Communications in Statistics—Theory and Methods, 36: 1319–1326, 2007

ISSN: 0361-0926 print/1532-415X online
DOI: 10.1080/03610920601076990

On Characterization of Distributions Through the Properties of Conditional Expectations of Order Statistics

ISMIHAN BAIRAMOV[1] AND TUĞBA ÖZKAL[2]

[1]Department of Mathematics, Izmir University of Economics, Balcova, Izmir, Turkey
[2]Department of Statistics, Dokuz Eylul University, Buca, Izmir, Turkey

For a general class of distributions some characterizations through the properties of conditional expectations of order statistics and progressively Type-II censored order statistics are given. Let $X_{1:n}, X_{2:n}, \ldots, X_{n:n}$ be the order statistics of the sample $X_1, X_2, \ldots, X_n$ from a continuous distribution and $X^{\mathbf{R}}_{1:m:n}, X^{\mathbf{R}}_{2:m:n}, \ldots, X^{\mathbf{R}}_{m:m:n}$ be the progressively Type-II censored order statistics with the progressively Type-II censoring scheme $\mathbf{R} = (R, R, \ldots, R)$. We show that in this case, the joint distribution of $X^{\mathbf{R}}_{1:m:n}, \ldots, X^{\mathbf{R}}_{m:m:n}$ can be reduced to the joint distribution of usual order statistics of a sample size m from a continuous random variable.

Keywords Characterizations of distributions; Order statistics; Progressively Type-II censored order statistics.

Mathematics Subject Classification Primary 62H05; Secondary 62G30.

1. Introduction

Let $X_1, X_2, \ldots, X_n$ be independent random variables with a common absolutely continuous distribution function (df) $F(x)$ and a probability density function (pdf) $f(x)$. Let $X_{1:n}, X_{2:n}, \ldots, X_{n:n}$ be the corresponding order statistics. Characterizations of F based on the properties of conditional expectations with the constant sample size n have been discussed by many authors.

Ruiz and Navarro (1996) (see also Navarro et al., 1998) gave a representation of any continuous df $F(x)$ based on the doubly truncated mean function $m(x, y)$

Received August 1, 2005; Accepted July 5, 2006
Address correspondence to Ismihan Bairamov, Department of Mathematics, Izmir University of Economics, 35330 Balcova, Izmir, Turkey; E-mail: ismihan.bayramoglu@ieu.edu.tr

defined by

$$m(x, y) = E(X \mid x \le X \le y) = \frac{1}{F(y) - F(x)} \int_x^y t\, dF(t),$$

whose domain of definition is the set $D = \{(x, y) \in \mathbb{R}^2$ such that $F(x) < F(y)\}$.

Ruiz and Navarro (1996) obtained the explicit expression of $F(x)$ starting from function $m(x, y)$ and define the relationship between $m(x, y)$ and order statistics as

$$E\left(\frac{1}{s-r-1} \sum_{i=r+1}^{s-1} h(X_{i:n}) \mid X_{r:n} = x, X_{s:n} = y\right) = E(h(X) \mid x \le X \le y),$$

for all $(x, y) \in D$, if $1 \le r < s \le n$, where $X_{1:n} \le X_{2:n} \le \cdots \le X_{n:n}$ are the order statistics from the sample of the random variable X. Balasubramanian and Beg (1992) presented similar results for a particular function $h(x)$. Gupta et al. (1993) consider the case when $r = 1$ and $s = n$.

In this work we characterize continuous distributions through the properties of conditional expectations of order statistics and progressively Type-II right-censored order statistics. In Sec. 2, characterization theorems for a general class of distributions are presented in terms of the function

$$E\{g(X_{j:n}) \mid X_{j-p:n} = x, X_{j+q:n} = y\} = A(x, y),$$

where p and q positive integers such that $p + 1 \le j \le n - q$ and $g(\cdot)$, $A(\cdot, \cdot)$ are real valued functions satisfying certain regularity conditions. The results obtained in the paper are closely related to works by Kotlarski (1972) and Shanbhag and Bhaskara Rao (1975). We show that with the progressively Type-II censoring scheme $R = (R, R, \ldots, R)$, the joint distribution of $X^{\mathbf{R}}_{1:m:n}, \ldots, X^{\mathbf{R}}_{m:m:n}$ can be reduced to the joint distribution of usual order statistics of $X_{1:m}, \ldots, X_{m:m}$ from a continuous random variable in Sec. 3.

2. Characterizations Through the Conditional Expectations of Order Statistics

Throughout the article we denote by $a_F = \inf\{x : F(x) > 0\}$ and $b_F = \sup\{x : F(x) < 1\}$, respectively, the left and right extremities of F, where F is a df of a random variable X. We assume that F is absolutely continuous and strictly increasing on (a_F, b_F).

Our characterization results are based on the following theorem.

Theorem 2.1. *Let $h(x)$ be a differentiable real valued function on $[0, 1]$ such that the condition*

$$h'(y) \ne \frac{h(y) - h(x)}{y - x}, \tag{1}$$

holds for all $0 < x < y < 1$. Furthermore, let G be an absolutely continuous and strictly increasing df with left and right extremities, $a_G = a_F$ and $b_G = b_F$, respectively.

Then $F = G$ if, and only if, the representation

$$E\{h'(G(X)) \mid x \le G(X) \le y\} = \frac{h(y) - h(x)}{y - x}, \tag{2}$$

is valid for all $0 < x < y < 1$.

Proof. It is clear that (2) is equivalent to

$$E\{h'(G(X)) \mid x \le X \le y\} = \frac{h(G(y)) - h(G(x))}{G(y) - G(x)},$$

for all $a_F < x < y < b_F$.

Necessity. If X have df $G(x)$, $G'(x) = g(x)$, then

$$\begin{aligned} E\{h'(G(X)) \mid x \le X \le y\} &= \frac{1}{G(y) - G(x)} \int_x^y h'(G(z))g(z)dz \\ &= \frac{h(G(y)) - h(G(x))}{G(y) - G(x)}. \end{aligned}$$

Sufficiency. Let (2) hold. Denote by $F(x)$ and $f(x)$ the df and pdf of X, respectively. Then from (2)

$$\int_x^y h'(G(z))f(z)dz = \frac{h(G(y)) - h(G(x))}{G(y) - G(x)}[F(y) - F(x)]. \tag{3}$$

Differentiating (3) with respect to y after some algebra we have

$$\{f(y)[G(y) - G(x)] - g(y)[F(y) - F(x)]\} \times \left\{h'(G(y)) - \frac{h(G(y)) - h(G(x))}{G(y) - G(x)}\right\} = 0.$$

Therefore,

$$\frac{f(y)}{F(y) - F(x)} = \frac{g(y)}{G(y) - G(x)}, \quad \forall a_F < x < y < b_F.$$

Taking limit as $x \to a_F$ we obtain

$$\frac{f(y)}{F(y)} = \frac{g(y)}{G(y)}, \quad \forall a_F < y < b_F.$$

Equality of hazard rate functions implies equality of F and G. □

Theorem 2.1 states that the representation (2) characterizes the df F in the class $\mathcal{F}$ all absolutely continuous and strictly increasing distribution functions with left and right extremities $l_F = a_F$ and $r_F = b_F$, respectively.

Remark 2.1. It is clear that $\frac{h(y)-h(x)}{y-x}$ is the slope of the straight line joining points $(x, h(x))$, $(y, h(y))$, and $h'(y)$ is the slope of the tangent line to the graph of $h(\cdot)$ at the point y. The condition (1) assumes these two slopes to be different. One

can observe that for any convex or concave function $h(\cdot)$ the property (1) is satisfied.

Let $X_1, X_2, \ldots, X_n$ be the independent copies of X, i.e., be i.i.d. random variables with df F and pdf f. Denote by $X_{1:n}, \ldots, X_{n:n}$ the corresponding order statistics. The following fact shows that the conditional distribution of $X_{j:n}$ given $X_{j-p:n} = x$ and $X_{j+q:n} = y$ does not depend on the sample size n. Note that it depends on x, y and F.

Theorem 2.2. *If $p + 1 \leq j \leq n - q$, then*

$$(X_{j:n} \mid X_{j-p:n} = x, X_{j+q:n} = y) \stackrel{d}{=} Z_{p:p+q-1}, \quad p + 1 \leq j \leq n - q,$$

where

$$Z \stackrel{d}{=} (X \mid x < X < y).$$

We use Theorems 2.1 and 2.2 to characterize several distributions by properties of conditional expectations of order statistic $X_{j:n}$ given $X_{j-p:n} = x$ and $X_{j+q:n} = y$.

Let us remind that we assume the df F of the random variable X is absolutely continuous and strictly increasing on (a_F, b_F), where $a_F = \inf\{x : F(x) > 0\}$ and $b_F = \sup\{x : F(x) > 1\}$.

Theorem 2.3. *Let G be an absolutely continuous and strictly increasing df and the left and right extremities of G be $a_G = a_F$ and $b_G = b_F$, respectively. Then, $F = G$ if, and only if, the representation*

$$\frac{1}{k}\sum_{p=1}^{k} E[h'(G(X_{j:n})) \mid X_{j-p:n} = x, X_{j+k+1-p:n} = y] = \frac{h(G(y)) - h(G(x))}{G(y) - G(x)} \tag{4}$$

holds for all $a_F < x < y < b_F$, where $h(x)$ satisfies conditions of Theorem 2.1 and j, n, k are fixed numbers such that

$$k + 1 \leq j \leq n - k. \tag{5}$$

Proof. Taking into account the results of Theorem 2.2 we have

$$\begin{aligned}
&\frac{1}{k}\sum_{p=1}^{k} E[h'(G(X_{j:n})) \mid X_{j-p:n} = x, X_{j+k+1-p:n} = y] \\
&\quad = E\left[\frac{1}{k}\sum_{p=1}^{k} h'(G(Z_{p:k}))\right] = E\left[\frac{1}{k}\sum_{i=1}^{k} h'(G(Z_i))\right] \\
&\quad = \frac{1}{k} k E[h'(G(Z))] = E[h'(G(Z))] \\
&\quad = E[h'(G(X)) \mid x \leq X \leq y].
\end{aligned}$$

Then the proof is completed by using Theorem 2.1. □

2.1. *Characterizations for Some Special Distributions*

The following characterizations for special distributions can be obtained from Theorem 2.3 by the appropriate choice of the function h.

2.1.1. *Uniform Distribution.* The absolutely continuous random variable X with strictly increasing df having support $[0, 1]$ has Uniform distribution over $(0, 1)$ if, and only if, the representation

$$\frac{1}{k}\sum_{p=1}^{k} E\left[X_{j:n}^{c} \mid X_{j-p:n} = x, X_{j+k+1-p:n} = y\right] = \frac{y^{c+1} - x^{c+1}}{(c+1)(y-x)}, \tag{6}$$

holds for all $0 < x < y < 1$. In (6) $j \in \{k+1, k+2, \ldots, n-k\}$ is fixed. This result follows from Theorem 2.3 by using $h(x) = \frac{x^{c+1}}{c+1}$, $c \geq 1$.

2.1.2. *Weibull Distribution.* Let $h(x) = (1-x)(\ln(1-x) - 1)$. Then $h'(x) = -\ln(1-x)$.

The absolutely continuous random variable X with strictly increasing df having support $[0, \infty)$, has Weibull distribution $F(x) = 1 - \exp(-\alpha x^{\beta})$, $x \geq 0$, $\alpha > 0$, $\beta > 0$ if, and only if, the representation

$$\frac{1}{k}\sum_{p=1}^{k} E\left[\alpha X_{j:n}^{\beta} \mid X_{j-p:n} = x, X_{j+k+1-p:n} = y\right]$$
$$= \frac{\exp(-\alpha x^{\beta})(\alpha x^{\beta} + 1) - \exp(-\alpha y^{\beta})(\alpha y^{\beta} + 1)}{\exp(-\alpha x^{\beta}) - \exp(-\alpha y^{\beta})},$$

holds for all $0 \leq x < y < \infty$. Here j satisfies (5).

The above characterization with $\beta = 1$ turns into a characterization of exponential distribution.

2.1.3. *Generalized Beta Distribution.* The absolutely continuous random variable X with strictly increasing df having support $[a, b]$ has Generalized Beta distribution $F(x) = 1 - \frac{(b-x)^{\theta}}{(b-a)^{\theta}}$, $a \leq x \leq b$, $\theta > 0$, $-\infty < a < b < \infty$ if, and only if, the representation

$$\frac{1}{k}\sum_{p=1}^{k} E\left[X_{j:n} \mid X_{j-p:n} = x, X_{j+k+1-p:n} = y\right] = b + \frac{(a-b)\theta}{\theta+1}\frac{a_1(x, y; \theta+1)}{a_1(x, y; \theta)},$$

holds for all $a < x < y < b$ where $a_1(x, y; \theta) = \left(\frac{b-x}{b-a}\right)^{\theta} - \left(\frac{b-y}{b-a}\right)^{\theta}$. Here j satisfies (5).

The result is obtained by using target function $h(x) = bx - \frac{(a-b)\theta}{\theta+1}(1-x)^{1+1/\theta}$, $h'(x) = b - (b-a)(1-x)^{1/\theta}$.

2.1.4. *Pareto Distribution.* The absolutely continuous random variable X with strictly increasing df having support $[\gamma, \infty)$ has Pareto distribution $F(x) = 1 - \frac{(\gamma+\delta)^{\theta}}{(x+\delta)^{\theta}}$, $x \geq \gamma$, $\theta > 0$, $\gamma + \delta > 0$, if, and only if, the representation

$$\frac{1}{k}\sum_{p=1}^{k} E\left[X_{j:n} \mid X_{j-p:n} = x, X_{j+k+1-p:n} = y\right] = -\frac{\theta}{\theta-1}\frac{b(x, y; \theta-1)}{b(x, y; \theta)} - \delta,$$

holds for all $\gamma \le x < y < \infty$ where $b(x, y; \theta) = \left(\frac{1}{x+\delta}\right)^{\theta} - \left(\frac{1}{y+\delta}\right)^{\theta}$, $\theta \neq 1$ and j satisfies (5). The result is obtained by using target function

$$h(x) = \frac{(\gamma+\delta)\theta}{\theta-1}(1-x)^{1-(1/\theta)} - \delta x.$$

3. Characterizations Through Conditional Expectations of Progressively Type-II Censored Order Statistics

Let $X_1, X_2, \ldots, X_n$ be continuous i.i.d. random variables with df F and pdf f. Denote by $X^{\mathbf{R}}_{1:m:n}, \ldots, X^{\mathbf{R}}_{m:m:n}$ the corresponding progressively Type-II Censored order statistics with the censoring scheme $\mathbf{R} = (R_1, R_2, \ldots, R_m)$. We consider the case when $R_1 = R_2 = \cdots = R_m = R$ and give some characterizations for a general class of distribution through the conditional expectations of progressively Type-II censored order statistics. First, we show that in the case $R_1 = R_2 = \cdots = R_m = R$, the joint distribution of $X^{\mathbf{R}}_{1:m:n}, \ldots, X^{\mathbf{R}}_{m:m:n}$ can be reduced to the joint distribution of usual order statistics of a sample size m from a continuous random variable.

Theorem 3.1. *Let $X_1, X_2, \ldots, X_n$ be a sample of size n from a population with continuous df F and pdf f and let $X^{\mathbf{R}}_{1:m:n} \le X^{\mathbf{R}}_{2:m:n} \le \cdots \le X^{\mathbf{R}}_{m:m:n}$ be the progressively Type-II censored order statistics under the censoring scheme $\mathbf{R} = (R_1, R_2, \ldots, R_m)$. Then in the special case when $R_1 = R_2 = \cdots = R_m = R$, we have*

$$\left(X^{\mathbf{R}}_{1:m:n}, \ldots, X^{\mathbf{R}}_{m:m:n}\right) \stackrel{d}{=} \left(Y_{1:m}, \ldots, Y_{m:m}\right),$$

where $Y_1, Y_2, \ldots, Y_m$ is the sample with df $Q(x) = 1 - (1 - F(x))^{n/m}$ and $R + 1 = n/m$.

Proof. The joint pdf of $(X^{\mathbf{R}}_{1:m:n}, X^{\mathbf{R}}_{2:m:n}, \ldots, X^{\mathbf{R}}_{m:m:n})$, $\mathbf{R} = (R_1, R_2, \ldots, R_m)$ is (see Balakrishnan and Aggarwala, 2000)

$$\begin{aligned} f^{\mathbf{R}}_{1,2,\ldots,m}(x_1, \ldots, x_m) &= n(n - R_1 - 1)\cdots(n - R_1 - R_2 - \cdots - R_{m-1} - (m-1)) \\ &\quad \times \prod_{i=1}^{m} f(x_i)(1 - F(x_i))^{R_i}, \end{aligned}$$

where $R_1 + R_2 + \cdots + R_m + m = n$.

If $R_1 = R_2 = \cdots = R_m = R$, then we have $R + 1 = n/m$. Since

$$\begin{aligned} &n(n - (R+1))(n - 2(R+1))\cdots(n - (m-1)(R+1)) \\ &= n\left(n - \frac{n}{m}\right)\left(n - 2\frac{n}{m}\right)\cdots\left(n - (m-1)\frac{n}{m}\right) = (R+1)^m m!, \end{aligned}$$

we have

$$\begin{aligned} f^{\mathbf{R}}_{1,2,\ldots,m}(x_1, \ldots, x_m) &= m!(R+1)^m(1 - F(x_1))^R \cdots (1 - F(x_m))^R f(x_1)\cdots f(x_m) \\ &= m! g(x_1)\cdots g(x_m), \end{aligned}$$

where $g(x) = (R+1)f(x)(1 - F(x))^R$, $G(x) = 1 - (1 - F(x))^{R+1}$. □

By using Theorems 2.3 and 3.1, we can give characterizations for absolutely continuous distributions in terms of conditional expectations of progressively Type-II censored order statistics.

Due to Theorem 3.1, any property of order statistics can be easily transformed to the corresponding property of progressively Type-II censored order statistics with censoring scheme $(R, R, \ldots, R)$. In particular, any characterization of a probability distribution in term of order statistics can be rewritten as a characterization in terms of progressively Type-II censored order statistics with censoring scheme $(R, R, \ldots, R)$. Theorem 3.2 can serve as an example.

Theorem 3.2. *Assume that the random variable X has absolutely continuous and strictly increasing df F with left and right extremities a_F and b_F, respectively. Let G be also an absolutely continuous and strictly increasing df with left and right extremities $a_G = a_F$ and $b_G = b_F$. Then $F = 1 - (1 - G)^{\frac{m}{n}}$ if, and only if, the representation*

$$\frac{1}{k}\sum_{p=1}^{k} E\left[h'(G(X^{\mathbf{R}}_{j:m:n})) \mid X^{\mathbf{R}}_{j-p:m:n} = x, X^{\mathbf{R}}_{j+k+1-p:m:n} = y\right] = \frac{h(G(y)) - h(G(x))}{G(y) - G(x)},$$

holds for all $a_X < x < y < b_X$. $\mathbf{R} = (R, R, \ldots, R)$ and the number j, n, k are fixed and satisfies (5).

Proof of Theorem 3.2 follows easily from Theorems 2.3 and 3.1.

The characterization of generalized beta distribution for order statistics and Theorem 3.1 immediately implies the characterization of this distribution in terms of progressively Type-II censored order statistics. Similarly, one can transform characterizations of Weibull (exponential), generalized beta, and Pareto distributions given in terms of order statistics into characterizations in terms of progressively Type-II censored order statistics. The following characterization of generalized beta distribution is given as an example.

3.1. *Generalized Beta Distribution*

The absolutely continuous random variable X with strictly increasing df having support $[a, b]$ has Generalized Beta distribution $F(x) = 1 - \frac{(b-x)^{\theta}}{(b-a)^{\theta}}$, $a \le x \le b$, $\theta > 0$, $-\infty < a < b < \infty$ if, and only if, the representation

$$\begin{aligned}\frac{1}{k}\sum_{p=1}^{k} E\left[X^{\mathbf{R}}_{j:m:n} \mid X^{\mathbf{R}}_{j-p:m:n} = x, X^{\mathbf{R}}_{j+k+1-p:m:n} = y\right]\\ = b + \frac{\theta(R+1)(a-b)}{\theta(R+1)+1}\frac{d_1(x, y; \theta+1)}{d_1(x, y; \theta)},\end{aligned}$$

holds for all $a \le x < y < b$, where $d_1(x, y; \theta) = \left(\frac{b-x}{b-a}\right)^{\theta(R+1)} - \left(\frac{b-y}{b-a}\right)^{\theta(R+1)}$. The result is obtained by using target function $h(x) = bx - \frac{(a-b)(1-x)^{1+1/(\theta(R+1))}}{1/(\theta(R+1))}$ and $h'(x) = b - (b-a)(1-x)^{1/(\theta(R+1))}$ and $G(x) = 1 - \frac{(b-x)^{\theta(R+1)}}{(b-a)^{\theta(R+1)}}$.

The above characterization with $a = 0$, $b = 1$, $\theta = 1$ turns into a characterization of uniform distribution.

References

Balasubramanian, K., Beg, M. I. (1992). Distributions determined by conditioning on a pair of order statistics. *Metrika* 39:107–112.

Balakrishnan, N., Aggarwala, R. (2000). *Progressive Censoring: Theory, Methods, and Applications.* Boston: Birkhauser.

Gupta, S. D., Goswami, A., Rao, B. V. (1993). On a characterization of uniform distribution. *J. Multivariate Anal.* 44:102–114.

Kotlarski, I. I. (1972). On a characterization of some probability distributions by conditional expectations. *Sankhyã Ser. A.* 34:461–466.

Navarro, J., Ruiz, J. M., Zoroa, N. (1998). A unified approach characterization problems using conditional expectations. *J. Statist. Plann. Infer.* 69:193–207.

Ruiz, J. M., Navarro, J. (1996). Characterizations based on conditional expectations of the doubly truncated distribution. *Ann. Instit. Statist. Math.* 48:563–572.

Shanbhag, D. N., Bhaskara Rao, M. (1975). A note on characterizations of probability distributions based on conditional expected values. *Sankhyã Ser. A.* 37: 297–300.