PROBABILITY
AND
MATHEMATICAL STATISTICS
Vol. 45, Fasc. 1 (2025), pp. 53–69
Published online 30.6.2025
doi:10.37190/0208-4147.00217

UNBIASED ESTIMATION OF EXPONENTIAL RELIABILITY FOR LEFT CENSORED DATA

BY

PIOTR BOLESŁAW NOWAK (WROCŁAW)

Abstract. The paper applies the Rao–Blackwell technique to construct an unbiased estimator of the reliability function in the exponential distribution under left censored data. The resulting estimator is a function of a sufficient statistic. However, we cannot conclude that it is a minimal variance estimator because the sufficient statistic is not complete. We also give an asymptotic confidence interval for reliability based on the estimator obtained. The performance of this estimator is compared to the maximum likelihood estimator via Monte Carlo simulation.

2020 Mathematics Subject Classification: Primary 62N01; Secondary 62F12.

Key words and phrases: exponential distribution, reliability function, unbiased estimation.

1. INTRODUCTION AND PRELIMINARIES

The paper deals with unbiased estimation of the reliability function $R(t)$ for left censored data from the exponential distribution with density function $f(x) = (1/\theta)e^{-x/\theta} \cdot \mathbf{1}_{(0,\infty)}(x)$ and cumulative distribution function $F(x) = (1 - e^{-x/\theta}) \cdot \mathbf{1}_{(0,\infty)}(x)$, where $\theta > 0$ and $\mathbf{1}_A$ denotes the indicator function of the set A. The reliability function, also known as survival function, has the form $R(t) = 1 - F(t) = e^{-t/\theta}$, $t > 0$. It is the most important characteristic used in life data analysis and reliability theory.

Let $(X_1, \ldots, X_n)$ be a sample from the exponential distribution. In many practical situations, for example in reliability engineering, medicine, sociology or economy, data are subject to left censoring, meaning that observations below a certain threshold are not recorded. Then we deal with left censored data, i.e. with random variables $\max(X_i, T)$ and $\delta_i = 1\{X_i > T\}$, $i = 1, \ldots, n$, where $T > 0$. For this case we need to have an unbiased estimator of reliability from the exponential distribution.

Now we provide examples regarding left censoring. As a first example, consider the data on Ammonia Nitrogen Concentrations in the Skagit River (available

in the dataset Skagit.NH3_N.df from the EnvStats package in R). A measuring device records pollutant concentrations only if they exceed a certain detection limit, determined by the instrument's sensitivity. If the actual concentration is below this threshold (here less than 0.01 mg/l), its exact value remains unknown. The reliability function represents the probability that the true pollutant concentration exceeds a given threshold. Another example comes from epidemiology, specifically in the study of incubation periods. When analyzing the incubation period for immunodeficiency syndrome (AIDS), researchers begin observing patients after a predefined post-dialysis period; see e.g. Lui et al. [9]. If a patient already exhibits symptoms at the start of the study, the exact incubation time remains unknown, and we only know it was shorter than the observation threshold. Here, the reliability function describes the probability that disease symptoms have not yet appeared by a given time. Another example arises in reliability theory. If a device fails almost instantaneously (e.g., within one second of operation), the failure may be undetected, because the recording system does not capture such short failure times. In this case, the reliability function represents the probability that the device remains operational beyond a given time. We discuss this case in Example 2 in Section 8.

For other applications of left censoring see e.g. Balakrishnan and Varadan [1], Klein and Moeschberger [5], and Mitra and Kundu [10]. The problem of unbiased estimation of exponential reliability was studied by many authors, but they mainly focused on the following cases: complete samples (Pugh [12]), right censored samples until a fixed time T (Bartoszewicz [2]) or right censored samples until the time of the rth failure (Basu [3], Laurent [6], Tate [14]).

In this paper we examine the problem of unbiased estimation of the reliability function in the exponential model under left censored data until a fixed time T. Let $D(t)$ denote the number of failures after time t, that is, $D(t) = \sum_{i=1}^{n} 1\{X_i > t\}$. This estimator applies to the case of complete data and to left censored data when $t > T$. Because $D(t)$ is a random variable with binomial distribution $b(n, e^{-t/\theta})$, the simplest unbiased estimator for reliability is $\hat{R}_0(t) = D(t)/n$. Its variance is $\mathrm{Var}[\hat{R}_0(t)] = e^{-t/\theta}(1 - e^{-t/\theta})/n$ for all $t > 0$. We give an unbiased estimator of $R(t)$, $t > T$, for left censored data, which has variance uniformly smaller than the variance of $\hat{R}_0(t)$.

2. THE SUFFICIENT STATISTIC AND ITS DISTRIBUTION

In this section we find the sufficient statistic for the parameter θ and we derive its distribution. This, by means of the Rao–Blackwell technique, allows us to construct a new reliability estimator based on the sufficient statistic, which is unbiased with variance uniformly smaller than the variance of the estimator $\hat{R}_0(t)$, $t > T$.

Let n identical items having exponentially distributed life time be placed simultaneously on life test according to the left censoring scheme. Let d denote the observed value of the random variable $D(T)$ and let $X_{(n-D(T)+1)} \leqslant \cdots \leqslant X_{(n)}$

be the moments of failures after time T. The joint probability density function of the vector $(X_{(n-D(T)+1)}, \dots, X_{(n)}, D(T))$ is

$$p(x_{(n-d+1)}, \dots, x_{(n)}, d; \theta) = \begin{cases} F(T)^n & \text{for } d = 0, \\ \dfrac{n!}{(n-d)!} F(T)^{n-d} \prod_{i=1}^{d} f(x_{(n-d+i)}) & \\ \quad \text{for } d = 1, \dots, n,\ x_{(n-d+1)} \leqslant \dots \leqslant x_{(n)}. & \end{cases}$$

Hence, for $d = 1, \dots, n$ and $x_{(n-d+1)} \leqslant \dots \leqslant x_{(n)}$ the log likelihood function denoted by $\log L(x_{(n-d+1)}, \dots, x_{(n)}, d; \theta)$ (or simply $\log L(\theta)$) is

$$\log L(\theta) = \frac{n!}{(n-d)!} \sum_{i=1}^{d} (-\log \theta - x_{(n-d+i)}/\theta) + (n-d) \log(1 - e^{-T/\theta}).$$

The factorization theorem implies that the sufficient statistic for the parameter θ is the vector $(D(T), S(T))$, where $S(T) = \sum_{i=1}^{D(T)} X_{(n-D(T)+i)}$, and $D(T)$ is the binomial random variable $b(n, e^{-T/\theta})$. The set of values of this sufficient statistic consists of the point $(0, 0)$ and n segments $\{(d, s) : s > dT,\ d = 1, \dots, n\}$. We have $P((D(T), S(T)) = (0, 0)) = (1 - e^{-T/\theta})^n$.

Now assume that $d \geqslant 1$ and we find the conditional distribution of $S(T)$ given $D(T) = d$. Let $Y_1, \dots, Y_d$ be independent random variables with exponential distribution truncated at T with density $f_Y(y) = (1/\theta) e^{-(y-T)/\theta} \cdot \mathbf{1}_{(T,\infty)}(y)$. Observe that the conditional distribution of $S(T)$ given $D(T) = d$ is the same as the unconditional distribution of the sum $S = Y_1 + \dots + Y_d$. Then the joint density of the vector $(Y_1, \dots, Y_d)$ is

$$p_{\theta d}(y_1, \dots, y_d) = \begin{cases} \theta^{-d} e^{-(y_1 + \dots + y_d - dT)/\theta} & \text{if } y_i > T \text{ for } i = 1, \dots, d, \\ 0 & \text{otherwise.} \end{cases} \tag{2.1}$$

For $d = 1$ we have $S = Y_1$ and from (2.1) we get $p_{\theta d}(s) = \theta^{-1} e^{-(s-T)/\theta}$ for $s > T$. For $d \geqslant 2$ let us consider the transformation $(Y_1, \dots, Y_d) \mapsto (Y_1, \dots, Y_{d-1}, S)$. Its Jacobian is equal to 1 and therefore, according to (2.1), the density of the vector $(Y_1, \dots, Y_{d-1}, S)$ is

$$q_{\theta d}(y_1, \dots, y_{d-1}, s) = \begin{cases} \theta^{-d} e^{-(s-dT)/\theta} & \text{if } y_i > T \text{ for } i = 1, \dots, d-1 \\ & \text{and } y_1 + \dots + y_{d-1} < s - T, \\ 0 & \text{otherwise.} \end{cases}$$

It follows that the density of the statistic S for $d \geqslant 2$ is

$$p_{\theta d}(s) = \int q_{\theta d}(y_1, \dots, y_{d-1}, s)\, dy_1 \cdots dy_{d-1} = V_d(s, T) \theta^{-d} e^{-(s-dT)/\theta},$$

where $V_d(s, T)$ is the volume of the region in $\mathbf{R}^{d-1}$ determined by the inequalities $y_i > T$ for $i = 1, \dots, d-1$ and $y_1 + \dots + y_{d-1} < s - T$. Now we prove the following lemma.

LEMMA 2.1. *The volume of the region in* $\mathbf{R}^{d-1}$*, where* $d \geqslant 2$*, determined by the inequalities* $y_i > T$ *for* $i = 1, \ldots, d-1$ *and* $y_1 + \cdots + y_{d-1} < s - T$ *equals*

$$V_d(s,T) = \frac{(s-dT)_+^{d-1}}{(d-1)!},$$

where $a_+^m = [\max(0,a)]^m$ *for* $m > 0$.

Proof. Because $\int_0^\infty p_{\theta d}(s)\,ds = 1$, we find that $\int_0^\infty V_d(s,T)e^{-s/\theta}\,ds = \theta^d e^{-dT/\theta}$. Observe also that the left side of the last equation is the Laplace transform of the function $V_d(s,T)$. Moreover, it is easy verify that $\int_0^\infty (s-dT)_+^{d-1} e^{-s/\theta} ds = (d-1)!\theta^d e^{-dT/\theta}$. Therefore, from the uniqueness of the Laplace transform we get $V_d(s,T) = (s-dT)_+^{d-1}/(d-1)!$. ∎

As an immediate consequence of Lemma 2.1 we get the following lemma.

LEMMA 2.2. *If* $Y_1, \ldots, Y_d$ *are independent, identically distributed random variables with density*

$$f_Y(y) = (1/\theta)e^{-(y-T)/\theta} \cdot \mathbf{1}_{(T,\infty)}(y),$$

then the random variable $S = Y_1 + \cdots + Y_d$ *has density*

$$p_{\theta d}(s) = \frac{\theta^{-d} e^{-(s-dT)/\theta}}{(d-1)!}(s-dT)_+^{d-1}. \tag{2.2}$$

In a similar manner we can prove the following lemma concerning the distribution of the sum of exponential random variables truncated on the set (a,b), i.e. truncated on the left and on the right.

LEMMA 2.3. *If* $Y_1, \ldots, Y_d$ *are independent, identically distributed random variables with density*

$$f_{(a,b)}(y) = \frac{e^{-y/\theta}}{\theta(e^{-a/\theta} - e^{-b/\theta})} \cdot \mathbf{1}_{(a,b)}(y),$$

then the density of $S = Y_1 + \cdots + Y_d$ *is the* d*-fold convolution of* $f_{(a,b)}$*, i.e.*

$$f_{(a,b)}^{d*}(s) = \frac{\theta^{-d} e^{-s/\theta}}{(d-1)!(e^{-a/\theta} - e^{-b/\theta})^d} \times \sum_{j=0}^{d} (-1)^j \binom{d}{j} [s - (d-j)a - jb]_+^{d-1}, \quad s > 0. \tag{2.3}$$

Proof. The assertion holds for $d = 1$. Let $d \geqslant 2$. Then the density of the vector $(Y_1, \ldots, Y_{d-1}, S)$ is

$$q_{\theta d}(y_1, \ldots, y_{d-1}, s) = \begin{cases} \dfrac{\theta^{-d} e^{-s/\theta}}{(e^{-a/\theta} - e^{-b/\theta})^d} & \text{if } a < y_i < b \text{ for } i = 1, \ldots, d-1, \\ & s - b < y_1 + \cdots + y_{d-1} < s - a, \\ 0 & \text{otherwise.} \end{cases}$$

Hence

$$f^{d*}_{(a,b)}(s) = \int q_{\theta d}(y_1, \ldots, y_{d-1}, s)\, dy_1 \cdots dy_{d-1} = V_d(s, a, b) \frac{\theta^{-d} e^{-s/\theta}}{(e^{-a/\theta} - e^{-b/\theta})^d},$$

where $V_d(s, a, b)$ is the volume of the region in $\mathbf{R}^{d-1}$ determined by the inequalities $a < y_i < b$ for $i = 1, \ldots, d-1$ and $s - b < y_1 + \cdots + y_{d-1} < s - a$. Because $\int_0^\infty f^{d*}_{(a,b)}(s)\, ds = 1$ we have

$$\int_0^\infty V_d(s, a, b) e^{-s/\theta}\, ds = \theta^d (e^{-a/\theta} - e^{-b/\theta})^d.$$

The left side of this equation is the Laplace transform of the function $V_d(s, a, b)$. One can easily check that

$$V_d(s, a, b) = \frac{1}{(d-1)!} \sum_{j=0}^{d} (-1)^j \binom{d}{j} [s - (d-j)a - jb]_+^{d-1}$$

satisfies the above equation. Therefore by the uniqueness of the Laplace transform the proof is complete. ∎

Note that the density of the convolution $f^{d*}_{(0,b)}$ has been derived by Hoem [4]. Letting $b \to \infty$ in (2.3) and assuming that $a = T$, we also get the density $p_{\theta d}$ in (2.2).

3. MAIN RESULT

In order to construct a new unbiased estimator of the reliability we apply the Rao–Blackwell theorem (see Voinov and Nikulin [15]).

Using this theorem we define an estimator $\hat{R}(t)$ as the conditional expectation $\hat{R}_0(t)$ given $(D(T), S(T))$, i.e. $\hat{R}(t) = E[\hat{R}_0(t) \mid D(T), S(T)]$ for $t > T$. It is an unbiased estimator of the exponential reliability with variance smaller than that of the estimator $\hat{R}_0(t)$.

Unfortunately, except for the trivial case $n = 1$, this new estimator is not a minimum variance unbiased estimator because the sufficient statistic is not complete (as proved below).

THEOREM 3.1. *For left censored data, the function*

$$\hat{R}(t) = \begin{cases} 0 & \text{for } d = 0, \\ \dfrac{d}{n}\left[1 - \dfrac{t-T}{s-dT}\right]_+^{d-1} & \text{for } d > 0, \end{cases} \tag{3.1}$$

is an unbiased estimator of the exponential reliability $R(t)$, $t > T$, where $a_+^m = [\max(0,a)]^m$ for $m > 0$ and

$$a_+^0 = \begin{cases} 0 & \text{for } a \leqslant 0, \\ 1 & \text{for } a > 0. \end{cases}$$

Notice that for $T = 0$ we have $d = n$, which implies that $\hat{R}(t) = \left[1 - \frac{t}{s}\right]_+^{n-1}$ for $t > 0$. In this case S is a complete statistic for θ. Therefore, by the Lehmann–Scheffé theorem [7, 8] $\hat{R}(t)$ is a minimum variance estimator in this case. This is a result of Pugh [12], who derived a minimum variance unbiased estimator of exponential reliability for a complete sample.

In order to prove Theorem 3.1 we first prove the following lemma.

LEMMA 3.1. *The conditional density of $S(T)$ given $D(T) = d$ and $D(t) = k$ for $t > T$ is*

$$h(s \mid d,k) = \frac{\theta^{-d} e^{-(s-kt)/\theta}}{(d-1)!(e^{-T/\theta} - e^{-t/\theta})^{d-k}} \times \sum_{j=0}^{d-k} (-1)^j \binom{d-k}{j} [s - (d-k-j)T - (k+j)t]_+^{d-1}.$$

Proof. The random variable $S(T)$ given $D(T) = d$ and $D(t) = k$ equals the sum of the lifetimes of those $d-k$ elements which failed in the interval (T,t) and the lifetimes of the k elements which failed in the interval (t,∞). Hence, the conditional density of the distribution of $S(T)$ given $D(T) = d$ and $D(t) = k$ equals

$$h(s \mid d,k) = \int_{-\infty}^{\infty} f_{(T,t)}^{(d-k)*}(s-v) f_{(t,\infty)}^{k*}(v)\, dv,$$

where the functions $f_{(T,t)}^{(d-k)*}$ and $f_{(t,\infty)}^{k*}$ are defined in (2.3) and (2.2), respectively. Therefore

$$h(s \mid d,k) = \frac{\theta^{-d} e^{-(s-kt)/\theta}}{(k-1)!(d-k-1)!(e^{-T/\theta} - e^{-t/\theta})^{d-k}} \sum_{j=0}^{d-k} (-1)^j \binom{d-k}{j} \cdot I,$$

where $I = \int_{-\infty}^{\infty} [v-kt]_+^{k-1} [s - v - (d-j-k)T - jt]_+^{d-k-1}\, dv$.

Easy algebra implies that

$$I = c_+^{d-1} B(d-k,k) = c_+^{d-1}(d-k-1)!(k-1)!/(d-1)!,$$

where B is the beta function and $c_+ = [s-(d-j-k)T-(j+k)t]_+$. ∎

Proof of Theorem 3.1. In order to derive an unbiased estimator of the reliability $R(t)$ for $t > T$ we need to calculate $E[D(t)/n \mid D(T) = d,\, S(T) = s]$, which equals

$$\begin{aligned}\hat{R}(t) &= \frac{1}{n}\sum_{k=0}^{d} kP[D(t)=k \mid D(T)=d,\, S(T)=s] \\ &= \frac{1}{n}\sum_{k=0}^{d} k\frac{h(s \mid d,k)P[D(T)=d,\, D(t)=k]}{p_{\theta d}(s)P[D(T)=d]},\end{aligned}$$

where $p_{\theta d}(s)$ denotes the density of the random variable $S = Y_1 + \cdots + Y_d$ given in (2.2).

Since $P[D(T)=0,\, D(t)=k] = 0$ for $k \geqslant 1$, we have $\hat{R}(t) = 0$ for $d = 0$. Otherwise, by Lemma 3.1 and the following facts:

(a) $P[D(T)=d] = \binom{n}{d} e^{-dT/\theta}(1-e^{-T/\theta})^{n-d}$, $0 \leqslant d \leqslant n$;

(b) $P[D(T)=d, D(t)=k]$

$$= \frac{n!}{k!(d-k)!(n-d)!} e^{-kt/\theta}(e^{-T/\theta}-e^{-t/\theta})^{d-k}(1-e^{-T/\theta})^{n-d}, 0 \leqslant k \leqslant d \leqslant n,$$

the function

$$\hat{R}(t) = \frac{d}{n}\sum_{k=1}^{d}\binom{d-1}{k-1}\sum_{j=0}^{d-k}(-1)^j\binom{d-k}{j}Q(j+k)$$

is an unbiased estimator of reliability, where

$$Q(j+k) = \left[\frac{s-(d-j-k)T-(j+k)t}{s-dT}\right]_+^{d-1}.$$

This can be written in the form

$$\begin{aligned}\hat{R}(t) &= \frac{d}{n}\sum_{m=1}^{d}\sum_{r=1}^{m}\binom{d-1}{m-1}\binom{m-1}{r-1}(-1)^{m-r}Q(m) \\ &= \frac{d}{n}\sum_{m=1}^{d}\binom{d-1}{m-1}Q(m)(1-1)^{m-1} = \frac{d}{n}Q(1) = \frac{d}{n}\left[1-\frac{t-T}{s-dT}\right]_+^{d-1},\end{aligned}$$

which completes the proof. ∎

One can show in a similar manner that for $t < T$ we have

$$E\left[\frac{D(t)}{n} \;\middle|\; D(T) = d,\ S(T) = s\right] = 1 - \left(1 - \frac{d}{n}\right)\frac{1 - e^{-t/\theta}}{1 - e^{-T/\theta}}.$$

Under left censoring, $D(t)/n$ for $t < T$ is not a function of the observations $(\max(X_i, T), \delta_i)$, $i = 1, \ldots, n$, and cannot be an estimator. Nevertheless, we can provide an asymptotically unbiased estimator of the reliability function for $t > 0$ which is based on the sufficient statistic $(D(T), S(T))$.

LEMMA 3.2. *For left censored data the function*

$$\tilde{R}(t) = \frac{d}{d+\varepsilon}\left[1 - \frac{t}{s - dT}\right]_+^{d-1}, \quad \varepsilon > 0, \tag{3.2}$$

is an asymptotically unbiased estimator of the reliability function $R(t)$, $t > 0$, i.e. it tends to $e^{-t/\theta}$ as $n \to \infty$.

Proof. We start from $\tilde{R}_0(t) = D(T+t)/(D(T)+\varepsilon)$, $\varepsilon > 0$, which is an asymptotically unbiased estimator of $R(t)$. This follows from

$$\begin{aligned} E[\tilde{R}_0(t)] &= E\{E[\tilde{R}_0(t) \mid D(T)]\} \\ &= E\left\{\frac{1}{D(T)+\varepsilon}E[D(T+t) \mid D(T)]\right\} = e^{-t/\theta}E\left[\frac{D(T)}{D(T)+\varepsilon}\right], \end{aligned}$$

because $[D(T+t) \mid D(T) = d]$ has the binomial distribution $b(d, e^{-t/\theta})$. Hence, as $n \to \infty$, $E[\tilde{R}_0(t)]$ tends to $e^{-t/\theta}$ with probability 1. The estimator $\tilde{R}_0(t)$ is biased but by the law of total variance the estimator $\tilde{R}(t)$ defined as $E[\tilde{R}_0(t) \mid D(T) = d,\ S(T) = s]$ has smaller variance than $\tilde{R}_0(t)$. In fact,

$$\begin{aligned} \mathrm{Var}[\tilde{R}_0(t)] &= \mathrm{Var}[E(\tilde{R}_0(t) \mid D, S)] + E[\mathrm{Var}(\tilde{R}_0(t) \mid D, S)] \\ &\geqslant \mathrm{Var}[E(\tilde{R}_0(t) \mid D, S)]. \end{aligned}$$

Finally, by Theorem 3.1 we get

$$\begin{aligned} \tilde{R}(t) &= E[\tilde{R}_0(t) \mid D(T) = d,\ S(T) = s] = \frac{n\hat{R}(T+t)}{d+\varepsilon} \\ &= \frac{d}{d+\varepsilon}\left[1 - \frac{t}{s - dT}\right]_+^{d-1}, \end{aligned}$$

which ends the proof. ∎

4. INCOMPLETENESS OF $(D(T), S(T))$

In this section we prove that the statistic $(D(T), S(T))$ is complete only for the trivial case $n = 1$. In order to prove that $(D(T), S(T))$ is not complete for $n \geqslant 2$ we use a probabilistic argument. This idea comes from Sarkadi [13].

THEOREM 4.1. *The statistic $(D(T), S(T))$ is not complete except for the trivial case $n = 1$.*

Proof. First, we show that for $n = 1$ and for any function $\varphi(d, s)$ the condition $E[\varphi(D, S)] = 0$ for every $\theta > 0$ implies $\varphi(d, s) = 0$. For this case

$$E[\varphi(D, S)] = (1 - e^{-T/\theta})\varphi(0, 0) + \int_T^\infty \varphi(1, s)(1/\theta)e^{-T/\theta}\, ds$$

and $E[\varphi(D, S)] = 0$ if and only if

$$\int_0^\infty \tilde{\varphi}(1, s)e^{-s/\theta} ds = -\varphi(0, 0)\theta(1 - e^{-T/\theta}), \tag{4.1}$$

where

$$\tilde{\varphi}(1, s) = \begin{cases} \varphi(1, s) & \text{for } s \geqslant T, \\ 0 & \text{for } s \in (0, T). \end{cases}$$

The left side of (4.1) is the Laplace transform of $\tilde{\varphi}(1, s)$. It follows from the uniqueness of the Laplace transform that $\tilde{\varphi}(1, s) = -\varphi(0, 0)$ for $s \in (0, T)$ and $\tilde{\varphi}(1, s) = 0$ for $s \geqslant T$. This yields $\varphi(1, s) = 0$ for $s \geqslant T$, and also $\varphi(0, 0) = 0$, which ends the proof of completeness for $n = 1$.

Now, let $n \geqslant 2$ and consider the following statistic (which does not depend on θ):

$$\varphi(D, S) = P(X_1 > T,\, X_2 > T \mid D, S) - P(X_1 > 2T \mid D, S).$$

Then

$$E[\varphi(D, S)] = P(X_1 > T,\, X_2 > T) - P(X_1 > 2T) = (e^{-T/\theta})^2 - e^{-2T/\theta} = 0.$$

Let us consider the event $A = \{D = n,\, S < (n + 1)T\}$. If A occurs, then $T < X_i < 2T$ for $i = 1, \ldots, n$, and hence $\varphi(D, S) = 1 - 0 = 1$. Now we show that $P(A) > 0$, which will complete the proof. Indeed, if $T < X_i < (n + 1)T/n$ for $i = 1, \ldots, n$ then $S < (n + 1)T$. Hence

$$P(A) > P(T < X_i < (n + 1)T/n,\, i = 1, \ldots, n) > 0. \quad \blacksquare$$

COROLLARY 4.1. *For $n = 1$ the unbiased estimator $\hat{R}(t) = 1\{X_1 > t\}$, $t > T$, of reliability is a minimum variance unbiased estimator, because it is a function of a complete and sufficient statistic. When $n \geqslant 2$ we cannot conclude that $\hat{R}$ has minimum variance, because the sufficient statistic $(D(T), S(T))$ is not complete.*

5. THE ASYMPTOTIC DISTRIBUTION OF $\hat{R}$

In order to derive the distribution of $\hat{R}(t)$ when $t > T$ and $n \to \infty$, we first provide an approximation of this estimator.

Let $\bar{D} = (1/n)\sum_{i=1}^{n} \delta_i$ and $\bar{S} = (1/n)\sum_{i=1}^{n} S_i$, where $S_i = 1\{X_i > T\}X_i$ for $i = 1, \ldots, n$. Because $S(T)/n = \bar{S}$, by the Strong Law of Large Numbers we know that $S(T)/D(T) = \bar{S} \cdot (D(T)/n)^{-1}$ tends to some constant with probability 1. Therefore, for large n we may approximate our estimator as follows:

$$\hat{R}(t) = \frac{d}{n}\left[1 - \frac{(t-T)/d}{s/d - T}\right]_{+}^{(d-1)} \simeq \frac{d}{n}\exp\left[-\frac{t-T}{s/d - T}\right]. \tag{5.1}$$

By the multivariate central limit theorem,

$$\sqrt{n}[(\bar{D}, \bar{S}) - E(\bar{D}, \bar{S})] \to_d N((0,0), \Sigma),$$

where the elements σ_{ij} of Σ are

$$\sigma_{11} = \mathrm{Var}(\delta_i) = e^{-T/\theta}(1 - e^{-T/\theta}),$$
$$\sigma_{12} = \mathrm{Cov}(\delta_i, S_i) = (\theta + T)(e^{-T/\theta} - e^{-2T/\theta}),$$
$$\sigma_{22} = \mathrm{Var}(S_i) = (T^2 + 2\theta T + 2\theta^2)e^{-T/\theta} - (\theta + T)^2 e^{-2T/\theta}.$$

Let $\hat{R}_a(t)$ be the approximate estimator of reliability for $t > T$ given in (5.1). Define

$$g(x, y) = x \exp\left[-\frac{t-T}{y/x - T}\right], \quad x, y > 0.$$

By the delta method,

$$\sqrt{n}[g(\bar{D}, \bar{S}) - g(\beta)] \to_d N(0, \nabla g(\beta)\Sigma\nabla g(\beta)^T),$$

where $\beta = (E(\bar{D}), E(\bar{S}))$ and $\nabla g(\beta) = [(\partial/\partial x)g(x,y), (\partial/\partial y)g(x,y)]|_{(x,y)=\beta}$.

One can check that $g(E(\bar{D}), E(\bar{S})) = e^{-t/\theta}$. Let $\mathrm{Var}_a(\theta)$ be the asymptotic variance of the statistic $g(\bar{D}, \bar{S})$. After direct calculations and replacing the parameter θ by $-t/\log R$ in the function Var_a, we get the approximate $100(1-\alpha)\%$ confidence interval for the reliability $R(t)$,

$$\hat{R}_a(t) \pm z_{\alpha/2}\sqrt{\frac{\mathrm{Var}_a(\hat{R}_a(t))}{n}}, \tag{5.2}$$

where $\mathrm{Var}_a(R) = R^{2-T/t}[(1 - R^{T/t}) + (1 - T/t)^2 \log^2 R]$ and $z_{\alpha/2}$ denotes the critical value of the standard normal distribution at $\alpha/2$.

Note that for $t \to T$ the confidence interval (5.2) reduces to the Wald confidence interval for proportion of the form

$$d/n \pm z_{\alpha/2}\sqrt{\frac{(d/n)(1 - d/n)}{n}}.$$

Using the same method, we derive the asymptotic properties of the modified estimator of reliability defined in (3.2). When n is large, so is d, and the estimator $\tilde{R}(t)$ has the following approximation:

$$\tilde{R}(t) = \frac{d}{d+\varepsilon}\left[1 - \frac{t}{s - dT}\right]_+^{d-1} \simeq \exp\left[-\frac{t}{s/d - T}\right]. \tag{5.3}$$

Denote the approximation of $\tilde{R}(t)$ given in (5.3) by $\tilde{R}_b(t)$. Applying the delta method and after direct computation we get the approximate $100(1-\alpha)\%$ confidence interval for the reliability $\hat{R}(t)$ which is based on the estimator $\tilde{R}_b(t)$ for $t > 0$:

$$\tilde{R}_b(t) \pm z_{\alpha/2}\sqrt{\frac{\mathrm{Var}_b(\tilde{R}_b(t))}{n}}, \tag{5.4}$$

where $\mathrm{Var}_b(R) = R^{2-T/t}\log^2 R$.

6. THE MAXIMUM LIKELIHOOD ESTIMATOR

Now we derive the maximum likelihood estimator (MLE) for reliability. This problem was considered by Pugh [12] in the case of a complete sample. Pugh showed that the MLE for reliability is biased. Based on the considerations in Section 2, the log-likelihood function is

$$\log L(\theta) \propto \sum_{i=1}^{d}(-\log\theta - x_{(n-d+i)}/\theta) + (n-d)\log(1 - e^{-T/\theta}).$$

The normal equation $\partial \log L(\theta)/\partial\theta = 0$ for deriving the MLE of θ (say $\hat{\theta}_{ml}$) is equivalent to $h(\theta) = \theta$, where

$$h(\theta) = \frac{1}{n}\left[\sum_{i=1}^{d} x_{(n-d+i)} + (n-d)\left(\theta + \frac{T}{1 - e^{T/\theta}}\right)\right].$$

Observe that if $d = 0$ then the MLE of θ does not exist. If $d = n$, then $\hat{\theta}_{ml}$ is the arithmetic mean of the sample. Otherwise, it can be easily proved that h is a mapping from $(0,\infty)$ to itself and $\sup_{\theta\in(0,\infty)} h'(\theta) = 1 - d/n$, which means that h is a contraction. Therefore h has a unique fixed point which can be obtained as a limit through an iterative procedure from any initial positive value θ_0, i.e. $\hat{\theta}_{ml} = \lim_{i\to\infty}\theta_i$, where $\theta_{i+1} = h(\theta_i)$, $i \geqslant 0$. The initial value of θ_0 can be taken as $\theta_0 = -T/\log(d/n)$. Hence, by the invariant principle the MLE of reliability is $\hat{R}_{ml}(t) = e^{-t/\hat{\theta}_{ml}}$. Let $V_{ml}(\theta)$ denote the asymptotic variance of $\hat{R}_{ml}(t)$. Then

$$V_{ml}(\theta) = \left[\frac{\partial}{\partial\theta}e^{-t/\theta}\right]^2\left\{-E\left[\frac{\partial^2}{\partial\theta^2}\log L(\theta)\right]\right\}^{-1}.$$

One can show that the approximate $100(1-\alpha)\%$ confidence interval for reliability which is based on $\hat{R}_{ml}(t)$ has the form

$$\hat{R}_{ml}(t) \pm z_{\alpha/2}\sqrt{\mathrm{Var}_{ml}(\hat{\theta}_{ml})}, \tag{6.1}$$

where $\mathrm{Var}_{ml}(\theta) = \dfrac{t^2 e^{-2t/\theta}(e^{T/\theta}-1)}{n[T^2+\theta^2(1-e^{-T/\theta})]}$.

7. ANOTHER UNBIASED ESTIMATOR OF RELIABILITY BASED ON $(D(T), S(T))$

In this section we derive another unbiased estimator of $R(t)$, which is also a function of the sufficient statistic $(D(T), S(T))$.

Assume that $t > nT$ and consider the estimator

$$\delta(t) = \begin{cases} 1 & \text{if } \min(X_1,\dots,X_n) > t/n, \\ 0 & \text{otherwise.} \end{cases}$$

The estimator δ is unbiased, because

$$\begin{aligned} E[\delta(t)] &= P[\min(X_1,\dots,X_n) > t/n] \\ &= \prod_{i=1}^{n} P[X_i > t/n] = (e^{-(t/n)/\theta})^n = e^{-t/\theta}. \end{aligned}$$

By the Rao–Blackwell theorem, our aim is to find a new estimator $\delta^*(t) = E[\delta(t) \mid D(T)=d,\ S(T)=s]$. We have

$$\begin{aligned} \delta^*(t) &= P[\min(X_1,\dots,X_n) > t/n \mid D(T)=d, S(T)=s] \\ &= P\Big[\min(Y_1,\dots,Y_d) > t/n \;\Big|\; \sum_{i=1}^{d} Y_i = s,\ Y_k > T,\ k=1,\dots,d\Big]. \end{aligned}$$

Observe that if $d \leqslant n-1$, then $\min(X_1,\dots,X_n) \leqslant T$ and therefore $\delta^*(t) = 0$. Otherwise, for $n > 1$ and $d = n$,

$$\delta^*(t) = \frac{\int_A q_{\theta n}(y_1,\dots,y_{n-1},s)\,dy_1\cdots dy_{n-1}}{\int_B q_{\theta n}(y_1,\dots,y_{n-1},s)\,dy_1\cdots dy_{n-1}},$$

where

(a) A is the volume of the region in $\mathbf{R}^{n-1}$ determined by

$$y_1+\cdots+y_{n-1} < s - t/n \quad \text{and} \quad y_i > t/n,\ i=1,\dots,n-1;$$

(b) B is the volume of the region in $\mathbf{R}^{n-1}$ determined by

$$y_1+\cdots+y_{n-1} < s - T \quad \text{and} \quad y_i > T,\ i=1,\dots,n-1.$$

Applying Lemma 2.1 for $d = n$, we get

$$\delta^*(t) = \frac{V_n(s, t/n)}{V_n(s, T)} = \left[\frac{s-t}{s-nT}\right]_+^{n-1}.$$

Note that this formula is also valid if $d = n = 1$, because then

$$\delta^*(t) = P(Y_1 > t \mid Y_1 = s,\, Y_1 > t) = 1\{s > t\}.$$

Finally, we summarize our considerations in the following theorem.

THEOREM 7.1. *For $t > nT$, the function*

$$\delta^*(t) = \begin{cases} 0 & \text{for } d \leqslant n-1, \\ \left[\dfrac{s-t}{s-nT}\right]_+^{n-1} & \text{for } d = n, \end{cases} \tag{7.1}$$

is an unbiased estimator for the exponential reliability $R(t)$.

The estimator δ^* has limited applications because it is defined only for $t > nT$. For $n \geqslant 2$ this estimator can be worse than $\hat{R}_0(t)$ because

$$\lim_{t \to nT^+} (\mathrm{Var}[\delta^*(t)] - \mathrm{Var}[\hat{R}_0(t)]) = \left(1 - \frac{1}{n}\right) e^{-nT/\theta}(1 - e^{-nT/\theta}) > 0.$$

Hence this estimator is of theoretical rather than practical importance and we will not investigate further.

8. NUMERICAL EXPERIMENTS

In this section we present results of a simulation study and a real life example. In the first example, all quantities were calculated based on 10000 samples generated from the exponential distribution with the following censoring scenario: censoring time $T = 1$ and $\theta = 1$ (i.e. censoring rate is 63.1%). The second example concerns the analysis of real data.

EXAMPLE 1. Our aim is to compare the following estimators of reliability: $\hat{R}_{ml}$ (maximum likelihood estimator), $\hat{R}$ (unbiased estimator) and $\tilde{R}$ (modified unbiased estimator when $t < T$) in terms of biases and MSEs (mean square errors). The times at which we calculate reliabilities are $t = 0.33, 0.67, 1.0, 1.33, 1.67$. For $t = 1$ we calculated $\hat{R}$ as d/n. Because for $t = 0.33$, 0.67 an unbiased estimator for reliability does not exist, we applied the modified unbiased estimator $\tilde{R}$ with $\varepsilon = 0.01$. The results are shown in Table 1. From the simulation results we observe that the estimator $\hat{R}$ has an empirical bias close to zero in all cases. However, the estimators $\tilde{R}$ and $\hat{R}_{ml}$ are biased, which is clearly seen in the case of small samples,

TABLE 1. Simulation results for reliability estimators: average biases and average mean square errors (MSE) for reliability estimators: the modified unbiased estimator $\tilde{R}$ (for $t < 1$), unbiased estimator $\hat{R}$ (for $t \geqslant 1$), and the maximum likelihood estimator $\hat{R}_{ml}$.

	Bias $(\tilde{R}, \hat{R}_{ml})$ for $t < 1$		**Bias** $(\hat{R}, \hat{R}_{ml})$ for $t \geqslant 1$		
n	$t = 0.33$	$t = 0.67$	$t = 1.0$	$t = 1.33$	$t = 1.67$
1	$(-0.455, 0.105)$	$(-0.322, 0.166)$	$(-0.005, 0.196)$	$(-0.001, 0.208)$	$(0.001, 0.207)$
2	$(-0.288, 0.041)$	$(-0.204, 0.067)$	$(-0.002, 0.080)$	$(0.001, 0.088)$	$(0.001, 0.088)$
5	$(-0.076, -0.018)$	$(-0.054, -0.016)$	$(0.001, -0.008)$	$(-0.001, 0.006)$	$(-0.001, 0.009)$
10	$(-0.009, -0.022)$	$(-0.005, -0.024)$	$(-0.001, -0.019)$	$(0.000, -0.011)$	$(0.000, -0.004)$
30	$(-0.001, -0.008)$	$(-0.001, -0.009)$	$(-0.001, -0.008)$	$(0.000, -0.005)$	$(0.000, -0.003)$
	MSE $(\tilde{R}, \hat{R}_{ml})$ for $t < 1$		**MSE** $(\hat{R}, \hat{R}_{ml})$ for $t \geqslant 1$		
1	$(0.401, 0.015)$	$(0.257, 0.038)$	$(0.231, 0.054)$	$(0.194, 0.062)$	$(0.154, 0.063)$
2	$(0.308, 0.008)$	$(0.232, 0.019)$	$(0.118, 0.026)$	$(0.092, 0.028)$	$(0.072, 0.027)$
5	$(0.140, 0.009)$	$(0.142, 0.018)$	$(0.047, 0.019)$	$(0.033, 0.017)$	$(0.025, 0.014)$
10	$(0.043, 0.008)$	$(0.060, 0.014)$	$(0.023, 0.014)$	$(0.015, 0.012)$	$(0.011, 0.009)$
30	$(0.006, 0.002)$	$(0.013, 0.004)$	$(0.008, 0.005)$	$(0.005, 0.004)$	$(0.004, 0.003)$

but this bias decreases as n increases. The estimator $\hat{R}_{ml}$ is better than $\hat{R}$ and $\tilde{R}$ in terms of MSE, but this difference decreases as n increases.

EXAMPLE 2 (Real data). In this example we analyze a real data set to demonstrate the computation of exponential reliability together with confidence intervals. The following data (see Nelson [11, p. 124]) represent the time to breakdown (in seconds) of an insulating fluid tested at 45kV constant voltage stress: $1^-, 1^-, 1^-$, 2, 2, 3, 9, 13, 47, 50, 55, 71. Some breakdowns occurred in less than one second, making them too early to be recorded. These times are left censored and are denoted here by the label " $-$ ".

The given times do not follow an exponential distribution. Nelson examined the validity of these data with the Weibull distribution using probability plots.

One can verify that transforming the data using the function $f(t) = \sqrt{t},\ t > 0$, results in the exponential distribution (the p-value of the Anderson–Darling test increased from 0.03 to 0.9).

Assume now that the data follow the distribution $Y = X^2$, where X is the exponential distribution with parameter θ. In order to estimate $P(Y > t)$, we estimate the exponential reliability $R(\sqrt{t})$, where $t = 0.5, 2, 5, 10, 20, 30, 50, 70$. For estimated exponential reliabilities we fitted the exponential function $e^{-\sqrt{t}/\theta}$ by the method of least squares, obtaining $\hat{\theta} = 3.66$ with residual sum of squares of 0.0002.

In order to find confidence intervals for these reliabilities we applied parametric bootstrap. For each time we generated 1000 bootstrap samples from the exponential distribution with $\theta = 3.66$ and calculated estimators of reliabilities. The results (denoted as CI_B) are displayed in Table 2.

The second approach to interval estimation presented here is based on solving the following equations with respect to R:

$$P(\hat{R} \geqslant R_0 \mid R = R_L) = \alpha/2, \quad P(\hat{R} \leqslant R_0 \mid R = R_U) = \alpha/2,$$

where R_0 is the observed value of the estimator $\hat{R}$. Then the interval (R_L, R_U) is the exact confidence interval for reliability at confidence level $1 - \alpha$.

For the estimator $\hat{R}$ we calculate its distribution as follows:

$$P(\hat{R} > R) = \sum_{d=1}^{n} P(\hat{R} > R \mid D = d) \cdot P(D = d),$$

where (for $d > 0$)

$$P(\hat{R} > R \mid D = d) = P\left(S > \frac{t - T}{1 - A} + dT\right), \quad A = \left(\frac{nR}{d}\right)^{1/(d-1)}.$$

Finally, for $S = \sum_{i=1}^{d} Y_i$ we easily conclude that

$$P(S > a) = \Gamma\left(d, \frac{a - dT}{\theta}\right)/(d-1)!.$$

Here Γ denotes the incomplete Gamma function. Replacing the estimator $\hat{R}$ by $\tilde{R}$, we also get confidence intervals for $t < T$. The numerical results (denoted as CI_E) are displayed in Table 2. Observe that the confidence intervals CI_B and CI_E are similar to each other.

TABLE 2. The estimators of the reliability and 95% bootstrap (CI_B) and exact (CI_E) confidence intervals for real data.

t	0.5	2	5	10	20	30	50	70
$\hat{R}(t)$	0.83	0.68	0.55	0.43	0.30	0.22	0.14	0.09
CI_B	(0.75,0.97)	(0.49,0.90)	(0.35,0.76)	(0.24,0.68)	(0.11,0.51)	(0.04,0.41)	(0.00,0.27)	(0.00,0.18)
CI_E	(0.73,0.92)	(0.44,0.86)	(0.34,0.74)	(0.23,0.64)	(0.13,0.52)	(0.08,0.45)	(0.04,0.36)	(0.02,0.30)

9. CONCLUDING REMARKS

We considered unbiased point estimation for exponential reliability. In the literature there are numerous works concerning parametric estimation for the exponential distribution, mainly for complete samples and censored samples (right or left), grouped samples, and many other types. The most common method of obtaining estimators is the maximum likelihood method, which is due to the asymptotic properties of these estimators like consistency, efficiency, normality. Constructing

unbiased estimators is technically more difficult than MLEs, because it involves solving integral equations or calculating conditional expectations.

Moreover, the problem of unbiased estimation of the survival function for the exponential function has not been considered yet. It was done for complete samples and right censored samples. Hence this paper fills a gap in the theory of point estimation. The unbiased estimator $\hat{R}$ for $T = 0$, which means the absence of censoring, coincides with the classical minimum variance unbiased estimator of exponential reliability. Moreover, when $T > 0$ and except for the trivial case $n = 1$, we cannot conclude that $\hat{R}$ is the unique unbiased estimator which is based on the sufficient statistic $(D(T), S(T))$ or a minimum variance unbiased estimator, because the sufficient statistic is not complete. A similar comment was made by Bartoszewicz [2] for the case of right censoring.

Unbiased estimators are preferred to MLEs especially for small samples since biases of $\hat{R}_{ml}$ are larger than those of $\hat{R}$. Furthermore, MLEs may not exist and then it is also reasonable to use the unbiased estimator $\hat{R}$. On the other hand, MLEs are generally better than unbiased estimators with respect to MSE, but the discrepancy decreases with the sample size.

Acknowledgments. The author thanks the referees for their comments and suggestions which greatly improved the paper.

REFERENCES

[1] N. Balakrishnan and J. Varadan, *Approximate MLEs for the location and scale parameters of the extreme value distribution with censoring*, IEEE Trans. Reliability 40 (1991), 146–151.

[2] J. Bartoszewicz, *Estimation of reliability in the exponential case (I)*, Appl. Math. (Warsaw) 14 (1974), 185–194.

[3] A. P. Basu, *Estimates of reliability for some distributions useful in life testing*, Technometrics 6 (1969), 215–219.

[4] J. Hoem, *The sampling distribution of an estimator arising in connection with the truncated exponential distribution*, Ann. Math. Statist. 40 (1969), 702–703.

[5] J. P. Klein and M. L. Moeschberger, *Survival Analysis: Techniques for Censored and Truncated Data*, 2nd ed., Springer, New York, 2003.

[6] A. G. Laurent, *Conditional distribution of order statistics and distribution of the reduced i-th order statistic of the exponential model*, Ann. Math. Statist. 34 (1963), 652–657.

[7] E. L. Lehmann and H. Scheffé, *Completeness, similar regions and unbiased estimation: Part I*, Sankhyā 10 (1950), 305–340.

[8] E. L. Lehmann and H. Scheffé, *Completeness, similar regions and unbiased estimation: Part II*, Sankhyā 15 (1955), 219–236.

[9] K.-J. Lui, D. N. Lawrence, W. M. Morgan, T. A. Peterman, H. W. Haverkos and D. J. Bregman, *A model-based approach for estimating the mean incubation period of transfusion-associated acquired immunodeficiency syndrome*, Proc. Nat. Acad. Sci. USA 83 (1986), 3051–3055.

[10] S. Mitra and D. Kundu, *Analysis of left censored data from the generalized exponential distribution*, J. Statist. Comput. Simul. 78 (2008), 669–679.

[11] W. B. Nelson, *Applied Life Data Analysis*, Wiley, New York, 1982.

[12] E. L. Pugh, *The best estimate of reliability in the exponential case*, Operations Research 11 (1963), 57–61.

[13] K. Sarkadi, *On two theorems of Bartoszewicz*, Appl. Math. (Warsaw) 17 (1980), 1–3.

[14] R. F. Tate, *Unbiased estimation: functions of location and scale parameters*, Ann. Math. Statist. 30 (1959), 341–366.
[15] V. G. Voinov and M. S. Nikulin, *Unbiased Estimators and Their Applications, Vol.1: Univariate Case*, Kluwer Acad. Publ., Dordrecht, 1993.

Piotr Bolesław Nowak
Institute of Economic Sciences
University of Wrocław
50-145 Wrocław, Poland
E-mail: piotr.nowak2@uwr.edu.pl

Received 11.11.2024;
accepted 3.6.2025