EUROPEAN
ACADEMIC
RESEARCH

Two-Step Sequential Procedure for the Point Estimation of the Exponential Scale Parameter

EDNA P. ARANGCON – CENITA, M.S.
Southern Philippines Agribusiness, Marine and Aquatic School of Technology
Malita, Davao del Sur
The Philippines
DAISY LOU LIM – POLESTICO, PHD
Mathematics Department
College of Science and Mathematics
MSU-Iligan Institute of Technology, Iligan City
The Philippines

Abstract:

To estimate the mean μ and variance σ^2 of an exponential distribution, Moors and Strijbosch [10] proposed a stopping rule C for which the observed sample variance is the extension criterion whether to double the initial sample size n at the second stage. A modified two-step sequential procedure for the point estimation of the exponential scale parameter with the inclusion of the cost function to the observed sample variance is proposed. As a measure of accuracy of the procedure, the bias and mean-squared error of the estimates are computed. Also shown is the loss of accuracy due to fixed sample size and estimate of the optimum sample size at stopping.

We derived the estimates and the properties of their mean and the variance. Biases remain in estimating the standard estimators. However, a two-stage sequential sampling plan is recommended to improve accuracy of the estimates of sample of fixed size.

Through simulations, it was found out that with a specific bound of the cost function c, the proposed two-step sequential scheme seems to be an appropriate procedure in the point estimation of the exponential scale parameter.

Simulations using the developed Scilab programs are included to illustrate the performance of the procedures in the study.

Key words: Two-Step Sequential Procedure, Point Estimation, Exponential Scale Parameter

1 Introduction

Deciding on the optimal sample size in advance may result to two possibilities; underestimation of the population parameter being estimated or overestimation of the estimates. When n is too small, the precision of the estimate can be low, depending on the variance. When n is too large, we use unnecessary large number of sample, and face the problem of costly collection of data. Sequential sampling has been developed to address this major problem in sampling. Moors and Strijbosch (2002) studied the two-step procedure in estimating point estimates for gamma distribution. Variance and mean squared error of $\overline{y}$ as accuracy measures has been explored. In principle, the basic idea to propose such two-stage sequential sampling by Moors and Srijbosch was deeply motivated by the works of Stein (1945), Anscombe (1952) and Chow and Robbins (1965). One may look at Mukhopadhyay et. al. (1994) for a review of sequential estimation theory and method.

In Section 2, we revisit Moors and Strijbosch derivations of all expectations and variances. We propose in Section 3 the stopping variable with the inclusion of cost per unit observation and give formulas for its expected value and variance. In Section 4, we present the expected value, bias, variance and mean squared error of $\overline{y}$ at stopping. In Section 5, we estimate the variance, bias of the variance, MSE and expectation of $\dfrac{s^2}{n}$ as measure of accuracy. We illustrate with numerical examples that theoretical value is in agreement with the simulated value.

In Section 6, we present a numerical analysis of Moors and Strijbosch stopping variable and the proposed stopping rule. It is concluded that the modified two-step procedure is a generalization of Moors and Strijbosch procedure for the case of the exponential distribution. It is deduced that when the bound of cost function $c > 0.75$, the procedure is effective and recommended.

2 Preliminary result

Let $X_1, X_2, ..., X_n$ be independent and identically distributed random variables according to an exponential distribution having the probability density function

$$f(x) = \frac{1}{\sigma} \exp\left(-\frac{x}{\sigma}\right) I_{\{x>0\}}(x), \ (2.1)$$

where the scale parameter $\sigma \in (0, \infty)$ is unknown.

Suppose we take a random sample size $n \geq 2$ independent observations x and z from a negative exponential distribution with $\mu = \sigma^2 = 1$. The joint density $f_{x,z}(x, z)$ is given by

$$f_{x,z}(x, z) = e^{-x-z}, \quad x > 0, \ z > 0$$
$$= e^{-(x+z)}.$$

With the transformation for the mean $u = (x + z)/2$, $y = x - z$ becomes the deviation between observations.
The Jacobian matrix J of this transformation is

$$J = \begin{pmatrix} \dfrac{\partial u}{\partial x} & \dfrac{\partial u}{\partial z} \\ \dfrac{\partial y}{\partial x} & \dfrac{\partial y}{\partial z} \end{pmatrix}$$

$$|J| = 1$$

Then the joint density $q(u, y)$ becomes

$$q_{u,y}(u,y) = e^{-2u}, |y| < 2u$$

and the conditional densities are derived and given as follows

$$q_1(u) = 4ue^{-2u}, u > o$$

$$q_2(y|u) = \frac{1}{4u}, |y| < 2u \qquad (2.2)$$

Now, from the mean $u = \dfrac{x+z}{2}$, we have

$$s^2 = \frac{y^2}{2}$$

and

$$P(s^2 \leq s^2 | u) = 2P\left(0 < y < \frac{\sqrt{2s^2}}{2u}\right) = \frac{\sqrt{2s^2}}{2u}, o < s^2 < 2u^2 \quad (2.3)$$

Consequently, it follows to derive the conditional density, marginal density, and conditional expectation.

Differentiating (2.3) with respect to the sample variance s^2 gives us the conditional density given as

$$g_2(s^2 | u) = \frac{1}{u\sqrt{8s^2}}, 0 < s^2 < 2u^2$$

The marginal density of s^2 can be obtained as

$$g(s^2) = \frac{1}{2s^2} e^{-\sqrt{2s^2}}, s^2 > 0$$

and the conditional distribution of μ and s^2

$$g(u | s^2) = 2e^{\sqrt{2s^2} - 2u}, 0 < s^2 < 2u^2$$

Consequently, conditional expectation of s^2 given u is derived.

$$E(u | s^2) = \frac{1 + \sqrt{2s^2}}{2}$$

3. **The Two - Step Sequential Procedure**

Given a random sample $X_1, X_2, ..., X_n$ from an exponential population defined in (2.1), we will consider the stopping variable

$$C^* = \left\{ s^2 + cn_1 > w \right\} \text{ or}$$

$$C^* = \left\{ s^2 > w - cn_1 \right\}, w - cn_1 > 0$$

where w is a predetermined set bound variance on $s^2 + cn_1$ and n_1 is the initial sample size. The preselected bound variance $w > cn_1$ is chosen in such a way as to achieve some type of balance between sample variability and sampling cost. Motivated by (3.1), the following is the proposed two-step procedure.

First step. Take a random sample of fixed size n from an exponential population defined in (2.1). We compute the sample variance $s_{n_1}^2 = (n_1 - 1)^{-1} \sum_{i=1}^{n_1} \left(x - \overline{x}_{n_1} \right)$. If it is observed from initial observation that

$$\left(C^* \right)' = \left\{ s^2 < w - cn_1 \right\} \quad (3.2)$$

with the inclusion of the cost per unit sample cn_1; we terminate sampling; otherwise we proceed to the second step.

Second Step. If the extension criterion holds, we proceed by recording additional observation, that is, take $X_{n+1}, X_{n+2}, ..., X_{2n}$ sample units until $2n$ sample is observed. Thus the final sample size n, sample mean $\overline{y}$ and sample variance s^2 is defined

$$\left(n, \overline{y}, s^2 \right) = \begin{cases} \left(n_1, \overline{y}_1, s_1^2 \right. & if & \left(C^* \right)' \\ n_3, \overline{y}_3, s_3^2 & if & C^* \end{cases} \quad (3.3)$$

Consequently, it follows that

$$n_3 = n_1 + n_2$$

$$\overline{y}_3 = \left(n_1 \overline{y}_1 + n_2 \overline{y}_2 \right) / n_3 \tag{3.4}$$

$$\left(n_3 - 1 \right) s_3^2 = \left(n_1 - 1 \right) s_1^2 + \left(n_2 - 1 \right) s_2^2 + n_1 n_2 \left(\overline{y}_1 - \overline{y}_2 \right)^2 / n_3$$

Now, the probability that C^* occurs is $P\left(C^* \right) = e^{-\sqrt{2(w - cn_1)}}$. The conditional expectation of the mean given the extension criterion as computed as follows

$$E\left(u \mid C^* \right) P\left(C^* \right) = \int_{w - cn_1}^{\infty} E\left(u \mid s^2 \right) g_1 \left(s^2 \right) ds$$

$$= \frac{2\sqrt{2(w - cn_1)}}{2} p$$

and

$$E\left(u^2 \mid C^* \right) P\left(C^* \right) = \int_{w - cn_1}^{\infty} E\left(u^2 \mid s^2 \right) g_1 ds^2$$

$$= \frac{\left(3 + \sqrt{2(w - cn_1)} + w - cn_1 \right) p}{2} \tag{3.5}$$

Noting the consequences of (2.5) and (3.4)

$$Var\left(u \mid s^2 \right) = \frac{1}{4}$$

$$Var\left(u \mid C^* \right) = \frac{1}{2}$$

This result would indicate that variance of the sample mean u and variance s^2 is independent of s^2 and the set bound $w - cn_1$.

4. Estimation of the Mean

Now, to estimate the mean μ and its measure of accuracy and precision, the expectation of (3.3) will yield the results given in proposition that follows:

Proposition 4.1: For $p = e^{-\sqrt{2(w - cn_1)}}$, with $w > cn_1$

i. $\quad E\left(\bar{y}\right) = \dfrac{4 - p\sqrt{2\left(w - cn_1\right)}}{4}$

ii. $\quad B\left(\bar{y}\right) = \dfrac{-p\sqrt{2\left(w - cn_1\right)}}{4}$

iii. $\quad Var\left(\bar{y}\right) = \dfrac{1}{2} - p\dfrac{3\left(w - cn_1\right) + \left(w - cn_1\right)p + 2}{8}$

iv. $\quad MSE\left(\bar{y}\right) = \dfrac{1}{2} - p\left(3\left(w - cn_1\right) + 2\right)/8$

Proof: (i) Getting the expected value of (3.3), we have

$$E\left(\bar{y}\right) = E\left(\bar{y} \mid C'\right)P\left(C'\right) + E\left(\bar{y}_3 \mid C\right)P\left(C\right)$$

and

$$0 = E\left(\bar{y}_1\right) = E\left(\bar{y}_1 \mid C'\right)P\left(C'\right) + E\left(\bar{y}_1 \mid C\right)P\left(C\right)$$

so that

$$E\left(\bar{y}\right) = E\left(\bar{y}_1\right) - E\left(\bar{y}_1 \mid C\right)P\left(C\right) + \frac{1}{2}E\left(\bar{y}_1 + \bar{y}_2 \mid C\right)P\left(C\right)$$

$$= E\left(\bar{y}_1\right) - E\left(\bar{y}_1 \mid C\right)P\left(C\right) + \frac{1}{2}\left[E\left(\bar{y}_1 \mid C\right)P\left(C\right) + E\left(\bar{y}_2 \mid C\right)P\left(C\right)\right]$$

$$= E\left(\bar{y}_1\right) + \frac{E\left(\bar{y}_2 \mid C\right)P\left(C\right) - E\left(\bar{y}_1 \mid C\right)P\left(C\right)}{2}$$

$$E\left(\bar{y}\right) = E\left(\bar{y}_1\right) + \frac{1}{2}E\left(\bar{y}_2 - \bar{y}_1 \mid C\right)P\left(C\right)$$

From (3.5), the expectation of the estimator $\bar{y}$ is defined,

$$E\left(\bar{y}\right) = \frac{4 - p\sqrt{2\left(w - cn_1\right)}}{4}, \qquad \text{w-cn}_1 > 0.$$

A negative bias remains in estimating the mean μ, which implies underestimation of the parameter being estimated.

5 Estimation of the Variance

The expectation of s_1^2 can be found similarly. Taking the conditional expectation of (3.3) and applying results of (3.5). However, since only high values of s_1^2 lead to extension of the original sample, it may be expected that s^2 is negatively biased as shown in the next proposition.

Proposition 5.1 For $w > cn_1$, and $v = \dfrac{s^2}{n}$, we have

i.) $E(s^2) = 1 - \left(\dfrac{3p(w - cn_1) - 4p\sqrt{w - cn_1}}{6} \right)$;

ii.) $B(s^2) = -\dfrac{p}{6}\left[3(w - cn_1) + 4\sqrt{2(w - cn_1)} \right]$;

iii.) $E(v) = \dfrac{1}{2} - \dfrac{p}{24}\left[9(w - cn_1) + 10\sqrt{2(w - cn_1)} + 6 \right]$; and

iv.) $B(v) = -\dfrac{5p}{12}\sqrt{2(w - cn_1)}$

Proof: (*i*) To show $E(s^2)$, where s^2 is the variance of the final sample, perform the expectation of (3.3) and gives

$$E(s^2) = E(s_1^2 \mid C')P(C') + E(s_3^2 \mid C)P(C),$$

s_1^2 is defined as the variance of the initial sample. With the relation

$$0 = E(s_1^2) = E(s_1^2 \mid C')P(C') + E(s_1^2 \mid C)P(C),$$

so that

$$E(s_1^2 \mid C')P(C') = E(s^2) - E(s_1^2 \mid C)P(C).$$

By substitution, we come up with

$$E(s^2) = E(s_1^2) - E(s_1^2 \mid C)P(C) + E(s_3^2 \mid C)P(C)$$
$$= \sigma^2 + E(s_3^2 \mid C)P(C) - E(s_1^2 \mid C)P(C)$$
$$= \sigma^2 - E(s_3^2 - s_1^2 \mid C)P(C)$$

From (3.3), we define
$$(n_3 - 1)s_3^2 = (n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + n_1 n_2 (\bar{y}_1 - \bar{y}_2)^2 / n_3. \text{ So that,}$$

$$E(s^2) = \sigma^2 - E\left[(n-1)s^2 + (n_2 - 1)s_2^2 + n_1 n_2 (\bar{y}_1 - \bar{y}_2)^2 - s_1^2 \mid C\right]P(C)$$

$$= \sigma^2 - E\left[s_2^2 + (\bar{y}_1 - \bar{y}_2)^2 - s_1^2 \mid C\right]P(C)$$

$$= 1 - \left(\frac{3p(w - cn_1) + 4\sqrt{2(w - cn_1)}}{6}\right).$$

The simulated and theoretical values of the estimator $\bar{y}$ for μ yield closer estimates to the parameter being estimated. Overall agreement between the simulated and the true value is observed for the estimator s^2 for σ^2 in Table 4.1.

Table 4.1 Numerical Analysis of the Proposed Sequential Procedure: Simulation and Theoretical Results when $n=500$, $\mu = \sigma^2 = 1$

	simulated	TRUE	simulated	TRUE	simulated	TRUE	simulated	TRUE
	$c = 1$		$c = 0.1$		$c = 0.01$		$c = 0.0001$	
$E(\bar{y})$	0.9918	0.9909	0.99437	0.9946	0.99478	0.99334	0.9998	0.9965
$B(\bar{y})$	−0.0082	−0.0091	−0.0056	−0.0054	−0.0052	−0.0067	−0.0002	−0.00349
$E(s^2)$	0.9703	0.9534	0.9822	0.9745	0.9799	0.9904	0.9994	0.991
$B(s^2)$	−0.02973	−0.0466	−0.0178	−0.0254	−0.0020	−0.0096	−0.00500	−0.0093

6 Simulation Results

The summary of simulation results are displayed in Table 6 of the two-step sequential sampling design for the parameter values $= \sigma^2 = 1$. As the cost is allowed to approach zero, the simulation values are derived from the stopping rule of this study while the simulation results for the column when $c = 0$, are simulation values obtained from the stopping rule of Moors and Strijbosch. The developed Scilab program runs the values when $c = 0$.

The estimates of the parameter as cost nears zero will approach the simulated values when $c = 0$ for all sample sizes.

However, the estimates are closer to the parameter being estimated for larger sample sizes. Furthermore, as the sample size increases, the rate of convergence is faster to the parameter being estimated. Results from t-test at α=0.05 revealed that when $c \leq 0.75$, either of the Moors and Strijbosch procedure or the proposed two – step sequential procedure is applicable . The cost in this case is not negligible but the performance of the procedure is not statistically significant. However, at $c > 0.75$ we cannot neglect anymore the cost of each observation. The proposed two-step sequential procedure should be used.

Table 6 Numerical Analysis of the Proposed Sequential Procedure as $c \rightarrow 0$ vs. $c = 0$ for $E(\bar{y})$

$E(\bar{y})$	c=1	c=0.90	c=0.75	c=0.5	c=0.01	c=0.001	c=0.0001	c=0
n=2	0.88477	**0.87804**	0.88613	0.89044	0.90835	0.91322	0.91389	**0.93312**
n=4	0.89352	**0.90203**	0.90541	0.90886	0.92404	0.92546	0.93318	**0.94457**
n=9	0.92901	**0.91753**	0.93645	0.93911	0.94103	0.94586	0.94880	**0.96454**
n=16	0.94086	**0.94688**	0.94858	0.95021	0.95985	0.96256	0.96376	**0.98008**
n=25	0.95766	**0.95719**	0.95599	0.95812	0.96766	0.96674	0.97133	**0.98167**
p-value	**0.047**	**0.048**	**0.07ns**	**0.085ns**	**0.553ns**	**0.591ns**	**0.840ns**	

ns – not significant

REFERENCES

[1] Lim, D.L. Isogai E .and Uno C. *On the Sequential Estimation of Functions of Exponential Scale Size Parameters*, (2006).

[2] Moors, J.J.A and Strijbosch, L.W.G. *Two - Step Sequential Sampling for Gamma Distributions*, Discussion Paper (2002) No. 2002-78.

[3] Mukhopadhyay, N. and Zacks, S. *Distributions of Sequential and Two-Stage Stopping Times for Fixed-Width Confidence Intervals in Bernoulli Trials: Application in Reliability*, Sequential Analysis, 26:4(2007), 425- 441.

[4] SCILAB. Copyright. 1989 - 2010. Digiteo. Scilab is trademark of INRIA. 25 May 2010. http://www.scilab.org.