Volume 3 Issue 1 May 2012
The MINDANAWAN
Journal of Mathematics
ISSN 2094-7380

A Proposed Two-Stage Sequential Point Estimation of the Hazard Rate Function

Aljo Clair P. Pingal, Daisy Lou L. Polestico
Department of Mathematics, College of Science and Mathematics, MSU-Iligan Institute of Technology, Iligan City
emal@mail.com

Abstract: We consider the problem of minimum risk point estimation of the exponential hazard rate. We assume that $\sigma > \sigma_L$, where $\sigma_L > 0$ is known to the experimenter from past experiences. A modified two-stage procedure is proposed which is shown to possess some first-order properties and is risk efficient. Simulation results are given and results indicate some small sample behaviour and provide support for the asymptotic behaviour of the sequential procedure as the cost c goes to zero. Hence, the proposed two-stage sequential procedure appears to be effective and useful.

1 Introduction

In sampling, deciding the optimal sample size in advance is a difficult problem. In most cases, researchers take a large sample size to assure that the population is well represented, and face the problem of a very costly collection of data. Often times, even if all the sample elements are drawn at random, it does not guarantee that the sample taken is a random sample and the estimates obtained are satisfactory. Sequential analysis has been developed in conquest of this major problem in sampling. The goal of this paper is to propose a two-stage procedure on estimating the exponential hazard rate.

Let $X_1, X_2, \ldots$ be a sequence of i.i.d. random variables having the following probability density function

$$f(x;\sigma) = \frac{1}{\sigma}e^{-x/\sigma},\ x > 0, \tag{3}$$

where the scale parameter $\sigma \in (0,\infty)$ is unknown. Taking samples of size n from an exponential population, we estimate $\theta = \dfrac{1}{\sigma}$ by $\hat{\theta} = \dfrac{1}{\overline{X}_n}$, where $\overline{X}_n = \dfrac{1}{n}\sum_{i=1}^{n} X_i$ is an estimate of σ.

Now, consider the loss function

$$L(\hat{\theta}_n) = (\hat{\theta}_n - \theta)^2 + cn, \tag{4}$$

where $c > 0$ is a known cost per unit sample. The risk as the expected loss, is given by

$$R_n(c) \quad = \quad E[L(\hat{\theta}_n)] = E[(\hat{\theta}_n - \theta)^2] + cn. \tag{5}$$

The risk $R_n(c)$ is approximately minimized at

$$n = n^* = \frac{1}{\sigma\sqrt{c}} \tag{6}$$

However since σ is unknown, so the optimal fixed sample size n^* is expected to be unknown. Uno et al. [3] showed a fully sequential procedure for the estimation of the exponential hazard rate $\theta = \frac{1}{\sigma}$. In this paper,we develop a two-stage procedure for the estimation of $\frac{1}{\sigma}$, we do this by first estimating the optimal fixed sample size n^*.

2 Proposed Two-Stage Sequential Procedure

Motivated by (6) the following is the proposed two-stage procedure for $\theta = \frac{1}{\sigma}$.

Stage 1: Under the assumption that $\sigma > \sigma_L$, $\sigma_L > 0$ is known from past experiment, take a random sample of size m, that is $X_1, X_2, X_3, \ldots, X_m$ where m is defined by

$$m = m_c = \max\left\{m_0, \left\lfloor \frac{1}{\sqrt{c}\cdot\sigma_L} \right\rfloor + 1\right\}, \qquad (7)$$

and $m_0 \geq 1$ is a fixed integer and $\lfloor x \rfloor$ is the greatest integer less than x. The initial sample size m is determined by (7).

Stage 2: Solve for the sample mean $\overline{X}_m$. Then the optimal fixed sample size that minimizes the risk is estimated by

$$N = N_c = \max\left\{m, \left\lfloor \frac{1}{\sqrt{c}\cdot\overline{X}_m} \right\rfloor + 1\right\}. \qquad (8)$$

This procedure was based on the proposed two stage sequential procedure of functions of the exponential scale parameter. Below are the the interesting properties of the two-stage sequential procedure.

Theorem 2.1 *If N is defined as in* (6), *then* $P(N < \infty) = 1$.

Theorem 2.1 assures that the sequential procedure would eventually stop. Thus, as we take observation at the second stage, there will be a finite sample size N as given in (6) and that taking samples of size N sequentially will result to a minimum risk in the estimation of $\theta = \frac{1}{\sigma}$.

Theorem 2.2 *(First-order Asymptotic Efficiency) If N is defined as in* (6)*, then as $c \to 0$*

(i) $\dfrac{N}{n^*} \xrightarrow{a.s.} 1$*; and*

(ii) $\displaystyle\lim_{c\to 0} \frac{E(N)}{n^*} = 1.$

Theorem 2.3 *If N is defined as in* (6)*, then the proposed two-stage procedure is asymptotically risk efficient, that is,*

$$\lim_{c\to 0} \frac{R_N}{R_{n^*}} = 1.$$

Theorem 2.2 shows that with the proposed two-stage procedure, as c approaches 0 the ratio between N and the optimal sample size n^* approaches 1 and the ratio between the expectation of N and the optimal sample size n^* also approaches 1. While theorem 2.3 shows that the risk using the two stage procedure converge to the risk of the optimal sample size n^* as c approaches 0, that is the estimates are asymptotically risk efficient.

3 Simulation Results

In this section, a simulation study is performed in order to verify some of the desirable properties of the proposed two-stage sequential procedure and to see the performance of the two stage procedure and the asymptotic behavior of the risk. The simulations have been carried out using SCILAB. The results of the simulation, based on the proposed two-stage sequential procedure for $g(\sigma) = \frac{1}{\sigma}$, are summarized in Table 1 which contain the average estimate of σ, the average stopping time, $E(N)$, and the ratio of $E(N)$ and n^*.

Consider the hazard rate $\frac{1}{\sigma}$ with $\sigma = 1$. Let c be sufficiently small, such that $n^* = \sigma^{-1}c^{-1/2} = 20, 50, 100, 200, 500$ and 1000. Set $\sigma_L = 0.8$, $m_0 = 2$ and consider $10,000$ repetitions by means of the stopping rule N in (6).

Table 1: Simulation Results for the Estimation of the Exponential Hazard Rate using the Proposed Two-Stage Sequential Procedure with $\sigma = 1$,$m_0 = 2$, and $\sigma_L = 0.8$

n^*	20	50	100	200	500	1,000
c	0.0025	0.0004	0.0001	0.000025	0.000004	0.000001
$E(N)$	22.1189	52.00072	101.7401	201.6937	501.7868	1002.0745
$E(\frac{1}{\overline{X}_n})$	1.0327487	1.0163115	1.0075063	1.0036208	1.0022869	1.0013599
$E(N)/n^*$	1.105945	1.040144	1.017401	1.0084685	1.035736	1.0020745
R_N/R_{n^*}	0.9907801	0.9909454	0.99488	1.016265	0.9976519	0.9991851

In Table 1, observe that the estimate $E(\frac{1}{\overline{X}_N})$ converges to the corresponding true value $\frac{1}{\sigma} = 1$ as $c \to 0$.Also, note that both $E(N)$ are based on the averages of the observed values given by (6) for $10,000$ independent experiments for each selected values of c. As depicted in table 1, $E(N)$ converges to n^* as the sampling cost per observation, c, gets smaller, that is the ratio of $E(N)$ and n^* converges to 1 as $c \to 0$ which implies that the procedure is first-order efficient and also the ratio R_N/R_{n^*} converges to 1 as $c \to 0$ which implies that the procedure is asymptotically risk efficient. These are desirable properties of N and its resulting estimates.

On the choice of σ_L, Mukhopadhay and Duggan [2] suggested that σ_L is at least 20 percent away from σ. For the chosen $\sigma_L = 0.8$ with $\sigma = 1$, $E(N)$ is observed to have a fast rate of convergence to n^*, as $c \to 0$.

References

[1] **Mukhopadhyay, N. and Duggan, W.T.** (1997a) *Can a two-stage procedure enjoy second-order properties?* Sankhya, Series A, 59, 435-448.

[2] **Mukhopadhyay, N and Duggan W.T.** (2001). *A two-stage point estimation procedure for the mean of an exponential distribution and second-order results*. Statistics and Decisions. 19, 155-171.

[3] **Uno, C., Isogai, E., and Lim, D.L.** (2004). *Sequential Point Estimation of a Function of the Exponential Scale Parameter*. Austrian Journal of Statistics, Vol. 33, No. 3, 281-291.