International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2015): 6.391

Minimum Risk Point Estimation of Location Parameter of a Galton Distribution: A Sequential Approach

Sandeep Singh Charak

University of Jammu, Department of Statistics, Jammu

Abstract: ***In this paper we have considered one important distribution useful in Reliability and Life Testing namely, Galton distribution or Lognormal distribution. The lognormal distribution is sometimes called the antilognormal distribution. The name has some logical basis in that it is not the distribution of the logarithm of a normal variable (this is not even always real) but of an exponential – that is, antilognormal-function of such a variable. It is occasionally referred to as the Galton distribution or Galton's distribution, after Francis Galton. In section I we give a brief introduction to their applications and in section II we explain the problem of estimation of minimum risk point estimation of mean of a Lognormal distribution or Galton distribution and failure of the fixed sample size procedure. In section III sequential solution for this problem have been provided along with study of their asymptotic properties.***

Keywords: Sequential procedure, Stopping time, regret, minimum risk, point estimation

1. Introduction

The Galton distribution or Lognormal distribution is sometimes called the antilognormal distribution. The name has some logical basis in that it is not the distribution of the logarithm of a normal variable (this is not even always real) but of an exponential – that is, antilognormal-function of such a variable. It is occasionally referred to as the Galton distribution or Galton's distribution, after Francis Galton. The use of the distribution in the description of psychophysical phenomena gave a graphical method for estimating parameters. Galton distribution or Lognormal distribution have been found to be applicable to distributions of particle size in naturally occurring aggregates. Further applications, in agricultural, entomological and even literary research that generate Galton distribution or Lognormal distribution in a variety of biological, pharmacological, modeling the weights of children, and construction of age - specific reference ranges for clinical variables. The sums of independent Galton or lognormal variables are used in telecommunication to study the effects of the atmosphere on radar signals. It has also been found to be a serious competitor to the Weibull distribution in representing lifetime distribution. Aitchison and Brown (1957) cite various applications such as the number of persons in a census occupation class and the distribution of incomes in econometrics, the distribution of stars in the universe and the distribution of the radical component of Chinese characters. The Galton or lognormal model for minerals present in low concentrations has been experimentally verified for many minerals. Chattopadhyay, B. & Kelley, K. (2016) they propose a general theory for a sequential estimation of the population coefficient of variation that considers both the sampling error and the study cost, importantly without specific distributional assumptions. In this paper the same distribution is revisited but we considered here the problem of minimum risk point estimation of location parameter of a Galton distribution or Lognormal distribution besides sequential solution for this problem have been provided along with study of their asymptotic properties.

2. The Set Up of the Estimation Problems and the Failure of the Fixed Sample size Procedures

Let us consider a sequence $\{X_i\}$, i=1,2,3,…. of independently distributed random variable from a Galton distribution or Lognormal distribution with the p.d.f

$$f(X;\,\mu,\sigma) = \frac{1}{\sigma X\sqrt{2\pi}} exp\left\{-\frac{1}{2\sigma^2}(logX-\mu)^2\right\};\ (X>0),$$

where $\mu \in (-\infty,\infty)$ and $\sigma \in (0,\infty)$ are the parameters.

The mean of the distribution is $\zeta = E(X_i) = exp(\mu + \sigma 2/2$. Given a random sample $X_1, X_2, X_3,\ldots\ldots,X_n$ of size n, let the loss incurred in estimating ζ by

$$\bar{X}_n = n^{-1}\sum_{i=1}^{n} X_i$$

be

$$L(\zeta\,,\bar{X}_n) = A(\,\bar{X}_n - \zeta)^2 + n, \qquad (2.1)$$

Having the associated risk

$$R_n(A) = \left(\frac{A}{n}\right)\zeta^2\left(e^{\sigma^2}-1\right) + n. \qquad (2.2)$$

The value $n = n_0$, which minimizes $R_n(A)$, is given by

$$n_0 = A^{\frac{1}{2}}\zeta\left(e^{\sigma^2}-1\right)^{\frac{1}{2}}, \qquad (2.3)$$

and substituting $n = n_0$ in (2.2) , the corresponding minimum risk is

$$R_{n_0}(A) = 2A^{\frac{1}{2}}\zeta\left(e^{\sigma^2}-1\right)^{\frac{1}{2}}. \,(2.4)$$

However, when μ and /or σ^2 is unknown, no fixed sample size procedure achieves the minimum risk (2.4).

Our problem is point estimation of the mean of a Galton distribution or Lognormal distribution.

3. The Sequential Procedure for the Point Estimation of Mean of Lognormal Distribution

We first consider the case when σ is known and, without any loss of generality, we assume that $\sigma = 1$. From (2.3) and (2.4),

$$n_0 = KA^{\frac{1}{2}}\zeta_0 \,, \qquad (3.1)$$

and

$$R_n(A) = 2KA^{\frac{1}{2}}\zeta_0 \,, \qquad (3.2)$$

where $K = (e-1)^{1/2}$ and $\zeta_0 = exp(\mu + 1/2)$. Motivated by (3.1), the stoping time N is defined by

$$N = \inf\left[n \geq 1 : KA^{\frac{1}{2}}\, \bar{X}_n\right]. \qquad (33)$$

After stopping, we estimating ζ_0 by $\bar{X}_N$, incurring the risk

$$R_N(A) = AE((\bar{X}_N - \zeta_0)^2) + E(N) \qquad (3.4)$$

Following Starr(1966b) and Star and Woodroofe(1969), we define the 'risk efficiency' and 'regret' of the sequential procedure(3.3), by

$$R_e(A) = \frac{R_N(A)}{R_{n_0}(A)} \qquad (3.5)$$

and

$$R_g(A) = R_N(A) - R_{n_0}(A)\,, \qquad (3.6)$$

respectively.

Now we prove the following theorem, which establishes the result that the sequential procedure (3.3) is asymptotically 'risk efficient'.

Theorem 1 For the stopping rule defined at (3.3) and all ζ_0 ,

$$\lim_{A\to\infty} R_e(A) = 1.$$

Proof

Denoting by

$$S_n = n\,\bar{X}_N = \sum_{i=1}^{n} X_i \,,$$

we can rewrite the stopping rule (3.3) as

$$N = \inf\left[n \geq 1 : S_n \leq K^{-1}A^{-\frac{1}{2}}n^2\right]. \qquad (3.7)$$

From Wald's lemma for cumulative sums,

$E((S_N - N\zeta_0)^2) = K^2{\zeta_0}^2E(N)$.

Hence , we obtain from (3.4) that

$$R_N(A) = AE\left((S_N - N\zeta_0)^2(AN^{-2} - K^{-2}{\zeta_0}^{-2})\right) + 2E(N). \; (3.8)$$

After substitutions from (3.2) and (3.8) in (3.5), we get

$$R_e(A) = \left(2KA^{\frac{1}{2}}\zeta_0\right)^{-1} E\left((S_N - N\zeta_0)^2(AN^{-2} - K^{-2}{\zeta_0}^{-2})\right) + E\left(\frac{N}{KA^{\frac{1}{2}}\zeta_0}\right). \; (3.9)$$

It can be seen that $A^{-\frac{1}{2}}N \xrightarrow{a.s} K\zeta_0$ as $A\to\infty$.

It now follows from a result of Gut (1974) that

$$\left[\left(A^{-\frac{1}{2}}N\right)^4 : A \geq 1\right] \text{ is uniformly integrable. } (3.10)$$

From(3.10) and dominated convergence theorem,

$$\lim_{A\to\infty} E\left[\frac{N}{KA^{\frac{1}{2}}\zeta_0}\right] = 1. \; (3.11)$$

From (3.9) and (3.11), we conclude that the result follows if we can prove that

$$E\left[A^{-\frac{1}{2}}(S_N - N\zeta_0)^2(AN^{-2} - K^{-2}{\zeta_0}^{-2})\right] = o(1) \; as \; A \to \infty. \; (3.12)$$

To this end, from Holder's inequality ,

$E\left[A^{-\frac{1}{2}}(S_N - N\zeta_0)^2(AN^{-2} - K^{-2}{\zeta_0}^{-2})\right] \leq E^{1/2}\left|A^{-\frac{1}{4}}(S_N - N\zeta_0 4E1/2AN-2-K-2\zeta_0-22.\right.$ (3.13)

From (3.10) and lemma 5 of Chow and Yu (1981),

$$\left[\left|A^{-\frac{1}{4}}(S_N - N\zeta_0)\right|^4 : A \geq 1\right] \text{ is uniformly integrable. } (3.14)$$

From the definition of N at (3.3), $AN^{-2} \leq A^{\frac{1}{2}}(KS_N)^{-1}$.

Hence, using the dominated ergodic theorem of MarcinKiewicz and Zygmund [see Chow and Teicher (1978,p.35)], $\left((A^2N^{-4}) : A \geq 1\right)$ is uniformly integrable. (3.15)

Since $AN^{-2} \xrightarrow{as} K^{-2}{\zeta_0}^{-2}$ as $A \to \infty$, using (3.14) and (3.15), we obtain from (3.13) that

$$A^{-\frac{1}{2}}E[(S_N - N\zeta_0)^2(AN^{-2} - K^{-2}{\zeta_0}^{-2})] \leq O(1)\, E^{1/2}\left[(AN^{-2} - K^{-2}{\zeta_0}^{-2})\right]^2 = o(1), as\; A \to \infty\,,$$

and (3.12) holds. This completes the proof of the theorem.

In the next theorem, we prove the bounded nature of the 'regret'.

Theorem 2 For the sequential procedure(3.3),

$$\lim_{A\to\infty} R_g(A) = o(A^{\frac{1}{2}}).$$

Proof

From (3.2) and (3.4), substituting the values of $R_{n_0}(A)$ and $R_N(A)$ in (3.6), we get

$$R_g(A) = E\left((S_N - N\zeta_0)^2(AN^{-2} - K^{-2}{\zeta_0}^{-2})\right) + 2A^{\frac{1}{2}}E\left[\left(A^{-\frac{1}{2}}N - K\zeta_0\right)\right]. \; (3.16)$$

By Holder's inequality,

$$E\left((S_N - N\zeta_0)^2(AN^{-2} - K^{-2}{\zeta_0}^{-2})\right) \leq A^{\frac{1}{2}}E^{\frac{1}{2}}\left|A^{-\frac{1}{4}}(S_N - N\zeta_0)\right|^4 E^{\frac{1}{2}}\left|(AN^{-2} - K^{-2}{\zeta_0}^{-2})\right|^2.$$

(3.17)

Utilizing (3.14), (3.15) and (3.17), we obtain from (3.16) that, as $A \to \infty$,

$$R_g(A) \leq A^{\frac{1}{2}}o(1),$$

and the theorem follows.

Now we consider the case when σ is also unknown. In this situation, we use the estimator of σ^2 as

$$\hat{\sigma}_n^{\,2} = n^{-1}\sum_{i=1}^{n}(Y_i - \bar{Y}_n)^2\,,$$

where $Y_i = \log X_i$ and

$$\bar{Y}_n = n^{-1}\sum_{i=1}^{n} Y_i .$$

Now we define the following stopping rule in conformity with (2.3). i.e.,

$$N = \inf\left[n \ge m : n \ge A^{\frac{1}{2}} \bar{X}_n \left(e^{\hat{\sigma}_n^2} - 1\right)^{1/2}\right]$$

$$= \inf\left[n \ge m : S_n \ge A^{-\frac{1}{2}} n^2 \left(e^{\hat{\sigma}_n^2} - 1\right)^{-\frac{1}{2}}\right], (3.18)$$

where 'm' is the starting sample size and is such that, for $\delta > 0$,

$$\delta A^{\frac{1}{4}} \le m = o\left(A^{\frac{1}{2}}\right) \text{ as } A \to \infty .$$

The 'risk-efficiency' of the sequential procedure (3.18) is defined by

$$R_e(A) = \frac{R_N(A)}{R_{n_o}(A)}, (3.19)$$

where $R_N(A) = A\,E\left(\bar{X}_N - \zeta\right)^2 + E(N)$. (3.20)

and $R_{n_o}(A) = 2KA^{\frac{1}{2}}\zeta$. (3.21)

In the following theorem, we prove that the sequential procedure (3.18) is 'asymptotically risk-efficient.'

Theorem 3 For the sequential procedure (3.18).

$$\lim_{A\to\infty} R_e(A) = 1.$$

Proof

Combining (3.19), (3.20) and (3.21), we can write

$$R_e(A) = \left(2KA^{\frac{1}{2}}\zeta\right)^{-1} E\left(A\left(\bar{X}_N - \zeta\right)^2\right) +$$

$$\left(2KA^{\frac{1}{2}}\zeta\right)^{-1} E(N)$$

$$= \left(2KA^{\frac{1}{2}}\zeta\right)^{-1}$$

$$E\left[\left(S_N - N\zeta\right)^2\left\{AN^{-2} - \zeta^{-2}\left(e^{\sigma^2} - 1\right)^{-1}\right\}\right] +$$

$$2\left(2KA^{\frac{1}{2}}\zeta\right)^{-1} E(N). (3.22)$$

From theorem 2 of Chow, Robbins and Teicher (1965), it is easy to see that

$$A^{-\frac{1}{2}}N \xrightarrow{a.s} \zeta\left(e^{\sigma^2} - 1\right)^{\frac{1}{2}} \text{ as } A \to \infty .$$

Under the condition imposed on the starting sample size 'm', it follows from lemma 3 of Martinsek (1983) that, for $p > 0$,

$$\left[\left(A^{-\frac{1}{2}}N\right)^{-p} : A \ge 1\right] \text{ is uniformly integrable. } (3.23)$$

Applying (3.23), we conclude that

$$\left(KA^{\frac{1}{2}}\zeta\right)^{-1} E(N) = \left(\zeta\left(e^{\sigma^2} - 1\right)^{\frac{1}{2}}\right)^{-1} E\left(A^{-\frac{1}{2}}N\right)$$

$\to 1$, as $A \to \infty$. (3.24)

From (3.22) and (3.24), we see that the theorem follows, if we can prove that

$$A^{-\frac{1}{2}}\, E\left[\left(S_N - N\zeta\right)^2\left\{AN^{-2} - \zeta^{-2}\left(e^{\sigma^2} - 1\right)^{-1}\right\}\right] =$$

$o(1)$, as $A \to \infty$. (3.25)

To this end, by Holder's inequality,

$$E\left[A^{-\frac{1}{2}}\left(S_N - N\zeta\right)^2\left(AN^{-2} - \zeta^{-2}\left(e^{\sigma^2} - 1\right)^{-1}\right)\right]$$

$$\le E^{\frac{1}{2}}\left|A^{-\frac{1}{4}}\left(S_N - N\zeta\right)\right|^4$$

$$E^{\frac{1}{2}}\left|\left(AN^{-2} - \zeta^{-2}\left(e^{\sigma^2} - 1\right)^{-1}\right)\right|^2$$

(3.26)

From lemma 5 of Chow and Yu (1981),

$$\left[\left|A^{-\frac{1}{4}}\left(S_N - N\zeta\right)\right|^4 : A \ge 1\right] \text{ is uniformly integrable.}$$

(3.27)

Since $AN^{-2} \xrightarrow{a.s} \zeta^{-2}\left(e^{\sigma^2} - 1\right)^{-1}$ as $A \to \infty$, we obtain from (3.26), on using (3.23) and (3.27), that

$$E\left[A^{-\frac{1}{2}}\left(S_N - N\zeta\right)^2\left(AN^{-2} - \zeta^{-2}\left(e^{\sigma^2} - 1\right)^{-1}\right)\right]$$

$$\le O(1)\, E^{\frac{1}{2}}\left|\left(AN^{-2} - \zeta^{-2}\left(e^{\sigma^2} - 1\right)^{-1}\right)\right|^2 .$$

$= o(1)$, as $A \to \infty$,

and (3.25) holds. This completes the proof of the theorem.

References

[1] Aitchison, J. and Brown, J. A. C. (1957): The Log-normal Distribution, ***Cambridge Unvi. Press.***

[2] Chattopadhyay, B. & Kelley, K. (2016): Estimation of the coefficient of variation with minimum risk: A sequential method for minimizing sampling error and study cost. ***Multivariate Behavioral Research, Vol.0, No.0, 1-22***

[3] Chow, Y. S. and Yu, K. F. (1981): The performance of a sequential procedure for estimation of the mean. ***Ann. Statist., 9, 184-189.***

[4] Chow, Y. S. and Teicher, H. (1978): Probability Theory. ***Spinger verlag, New York.***

[5] Chow, Y. S., Robbins, H. and Teicher, H. (1965): Moments of Randomly stopped sums. ***Ann. Math. Statist., 36, 789-799.***

[6] Gut, A. (1974): On the moments and limit distribution of some first passage time. ***Ann. Prob., 2, 308-277.***

[7] Martinsek, A. T. (1983): Second order approximation to the risk of a sequential procedure. ***Ann. Statist., 11, 827-836.***

[8] Starr, N. (1966 b): On the asymptotic efficiency of a sequential procedure for estimating the mean. ***Ann. Math. Statist., 37, 1137-1185.***

[9] Starr, N. and Woodroofe, N. (1969): Remarks on sequential point estimation. ***Proc. Nat. Acad. Sci., 63, 285-288.***