Eureka! It Fits a Pearson Type 3 Distribution

N. C. MATALAS

U.S. Geological Survey, Washington, D. C. 20242

J. R. WALLIS

IBM Thomas J. Watson Research Center, Yorktown Heights, New York 10598

Under the assumption that a random variable is distributed as Pearson type 3, a comparison was made between moment and maximum likelihood estimates of the parameter values of the distribution and the variate values at specified probability levels. For the region where maximum likelihood solutions may be obtained, maximum likelihood estimates yield solutions that are less biased and less variable than the comparable moment estimates. When these results are extended to quite small samples, say, $N \approx 25$, they become quite pronounced as the probability p becomes greater than $N/(N + 1)$.

Streamflow sequences represent samples drawn from a population whose distribution function is apparently skewed. For those concerned with the design of hydraulic structures for flood control or low-flow augmentation, the distribution of the flows may be quite important. Within the design process it may be necessary to estimate the probabilities of certain events, such as the 100-year flood or the 10-year, 7-day low flow. In some cases, even rarer events may be of concern. Unfortunately, with small samples from a skewed distribution there is seemingly a one-to-one correspondence between the number of hydrologists and the number of different 'best engineering judgments' of these probability estimates.

To estimate the probabilities of hydrologic events, one generally assumes that (1) the observed flows are independently distributed in time, (2) the flows follow a specified distribution function, (3) the estimates of the parameter values of the distribution are unbiased, (4) the sample size is large enough to warrant estimation of the parameter values, and (5) no operational decisions, for example how to treat zero flows, markedly influence the results. Although the validity of each of these assumptions is questionable, seldom have probability estimates been subjected to sensitivity analysis to assess the effect of the assumptions on the results.

Streamflow sequences tend to be persistent; that is, high flows tend to follow high flows, and low flows tend to follow low flows. For persistent sequences the assumption of temporal independence does not hold. Although a specific distribution can be selected to represent the underlying distribution of flows, no a priori knowledge exists to support the assumption that the specified distribution is indeed the underlying distribution. Classical statistical tests of goodness of fit are not powerful enough to discriminate among reasonable choices of distributions, and, more often than not, by default the choice is made by fiat. Perhaps the best-known example of choice by fiat is that of the work group on flow frequency methods of the Water Resources Council, reported on by *Benson* [1968], which concluded that 'The log Pearson type 3 distribution has been selected as the base method, with provisions for departures from the base method where justified.'

Because of the extensive use of the Pearson type 3 distribution in hydrologic studies, an investigation of its usefulness was made under the assumption of temporal independence in reference to the third and fourth assumptions above.

PEARSON TYPE 3 DISTRIBUTION

The Pearson type 3 density function may be expressed as

$$f[X: m, a, b] = \frac{1}{|a|\Gamma(b+1)}\left(\frac{X-m}{a}\right)^b \cdot \exp\left[-\left(\frac{X-m}{a}\right)\right] \quad (1)$$

where m, a, and b are parameters. If the distribution is positively skewed, a is positive, and $X \geq m$; otherwise, a is negative, and $X \leq m$. In the special case for which $m = 0$ the Pearson type 3 distribution is referred to as the gamma distribution. The parameters m, a, and b are related to the first four moments of the random variable X as follows:

$$\mu = m + a(b+1) \quad (2)$$

$$\sigma = |a|(b+1)^{1/2} \quad (3)$$

$$\gamma = 2(b+1)^{-1/2} \quad (4)$$

$$\lambda = 6(b+1)^{-1} + 3 \quad (5)$$

where μ, σ, γ, and λ denote the mean, standard deviation, coefficient of skewness, and coefficient of kurtosis of X. Since the standard deviation is positive, by definition the absolute value of a is used to define σ in terms of a and b. As can be seen from (3) and (4), $b \geq -1$.

From an observed sequence of flows of length N, $X(1), \cdots, X(N)$, moment estimates of μ, σ, and γ, denoted as μ', σ', and γ', may be obtained. By use of the estimates μ', σ', and γ' the estimates of the parameters of the distribution m', a', and b' may be derived from (2), (3), and (4). The estimates m', a', and b', referred to as moment estimates, are used extensively in fitting the Pearson type 3 distribution to the observed flows. Estimates of m, a, and b could be derived by using the estimate λ' and, say, μ' and σ'. However, this is seldom done because λ' is subject to larger sampling errors than μ', σ', and γ' are.

Another set of estimates, referred to as maximum likelihood estimates, are given as follows. The likelihood function $L[X]$ is defined as

$$L[X] = \sum_{i=1}^{N} \ln f[X_i] \quad (6)$$

where $f[X_i]$ denotes the value of $f[X]$ for $X = X_i$. The maximum likelihood estimates m'', a'', and b'' are given by the solution of the following three simultaneous nonlinear equations:

$$\frac{\partial L}{\partial m} = -b'' \sum_{i=1}^{N}\left(\frac{1}{(X_i - m'')}\right) + \frac{N}{a''} = 0 \quad (7)$$

$$\frac{\partial L}{\partial a} = -\frac{N}{a''}(b''+1) + \frac{1}{(a'')^2}\sum_{i=1}^{N}(X_i - m'') = 0 \quad (8)$$

$$\frac{\partial L}{\partial b} = -N\frac{d \ln \Gamma(b''+1)}{db''} + \sum_{i=1}^{N} \ln\left(\frac{X_i - m''}{a''}\right) = 0 \quad (9)$$

From (9) it is seen that the quantity $[(X_i - m'')/a''] \geq 0$ for all i. If the distribution is negatively skewed, a'' is negative, and $(X_i - m'') \leq 0$ for all i. On the other hand, if the distribution is positively skewed, a'' is positive, and $(X_i - m'') \geq 0$ for all i. Thus from (7) it is seen that b'' is constrained in such a way that $b'' \geq 0$, in which case $|\gamma''| \leq 2$. If $|\gamma| \geq 2$, the distribution is exponential with $P[X] = \infty$ at $X = m$. If the underlying distribution has $|\gamma| > 2$ and therefore $b < 0$, maximum likelihood solutions for γ from sample sequences will yield biased estimates of $|\gamma|$, since no sample sequence will yield a value $|\gamma''| > 2$.

A half century or so ago considerable argument existed concerning the relative merits of moment and maximum likelihood estimates. One of the colorful chapters in the history of statistical theory was the arguments of Pearson, an advocate of moment estimates, and Fisher, an advocate of maximum likelihood estimates. *Fisher* [1922] demonstrated that maximum likelihood estimates are asymptotically more efficient than moment estimates. Over the years much of the argument has vanished with the general acceptance that, at least from the point of view of efficiency, maximum likelihood estimates are better than moment estimates. Maximum likelihood estimates have not become dominant in hydrology, although a cursory search of the literature produced a few instances where such estimates have been used for fitting

distribution functions to streamflow data [*Matalas*, 1963; *Markovic*, 1965; *Domokos and Szasz*, 1968].

Perhaps the principal reason for the reluctance to use maximum likelihood estimates is the difficulty in solving the maximum likelihood equations. Also, a very limited literature exists on the statistical sampling properties of maximum likelihood estimates, particularly for small samples. Maximum likelihood estimates are not as mathematically tractable as moment estimates and are seemingly more difficult to use in operational programs such as the generation of synthetic streamflow sequences. With the general availability of digital computers, solutions of the maximum likelihood equations can be obtained, and, moreover, Monte Carlo experiments can be conducted to assess the statistical properties of the maximum likelihood estimates.

EXPERIMENTAL DESIGN

To assess the relative goodness of moment and maximum likelihood estimates, Pearsonian numbers were generated for various values of N, and $m = 1000$, $a = 400$, and $b = -0.5$ to 399. However, only the results for $b = 3$ are presented in this paper. On the basis of various sets of size N, the means and standard deviations of the moment and maximum likelihood estimates of m, a, and b were determined. The Pearson distribution was fitted to the data for the various sets, whereby estimates of the 'flows' at probability levels 0.96, 0.98, 0.99, and 0.999 were determined. The means and standard deviations for these estimates were determined and served as a basis for assessing the relative goodness of the two types of estimates.

In practice, one may wish to fit the Pearson distribution not to the flows but to the logarithms of the flows, the resulting fit being referred to as the log Pearson distribution. To assess the effect of the logarithmic transform, Pearsonian numbers were logarithmically transformed and then fitted with a Pearson distribution. In these experiments the 'world' was known to be Pearson, whereas in practice the world is unknown. Nonetheless, these experiments afford an opportunity to assess partially the effect of the logarithmic transform on the estimated moments of the flows.

To generate Pearsonian numbers, the following procedure was used. Let Y denote a random variable normally distributed with zero mean and unit variance. The random variable Z, defined as

$$Z = \sum_{i=1}^{n} Y^2(i) \tag{10}$$

is distributed as gamma (in effect χ^2) with

$$\mu(Z) = n \tag{11}$$

$$\sigma^2(Z) = 2n \tag{12}$$

$$\gamma(Z) = 2(2/n)^{1/2} \tag{13}$$

where $\mu(Z)$, $\sigma^2(Z)$, and $\gamma^2(Z)$ denote the mean, variance, and coefficient of skewness of z. The random variable X, defined as

$$X = \mu(X) + \sigma(X)[(2n)^{-1/2}Z - (n/2)^{1/2}] \tag{14}$$

is distributed as gamma with mean $\mu(X)$, standard deviation $\sigma(X)$, and coefficient of skewness

$$\gamma(X) = (8/n)^{1/2} \tag{15}$$

In terms of the parameters m, a, and b (2, 3, and 4),

$$X = m + a(b + 1) + |a|(b + 1)^{1/2}[(2n)^{-1/2}Z - (n/2)^{1/2}] \tag{16}$$

is distributed as Pearson type 3, the parameters m, a, and b being related to $\mu(X)$, $\sigma(X)$, and $\gamma(X)$ as defined by (2)–(4).

For the various experiments reported on below, sample sequences of varying length N were generated on the basis of a Pearson type 3 distribution with $m = 1000$, $a = 400$, and $b = 3$, where $\mu(X) = 2600$, $\sigma(X) = 800$, and $\gamma(X) = 1$. Generation was carried out by means of (16), where, by (13), $n = 8$. Given a sequence of length N, moment and maximum likelihood estimates of m, a, and b were obtained. Let m^*, a^*, and b^* denote one of the other types of estimates. The flow $X^*(p)$ corresponding to $F[X \leq X^*(p)] = p$ was determined as follows.

Assume a^* to be positive, that is, the distribution to be positively skewed. Then

$$F[X \leq X^*(p)] = p = \int_{m^*}^{X^*(p)} f(X: m^*, a^*, b^*]\, dX \tag{17}$$

where $f[K: m^*, a^*, b^*]$ is defined by (1). For a^* negative, the distribution negatively skewed,

$$F[X \leq X^*(p)] = p$$

$$= 1 - \int_{-m^*}^{-X^*(p)} f[Y: -m^*, -a^*, b^*]\, dY \quad (18)$$

where $Y = -X$. Methods of numerical integration were used to carry out the solutions of (17) and (18), the incomplete gamma functions, for $p = 0.96$, 0.98, 0.99, and 0.999.

To test the Pearsonian number generation technique, 10 sets, each of 50,000 numbers, were generated, and by the method of moments the estimates of $\mu(X)$, $\sigma(X)$, $\gamma(X)$, and $\lambda(X)$ were derived and compared with the values $\mu(X) = 2600$, $\sigma(X) = 800$, $\gamma(X) = 1$, and $\lambda(X) = 4.5$. The test results are given in Table 1. For these Pearsonian numbers the moment estimates of m, a, and b are given in Table 2.

The test results indicate that the generated numbers exhibit the characteristics of Pearsonian numbers quite well, at least in terms of the moments of the numbers and the parameter values of the distribution. In effect, the results indicate the goodness of the random number generator. A well-tested uniform random number generator [*Lewis et al.*, 1969] and a carefully programed Box-Muller transform [*Box and Muller*, 1958] were used to obtain normal numbers, which were transformed into Pearsonian numbers by the technique described above. Even closer agreement between the sample and the expected values could have been achieved by increasing the size of the sets. In the subsequent experiments, samples of size $N = 25$, 50, 100, 250, 500, 1000, and 50,000 were used.

TABLE 1. Sample Moments for 10 Sets of 50,000 Pearsonian Numbers Each

Set	$\mu'(X)$	$\sigma'(X)$	$\gamma'(X)$	$\lambda'(X)$
1	2596.45	650,421	1.03300	4.68076
2	2600.13	639,682	0.99757	4.50857
3	2609.00	639,268	0.97808	4.35183
4	2594.74	633,926	1.02993	4.71594
5	2598.34	642,871	0.97640	4.34393
6	2598.13	638,229	0.99789	4.44309
7	2596.37	639,059	1.00767	4.51248
8	2598.87	636.362	1.02742	4.66957
9	2590.07	625,381	0.98851	4.45352
10	2607.91	659,137	1.01129	4.46216
Average	2599.00	640,433	1.00478	4.51419
Theoretic value	2600.	640,000	1.00000	4.50000

The corresponding moment estimates of m, a, and b are given in Table 2.

TABLE 2. Moment Estimates m', a', and b' for 10 Sets of 50,000 Pearsonian Numbers Each

Set	Observed Minimum Value	m'	a'	b'
1	1018	1035*	416.6	2.748
2	1059	997	398.9	3.019
3	1051	974	391.0	3.181
4	1030	1049*	410.0	2.771
5	1070	956	391.4	3.195
6	1026	997	398.6	3.017
7	1023	1010	402.8	2.939
8	1067	1046	390.9	2.789
9	1043	990	390.9	3.093
10	1055	1002	410.5	2.911
Mean		1005.6	402.1	2.966
Standard deviation		30.23	9 36	0.163
Theoretic value		1000	400	3.000

* The value of m' is larger than the minimum value in the particular sequence of 50,000 flows.

As was noted above, for a Pearson type 3 distribution $b \geq -1$. However, if $b \leq 0$ (that is, $|\gamma| \geq 2$), the method of maximum likelihood leads to values of $b'' \geq 0$, whereas the method of moments admits values of $b' \leq 0$. When the incomplete gamma functions (17 and 18) are being solved for values of $X^*(p)$, it is extremely difficult to obtain solutions if $b^* < -0.75$ [*Pearson*, 1951], and available computer algorithms exclude this region from the solution set. This computational constraint does not affect the solution by the method of maximum likelihood when negative values of b'' are analytically excluded by the likelihood equations themselves. Although the method of moments admits values of $b' < -0.75$, solutions of corresponding values of $X'(p)$ could not be attained.

MOMENT ESTIMATES

Although the method of moments gives good estimates, small biases, and standard deviations

TABLE 3. Mean and Range of Moment Estimates of m, a, and b Based on 500 Sets of Size N

	Mean Values			Largest Observed Values			Smallest Observed Values		
N	m'	a'	b'	m'	a'	b'	m'	a'	b'
25	4012	333	6654	$>10^6$	1524	$>10^6$	−182,787	0.4	−0.61
50	330	359	141	121,431	1356	35,414	−101,542	3.4	−0.55
100	672	379	8.5	1902	1135	1393	−20,582	16.5	−0.34
250	881	400	4.0	1724	879	19.7	−691	159	0.07
500	953	404	3.4	1649	730	11.8	−19	201	0.36
1000	977	402	3.2	1989	695	7.2	393	267	0.45

of the moments and parameter values of the distribution for very large sample sizes, it fails to do so for sample sizes near those encountered in hydrology, say, $N \leq 50$. For small sample sizes the estimates m', a', and b' are highly biased and extremely variable, as can be seen from Table 3. For other values of b ranging from $b = -0.5$ to $b = 399$ the results, although they are not presented, were found to be similarly highly biased and variable.

For many hydrologic problems the individual values m', a', and b' for a particular flow sequence are intrinsically important only in that they are used to derive estimates of flows at particular probability levels. However, m has some physical interpretation for streamflow. For a positively skewed distribution, m represents the lower bound of admissible flows and must therefore be positive. A negatively skewed distribution, where m denotes the upper bound on admissible flow, is not physically meaningful, for this implies that negative flows attain and that there is a 'maximum certain flood.' For additional properties of the negatively skewed Pearson type 3 distribution see *Gilroy* [1972]. Given positive values of skewness and m, finite samples may yield flows for which the fitted Pearson type 3 distribution is physically not realizable. Nonetheless, the fitted distributions may yield 'reasonable' estimates of $X(p)$.

The probability levels 0.96, 0.98, 0.99, and 0.999 were considered for various values of N. The procedure outlined above for determining the value of $X'(p)$ for which $F[X \leq X'(p)] = p$ was checked by using a χ^2 table [*Pearson*, 1934] for seven exact entries, and errors were found to be $<1\%$.

The results for 400 sets each for $N = 25$, 50, 100, 250, 500, and 1000 and for 10 sets of $N = 50{,}000$ are given in Table 4.

Although the estimates m', a', and b' are highly biased and variable, on the average, reasonably good estimates of the theoretic values of $X'(p)$ may be derived by the method of moments for $N \geq 50$. For a particular sequence, however, $X'(p)$ is subject to large sampling errors, particularly for small N, say, <50. Estimates of $X(p)$ agree rather closely

TABLE 4. Mean and Standard Deviation of Values of $X'(p)$ Corresponding to Probability Level p and Sample Size N

	$p = 0.96$		$p = 0.98$		$p = 0.99$		$p = 0.999$	
N	Mean	s.d.	Mean	s.d.	Mean	s.d.	Mean	s.d.
25	4138	419	4503	541	4853	675	5944	1158
50	4207	312	4595	401	4968	497	6135	847
100	4216	248	4609	326	4988	411	6174	717
250	4229	156	4626	204	5008	256	6210	445
500	4233	113	4634	150	5018	191	6227	336
1,000	4240	80	4642	107	5030	136	6246	240
50,000	4220	50	4618	56	5002	62	6206	82
Theoretic value	4234		4634		5018		6225	

with their respective theoretic values for $p \leq N/(N+1)$, where N is the length of the observed sequence. Both the bias and the standard deviation of the estimates of $X(p)$ increase as p becomes greater than $N/(N+1)$. Although additional computer experiments are not presented in this report, they have shown that, as γ increases with very small N, say, <25, the individual estimates of $X'(p)$ increase in bias and become even more variable.

LOG TRANSFORM

The work group on flow frequency methods of the Water Resources Council [*Benson*, 1968] recommended that sequences of flows be logarithmically transformed and then fitted with a Pearson type 3 distribution. The justification for the log Pearson distribution was apparently threefold. First, it was in common use by some federal agencies. Second, the log Pearson distribution has a parameter relating to skewness and is therefore more flexible for curve fitting than the log normal distribution. Third, if the skewness of the logs of the flows were 0, the log Pearson distribution could be regarded as a special case of the log normal distribution in that a Pearson distribution for which the skewness is 0 is the normal distribution, and the log normal distribution has a long history of use in hydrology.

Ten sets of 50,000 Pearsonian numbers were generated with $m = 1000$, $a = 400$, and $b = 3$. For each set the log transform was applied, and, with the transformed numbers, estimates of their moments and values of m, a, and b were derived by the method of moments. These estimates are given in Table 5.

By means of the log transform the Pearsonian numbers are transformed to approximately normal numbers in terms of skewness and kurtosis. The Pearsonian numbers were characterized by a skewness equal to 1 and kurtosis equal to 4.5, whereas the log transform of the numbers yielded a skewness and kurtosis of about 0.16 and 2.76, respectively.

The log Pearson estimates of $X(p)$ for probability level p and sample size N are given in Table 6, where the estimates are based on 10 sets for $N = 50{,}000$ and 400 sets for the other values of N.

The estimates of $X(p)$ were derived as follows. For a given sequence the 'observations' were transformed into log space and by the method of moments fitted by a Pearson type 3 distribution. In log space, estimates of the flow at probability level p were determined as is described above. Values in real space were obtained by exponentiating the log space estimates.

Given that the world is Pearsonian, the assumption that it is log Pearson leads to estimates of $X(p)$ that are biased and quite variable, the bias and variability increasing markedly as p becomes larger than $N/(N+1)$. In addition, other tests not reported here showed that for values of $p > N/(N+1)$ the variability in $X'(p)$ increased with increasing skewness in the untransformed data. A comparison of

TABLE 5. Sample Estimates by the Method of Moments for Logarithmically Transformed Pearsonian Numbers

Set	μ'	σ'	γ'	λ'	m'	a'	b'
1	3.396	0.0165	0.1574	2.769	1.765	0.0101	160.4
2	3.395	0.0167	0.1688	2.767	1.862	0.0109	139.4
3	3.396	0.0165	0.1471	2.761	1.647	0.0095	183.9
4	3.397	0.0164	0.1540	2.729	1.733	0.0099	167.7
5	3.395	0.0164	0.1624	2.795	1.820	0.0104	150.6
6	3.395	0.0167	0.1465	2.729	1.633	0.0095	185.4
7	3.395	0.0165	0.1602	2.754	1.793	0.0103	154.9
8	3.395	0.0165	0.1626	2.754	1.815	0.0104	150.3
9	3.396	0.0163	0.1769	2.756	1.951	0.0113	126.8
10	3.394	0.0163	0.1501	2.757	1.693	0.0096	176.6
Mean	3.395	0.0165	0.1586	2.757	1.771	0.0102	159.6
Standard deviation	0.001	0.0001	0.0097	0.019	0.099	0.0006	19.1

TABLE 6. Mean and Standard Deviation of Log Pearson Values of $X(p)$ Corresponding to Probability Level p and Sample Size N

	$p = 0.96$		$p = 0.98$		$p = 0.99$		$p = 0.999$	
N	Mean	s.d.	Mean	s.d.	Mean	s.d.	Mean	s.d.
25	4193	414	4627	566	5070	757	6661	1715
50	4245	308	4693	409	5147	531	6739	1103
100	4230	316	4672	390	5119	481	6673	907
250	4245	144	4688	189	5134	244	6673	494
500	4243	103	4687	138	5132	178	6666	364
1,000	4246	72	4688	96	5131	124	6656	255
50,000	4260	54	4702	59	5147	62	6666	82
Theoretic value	4234		4634		5018		6225	

Tables 4 and 6 shows that for estimates of $X(p)$, where $p \leq N/(N + 1)$, little is gained by the log transformation. The log transform yields estimates of $X(p)$ that are slightly less biased and variable than those derived from the untransformed data themselves, whereas the opposite occurs as p becomes larger than $N/(N + 1)$.

As p becomes much greater than $N/(N + 1)$, the transform-derived estimates of $X(p)$ relative to the estimate derived from the untransformed data continue to increase in bias and variability.

In practice, the real world is unknown, so that a complete assessment of the goodness of estimates of $X(p)$ derived under the assumption of log Pearson is not easy to make. Log Pearson numbers could have been generated, whereby the sampling properties of their moments and parameter values could have been investigated. Although this investigation would be a useful exercise, it would not by itself suffice to assess the errors likely to be incurred in the real world if streamflow is not log Pearson distributed. The above experiment indicates that the errors could be large, particularly for small N. A more detailed assessment of these errors could be pursued within a statistical decision theoretic framework.

MAXIMUM LIKELIHOOD ESTIMATES

Solution of (7), (8), and (9) yields the maximum likelihood estimates of m, a, and b. With these estimates and (2), (3), and (4) the maximum likelihood estimates of μ, σ, and γ may be obtained. Unfortunately, the calculations involved in deriving the maximum likelihood estimates are not trivial. They entail a time-consuming iterative solution of three nonlinear simultaneous equations in three unknowns, and a completely general iterative fast and accurate computer solution was difficult to obtain. In particular, with small samples the response surface occasionally may be inordinately flat, and, if a sample exhibits a very small absolute value of skewness, a solution may not be possible because of computer-imposed constraints. A further constraint is imposed by the value of b itself. As was noted above, a maximum likelihood solution is not possible if $b < 0$.

Given a value of m assigned a priori, a mathematical solution of the simultaneous equations (8) and (9) exists [*Greenwood and Durand*, 1960]. However, as was noted above, the solution may not be feasible in terms of computational time and cost. A better procedure, suggested by Robin T. Clarke of the Institute of Hydrology, Wallingford, England (personal communication, 1972), is to assume m and solve (7) and (8) explicitly for b' and a':

$$b' = \left[N^2 \Big/ \sum_{i=1}^{N} (X_i - m'')\right] \cdot \left[\sum_{i=1}^{N} \left(\frac{1}{X_i - m''}\right) - N^2 \Big/ \sum_{i=1}^{N} (X_i - m)\right] \tag{19}$$

$$a' = \sum_{i=1}^{N} (X_i - m'') \Big/ \left[N(b' + 1)\right] \tag{20}$$

estimates a'', b'', and m'' being substituted into (9) to yield a value of $\partial b/\partial b = R$. If $|R| >$

10^{-6}, the procedure was repeated with a new assigned value for m. If after $k = 30$ iterations $|R| > 10^{-6}$, the solution effort was considered a failure. Failures of course can be reduced by increasing the $|R|$ failure criterion or by adopting some exogenous rule for assigning a value to m. This, however, was not done, since we wished to make our comparisons by using accurate parameter estimates (parameter value estimates resolved to 5 decimal places) rather than computational speed. Failure samples were discarded. Only the nonfailure samples were used in assessing the statistical properties of m'', a'', and b''.

Mean values of the maximum likelihood estimates m'', a'', and b'' based on 400 sets of size $N \leq 1000$ and 10 sets of size $N = 50{,}000$ are given in Table 7. Table 8 gives means and standard deviations of $X''(p)$ based on 50 sets of size $N \leq 1000$ and 10 sets of size $N = 50{,}000$.

A comparison of the results given in Tables 3 and 4 with those given in Tables 7 and 8 emphasizes the goodness of maximum likelihood estimates relative to moment estimates in terms of bias and variability. For large N both methods yield good estimates of m, a, b, and $X(p)$. However, for small N the maximum likelihood estimates have both smaller bias and smaller standard deviation than the moment estimates.

For values of $p \leq N/(N + 1)$ the moment estimates of $X(p)$ have somewhat larger bias and variability than the corresponding maximum likelihood estimates. As p becomes larger than $N/(N + 1)$, the bias and variability in the moment estimates of $X(p)$ become much larger than those in the maximum likelihood. From a practical point of view, maximum likelihood offers no great advantage in estimating $X(p)$ for $p \leq N/(N + 1)$. However, for $p > N/(N + 1)$ the bias and variability in moment estimates of $X(p)$ are large enough to warrant consideration of using maximum likelihood estimates.

TABLE 7. Mean of Maximum Likelihood Estimates Based on 400 Sets of Size $N \leq 1000$ and 10 Sets of Size $N = 50{,}000$

N	m''	a''	b''
25	1025	414	3.53
50	1033	425	3.40
100	1005	413	3.26
250	996	407	3.12
500	1018	411	2.93
1,000	1006	403	2.95
50,000	1000	400	3.00
Theoretic value	1000	400	3.00

It should be noted that the estimate of $X(p)$ is based on the estimates of μ, σ, and γ. The biases in the estimates of σ and γ contribute to the bias in the estimate of $X(p)$. Corrections for the biases in the moment estimates of σ and γ require the underlying distribution function as well as the values of σ and γ to be known. The method of maximum likelihood in effect makes these corrections in relation to the assumed underlying distribution function.

SUMMARY AND CONCLUSIONS

Although in terms of asymptotic efficiency the relative goodness of maximum likelihood

TABLE 8. Mean and Standard Deviation of Maximum Likelihood Estimates of $X(p)$ Corresponding to Probability Level p and Sample Size N

	$p = 0.96$		$p = 0.98$		$p = 0.99$		$p = 0.999$	
N	Mean	s.d.	Mean	s.d.	Mean	s.d.	Mean	s.d.
25	4177	422	4590	512	4989	602	6256	903
50	4242	277	4651	330	5046	387	6292	586
100	4226	250	4630	310	5018	372	6241	587
250	4220	147	4621	187	5008	227	6225	371
500	4249	110	4654	132	5043	156	6269	240
1,000	4242	74	4646	89	5035	105	6258	163
50,000	4233	17	4632	21	5010	25	6225	35
Theoretic value	4234		4634		5018		6225	

estimates over moment estimates has been known for nearly a half century, little use has been made of maximum likelihood estimates in hydrology. The reason perhaps for this neglect has been the formidable computations needed to obtain maximum likelihood estimates. Although the computation task remains, the advent of digital computers has made the solution task feasible.

The above comparisons between moment and maximum likelihood estimates were made on the basis of the Pearson type 3 distribution. If one is interested in only the estimates of μ, σ, and γ, there might be some advantage in using moments rather than maximum likelihood estimators. For moment estimates of μ, σ, and γ one need not make any assumption about the underlying distribution function. Often principal interest is not in the values of μ, σ, and γ per se but in the values of $X(p)$ for various probability levels or in the estimate of the lower bound m, and the values of μ, σ, and γ are of interest only insofar as they are the basis for mathematical estimates of $X(p)$ or m. For such estimates, either by the method of moments or the method of maximum likelihood, an assumption concerning the underlying distribution function must be made.

In many cases one is interested in extrapolating beyond the limits of the sample. For example, with a 25-year record of annual flood peaks one may wish to estimate the 50- or 100-year flood peaks. By the method of moments, estimates of a flood peak $X(p)$ with return period T (equal to $1/p$) will be biased. The bias for the estimates will decrease as the record size N increases, whereas the bias for fixed N will increase as T increases. Thus the estimate of the T-year flood is in effect the estimate of the flood when $T' < T$ years. With the method of maximum likelihood, unbiased and minimum variance estimates of the T-year flood with records of size N may be made.

During this study several difficulties related to the properties of the Pearson type 3 distribution were encountered in deriving both moment and maximum likelihood estimates. Although these difficulties may not seriously affect the use of this distribution from the point of view of curve fitting, the difficulties are sufficiently troublesome that using other distributions warrants consideration, particularly if maximum likelihood estimates are of interest.

REFERENCES

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

Box, G. E. P., and M. E. Muller, A note on the generation of random normal deviates, *Ann. Math. Statist.*, *29*, 610–611, 1958.

Domokos, M., and D. Szasz, Generation of fitting probability distribution functions of discharges by electronic computer, *Int. Ass. Sci. Hydrol. Publ. 81*, 535–545, 1968.

Fisher, R. A., On the mathematical foundations of theoretical statistics, *Phil. Trans., A222*, 309–368, 1922.

Gilroy, E. J., The upper bound of a log-Pearson type 3 variable with negatively skewed logarithms, *U.S. Geol. Surv. Prof. Pap. 800-B*, 273–275, 1972.

Greenwood, J. A., and D. Durand, Aids for fitting the gamma distribution by maximum likelihood, *Technometrics*, *2*(1), 55–65, 1960.

Lewis, P. A. W., A. S. Goodman, and J. M. Miller, A pseudorandom number generator for the System/360, *IBM Syst. J.*, *8*(2), 136–146, 1969.

Markovic, R. D., Probability functions of best fit to distributions of annual precipitation and runoff, *Hydrol. Pap. 8*, 33 pp., Colo. State Univ., Fort Collins, 1965.

Matalas, N. C., Probability distribution of low flows, *U.S. Geol. Surv. Prof. Pap. 434-A*, 1–27, 1963.

Pearson, K., *Tables of Incomplete Gamma Function*, Her Majesty's Stationery Office, London, 1934.

Pearson, K., *Tables of the Incomplete Γ-Function*, Cambridge University Press, London, 1951.

(Received October 20, 1972.)