-1.

Van Noortwijk, J.M., 1999. Quantiles of Generalised Gamma Distributions from a Bayesian Point of View, HKV-report.

Vrijling, J.K., and Van Gelder, P.H.A.J.M., 1995. Probabilistic design of berm breakwaters, in: Engineering Probabilistic Design and Maintenance for Flood Protection, pp. 181-198, Discussion: pp. 199-213.

Vrijling, J.K., and Van Gelder, P.H.A.J.M., 1997a. Societal risk and the concept of risk aversion, Advances in Safety and Reliability, Vol. 1, pp. 45-52.

Wang, Q.J., 1998. Approximate goodness-of-fit tests of fitted generalized extreme value distributions using LH moments, Water Resources Research, 34: (12) 3497-3502 DEC 1998.

WMO, 1989. Statistical distributions for flood frequency analysis, Operational Hydrology Report No. 33, World Meteorological Organization---No. 718, Geneva, Switzerland.

Wu B., Hou Y., and Ding J., 1991. Method of lower-bound to estimate the parameters of a Pearson type III distribution. Hydrol-Sci-J. v 36 n 3 Jun 1991, p 271-280.

Yamaguchi, M., 1997. Intercomparison of parameter estimation methods in extremal wave analysis. In B.L. Edge, editor, 25th International Conference on Coastal Engineering, Orlando, Florida, U.S.A., 1996, pages 900-913, New York: American Society of Civil Engineers (ASCE).