10th National Convention on Statistics (NCS)
EDSA Shangri-La Hotel
October 1-2, 2007

Sequential Confidence Intervals for The Exponential Hazard Rate

by

Junvie M. Pailden and Daisy Lou L. Polestico

For additional information, please contact:

Author's name	:	Junvie M. Pailden
Designation	:	Student, MS Statistics
Affiliation	:	MSU-Iligan Institute of Technology College of Science and Mathematics
Address	:	Iligan City, Philippines
Tel. no.	:	(063) 221-4068; (063917) 716-2941
E-mail	:	

Co-Author's Name	:	Daisy Lou L. Polestico
Designation	:	Faculty Member
Affiliation	:	MSU-Iligan Institute of Technology College of Science and Mathematics
Address	:	Iligan City, Philippines
Tel. no.	:	(063) 221-4068 /(063917) 716-2941
E-mail	:	

Sequential Confidence Intervals For The Exponential Hazard Rate

by

Junvie M. Pailden and Daisy Lou L. Polestico

ABSTRACT

Let $X_1; X_2\ldots$ be independent and identically distributed random variables from an exponential distribution with unknown scale parameter $\sigma \in (0, \infty)$. Given $d > 0$ and $\alpha \in (0, 1)$, we construct confidence intervals I_n for the hazard rate $\theta = 1/\sigma$, with length $2d$ and coverage probability $1 - \alpha$, based on random samples of size n from the exponential population. Since σ is unknown, then it is known that no best fixed sample size procedure can be used to estimate the unknown hazard rate. Thus, we propose a fully sequential procedure to estimate the optimum sample size n^* and consequently, the hazard rate θ. The proposed stopping rule is motivated by finding the sample size so that the coverage probability of the sequential confidence intervals is at least $1 - \alpha$, that is, $P(\theta \in I_n) \geq 1 - \alpha$. In this paper, we derive first-order approximations of the expected sample size and improves the procedure by developing second-order expansions of the rate of convergence of the coverage probability.

I. Introduction

Sequential estimation refers to techniques for estimating parameters when the sample size is not fixed in advance but is determined during the course of the experiment by some criteria which depend on the observations as they occur. This criterion is often called the *stopping rule* and means that sampling is continued until this criterion is satisfied. In a seminal paper, Chow and Robbins [2] proposed a fully sequential procedure in finding a confidence interval of prescribed width and prescribed coverage probability when the underlying distribution function is fixed but unknown. They showed that their proposed sequential procedure is *asymptotically consistent* and *asymptotically efficient*, which are desirable properties in any experiment. These salient properties have attracted many researchers in the field and since then substantial attention have been placed in improving fully sequential procedures. In a recent paper, Lim et al. [3] proposed a fully sequential procedure in constructing confidence intervals for functions of two unknown exponential scale parameters. They have been successful in showing the asymptotic consistency of the procedure and obtained a second-order approximation of the expected sample size. In this present paper, we shall simplify the results in [3] to the one-sample case and consider the estimation of the exponential hazard rate which is the reciprocal of the exponential scale parameter.

The *exponential distribution* is often used to model the time between independent events that happen at a constant average rate. In application, the exponential distribution can be used to model lifetimes of various practical situations including but not limited to lengths of time between successive catastrophic events (floods, earthquakes and so on) and lengths of time between emergency arrivals at a hospital, to cite a few. The *hazard rate* is defined as the inverse of the mean time between independent events.

Let $X_1, X_2, ...$ be independent and identically distributed random variables having an exponential distribution with probability density function given as

$$f_s(x) = \frac{1}{s}\exp\left(-\frac{x}{s}\right) \qquad x > 0,$$

where the scale parameter $s > 0$ is unknown. Let us consider the exponential hazard rate s^{-1}. Taking a random sample $X_1, X_2, ..., X_n$, let $\overline{X}_n$ denote the sample mean. Our goal is to estimate s^{-1} by its natural estimator $\overline{X}_n^{-1}$ and to find a confidence interval $I_n = \left[\overline{X}_n^{-1} - d, \overline{X}_n^{-1} + d\right]$ for s^{-1} of prescribed confidence coefficient $1-a$, and length $2d$, where $d > 0$ and $0 < a < 1$. It is known that no best fixed sample size procedure exists in constructing confidence intervals of prescribed length $2d$. Throughout, we shall assume that Z denotes a standard normal random variable and '$\xrightarrow{d}$', '$\xrightarrow{p}$' and '$\xrightarrow{a.s.}$' stands for convergence in distribution, convergence in probability and almost sure convergence, respectively.

The following result gives the asymptotic distribution of $\sqrt{n}(\overline{X}_n^{-1} - s^{-1})$.

Theorem 1.1. *Let $X_1, X_2, ..., X_n$ be independent and identically distributed random variables from an exponential distribution with scale parameter $s > 0$. Then*

$$\sqrt{n}(\overline{X}_n^{-1} - s^{-1}) \xrightarrow{d} N(0, s^{-2}) \quad \text{as} \quad n \to \infty.$$

Proof: Let $h(\overline{X}_n) = (\overline{X}_n)^{-1}$. By Taylor's Expansion of h around s,

$$\overline{X}_n^{-1} = h(\overline{X}_n) = h(s) + h'(s)(\overline{X}_n - s) + \frac{1}{2}h''(n)(\overline{X}_n - s)^2$$

$$= s^{-1} - s^{-2}(\overline{X}_n - s) + n^{-3}(\overline{X}_n - s)^2,$$

where n is a random variable lying between $\overline{X}_n$ and s.

Let $D_n = \sqrt{n}(\overline{X}_n - s)/s$ and $R_n = n^{-3}(\overline{X}_n - s)^2$. We then have

$$\sqrt{n}(\overline{X}_n^{-1} - s^{-1}) = -s^{-1}D_n + \sqrt{n}R_n . \qquad (1.1)$$

First, let us consider the second term $\sqrt{n}R_n$. By the Strong Law of Large Numbers, $\left|\overline{X}_n - s\right| \xrightarrow{a.s.} 0$ as $n \to \infty$. Thus,

$$\sqrt{n}R_n \xrightarrow{P} 0 \text{ as } n \to \infty .$$

By the Central Limit Theorem, $D_n \xrightarrow{d} Z$ as $n \to \infty$. Let $f(x) = -s^{-1}(x)$. Since f is continuous, it follows that

$$f(D_n) \xrightarrow{d} f(Z) \text{ as } n \to \infty ,$$

where $f(Z)$ is a normal random variable with mean 0 and variance s^{-2}. Together, we have

$$\sqrt{n}(\overline{X}_n^{-1} - s^{-1}) \xrightarrow{d} N(0, s^{-2}) \quad \text{as } n \to \infty . \qquad ¦$$

II. Fixed Width Confidence Interval

Let $h(x)$ be a positive, real-valued and three-times continuously differentiable function defined by $h(x) = x^{-1}$. We estimate $h(s) = s^{-1}$ by $\overline{X}_n^{-1}$ and construct a confidence interval $I_n = \left[\overline{X}_n^{-1} - d, \overline{X}_n^{-1} + d\right]$. To illustrate the construction, we choose a standard normal variate $a = a_a > 0$ such that $F(a) = 1 - a/2$, where F is the standard normal distribution function. Set

$$n^* = \frac{a^2}{d^2 s^2} . \qquad (2.1)$$

Now, the coverage probability for s^{-1} with I_n as the confidence interval is given by

$$P\left\{s^{-1} \in I_n\right\} = P\left\{\left|\overline{X}_n^{-1} - s^{-1}\right| \le d\right\}$$
$$= P\left\{\sqrt{n}s\left|\overline{X}_n^{-1} - s^{-1}\right| \le d\sqrt{n}s\right\}$$

Then for all $n \ge n^*$,

$$P\left\{s^{-1} \in I_n\right\} \ge P\left\{\sqrt{n}s\left|\overline{X}_n^{-1} - s^{-1}\right| \le a\right\} \approx 1 - a , \qquad (2.2)$$

for sufficiently large n^*. In the language of Chow and Robbins [2], we say that n^* is the *optimal sample size* which satisfies (2.2). Observe that as $d \to 0$, $n^* \to \infty$.

III. Main Results

In this section, we will propose a stopping rule used as an estimate of the optimum sample size used in constructing the confidence interval $I_n = \left[\overline{X}_n^{-1} - d, \overline{X}_n^{-1} + d\right]$ for the exponential hazard rate, s^{-1}. Recall from the previous section that n^* is the optimum sample size needed in estimating I_n with coverage probability of at least $1-a$. Now, since s is unknown, n^* is also unknown. Takada [4] pointed out that fixed sample size procedures are not available for scale families. To circumvent this problem we propose a sequential procedure in estimating I_n. We thus propose the following stopping rule:

$$N_d = \inf\left\{ n \geq m : n \geq \frac{a^2}{d^2 \overline{X}_n^2} \right\}, \qquad (3.1)$$

where $m \geq 2$ is the initial sample size.

Theorem 3.1. Let n^* and N_d be defined as in (2.1) and (3.1), respectively. Then the following statements hold:

(i) $P\{N_d < \infty\} = 1$ for each $d > 0$;

(ii) $N_d \xrightarrow{a.s.} +\infty$ as $d \to 0$; and

(iii) $\frac{N_d}{n^*} \xrightarrow{a.s.} 1$ as $d \to 0$.

Definition 3.2. A sequence $\langle Y_n \rangle$, $n \geq 1$, of random variables is said to be *uniformly continuous in probability* (u.c.i.p.) if and only if for every $e > 0$ there is a $d > 0$ for which

$$P\left\{ \max_{0 \leq k \leq nd} \left| Y_{n+k} - Y_n \right| \geq e \right\} < e, \quad \forall \ n \geq 1.$$

Definition 3.3. A sequence $\langle Y_n \rangle$, $n \geq 1$, of random variables is said to be *stochastically bounded* if and only if for every $e > 0$ there is a $C > 0$ for which

$$P\{|Y_n| > C\} < e\ , \quad \forall \quad n \geq 1.$$

The following theorem gives the asymptotic distribution of $\sqrt{N_d}\left(\overline{X}_{N_d}^{-1} - s^{-1}\right)$, which is needed in showing the asymptotic consistency of the proposed sequential procedure.

Theorem 3.4. Let N_d be defined as in (3.1). As $d \to 0$,

$$\sqrt{N_d}\left(\overline{X}_{N_d}^{-1} - s^{-1}\right) \xrightarrow{d} N\left(0, s^{-2}\right).$$

Proof: From (1.1), we have

$$\sqrt{n}(\overline{X}_n^{-1} - s^{-1}) = -s^{-1}D_n + \sqrt{n}R_n\ ,$$

where $D_n = \sqrt{n}\left(\overline{X}_n - s\right)/s$, f is a linear function $f(x) = -s^{-1}x$, $x \in (0, +\infty)$, and $R_n = n^{-3}\left(\overline{X}_n - s\right)^2$, n is a random variable lying between $\overline{X}_n$ and s .

By the Central Limit Theorem, $D_n \xrightarrow{d} Z$. From Example 1.8 of Woodroofe [6], D_n is u.c.i.p. and stochastically bounded. Since f is continuous, it follows from Lemma 1.4 of Woodroofe [6] that $f\left(D_n\right)$ is u.c.i.p.. Further, since $f(D_n) \xrightarrow{d} f(Z)$ as $n \to \infty$, Theorem 3.1 (iii) and Theorem 1.4 of Woodroofe [6] imply that

$$f\left(D_{N_d}\right) \xrightarrow{d} f(Z) \text{ as } d \to 0\ ,$$

where $f(Z)$ is normally distributed with mean 0 and variance s^{-2}. From Theorem 3.1 (ii), $N_d \to \infty$ as $d \to 0$. Thus, by the Strong Law of Large Numbers $\left|\overline{X}_{N_d} - s\right| \xrightarrow{a.s.} 0$. As a result, $\sqrt{N_d}R_{N_d} \xrightarrow{p} 0$ as $d \to 0$. Consequently,

$$\sqrt{N_d}\left(\overline{X}_{N_d}^{-1} - s^{-1}\right) \xrightarrow{d} f(Z) \text{ as } d \to 0.$$ ¦

The following theorem establishes the asymptotic consistency of the sequential procedure.

Theorem 3.5. (*Asymptotic Consistency*) Let N_d be defined as in (3.1). Then,

$$\lim_{d \to 0} P\left\{s^{-1} \in I_{N_d}\right\} = 1 - a \ .$$

Proof: Let $Z_d = \sqrt{N_d}\left(\overline{X}_{N_d}^{-1} - s^{-1}\right)/s$. By Theorem 3.1 (i) we are assured that N_d is finite; and write

$$P\left\{s^{-1} \in I_{N_d}\right\} = P\left\{\left|\overline{X}_{N_d}^{-1} - s^{-1}\right| \le d\right\}$$
$$= P\left\{\left|Z_d\right| \le d\sqrt{N_d}\, s\right\}.$$

To prove our assertion, we proceed in the following manner. First we show that

$$\liminf_{d \to 0} P\left\{s^{-1} \in I_{N_d}\right\} \ge 1 - a \ ,$$

and then

$$\limsup_{d \to 0} P\left\{s^{-1} \in I_{N_d}\right\} \le 1 - a \ .$$

From the definition of N_d in (3.1), we have $d\sqrt{N_d}\, s \ \ge a s \overline{X}_{N_d}^{-1}$ and

$$P\left\{s^{-1} \in I_{N_d}\right\} \ge P\left\{\left|Z_d\right| \le a s \overline{X}_{N_d}^{-1}\right\} \equiv A_d, \ \text{ say.}$$

Now, for any fixed $e \in (0,1)$,

$$A_d = P\left\{\left|Z_d\right| \le a s \overline{X}_{N_d}^{-1}, \left|s\overline{X}_{N_d}^{-1} - 1\right| \le e\right\} + P\left\{\left|Z_d\right| \le a s \overline{X}_{N_d}^{-1}, \left|s\overline{X}_{N_d}^{-1} - 1\right| > e\right\}$$
$$\le P\left\{\left|Z_d\right| \le a s \overline{X}_{N_d}^{-1}, \left|s\overline{X}_{N_d}^{-1} - 1\right| \le e\right\} + P\left\{\left|s\overline{X}_{N_d}^{-1} - 1\right| > e\right\}$$
$$\le P\left\{\left|Z_d\right| \le a(1+e)\right\} + P\left\{\left|s\overline{X}_{N_d}^{-1} - 1\right| > e\right\},$$

since $s\overline{X}_{N_d}^{-1} \le 1 + e$. By the Strong Law of Large Numbers, $\overline{X}_n \xrightarrow{a.s.} s$ as $d \to 0$. Hence, $s\overline{X}_{N_d}^{-1} \xrightarrow{p} 1$ as $d \to 0$ and

$$\lim_{e \to 0} P\left\{\left|s\overline{X}_{N_d}^{-1} - 1\right| > e\right\} = 0 \, .$$

Now, by Theorem 3.4, $Z_d \xrightarrow{d} Z$ as $d \to 0$. Thus, letting $d \to 0$ and taking $e \to 0$ we have $\limsup_{d \to 0} A_d \le 1 - a$. Similarly, we have

$$A_d = P\left\{|Z_d| \le a s \overline{X}_{N_d}^{-1}, \left|s\overline{X}_{N_d}^{-1} - 1\right| \le e\right\} - P\left\{\left|s\overline{X}_{N_d}^{-1} - 1\right| > e\right\}$$
$$+ P\left\{|Z_d| \le a s \overline{X}_{N_d}^{-1}, \left|s\overline{X}_{N_d}^{-1} - 1\right| > e\right\} + P\left\{\left|s\overline{X}_{N_d}^{-1} - 1\right| > e\right\}$$
$$\ge P\left\{|Z_d| \le a(1-e)\right\} - P\left\{\left|s\overline{X}_{N_d}^{-1} - 1\right| > e\right\},$$

since $1 - e \le s\overline{X}_{N_d}^{-1}$. Hence, taking $e \to 0$ we get $\liminf_{d\to 0} A_d \ge 1 - a$ and $\lim_{d\to 0} A_d = 1 - a$. As a result, $\liminf_{d\to 0} P\left\{s^{-1} \in I_{N_d}\right\} \ge 1 - a$.

Now, observe that

$$N_d - 1 < \frac{a^2}{d^2 \overline{X}_{N_d - 1}^2}$$
$$\frac{d\sqrt{N_d}\,s}{a} < \sqrt{\frac{s^2}{\overline{X}_{N_d-1}^2} + \frac{d^2 s^2}{a^2}} \equiv k_d, \text{ say,}$$

for all $N_d > m$. Now, for any $e \in (0,1)$,

$$P\left\{s^{-1} \in I_n\right\} = P\left\{|Z_d| \le d\sqrt{N}s,\ N > m\right\} + P\left\{|Z_d| \le d\sqrt{N}s,\ N = m\right\}$$
$$\le P\left\{|Z_d| \le a \cdot k_d,\ |k_d - 1| \le e\right\} + P\left\{|Z_d| \le a \cdot k_d,\ |k_d - 1| < e\right\} + P\{N_d = m\}$$
$$\le P\left\{|Z_d| \le a(1+e)\right\} + P\left\{|k_d - 1| < e\right\} + P\{N_d = m\},$$

since $k_d \le 1 + e$.

Observe that $P\{N_d = m\} = P\left\{m \ge \dfrac{a^2}{d^2 \overline{X}_m^2}\right\} \to P\{m \ge \infty\} = 0$, as $d \to 0$. Also, by the Strong Law of Large Numbers, $\overline{X}_{N_d} \xrightarrow{a.s.} s$ as $d \to 0$, so that $k_d \to 1$ as $d \to 0$. Finally, by Theorem 3.4, $Z_d \xrightarrow{d} Z$ as $d \to 0$. Thus, letting $d \to 0$ and taking $\epsilon \to 0$, it follows that

$$\limsup_{d\to 0} P\left\{s^{-1} \in I_{N_d}\right\} \le 1 - a.$$

Therefore, $\lim_{d\to 0} P\left\{s^{-1} \in I_{N_d}\right\} = 1 - a$. ¦

Let f be a real-valued function defined by $f(x) = x^2 / s^2$. Let $Z_n = nf(\overline{X}_n)$. From (2.1) and (3.1), we have $Z_n \ge n^*$. Hence, we can rewrite (3.1) as

$$N_d = \inf\left\{n \ge m : Z_n \ge n^*\right\}. \tag{3.2}$$

Now, by taking the Taylor's Expansion of $f(\overline{X}_n)$ around s , we obtain

$$f(\overline{X}_n) = f(s) + f'(s)(\overline{X}_n - s) + \frac{1}{2} f''(h)(\overline{X}_n - s)^2$$

$$= 1 + 2\left(\frac{\overline{X}_n - s}{s}\right) + \left(\frac{\overline{X}_n - s}{s}\right)^2,$$

where h is a random variable lying between $\overline{X}_n$ and s . Thus, Z_n can be written as

$$Z_n = n + 2S_n + x_n,$$

where

$$S_n = n\left(\frac{\overline{X}_n - s}{s}\right) \quad \text{and} \quad x_n = n\left(\frac{\overline{X}_n - s}{s}\right)^2 . \tag{3.3}$$

Lemma 3.6. Suppose S_n and $?_n$ are defined as in (3.3) above. Then

$$S_n^* = \frac{S_n}{\sqrt{n}} \xrightarrow{d} Z \quad \text{and} \quad x_n \xrightarrow{d} x,$$

both holding as $n \to \infty$, where x has a chi-square distribution with 1 degree of freedom.

The results used in the proof of the next theorem, which gives us the expected sample size resulting from the proposed sequential procedure, are taken from the results of Aras and Woodroofe [1]. We use the expected sample size given below as an estimate of the unknown optimum sample size n^*.

Theorem 3.7. Let N_d be defined as in (3.1), then

$$E(N_d) = n^* + r - 1 + o(1) \quad \text{as} \quad d \to 0,$$

where n^* is defined as in (2.1) and r satisfies $0 < ? < 5/2$.

Proof: Our aim is to check conditions (C1)-(C6) of Aras and Woodroofe [1]. To start with, let $Y_n = (X_n - s)/s$, $n \geq 1$. Note that we can write Z_n in (3.2-3.3) as $Z_n = n\overline{X}_n^2 / s^2 = ng(\overline{Y}_n)$, where $g(x) = (s + sx)^2 / s^2$. Clearly (C1) holds since $E(Y_1) = 0$, $E(Y_1^2) \leq 1 < \infty$, and $E(2Y_1^3) = 16 < \infty$. Since g is convex and $E\left[\{g(Y_1)\}^+\right] = E(X_1 / s)^6 = 720 < \infty$, (C2) and (C3) now follows from Proposition 5 of Aras and Woodroofe [1]. Finally, note that $g(0) = 1$, $g'(0) = 2$, and g is twice continuously differentiable from which (C4), (C5), and (C6) follows from Proposition 4 of Aras and Woodroofe [1]. Moreover, from Corollary 2.7 of Woodroofe [6],

$$r = \frac{5}{2} - \sum_{n=1}^{\infty} \frac{1}{n} E\left\{(n + 2S_n)^-\right\}$$

where $x^- = \max(-x, 0)$. Therefore, from Theorem 1 of Aras and Woodroofe [1], we have

$$E(N_d) = n^* + r - 1 + o(1) \quad \text{as} \quad d \to 0.$$

¦

IV. Simulation Results

In this section we consider simulation results of the proposed fully sequential procedure. Specifically, we will illustrate that the proposed sequential procedure possess the properties derived in Theorem 3.1, Theorem 3.5, and Theorem 3.7. As an example, consider the special case when the likelihood of hazard (failure) of a certain component remains constant with respect to time (for example, in some product like a brick or protected steel beam); the hazard rate is simply the inverse of the mean time between hazards. Assume that for this component, the mean time between hazards follows an exponential distribution with unknown scale parameter $s > 0$, then the hazard rate is given by s^{-1}. One aim is to estimate the hazard rate and thus find a confidence interval for s^{-1} with some pre-specified length $2d > 0$. We consider the stopping rule N_d in (3.1) in constructing sequential confidence intervals $I_{N_d} = \left[\overline{X}_{N_d}^{-1} - d, \overline{X}_{N_d}^{-1} + d\right]$ for s^{-1}. We shall give simulation results for the case when $s = 2$ and with level of confidence $a = 0.05$. The following results are based on 10,000 repetitions.

Table 4.1 Simulation Results Using N_d

n^*	20	100	200	500	1000
d	0.2191	0.0980	0.0693	0.0438	0.0310
$E(N_d)$	21.8898	101.3951	201.4857	501.2788	1011.03
$E(N_d / n^*)$	1.0945	1.0140	1.0074	1.0026	1.0011
$E(\overline{X}_{N_d})$	2.1421	2.0384	2.0186	2.0077	2.0039

$E\left(\overline{X}_{N_d}^{-1}\right)$	0.4875	0.4960	0.4981	0.4991	0.4995
$\left\|E\left(\overline{X}_{N_d}^{-1}\right) - s^{-1}\right\|$	0.0125	0.0040	0.0019	0.0009	0.0005
$P\left(s^{-1} \in I_{N_d}\right)$	0.9820	0.9476	0.9451	0.9470	0.9522

Table 4.1 seems to show that N_d, $\overline{X}_{N_d}$, and $\overline{X}_{N_d}^{-1}$ are unbiased estimators of n^*, s, and s^{-1}, respectively. Notice also that N_d / n^* converges to 1 which confirms the asymptotic efficiency of the sequential procedure. Finally, note that the coverage probability $P\left(s^{-1} \in I_{N_d}\right)$ converges to $1 - a = 0.95$, showing that the sequential confidence intervals are asymptotically consistent. Together, these properties suffice that of a fully sequential procedure.

References

[1] Aras, G. and M. Woodroofe. *Asymptotic expansions for the moments of a randomly stopped average*, Annals of Statistics (1993) 21:503-519.

[2] Chow, Y.S. and H. Robbins. *On the asymptotic theory of fixed width confidence intervals for the mean*, Annals of Mathematical Statistics, 36:457-462, 1965.

[3] Lim, D.L., E. Isogai, and C. Uno. *Two-sample fixed width confidence intervals for a function of exponential scale parameters*. Far East Journal of Statistics, 14:215-227, 2004.

[4] Takada, Y. *Non-existence of fixed sample size procedures for scale families.* Sequential Analysis, 5: 99-100, 1986.

[5] Uno, C., E. Isogai, and D.L. Lim. *Sequential point estimation of a function of the exponential scale parameter*, Austrian Journal of Statistics 33, 2004.

[6] Woodroofe, M. *Nonlinear renewal theory in sequential analysis*, CBMS Monograph No. 39, Society for Industrial and Applied Mathematics, 1982.