Original Research Article

WEIGHTED LOG-PEARSON TYPE III DISTRIBUTION: PROPERTIES, ESTIMATION AND APPLICATION

Abstract: In this paper, we have proposed a new version of three parameter log-Pearson type III distribution known as weighted Log-Pearson type III distribution. The different structural properties of the newly model have been studied. The maximum likelihood estimators of the parameters and the Fishers information matrix have been discussed. Finally, a real-life data set has been analyzed.

Keywords: Weighted distribution, Three parameter log-Pearson type III distribution, Reliability analysis, Maximum likelihood estimator, Order statistics, Entropies.

Comment [TG1]: parameters

Comment [TG2]: delete

1. INTRODUCTION

The log-Pearson type (LPT) III distribution is extensively used in hydrologic frequency analysis. LPT-III distribution is nothing but the generalized version of Gamma distribution. Many researchers have studied this distribution by estimating in various method of estimation along with the applicability in hydrological analysis. Condie (1971) has discussed the various properties and estimation of Log Pearson Type 3 distribution by using the method of maximum likelihood. Bobee (1975) has discussed the Log Pearson Type 3 distribution and its application in Hydrology. Nozdryn-Plotnicki and Watt (1979) have analyzed the fitting techniques for the Log Pearson Type 3 distribution using Monte Carlo Simulation. McMahon and Srikanthan (1981) have researched whether the Log Pearson III distribution is applicable to flood frequency analysis of Australian streams. As a result, they have found that Log Pearson Type III distribution is appropriate for estimation of extreme flood discharges. Srikanthan and McMahon (1981) discussed the estimation of the parameters Log Pearson III distribution and estimated the flood annual peak. Phien and Hira (1983) have estimated the parameters of the Log Pearson Type 3 distribution by the method of maximum likelihood and method of (direct and indirect) moments. Ashkar and Bobee (1987) have applied the generalized method of moments for some practical results for the log-Pearson type 3 distribution specially in flood frequency analysis. Singh and Singh (1988) have discussed the principle of maximum entropy (POME) to estimate the parameters of Log Pearson III distribution along with the method of moment (MOM) and maximum likelihood estimation (MLE) methods. Another distribution which is used to study the hydrological analysis is Weighted distribution. This concept of weighted distributions was given by Fisher (1934) to model the ascertainment bias. Later Rao (1965) developed this concept in a unified manner while modelling the statistical data when the standard distributions were not appropriate to record these observations with equal probabilities. As a result, weighted models were formulated in such situations to record the observations according to some weighted function. The weighted distribution reduces to length biased distribution when the weight function considers only the length of the units. The concept of length biased sampling was first introduced by Cox (1969) and Zelen (1974). More generally, when the sampling mechanism selects units with probability proportional to some measure of the unit size, resulting distribution is called size-biased. There are various good sources, which provide the detailed description of weighted distributions.

The main purpose of this paper is to introduce a new distribution, which is Weighted Log-Pearson Type-III distribution. Generally, Log-Pearson Type-III distribution used to model hydrological analysis. In this distribution, we introduced a new version to model the distribution in temperature analysis.

2. WEIGHTED LOG-PEARSON TYPE III (WLPT-III) DISTRIBUTION

The probability density function (pdf) of three parameter Log-Pearson Type-III distribution is given by

$$f(x;a,b,c) = \frac{1}{ax\Gamma(b)}\left(\frac{\ln x - c}{a}\right)^{b-1} \exp\left[-\left(\frac{\ln x - c}{a}\right)\right] \quad ; a > 0, \mathrm{b} > 0, \text{and } 0 < \mathrm{c} < \ln \mathrm{x} \qquad (2.1)$$

Suppose X is a non-negative random variable with probability density function $f(x)$. Let $w(x)$ be the non-negative weight function, then the probability density function of the weighted random variable X_w is given by:

$$f_w(x) = \frac{w(x)f(x)}{E(w(x))}, \quad x > 0,$$

where $w(x)$ be a non-negative weight function and $E(w(x)) = \int w(x)f(x)dx < \infty$.

In this paper, we will consider the weight function as $w(x) = x$ to obtain the weighted three parameter Log-Pearson type III distribution. The probability density function of weighted three parameter Log-Pearson Type III distribution is given as:

Comment [TG3]: parameters

$$f_w(x) = \frac{xf(x)}{E(x)}; \qquad x > 0 \qquad (2.2)$$

$$\text{where } E(x) = \int_{e^c}^{\infty} xf(x;a,b,c)\,dx$$

$$E(x) = \frac{e^c}{(1-a)^b} \qquad (2.3)$$

Substitute (2.1) and (2.3) in equation (2.2), we will get the required pdf of weighted three parameter Log-Pearson type III distribution as

$$f_w(x) = \frac{(1-a)^b}{a\Gamma(b)e^c}\left(\frac{\ln x - c}{a}\right)^{b-1} e^{-\left(\frac{\ln x - c}{a}\right)} \qquad (2.4)$$

and the cumulative density function of WTLPT-III distribution is obtained as

$$F_w(x) = \int_0^x f_w(x)dx$$

$$= \int_{e^c}^{x} \frac{(1-a)^b}{a\Gamma(b)e^c}\left(\frac{\ln x - c}{a}\right)^{b-1} e^{-\left(\frac{\ln x - c}{a}\right)} dx$$

After simplification, we will get the cumulative distribution function of WLP3 distribution

$$F_w(x) = \frac{1}{(1-a)\Gamma(b)}\gamma\left(b, \frac{\ln x - c}{a}\right) \tag{2.5}$$

3. RELIABILITY ANALYSIS

In this section, we will discuss about the survival function, failure rate, reverse hazard rate and Mill's ratio of the WS distribution.

The survival function or the reliability function of the weighted Log-Pearson Type-III distribution is given by

Comment [TG4]: It must be reformulated

$$S(x) = 1 - \frac{1}{(1-a)\Gamma(b)}\gamma\left(b, \frac{\ln x - c}{a}\right)$$

The hazard function is also known as the hazard rate, instantaneous failure rate or force of mortality and is given by

$$h(x) = \frac{(1-a)^{b+1}\left(\frac{\ln x - c}{a}\right)^{b-1} e^{-\left(\frac{\ln x - c}{a}\right)}}{a(1-a)e^c\Gamma(b) - \gamma\left(b, \frac{\ln x - c}{a}\right)}$$

The reverse hazard rate is given by

$$h_r(x) = \frac{(1-a)^{b+1}\left(\frac{\ln x - c}{a}\right)^{b-1} e^{-\left(\frac{\ln x - c}{a}\right)}}{ae^c\gamma\left(b, \frac{\ln x - c}{a}\right)}$$

and the Mills ratio of the WLPT-III distribution is

$$\frac{1}{h_r(x)} = \frac{ae^c\gamma\left(b, \frac{\ln x - c}{a}\right)}{(1-a)^{b+1}\left(\frac{\ln x - c}{a}\right)^{b-1} e^{-\left(\frac{\ln x - c}{a}\right)}}$$

4. MOMENTS AND ASSOCIATED MEASURES

Let *X* denotes the random variable of WLPT-III distribution with parameters *α, b* and *c* then the r^{th} order moment $E(X^r)$ of WLP3 distribution can be obtained as

$$E(X^r) = \mu_r' = \int_0^\infty x^r f_w(x)dx = \int_{e^c}^\infty x^r \frac{(1-a)^b}{a\Gamma(b)e^c}\left(\frac{\ln x - c}{a}\right)^{b-1} e^{-\left(\frac{\ln x - c}{a}\right)}dx$$

$$E(X^r) = e^{cr}\left(\frac{1-a}{1-a-ar}\right)^b \qquad (4.1)$$

Putting *r* =1 in equation (4.1), we will get the mean of WLPT-III distribution which is given by

$$E(X) = \mu_1' = e^c\left(\frac{1-a}{1-2a}\right)^b$$

and putting *r* = 2, we obtain the second moment as

$$E(X^2) = e^{2c}\left(\frac{1-a}{1-3a}\right)^b$$

Therefore,

$$\text{Variance } = e^{2c}\left\{\left(\frac{1-a}{1-3a}\right)^b - \left(\frac{1-a}{1-2a}\right)^{2b}\right\}$$

$$\text{S.D.}(\sigma) = e^c\sqrt{\left\{\left(\frac{1-a}{1-3a}\right)^b - \left(\frac{1-a}{1-2a}\right)^{2b}\right\}}$$

$$\text{Coefficient of Variation } (C.V.) = \frac{\sigma}{\mu_1'} = \frac{(1-2b)^b\sqrt{\left\{\left(\frac{1-a}{1-3a}\right)^b - \left(\frac{1-a}{1-2a}\right)^{2b}\right\}}}{(1-a)^b}$$

$$\text{Coefficient of Dispersion } (\gamma) = \frac{\sigma^2}{\mu_1'} = \frac{e^c(1-2b)^b\left\{\left(\frac{1-a}{1-3a}\right)^b - \left(\frac{1-a}{1-2a}\right)^{2b}\right\}}{(1-a)^b}$$

Moment generating function and Characteristic function

Let X have a weighted Log-Pearson type III distribution, then the MGF of X is obtained as

$$M_X(t) = E(e^{tx}) = \int_0^\infty e^{tx} f_w(x)dx$$

Using Taylor's series

$$M_X(t) = E(e^{tx}) = \int_0^\infty \left(1 + tx + \frac{(tx)^2}{2!} + \cdots\right) f_w(x)dx$$

$$= \int_0^\infty \sum_{j=0}^\infty \frac{t^j}{j!} x^j f_w(x)dx$$

$$= \sum_{j=0}^\infty \frac{t^j}{j!} \mu_j'$$

$$= \sum_{j=0}^\infty \frac{t^j}{j!} \left\{ e^{cj} \left(\frac{1-a}{1-a-aj} \right)^b \right\}$$

$$\Rightarrow M_x(t) = \sum_{j=0}^\infty \frac{t^j}{j!} \left\{ e^{cj} \left(\frac{1-a}{1-a-aj} \right)^b \right\}$$

Similarly, the characteristic function of WLP-III distribution can be obtained as

$$\varphi_x(t) = M_x(it)$$

$$\Rightarrow M_x(it) = \sum_{j=0}^\infty \frac{(it)^j}{j!} \left\{ e^{cj} \left(\frac{1-a}{1-a-aj} \right)^b \right\}$$

5. ENTROPIES

The concept of entropy is important in different areas such as probability and statistics, physics, communication theory and economics. Entropies quantify the diversity, uncertainty, or randomness of a system. Entropy of a random variable X is a measure of variation of the uncertainty.

5.1. Renyi Entropy

The Renyi entropy is important in ecology and statistics as index of diversity. The Renyi entropy is also important in quantum information, where it can be used as a measure of entanglement. For a given probability distribution, Renyi entropy is given by

$$e(\beta) = \frac{1}{1-\beta} \log\left(\int f^\beta(x)dx \right)$$

where, $\beta > 0$ and $\beta \neq 1$

$$e(\beta)=\frac{1}{1-\beta}\log\int_{e^c}^{\infty}\left\{\frac{(1-a)^b}{a\Gamma(b)e^c}\left(\frac{\ln x-c}{a}\right)^{b-1}e^{-\left(\frac{\ln x-c}{a}\right)}\right\}^{\beta}dx$$

$$e(\beta)=\frac{1}{1-\beta}\log\left\{\left(\frac{(1-a)^b}{a\Gamma(b)e^c}\right)^{\beta}\int_{e^c}^{\infty}\left\{\left(\frac{\ln x-c}{a}\right)^{b-1}\right\}^{\beta}\left\{e^{-\left(\frac{\ln x-c}{a}\right)}\right\}^{\beta}dx\right\}$$

$$e(\beta)=\frac{1}{1-\beta}\log\left\{\frac{(1-a)^b}{a\Gamma(b)e^c}\right\}^{\beta}\frac{1}{a^{\beta(b-1)}}\int_{e^c}^{\infty}((-c)+\ln x)^{\beta(b-1)}e^{-\left(\frac{\ln x-c}{a}\right)^{\beta}}dx$$

Using binomial expansion

$$\{(-c)+\ln x\}^{\beta(b-1)}=\sum_{k=0}^{\infty}\binom{\beta(b-1)}{k}(-c)^{\beta(b-1)-k}(\ln x)^k$$

in equation (8), we get

$$e(\beta)=\frac{1}{1-\beta}\log\left[\left\{\frac{(1-a)^b}{a\Gamma(b)e^c}\right\}^{\beta}\frac{1}{a^{\beta(b-1)}}\int_{e^c}^{\infty}\sum_{k=0}^{\infty}\binom{\beta(b-1)}{k}(-c)^{\beta(b-1)}\ln x^k e^{-\left(\frac{\ln x-c}{a}\right)^{\beta}}dx\right]$$

$$e(\beta)=\frac{1}{1-\beta}\log\left[\left(\frac{(1-a)^b}{a\Gamma(b)e^c}\right)^{\beta}\frac{ae^c}{a^{\beta(b-1)}}\left(\frac{a+c(\beta-a)}{(\beta-a)^2}\right)\sum_{k=0}^{\infty}k\binom{\beta(b-1)}{k}(-c)^{\beta(b-1)-k}\right]$$

5.2. Tsallis Entropy

A generalization of Boltzmann-Gibbs (B-G) statistical mechanics initiated by Tsallis has focussed a great deal to attention. This generalization of B-G statistics was proposed firstly by introducing the mathematical expression of Tsallis entropy (Tsallis, 1988) for a continuous random variable is defined as follows

$$S_\lambda=\frac{1}{\lambda-1}\left(1-\int_0^{\infty}f^{\lambda}(x)dx\right)$$

$$S_\lambda=\frac{1}{\lambda-1}\left(1-\int_{e^c}^{\infty}\left(\frac{(1-a)^b}{a\Gamma(b)e^c}\left(\frac{\ln x-c}{a}\right)^{b-1}e^{-\left(\frac{\ln x-c}{a}\right)}\right)^{\lambda}dx\right)$$

By simplification,

$$S_\lambda=\frac{1}{\lambda-1}\left[1-\left\{\left(\frac{(1-a)^b}{a\Gamma(b)e^c}\right)^{\lambda}\frac{ae^c}{a^{\lambda(b-1)}}\left(\frac{a+c(\lambda-a)}{(\lambda-a)^2}\right)\sum_{k=0}^{\infty}k\binom{\lambda(b-1)}{k}(-c)^{\lambda(b-1)k}\right\}\right]$$

Comment [TG5]: focosed

6. ORDER STATISTICS

Let $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ be the order statistics of a random sample $X_1, X_2, \ldots, X_n$ drawn from the continuous population with probability density function $f_x(x)$ and cumulative density function with $F_x(x)$, then the pdf of r^{th} order statistics $X_{(r)}$ is given by

$$f_{X(r)}(x) = \frac{n!}{(r-1)!(n-r)!} f_X(x)\left[F_X(x)\right]^{r-1}\left[1-F_X(x)\right]^{n-r} \tag{6.1}$$

Using the equations (2.4) and (2.5) in equation (6.1), the probability density function of r^{th} order statistics $X_{(r)}$ of WLPT-III distribution is given by

$$\begin{aligned} f_{X(r)}(x) = {} & \frac{n!}{(r-1)!(n-1)!}\left(\frac{(1-a)^b}{ae^c\Gamma(b)}\left(\frac{\ln x-c}{a}\right)^{b-1} e^{-\left(\frac{\ln x-c}{a}\right)}\right) \\ & \times\left(\frac{1}{(1-a)\Gamma(b)}\gamma\left(b,\frac{\ln x-c}{a}\right)\right)^{r-1} \\ & \times\left(1-\left(\frac{1}{(1-a)\Gamma(b)}\gamma\left(b,\frac{\ln x-c}{a}\right)\right)\right)^{n-r} \end{aligned}$$

Therefore, the probability density function of higher order statistics $X_{(n)}$ can be obtained as

$$f_{X(n)}(x) = n\left(\frac{(1-a)^b}{ae^c\Gamma(b)}\left(\frac{\ln x-c}{a}\right)^{b-1} e^{-\left(\frac{\ln x-c}{a}\right)}\right)\left(\frac{1}{(1-a)\Gamma(b)}\gamma\left(b,\frac{\ln x-c}{a}\right)\right)^{n-1}$$

and the pdf of 1^{st} order statistic $X_{(1)}$ can be obtained as

$$f_{X(1)}(x) = n\left(\frac{(1-a)^b}{ae^c\Gamma(b)}\left(\frac{\ln x-c}{a}\right)^{b-1} e^{-\left(\frac{\ln x-c}{a}\right)}\right)\left(1-\frac{1}{(1-a)\Gamma(b)}\gamma\left(b,\frac{\ln x-c}{a}\right)\right)^{n-1}$$

7. BONFERRONI AND LORENZ CURVES

The Bonferroni and the Lorenz curves are not only used in economics in order to study the income and poverty, but it is also being used in other fields like reliability, medicine, insurance and demography.

The Bonferroni and Lorenz curves are given by

$$B(p)=\frac{1}{p\mu_1'}\int_0^q xf(x)dx$$

and
$$L(p)=\frac{1}{\mu_1'}\int_0^q xf(x)dx$$

Where $\mu_1'=E(x)=e^c\left(\frac{1-a}{1-2a}\right)^b$ and $q=F^{-1}(p)$

$$B(p)=\frac{(1-2a)^b}{pe^c(1-a)^b}\int_{e^c}^q \frac{(1-a)}{ae^c\Gamma(b)}x\left(\frac{\ln x-c}{a}\right)^{b-1}e^{-\left(\frac{\ln x-c}{a}\right)}dx$$

After simplification, we get

$$B(p)=\frac{(1-2a)}{ap\Gamma(b)}\gamma\left(b,\frac{\ln q-c}{a}\right)$$

and

$$L(p)=p\,B(p)=\frac{(1-2a)}{a\Gamma(b)}\gamma\left(b,\frac{\ln q-c}{a}\right)$$

8. MAXIMUM LIKELIHOOD ESTIMATOR AND FISHER'S INFORMATION MATRIX

In this section, we will discuss the maximum likelihood estimators of the parameters of weighted log-Pearson type III distribution. Consider $X_1,X_2,\ldots,X_n$ be the random sample of size n from the WLPT-III distribution, then the likelihood function is given by

$$L(x;a,b,c)=\left(\frac{(1-a)}{ae^c\Gamma(b)}\right)^n\prod_{i=1}^n\left[\left(\frac{\ln x-c}{a}\right)^{b-1}e^{-\left(\frac{\ln x-c}{a}\right)}\right]$$

The log likelihood function is

$$\log L=n\log(1-a)-n\log a-nc-n\log\Gamma(b)+(b-1)\sum_i(\ln x_i-c)-(b-1)\log a-\left(\frac{\sum_i \ln x_i-c}{a}\right)$$

The maximum likelihood estimates of a,b,c can be obtained by differentiating log L with respect to a,b,c and must satisfy the normal equation

$$\frac{\partial \log L}{\partial a} = -\frac{nb}{(1-a)} - \frac{n}{a} - \frac{b}{a} + \left(\frac{\ln x - c}{a^2}\right) = 0$$

$$\frac{\partial \log L}{\partial b} = n\log(1-a) - n\Psi(b) + \sum_i (\log x_i - c) - \log a = 0$$

$$\frac{\partial \log L}{\partial c} = -n - n(b-1) + \frac{n}{a} = 0$$

Where ψ(.) is the digamma function.

Because of the complicated form of likelihood equations, algebraically it is very difficult to solve the system of nonlinear equations. Therefore we use R and wolfram mathematics for estimating the required parameters.

To obtain confidence interval we use the asymptotic normality results. We have that, if $\hat{\lambda} = (\hat{a}, \hat{b}, \hat{c})$ denotes the MLE of $\lambda = (a, b, c)$, we can state the results as follows:

$$\sqrt{n}(\hat{\lambda} - \lambda) \rightarrow N_3(0, I^{-1}(\lambda))$$

Where $I(\lambda)$ is Fisher's Information Matrix, i.e.,

$$I(\lambda) = -\frac{1}{n}\begin{pmatrix} E\left(\frac{\partial^2 \log L}{\partial a^2}\right) & E\left(\frac{\partial^2 \log L}{\partial a\,\partial b}\right) & E\left(\frac{\partial^2 \log L}{\partial a\,\partial c}\right) \\ E\left(\frac{\partial^2 \log L}{\partial b\,\partial a}\right) & E\left(\frac{\partial^2 \log L}{\partial b^2}\right) & E\left(\frac{\partial^2 \log L}{\partial b\,\partial c}\right) \\ E\left(\frac{\partial^2 \log L}{\partial c\,\partial a}\right) & E\left(\frac{\partial^2 \log L}{\partial c\partial b}\right) & E\left(\frac{\partial^2 \log L}{\partial c^2}\right) \end{pmatrix}$$

Since λ being unknown, we estimate $I^{-1}(\lambda)$ by $I^{-1}(\hat{\lambda})$ and this can be used to obtain asymptotic confidence intervals for α, b and c.

9. DATA ANALYSIS

In this study, monthly mean temperature series of Silchar city, Assam, India from January 1981-December 2018 (37 years) has been collected from India Meteorological Department, Pune, India. The data has been analyzed using the LPT distribution. The respective values of the parameters by using the method of maximum likelihood estimation are $\hat{a} = 0.1000, \hat{b} = 2.4479$ and $\hat{c} = 2.2552$.

10. CONCLUSION

In the present study, we have introduced a new generalization of the Log-Pearson Type-III distribution as weighted Log-Pearson Type-III distribution with three parameters. The subject distribution is generated by using the weighted technique and the parameters have been obtained by using maximum likelihood estimator. Some statistical properties along with reliability measures are discussed. The proposed distribution have its applications in temperature data of Silchar City, Assam, India.

REFERENCES

1. Ashkar, F. and Bobee, 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.
2. Bobee, B. , 1975. The Log Pearson Type 3 Distribution and Its Application in Hydrology. WATER RESOURCES RESEARCH, 5: 681-689.
3. Condie, R., 1977. The Log Pearson Type 3 Distribution: The T-Year Event and Its Asymptotic Standard Error by Maximum Likelihood Theory. WATER RESOURCES RESEARCH, 6: 987-991.
4. Dey, S., Ali, S., Park, C., 2015, Weighted exponential distribution: properties and different methods of estimation. Journal of Statistical Computation and Simulation. 3641-3661.
5. McMahon, T.A. and Srikanthan, R., 1981. Log Pearson III distribution -- is it applicable to flood frequency analysis of Australian streams? J. Hydrol., 52: 139-147.
6. 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 RESOURCES RESEARCH, 3: 714-718.
7. Phien, H.N. and Hira, M.A., 1983. Log Pearson type-3 distribution: parameter estimation. J. Hydrol., 64: 25-37.
8. Srikanthan, R. and McMahon, T.A., 1981. Log Pearson III distribution -- effect of dependence, distribution parameters and sample size on peak annual flood estimates. J. Hydrol., 52: 149--159.
9. Singh, V.P. and Singh, K., 1988. Parameter Estimation For Log-Pearson Type Iii Distribution By Pome. J. Hydraul. Eng., 114(1): 112-122.

Comment [T6]: Rephrase in one uniform