PARAMETER ESTIMATION FOR LOG-PEARSON TYPE III DISTRIBUTION BY POME

By Vijay P. Singh,¹ Member, ASCE and Kulwant Singh²

ABSTRACT: The principle of maximum entropy (POME) is employed to derive an alternative method of parameter estimation for the log-Pearson type (LPT) III distribution. Historical flood data are used to evaluate this method and compare it with the methods of moments (MOM) and maximum likelihood estimation (MLE). The parameter estimates yielded by POME are comparable to those by MOM and MLE.

INTRODUCTION

The log-Pearson type (LPT) III distribution is extensively used in hydrologic frequency analysis (2-4,12,15,16,18,24). Its use was recommended by the working group of the Water Resources Council on flow frequency methods, as reported by Benson (1), which concluded that "The log-Pearson III distribution has been selected as the base method, with provisions for departures from the base method where justified."

The fitting procedure (16), used frequently, consists of the logarithmic transformation of natural data and then fitting the Pearson type (PT) III distribution (3-5,10,18) by the method of moments or fitting the LPT III distribution directly (12,22). The PT III distribution has a lower bound when skew is positive and an upper bound when skew is negative. It enters the region of negative values for some parameter values when skew is positive and has minus infinity as the lower bound when skew is negative. Its form varies from a J shape to a bell shape. An interpretation of these characteristics is helpful for its proper application to flood flows.

Bobee (3) studied mathematical and statistical properties of the LPT III distribution at length. He presented the various forms of its density function and relationships between its parameters and moments, coefficient of variation, and coefficient of skewness. The parameters of the LPT III distribution are estimated using a number of methods (6,13) that produce different estimates and confidence intervals. Condie (6) applied the method of maximum likelihood estimation and illustrated it with data for Canadian rivers. Bobee (2) used the method of moments, whereas Rao (16) used the method of mixed moments.

The objective of this study was to derive an alternative procedure for the estimation of its parameters for the LPT III distribution using the principle of maximum entropy (POME), and then to test the method using annual maximum flood data for a number of rivers. This method of parameter

¹Prof. and Coordinator, Water Resour. Prog., Dept. of Civ. Engrg., Louisiana State Univ., Baton Rouge, LA 70803-6405.

²Grad. Resear. Asst., Dept. of Civ. Engrg., Louisiana State Univ., Baton Rouge, LA 70803.

Note. Discussion open until June 1, 1988. To extend the closing date one month, a written request must be filed with the ASCE Manager of Journals. The manuscript for this paper was submitted for review and possible publication on May 12, 1986. This paper is part of the *Journal of Hydraulic Engineering*, Vol. 114, No. 1, January, 1988. ©ASCE, ISSN 0733-9420/88/0001-0112/1.00 + 1.15 per page. Paper No. 22152.

estimation was also compared with the methods of moments (MOM) and maximum likelihood estimation (MLE).

DERIVATION OF PARAMETER ESTIMATION METHOD

Let $y = \ln x$ where x is a positive random variable. If y has a PT III distribution, then x will have an LPT III distribution with the probability density function given by

where a > 0, b > 0, and $0 < c < \ln x$ are parameters. $\Gamma(\cdot) =$ the gamma function. To derive a method for the estimation of the parameters a, b, and c of Eq. 1, three steps are involved: (1) Specification of appropriate constraints; (2) construction of the partition function or zeroth Lagrange multiplier; and (3) defining the relation between Lagrange multipliers and constraints. Complete mathematical discussion of this method can be found in Jaynes (7), Levine and Tribus (11), Reza (17), Shannon (19), and Tribus (23). Two illustrative applications have been presented by Jowitt (7), Jowitt and Munro (8), Munro and Jowitt (14), and Sonuga (20,21) for extreme-value type I and normal distributions.

Specification of Constraints

Following Jaynes (7) and Tribus (23), constraints appropriate for Eq. 1 can be written as

$$\int_{c_{\star}}^{\infty} f(x)dx = 1 \qquad (2)$$

$$\int_{c_{\star}}^{\infty} \ln x f(x) dx = E(\ln x) = \bar{y} \quad \dots \quad (3)$$

$$\int_{c_{\star}}^{\infty} \ln (\ln x - c) f(x) \, dx = E[\ln (\ln x - c)] \quad \dots \quad \dots \quad \dots \quad \dots \quad (4)$$

in which $E[\cdot]$ denotes expectation of the bracketed quantity; $c_* = \exp(c)$; and $\bar{y} =$ the mean of y. These constraint equations specify the information sufficient for the LPT III distribution. Because this information is determined from data in terms of expectations, the parameters and other statistics of the distribution can be physically interpreted.

Construction of Zeroth Lagrange Multiplier, λ_0

The probability density function f(x) corresponding to POME and consistent with Eqs. 2-4 takes the following form:

where λ_0 , λ_1 , and λ_2 = Lagrange multipliers. The mathematical rationale for Eq. 5 has been presented by Tribus (23). From the total probability

113

c ...

condition in Eq. 1

$$\int_{c_{\star}}^{\infty} f(x) \ dx = \int_{c_{\star}}^{\infty} \exp\left[-\lambda_0 - \lambda_1 \ln x - \lambda_2 \ln \left(\ln x - c\right)\right] dx \dots \dots \dots \dots (6)$$

so that

Consider $\ln x - c = y$. Then $x = \exp(y + c)$, and $dx = \exp(y + c)dy$. Eq. 7 can be written as

$$\exp (\lambda_0) = \exp \left[-c(\lambda_1 - 1)\right] \int_0^\infty \exp \left[-y(\lambda_1 - 1)\right] y^{-\lambda_2} dy \dots (8)$$

Inserting $z = y(\lambda_1 - 1)$ and carrying out the integration yields

$$\exp (\lambda_0) = \frac{\exp \left[-c(\lambda_1-1)\right]}{(\lambda_1-1)^{1-\lambda_2}} \int_0^\infty z^{-\lambda_2} e^{-z} dz \qquad (9a)$$

$$\exp (\lambda_0) = \frac{\exp \left[-c(\lambda_1 - 1)\right]}{(\lambda_1 - 1)^{1 - \lambda_2}} \Gamma(1 - \lambda_2) \dots \dots \dots \dots \dots \dots \dots \dots (9b)$$

The zeroth Lagrange multiplier is given as

$$\lambda_0 = -c(\lambda_1 - 1) + (\lambda_2 - 1) \ln (\lambda_1 - 1) + \ln \Gamma(1 - \lambda_2) \dots \dots \dots \dots \dots \dots (10)$$

From Eq. 7

$$\lambda_0 = \ln \int_{c_{\star}}^{\infty} \exp\left[-\lambda_1 \ln x - \lambda_2 \ln (\ln x - c)\right] dx \quad \dots \quad \dots \quad \dots \quad (11)$$

The zeroth Lagrange multiplier is also referred to as the partition function (7).

Relation between Lagrange Multipliers and Constraints

According to Tribus (23), the relation between Lagrange multipliers and constraints is obtained by taking partial derivatives of the zeroth Lagrange multiplier and then equating these derivatives to the constraints indicated by Eq. 5. To that end, differentiating Eq. 11 with respect to λ_1 and λ_2 yields

$$\frac{\partial \lambda_0}{\partial \lambda_1} = -\frac{\int_{c_*}^{\infty} \ln x \exp\left[-\lambda_1 \ln x - \lambda_2 \ln (\ln x - c)\right] dx}{\int_{c_*}^{\infty} \exp\left[-\lambda_1 \ln x - \lambda_2 \ln (\ln x - c)\right] dx} \dots \dots \dots \dots \dots (12a)$$

$$\frac{\partial \lambda_0}{\partial \lambda_1} = -\int_{c_*}^{\infty} \ln x \exp\left[-\lambda_0 - \lambda_1 \ln x - \lambda_2 \ln (\ln x - c)\right] dx \quad \dots \quad (12b)$$

J. Hydraul. Eng., 1988, 114(1): 112-122

$$\frac{\partial \lambda_0}{\partial \lambda_1} = -\int_{c_*}^{\infty} \ln x f(x) dx = -E(\ln x) \dots (12c)$$

$$\frac{\partial \lambda_0}{\partial \lambda_2} = -\frac{\int_{c_*}^{\infty} \ln (\ln x - c) \exp [-\lambda_1 \ln x - \lambda_2 \ln (\ln x - c)] dx}{\int_{c_*}^{\infty} \exp [-\lambda_1 \ln x - \lambda_2 \ln (\ln x - c)] dx} \dots (13a)$$

$$\frac{\partial \lambda_0}{\partial \lambda_2} = -\int_{c_*}^{\infty} \ln (\ln x - c)$$

$$\cdot \exp [-\lambda_0 - \lambda_1 \ln x - \lambda_2 \ln (\ln x - c)] dx \dots (13b)$$

$$\frac{\partial \lambda_0}{\partial \lambda_2} = -\int_{c_*}^{\infty} \ln (\ln x - c) f(x) dx = -E[\ln (\ln x - c)] \dots (13c)$$
Also differentiating Eq. 10 with respect to λ_1 and λ_2

$$\frac{\partial \lambda_0}{\partial \lambda_1} = -c + \frac{\lambda_2 - 1}{\lambda_1 - 1} \dots (14)$$

Equating Eqs. 12a-c to 14 and Eqs. 13a-c to 15

$$E(\ln x) = c + \frac{1 - \lambda_2}{\lambda_1 - 1} \quad (16)$$

$$E[\ln (\ln x - c)] = -\ln (\lambda_1 - 1) - \frac{\partial}{\partial \lambda_2} \ln \Gamma(1 - \lambda_2) \quad \dots \quad \dots \quad (17)$$

Differentiating λ_0 twice with respect to λ_1

which satisfies

where s_x^2 = variance of ln x. Let $b = (1 - \lambda_2)$. In terms of b, we can write Eqs. 16–18a as

$$E(\ln x) = c + \frac{b}{\lambda_1 - 1} \qquad (19)$$

where $\Psi(b) = d[\ln \Gamma(b)]/db$ is the digamma function. Substituting the value of λ_0 from Eq. 10 in Eq. 5, we get $f(x) = \exp\{[c(\lambda_1 - 1) - (\lambda_2 - 1) \ln(\lambda_1 - 1) - \ln \Gamma(1 - \lambda_2)] - \lambda_1 \ln x - \lambda_2 \ln(\ln x - c)\}$. In order for f(x) to be the LPT III distribution, $\lambda_1 - 1 = 1/a$. Therefore, expressing λ_1 and λ_2 in terms of *a* and *b*, Eqs. 19–21 can be written as

$$E(\ln x) = c + ab$$
(22)

$$E[\ln (\ln x - c)] = \ln a + \Psi(b)$$
(23)

$$s_x^2 = a^2b$$
(24)

Eqs. 22–24 constitute the POME method of parameter estimation for the LPT III distribution.

Two Other Methods of Parameter Estimation

Two of the most popular methods of parameter estimation are the method of moments (MOM) and the method of maximum likelihood estimation (MLE). Several workers (3,4,13) have compared these two methods and have found that neither is always superior. The POME does not appear to have been used for estimating parameters of the LPT III distribution. Therefore, virtually no literature exists on the comparison of parameter estimates by POME with those by MLE and MOM. To this end we briefly summarize these methods.

Method of Moments (MOM)

1.

The rth moment of Eq. 1 about origin is

$$M_r^0 = \int_{c_*}^{\infty} x^r \frac{1}{ax\Gamma b} \left(\frac{\ln x - c}{a}\right)^{b-1} \exp\left[-\left(\frac{\ln x - c}{a}\right)\right] dx \quad \dots \quad \dots \quad (25)$$

$$M_r^0 = \frac{\exp(cr)}{(1-ra)^b}$$
 (26)

Therefore

$$M_{1}^{0} = \frac{\exp(c)}{(1-a)^{b}} \dots (27a)$$

$$M_{2}^{0} = \frac{\exp(2c)}{(1-2a)^{b}} \dots (27b)$$

$$M_{3}^{0} = \frac{\exp(3c)}{(1-3a)^{b}} \dots (27c)$$
or
$$\ln M_{1}^{0} = c - b \ln(1-a) \dots (28a)$$

$$\ln M_2^0 = 2c - b \ln (1 - 2a) \dots (28b)$$

$$\ln M_3^0 = 3c - b \ln (1 - 3a) \dots (28c)$$

Solving Eqs. 28a-c simultaneously yields

from which a can be determined by a numerical procedure. The other two parameters, b and c, can be determined from either Eqs. 27a-c or 28a-c.

Method of Maximum Likelihood (MLE)

The likelihood function of receiving the sample data $D \equiv \{x_1, x_2, \ldots, x_n\}$ from the LPT III population given the values of a, b, and c is

Therefore

If L(D|a, b, c) is maximal, then so is $\ln L(D|a, b, c)$ so estimates of a, b, and c are sought, producing

$$\frac{\partial}{\partial a} \left[\ln L(D|a, b, c) \right] = 0 \qquad (32a)$$

$$\frac{\partial}{\partial b} \left[\ln L(D|a, b, c) \right] = 0 \qquad (32b)$$

$$\frac{\partial}{\partial c} \left[\ln L(D|a, b, c) \right] = 0 \qquad (32c)$$
Hence, the estimation equations are
$$\sum_{i=1}^{n} (\ln x_i - c) = n \ ab \qquad (33)$$

$$a(b-1)\sum_{i=1}^{n}\frac{1}{(\ln x_{1}-c)}=n$$
 (35)

Eqs. 33-35 are nonlinear in a, b, and c but can be easily solved using a standard numerical procedure. Note that Eq. 33 is equivalent to Eq. 22,

River gaging station (1)	Drainage area (km²) (2)	Length of record, <i>N</i> (3)	Mean, <i>Q</i> (m ³ /s) (4)	Standard deviation, <i>S_Q</i> (5)	Skewness, <i>C_s</i> (6)	Kurtosis, <i>K</i> s (7)
Amite River at Magnolia, Louisiana Sebasticook River at Pitts- field, Maine	1,804	31	698.0	365.9	0.16	2.15
	2,500	52	193.4	63.4	0.80	4.71
Oyster River at Durham, New Hampshire	140	44	9.1	4.4	1.16	5.22
Squannacook River at West Groton, Massa- chusetts Parker River at Byfield, Massachusetts HOP River at Columbia, Connecticut	585	32	43.8	23.0	0.95	4.40
	50	37	6.5	2.8	0.91	3.88
	65	49	64.9	41.1	1.73	5.81

TABLE 1. Pertinent Characteristics of Six Selected Rivers

and Eq. 34 to Eq. 23. The MLE and POME methods differ in their third equations.

APPLICATION TO FIELD DATA

The preceding three methods of parameter estimation were applied to annual peak discharge data for six selected rivers. Pertinent characteristics of the data are given in Table 1. These data were selected on the basis of length, completeness, homogeneity, and independence of record. Each gaging station had a record length of more than 30 years. The parameters estimated by the three methods are summarized in Table 2. For two sample gaging stations, a comparison of observed and computed frequency curves

Biver gaging	MOM				MLE		POME		
station (1)	a (2)	ь (З)	с (4)	a (5)	b (6)	с (7)	a (8)	b (9)	с (10)
Amite River at Magnolia, Louisiana	0.2866	5.076	4.921	0.1320	27.32	2.769	0.1041	38.49	2.369
Sebasticook River at Pitts- field, Maine	0.1320	6.735	4.320	0.0501	53.54	2.527	0.0422	65.94	2.427
Oyster River at Durham, New Hampshire	0.0802	37.28	-0.8925	0.0865	34.59	-0.8925	0.0802	37.28	-0.8925
Squannacook River at West Groton, Massa- chusetts	0.1768	10.20	1.834	0.1088	30.34	0.3343	0.0937	36.32	0.2343
Parker River at Byfield, Massachusetts	0.01467	845.2	-10.618	0.01422	872.0	-10.618	0.01467	845.2	-10.618
HOP River at Columbia, Connecticut	0.15886	11.71	2.157	0.1530	12.15	2.157	0.0906	35.97	0.7576

TABLE 2. Parameter Estimates by MOM, MLE, and POME Methods



FIG. 1. Frequency Curve Using Annual Maximum Discharge Series for Amite River at Magnolia, Louisiana



FIG. 2. Frequency Curve Using Annual Maximum Discharge Series for Sebasticook at Pittsfiled, Maine

119

	RAE			RME			
Station (1)	MOM (2)	MLE (3)	POME (4)	MOM (5)	MLE (6)	POME (7)	
Amite River at Magnolia, Louisiana	23.40	17.86	17.21	10.5	6.98	5.23	
Sebasticook River at Pitts- field, Maine	6.86	6.11	4.87	1.69	0.89	0.85	
Oyster River at Durham, New Hampshire	6.18	6.59	6.18	0.68	0.73	0.68	
Squannacook River at West Groton, Massa- chusetts	11.06	9.61	8.60	2.71	1.81	1.58	
Parker River at Byfield, Massachusetts	4.94	4.97	4.94	0.37	0.38	0.37	
HOP River at Columbia, Connecticut	4.66	4.65	4.92	0.48	0.48	0.51	

TABLE 3. Relative Mean Error and Relative Absolute Error by MOM, MLE, and POME Methods for Six Selected Rivers

is shown in Figs. 1 and 2. The observed frequency curve was obtained by using the Gringorton plotting position formula.

The parameter estimates obtained by the POME and MLE methods are closer to each other than those for MOM. Consequently, their corresponding frequency curves are also closer. POME does not require the use of a coefficient of skewness, whereas MOM does. In this way, the bias is reduced when POME is used to estimate the parameters of LPT III distribution.

To compare these methods further, relative mean error (RME) and relative absolute error (RAE) were computed as given in Table 3. These were computed as

and

RAE =
$$\frac{1}{N} \sum_{i=1}^{N} \left| \frac{Q_0 - Q_c}{Q_0} \right|$$
 (37)

in which N = sample size; $Q_0 =$ observed annual peak discharge of a given probability; and $Q_c =$ computed annual peak discharge of the same probability.

For five of the six data sets, both RME and RAE yielded by POME were less than or equal to those of MLE. For only one data set (the HOP River at Columbia, Connecticut), values of these measures were lower for MOM than those for POME, but the differences were marginal. For only two rivers (the Amite River at Magnolia, Louisiana; and the Squannacook River at West Groton, Massachusetts), values of these measures produced

by POME and MOM were significantly different. For all six data sets. POME and MLE yielded comparable values of these measures. This analysis suggests that POME is a good alternative method of parameter estimation. More testing, however, is needed for defining comparative limitations and strengths of this method.

CONCLUSIONS

The following conclusions can be drawn from this study:

1. POME offerred an alternative method for estimating parameters of the LPT III distribution.

2. The parameter estimates yielded by POME were comparable to those by MLE and MOM.

3. For three of the six selected rivers, POME produced the least RAE and RME.

ACKNOWLEDGMENTS

This study was supported in part by funds provided by the Geological Survey, United States Department of Interior, through the Louisiana Water Resources Research Institute under the project titled "Assessment of Uncertainty in Hydrologic Models for Flood Frequency Analysis."

APPENDIX I. REFERENCES

- 1. Benson, M. A. (1968). "Uniform flood-frequency estimating methods for federal agencies." Water Resour. Resear., 4(5), 891–908.
- 2. Bobee, B. (1975). "The log-Pearson type 3 distribution and its application in hydrology." Water Resour. Resear., 11(5), 681-689.
- 3. Bobee, B., and Robitaille, R. (1975). "Correction of bias in the estimation of the coefficient of skewness." Water Resour. Resear., 11(6), 851-854.
- 4. Bobee, B., and Robitaille, R. (1977). "The use of the Pearson type 3 and log-Pearson type 3 distributions revisited." Water Resourc. Res., 13(2), 427-443.
- 5. Buckett, J., and Oliver, F. R. (1977). "Fitting the Pearson type 3 distribution in practice." Water Resourc. Res., 13(5), 851-852.
- 6. Condi, R. (1977). "The log-Pearson type 3 distribution: The T-year event and its asympototic standard error by maximum likelihood theory." Water Resourc. Res., 13(6), 987-991.
- 7. Jaynes, E. T. (1968). "Prior probabilities." IEEE Trans. Syst., Man. Cybern., 3(SSC-4), 227-241.
- Jowitt, P. W. (1979). "The extreme-value type-I distribution and the principle of maximum entropy." J. Hydr., 42, 23–38.
- 9. Jowitt, P. W., and Munro, J. (1975). "The influence of void distribution and entropy on the engineering properties of granular media." Proc. 2nd Int. Conf. Applic. Stat. Probability Soil Struct. Engrg., Aachen, W. Germany, 365-385.
- 10. Lall, U., and Beard, L. R. (1982). "Estimation of Pearson type 3 moments." Water Resour. Res., 18(5), 1563-1569.
- 11. Levine, R. D., and Tribus, M. (1979). The maximum entropy formalism. M.I.T. Press, Cambridge, Mass.
- 12. McMahon, T. A., and Srikanthan, R. (1981). "Log-Pearson III distribution-Is it applicable to flood frequency analysis of Australian streams?" J. Hydr., 52, 139–142.
- 13. Matalas, N. C., and Wallis, J. R. (1973). "Eureka! It fits a Pearson type 3 distribution." Water Resourc. Res., 9(2), 281–289.
 14. Munro, J., and Jowitt, P. W. (1978). "Decision analysis in the ready-mixed

concrete industry." Proc. Inst. Civ. Engrg., 65(2), 41-52.

- 15. Nozdryn-Plotnieko, M. J., and Watt, W. E. (1979). "Assessment of fitting techniques for the log-Pearson type 3 distribution using Monte Carlo simulation." Water Resourc. Res., 15(3), 719-718.
- 16. Rao, D. V. (1980). "Log-Pearson type 3 distribution: a generalized evaluation." J. Hydr. Div., ASCE, 106(5), 853-872.
- 17. Reza, F. M. (1961). An introduction to information theory. McGraw-Hill Book Co., Inc., New York, N.Y.
- Shaligram, V. M., and Lele, V. S. (1978). "Analysis of hydrologic data using Pearson type III distribution." Nordic Hydr., 9, 31-42.
- 19. Shannon, C. E. (1948). "The mathematical theory of communication." Bell System Tech. J., 27, 279-428, 612-656.
- 20. Sonuga, J. O. (1972). "Principle of maximum entropy in hydrologic frequency analysis." J. Hydr., 17, 177-191.
- 21. Sonuga, J. O. (1976). "Entropy principle applied to rainfall-runoff process." J. Hydr., 30, 81–94.
- 22. Srikanthan, R., and McMahon, T. (1981). "Log-Pearson III distribution-Effect of dependence distribution parameters and sample size on peak annual flood estimates." J. Hydr., 52, 149-159.
- 23. Tribus, M. (1969). Rational descriptors, decisions and designs. Pergamon Press, New York, N.Y.
- 24. Water Resources Council (1977). "Guidelines for determining flood flow frequency." Bull. No. 17A Hydr. Comm., 1-26. and a second second

APPENDIX II. NOTATION

The following symbols are used in this paper:

parameter; parameter; bias statistic; parameter; exp(c); covariance (\cdot); expectation of [\cdot]; exponential; probability density function of x; entropy of (\cdot); а = b = BIAS = С = C_* = $cov(\cdot)$ = $E[\cdot]$ = exp = f(x)= $H(\cdot)$ = entropy of (\cdot) ; scale parameter in definition of entropy; K = $L(\cdot)$ maximum likelihood function of (\cdot) ; = ln == logarithm; $M_r^0(\cdot)$ = *r*th moment of (\cdot) about origin 0; probability of $x = x_i$, i = 0, 1, 2, ...; $p(x_i)$ = Q_0 observed flood discharge; = \tilde{Q}_c computed flood discharge; === RMSE root mean square error; = s_x^2 = variance of ln x; var(·) === variance of (\cdot) ; *j*th constraint as function of x, j = 1, 2, ...; $w_j(x)$ = == random variable; x = $\ln x$; y ith Lagrange multiplier; λi = $\Gamma(\cdot)$ = gamma of (\cdot) ; and $\Psi(b)$ = $d[\ln \Gamma(b)]/db =$ digamma function.