Estimating 100-year flood confidence intervals

R. J. Whitley

Department of Mathematics, University of California, Irvine, CA 92717, USA

T. V. Hromadka

Williamson and Schmid, 17782 Sky Park Blvd., Irvine, CA 92714, USA

The estimation of the 100-year flood, or more generally the *T*-year flood, is a basic problem in hydrology. An important source of uncertainty in this estimate is that caused by the uncertain estimation of parameters of the flood distribution. This uncertainty can have a significant effect on the flood design value, and its quantification is an important aspect of evaluating the risk involved in a chosen level of flood protection. In this paper, simulation is used to determine confidence intervals for the flood design value. The simulation allows verification of Stedinger's formula not only as it applies to confidence intervals, but also verifies the formula as an approximation to percentiles as well.

INTRODUCTION

The estimation of the 100-year flood, or more generally the *T*-year flood, is a basic problem in hydrology. An important source of uncertainty in this estimate is that caused by the uncertain estimation of parameters of the flood distribution. This uncertainty can have an significant effect on the flood design value, and its quantification is an important aspect of evaluating the risk involved in a chosen level of flood protection.

It is noted in Bulletins 17A and 17B[7,8], that in the case of a flood distribution whose logarithm is normally distributed, confidence intervals for the *T*-year flood can be obtained by the use of the noncentral student's *t*-distribution. The more general case, following the guidelines of Bulletins 17A and 17B, is when the logarithm of the flood distribution has a Pearson III distribution with a nonzero skew parameter. This case of nonzero skew is more complicated than the case of a lognormally distributed flood which is the case of zero skew[3,4,6,10,13,14,18]. In an important paper[18] Stedinger shows that the confidence intervals, for quantiles, which are given in the US Water Resources Council guidelines[7,8] are not satisfactory. He uses a variance ratio formula due to Kite[13] and derives an expression for confidence intervals for the quantiles which he shows is satisfactory in several simulations.

In the following, confidence intervals for the flood design value are found by simulation. The simulation gives an approximate distribution which can be used to obtain confidence intervals of various levels. It also allows the verification of Stedinger's formula not only as it applies to confidence intervals, but extends the results contained in Stedinger[18] in showing that the formulae are good approximations to percentiles of the sampling distribution.

Accepted February 1987. Discussion closes February 1988.

Equations

First consider the case where X, the logarithm of the maximum annual discharge, has a normal distribution, i.e., the case of zero skew. For the *T*-year flood, take $p=1-1/T$, and let y_p be the *p*th quantile of X; i.e.,

$$P(X \leq y_p) = p \tag{1}$$

It is y_p that we want to estimate. The usual estimates for the mean μ and standard deviation σ of X, based on m data points, are denoted by $\hat{\mu}$ and $\hat{\sigma}$. Then

$$(y_p - \hat{\mu})/\hat{\sigma} = ([(\mu-\hat{\mu})/\sigma] + z_p)/(\hat{\sigma}/\sigma) \tag{2}$$

where z_p is the *p*th quantile for a normal distribution with mean 0 and standard deviation 1. This can be written as

$$(1/\sqrt{m})[(Z + z_p\sqrt{m})/\sqrt{W}] \tag{3}$$

The random variable in brackets in (3) has a noncentral *t*-distribution, with noncentrality parameter $\delta = z_p\sqrt{m}$; the special case $\delta=0$ is the student's *t*-distribution. Thus confidence intervals for y_p can be computed using equations (2) and (3).

Second consider the case of nonzero skew, where the logarithm X of the yearly peak discharge is assumed to have a Pearson type III distribution with density function

$$f(x) = (1/|a|\Gamma(b))[(x-c)/a]^{b-1} \exp -[(x-c)/a] \tag{4}$$

where, in the case of positive a, the density is given by (4) for $x>c$ and is zero for $x<c$, while in the case of negative a the density is given by (4) for $x<c$ and is zero for $x>c$. Note that $\gamma^2 = 4/b$ where a has the same sign as γ (see, for example Ref. 9).

In following the guidelines of Bulletins 17A and 17B[7,8], the skew coefficient γ is estimated either from a map of regional skews or from a large pool of data from that region. Consequently the error in estimating γ is of an entirely different kind than that which arises in estimating μ and σ by means of m data points for the specific area for

Table 1. Percent relative error in Stedinger-Kite formula – five data points – 100 year flood

Percentiles	5%	25%	75%	95%
Skew = −0.75	−5.5	−2.6	0.1	1.4
Skew = −0.50	−1.9	−0.2	0.2	0.4
Skew = −0.25	0.5	0.3	−0.1	−0.9
Skew = 0.00	0.0	−0.5	0.4	0.7
Skew = 0.25	−1.4	−1.0	0.2	1.0
Skew = 0.50	−5.8	−2.8	0.0	2.3
Skew = 0.75	−9.6	−4.8	0.0	2.9

Table 2. Percent relative error in Stedinger-Kite formula – ten data points – 100 year flood

Percentiles	5%	25%	75%	95%
Skew = −0.75	−2.8	−1.8	−0.2	1.1
Skew = −0.50	−0.9	−0.4	−0.2	0.1
Skew = −0.25	−0.7	−0.1	0.1	0.4
Skew = 0.00	0.1	0.0	−0.2	0.1
Skew = 0.25	−0.7	−0.6	−0.5	0.0
Skew = 0.50	−1.4	−1.0	−0.3	0.3
Skew = 0.75	−3.4	−2.3	−0.8	−0.1

which the T-year flood is being estimated, and what is usually done to simplify this complicated situation, and what we will do, is to suppose that γ is given exactly. This focuses attention on that part of the variability in the estimate of the T-year flood which arises from the uncertainty in the estimation of μ and σ, and ignores the variability which comes from the uncertainty in the estimate of γ.

A computation shows that, in terms of distributions,

$$(y_p-\hat{\mu})/\hat{\sigma}=(t_p-\hat{\mu}_Z)/\hat{\sigma}_Z \qquad \text{for } a>0$$

$$(y_p-\hat{\mu})/\hat{\sigma}=(\hat{\mu}_Z-t_{1-p})/\hat{\sigma}_Z \qquad \text{for } a<0 \tag{5}$$

Where t_q is the $100q$%th percentile for the gamma distribution Z which has the one parameter density

$$g(x)=(1/\Gamma(b))x^{b-1}e^{-x} \tag{6}$$

And where $\hat{\mu}_Z$ is the estimator for the mean of Z and $\hat{\sigma}_Z^2$ is the customary estimator for the variance of Z. Confidence intervals for y_p were obtained by simulating the distribution of

$$(\delta-\hat{\mu}_Z)/\hat{\sigma}_Z \tag{7}$$

for Z having the gamma distribution (6), and applying this to equation (5).

COMPUTATION AND COMPARISON

In Ref. 10 Hardison simulated the random variable in (7), in order to compare it with the noncentral t-distribution. His simulation involved sample sizes 10, 20 and 40, skew coefficients of −1.0 and 1, levels of significance 0.01, 0.05, 0.10, 0.90, 0.95, and 0.99, and were each determined by 1000 point simulations.

To obtain more accuracy and to cover a more extensive range of sample sizes, skews, and levels of significance than Ref. 10, we simulated a set of values of the random variable (7); our simulation gives percentiles for the range 5%(5%)95%. For the details of this simulation see Ref. 19.

Stedinger, in Ref. 18, uses Kite's formula for the variance of $\hat{y}_p$ and derives an approximate confidence interval for y_p; see his equations (20) and (21). His success in using this formula to construct correct confidence intervals suggests that his formula might even be accurate for determining the qth percentile Γ_q. When restated in terms of the percentiles, his formula is:

$$\Gamma_q=K_p+\lambda(\zeta_q(p)-z_p) \tag{8}$$

The index p is related to the T-year flood, as above, by $p=1-1/T$, and the constant z_p is the pth percentile for a normal $N(0, 1)$ distribution. The constant K_p is given by

$$K_p=(t_p-b)/\sqrt{b} \qquad \text{for } a>0$$

$$K_p=(b-t_{1-p})/\sqrt{b} \qquad \text{for } a<0 \tag{9}$$

in which t_p or t_{1-p} can be obtained by applying the accurate Wilson-Hilferty transformation[illegible] to either z_p or z_{1-p}. The factor λ is a positive number given by Kite's variance ratio formula.

$$\lambda^2=[1+\gamma K_p+(0.5)(1+0.75\gamma^2)K_p^2]/[1+0.5z_p^2] \tag{10}$$

The use of equation (8) is then reduced to finding $\zeta_q(p)$, where $\zeta_q(p)\sqrt{m}$ is the qth percentile for the noncentral t-distribution with noncentrality parameter $\delta=z_p\sqrt{m}$, which was discussed above in the case of zero skew.

The entries in the tables below are the relative percent difference between the simulation values for the percentiles of the distribution (7) and the values given by equation (8), i.e., 100 × [(equation (8) value − simulation value)/simulation value]. The tables are for the 100 year flood. The numbers of points in the simulations are: for 5 data points, 30 000 sets of 5 data points (this is for comparison purposes; 5 data points are not enough to use in estimating a 100-year event); for 10 data points, 20 000 sets of 10 data points; for 20 data points, 15 000 sets of 20 data points; and for 30 data points, 10 000 sets of 30 data points. For the details of the computations, see Ref. 19.

Table 3. Percent relative error in Stedinger-Kite formula – twenty data points – 100 year flood

Percentiles	5%	25%	75%	95%
Skew = −0.75	−1.3	−1.0	−0.4	−0.7
Skew = −0.50	−0.7	−0.3	−0.1	−0.1
Skew = −0.25	−0.2	0.3	0.0	0.0
Skew = 0.00	0.2	0.0	0.3	0.2
Skew = 0.25	−0.2	−0.6	−0.3	0.0
Skew = 0.50	−1.2	−0.7	−0.4	0.9
Skew = 0.75	−2.0	−1.4	−0.7	−0.3

Table 4. Percent relative error in Stedinger-Kite formula – thirty data points – 100 year flood

Percentiles	5%	25%	75%	95%
Skew = −0.75	−1.0	−0.8	−0.4	0.0
Skew = −0.50	0.1	−0.1	−0.1	−0.4
Skew = −0.25	0.2	0.3	0.0	−0.4
Skew = 0.00	0.0	0.1	0.0	−0.6
Skew = 0.25	−0.6	−0.1	−0.3	−0.3
Skew = 0.50	−0.5	−0.5	−0.4	−0.4
Skew = 0.75	−0.9	−0.6	−0.1	0.2

Formula (8) is therefore an acceptable way to compute the percentiles for parameters in the usual range for hydrology problems.

In the case of zero skew, equation (8) is exact and so the entry in the skew = 0.00 row of the tables gives the error in the simulation percentiles, which indicates the general overall size of the simulation errors.

REFERENCES

1 Beard, L. R. Impact of hydrological uncertainties on flood insurance, *J. Hydraulics Div., ASCE*, 1978, **104**, 1473–1484
2 Beard, L. R. Statistical Methods in Hydrology, US Army Corp of Engineers, Sacramento, CA, 1962
3 Bobee, B. Comment on The log Pearson type 3 distribution: The *T*-year event and its asymptotic standard error by R. Condie, *Water Res. Research*, 1979, **15**, 189–190
4 Bobee, B. Sample error of *T*-year events computed by fitting a Pearson type 3 distribution, *Water Res. Research*, 1973, **9**, 1264–1270
5 Breiman, L. Statistics: With a View Towards Applications, Houghton-Mifflin, Boston, 1973
6 Condie, R. The log Pearson type 3 distribution: The *T*-year event and its asymptotic standard error by maximum likelihood theory, *Water Res. Research*, 1977, **13**, 987–991
7 Guidelines for Determining Flood-Flow Frequency, Water Resources Council Hydrology Committee, Bulletin #17A, Washington DC, 1977
8 Guidelines for Determining Flood-Flow Frequency, Water Resources Council Hydrology Committee, Bulletin #17B, Washington DC, 1981
9 Hall, M. J. Urban Hydrology, Elsevier, London, 1984
10 Hardison, C. Confidence limits for flood frequency curves computed from samples from Pearson type III populations, *J. of Research U.S.G.S.*, 1976, **4**, 545–547
11 Hromadka II, T. V. and Whitley, R. J. Calibrating a Design Storm Method, in preparation (1986)
12 Hoshi, K. and Burges, S. J. Approximate estimation of the derivative of a standard gamma quantile for use in confidence interval estimates, *J. of Hydrology*, 1981, **53**, 317–325
13 Kite, G. W. Confidence limits for design events, *Water Res. Research*, 1975, **11**, 48–53
14 Kite, G. W. Frequency and Risk Analysis in Hydrology, Water Resources Pub., Fort Collins, CO, 1977
15 Mathur, R. K. A Note on the Wilson-Hilferty transformation of chi-squared, *Bull. Calcutta Stat. Assoc.*, 1961, **10**, 103–105
16 Moran, P. A. P. The Statistical treatment of floods, *Trans. Amer. Geophysical Union*, 1957, **38**, 519–523
17 Resnikoff, G. J. and Lieberman, G. J. Tables of the Non-Central *t*-Distribution, Stanford University Press, CA, 1957
18 Stedinger, J. R. Confidence intervals for design events, *J. of Hydraulic Engineering*, 1983, **109**, 13–27
19 Whitley, R. J. and Hromadka II, T. V. Computing Confidence Limits for Floods I and II, *Microsoftware for Engineers*, 1986, **2**, 138–158
20 Wilson, E. B. and Hilferty, M. M. The distribution of chi-square, Proc. Nat. Acad. Sci., USA, 1931, **17**, 684–688