Journal of Hydrology, 118 (1990) 19–37
Elsevier Science Publishers B.V., Amsterdam — Printed in The Netherlands

[4]

BIAS IN LOG-TRANSFORMED FREQUENCY DISTRIBUTIONS

BRUCE G. WILSON[1], BARRY J. ADAMS[2*] and BRYAN W. KARNEY[2]

[1]*Proctor and Redfern Group, Toronto, Ont. (Canada)*
[2]*Department of Civil Engineering, University of Toronto, Toronto, Ont. M5S 1A4 (Canada)*

(Received September 20, 1989; accepted for publication November 6, 1989)

ABSTRACT

Wilson, B.G., Adams, B.J. and Karney, B.W., 1990. Bias in log-transformed frequency distributions. J. Hydrol., 118: 19–37.

Estimates of the parameters of frequency distributions are often derived from the logarithms of a series of measured data rather than from the original series itself. This is common practice, for example, in estimating flood frequencies from measured streamflow data, extreme wave frequencies from wave height data, as well as in many other contexts. This procedure leads to biased estimates of the moments and quantiles of the frequency distribution in arithmetic space. This bias is the result of the non-linear inverse transformation process. The case of the two parameter lognormal distribution is examined in detail and analytical expressions for the bias are derived. The bias introduced by the logarithmic transformation is dependent on the sample size and the population parameters. The derived analytical expressions for bias agree with the results of a series of Monte Carlo simulations.

INTRODUCTION

It is common practice in many applications to fit a frequency distribution to the logarithms of a series of data rather than to the measured arithmetic series. For instance, log-transformed frequency distributions have been used to determine extreme wave heights for use in designing offshore structures (Isaacson and MacKenzie, 1981), to describe the distribution of hydraulic conductivity values in a porous medium (Freeze, 1975) and to fit flood frequency data (Benson, 1968; Stedinger, 1980; U.S. Water Resources Council (USWRC), 1981). In flood frequency analysis, the lognormal distribution has been used to represent flood flow probabilities for many years (Hazen, 1914) and other theoretical frequency distributions are routinely fitted to the logarithms of flood flow data. The use of the log Pearson type 3 distribution has been recommended by the USWRC (1981). In this procedure, a Pearson type 3 distribution is fitted to the base 10 logarithms of a series of maximum annual flood flows.

When dealing with a lognormal distribution it seems quite natural to work in log space because the properties of the normal distribution are well known and the mathematical manipulations are easily performed. However, the logarithmic transformation is simply a mathematical convenience, and the logarithm of a flood flow has little, if any, physical meaning. The concern of the

0022-1694/90/$03.50

flood frequency analyst is with actual flows, not with the logarithms of flows. Therefore, the goodness-of-fit of any distribution must be measured in arithmetic space, not in log space.

The logarithmic and exponential functions are, of course, very non-linear. The assumption implicit in applying any fitting procedure in log space is that these non-linear transformations have little or no effect on the fit of the model to the original data. This assumption does not appear to have been thoroughly tested in the literature. In fact, the logarithmic transformation can be expected to cause a bias: in general, $E[g(x)]$ is not equal to $g(E[x])$ for most functions of a random variable, $g(x)$. The purpose of this paper is to determine the magnitude of the bias caused by employing a logarithmic transformation.

There is some empirical evidence reported in the hydrologic literature concerning how well distributions, estimated from the logarithms of a series, fit the original data but the conclusions that have been drawn are somewhat ambiguous. Bobeé and Robitaille (1977) and Ashkar and Bobeé (1987) report that estimating the parameters of a log Pearson type 3 distribution by a method of moments in arithmetic space provides a better fit than when the parameters are estimated from the logarithms of the flows. However, Nozdryn-Plotnicki and Watt (1979) found that the method which provided the best fit depended on the parameters of the distribution in question. They found that Bobeé's method worked well if the skew of a sample in log space was negative, but that fitting the distribution in log space is preferable if the skew is positive.

Similar results are reported for lognormal distributions. Stedinger (1980) found that fitting a two parameter lognormal distribution using the moments of the untransformed variate was preferable to fitting the distribution in log space for certain combinations of sample size and population coefficient of variation. However, different combinations of sample size and population coefficient of variation led to the opposite conclusion. Overall, Stedinger recommended that the parameters of a two parameter lognormal distribution should be estimated in log space, in part because this leads to unbiased estimates of the arithmetic population mean and standard deviation.

Work by Koch and Smillie (1986) on biases in log transformed regression models has focused attention on the logarithmic transformation of frequency distributions. Koch and Smillie demonstrated that fitting a linear regression model to a logarithmic or power transformation of observed data produced bias in the model when transformed back into arithmetic space. This bias exists unless there is a perfect fit of the regression model and is the direct result of the inverse transformation process. The magnitude of the bias is determined by the variance of the error in the regression model.

A similar phenomenon occurs when a frequency distribution is fitted to the logarithms of a series of data. All parameter estimates based on this transformed data are random variables and as such are described by probability distributions. Even if these log space parameter estimates are unbiased, some error in any particular parameter estimate can be expected. Because of this

random error in parameter estimates, a bias can be expected when the model is transformed back into arithmetic space. The purpose of this paper is to outline and quantify the sources of bias.

LOG TRANSFORM BIAS IN THE LOGNORMAL DISTRIBUTION

To investigate logarithmic transformation bias in frequency distributions, the two parameter lognormal distribution is examined in detail. The properties of the lognormal distribution are well known (e.g. Aitchison and Brown, 1966) and analytical solutions for parameters are possible. The following discussion reviews properties of parameters and their estimates since this information is relevant to subsequent derivations.

Consider a random variable, y, normally distributed with mean μ_y and variance σ_y^2. The probability density function of y is:

$$f_Y(y) = \frac{1}{\sigma_y\sqrt{2\pi}} \exp\left\{ - \frac{(y - \mu_y)^2}{2\sigma_y^2} \right\} \tag{1}$$

The random variable $x = \exp(y)$ is lognormally distributed with the following probability density function:

$$f_X(x) = \frac{1}{x\sigma_y\sqrt{2\pi}} \exp\left\{ - \frac{[\ln(x) - \mu_y]^2}{2\sigma_y^2} \right\} \tag{2}$$

The following relationships hold between the arithmetic moments of this distribution, μ_x and σ_x^2, and the log space parameters μ_y and σ_y^2 (Aitchison and Brown, 1966):

$$\mu_x = \exp\left(\mu_y + \frac{\sigma_y^2}{2}\right) \tag{3}$$

$$\sigma_x^2 = \mu_x^2 [\exp(\sigma_y^2) - 1] \tag{4}$$

Suppose that a random sample of size n, $(x_1, x_2, \ldots, x_n)$, is drawn from a population described by eqn. (2). The mean and variance of the logarithms of the series are estimated following Stedinger's (1980) recommendation as:

$$\hat{\mu}_y = \bar{y} = \frac{1}{n}\sum_{i=1}^{n} \ln(x_i) \tag{5}$$

and

$$\hat{\sigma}_y^2 = v_y^2 = \frac{1}{n}\sum_{i=1}^{n} [\ln(x_i) - \bar{y}]^2 \tag{6}$$

Since the transformed variate, y, is normally distributed, the sampling distributions and other important properties of these parameter estimates are well known. The statistic $\bar{y}$ is distributed normally with a mean μ_y and variance σ_y^2/n. The statistic v_y^2 is distributed as a χ^2 distribution which is a special case of the

gamma distribution with a mean of $(n-1)\sigma_y^2/n$ and variance $2(n-1)\sigma_y^4/n^2$. Furthermore, since y is normally distributed, $\bar{y}$ and v_y^2 are independent (Guttman et al., 1982).

Substituting eqns. (5) and (6) into eqn. (3), the estimate of the mean of the fitted distribution in arithmetic space is:

$$\hat{\mu}_x = \exp\left(\bar{y} + \frac{v_y^2}{2}\right) \tag{7}$$

The estimate of the second central moment of the fitted distribution in arithmetic space, found by substituting eqns. (5) and (6) into eqn. (4) is:

$$\hat{\sigma}_x^2 = \exp(2\bar{y} + v_y^2)[\exp(v_y^2) - 1] \tag{8}$$

To determine if eqns. (7) and (8) are biased, the expected value of each expression must be compared with the population value of the moment being estimated. Taking the expectation of eqn. (7) yields:

$$E(\hat{\mu}_x) = E\left[\exp\left(\bar{y} + \frac{v_y^2}{2}\right)\right] = E\left[\exp(\bar{y})\exp\left(\frac{v_y^2}{2}\right)\right] \tag{9}$$

since $\bar{y}$ and v_y^2 are independent, their covariance is equal to zero and, accordingly:

$$E(\hat{\mu}_x) = E[\exp(\bar{y})]\, E\left[\exp\left(\frac{v_y^2}{2}\right)\right] \tag{10}$$

Since $\bar{y}$ is normally distributed, $\exp \bar{y}$ is lognormally distributed. Then from eqn. (3):

$$E[\exp \bar{y}] = \exp\left(\mu_y + \frac{\sigma_y^2}{2n}\right) \tag{11}$$

Recognizing that $v_y^2/2$ follows a gamma distribution with a mean of $(n-1)\sigma_y^2/2n$ and a variance of $(n-1)\sigma_y^4/2n^2$ and using the properties of the moment generating function of a gamma distribution (Blake, 1979), it can be easily shown that:

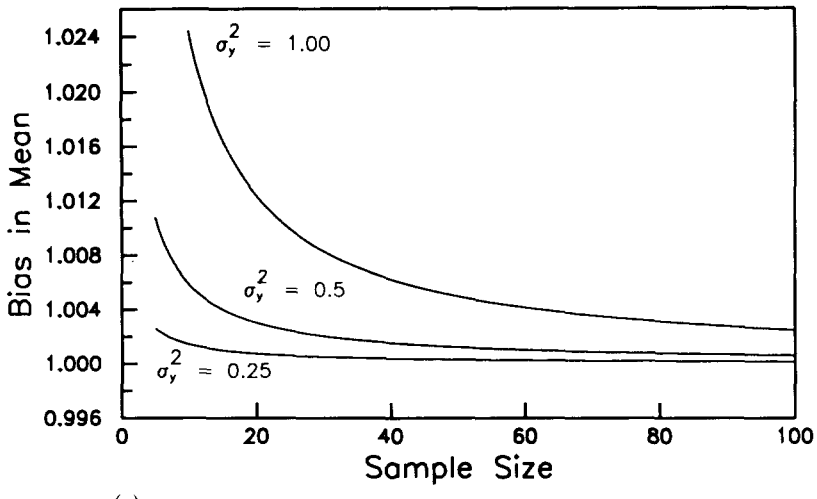
$$E\left[\exp\left(\frac{v_y^2}{2}\right)\right] = \left(1 - \frac{\sigma_y^2}{n}\right)^{-\left(\frac{n-1}{2}\right)} \tag{12}$$

It is noted that as n becomes large, the value of eqn. (12) approaches $\exp\left(\frac{\sigma_y^2}{2}\right)$.

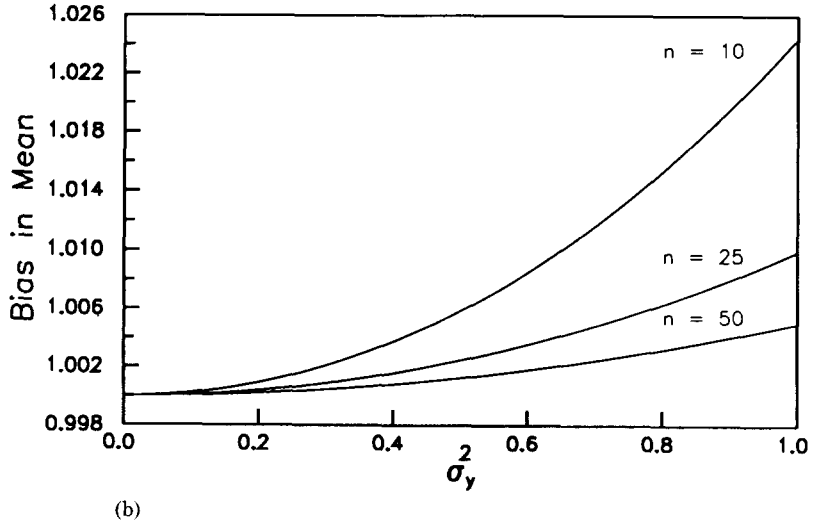
Substituting eqns. (11) and (12) into eqn. (10) yields:

$$E[\hat{\mu}_x] = \left[\exp\left(\mu_y + \frac{\sigma_y^2}{2n}\right)\right]\left(1 - \frac{\sigma_y^2}{n}\right)^{-\left(\frac{n-1}{2}\right)} \tag{13}$$

If the estimate of μ_x from eqn. (7) is unbiased then the ratio of eqn. (13) to eqn.



(a)



(b)

Fig. 1. Bias in arithmetic mean as (a) a function of sample size, (b) as a function of σ_y^2.

(3) would be unity. Forming this ratio results in the following (Kendall and Stuart, 1961, vol. 2, p. 74):

$$\frac{E(\hat{\mu}_x)}{\mu_x} = \left\{\exp\left(\frac{1-n}{2n}\sigma_y^2\right)\right\}\left(1-\frac{\sigma_y^2}{n}\right)^{-\left(\frac{n-1}{2}\right)} \neq 1 \tag{14}$$

This equation arises as a direct result of using eqns. (5), (6) and (7) to estimate the mean of the fitted distribution as suggested by Stedinger (1980). Equation (14) states that the arithmetic mean of a two parameter lognormal distribution

fitted in log space is a biased (but consistent) estimator. The magnitude of the bias depends only upon the sample size (n) and the (generally unknown) variance of the transformed variate population (σ_y^2). An examination of eqn. (14) indicates that for all combinations of n and σ_y^2, the bias is greater than unity. This suggests that estimating parameters in log space tends to overestimate the arithmetic mean. The bias predicted by eqn. (13) is plotted as a function of sample size (n) and the variance of the log transformed variate (σ_y^2) in Fig. 1a and b, respectively. These figures reveal that the bias in the arithmetic mean is usually small, decreases with increasing sample size and increases with increasing σ_y^2.

The expected value of the variance of a two parameter lognormal distribution estimated using eqn. (8) is now considered as follows:

$$E(\hat{\sigma}_x^2) = E\{\exp(2\bar{y} + v_y^2)[\exp(v_y^2) - 1]\}$$

$$= E[\exp(2\bar{y}) \exp(2v_y^2)] - E[\exp(2\bar{y}) \exp(v_y^2)]$$

and, since $\bar{y}$ and v_y^2 are independent:

$$E(\hat{\sigma}_x^2) = E[\exp(2\bar{y})]\, E[\exp(2v_y^2)] - E[\exp(2\bar{y})]\, E[\exp(v_y^2)] \tag{15}$$

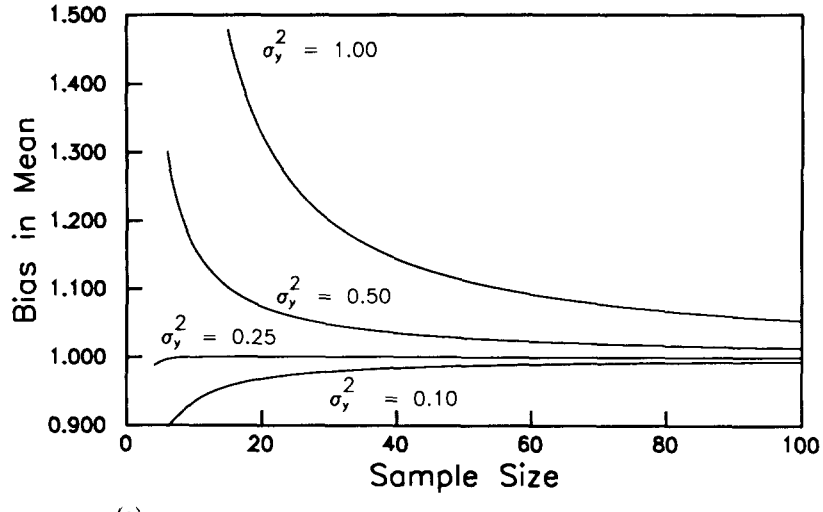
Also, since $2\bar{y}$ is distributed normally and v_y^2 follows a gamma distribution and again utilizing the properties of the moment generating function of the gamma distribution (Blake, 1979), eqn. (15) yields:

$$E(\hat{\sigma}_x^2) = \left[\left(1 - \frac{4\sigma_y^2}{n}\right)^{-\left(\frac{n-1}{2}\right)} - \left(1 - \frac{2\sigma_y^2}{n}\right)^{-\left(\frac{n-1}{2}\right)}\right] \exp\left\{2\mu_y + \frac{2\sigma_y^2}{n}\right\} \tag{16}$$

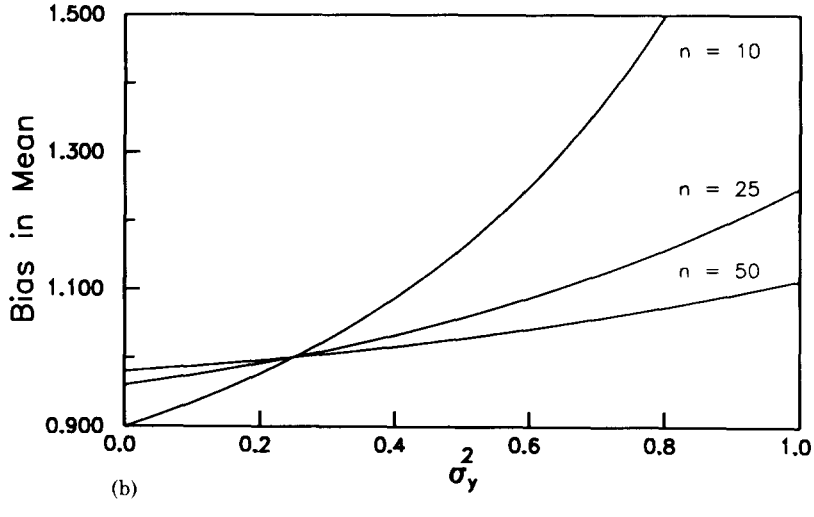
This estimate of σ_x^2 is unbiased if the ratio of eqn. (16) to eqn. (4) is unity. This ratio is:

$$\frac{E(\hat{\sigma}_x^2)}{\sigma_x^2} = \frac{\left[\left(1 - \frac{4\sigma_y^2}{n}\right)^{-\left(\frac{n-1}{2}\right)} - \left(1 - \frac{2\sigma_y^2}{n}\right)^{-\left(\frac{n-1}{2}\right)}\right] \exp\left(\frac{2-n}{n}\sigma_y^2\right)}{\exp(\sigma_y^2) - 1} \tag{17}$$

Equation (17) is a consequence of estimating the variance of a lognormal distribution using eqns. (5), (6) and (8) as suggested by Stedinger (1980). Equation (17) indicates that for most combinations of sample size and population parameters, fitting a two parameter lognormal distribution in log space leads to biased estimates of the variance of the distribution in arithmetic space. Again, the magnitude of the bias depends only upon the sample size (n) and the variance of the log transformed variate (σ_y^2). Unlike the case of the arithmetic mean, the bias given by eqn. (17) is not always small. Indeed, Fig. 2a and b, which plot eqn. (17) for various combinations of n and σ_y^2, respectively, reveal that the bias can be quite substantial. However, for certain combinations of sample size and population parameters the bias is negligible.



(a)



(b)

Fig. 2. Bias in arithmetic variance as (a) a function of sample size, (b) as a function of σ_y^2.

The effect of 'unbiased' estimators

The biases given by eqns. (14) and (17) arise because the log space parameter estimates are random variables and not because of bias in the log space estimators themselves. To emphasize this point, eqn. (6) can be replaced by an unbiaised estimator of the variance of the logarithms of the flows and the resulting bias in the arithmetic moments determined.

An unbiased estimator of the variance of the log transformed variate is:

$$\hat{\sigma}^2 = s_y^2 = \frac{1}{n-1}\sum_{i=1}^{n}[\ln(x_i) - \bar{y}]^2 \tag{18}$$

While this estimator is unbiased, for a normally distributed population it has a larger mean square error than eqn. (6) (Stedinger, 1980).

Substituting eqns. (5) and (18) into eqns. (3) and (4) produces the following estimates of the mean and variance of the fitted distribution in arithmetic space:

$$\hat{\mu}_x = \exp\left(\bar{y} + \frac{s_y^2}{2}\right) \tag{19}$$

and

$$\hat{\sigma}_x^2 = \exp(2\bar{y} + s_y^2)[\exp(s_y^2) - 1] \tag{20}$$

Following derivations parallel to those for eqns. (14) and (17), it can be shown (Wilson, 1988) that:

$$\frac{E(\hat{\mu}_x)}{\mu_x} = \exp\left(\frac{1-n}{2n}\sigma_y^2\right)\left(1 - \frac{\sigma_y^2}{n-1}\right)^{-\left(\frac{n-1}{2}\right)} \tag{21}$$

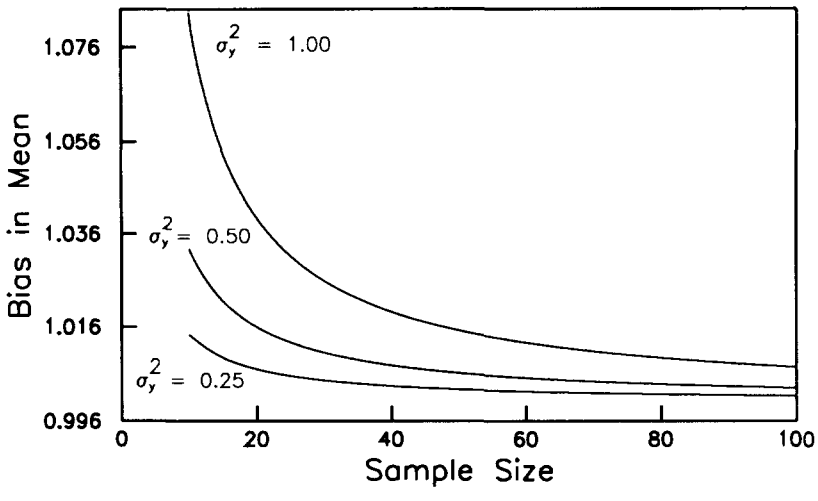
and

$$\frac{E(\hat{\sigma}_x^2)}{\sigma_x^2} = \frac{\left[\left(1 - \frac{4\sigma_y^2}{n-1}\right)^{-\left(\frac{n-1}{2}\right)} - \left(1 - \frac{2\sigma_y^2}{n-1}\right)^{-\left(\frac{n-1}{2}\right)}\right]\exp\left(\frac{2-n}{n}\sigma_y^2\right)}{\exp(\sigma_y^2) - 1} \tag{22}$$

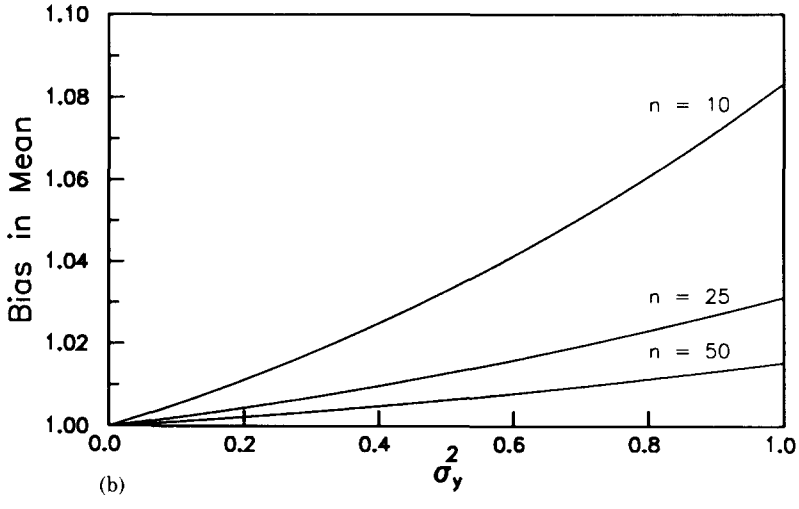
Equation (21), which arises from eqns. (5), (18) and (19), states that the arithmetic mean of a two parameter lognormal distribution fitted in log space is biased and that the bias depends on n and σ_y^2. Bias in the arithmetic mean predicted by eqn. (19) is plotted as a function of sample size (n) and the variance of the log transformed variate (σ_y^2) in Fig. 3a and b, respectively. These figures show that the bias in the arithmetic mean calculated using an unbiased estimate of σ_y^2 is larger than the bias introduced when a biased estimate of σ_y^2 is used.

Equation (22) indicates that the variance of a two parameter lognormal distribution fitted in log space, arising from eqns. (5), (18) and (20), is biased. The magnitude of the bias depends only upon n and σ_y^2. Figure 4a and b plot eqn. (22) for various combinations of n and σ_y^2, respectively. These figures show that although eqn. (22) appears to be quite similar to eqn. (17), the biases given by eqn. (22) are larger.

The bias introduced in the arithmetic space parameters by employing an unbiased estimate of σ_y^2 is significantly larger than that introduced by using a biased estimator in log space. It is ironic that an attempt to reduce bias by using



(a)



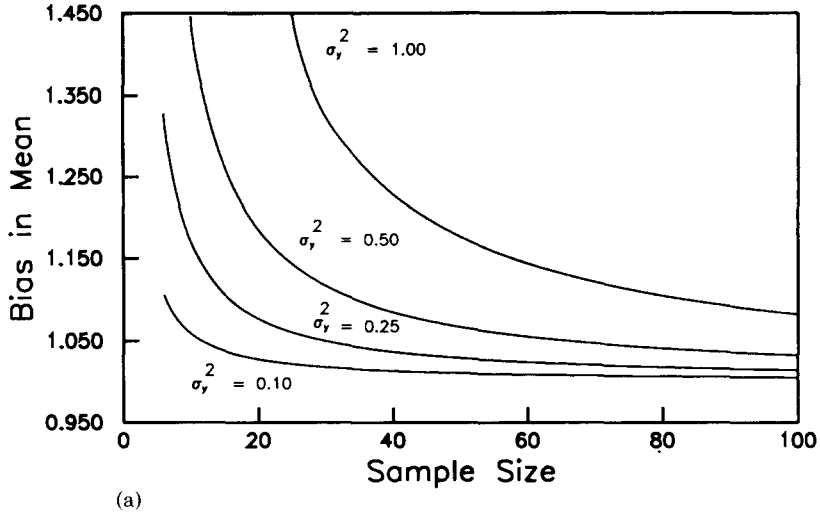
(b)

Fig. 3. Bias in arithmetic mean as (a) a function of sample size, (using unbiased log space estimators), and (b) as a function of σ_y^2 (using unbiased log space estimators).

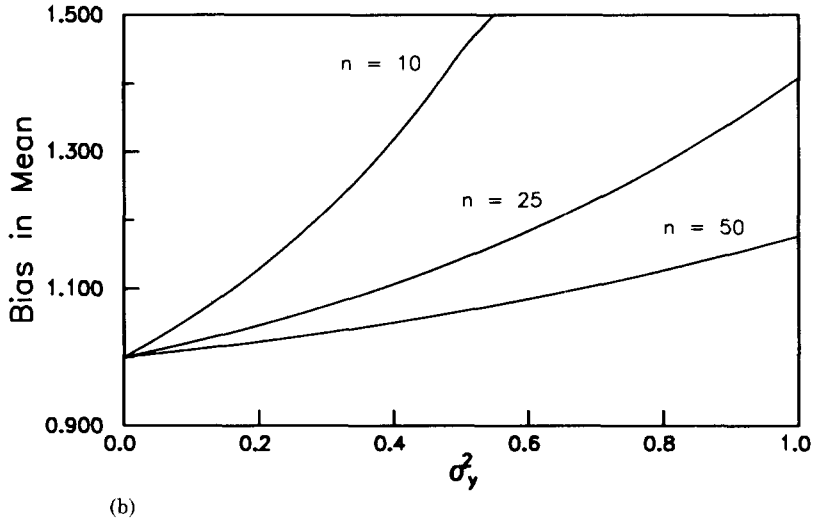
unbiased estimators in log space actually results in a much larger bias in arithmetic space moment estimates.

Bias in quantile estimates

While estimates of the moments of a distribution provide useful information, the flood frequency analyst is usually more concerned about estimating a



(a)



(b)

Fig. 4. Bias in arithmetic variance as (a) a function of sample size (using unbiased log space estimators), and (b) a function of σ_y^2, (using unbiased log space estimators).

particular quantile of a distribution such as the 100 year flood. The pth quantile of a two parameter lognormal distribution is defined as (Aitchison and Brown, 1966):

$$x_p = \exp(\mu_y + \sigma_y z_p) \tag{23}$$

in which σ_y is the standard deviation of the log space variate and z_p is the 100pth

percentile of the standard normal distribution. The pth quantile is usually estimated as:

$$\hat{x}_p = \exp(\bar{y} + v_y z_p) \tag{24}$$

in which $v_y = \sqrt{v_y^2}$.

The expected value of this estimate of $\hat{x}_p$ is:

$$E(\hat{x}_p) = E[\exp(\bar{y} + v_y z_p)] = E[\exp(\bar{y})]\, E[\exp(v_y z_p)] \tag{25}$$

To determine the expected value of the estimate of the pth quantile from eqn. (25), in addition to eqn. (11), the distribution of v_y is required. The exact functional form of the distribution of v_y was derived by Student (1908) who also showed that for moderate sample sizes the distribution is approximately normal with mean:

$$E(v_y) \approx \sqrt{1 - \frac{3}{2n} + \frac{1}{8n^2}}\,\sigma_y \tag{26}$$

and variance

$$V(v_y) \approx \left(1 - \frac{1}{4n}\right)\frac{\sigma_y^2}{2n} \tag{27}$$

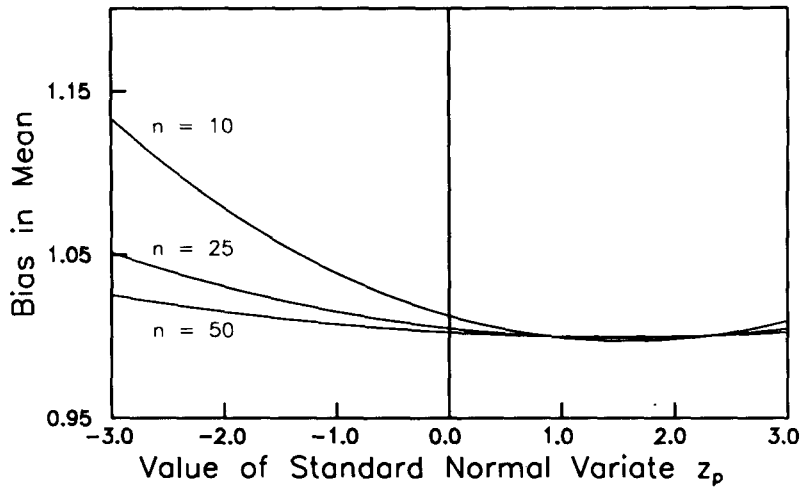
Assuming v_y is distributed normally, $\exp(v_y t_p)$ will be lognormally distributed. Therefore, substituting (26) and (27), eqn. (25) becomes:

$$E(\hat{x}_p) \approx \exp\left\{\mu_y + \frac{\sigma_y^2}{2n}\right\} \exp\left\{\sqrt{1 - \frac{3}{2n} + \frac{1}{8n^2}}\, z_p \sigma_y + \frac{z_p^2 \sigma_y^2 (4n - 1)}{16n^2}\right\} \tag{28}$$

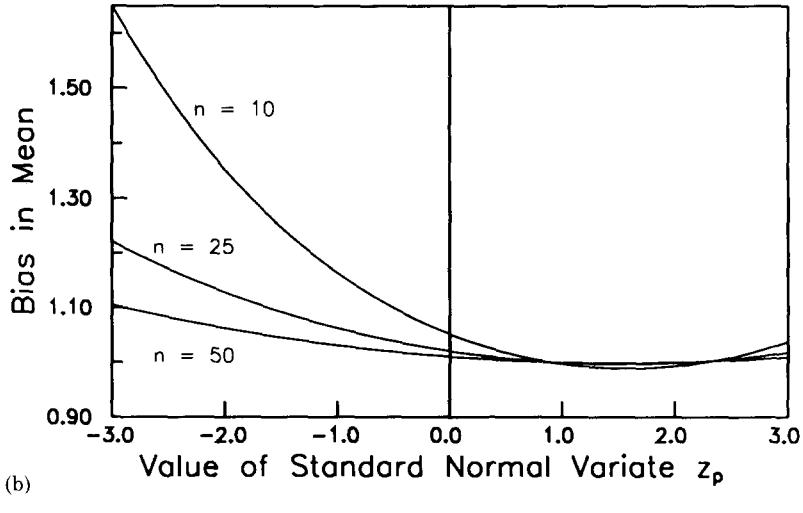
The estimator of the pth quantile is unbiased if the ratio of eqn. (28) to eqn. (23) is unity. This ratio is:

$$\frac{E(\hat{x}_p)}{x_p} \approx \exp\left\{\frac{\sigma_y^2}{2n}\right\} \exp\left\{\left[\sqrt{1 - \frac{3}{2n} + \frac{1}{8n^2}} - 1\right] z_p \sigma_y\right\} \times \exp\left\{\frac{z_p^2 \sigma_y^2 (4n - 1)}{16n^2}\right\} \tag{29}$$

Equation (29) indicates that quantile estimates based on eqn. (24) are biased (but consistent). The magnitude of the bias depends not only upon n and σ_y^2 but also upon the quantile of interest. Figure 5a and b plot the bias in quantile estimates as a function of the quantile of interest and sample size for $\sigma_y^2 = 0.25$ and $\sigma_y^2 = 1.00$, respectively. These figures show that the bias in some quantile estimates can be quite large while the bias in other quantile estimates is relatively small. The bias is generally positive, indicating that estimating quantiles using eqn. (24) tends to overestimate the quantile of interest. However, it is noted that for some combinations of population parameters, eqn. (24) underestimates the quantile of interest.



(a)



(b)

Fig. 5. Bias in quantile estimates (a) $\sigma_y^2 = 0.25$, (b) $\sigma_y^2 = 1.00$.

Fitting a two parameter lognormal distribution in log space can result in rather unbiased estimates of some quantiles (in the range of 10 to 100 year return periods). It is fortuitous that these quantiles are generally of greatest interest to flood frequency analysts. Of course this is not a general result and distributions other than the two parameter lognormal may exhibit larger biases when fitted in log space.

In addition, since $\bar{y}$ is normally distributed and v_y is approximately normally distributed, an inspection of eqn. (24) reveals that individual quantile estimates

will be approximately lognormally distributed. This suggests that even in the range of return periods where quantile estimates are relatively unbiased, these estimates are likely to vary widely about the expected value given by eqn. (28) and to exhibit a positive skew. This agrees with the finding that confidence limits on two parameter lognormal quantile estimates are not equally spaced above and below the mean (Stedinger, 1983).

Summary of bias equations

Since a large number of equations have been presented, it is convenient to summarize the various bias relations and the assumptions that led to them in tabular form. Table 1 conveniently displays the estimates and the estimator bias for the mean, variance and quantile. The table distinguishes between using the sample variance (v_y^2) to estimate the population variance and using an 'unbiased' estimator of the variance in log space (s_y^2). The table allows the difference between the estimators and their resulting biases to be compared at a glance, thus summarizing the primary results of this paper.

COMPARISON WITH MONTE CARLO SIMULATIONS

In order to verify the above derivations, a series of Monte Carlo simulation experiments were performed. Samples of size $n = 10$, 25 and 50 of a random variable, y, were generated from normal distributions with parameters $\mu_y = 1.00$ and $\sigma_y^2 = 0.25$, 0.50, 0.75 and 1.00. The mean and variance of each sample were calculated and used to estimate the mean, variance and three quantiles of $x = \exp(y)$ using eqns. (7), (8) and (24). Averages of the arithmetic mean, variance and quantile estimates over 25 000 repetitions of the experiment were calculated.

Table 2 compares the average arithmetic mean, variance and quantile estimates from the Monte Carlo simulations with the expected values of these quantities given by eqns. (13), (16) and (28), respectively, and with the true population values given by eqns. (3), (4) and (23), respectively. The simulation results are in very good agreement with the theoretical results derived in this paper. As expected, there is poor agreement between the parameter estimates from the simulation and the true population values of the arithmetic moments and quantiles. These simulation results confirm the existence of a log transform bias and the effect of sample size (n) and the variance of the log transformed variate (σ_y^2) on that bias.

LOG TRANSFORM BIAS IN OTHER FREQUENCY DISTRIBUTIONS

Unfortunately, analytical solutions for log transform bias for distributions other than the two parameter lognormal can be difficult, if not impossible to achieve. This is the case when an attempt is made to determine the bias resulting from a logarithmic transformation of a Pearson type 3 distribution. Unlike the lognormal distribution, the sampling properties of the parameter

TABLE 1

Bias in moment and parameter estimates

Estimate of μ_y and σ_y^2	Quantity	Relation	Eqn. No.
$\hat{\mu}_y = \bar{y}$	Estimate of μ_x	$\exp\left(\bar{y} + \frac{v_y^2}{2}\right)$	(7)
and			
$\hat{\sigma}_y^2 = v_y^2$	Bias of $\hat{\mu}_x$	$\exp\left(\frac{1-n}{2n}\sigma_y^2\right)\left(1 - \frac{\sigma_y^2}{n}\right)^{-\left(\frac{n-1}{2}\right)}$	(14)
$\hat{\mu}_y = \bar{y}$	Estimate of μ_x	$\exp\left(\bar{y} + \frac{s_y^2}{2}\right)$	(19)
and			
$\hat{\sigma}_y^2 = s_y^2$	Bias of $\hat{\mu}_x$	$\exp\left(\frac{1-n}{2n}\sigma_y^2\right)\left(1 - \frac{\sigma_y^2}{n-1}\right)^{-\left(\frac{n-1}{2}\right)}$	(21)
$\hat{\mu}_y = \bar{y}$	Estimate of σ_x^2	$\exp(2\bar{y} + v_y^2)[\exp(v_y^2) - 1]$	(8)
and			
$\hat{\sigma}_v^2 = v_y^2$	Bias of $\hat{\sigma}_x^2$	$\dfrac{\left[\left(1 - \frac{4\sigma_y^2}{n}\right)^{-\left(\frac{n-1}{2}\right)} - \left(1 - \frac{2\sigma_y^2}{n}\right)^{-\left(\frac{n-1}{2}\right)}\right] \exp\left(\frac{2-n}{n}\sigma_y^2\right)}{\exp(\sigma_y^2) - 1}$	(17)

$\hat{\mu}_y = \bar{y}$	Estimate of σ_x^2	$\exp(2\bar{y} + s_y^2)[\exp(s_y^2) - 1]$	(20)
and			
$\hat{\sigma}_y^2 = s_y^2$	Bias of $\hat{\sigma}_x^2$	$\dfrac{\left[\left(1 - \dfrac{4\sigma_y^2}{n-1}\right)^{-\left(\frac{n-1}{2}\right)} - \left(1 - \dfrac{2\sigma_y^2}{n-1}\right)^{-\left(\frac{n-1}{2}\right)}\right] \exp\left(\dfrac{2-n}{n}\,\sigma_y^2\right)}{\exp(\sigma_y^2) - 1}$	(22)
$\hat{\mu}_y = \bar{y}$	Estimate of x_p	$\exp(\bar{y} + v_y z_p)$	(24)
and			
$\hat{\sigma}_y^2 = v_y^2$	Bias of $\hat{x}_p$	$\approx \exp\left\{\dfrac{\sigma_y^2}{2n}\right\} \exp\left\{\left[\sqrt{1 - \dfrac{3}{2n} + \dfrac{1}{8n^2}} - 1\right] z_p \sigma_y\right\} \exp\left\{\dfrac{z_p^2 \sigma_y^2 (4n - 1)}{16n^2}\right\}$	(29)

TABLE 2

Comparison of Monte Carlo simulation with theoretical results

	$\hat{\mu}_x$		$\hat{\sigma}_x^2$		$\hat{x}_{0.01}$		$\hat{x}_{0.50}$		$\hat{x}_{0.99}$	
	Sim.	Theory	Sim.	Theory	Sim.	Theory	Sim.	Theory	Sim.	Theory
$\sigma_y^2 = 0.25$										
$n = 10$	3.084	3.085	2.68	2.69	0.972	0.971	2.753	2.752	8.32	8.33
$n = 25$	3.083	3.082	2.70	2.70	0.894	0.895	2.732	2.732	8.57	8.57
$n = 50$	3.082	3.081	2.70	2.70	0.871	0.871	2.726	2.725	8.64	8.64
Population	3.080		2.69		0.848		2.718		8.72	
$\sigma_y^2 = 0.50$										
$n = 10$	3.51	3.51	9.2	9.2	0.647	0.651	2.79	2.79	13.7	13.6
$n = 25$	3.50	3.50	8.38*	8.37	0.571	0.571	2.75	2.75	13.9	13.9
$n = 50$	3.49	3.49	8.14	8.13	0.547	0.547	2.73	2.73	14.0	14.0
Population	3.49		7.90		0.523		2.72		14.1	
$\sigma_y^2 = 0.75$										
$n = 10$	4.01	4.01	24.7*	24.9	0.483	0.484	2.82	2.82	20.1	20.1
$n = 25$	3.98	3.98	19.9	19.9	0.406	0.406	2.76	2.76	20.4	20.3
$n = 50$	3.97	3.97	18.6	18.6	0.383	0.383	2.74	2.74	20.4	20.4
Population	3.96		17.5		0.361		2.72		20.4	
$\sigma_y^2 = 1.00$										
$n = 10$	4.59	4.59	67.0*	65.0	0.378	0.380	2.86	2.86	28.1	28.0
$n = 25$	4.53	4.53	43.2*	43.1	0.305	0.306	2.77	2.77	28.1	28.0
$n = 50$	4.51	4.50	38.5	38.4	0.285	0.284	2.75	2.75	28.0	28.0
Population	4.48		34.5		0.264		2.72		27.9	

Notes: Columns 1, 3, 5, 7 and 9 are averages based on 25 000 replicates with $\mu_y = 1.00$.
Approximate 90% confidence intervals for columns 1, 3, 5, 7 and 9 are less than ± 2 in least significant digit reported except for entries marked (*) for which 90% confidence intervals are less than ± 5 in least significant digit reported.
Columns 2 and 4 are based on eqns. (13) and (16), respectively.
Columns 6, 8 and 10 are based on eqn. (28).
Population mean, variance and quantiles are based on eqns. (3), (4) and (23), respectively.

estimates of the Pearson type 3 distribution are not well known. In fact, many studies have shown that the distributions of these parameter estimates are biased, bounded, and skewed (Matalas and Wallis, 1973; Kirby, 1974; Wallis et al., 1974; Nozdryn-Plotnicki and Watt, 1979; Lall and Beard, 1982). To add to these difficulties, these parameter estimates are not necessarily independent. For example, Ashkar and Bobée (1987) have reported positive covariances between the moments of the logarithms of a sample drawn from a log Pearson type 3 distribution.

Not only do these biased and bounded parameter estimates make an analytical solution for log transform bias difficult to derive, they possibly contribute to bias in the estimates of the arithmetic parameters of a fitted distribution as well. The results derived for the two parameter lognormal distribution suggest that estimates of the arithmetic moments and quantiles of a log Pearson type 3 distribution fitted in log space would also be biased. This could account for many of the problems encountered when applying the log Pearson type 3 distribution frequently reported in the literature (e.g. Wallis and Wood, 1985).

CONCLUSIONS

It is commonly assumed that fitting a frequency distribution to the logarithms of a series has no substantial effect on the fit of that distribution to the original data. This assumption has been shown to be ill-founded. The log space parameters estimated from the sample data are random variables and these estimates typically have large variances or standard errors. When arithmetic space parameters are estimated by an inverse transformation (antilog) of a function of the log space parameter estimates, the arithmetic space parameters are estimated as functions of functions of random variables. Algebraic manipulations which are valid for parameters of theoretical distributions, which are constants, are not necessarily valid for parameter estimates from sample data, which are random variables, (i.e. in general $E[g(x)] \neq g(E[x])$. The variance of log space parameter estimates combined with the inverse transformation lead to bias in the estimates of the arithmetic moments of the series. Since the moments of a distribution describe the shape of that distribution, this bias in moment estimates suggest that distributions fitted in log space may not fit the arithmetic series well.

Analytical expressions for the bias of the mean, variance and quantile estimates for the two parameter lognormal distribution have been derived. These analytical expressions show that the bias is a function of the sample size and the population parameter values. Since the population parameter values are generally unknown, the magnitude of the bias is difficult to determine a priori. For some distributions, such as the log Pearson type 3, analytical

expressions for the magnitude of the bias may be difficult or impossible to determine. This observation should be heeded by practitioners of flood frequency analysis and dictate that caution be exercised in the application of procedures recommended by the USWRC (1981) and others.

ACKNOWLEDGMENT

The research on which this paper is based was supported in part by the Natural Sciences and Engineering Research Council of Canada. This financial support is gratefully acknowledged.

REFERENCES

Aitchison, J. and Brown, J.A.C., 1966. The Lognormal Distribution. Cambridge University Press, London.

Ashkar, F. and Bobeé B., 1987. The generalized method of moments as applied to problems of flood frequency analysis: some practical results for the log-Pearson type 3 distribution. J. Hydrol., 90: 199–217.

Benson, M.A., 1968. Uniform flood-frequency estimating methods for federal agencies. Water Resour. Res., 4(5): 891–908.

Blake, I.F., 1979. An Introduction to Applied Probability. John Wiley, New York, p. 146.

Bobeé B. and Robitaille, R., 1977. The use of the Pearson type 3 and log Pearson type 3 distributions revisited. Water Resour. Res., 13(2); 427–443.

Freeze, R.A., 1975. A stochastic-conceptual analysis of one-dimensional groundwater flow in a nonuniform homogeneous media. Water Resourc. Res., 11(5) 725–741.

Guttman, I., Wilks, S.S. and Hunter, J.S., 1982. Introductory Engineering Statistics. John Wiley, New York, 3rd ed., p. 521.

Hazen, A., 1914. Discussion on 'Flood Flows' by W.E. Fuller. Trans Am. Soc. Civil Eng., 77: 626–632.

Isaacson, M. and MacKenzie, N.G., 1981. Long term distribution of ocean waves: A review. J. Waterway, Port, Coastal and Ocean Div., ASCE, 107: 93–109.

Kendall, M.G. and Stuart, C., 1961. The Advanced Theory of Statistics. Vol. 2. Charles Griffin, London, p. 74.

Kirby, W., 1974. Algebraic Boundedness of Sample Statistics. Water Resour. Res., 10(2): 220–222.

Koch, R.W. and Smillie, G.M., 1986. Bias in hydrologic prediction using log-transformed regression models. Water Resour. Bull., 22(5): 717–723.

Lall, U. and Beard, L.R., 1982. Estimation of Pearson type 3 moments. Water Resour. Res., 18(5): 1563–1569.

Matalas, N.C. and Wallis, J.R., 1973. Eureka! It fits a Pearson type 3 distribution. Water Resour. Res., 9(2): 281–289.

Nozdryn-Plotnicki, M.J. and Watt, W.E., 1979. Assessment of fitting techniques for the log Pearson type 3 distribution using Monte Carlo simulation. Water Resour. Res., 15(3): 714–718.

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

Stedinger, J.R., 1983. Confidence intervals for design events. J. Hydraul. Eng., ASCE, 109(1): 13–27.

Student (W.S. Gosset), 1908. The probable error of a mean. Biometrika, 6(1): 1–25.

U.S. Water Resources Council. Guidelines for determining flood flow frequency. Hydrol. Comm., Bull. 17B, Washington, DC, 1981.

Wallis, J.R., Matalas, N.C. and Slack, J.R., 1974. Just a moment! Water Resour. Res., 10(2): 211–219.

Wallis, J.R. and Wood, E.F., 1985. Relative accuracy of log Pearson 3 procedures. J. Hydraul. Eng., ASCE, 111(7): 1042–1056.

Wilson, B.G., 1988. A critical review of the method of moments in hydrology. M.A.Sc. Thesis, Dep. Civ. Eng., University of Toronto.