MINIMUM VARIANCE UNBIASED ESTIMATION OF THE SCALE PARAMETER OF EXPONENTIAL DISTRIBUTIONS AND RELATED LOGARITHMIC INTEGRALS

Roberto N. Padua[1]
https://orcid.org/0000-0001-5592-8867
Liceo de Cagayan University, Cagayan de Oro City, Philippines

Dionisel Y. Regalado[2]
Mindanao University of Science and Technology
Cagayan de Oro City

Kenneth P. Perez[3]
Department of Education
Tangub City Division

Rodel B. Azura[4]
Agusan del Sur State College of Agriculture and Technology
Bunawan , Agusan del Sur

ABSTRACT

The paper tackles two (2) problems related to the exponential distribution. The first concerns a detailed derivation of the minimum variance unbiased estimator of the scale parameter. The second focuses on the relationship of the expected value of the reciprocal of an exponential random variable which is shown to be equivalent to evaluating the logarithmic integral and the density of primes as found in the Prime Number Theorem. In the first problem, we showed that the minimum variance unbiased estimator of the scale parameter has a variance larger than the Cramer-Rao lower bound. In the second problem, we demonstrated that the expected value of the reciprocal of an exponential random variable also obtains the density of primes less or equal to a given large number x. The minimum variance unbiased estimator found in the first problem can then be utilized to find such an approximation to the density of primes for the second problem. The second problem provides a new way of viewing the problem of finding the density of primes less or equal to x.

Keywords: *unbiased estimator, minimum variance, exponential scale parameter, logarithmic integrals, density of primes*

1.0 Two Problems

A random sample of size n, $x_1, x_2, \dots, x_n$, is given from an exponential distribution:

$$f(x, \theta) = \theta e^{-\theta x}, \quad \theta > 0, x > 0 \tag{1}$$

It is wanted to find an unbiased estimator of θ with minimum variance. The maximum likelihood estimator (MLE) of θ can be achieved from the likelihood function:

$$L(x_1, \dots, x_n;\ \theta) = \theta^n e^{-\theta\ \sum_{i=1}^{n} x_i} \quad (2)$$

by setting the derivative equal to zero:

$$\frac{\partial L}{\partial \theta} = 0 \quad (3)$$

The maximum likelihood estimator of θ, namely,

$$\hat{\theta} = \arg\left(\max_{\theta} L(x,\theta)\right) = \frac{1}{\bar{x}} \quad (4)$$

where $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$, is a biased estimator of θ, since:

$$E(\hat{\theta}) = E\left(\frac{1}{\bar{x}}\right) \geq \frac{1}{E(\bar{x})} = \theta \quad (5)$$

by Jensen's inequality (Lehman, 1984). The first problem is to find an unbiased estimator of θ with minimum variance. This is not a new problem (see Rao, 2010) but is seldom discussed in statistical inference classes and, so, we deemed it important to provide a formal exposition of the solution.

The second problem, which has application in analytic number theory in relation to the topic on the density of primes, considers the exact instance of (1) for which $n = 1$. In particular, let $x \underset{\sim}{d} \exp(\theta)$ with density (1) and let $S = \frac{1}{x}$. The problem is to evaluate:

$$E\left(\frac{1}{x}\right) = \theta \int_0^{\infty} \frac{e^{-\theta x}}{x} dx \quad (6)$$

We show that (6) is closely related to a variant of the logarithmic integral:

$$\ell_i(u) = \int_0^{\infty} \frac{du}{\ln(u)} \quad (7)$$

The logarithmic integral, in turn, is related to the offset logarithmic integral which estimates the density of primes less or equal to x:

$$L_i(x) = \int_2^{x} \frac{du}{\ln(u)} = \ell_i(x) - \ell_i(2) \quad (8)$$

(Tao, 2013; Schoenfield, 1976).

2.0 Solution to the First Problem

From (2), we find that a sufficient statistic for estimating θ is given by:

$$S = \sum_{i=1}^{n} x_i \tag{9}$$

By using the moment – generating function (mgf) technique (Abramowitz et al., 1983)

$$M_s(t) = \left[m_{x_i}(t)\right]^n = \frac{1}{(1-\theta t)^n} \tag{10}$$

We recognize that (10) is the mgf of a Gamma Random Variable with $\alpha = n$, $\beta = \theta$, hence:

$$f_s(s) = \frac{\theta^n}{\Gamma(n)}\, s^{n-1} e^{-\theta s},\;\; s > 0, \theta > 0 \tag{11}$$

We calculate the expected value of s^{-1}:

$$E(s^{-1}) = \int_0^\infty \frac{\theta^{n-1}}{\Gamma(n)} s^{n-2} e^{-\theta s} ds \tag{12}$$

$$= \frac{\theta^n}{(n-1)} \int_0^\infty \frac{\theta^{n-1}}{\Gamma(n-1)} s^{n-2} e^{-\theta s} ds$$

$$= \frac{\theta}{n-1} \cdot 1$$

Since

$$\frac{\theta^{n-1}}{\Gamma(n-1)} s^{n-2} e^{-\theta s} \underset{\sim}{d} \Gamma(n-1, \theta)$$

It follows that an unbiased estimator of θ is:

$$\tilde{\theta} = \frac{n-1}{s} = \frac{n-1}{\sum_{i=1}^{n} x_i} = \left(\frac{n-1}{n}\right) . \frac{1}{\bar{x}} = \frac{n-1}{n} . \hat{\theta} \tag{13}$$

by the Lehman – Scheffe's theorem, $\tilde{\theta}$ is an unbiased estimator of θ which is a function of a complete and sufficient statistic s, $\tilde{\theta}$ has the smallest variance among all unbiased estimators of θ.

Theorem 1. Let $x_1, x_2, \ldots, x_n$ be independent and identically distributed random observations from an exponential distribution with scale parameter θ. The uniformly minimum variance unbiased estimator of θ is:

$$\tilde{\theta} = \frac{n-1}{n}.\frac{1}{\bar{x}}$$

where $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$.

Proof. The unbiasedness of $\tilde{\theta}$ is shown in (12). Sufficiency is established by the factorization theorem applied to (2). We need to demonstrate completeness:

$$\int_0^\infty z\,(t)\theta e^{-\theta t}dt = 0, \quad \theta > 0, t > 0$$

Then

$$\int_0^\infty z\,(t)\theta e^{-\theta t}dt = 0 \int_0^\infty z\,(t)\theta e^{-\theta t}dt = 0 \tag{14}$$

However, (14) implies $z(t) = 0 \quad \forall t$. The result follows by the Lehman – Scheffe's theorem. ■

It will be of interest to find the variance of $\tilde{\theta}$. For this, we require $E(\tilde{\theta}^2)$:

$$E\left(\tilde{\theta}^2\right) = (n-1)^2 E\left(\frac{1}{s^2}\right)$$

$$= (n-1)^2 \int_0^\infty \frac{\theta^n}{\Gamma(n)}\; s^{n-3} e^{-\theta s} ds$$

$$= \frac{(n-1)^2}{(n-1)(n-2)}.\theta^2 \int_0^\infty \frac{\theta^{n-2}}{\Gamma(n)}\; s^{n-3} e^{-\theta s} ds$$

$$= \frac{(n-1)}{(n-2)}\theta^2.1 \tag{15}$$

Hence:

$$var\left(\tilde{\theta}\right) = E\left(\tilde{\theta}^2\right) - \left[E\left(\tilde{\theta}\right)\right]^2$$

$$= \left(\frac{n-1}{n-2} - 1\right)\theta^2$$

$$var\left(\tilde{\theta}\right) = \frac{\theta^2}{n-2} \tag{16}$$

The Cramer – Rao lower bound can be calculated and compared to (16):

$$logL(x, \theta = n \log\theta - \theta \sum_{i=1}^{n} x_i)$$

$$\frac{\partial \log L}{\partial \theta} = \frac{n}{\theta} - \sum_{i=1}^{n} x_i$$

$$\frac{\partial^2 \log L}{\partial \theta^2} = \frac{-n}{\theta^2}$$

$$I(\theta) = E\left(-\frac{\partial^2 \log L}{\partial \theta^2}\right) = \frac{n}{\theta^2} \quad (17)$$

It follows that:

$$var\left(\tilde{\theta}\right)_{CR} = \frac{1}{\frac{n}{\theta^2}} = \frac{\theta^2}{n}$$
(18)

The variance of $\tilde{\theta}$ is larger than the Cramer – Rao lower bound, however, they become similar for large n.

3.0 Solution to the Second Problem.

In the second problem, we want to begin a connection between the exponential distribution (θ) and the logarithmic integral (7) which offers a good estimate to the number of primes less or equal to $x \; or \; \pi(x)$..

Cramer (1936) averred that the prime gaps x_n:

$$x_n = P_{n+1} - P_n \quad , n = 1, 2, \ldots \quad (19)$$

have an estimated exponential distribution with $\theta = \log(n)$. The prime gaps $\{x_n\}$ can be viewed as arrival times (of primes) and so the reciprocal $\{x_n\}_{n=1}^{\infty}$ can be interpreted as rates of arrival i.e. one arrival per time. If the density of prime gaps is:

$$f_{x_n}(x) = \theta e^{-\theta x} \qquad , x > 0, \theta > 0$$
(20)

then, we are concerned on the behavior of $\left\{\frac{1}{x_n}\right\}_{n=1}^{\infty}$. In particular,

$$E\left(\frac{1}{x}\right) = \theta \int_0^{\infty} \frac{e^{-\theta x}}{x} dx \quad (21)$$

Let $t = e^{-\theta x}$, then (21) becomes:

$$\lim_{x\to\infty} \theta \int_0^{x} \frac{e^{-\theta x}}{x} dx = \lim_{x\to\infty} \theta \int_0^{t} \frac{dt}{\ln(t)} = \lim_{t\to\infty} [l_i(t)] \quad (22)$$

The logarithmic integral $\int_0^\infty \frac{dt}{\ln(t)}$ is divergent but the asymptotic expansion:

$$l_i(t) = \int_0^t \frac{dt}{\ln(t)} \sim \frac{t}{\ln(t)} \sum_{k=0}^{\infty} \frac{k!}{(\ln(t))^k} \tag{23}$$

is valid for large t. Avoiding the singularity at $t = 0$, we introduce the offset logarithmic integral:

$$l_i(t) = \int_2^t \frac{du}{\ln(u)} = l_i(t) - l_i(2) \tag{24}$$

$$\sim \frac{t}{\ln(t)} \sum_{k=0}^{\infty} \frac{k!}{(\ln(t))^k} = \frac{t}{\log t} + \frac{t}{(\log t)^2} + \frac{2t}{(\log t)^3} + \frac{6t}{(\log(t))^4} + \cdots \text{ for } t \geq 2$$

The number of primes less than or equal to t, denoted by $\pi(t)$, is given by the Prime Number Theorem (PNT) as:

$$\pi(t) \sim \frac{t}{\ln(t)} \tag{25}$$

proven independently by Hadamard and dela Vallee Poussin (1896). Van Koch (1901) proved that:

$$|\pi(t) - L_i(t)| = 0(\sqrt{t} \log t) \tag{26}$$

and the constant in the big 0 notation was estimated by Schoenfeld (1976) to be:

$$|\pi(t) - l_i(t)| < \frac{\sqrt{t} \log t}{8\pi} \tag{27}$$

The prime number theorem (PNT) assumes that primes are uniformly distributed on intervals of length **t** with probability $\frac{1}{\ln(t)}$. Hence, the estimated number of primes on such intervals is $\frac{t}{\ln(t)}$. On the other hand, if the prime gaps x_n are expected to obey an exponential distribution with parameter $\theta = \log(t)$, then the expected number of primes less or equal to t is precisely $E\left(\frac{1}{x_n}\right)$, i.e. the expected number of primes per arrival time is $E\left(\frac{1}{x_n}\right)$.

On the Density of $\frac{1}{x_n}$. The density of $y = \frac{1}{x_n}$ is useful for estimating the higher order moments of y.

Let $= \frac{1}{x_n}$, then:

$$F_y(y) = P(Y \le y) = P\left(\frac{1}{x_n} \le y\right)$$

$$= P\left(\frac{1}{y} \le x_n\right) = P\left(x_n \ge \frac{1}{y}\right)$$

$$= 1 - P\left(x_n < \frac{1}{y}\right)$$

$$1 - \left(1 - e^{-\frac{\theta}{y}}\right)$$

$$F_y(y) = e^{-\frac{\theta}{y}} \quad , y > 0 \tag{28}$$

The density is found by differentiation:

$$f_y(y) = \frac{\theta}{y^2} e^{-\frac{\theta}{y}}, \; y > 0 \tag{29}$$

On an Estimate for the Density of Primes.

Using the results of the previous section, we find that an estimator for the density of primes less or equal to x is:

$$\pi(x) \sim \left(\frac{n-1}{n}\right) . \frac{1}{\bar{x}} \;\; l_i(x) \tag{30}$$

$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ is the sample mean of all prime gaps in the interval $[1, x]$.

ACKNOWLEDGMENT

The authors wish to acknowledge the assistance of Dr. Roberto N. Padua, consultant and professor of Mindanao University of Science and Technology and Liceo de Cagayan University in formulating the problems and in carefully examining the results obtained.

References:

Abramowitz, Milton; Stegun, Irene Ann, eds. (1983) [June 1964]. "Chapter 5". Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables. Applied Mathematics Series **55** (Ninth reprint with additional corrections of tenth original printing with corrections (December

1972); first ed.). Washington D.C., USA; New York, USA: United States Department of Commerce, National Bureau of Standards; Dover Publications. p. 228.

Granville, Andrew (1995). "Harold Cramér and the distribution of prime numbers" (PDF). Scandinavian Actuarial Journal 1: 12–28.

Hardy, G. H. & Littlewood, J. E. (1916). "Contributions to the Theory of the Riemann Zeta-Function and the Theory of the Distribution of Primes". Acta Mathematica 41: 119–196.

Lehman, P. (1984). Theory of Point Estimation. Wiley Series, New York, p 151-153.

Padua, R. N., & Libao, M. F. (2017). ON STOCHASTIC APPROXIMATIONS TO THE DISTRIBUTION OF PRIMES AND PRIME METRICS. *Journal of Higher Education Research Disciplines*, *1*(1), 33-38.

Rao, S. (2010) . Advanced Statistical Inference. (Unpublished Lecture Notes, 7-10).

Regalado, D. Y. & Azura, R. B. "ON A MIXED REGRESSION ESTIMATOR FOR DENSITY OF PRIME GAPS". *Journal of Higher Education Research Disciplines*, 1(2), 48 – 57, 2017.

Schoenfeld, Lowell (1976). "Sharper Bounds for the Chebyshev Functions θ(x) and ψ(x). II". Mathematics of Computation **30** (134): 337–360.

Tao, T. (2013). "Structure and Randomness in the Distribution of Primes" (Video Lectures, University of California)

Temme, N. M. (2010), "Exponential, Logarithmic, Sine, and Cosine Integrals", in Olver, Frank W. J.; Lozier, Daniel M.; Boisvert, Ronald F.; Clark, Charles W., NIST Handbook of Mathematical Functions, Cambridge University Press

Von Koch, Helge (1901). "On the distribution of prime numbers". Acta Mathematica (in French) **24** (1): 159–182