Some Applications of Near-Order Statistics in Two-Parameter Exponential Distribution

Masoumeh Akbari¹*, Mahboubeh Akbari²

¹Department of Statistics, University of Mazandaran, Babolsar, Mazandaran 47416-95447, Iran
²Department of Statistics, University of Birjand, Birjand, 97175-615, Iran

ABSTRACT

In this paper, some characterization results for exponential distribution are established. The results are concluded in terms of number of observations near of order statistics. It is shown that its probability mass function and its first moment can characterize the exponential distribution. Also, an estimator based on near-order statistics is introduced for tail thickness of exponential distribution.

1. INTRODUCTION

Two-parameter exponential distribution is the simplest lifetime distributions that is useable in survival analysis and reliability theory. So far, more results of characterization of exponential distribution have been obtained that some of them are based on order statistics. Let $X_1, X_2, \cdots, X_n$ be independent and continuous random variables. The corresponding order statistics are the $X_i$’s arranged in non-decreasing order, denoted by $X_{1:n} < X_{2:n} < \cdots < X_{n:n}$. More details concerning order statistics can be seen in David and Nagaraja [1] and Arnold et al. [2]. Desu [3] proved that distribution of population is exponential if and only if $nX_{1:n} \overset{d}{=} X_1$, for all $n \geq 1$, where the notation $\overset{d}{=}$ states the equality in distribution. Also, more characterization results of exponential distribution can be seen in Galambos and Kotz [4] and Ahsanullah and Hamedani [5]. Recently, the problem of number of observations near the order statistics is considered. At first, Pakes and Stutel [6] defined the number of observations within $a$ of the sample maximum $X_{n:n}$ as

$$K_n(a) = \#\{j = 1, \ldots, n, X_j \in (X_{n:n} - a, X_{n:n})\}.$$ 

Then, this definition was developed for the number of observations falling in the open left and right $a$–vicinity of the $k$th order statistics by Pakes and Li [7] and Balakrishnan and Stepanov [8], respectively. After that, following two random variables have been considered in the literature

$$K_-(n, k, a) = \#\{j = 1, \ldots, n, X_j \in (X_{k:n} - a, X_{k:n})\},$$

$$K_+(n, k, a) = \#\{j = 1, \ldots, n, X_j \in (X_{k:n}, X_{k:n} + a)\},$$

where $k = 1, 2, \ldots, n$ and $a > 0$ is a constant. So far, a lot of results for counting random variables in (1) and (2) have been obtained such that the most of them are focused on their asymptotic behavior under different conditions. For example, Pakes and Li [7], Balakrishnan and Stepanov [8,9], Dembińska et al. [10], Dembińska [11–14], Dembińska and Iliopoulos [15], Pakes [16–18] and Iliopoulos et al. [19]. There are few results of statistical inferences based on (1) and (2), for example, Müller [20] and Hashorva and Hüsler [21] considered the estimation of tails based on them. Also, some results have been obtained in terms of near-order insurance claims (See, e.g., Li and Pakes [22], Hashorva [23,24]).

*Corresponding author. Email: m.akbari@umz.ac.ir
In this paper, we will prove some characterization results of two-parameter exponential distribution based on these counting random variables which are stated in sections 2 and 3. An estimator of $e^{-\sigma}$ is introduced in section 4 and some properties of this estimator are discussed. Further, its performance is compared with the maximum likelihood estimator (MLE) through simulation.

2. CHARACTERIZATION BASED ON DISTRIBUTIONAL RESULTS

Let $X$ be a random variable having two-parameter exponential distribution with parameters $\mu$ and $\sigma$, denoted by Exp($\mu, \sigma$). Then the cumulative distribution function (CDF) of $X$ is

$$F(x) = 1 - e^{-\sigma(x-\mu)}, \quad x \geq \mu.$$  \hspace{1cm} (3)

According to (1), the probability mass function (pmf) of $K_+(n, k, a)$ for any $j = 0, 1, \cdots, n-k$, have been obtained as (See Dęmbińska et al. [10])

$$P(K_+(n, k, a) = j) = \frac{n!}{j!(k-1)!(n-k-j)!} \int_{\mathbb{R}} F^{n-k-j}(x+a)(F(x)-\tilde{F}(x+a))^jF^{k-1}(x)dF(x). \hspace{1cm} (4)$$

Further, it is easy to verify that the pmf of $K_-(n, k, a)$ for any $j = 0, 1, \cdots, k-1$ is

$$P(K_-(n, k, a) = j) = \frac{n!}{j!(k-j-1)!(n-k)!} \int_{\mathbb{R}} F^{k-j-1}(x-a)(F(x)-F(x-a))^j\tilde{F}^{n-k}(x)dF(x). \hspace{1cm} (5)$$

Now, assume that $F(\cdot)$ has a form as (3). Substituting in Eq. (4), we conclude easily that $K_+(n, k, a)$ has binomial distribution with parameters $(n-k)$ and $(1 - e^{-\sigma a})$, that is

$$P(K_+(n, k, a) = j) = \binom{n-k}{j} e^{-\sigma a(n-k-j)}(1 - e^{-\sigma a})^j,$$  \hspace{1cm} (6)

for any $j = 0, 1, \ldots, n-k$. Therefore, the expected value of $K_+(n, k, a)$ is given by

$$E(K_+(n, k, a)) = (n-k)(1 - e^{-\sigma a}). \hspace{1cm} (7)$$

In this section, we will show that Eqs. (6) and (7) can characterize exponential distribution. The results are proved through properties of completeness sequence function. So, we firstly define complete sequence function and recall some well-known theorems.

**Definition 2.1.** A sequence $\{\Phi_n\}_{n \geq 1}$ of elements of a Hilbert space $H$ is called complete if the only element which is orthogonal to every $\{\Phi_n\}$ is the null element, that is,

$$\langle f, \Phi_n \rangle \Rightarrow f = 0,$$  \hspace{1cm} (8)

where $\langle \cdot, \cdot \rangle$ denotes the inner product of $H$. The Hilbert space $L^2(0, 1)$ is considered here whose inner product is given by $\langle f, g \rangle = \int_0^1 f(x)g(x)dx$, where $f$ and $g$ are two real-valued square integrable functions defined on $(0, 1)$.

The sequence $\{x^n, \ n \geq 1\}$ is the most important complete sequence function. Even under conditions a subsequence of it is a complete sequence that is stated in the following theorem and is well-known as the Müntz theorem.

**Theorem 2.1.** Higgins ([25], p. 95) The set $\{x^{n_1}, x^{n_2}, \cdots; 1 \leq n_1 < n_2 < \cdots\}$ forms a complete sequence in $L^2(0, 1)$ if and only if

$$\sum_{j=1}^{\infty} n_j^{-1} = \infty.$$  \hspace{1cm} (9)

We refer the reader to Higgins [25] for Hilbert space and complete sequence function.

**Theorem 2.2.** Let $X_1, X_2, \ldots, X_n$ be continuous random variables with CDF $F$. Then $F$ has exponential distribution Exp($\mu, \sigma$) if and only if

$$P(K_+(n, k, a) = j_0) = (1 - e^{-\sigma a})^{j_0}, \quad a > 0,$$  \hspace{1cm} (10)

for a fixed $j_0 \in \{0, 1, \ldots, n-1\}$ and $k = n - j_0$. 
Proof. If $X$ has exponential distribution, then Eq. (10) is easily obtain. Conversely, let (10) holds, then

$$0 = P(K_4(n, k, a) = j_0) - (1 - e^{-\sigma a})^b.$$ 

From (4) we have

$$0 = \frac{nt}{j_0! (n - j_0 - 1)!} \int_0^1 (F(x + a) - F(x)) h_{F(x)} F^{n-1}(x) dF(x) - (1 - e^{-\sigma a})^b.$$ 

Further, using Minkowski's inequality for the quantity

$$= \frac{nt}{j_0! (n - j_0 - 1)!} \int_0^1 (F(F^{-1}(u) + a) - u) h_{F^{-1}}(1 - u)^b a du - (1 - e^{-\sigma a})^b \int_0^1 (1 - u)^b du$$

$$= \int_0^1 [(F(F^{-1}(u) + a) - u) h_{F^{-1}}(1 - u)^b a du.]$$

Taking $\eta(u) = (F(F^{-1}(u) + a) - u)^b$, we have

$$\int_0^1 \eta^2(u) du = \int_0^1 (F(F^{-1}(u) + a) - u)^{2b} du \leq 1.$$ 

The above inequality shows that $\eta(u) \in L^2(0, 1)$. Similarly we have the quantity

$$(1 - e^{-\sigma a})^b(1 - u)^b \in L^2(0, 1).$$ 

Further, using Minkowski's inequality for the quantity

$$\Theta(u) = [F(F^{-1}(u) + a) - u]^b = (1 - e^{-\sigma a})^b(1 - u)^b,$$ 

we get $\Theta(u) \in L^2(0, 1)$. So, if (11) holds for any $n \geq j_0 + 1$, then using complete property of sequence $u^{n-1}$, $n \geq j_0 + 1$, we have

$$(1 - e^{-\sigma a})^b(1 - u)^b = 0.$$ 

From (12) is concluded that

$$\frac{F(t + a)}{F(t)} = e^{-\sigma t}, \quad a > 0, \quad t \in \mathbb{R},$$

where $t = F^{-1}(u)$. The most general solution of (13) is the function $F(x) = ce^{-\sigma x}$, $x > 0$, where $c$ is a constant. (See Aczél [26], pp. 17–18). Taking $c = e^{\sigma \mu}$, the proof is completed.

The other characterization of exponential distribution is based on first moment of $K_4(n, k, a)$ which is stated in the next theorem.

**Theorem 2.3.** Let $X_1, X_2, \ldots, X_n$ be continuous random variables with CDF $F$. Then $F$ has exponential distribution $\text{Exp}(\mu, \sigma)$ if and only if for a fixed $k \geq 1$ and every $a > 0$, following quantity holds.

$$E(K_4(n, k, a)) = (n - k)(1 - e^{-\sigma a}), \quad n \geq k.$$ 

**Proof.** If $X$ has $\text{Exp}(\mu, \sigma)$, then from (7) proof of the necessity is concluded. Conversely, Let us assume that

$$E(K_4(n, k, a)) = (n - k)(1 - e^{-\sigma a}).$$

On the other hand, from (i) the first moment of $K_4(n, k, a)$ is given by

$$E(K_4(n, k, a)) = n - k - k(n - k) \left( \frac{n}{k} \right) \int_0^1 \frac{F(F^{-1}(u) + a) u^{n-1}(1 - u)^{n-1}}{1 - u} du.$$ 

The quantities (14) and (15) result in

$$k \left( \frac{n}{k} \right) \int_0^1 \left( \frac{F(F^{-1}(u) + a)}{1 - u} - e^{-\sigma a} \right) u^{n-1}(1 - u)^{n-1} du = 0.$$
If Eq. (16) holds for any \( n \geq k \), then by completeness property of sequence \(((1 - u)^{n-k}, n \geq k)\), we have

\[
\left\{ \frac{F^{-1}(u) + a}{1 - u} - e^{-\sigma a} \right\} d^{k-1} = 0.
\]

(17)

Similar to the proof of Theorem 2.2, \( \tilde{F}(x) = ce^{-\sigma x} \) is the most general solution of (17) and this completes the proof. □

**Remark 2.1.** According to Müntz theorem that is stated in Theorem 2.1, the all results of this section are true for any increasing subsequence \( \{n_p, j \geq 1\} \) which satisfies in (9) instead of for all \( n \geq 1 \).

### 3. Characterization Based on Dependency Assumptions

It is known that only for the exponential distribution, any two non-overlapping spacings will be independent (See, e.g., Arnold et al. [2]). Let us define two spacings \( W_1 \) and \( W_2 \) as follows

\[
W_1 = X_{k+1:n} - X_{k+1-j:n} \quad \text{and} \quad W_2 = X_{k+2:j:n} - X_{k+1:j:n},
\]

for any \( j_1 = 2, \cdots, k - 1 \) and \( j_2 = 2, \cdots, n - k \). According to definitions \( K_+ (n, k, b) \) and \( K_- (n, k, a) \), we can write the following equivalent events

\[
\{K_+(n, k, a) \leq j_1 - 1\} = \{W_1 \geq a\},
\]

and

\[
\{K_-(n, k, a) \leq j_2 - 1\} = \{W_2 \geq b\}.
\]

So,

\[
P(K_-(n, k, a) \leq j_2 - 1) = P(W_1 \geq a),
\]

(18)

and

\[
P(K_+(n, k, b) \leq j_2 - 1) = P(W_2 \geq b).
\]

(19)

From (1) and (2), one can obtain easily the probability generating functions (pgf) of \( K_-(n, k, a) \) and \( K_+(n, k, b) \) as follows (see Balakrishnan and Stepanov [8])

\[
\Pi_-(n, k, a, s) = E \left( s^{K_-(n,k,a)} \right) = n \left( \frac{n-1}{n-k} \right) \int \{(1 - s)F(x - a) + sF(x)\}^{k-1} (1 - F(x))^{n-k} dF(x).
\]

(20)

\[
\Pi_+(n, k, b, t) = E \left( t^{K_+(n,k,b)} \right) = n \left( \frac{n-1}{n-k} \right) \int [tF(x + b) - F(x)] + 1 - F(x + b) \}^{n-k} F(x)^{k-1} dF(x).
\]

(21)

It also follows that the joint pgf \( K_- (n, k, a) \) and \( K_+ (n, k, b) \) is

\[
E \left( s^{K_-(n,k,a)} t^{K_+(n,k,b)} \right) = n \left( \frac{n-1}{n-k} \right) \int \{(1 - s)F(x - a) - sF(x)\}^{k-1} \times \{(1 - t)F(x + b) + tF(x)\}^{n-k} dF(x).
\]

(22)

In the next theorem, we show an another characterization for exponential distribution based on independent near-order statistics.

**Theorem 3.1.** Let \( X_1, X_2, \cdots, X_n \) be continuous random variables with CDF \( F \). Then \( F \) has \( \text{Exp} (\mu, \sigma) \) if and only if \( K_- (n, k, a) \) and \( K_+ (n, k, b) \) be independent for a fixed \( k \geq 1 \) and for any \( a > 0 \) and \( b > 0 \).

**Proof of the necessity.** It is enough to show that joint pgf of \( K_- (n, k, a) \) and \( K_+ (n, k, b) \) is equal with multiplication of their pgfs. Let \( F \) be \( \text{Exp} (\mu, \sigma) \). Substituting it in Eqs. (20–22), we have

\[
\Pi_{\mu}^F(n, k, a, s) = n \left( \frac{n-1}{n-k} \right) \int \{(1 - s)(1 - u)e^{\sigma u} + s(1 - u)e^{\sigma u} \}^{k-1} u^{n-k} du,
\]

(23)
Proof of the sufficiency. and this completes the proof of necessity.

Pakes and Stutel [6] introduced following index of tail thickness when

4. AN ESTIMATOR BASED ON NEAR-ORDER STATISTIC

The quantity (26) shows that the spacings

The above equality can be written as

and this completes the proof of necessity.

Proof of the sufficiency. Suppose that counting random variables \( K_-(n, k, a) \) and \( K_+(n, k, b) \) be independent. Using (18) and (19), we have

\[
P(W_1 \geq a, W_2 \geq b) = P(K_-(n, k, a) \leq j_1 - 1, K_+(n, k, b) \leq j_2 - 1)
\]

\[
= P(K_-(n, k, a) \leq j_1 - 1) P(K_+(n, k, b) \leq j_2 - 1)
\]

\[
= P(W_1 \geq a) P(W_2 \geq b).
\]

The quantity (26) shows that the spacings \( W_1 \) and \( W_2 \) are independent. So, the proof is completed.

4. AN ESTIMATOR BASED ON NEAR-ORDER STATISTIC

Pakes and Stutel [6] introduced following index of tail thickness when \( r_x = \inf \{ x, F(x) = 1 \} = \infty \), as

\[
\gamma(a) = \lim_{x \to \infty} \frac{F(x) - F(x + a)}{F(x)}.
\]  (27)

A tail is called as “thick” tail when \( \gamma(a) = 0 \), “medium” tail when \( 0 < \gamma(a) < 1 \) and “thin” tail when \( \gamma(a) = 1 \). Assume that \( X \) has exponential distribution. Substituting its CDF into (27), imply that

\[
\gamma(a) = 1 - e^{-\sigma a}.
\]  (28)

It is well-known that the MLE of unknown scale parameter \( \sigma \) is \( n \left( \sum_{i=1}^{n} (X_i - X_{1:n}) \right)^{-1} \), when the underlying distribution is \( \text{Exp}(\mu, \sigma) \). So, one estimator for \( e^{-\sigma a} \) based on MLE can be considered as

\[
T_1 = \exp \left\{ -na \left( \sum_{i=1}^{n} (X_i - X_{1:n}) \right)^{-1} \right\}.
\]  (29)

Following, we introduce an estimator for \( e^{-\sigma a} \) based on near-order statistic. From (6), the pmf of \( K_+(n, k, a) \) can be written as

\[
P(K_+(n, k, a) = j) = \binom{n - k}{j} a^{-\sigma a(n - k)} \exp \left( \frac{1 - e^{-\sigma a}}{e^{-\sigma a}} \right).
\]  (30)

Eq. (30) shows that the counting random variable \( K_+(n, k, a) \) belongs to the one-parameter exponential family. Therefore, \( K_+(n, k, a) \) is a sufficient and complete statistic for \( e^{-\sigma a} \). According to expectation of \( K_+(n, k, a) \), an unbiased estimator for \( e^{-\sigma a} \) is equal to

\[
T_2 = 1 - \frac{K_+(n, k, a)}{n - 1}.
\]  (31)
So, the estimator $T_2$ is uniformly minimum-variance unbiased estimator (UMVUE) and its variance or minimum square error (MSE) is as follows

$$MSE(T_2) = \frac{n-k}{(n-1)^2}(1-e^{-\sigma a})e^{-\sigma a}. \tag{32}$$

Further, (31) and (32) imply that $T_2$ is a consistent estimator for $e^{-\sigma a}$. The performance of two estimators $T_1$ and $T_2$ is comparable through their MSE. But it is difficult to calculate MSE of $T_1$ theoretically. So, we compare them numerically. In this study, we explore the MSE of $T_1$ and $T_2$ under different $\mu$, $a$ and $k$ which are stated in Figure 1. Also, this results are obtained based on 2000 bootstrap samples.

Our simulation results demonstrate that the performance of $T_1$ and $T_2$ has little differences with increasing $a$.

So, the obtained results show that with choosing appropriate $k$, the estimator $T_2$ can be considered as a good estimator for parameter $e^{-\sigma a}$.

An exact confidence interval for $e^{-\sigma a}$ when $a$ is known can be obtained by this fact that a confidence interval is available for $\sigma$ in two-parameter exponential distribution. So, the $100 (1-\alpha)$% interval confidence for $e^{-\sigma a}$ is given by

$$\exp\left\{\exp\left(-2a\chi^2\right)\right\} \leq e^{-\sigma a} \leq \exp\left\{\exp\left(-2a\chi^2\right)\right\},$$

where $\chi^2_{m,\alpha}$ denotes the $100 (1-\alpha)$th percentile of the central chi-square distribution with $m$ degree of freedom.

Now, we present an asymptotic confidence interval for $e^{-\sigma a}$ based on counting random variable $K_+ (n, k, a)$ which is stated in the following remark.

**Remark 4.1.** According to distribution of $K_+ (n, k, a)$, it can be considered as sum of independent and identically distributed random variables from binomial $(1, 1-e^{-\sigma a})$. So the conditions of central limit theorem for random variable $T_2$ hold and we have

$$\frac{(n-1)(T_2 - e^{-\sigma a})}{\sqrt{(n-k)e^{-\sigma a}(1-e^{-\sigma a})}} \xrightarrow{d} \mathcal{N}(0, 1). \tag{33}$$

Therefrom (33), we can construct asymptotically confidence interval for $e^{-\sigma a}$ by solving following inequality

$$e^{-2\sigma a}\left(\frac{Z^2_\alpha}{2} + n - 1\right) - e^{-\sigma a}\left(\frac{Z^2_\alpha}{2} + 2(n-1)T_2\right) + (n-1)T_2^2 < 0.$$

## 5. CONCLUSION

In this paper, we have shown some applications of counting random variable $K_+ (n, k, a)$ for two-parameter exponential distribution. We believe that the results of the second and third section can be used in the construction of testing goodness-of-fit for exponentiality which sometimes can be more efficient or more robust than others. See, Nikitin [27] for more details on application of characterization in goodness-of-fit test.
CONFLICT OF INTEREST
The authors declare that there is no potential conflict of interest related to this study.

AUTHORS’ CONTRIBUTIONS
The authors contributed equally to this work.

ACKNOWLEDGMENTS
The authors would like to thank the Editor in Chief, the Associate Editor and two anonymous reviewer for their valuable comments.

REFERENCES