Metrika (1992) 39:107 – 112

Distributions Determined by Conditioning on a Pair of Order Statistics

By K. Balasubramanian and M. I. Beg [1]

Abstract: Let $X_1, X_2, \ldots, X_n$ $(n \geq 3)$ be a random sample on a random variable X with distribution function F having a unique continuous inverse F^{-1} over (a,b), $-\infty \leq a < b \leq \infty$ the support of F. Let $X_{1:n} < X_{2:n} < \ldots < X_{n:n}$ be the corresponding order statistics. Let g be a nonconstant continuous function over (a,b). Then for some function G over (a,b) and for some positive integers r and s, $1 < r+1 < s \leq n$

$$E\left\{\frac{1}{s-r+1}\sum_{i=r}^{s} g(X_{i:n}) \,|\, X_{r:n}=x,\ X_{s:n}=y\right\} = \frac{G(x)+G(y)}{2}, \quad \forall\ x,y \in (a,b)$$

iff g and G are bounded, increasing and continuous, $G = g$ and $F(x) = \dfrac{g(x)-g(a+)}{g(b-)-g(a+)}$. This leads to characterization of several distributions.

1 Introduction

Let $X_1, X_2, \ldots, X_n$ be a random sample on a random variable X with distribution function F having a unique continuous inverse F^{-1} over (a,b), $-\infty \leq a < b \leq \infty$, the support of F. Let $X_{1:n} < X_{2:n} < \ldots < X_{n:n}$ be the corresponding order statistics. Ferguson (1967) characterized distributions using the fact $E\{X_{i:n} | X_{i+1:n}\} = \alpha x - \beta$. Beg and Kirmani (1974) characterized the same distributions through the condition $E\{X_i | X_{n:n} = x\} = \alpha x - \beta$, where α and β are constants. For related results we refer to Galambos and Kotz (1978) and also Azlarov and Volodin (1986).

Let g be a nonconstant continuous function over (a,b) with finite $g(a+)$ and $E\{g(X)\}$. By suitably choosing g, Beg and Balasubramanian (1990) characterized all distributions for which the explicit form of the distribution function is known, continuous and strictly increasing in its support (a,b) through the property

[1] K. Balasubramanian, Indian Statistical Institute, Delhi Centre, 7 S.J.S. Sansanwal Marg, New Delhi- 110016, India. M. I. Beg, School of Mathematics, University of Hyderabad, Hyderabad-500134, India.

$$E\left\{\frac{1}{s-1}\sum_{i=1}^{s-1} g(X_{i:n})\,|\,X_{s:n}=x\right\}=\frac{g(x)+g(a+)}{2}\,,\quad \forall\ x\in(a,b)\ .$$

Here the conditional expectation is assumed to exist. But in the present paper such an assumption is unnecessary as it exists anyway. Moreover, two functions g and G make the result stronger, in the sense that G should necessarily be equal to g. Thus the present result, all things considered, is considerably stronger.

Definition: Distribution generated by a function g:

Let g be a right continuous, increasing and bounded function on (a,b), $-\infty\le a<b\le\infty$. The distribution generated by g on (a,b) is $\dfrac{g(x)-g(a+)}{g(b-)-g(a+)}$.

If F is a distribution function, then 'F is generated by g' is equivalent to 'F is a linear function of g'.

Theorem 1: For some function G over (a,b) and for some positive integers r and s, $1<r+1<s\le n$

$$E\left\{\frac{1}{(s-r+1)}\sum_{i=r}^{s} g(X_{i:n})\,|\,X_{r:n}=x,\,X_{s:n}=y\right\}=\frac{G(x)+G(y)}{2}\,,\quad \forall\ x,y\in(a,b) \tag{1}$$

if and only if g and G are bounded, increasing and continuous, $G=g$ and F is the distribution generated by g in (a,b).

The proof of Theorem 1 follows immediately after a lemma.

2 A Lemma

Lemma 1: Let ϕ and h be functions defined over (α,β) and let h be continuous. If

$$\phi(x)-\phi(a)=(x-a)\left[\frac{h(a)+h(x)}{2}\right]\,,\quad \forall\ a,x\in(\alpha,\beta) \tag{2}$$

then h is a linear function.

Proof: In view of continuity of h,

$$\lim_{a\to x}\frac{\phi(x)-\phi(a)}{(x-a)}=h(x)\ .$$

Hence ϕ is differentiable and

$$\phi'(x)=h(x)\ ,\quad \forall\ x\in(\alpha,\beta)\ .$$

Thus from (2),

$$\phi(x)-\phi(a)=(x-a)\left[\frac{\phi'(a)+\phi'(x)}{2}\right]$$

or

$$\phi'(x)=2\left[\frac{\phi(x)-\phi(a)}{(x-a)}\right]-\phi'(a)\ ,\quad \forall\ a,x\in(\alpha,\beta)\ . \tag{3}$$

(3) is a linear differential equation in $\phi(x)$ with the general solution of the form $\phi(x)=a_0x^2+b_0x+c_0$, where $b_0=\phi'(a)-2a_0a$, $c_0=\phi(a)-a\phi'(a)+a_0a^2$ and a_0 is an arbitrary constant. Hence $h(x)=\phi'(x)$ is a linear function.

3 Proof of Theorem 1

The joint probability density function of $X_{r:n}$, $X_{i:n}$ and $X_{s:n}$ $(1\le r<i<s\le n)$ is for $x<t<y$

$$\frac{n!}{(r-1)!\,(i-r-1)!\,(s-i-1)!\,(n-s)!}[F(x)]^{r-1}[F(t)-F(x)]^{i-r-1}[F(y)-F(t)]^{s-i-1}$$
$$\cdot[1-F(y)]^{n-s}f(x)f(t)f(y)$$

and that of $X_{r:n}$ and $X_{s:n}$ $(1\le r<s\le n)$ is for $x<y$

$$\frac{n!}{(r-1)!\,(s-r-1)!\,(n-s)!}[F(x)]^{r-1}[F(y)-F(x)]^{s-r-1}[1-F(y)]^{n-s}f(x)f(y)\ .$$

The conditional probability density function of $X_{i:n}$ given $X_{r:n}=x$ and $X_{s:n}=y$ $(1\le r<i<s\le n)$ is for $x<t<y$

$$\frac{(s-r-1)!}{(i-r-1)!(s-i-1)!}\left[\frac{F(t)-F(x)}{F(y)-F(x)}\right]^{i-r-1}\left[\frac{F(y)-F(t)}{F(y)-F(x)}\right]^{s-i-1}\left[\frac{f(t)}{F(y)-F(x)}\right]$$

which is the distribution of the $(i-r)$-th order statistic in a sample of size $s-r-1$ drawn from $f(t)/[F(y)-F(x)]$ $(x<t<y)$, i.e., from the parent population truncated on the left at x on the right at y.

In view of the above relation it is easy to see that

$$\sum_{i=r}^{s}\frac{1}{(s-r+1)}E\{g(X_{i:n})\,|\,X_{r:n}=x,\ X_{s:n}=y\}$$

$$=\frac{1}{(s-r+1)}[(g(x)+g(y))+(s-r-1)E\{g(X)\,|\,x<X<y\}]\ . \tag{4}$$

If F is generated by g, taking $G=g$ in (a,b), it is easy to verify that the right hand side of (4) reduces to $(1/2)\ (G(x)+G(y))$.

To prove the converse, from (1) and (4), we have

$$E\{g(X)\,|\,x<X<y\}=\int_x^y g(t)\frac{dF(t)}{(F(y)-F(x))}$$

$$=\frac{(s-r+1)(G(x)+G(y))}{2(s-r-1)}-\frac{g(x)+g(y)}{(s-r-1)}=\frac{H(x)+H(y)}{2}\ ,$$

where $H(\cdot)=\dfrac{(s-r+1)G(\cdot)}{(s-r-1)}-\dfrac{2g(\cdot)}{(s-r-1)}$.

Putting $F(t)=u$, $F(x)=c$ and $F(y)=z$, we get

$$\int_c^z g(F^{-1}(u))du=\left[\frac{H(F^{-1}(c)+H(F^{-1}(z))}{2}\right](z-c)\ ,\quad \forall\ c,z\in(0,1)\ . \tag{5}$$

Writing $\int g(F^{-1}(u))du=\phi(u)$ and $H(F^{-1}(\cdot))=h(\cdot)$, (5) reduces to

$$\phi(z)-\phi(c)=\left[\frac{h(c)+h(z)}{2}\right](z-c)\ ,\quad \forall\ c,z\in(0,1)\ .$$

By Lemma 1, $\phi'(u) = h(u)$, $\forall\ u \in (0,1)$ and h is a linear function. Therefore, $g(F^{-1}(u)) = h(u) = H(F^{-1}(u))$. This shows $G = g$ and that g is a linear function of F. The theorem follows.

A simple interesting consequence of Theorem 1 is the following.

Corollary 1: Under the assumptions of Theorem 1,

$$E\{g(X_i)\,|\,X_{1:n}=x,\ X_{n:n}=y\} = E\left\{\frac{1}{n}\sum_{i=1}^{n} g(X_i)\,|\,X_{1:n}=x,\ X_{n:n}=y\right\}$$

$$= \frac{G(x)+G(y)}{2}, \quad \forall\ x,y \in (a,b)$$

if and only if F is a distribution function generated by g.

Corollary 1 remains valid if we replace

$$E\left\{\frac{1}{n}\sum_{i=1}^{n} g(X_i)\,|\,X_{1:n}=x,\ X_{n:n}=y\right\} \quad \text{by}$$

$$E\left\{\sum_{i=1}^{n} \alpha_i g(X_i)\,|\,X_{1:n}=x,\ X_{n:n}=y\right\}$$

where $\alpha_1, \alpha_2, \ldots, \alpha_n$ are any real numbers such that $\sum_{i=1}^{n} \alpha_i = 1$.

4 Application of Theorem 1

By suitably choosing g we can characterize all distributions for which the explicit form of the distribution function is known, continuous and strictly increasing in its support (a,b). These are the same as in Beg and Balasubramanian (1990).

Remark 1: For completeness, we state the following equivalent forms of Theorem 1, the latter avoiding the use of order statistics.

*Theorem 1 *:* For some function G over (a,b) and for some positive integers r and s, $1 < r+1 < s \leq n$

$$E\left\{\frac{1}{(s-r-1)}\sum_{i=r+1}^{s-1} g(X_{i:n}) \mid X_{r:n}=x,\ X_{s:n}=y\right\}=\frac{G(x)+G(y)}{2},$$

$\forall\ x,y \in (a,b)$

if and only if g and G are bounded, increasing and continuous, $G=g$ and F is the distribution generated by g in (a,b).

*Theorem 1***: For some function G over (a,b)

$$E[g(X)\mid x<X<y]=\frac{G(x)+G(y)}{2}, \quad \forall\ x,y\in(a,b)$$

if and only if g and G are bounded, increasing and continuous, $G=g$ and F is the distribution generated by g in (a,b).

Remark 2: In Beg and Balasubramanian (1990) conditioning on one order statistic is used for characterization. In the present paper conditioning on two order statistics is used. Conditioning on more than two order statistics is unnecessary in view of Markovian property of order statistics from continuous random variables (see David 1981, p. 20).

Acknowledgement: We thank the referee for some useful comments.

References

Azlarov TA, Volodin NA (1986) Characterization problems associated with the exponential distribution. Springer

Beg MI, Balasubramanian K (1990) Distributions determined by conditioning on a single order statistic. Metrika 37:37–43

Beg MI, Kirmani SNUA (1974) On a characterization of exponential and related distributions. Austral J Statist 16:163–166 (correction in Austral J Statist (1976) 18:85)

David HA (1981) Order statistics. Wiley

Ferguson TS (1967) On characterization of distributions by properties of order statistics. Sankhya, Ser A 29:265–278

Galambos J, Kotz S (1978) Characterizations of probability distributions. Lecture Notes in Mathematics 675. Springer

Received 25 January 1991
Revised version 26 April 1991