Stochastic Hydrology and Hydraulics 11: 51-63

Chowdhury and Stedinger's approximate confidence intervals for design floods for a single site

R. Whitley
Department of Mathematics, University of California, Irvine, CA 92697-3875, USA

T. V. Hromadka II
Failure Analysis Associates, 4590 MacArthur Blvd., Suite 400, Newport Beach, CA 92660-2027, USA

Abstract: A basic problem in hydrology is computing confidence levels for the value of the T-year flood when it is obtained from a Log Pearson III distribution in terms of estimated mean, estimated standard deviation, and estimated skew. In an important paper Chowdhury and Stedinger [1991] suggest a possible formula for approximate confidence levels, involving two functions previously used by Stedinger [1983] and a third function, λ, for which asymptotic estimates are given. This formula is tested [Chowdhury and Stedinger, 1991] by means of simulations, but these simulations assume a distribution for the sample skew which is not, for a single site, the distribution which the sample skew is forced to have by the basic hypothesis which underlies all of the analysis, namely that the maximum discharges have a Log Pearson III distribution. Here we test these approximate formulas for the case of data from a single site by means of simulations in which the sample skew has the distribution which arises when sampling from a Log Pearson III distribution. The formulas are found to be accurate for zero skew but increasingly inaccurate for larger common values of skew. Work in progress indicates that a better choice of λ can improve the accuracy of the formula.

1 Introduction

A basic problem in hydrology is the estimation, for design purposes, of the 100 year flood, or more generally the T-year flood. A major source of uncertainty in this estimation is the choice of underlying distribution for maximum discharge [Bobee, et. al, 1993, Cohon, et. al., 1988, and World Meteorological Organization, 1989]. Once this distribution is chosen, an important further source of uncertainty is that caused by the estimation of the parameters of the distribution. To give a more complete picture of the level of risk involved in a chosen level of flood protection it is necessary to quantify this uncertainty by means of confidence intervals for the T-year flood estimates. Water Resource Council Bulletin 17B [Advisory Council on Water Data, 1982] recommends the use of a log Pearson III distribution, fit to yearly maximum discharge data, for the prediction of T-year events. Other methods have been proposed, see for example the discussion in [Bobee, et. al., 1993, Cohon, et. al., 1988, and World Meteorological Organization, 1989], and an important area of research is in obtaining more accurate methods for estimating extreme floods.

However, in practice, because of the authority of the U.S. Water Resource Council, the procedure given in Bulletin 17B is used extensively.

The log Pearson III distribution, which will be discussed in more detail later, contains three parameters which are estimated by the procedure of Bulletins 17A and 17B in terms of the estimated mean, estimated standard deviation, and estimated skew of the distribution of logarithms of the yearly maximal discharge data. The estimators used are, except for the scaling factor appearing in front of the bracket in the formula (3) for γ, the usual ones.

Accurate confidence intervals can be given when the mean and standard deviation are estimated but the skew is known to be zero. In this case the log Pearson III distribution is actually a lognormal distribution, and confidence intervals can be obtained from the non-central t-distribution [Advisory Committee on Water Data, 1982, Resnikof and Lieberman, 1957, Stedinger, 1980].

The case of non-zero skew is more complicated than the case of zero skew [Bobee and Robitaille, 1977, Hu, 1987, Kite, 1975, Phien and Hsu, 1985]. Stedinger [1983] showed that the method of computing confidence intervals suggested in [Advisory Committee on Water Data, 1982] is not satisfactory; also see the general discussion in [Chowdhury and Stedinger, 1991].

If the skew is known (but not zero), and the mean and standard deviation are estimated, we showed [Whitley and Hromadka, 1986a] how to obtain confidence levels for the T-year flood by means of simulations. Stedinger [1983] gave an approximate expression for confidence intervals for the quantiles of the log Pearson III distribution using an asymptotic variance formula [Bobee, 1973, Kite, 1976]. The accuracy of this approximate formula for skew values γ in the range $-0.75 \leq \gamma \leq 0.75$ was discussed in [Whitley and Hromadka, 1986b, 1987].

A major attack on the complete problem of computing confidence intervals for the T-year flood when the mean, standard deviation and the skew are not known but are estimated was made by Chowdhury and Stedinger in [1991], the basic idea of which is to modify Stedinger's approximate formula for the case of known skew by replacing the variance ratio in that formula by first order asymptotic expansions which take into account some of the variation in the estimation of the skew. The purpose of this paper is to discuss the accuracy of these Chowdhury-Stedinger formulas for the basic case of data from a single site.

2 Basic equations

When yearly maximum discharge is fit by a log Pearson III distribution, for the prediction of T-year events, the logarithm of the yearly peak discharge is assumed to have a density function of the form:

$$f(x) = (1/|a|\Gamma(b))[(x-c)/a]^{b-1}\exp[-(x-c)/a] \tag{1}$$

where, in the case of positive a, the density is given by the expression (1) for x>c and is zero for x<c, while in the case of negative a the density is given by (1) for x<c and is zero for x>c. Computing the mean μ, standard deviation σ, and skew γ from equation (1) shows that

$$\begin{aligned} \sigma^2 &= a^2 b \\ \gamma^2 &= 4/b \\ \mu &= c + ab \end{aligned} \tag{2}$$

where a has the same sign as γ.

In the case of zero skew, which is the limiting case where the positive parameter b tends to infinity, the density in equation (1) tends to the density for the normal distribution.

It is further recommended in Bulletin 17B that the parameters a,b, and c be estimated by using equations (2) and the usual moment estimators for μ, σ, and γ, with the moment estimator for γ scaled to make it less biased [Bobee and Robitaille, 1975, Lettenmaier and Burges, 1980]. Specifically

$$\hat{\mu} = \sum_{i=1}^{m} x_i/m$$

$$\hat{\sigma} = \{m/m-1\}^{1/2}\left[\sum_{i=1}^{m} x_i^2/m - \hat{\mu}^2\right]^{1/2}$$

$$\hat{\gamma} = \{(m(m-1))^{1/2}/(m-2)\}\left[\sum_{i=1}^{m} x_i^3/m - 3\hat{\mu}\hat{\sigma} - \hat{\mu}^3\right]/\sigma^3 . \tag{3}$$

Two *observations* follow directly from these formulas. The first is that if the maximum yearly discharge is Q, and X=log(Q), then (X-c)/a has a gamma distribution with parameter b, i.e. with density

$$g(x) = (1/\Gamma(b))x^{b-1}e^{-x} \tag{4}$$

for x greater then zero, and $g(x) = 0$ for x less than zero. This shows that the parameters a and c can be scaled out of the problem, but the parameter b, or equivalently the skew, enters into the problem in a complex way as the parameter of a gamma distribution, the estimation of which contains many difficulties [Bowman and Shenton, 1988].

The second observation concerns the case of negative skew. *To introduce notation* which will be needed later, given a value T>1 of the T-year flood, e.g. T=100, set p=1-1/T. The T-year flood value for X=logQ is the number t_p having the property that

$$\Pr(x \leq t_p) = p \tag{5}$$

i.e. the value of the log of the maximum discharge will not, with probability p, *exceed the value* t_p. *In the case of positive skew*, $\Pr((X-c)/a \leq (t_p - c)/a) = p$ and therefore $(t_p - c)/a$ is the p-th percentile for the gamma distribution (4). If the skew is negative,

$$\Pr(x \leq t_p) = \Pr[(X-c)/a \geq (t_p - c)/a] \tag{6}$$

the reversal of the inequality occurring because a is negative.

Let a′, b′, and c′ be the parameters for -X=X′, which also has a Pearson III distribution. Equations (2) show that a′=-a, b′=b, and c′=-c. Therefore (6) can be rewritten as $\Pr[(X'-c')/a' \geq (-t_p - c')/a'] = \Pr(X' \geq -t_p) = 1 - \Pr(X' \leq -t_p)$, and then

$$\Pr(X' \leq -t_p) = 1 - p \tag{7}$$

or, changing p to 1-p,

$$\Pr(X' \leq -t_{1-p}) \;=\; p\,. \tag{8}$$

Letting t'_p be the p-th percentile for X'=-X,

$$t'_p \;=\; -t_{1-p}\,. \tag{9}$$

The relation (9) is useful in numerical calculations: problems involving negative skew can be solved as a related problem for positive skew.

3 Chowdhury and Stedinger's approximate formula

The approximate formula given in [Chowdhury and Stedinger, 1991] for a confidence interval for data taken at a single site can be written as

$$c\,(m,q,T,\hat{\mu},\hat{\sigma},\hat{\gamma}) \;=\; \hat{\mu}+\hat{\sigma}\,[K(\hat{\gamma},T)+\lambda(\hat{\gamma},T)B(m,q,T)]\;. \tag{10}$$

The parameters are m, the number of observations of maximum discharge at the site, the estimates of equations (3) based on these observations, and three functions: $K(\hat{\gamma},T)$ which depends, as indicated, on the values of $\hat{\gamma}$ and the value T of the T-year flood; B(m,q,T), which depends on m,T, and the confidence level q as described below; and $\lambda(\hat{\gamma},T)$, the variance scale factor. Letting X=logQ, the purpose of this formula is to provide a q-th percentile confidence value for the random variable X with true T-year flood value t_p, i.e. in the sense of repeated sampling

$$\Pr\,[t_p \leq c(m,q,T,\hat{\mu},\hat{\sigma},\hat{\gamma})] \;=\; q\,. \tag{11}$$

For example, with q=.85, repeated use of this formula at a series of sites, each of which has the same log Pearson III distribution and which is sampled by m data points, is supposed to give an "85% safe estimate" for the (logarithm of the) T-year flood value, i.e. one which is at least as large as the true T-year flood value t_p, 85% of the time.

4 General simulation considerations

To understand how these formulas are tested by simulation it is necessary to discuss some specific details. Results are presented for the T=100 year flood as being that of most general interest. A range of skew values are chosen for testing: γ=-1(1/2)1; a range of values of numbers m of data points at the site are chosen: m=10(10)100; and a number of confidence levels q = 5(5)95 are chosen. These parameter choices cover a wide range of practical values.

The simulation proceeds by choosing one of the skews γ from the set (-1.0, -0.5, 0.0, 0.5, 1.0). It is easy to see that the mean and standard deviation can be scaled out of the formulas so that it is completely general to suppose also that μ=0 and σ=1; of course that does not mean that $\hat{\mu}=0$ or that $\hat{\sigma}=1$. Then 100 points are drawn from a Pearson III distribution with skew γ, mean 0, and standard deviation 1, and these points are used to compute $\hat{\mu}$, $\hat{\sigma}$, and $\hat{\gamma}$ from (3) for the ten values m=10(10)100. For each of the confidence levels q=5(5)95 the value of $c(m,q,T,\hat{\mu},\hat{\sigma},\hat{\gamma})$ is computed from equation (10) and it is compared to the true T=100 year flood value t_p, p=.99, which is known exactly because the values of μ, σ, and γ are specified in the simulation. This is done a repeated number of times (30,000) and each time it is noted whether the values of $c(m,q,T,\hat{\mu},\hat{\sigma},\hat{\gamma})$ are indeed greater than t_p. The resulting relative frequency counts of these occurrences show for that value of skew and T=100 what confidence level is actually given by the formula (which is supposed to give the q-th confidence level). Looking at the set of tables for the range of skews tested will then indicate how accurate the use of formula (10) would be for a site whose log maximum discharges have the hypothesized Pearson III distribution and whose (unknown) skew lies in the interval of skews tested by the simulations.

The important point to note in this general outline of the simulations is that there are no hypotheses other than that logQ has a Pearson III distribution. In the simulations done by Chowdhury and Stedinger [1991], generating models for γ are chosen, the first is rejected as giving unreasonable results, while the second, which is used, is described as "not intended to be the true model of the distribution of γ in a region" [Chowdhury and Stedinger, 1991, pg 820]. A major problem (among others) with this approach is that in the case of sampling from a given site, the estimator $\hat{\gamma}$ has a specific distribution from the fact that it is sampled from a Pearson III distribution. If another model of the distribution of sample skews is assumed, that distribution may not be similar enough to the real distribution of skews from a Pearson III to make sampling from that model relevant to the Pearson III problem.

In Chowdhury and Stedinger [1991] three situations are considered: (A) Skew given by sampling from a site, which we discuss here; (B) Skew given by a regional skew G; and (C) Skew given by a combination of site skew and G. The formulas of (B) and (C) are then derived in [Chowdhury and Stedinger, 1991] using assumptions about the regional skew G and the distribution of the values of skew γ about the value of G [Advisory Council on Water Data, 1982, Tasker, 1978, Tasker and Stedinger, 1986]. These assumptions, additional to the basic assumption that logQ has a yearly maximum which has a Pearson III distribution, are also used in generating the random variables used in the simulations of Chowdhury and Stedinger [1991], making the results in the cases (B) and (C) critically dependent on the validity of these extra assumptions. We do not discuss cases (B) and (C) here, partially because of questions concerning the use and distribution of regional skews, see e.g. [Hardison, 1974, McCuen, 1979, Tasker, 1978, Tasker and Stedinger, 1986], but mainly because case (A) is basic to any computation of T-year floods.

5 Computing B, K, and λ

The functions B and K of equations (10) are those used in [Stedinger 1983], whose computation was discussed in [Whitley and Hromadka,1986b]. The function K was computed in [Whitley and Hromadka, 1986b] by use of the Wilson-Hilferty approximation [Wilson and Hilferty, 1931], whereas here it is done more accurately by inverting the incomplete gamma function [Press et. al., 1989], except for small $\gamma \leq 0.2$, equivalently large $b \geq 100$, for which the computation of the incomplete gamma function becomes unmanageable and some approximation must be used, and in this case we use the Wilson-Hilferty approximation. Concerning the use of the Wilson-Hilferty transformation, note that it was derived as an approximation to the Chi-Square distribution which, being the sum of squares of normal distributions, is positive; when, by means of a change of variable, it is applied to a Pearson III distribution it is only valid for non-negative skew. While the formula is not accurate when small negative values of skew are directly substituted into the formula, as noted by Chowdhury and Stedinger [1991], it is accurate to use equation (9) and then the Wilson-Hilferty approximate formula for positive skew.

The function B can be computed from the non-central t-distribution, as described by Whitley and Hromadka [1986b].

The computation of λ, in the forms given in equations (17), (18), or (19') of Chowdhury and Stedinger [1991] applying to cases (A), (B), and (C) discussed above, can be easily computed from a knowledge of K and $\frac{\partial K}{\partial \gamma}$. Although it is possible to derive an exact formula for $\frac{\partial K}{\partial \gamma}$ from its definition in terms of a percentile of an incomplete gamma function, that formula opens a Pandora's Box of numerical analysis problems. It is much easier to use Richardson extrapolation [Burden and Faires, 1993, pgs 168-173] and thereby accurately compute $\frac{\partial K}{\partial \gamma}$ from the values of K.

The graphs given below of B,K, and λ will allow the reader to calculate the confidence intervals of equation (10) for T=100 and a representative range of values of m, q, $\hat{\gamma}$ with enough accuracy to get a feel for the magnitude of the numbers so obtained. The curves in Graph 1 for m=10,25,50, and 75 can be distinguished because the slope of the curves decreases as m increases, i.e. extra data points lower the value of a given confidence level.

Graph 1: B(m,q,T), T=100, m = 10,25,50,75

Graph 2: K($\hat{\gamma}$,T), T=-100, $-2 \leq \hat{\gamma} \leq 2$

Graph 3: λ($\hat{\gamma}$,T), T=100, $-2 \leq \hat{\gamma} \leq 2$

Another remark concerning negative skew: arguments similar to those used in deriving equation (9), show that if equation (10) gives the required confidence levels, then no matter how the functions K, λ, and B are defined, we must have

$$K(\gamma, p) = -K(-\gamma, 1-p), \lambda(\gamma, p) = \lambda(-\gamma, 1-p), B(m, p, q) = -B(m, 1-p, 1-q) \,. \tag{12}$$

To simplify the expressions (12), K,B, and λ have been regarded as a functions of the variable p=1-1/T. If they are regarded as functions of T, as has been done above, the first equation would take the form $K(\gamma,T) = -K(-\gamma,(1-1/T)^{-1})$, etc. These theoretically derived relationships (12), are exactly satisfied by the formulas given in Chowdhury and Stedinger [1991], even though those formulas are obtained by means of approximate asymptotic expansions.

6 Simulations

The results of the simulations are given below in Tables 1-5. The values given are the errors made in using the formulas (10) for the T=100 year flood, the value of skew specified for each table, and the tabulated values of m and q. That is, if $\hat{q}$ is the observed frequency that (11) holds in the simulation, with m sample points, then the number tabulated in the (m,q) matrix of the table for that skew value is

$$100(\hat{q} - q) \,. \tag{13}$$

For each table, the values were obtained by simulating 30,000 sites. For each site 100 gamma distributed random variables were generated, using the techniques described in [Devroye, 1986]. The first 10 of these values were used to compute $\hat{\mu}$, $\hat{\sigma}$, and $\hat{\gamma}$ to be substituted into (10) for the case m=10, the first 20 values were used for the case

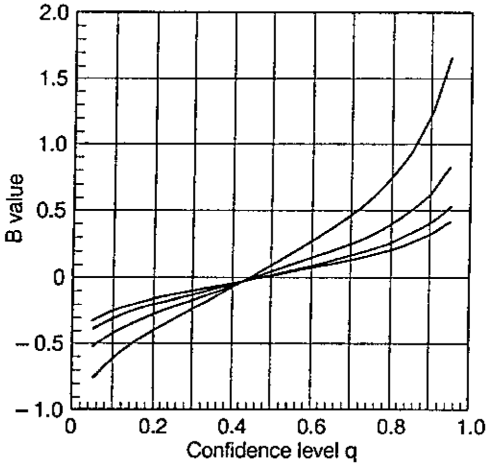

Figure 1. B(m, q, T) for T = 100, m = 10, 25, 50, 75

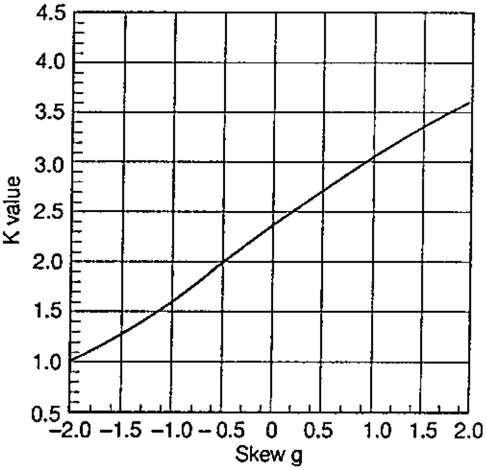

Figure 2. K(g, T) for T = 100

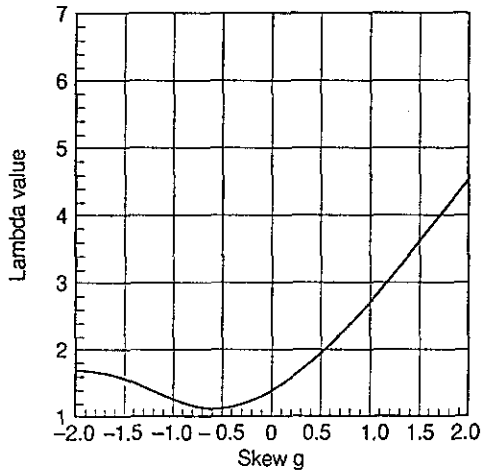

Figure 3. Lambda(g, T) for T = 100

m=20, etc, so that the values used for some value of m less than 100 is also used for all higher values of m.

Tables of the values of B, K and λ in intervals of 0.01 skew are used to compute the values of K and λ by tabular linear interpolation, not by computing $K(\hat{\gamma})$ and $\lambda(\hat{\gamma})$ for each simulated value of $\hat{\gamma}$.

The skew values in the simulations were truncated to lie in the interval [-2,2] (They are truncated in [Chowdhury and Stedinger, 1991] to lie in [-1.5,1.5]), reflecting the fact that in practice a very large estimated skew would probably be truncated. The exact nature of the truncation is not significant, there being very little difference between results for skew truncated to [-2,2] and [-4,4], even for $\gamma = \pm 1$. (This is not because there are no estimated skew values outside of these intervals for most m under consideration; Kirby's [1975] bounds for the absolute value of the unscaled skew estimates are 4.12 for m=20, 6.86 for m=50, and 9.8 for m=100.)

The confidence levels that are in common use are 50% and higher, which are depicted in Tables 1-5 by the last 10 rows i.e. q=50(5)95. With this in mind, Table 1 shows that the accuracy of (10) is acceptable for a skew of zero. For a skew of 1/2, Table 2 shows a maximum error in the last 10 rows occurring for m=10 and q=50. Using (10) in this case will provide a confidence level of 43.7%, not 50%. Similarly, the entry for m=10, q=50 in Table 3, shows that in this case of skew = -1/2, the use of (10) gives a confidence level of 59.9%, not 50%. The errors for skew ± 1 are shown in Tables 4-5 and are considerably larger.

Table 1. T = 100, skew = 0.0, 30,000 sites simulated columns: number of data points m = 10 (10) 100 rows: confidence levels q = 5 (5) 95.

Error in Confidence Levels, Skew = 0.0

	m=10	m=20	m=30	m=40	m=50	m=60	m=70	m=80	m=90	m=100
q=5	-4.7	-4.3	-4.0	-3.6	-3.3	-3.0	-2.9	-2.7	-2.4	-2.4
q=10	-8.1	-6.5	-5.7	-4.8	-4.4	-3.9	-3.6	-3.4	-3.1	-3.0
q=15	-9.4	-7.1	-5.7	-5.0	-4.2	-3.9	-3.4	-3.2	-3.0	-2.7
q=20	-9.1	-6.3	-5.4	-4.2	-3.7	-3.5	-3.0	-2.6	-2.5	-2.5
q=25	-7.7	-5.2	-4.5	-3.8	-3.1	-2.8	-2.5	-2.2	-1.7	-1.8
q=30	-5.9	-4.0	-3.5	-2.8	-2.3	-2.2	-1.9	-1.6	-1.3	-1.2
q=35	-4.3	-3.1	-2.6	-2.0	-1.7	-1.7	-1.4	-1.1	-0.8	-0.8
q=40	-3.0	-2.3	-1.6	-1.3	-1.1	-1.0	-0.7	-0.6	-0.2	-0.7
q=45	-1.7	-1.7	-1.0	-0.7	-0.5	-0.6	-0.4	-0.1	+0.0	-0.3
q=50	-0.5	-1.0	-0.4	-0.2	-0.1	-0.5	-0.1	-0.0	+0.1	-0.3
q=55	+0.6	-0.4	+0.1	+0.1	+0.1	-0.2	+0.2	-0.1	+0.1	-0.3
q=60	+1.2	-0.1	+0.1	+0.3	+0.2	-0.1	+0.0	+0.1	+0.1	+0.0
q=65	+1.7	+0.0	+0.2	+0.2	+0.1	+0.1	+0.1	-0.1	-0.1	-0.0
q=70	+2.2	+0.1	+0.2	+0.2	+0.2	-0.1	-0.1	-0.1	-0.4	-0.3
q=75	+2.5	+0.3	-0.0	+0.0	+0.1	-0.2	-0.4	-0.4	-0.7	-0.7
q=80	+3.0	+0.3	-0.1	-0.1	-0.1	-0.4	-0.8	-0.4	-0.6	-0.9
q=85	+2.9	+0.5	-0.2	-0.3	-0.4	-0.6	-0.6	-0.8	-1.1	-1.2
q=90	+2.6	+0.7	-0.1	-0.4	-0.5	-0.7	-0.9	-1.0	-1.2	-1.3
q=95	+1.9	+0.7	-0.0	-0.4	-0.6	-0.8	-1.0	-0.8	-0.9	-1.0

Table 2. T = 100, skew = 0.5, 30,000 sites simulated columns: number of data points m = 10 (10) 100 rows: confidence levels q = 5 (5) 95.

Error in Confidence Levels, Skew = 0.5

	m=10	m=20	m=30	m=40	m=50	m=60	m=70	m=80	m=90	m=100
q=5	-4.8	-4.7	-4.6	-4.3	-4.2	-4.0	-3.9	-3.7	-3.5	-3.4
q=10	-8.7	-8.0	-7.2	-6.7	-6.2	-5.7	-5.4	-4.9	-4.6	-4.5
q=15	-10.9	-9.5	-8.3	-7.3	-6.8	-6.3	-5.8	-5.3	-5.1	-4.8
q=20	-11.3	-9.3	-8.1	-7.2	-6.5	-6.2	-5.8	-5.3	-5.0	-4.5
q=25	-10.7	-8.8	-7.5	-6.5	-6.0	-5.8	-5.5	-4.9	-4.6	-4.3
q=30	-9.7	-8.1	-6.9	-5.9	-5.5	-5.3	-5.0	-4.7	-4.2	-3.8
q=35	-8.6	-7.3	-6.4	-5.4	-5.1	-4.9	-4.6	-4.3	-4.1	-3.6
q=40	-7.6	-6.5	-5.9	-5.2	-4.7	-4.6	-4.3	-3.9	-3.9	-3.5
q=45	-7.0	-6.0	-5.4	-4.8	-4.4	-4.2	-4.1	-3.8	-3.7	-3.4
q=50	-6.3	-5.6	-5.0	-4.5	-4.4	-4.1	-4.1	-3.8	-3.5	-3.3
q=55	-6.0	-5.3	-4.7	-4.5	-4.5	-4.1	-3.9	-3.8	-3.5	-3.3
q=60	-5.7	-5.3	-4.6	-4.5	-4.4	-3.9	-3.9	-3.7	-3.6	-3.3
q=65	-5.9	-5.1	-4.6	-4.5	-4.4	-3.9	-4.0	-3.9	-3.7	-3.5
q=70	-5.9	-5.2	-4.7	-4.6	-4.5	-4.0	-4.0	-4.0	-3.9	-3.5
q=75	-5.6	-5.3	-4.8	-4.7	-4.8	-4.3	-3.8	-3.9	-4.1	-3.6
q=80	-5.8	-5.6	-5.0	-4.8	-4.8	-4.4	-4.0	-3.8	-3.8	-3.6
q=85	-5.3	-5.5	-5.0	-4.8	-4.8	-4.3	-4.0	-4.0	-3.6	-3.5
q=90	-4.6	-5.3	-4.9	-4.5	-4.5	-4.2	-3.9	-3.7	-3.3	-3.3
q=95	-2.7	-4.2	-4.3	-3.9	-3.5	-3.5	-3.3	-3.1	-2.9	-2.6

Table 3. T = 100, skew = -0.5, 30,000 sites simulated columns: number of data points m = 10 (10) 100 rows: confidence levels l = 5 (5) 95.

Error in Confidence Levels, Skew = -0.5

	m=10	m=20	m=30	m=40	m=50	m=60	m=70	m=80	m=90	m=100
q=5	-4.5	-4.0	-3.5	-3.1	-2.7	-2.4	-2.0	-2.0	-1.9	-1.8
q=10	-6.9	-5.0	-4.0	-3.4	-2.8	-2.4	-2.1	-1.7	-1.5	-1.4
q=15	-6.6	-4.3	-3.1	-2.4	-1.8	-1.4	-1.4	-1.1	-0.8	-0.6
q=20	-4.6	-2.3	-1.5	-1.0	-0.6	-0.3	-0.2	+0.1	+0.2	+0.4
q=25	-1.7	-0.2	+0.1	+0.5	+0.7	+0.8	+0.9	+0.9	+1.3	+1.2
q=30	+1.1	+1.6	+1.8	+2.1	+2.0	+2.0	+2.2	+2.2	+2.2	+2.2
q=35	+3.5	+3.6	+3.4	+3.4	+3.2	+3.1	+3.1	+2.9	+2.9	+3.0
q=40	+5.7	+5.1	+4.8	+4.5	+4.4	+4.0	+3.9	+3.7	+3.7	+3.7
q=45	+8.1	+6.4	+5.9	+5.4	+5.3	+4.7	+4.6	+4.3	+4.3	+4.5
q=50	+9.9	+7.8	+7.0	+6.1	+6.0	+5.6	+5.1	+4.9	+4.9	+4.9
q=55	+11.8	+9.0	+7.7	+6.7	+6.5	+6.0	+5.6	+5.2	+5.1	+5.1
q=60	+13.1	+9.9	+8.2	+7.4	+6.9	+6.4	+6.2	+5.6	+5.5	+5.1
q=65	+13.9	+10.7	+8.8	+8.0	+7.0	+6.7	+6.5	+5.8	+5.6	+5.2
q=70	+14.0	+11.4	+9.4	+8.2	+7.1	+6.8	+6.5	+6.0	+5.7	+5.4
q=75	+13.5	+11.6	+9.7	+8.4	+7.4	+6.9	+6.3	+6.0	+5.7	+5.4
q=80	+12.3	+10.8	+9.5	+8.4	+7.4	+6.9	+6.1	+5.7	+5.6	+5.2
q=85	+10.4	+9.5	+8.7	+7.9	+7.0	+6.5	+5.8	+5.3	+5.1	+4.8
q=90	+7.7	+7.3	+7.0	+6.4	+5.7	+5.3	+5.1	+4.6	+4.3	+4.0
q=95	+4.2	+4.2	+4.1	+4.0	+3.8	+3.5	+3.3	+3.3	+2.9	+2.8

Table 4. T = 100, skew = 1.0, 30,000 sites simulated columns: number of data points m = 10 (10) 100 rows: confidence levels q = 5 (5) 95.

Error in Confidence Levels, Skew = 1.0

	m=10	m=20	m=30	m=40	m=50	m=60	m=70	m=80	m=90	m=100
q=5	-4.9	-4.9	-4.9	-4.7	-4.7	-4.6	-4.6	-4.5	-4.4	-4.3
q=10	-9.2	-8.8	-8.4	-8.0	-7.8	-7.6	-7.4	-7.2	-6.9	-6.7
q=15	-12.2	-11.4	-10.5	-9.8	-9.3	-8.9	-8.7	-8.2	-8.0	-7.6
q=20	-13.6	-12.2	-11.5	-10.7	-9.9	-9.4	-8.9	-8.5	-8.1	-7.7
q=25	-13.8	-12.5	-11.5	-10.7	-10.1	-9.3	-8.8	-8.4	-7.9	-7.6
q=30	-13.3	-12.1	-11.0	-10.3	-9.5	-9.0	-8.6	-8.1	-7.6	-7.4
q=35	-13.0	-11.5	-10.4	-9.8	-9.2	-8.4	-8.3	-7.4	-7.3	-7.0
q=40	-12.3	-10.8	-9.7	-9.1	-8.7	-7.9	-7.8	-7.1	-6.8	-6.6
q=45	-11.7	-10.1	-9.3	-8.7	-8.0	-7.4	-7.2	-6.8	-6.4	-6.3
q=50	-11.4	-9.5	-8.9	-8.3	-7.7	-7.2	-6.7	-6.5	-6.4	-6.0
q=55	-10.9	-9.5	-8.2	-8.0	-7.7	-6.9	-6.5	-6.1	-6.2	-5.9
q=60	-10.9	-9.3	-8.0	-7.6	-7.4	-6.8	-6.4	-5.9	-5.9	-5.9
q=65	-10.7	-9.1	-8.1	-7.6	-7.3	-6.7	-6.1	-5.8	-5.9	-5.7
q=70	-10.7	-9.2	-7.9	-7.3	-7.1	-6.4	-5.8	-5.6	-5.6	-5.7
q=75	-10.9	-9.2	-7.9	-7.2	-6.9	-6.5	-5.9	-5.8	-5.5	-5.5
q=80	-11.0	-9.0	-7.9	-7.2	-6.9	-6.3	-5.8	-5.7	-5.4	-5.4
q=85	-10.4	-8.9	-7.8	-7.0	-6.6	-6.0	-5.7	-5.6	-5.2	-5.1
q=90	-9.5	-8.6	-7.3	-6.6	-6.1	-5.6	-5.1	-5.1	-5.0	-4.6
q=95	-7.7	-7.1	-6.1	-5.5	-5.0	-4.7	-4.2	-4.0	-4.0	-3.8

Table 5. T = 100, skew = -1.. [illegible] sites simulated columns: number of data points *m* = 10 (10) 100 rows: confidence levels q = 5 (5) 95.

Error in Confidence Levels, Skew = -1.0

	m=10	m=20	m=30	m=40	m=50	m=60	m=70	m=80	m=90	m=100
q=5	-4.6	-3.4	-2.5	-1.9	-1.5	-1.1	-0.8	-0.6	-0.4	-0.1
q=10	-5.6	-2.7	-1.2	-0.1	+0.3	+0.6	+1.1	+1.3	+1.4	+1.4
q=15	-3.1	+0.3	+1.3	+2.5	+2.6	+2.8	+3.1	+3.3	+3.3	+3.4
q=20	+1.0	+3.7	+4.4	+4.9	+4.9	+5.0	+5.3	+5.2	+5.0	+5.0
q=25	+5.6	+6.9	+7.2	+7.5	+7.3	+7.2	+7.3	+7.3	+6.9	+6.7
q=30	+9.7	+9.9	+9.8	+9.9	+9.4	+9.1	+9.3	+9.1	+8.4	+8.2
q=35	+13.9	+13.1	+12.5	+11.9	+11.6	+11.0	+11.0	+10.7	+10.1	+9.8
q=40	+18.2	+16.7	+15.3	+14.2	+13.7	+12.8	+12.4	+12.2	+11.6	+11.1
q=45	+22.5	+19.9	+18.0	+16.6	+15.6	+15.0	+13.9	+13.7	+13.0	+12.3
q=50	+25.7	+22.9	+20.6	+18.9	+17.7	+16.8	+15.5	+14.8	+14.1	+13.4
q=55	+27.4	+25.5	+23.1	+21.3	+19.9	+18.2	+16.9	+15.9	+15.3	+14.5
q=60	+27.7	+26.7	+24.6	+23.1	+21.5	+20.0	+18.6	+17.3	+16.2	+15.6
q=65	+26.6	+26.4	+25.0	+23.6	+22.3	+21.0	+19.5	+18.5	+17.4	+16.4
q=70	+24.6	+24.8	+24.1	+23.1	+22.2	+21.2	+20.0	+18.9	+18.0	+17.0
q=75	+21.6	+22.1	+21.8	+21.3	+20.8	+20.1	+19.2	+18.5	+17.7	+16.9
q=80	+18.0	+18.6	+18.5	+18.3	+18.1	+17.9	+17.3	+16.8	+16.3	+15.9
q=85	+13.9	+14.3	+14.4	+14.4	+14.3	+14.3	+14.1	+13.8	+13.6	+13.5
q=90	+9.5	+9.8	+9.8	+9.8	+9.8	+9.8	+9.7	+9.7	+9.7	+9.6
q=95	+4.9	+5.0	+5.0	+5.0	+5.0	+5.0	+5.0	+5.0	+5.0	+5.0

The errors shown in Tables 1-5 show inaccuracies which make the use of (10) questionable for larger (but frequently occurring) values of skew. This creates a significant problem concerning their use in general since even if the site estimate for skew is very small, say zero, it is usually the case that the number of data points available do not allow one to rule out with high confidence the possibility that the skew is really large enough to fall in the range where the simulations above show considerable inaccuracy [Whitley and Hromadka, 1993].

7 Conclusions

Approximate formulas are given in Chowdhury and Stedinger [1991] for confidence levels for estimating the T-year flood when the mean, standard deviation, and skew are all estimated from data at a single site. The accuracy of these formulas is tested by means of simulations for the values of skews γ=-1(1/2)1, data points at the site m=10(10)100, and confidence levels q=5(5)(95). These simulations show that the use of these formulas is problematic in most circumstances.

However, a major contribution of Chowdhury and Stedinger [1991] which should not be overlooked is in suggesting a general form that an approximation formula might take: namely, equation (10). The derivation of λ in [Chowdhury and Stedinger, 1991] by means of relatively crude first order asymptotic expressions suggests that a better λ might make the Chowdhury and Stedinger formula (10) more accurate.

References

Advisory Committee on Water Data. 1982: Guidelines for Determining Flood Flow Frequency, Bulletin #17B of the Hydrology Subcommittee, OWDC, US Geological Survey, Reston, VA

Bobee, B., et. al. 1993: Towards a systematic approach to comparing distributions used in flood frequency analysis, J. of Hydrology 142, 121-136

Bobee, B. 1973: Sample error of T-year events computed by fitting a Pearson type 3 distribution, Water Resour. Res. 5, 1264-1270

Bobee, B.; Robitaille, R. 1975: Correction of bias in the estimation of the coefficient of skewness, Water Resour. Res. 11, 851-854

Bobee, B.; Robitaille, R. 1977: The use of the Pearson type 3 and log-Pearson type 3 distributions revisited, Water Resour. Res. 13, 427-443

Bowman, R.; Shenton, L. 1988: Properties of Estimators for the Gamma Distribution, Marcel Dekker, New York

Burden, R., Faires, J. 1993: Numerical Analysis, 5th ed., PW-Kent, Boston

Chowdhury, J.; Stedinger, J. 1991: Confidence intervals for design floods with estimated skew coefficient, ASCE J. Hyd. Eng. 11, 811-831

Cohon, J. et. al. 1988: Estimating Probabilities of Extreme Floods, National Academy Press, Washington D.C.

Devroye, L. 1986: Non-Uniform Random Variable Generation, Springer, New York

Hardison,C. 1974: Generalized skew coefficients of annual floods in the United States and their application, Water Resour. Res. 10, 745-752

Hu, S. 1987: Determination of confidence intervals for design floods, J. Hydrol. 96, 201-213

Kirby, W. 1974: Algebraic boundedness of sample statistics, Water Res. Research 10, 220-222

Kite, G. 1975: Confidence intervals for design events, Water Resour. Res. 11, 48-53

Kite, G. 1976: Reply to comments on "Confidence limits for design events", Water Resour. Res. 12, 826

Lettenmaier, D.; Burges, S. 1980: Correction for bias in estimation of the standard deviation and coefficient of skewness of the log Pearson 3 distribution, Water Resour. Res. 16, 762-766

McCuen, R. 1979: Map skew, J. Water Resour. Plann. Manage., ASCE 105(2) 269-277

Phien, H.; Hsu, L. 1985: Variance of the T-year event in the log-Pearson type 3 distribution, J. Hydrol. 77, 141-158

Press, W. et. al. 1989: Numerical Recipes in Pascal, Cambridge Univ. Press, New York

Resnikoff, G.; Lieberman, G. 1957: Tables of the Non-Central t-Distribution, Stanford Univ. Press, CA

Stedinger, J. 1980: Fitting log normal distributions to hydrologic data, Water Resour. Res. 16(3) 481-490

Stedinger, J. 1983: Confidence intervals for design events, CEJ. Hyd. Eng. 109, 13-27

Tasker, G. 1978: Flood frequency analysis with a generalized skew coefficient, Water Resour. Res. 14, 373-376

Tasker, G.; Stedinger, J. 1986: Regional skew with weighted LS regression, ASCE J. Water Resour. Planning Mgmt. 112, 709-722

Whitley, R.; Hromadka II, T. 1986: Computing confidence intervals for floods I, Microsoftware for Engineers 2(3) 138-150

Whitley, R.; Hromadka II, T. 1986: Computing confidence intervals for floods II, Microsoftware for Engineers 2(3) 151-158

Whitley, R.; Hromadka II, T. 1987: Estimating 100-year flood confidence intervals, Adv. Water. Resour. 10, 225-227

Whitley, R.; Hromadka II, T. 1993: Testing for non-zero skew in maximum discharge runoff data, Water Resour. Res. 29, 531-534

Wilson, E.; Hilferty, M. 1931: The distribution of chi-square, Proc. Nat. Acad. Sci. U.S.A. 17, 684-688

World Meteorological Organization. 1989: Operational Hydrology Report 33, Secretariat of the WMO, Geneva, Switzerland