Characterization of Exponential Distribution Through Normalized Spacing of Generalized Order Statistics

M. J. S. Khan*, S. Iqrar, M. Faizan

Department of Statistics and Operations Research, Aligarh Muslim University, Aligarh-202 002, India

ARTICLE INFO

Article History
Received 19 Oct 2017
Accepted 24 Dec 2018

Keywords
Continuous distribution
Generalized order statistics
Meijer’s G-function
Characterization
AMS Subject Classification: 62E10,
62G30, 60E05

ABSTRACT

In this paper, exponential distribution is characterized by normalized spacing of generalized order statistics (gos) using Meijer’s G-function. While the necessary part of the theorem was given by U. Kamps, E. Cramer, Statistics. 35 (2001), 269–280, we have given an easy proof of sufficient part in this paper. This paper contains the result of characterization of exponential distribution through normalized spacing of order statistics, sequential order statistics, progressive type II censoring and record values. Also, by simulation study, we have shown that the confidence interval based on upper records is shorter in length than asymptotic confidence interval constructed by maximum likelihood estimate.

1. INTRODUCTION

Ordered random variables like order statistics, record values, sequential order statistics and progressive type II censored order statistics frequently arise in real life situations. For application of ordered random variables, one may refer to [1], to [2] for order statistics, to [3] and [4] for record values and to [5] for progressive type II censored order statistics and reference cited therein. The concept of generalized order statistics (gos) was first introduced by [6]. Goss includes various important models related to ordered random variables like order statistics, record values, sequential order statistics, order statistics with non-integral sample size and progressive type II censored order statistics etc.

Let \(X_1, X_2, \ldots, X_n\) be a sequence of independent and identically distributed (iid) random variables (rv) with the distribution function (df) \(F(x)\) and the probability density function (pdf) \(f(x)\). Let \(n \in \mathbb{N}, k > 0, \bar{m} = (m_1, m_2, \ldots, m_{n-1}) \in \mathbb{R}^{n-1}\) and \(M_r = \sum_{i=r}^{n-1} m_i\), such that \(\gamma_r = k + n - r + M_r > 0\) for all \(r \in \{1, 2, \ldots, n - 1\}\), then \(X(1, n, \bar{m}, k), X(2, n, \bar{m}, k), \ldots, X(n, n, \bar{m}, k)\) are said to be the gos if their joint pdf is given by

\[k \left( \prod_{j=1}^{n-1} \gamma_j \right) \left( \prod_{i=1}^{n-1} [1 - F(x_i)]^{m_i} f(x_i) \right) \left[ 1 - F(x_n) \right]^{k-1} f(x_n) \tag{1}\]

on the cone \(F^{-1}(0+) < x_1 \leq x_2 \leq \cdots \leq x_n < F^{-1}(1)\) of \(\mathbb{R}^n\). By adjusting the parameters of gos, one may get almost all important models related to ordered random variables. For example, at \(m = 0, k = 1\) i.e. \(\gamma_r = n - i + 1; \beta_i > 0\), we get ordinary order statistics. At \(m = -1, k \in \mathbb{N}\) i.e. \(\gamma_r = k\), we shall get \(k^{th}\) record values. Similarly sequential order statistics (\(\eta_r = (n - i + 1) \beta_i; \beta_1, \beta_2, \cdots, \beta_n > 0\)), order statistics with non-integral sample size (\(\gamma_r = \beta_i - i + 1; \beta > 0\)), Pfeifer record values (\(\gamma_r = \beta_i; \beta_1, \beta_2, \cdots, \beta_n > 0\)) and progressive type II censored order statistics \((N, n \in \mathbb{N}, N = n + \sum_{j=1}^{n} R_j, R_j \in \mathbb{N}_0\) and \(\gamma_r = N - \sum_{j=1}^{n-1} R_j - j + 1, 1 \leq j \leq n\) ) can be obtained as specific cases of gos.

Let \(P_S\) stands for the probability measure on \(\mathbb{R}\) determined by \(F(x)\). Define \(\alpha = \inf \{x \in \mathbb{R} : F(x) > 0\}\) and \(\beta = \sup \{x \in \mathbb{R} : F(x) < 1\}\), then the pdf of \(X(r, n, \bar{m}, k)\) with respect to a measure \(P_S\) can be expressed as (see [7]):

\[f_r(x) = c_{r-1} G_r \left( F(x) \right) \prod_{i=1}^{r} \gamma_i I_{\alpha, \beta}(x), \tag{2}\]
Lemma 2.1. Let \( X \) be the \( r \)-th gos defined in (2), then
\[
\frac{1}{c_{r-1}} = \int_{\alpha}^{\beta} G_{r} \left( \frac{F(x)}{F(x)} \right) |y_{1}, \ldots, y_{r}| f(x) \, dx. \tag{5}
\]

Lemma 2.2. Let \( X(s-r, n-r, \bar{\mu}, k) \) and \( X(s-r-1, n-r, \bar{\mu}, k) \) be the two adjacent gos and \( \bar{\mu} = (m_{r+1}, \ldots, m_{n-1}) \in \mathbb{R}^{n-r-1}, \) then
\[
\bar{F}_{s-r, n-r, \bar{\mu}, k}(u) - \bar{F}_{s-r-1, n-r, \bar{\mu}, k}(u) = \frac{c_{r-2}}{c_{r-1}} G_{r-1} \left( \frac{\bar{F}(u)}{\bar{F}(u)} \right) |y_{r+1}, \ldots, y_{r}| \tag{6}
\]

Theorem 2.1. Let \( X_{1}, X_{2}, \ldots, X_{n} \) be non-negative iid random variables with absolutely continuous df (with respect to Lebesgue measure) \( F(x) \) and pdf \( f(x) \), where \( F(x) \) is strictly increasing over the support \((0, \infty)\), then for \( 1 \leq r < s < n \),
\[
X(l, n, \bar{\mu}, k) - X(r, n, \bar{\mu}, k) \overset{d}{=} X(l - r, n, \bar{\mu}, k), \quad l = s - 1, s
\]  
if and only if
\[
f(x) = 1 - e^{-x^\theta}, \quad \theta > 0. \tag{8}
\]

where \( c_{r-1} = \Pi_{j=1}^{r} y_{j} \), \( F(x) = 1 - F(x) \) and \( I_{A} \) denotes the indicator function of the set \( A \). Further \( G_{r} \left( \frac{F(x)}{F(x)} \right) = G_{r-1}^{0} \left( x \right| y_{1}, \ldots, y_{r} \right) \) is the particular Meijer's G-function defined by
\[
G_{r}^{0} \left( x \right| y_{1}, \ldots, y_{r} \right) = \frac{1}{2\pi i} \int_{L} \prod_{j=1}^{r} \frac{x^{-z}}{(y_{j} - z - 1)} \, dz, \quad |x| < 1 \tag{3}
\]
and \( L \) is an appropriately chosen contour of integration. For the definition of G-function and its numerous properties and applications, one may refer to [8] (Chapter 2). The joint \( P_{r} \otimes P_{s} \) density of \( X(r, n, \bar{\mu}, k) \) and \( X(s, n, \bar{\mu}, k) \), \( 1 \leq r < s \leq n \) is given by [7]
\[
f_{r,s} \left( x, y \right) = c_{r} G_{r} \left[ \frac{\bar{F}(y)}{\bar{F}(x)} \right] |y_{r+1}, \ldots, y_{s}| G_{s} \left[ \frac{\bar{F}(x)}{\bar{F}(x)} \right] |y_{1}, \ldots, y_{r}| I_{A} \left( x, y \right), \tag{4}
\]
where the set \( A_{1} \) is defined as \( A_{1} = \{(x, y) \in \mathbb{R}^{2} : \alpha < x < y < \beta \} \).

Suppose \( X_{1}, X_{2}, \ldots, X_{n} \) are iid random variables from exponential distribution with scale parameter \( \theta \), i.e., \( F(x) = 1 - e^{-x} \), \( x > 0 \), \( \theta > 0 \) and \( X_{1,1} \leq X_{2,1} \leq \cdots \leq X_{n,1} \) are the corresponding order statistics. The normalized spacing of order statistics (when the random variables are iid and from exponential distribution) was introduced by [9] as \( D_{1,1} = nX_{1,1} \) and \( D_{r,1} = (n - r + 1)(X_{r,1} - X_{r-1,1}), \) \( 2 \leq r \leq n \). [9] proved that \( D_{r,1} \) \( (1 \leq r \leq n) \) are iid from exponential distribution with scale parameter \( \theta \). Utilizing the above transformation, it can be shown that \( X_{r,1} - X_{r-1,1} \) and \( X_{s,1} - X_{r,1} - r \) are identically distributed across all \( 1 \leq r < s \leq n \). Based on order statistics, the characterization of exponential distribution through normalized spacing was discussed by [10–13] and [14]. The related results of characterization of exponential distribution for record values through spacing was established by [11,13,15,16] and [14,17] has shown that using linear function of spacings instead of linear function of order statistics is more suitable for theoretical and applied purposes. Normalized spacing for gos was given by [6] (p. 81). Suppose \( X_{1}, X_{2}, \ldots, X_{n} \) be the iid random variables from exponential distribution and \( X(1, n, \bar{\mu}, k), X(2, n, \bar{\mu}, k) \cdots X(n, n, \bar{\mu}, k) \) be the corresponding gos. [6] has shown that \( D(1, n, \bar{\mu}, k) = \gamma_{1}X(1, n, \bar{\mu}, k) \) and \( D(r, n, \bar{\mu}, k) = \gamma_{r}X(r, n, \bar{\mu}, k) - X(r - 1, n, \bar{\mu}, k) \), \( 2 \leq r \leq n \) are iid from exponential distribution. Further, [18] and [7] characterized the exponential distribution using normalized spacing of gos based on increasing failure rate and decreasing failure rate distributions. [19] characterized Weibull and Uniform distributions using record values.

In this paper, we have extended the result of [16] and [14] and characterized the exponential distribution using normalized spacing of gos. While the necessary part of the main theorem was given by [20], in this paper, we have given a simple and easy proof for the sufficient part of the theorem. The paper is divided into four sections. In Section 2, based on gos, we have characterized the exponential distribution. Section 3 contains a simulation study where we have shown that confidence interval based on upper records is shorter in length than asymptotic confidence interval constructed by maximum likelihood estimate. Concluding remark focusing on the use of the result given in Section 2, is contained in Section 4.

2. Characterization of exponential distribution

In this section, exponential distribution has been characterized through the normalized spacing based on gos.

The following two Lemmas are given, which have been used in the proof of the theorem. Proof of the Lemmas are straightforward and hence omitted.

Lemma 2.1. Let \( X(r, n, \bar{\mu}, k) \) be the \( r \)-th gos with pdf defined in (2), then
\[
\frac{1}{c_{r-1}} = \int_{\alpha}^{\beta} G_{r} \left( \frac{F(x)}{F(x)} \right) |y_{1}, \ldots, y_{r}| f(x) \, dx. \tag{5}
\]

Lemma 2.2. Let \( X(s-r, n-r, \bar{\mu}, k) \) and \( X(s-r-1, n-r, \bar{\mu}, k) \) be the two adjacent gos and \( \bar{\mu} = (m_{r+1}, \ldots, m_{n-1}) \in \mathbb{R}^{n-r-1}, \) then
\[
\bar{F}_{s-r, n-r, \bar{\mu}, k}(u) - \bar{F}_{s-r-1, n-r, \bar{\mu}, k}(u) = \frac{c_{r-2}}{c_{r-1}} G_{r-1} \left( \frac{\bar{F}(u)}{\bar{F}(u)} \right) |y_{r+1}, \ldots, y_{r}| \tag{6}
\]

Theorem 2.1. Let \( X_{1}, X_{2}, \ldots, X_{n} \) be non-negative iid random variables with absolutely continuous df (with respect to Lebesgue measure) \( F(x) \) and pdf \( f(x) \), where \( F(x) \) is strictly increasing over the support \((0, \infty)\), then for \( 1 \leq r < s < n \),
\[
X(l, n, \bar{\mu}, k) - X(r, n, \bar{\mu}, k) \overset{d}{=} X(l - r, n, \bar{\mu}, k), \quad l = s - 1, s
\]  
if and only if
\[
f(x) = 1 - e^{-x^{\theta}}, \quad \theta > 0. \tag{8}
\]
Proof. The necessary part was established by [20] as follows:

\[ X(l, n, \tilde{m}, k) - X(r, n, \tilde{m}, k) = \sum_{j=r+1}^{l} \left( \frac{z_j}{n} \right), \quad l = s - 1, \text{ where } z_j \sim \exp(\Theta). \]

This implies that \( X(l, n, \tilde{m}, k) - X(r, n, \tilde{m}, k) \overset{d}{=} X(l - r, n - r, \tilde{m}, k), l = s - 1, \text{ where the symbol } \overset{d}{=} \text{ represents equal in distribution}. \)

To prove the sufficient part, we have for any positive and finite \( u \)

\[
P[X(s, n, \tilde{m}, k) - X(r, n, \tilde{m}, k) \geq u] = c_{s-1} \int_{0}^{\infty} \int_{x+u}^{\infty} G_{s-r} \left[ \frac{\overline{F}(y)}{F(x)} \right] |\gamma_{s+1}, ..., \gamma_{s}| \frac{f(y)}{F(x)} G_{s} \left[ \frac{\overline{F}(x)|\gamma_{1}, ..., \gamma_{s}|}{F(x)} \right] f(x) dx dy.
\]

(9)

Let

\[ A(x, y) = G_{s-r} \left[ \frac{\overline{F}(y)}{F(x)} \right] |\gamma_{s+1}, ..., \gamma_{s}| \frac{f(y)}{F(x)}. \]

Integrating \( A(x, y) \) w.r.t. \( y \) over \( (x + u, \infty) \), we have

\[
\int_{x+u}^{\infty} A(x, y) dy = \int_{x+u}^{\infty} G_{s-r} \left[ \frac{\overline{F}(y)}{F(x)} \right] |\gamma_{s+1}, ..., \gamma_{s}| \frac{f(y)}{F(x)} dy
\]

\[
= -G_{s-r} \left[ \frac{\overline{F}(y)}{F(x)} \right] |\gamma_{s+1}, ..., \gamma_{s}| \frac{f(y)}{F(x)} \bigg|_{x+u}^{\infty} + \int_{x+u}^{\infty} \frac{\partial}{\partial y} G_{s-r} \left[ \frac{\overline{F}(y)}{F(x)} \right] |\gamma_{s+1}, ..., \gamma_{s}| \frac{\overline{F}(y)}{F(x)} dy.
\]

(10)

Now, we know from [8] (p. 130), that

\[ x^a G_r(x|\gamma_1, ..., \gamma_r) = G_r(x|\gamma_1 + a, ..., \gamma_r + a), \quad a \in \mathbb{R}. \]

(11)

Using (11) in (10), we have

\[
A(x, y) = -G_{s-r} \left[ \frac{\overline{F}(y)}{F(x)} \right] |\gamma_{s+1}, ..., \gamma_{s}| \bigg|_{x+u}^{\infty} + \int_{x+u}^{\infty} \frac{\partial}{\partial y} G_{s-r} \left[ \frac{\overline{F}(y)}{F(x)} \right] |\gamma_{s+1}, ..., \gamma_{s}| \frac{\overline{F}(y)}{F(x)} dy.
\]

(12)

where \( \gamma_{s+i}^* = \gamma_{s+i} + 1, i = 1, 2, ..., s - r. \)

Let \( \gamma_{1,r} \leq ... \leq \gamma_{s,r} \) denote the ordered parameters of \( \gamma_1, ..., \gamma_r, r \geq 2 \), then

\[
\lim_{s \to 1} G_r(x|\gamma_1, ..., \gamma_s) = \lim_{s \to 0^+} G_r(x|\gamma_1, ..., \gamma_s) = 0, \text{ if } \gamma_{1,r} > 1.
\]

(13)

[21] and

\[
\frac{d}{dx} G_r(x|\gamma_1, ..., \gamma_s) = \frac{1}{x} \left( \gamma_r - 1 \right) G_s(x|\gamma_1, ..., \gamma_s) - G_{s-1}(x|\gamma_1, ..., \gamma_{s-1}).
\]

(14)

[8] (p. 94)

Using (13) and (14) in (12), we get

\[
A(x, y) = G_{s-r} \left[ \frac{\overline{F}(x + u)}{F(x)} \right] |\gamma_{s+1}, ..., \gamma_{s}| + \int_{x+u}^{\infty} G_{s-r-1} \left[ \frac{\overline{F}(y)}{F(x)} \right] |\gamma_{s+1}, ..., \gamma_{s}| \frac{f(y)}{F(x)} dy
\]

\[
- \left( \gamma_r - 1 \right) \int_{x+u}^{\infty} G_{s-r} \left[ \frac{\overline{F}(y)}{F(x)} \right] |\gamma_{s+1}, ..., \gamma_{s}| \frac{f(y)}{F(x)} dy.
\]

This implies that

\[
P[X(s, n, \tilde{m}, k) - X(r, n, \tilde{m}, k) \geq u]
\]

\[
= c_{s-1} \int_{0}^{\infty} \int_{x+u}^{\infty} G_{s-r} \left[ \frac{\overline{F}(y)}{F(x)} \right] |\gamma_{s+1}, ..., \gamma_{s}| G_{s} \left[ \frac{\overline{F}(x)|\gamma_{1}, ..., \gamma_{s}|}{F(x)} \right] f(x) dx
\]

\[
+ c_{s-1} \int_{0}^{\infty} \int_{x+u}^{\infty} G_{s-r} \left[ \frac{\overline{F}(y)}{F(x)} \right] |\gamma_{s+1}, ..., \gamma_{s-1}| \frac{f(y)}{F(x)} G_{s} \left[ \frac{\overline{F}(x)|\gamma_{1}, ..., \gamma_{s-1}|}{F(x)} \right] f(x) dy dx
\]

\[
- \left( \gamma_r - 1 \right) c_{s-1} \int_{0}^{\infty} \int_{x+u}^{\infty} G_{s-r} \left[ \frac{\overline{F}(y)}{F(x)} \right] |\gamma_{s+1}, ..., \gamma_{s}| \frac{f(y)}{F(x)} G_{s} \left[ \frac{\overline{F}(x)|\gamma_{1}, ..., \gamma_{s}|}{F(x)} \right] f(x) dy dx
\]
Further, the value of mode only depends upon the parameter \(K\) for

\[s = \frac{\gamma}{2}, \ldots, \frac{\gamma}{1} + 1,\]  

the function \(G\) is unimodal, strictly increasing up to mode and then strictly decreasing.

Since both \(G, \frac{\bar{F}(x)}{F(x)}|\gamma|\) and \(f(x)\) are positive, using the generalization of Muntz-Swartz theorem (see \([22]\)), the above integral will be zero only if

\[G_{s-r} \left[ \frac{\bar{F}(x+u)}{F(x)} |\gamma+1| \right.\]  

and

\[G_{s-r} \left[ \bar{F}(u) |\gamma+1\right) = 0.\]  

For \(s > r + 1\), the function \(\gamma+1|\) is unimodal, strictly increasing up to mode and then strictly decreasing. Further, the value of mode only depends upon the parameter \(\gamma+1\), hence for the functions

\[G_{s-r} \left[ \frac{\bar{F}(x+u)}{F(x)} |\gamma+1\right)\]  

and

\[G_{s-r} \left[ \bar{F}(u) |\gamma+1\right)\]

mode occurs at the same point (see \([21];\) Theorem 2.1). The value of \(G_{s-r} \left[ \frac{\bar{F}(x+u)}{F(x)} |\gamma+1\right)\) and \(G_{s-r} \left[ \bar{F}(u) |\gamma+1\right)\) are equal only at the mode. At mode, the random variable \(X\) satisfies \(\frac{\bar{F}(x+u)}{F(x)} = \bar{F}(u)\), which imply that \(X\) satisfies the memoryless property. Only continuous distribution over the support \((0, \infty)\), which satisfies the memoryless property is exponential distribution and hence the theorem.

3. SIMULATION STUDY

Here we have followed the procedure adopted by \([19]\) and carried out the simulation to construct the confidence interval for scale parameter \(\theta\) of exponential distribution.

Let \(X_1, \ldots, X_n\) is a random sample from pdf \(f(x) = \frac{1}{\theta} e^{-\frac{x}{\theta}}, x > 0, \theta > 0\). Based on maximum likelihood estimator of exponential distribution, 100 \((1 - \alpha)\%\), \(0 < \alpha < 1\), asymptotic confidence interval \((CI_{\text{Asymptotic}})\) for \(\theta\) is given by

\[
CI_{\text{Asymptotic}} = \left( \frac{\bar{X}}{1 + \frac{2\alpha}{\sqrt{n}}}, \frac{\bar{X}}{1 - \frac{2\alpha}{\sqrt{n}}} \right)
\]
where \( \bar{X} \) is sample mean of random variables and \( Z_{\alpha/2} \) is a positive constant satisfying the relation \( \Phi \left( Z_{\alpha/2} \right) - \Phi \left( -Z_{\alpha/2} \right) = 1 - \alpha \) and \( \Phi(.) \) is the cumulative distribution function of the standard normal distribution. Using Theorem 2.1, for record statistics, we have \( X_{U(n)} - X_{U(n-m)} \) as \( \Delta X_{U(n-m)} \). As \( Y = X_{U(n-m)} \sim G(n-m, \vartheta) \), based on \( Y \), a 100 \( (1 - \alpha) \) % confidence interval for \( \vartheta \) is given by

\[
CI = \left( \frac{2Y}{X_{2n,1-\alpha/2}}, \frac{2Y}{X_{2n,\alpha/2}} \right).
\] (21)

To construct these confidence intervals we simulate 10,000 samples of size \( n \) from (8) for \( n = 4, 5, ..., 100 \), \( m = 1 \) and \( \vartheta = 0.2, 0.5, 1, 2 \). For each sample, we calculate the limits of confidence interval given in (20) and (21). The plot of coverage lengths versus \( n = 4, 5, ..., 100 \), is shown in Figs. 1–4. From the figures, it can be seen that the confidence interval obtained from (21) is smaller than the confidence interval obtained from (20) for \( n \leq 60 \). Thus, we get an improved estimator based on \( X_{U(n-m)} \).

![Plot of confidence interval for \( \vartheta = 0.2 \).](image1.png)

![Plot of confidence interval for \( \vartheta = 0.5 \).](image2.png)

![Plot of confidence interval for \( \vartheta = 1 \).](image3.png)
4. CONCLUDING REMARK

In real life, a statistician is often interested in guessing the probability distribution from which the true data is obtained. Characterization problem is a theoretical approach to obtain the distribution function if certain conditions are fulfilled. A probability distribution can be characterized in many ways and the method under study here is one of them. Further this result may be used to ease statistical computation. For example, if we are computing moments of difference between two gos (order statistics, records, sequential order statistics, progressive type II censored order statistics), when the random variables $X_1, X_2, \ldots, X_n$ follow exponential distribution, then we may utilize the result of Theorem 2.1. This result may also be utilized for calculating the sample quasi range, when the random variables $X_1, \ldots, X_n$ are iid having pdf $f(x) = \frac{1}{\theta} e^{-\frac{x}{\theta}}, \ x > 0, \ \theta > 0$. Various statistical properties of quasi range can be established using Theorem 2.1. Sample quasi range is used in estimating the population standard deviation if the sample size $n$ is small (see [23]). [24] used the sample quasi range in setting confidence intervals for the population standard deviation. Further, [25] and [26] have used the sample quasi range in estimating the scale parameter for a location scale family of distribution, if the parent distribution is symmetric about the location parameter. Sample quasi range and range are also extensively used in quality control.

REFERENCES