Available at WWW.MATHEMATICSWEB.ORG
POWERED BY SCIENCE @ DIRECT®

ELSEVIER

Statistics & Probability Letters 63 (2003) 435–440

STATISTICS & PROBABILITY LETTERS

www.elsevier.com/locate/stapro

# A characterization by linearity of the regression function based on order statistics

N. Balakrishnan*, I.S. Akhundov

*Department of Mathematics & Statistics, McMaster University, 1280 Main Street West, Hamilton, Ont., Canada L8S 4K1*

Received December 2002; received in revised form March 2003

**Abstract**

We consider an infinite sequence $X_1, X_2, \ldots$ of independent random variables having a common continuous distribution function $F(x)$. For $1 \leqslant i \leqslant n$, let $\{X_{i:n}\}$ denote the $i$th order statistic among $X_1, \ldots, X_n$. In this paper, we characterize the distributions for which the regression $E((X_{1:n} + \cdots + X_{n:n})/n | X_{k:n} = x)$ is a linear function of $x$.


*Keywords:* Order statistics; Regression function; Characterization

## 1. Introduction

Let $X_1, X_2, \ldots, X_n$ be independent random variables with a common continuous cumulative distribution function (c.d.f.) $F(x)$ and a probability density function $f(x)$. Let $X_{1:n} \leqslant X_{2:n} \leqslant \cdots \leqslant X_{n:n}$ be the corresponding order statistics. Characterizations of $F$ based on the properties of certain regression functions associated with order statistics have been discussed by many authors. The main problem considered in this area is to characterize the distributions for which the relation

$$E(g(X_{1:n}, \ldots, X_{n:n}) | X_{k:n}) = h(X_{k:n}) \tag{1}$$

holds for some functions $g : \mathbb{R}^n \to \mathbb{R}$ and $h : \mathbb{R} \to \mathbb{R}$ and fixed $k \in \{1, 2, \ldots, n\}$. Unfortunately, there is no solution for this problem in this general case. However, many characterizations were established in some special cases by choosing in different ways the functions $g$ and $h$.

* Corresponding author.
*E-mail addresses:* bala@mcmall.cis.mcmaster.ca (N. Balakrishnan), akhund@mcmaster.ca (I.S. Akhundov).

0167-7152/03/$ - see front matter 
doi:10.1016/S0167-7152(03)00123-8

Ferguson (1967) was the first author who treated this problem with $g(x_1,\dots,x_n)=x_{k-1}$ and $h(x)=ax+b$, for continuous distributions. The corresponding case for discrete distributions was discussed by Nagaraja (1988). Wesolowsky and Ahsanullah (1997) recently obtained characterizations of the exponential, power and Pareto distributions by considering this problem with $g(x_1,\dots,x_n)=x_{k+2}$, for some $k\in\{1,\dots,n-2\}$ and $h(x)=ax+b$. A general case with $h(x)$ that is not necessarily linear was considered by Bairamov and Balakrishnan (2001). We refer to Johnson et al. (1994) and Arnold et al. (1992) for recent surveys on this problem and references to pertinent literature.

In this paper, introducing the quantile function $G(x)$ such that $F(G(x))=x,\ 0<x<1$, we study the case when

$$g(x_1,x_2,\dots,x_n)=\frac{x_1+x_2+\cdots+x_n}{n} \quad \text{and} \quad h(x)=x.$$

We characterize a general family of distributions with quantile function

$$G(x;\lambda,c,d)=\frac{c(x-\lambda)}{\lambda(1-\lambda)(1-x)^{\lambda}x^{1-\lambda}}+d, \quad 0<x<1,\ 0<\lambda<1,\ c>0,\ -\infty<d<\infty$$

for which condition (1) holds. We will also discuss some simple properties of this family of distributions.

## 2. A characterization by linearity of regression

Suppose that $E(X_1)$ exists and

$$\varphi_{k:n}(x)=E\left(\frac{S_n}{n}\,|X_{k:n}=x\right)=E\left(\frac{X_{1:n}+\cdots+X_{n:n}}{n}\,|X_{k:n}=x\right).$$

**Theorem.** *The following statements are equivalent*:

(1) *For some $k$, the regression function $\varphi_{k:n}(x)$ satisfies almost surely the relation*

$$\varphi_{k:n}(x)=x;$$

(2) *The distribution $F$ is given by the quantile function*

$$G(x;\lambda,c,d)=c\,G(x;\lambda,1,0)+d,$$

*where $c(c>0)$ and $d(-\infty<d<\infty)$ are some constants, $0<\lambda<1$, and*

$$G(x;\lambda,1,0)=\frac{x-\lambda}{\lambda(1-\lambda)(1-x)^{\lambda}x^{1-\lambda}}, \quad 0<x<1.$$

**Proof.** To show that (1) implies (2) we use the following Lemma:

**Lemma** ([1, p. 23]). *If $F$ is continuous, then the conditional joint distribution of $X_{1:n},X_{2:n},\dots,X_{k-1:n}$, given $X_{k:n}=x$, is identical to the joint distribution of the order statistics $Y^x_{1:k-1},\dots,Y^x_{k-1:k-1}$ from* *a random sample of size $k-1$ with c.d.f. $H_x(u)$ having the form*

$$H_x(u)=\begin{cases}\dfrac{F(u)}{F(x)}, & u<x,\\ 1, & u\geqslant x;\end{cases}$$

*the conditional joint distribution of $X_{k+1:n},X_{k+2:n},\ldots,X_{n:n}$, given $X_{k:n}=x$, is identical to the joint distribution of the order statistics $Z^x_{1:n-k},\ldots,Z^x_{n-k:n-k}$ from a random sample of size $n-k$ with c.d.f. $R_x(v)$ having the form*

$$R_x(v)=\begin{cases}\dfrac{F(v)-F(x)}{1-F(x)}, & v>x,\\ 0, & v\leqslant x.\end{cases}$$

Using the above Lemma, we can write

$$\begin{aligned}E(X_{1:n}+\cdots+X_{k-1:n}|X_{k:n}=x)&=E(Y^x_{1:k-1}+\cdots+Y^x_{k-1:k-1})\\&=(k-1)EY^x_1=\frac{k-1}{F(x)}\int_{-\infty}^{x}u\,\mathrm{d}F(u)\end{aligned} \tag{2}$$

and

$$\begin{aligned}E(X_{k+1:n}+\cdots+X_{n:n}|X_{k:n}=x)&=E(Z^x_{1:n-k}+\cdots+Z^x_{n-k:n-k})\\&=(n-k)EZ^x_1=\frac{n-k}{1-F(x)}\int_{x}^{\infty}u\,\mathrm{d}F(u).\end{aligned} \tag{3}$$

From (2) and (3), we readily obtain

$$E(X_{1:n}+\cdots+X_{n:n}|X_{k:n}=x)=\frac{k-1}{F(x)}\int_{-\infty}^{x}u\,\mathrm{d}F(u)+x+\frac{n-k}{1-F(x)}\int_{x}^{\infty}u\,\mathrm{d}F(u).$$

Consequently,

$$\begin{aligned}\varphi_{k:n}(x)&=E\left(\frac{X_{1:n}+\cdots+X_{n:n}}{n}\,\middle|\,X_{k:n}=x\right)\\&=\frac{x}{n}+\frac{k-1}{nF(x)}\int_{-\infty}^{x}u\,\mathrm{d}F(u)+\frac{n-k}{n(1-F(x))}\int_{x}^{\infty}u\,\mathrm{d}F(u).\end{aligned} \tag{4}$$

Assuming the existence of $EX_1$, we have

$$\lim_{u\to-\infty}uF(u)=0\quad\text{and}\quad\lim_{u\to\infty}u(1-F(u))=0,$$

and so integration by parts in (4) yields

$$\begin{aligned}\varphi_{k:n}(x) &= \frac{x}{n} + \frac{k-1}{nF(x)}\left\{xF(x) - \int_{-\infty}^{x} F(u)\,\mathrm{d}u\right\} \\ &\quad + \frac{n-k}{n(1-F(x))}\left\{x(1-F(x)) + \int_{x}^{\infty}(1-F(u))\,\mathrm{d}u\right\} \\ &= x - \frac{k-1}{nF(x)}\int_{-\infty}^{x} F(u)\,\mathrm{d}u + \frac{n-k}{n(1-F(x))}\int_{x}^{\infty}(1-F(u))\,\mathrm{d}u,\end{aligned}$$

which means that $\varphi_{k:n}(x) = x$ if and only if

$$(k-1)(1-F(x))\int_{-\infty}^{x} F(u)\,\mathrm{d}u = (n-k)F(x)\int_{x}^{\infty}(1-F(u))\,\mathrm{d}u. \tag{5}$$

Let $\alpha = (k-1)/(n-k)$, $I_1(x) = \int_x^\infty (1-F(u))\,\mathrm{d}u$ and $I_2(x) = \int_{-\infty}^{x} F(u)\,\mathrm{d}u$. Then from (5), we have

$$-\alpha I_1'(x) I_2(x) = I_1(x) I_2'(x)$$

from which we can obtain

$$-\alpha \ln I_1(x) = \ln I_2(x) + c.$$

This in turn yields

$$\int_{-\infty}^{x} F(u)\,\mathrm{d}u \left(\int_{x}^{\infty}(1-F(u))\,\mathrm{d}u\right)^{\alpha} = c$$

or

$$\int_{-\infty}^{x} F(u)\,\mathrm{d}u = \frac{c}{(\int_x^\infty (1-F(u))\,\mathrm{d}u)^{\alpha}}. \tag{6}$$

Differentiating (6) with respect to $x$, we obtain

$$F(x) = \frac{c\alpha(1-F(x))}{(\int_x^\infty (1-F(u))\,\mathrm{d}u)^{\alpha+1}},$$

or equivalently

$$\int_{x}^{\infty}(1-F(u))\,\mathrm{d}u = \left(c\frac{(1-F(x))}{F(x)}\right)^{1/(\alpha+1)} = c\left(\frac{1}{F(x)} - 1\right)^{1/(\alpha+1)}.$$

Differentiation of both sides of this equality gives

$$-(1-F(x)) = c\left(\frac{1}{F(x)} - 1\right)^{(1/(\alpha+1))-1}\left(-\frac{f(x)}{F^2(x)}\right),$$

which can be rewritten as

$$(1-F(x))^{(n+k-2)/(n-1)} F(x)^{(2n-k-1)/(n-1)} = cf(x), \quad c > 0. \tag{7}$$

Introducing now the quantile function $G(x)$, (7) can be expressed as

$$(1-x)^{(n+k-2)/(n-1)}x^{(2n-k-1)/(n-1)} = cf(G(x)). \tag{8}$$

Since $G'(x) = 1/f(G(x))$, we find from (8) that

$$G'(x) = \frac{c}{(1-x)^{(n+k-2)/(n-1)}x^{(2n-k-1)/(n-1)}}, \quad c > 0. \tag{9}$$

Note that $G(1) = \infty$ and $G(0) = -\infty$, since $G(\cdot)$ is a quantile function. Now substituting $\lambda = (k-1)/(n-1)$ in (9), it is not difficult to see that for $0 < x < 1$

$$G(x) = G(x; \lambda, c, d) = c\int_{\lambda}^{x} \frac{\mathrm{d}u}{(1-u)^{1+\lambda}u^{2-\lambda}} + d$$

$$= \frac{c(x-\lambda)}{\lambda(1-\lambda)(1-x)^{\lambda}x^{1-\lambda}} + d, \quad c > 0,\ 0 < \lambda < 1,\ -\infty < d < \infty.$$

Next, looking at the steps above, it can be easily checked that (2) implies (1). Hence, the theorem is proved. □

## 3. Some remarks on the family of distributions

Note that $n = 2k - 1$ in (7) yields

$$(1 - F(x))^{3/2}F(x)^{3/2} = cf(x), \quad c > 0,$$

which is the case considered recently by Nevzorov et al. (2003). This case of $\lambda = \frac{1}{2}$ corresponds to the Student's $t$ distribution with 2 degrees of freedom ($t_2$), as mentioned by these authors.

It is easy to show that $E(X) = \int_0^1 G(x)\,\mathrm{d}x = d$ and $P\{X < d\} = \lambda$, since $G(\lambda) = d$. Further, the tails of this family of distributions have the following asymptotic behavior:

$$1 - F(x) \sim c'x^{-1/\lambda} \quad \text{as } x \to \infty,$$

and

$$F(x) \sim -c''x^{-1/(1-\lambda)} \quad \text{as } x \to -\infty.$$

We plotted the function $G(x)$ for some values of $\lambda$, and these plots show that the family of distributions can approximate a number of common distributions like Tukey lambda, Cauchy and Gumbel (for the maximum).

## Acknowledgements

The authors would like to thank Professor V.B. Nevzorov for many discussions they had with him while he visited McMaster University during the summer of 2002. They also thank the referee whose suggestions were very helpful in improving the paper.

## References

Arnold, B.C., Balakrishnan, N., Nagaraja, H.N., 1992. A First Course in Order Statistics. Wiley, New York.

Bairamov, I., Balakrishnan, N., 2001. Regression of order statistics and record values and characterization of distributions. Proceedings of the International Conference on Characterizations, Modeling and Applications, Antalya, Turkey, pp. 15–22.

Ferguson, T.S., 1967. On characterizing distributions by properties of order statistics. Sankhyá Ser. A 29, 265–278.

Johnson, N.L., Kotz, S., Balakrishnan, N., 1994. Continuous Univariate Distributions, Vol. 1, 2nd Edition. Wiley, New York.

Nagaraja, H.N., 1988. Some characterizations of discrete distributions based on linear regressions of adjacent order statistics. J. Statist. Planning Inference 20, 65–75.

Nevzorov, V.B., Balakrishnan, N., Ahsanullah, M., 2003. Simple characterization of Student's $t_2$ distribution. J. Roy. Statist. Soc. Ser. D, to appear.

Wesolowsky, J., Ahsanullah, M., 1997. On characterizing distributions via linearity of regression for order statistics. Austral. J. Statist. 39 (1), 69–78.