DEMONSTRATIO MATHEMATICA
Vol. XXX No 4 1997

Anna Dembińska, Jacek Wesołowski

ON CHARACTERIZING THE EXPONENTIAL DISTRIBUTION BY LINEARITY OF REGRESSION FOR NON-ADJACENT ORDER STATISTICS

Abstract. Let $X_1, X_2, \ldots, X_n$ be a random sample from an absolutely continuous distribution with the corresponding order statistics $X_{1:n} \leq X_{2:n} \leq \ldots \leq X_{n:n}$. It is shown that the distribution is exponential if and only if $E(X_{k+r:n}|X_{k:n}) = X_{k:n} + b_{(k,r,n)}$ for a pair of triplets (k, r, n); $r \geq 3$ - two cases are considered.

1. Introduction

There are many characterizations of the exponential distribution based on properties of order statistics, among them characterizations involving some dependency assumptions - see for instance Galambos and Kotz (1978) or Azlarov and Volodin (1986).

It is well known that for a sample from an exponential distribution $X_{1:n}$, $X_{2:n} - X_{1:n}, \ldots, X_{n:n} - X_{n-1:n}$ are mutally independent. Consequently, for an exponential sample, the regression of $X_{j:n}$ on $X_{i:n}$ $(j > i)$ is linear. A natural question to ask is if this happens only for the exponential distribution.

Fisz (1958) proved that if X_1 and X_2 are independent identically distributed with an absolutely continuous distribution and if $X_{1:2}$ and $X_{2:2} - X_{1:2}$ are independent then X_1 has an exponential distribution. Further extensions of this theorem were done by Rogers (1963) and Tanis (1964).

Ferguson (1967) proved the following theorem:

if $X_1, X_2, \ldots, X_n$ is a sample from a continuous distribution and
$E(X_{k+1:n}|X_{k:n}) = aX_{k:n} + b$ for some $1 \leq k < n$,
then only the following three cases are possible:

1991 *Mathematics Subject Classification*: 60E05, 62E10, 62E15.

Key words and phrases: order statistics, characterization, exponential distribution, linearity of regression.

1. $a = 1 \Rightarrow X_1$ has an exponential distribution,
2. $a > 1 \Rightarrow X_1$ has a Pareto distribution,
3. $a < 1 \Rightarrow X_1$ has a power distribution.

(One comment is in order. Ferguson states his result assuming that for some $1 \leq k < n$, $E(X_{k:n}|X_{k+1:n}) = aX_{k+1} - b$ instead of assuming the regression $X_{k+1:n}$ on $X_{k:n}$ and arrrives at distributions dual to that given in 1-3. Here the regression $X_{k+1:n}$ on $X_{k:n}$ is used to show the resemblance to the next theorem.)

As pointed out in the monograph Arnold, Balakrishnan, Nagaraja (1992) the question raised by Ferguson (1967) about analogous characterizations for non-adjacent order statistics has not been settled until the very recent paper by Wesołowski and Ahsanullah (1997). They solved the problem considering linearity of regression of $X_{k+2:n}$ on $X_{k:n}$:

if $X_1, X_2, \ldots, X_n$ is a sample from an absolutely continuous distribution and $E(X_{k+2:n}|X_{k:n}) = aX_{k:n} + b$ for some $1 \leq k \leq n-1$,
then the same three cases 1. - 3. are the only possible.

In this paper linearity of regression of $X_{k+r:n}$ on $X_{k:n}$ for $r > 2$ is considered. However instead of a single regression condition, a pair of identities $E(X_{k_i+r_i:n_i}|X_{k_i:n_i}) = = X_{k_i:n_i} + b_i, i = 1, 2$, is considered, with suitably chosen triplets $(k_i, r_i, n_i), i = 1, 2$. In Section 2 the case of $r_i = 3, 1 \leq k_i \leq n_i - 3, i = 1, 2, n_1 - k_1 \neq n_2 - k_2$ is studied. In this case a complete characterization of exponential distribution by a pair of regression conditions is given. In Section 3 spacings greater than 3 are also allowed but only for special choices of n_i and k_i i.e. $n_1 - k_1 = n_2 - k_2 + 1$.

Let us point out that characterizations connected with regressional properties of order statistics are widely investigated nowadays. See for instance: Beg and Balasubramanian (1990), Roy and Mukherjee (1991), El-Din, Mahmoud and Youssef (1991), Swanepoel (1991), Pakes, Fakhry, Mahmoud and Ahmad (1996).

2. Any triplets with spacing equal 3

Consider i.i.d. random variables $X_1, \ldots, X_n$ with a corresponding distribution function (df) F. As mentioned in Introduction for an exponential distribution we have $E(X_{k+r:n}|X_{k:n}) = X_{k:n} + b$. Here we treat the converse implication in the case $r = 3$.

THEOREM 1. *Assume that F is absolutely continous and $E(|X_{k_i+3:n}|) < \infty$, $i = 1, 2$, where $k_1, k_2 \in \{1, \ldots, n-3\}$, $k_1 \neq k_2$. If $E(X_{k_i+3:n}|X_{k_i:n}) = X_{k_i:n} + b_i$, $i = 1, 2$, where b_1, b_2 are some real numbers then F is an exponential df.*

P r o o f. Using the formulas for the joint density function of $X_{i:n}$ and $X_{j:n}$ and the density function of $X_{i:n}$ (see for instance the monograph Arnold, Balakrishnan, Nagaraja (1992)) we can write

$$(1)\quad E(X_{k+3:n}|X_{k:n}=x)= \\ =\frac{(n-k)(n-k-1)(n-k-2)}{2\bar{F}^{n-k}(x)}\int_x^{\infty} y[\bar{F}(x)-\bar{F}(y)]^2\bar{F}^{n-k-3}(y)f(y)\,dy$$

where $\bar{F}=1-F$ and f is the density of F. Let $\mu=inf\{x: F(x)>0\}$ and $\nu=sup\{x: F(x)<1\}$. Consequently by (1)

$$(2)\quad x+b= \\ =\frac{(n-k)(n-k-1)(n-k-2)}{2\bar{F}^{n-k}(x)}\int_x^{\infty} y[\bar{F}(x)-\bar{F}(y)]^2\bar{F}^{n-k-3}(y)f(y)\,dy$$

and then there does not exist an interval (c,d), $\mu<c<d<\nu$, over which F is constant since the left hand side of (2) is increasing in such an interval and the right side is constant, while both sides are continuous, so that they could not possibly be equal at the next point of increase of F. Thus (μ,ν) is the support of F and the equation

$$(3)\quad \frac{2\bar{F}^{n-k}(x)(x+b)}{(n-k)(n-k-1)(n-k-2)}=\int_x^{\nu} y[\bar{F}(x)-\bar{F}(y)]^2\bar{F}^{n-k-3}(y)f(y)\,dy$$

holds true for any $x\in(\mu,\nu)$.

The df F is absolutely continuous so $f=F'$ L-almost everywhere on $\mathbf{R}$, where L denotes the Lebesgue measure. Differentiating both sides of (3) with respect to x, we get for L-almost all $x\in(\mu,\nu)$

$$(4)\quad -2f(x)\int_x^{\nu} y[\bar{F}(x)-\bar{F}(y)]\bar{F}^{n-k-3}(y)f(y)\,dy= \\ =\frac{-2\bar{F}^{n-k-1}(x)f(x)(x+b)}{(n-k-1)(n-k-2)}+\frac{2\bar{F}^{n-k}(x)}{(n-k)(n-k-1)(n-k-2)}.$$

Observe that (4) implies that f equals to a function continuous on (μ,ν) so we can extend f to the whole interval (μ,ν) in such a way that $f(x)=F'(x)$ and (4) hold true for any $x\in(\mu,\nu)$. Since (4) implies $f>0$ in (μ,ν), thus we can divide both sides of (4) by $-2f(x)$ obtaining

$$(5)\quad \int_x^{\nu} y[\bar{F}(x)-\bar{F}(y)]\bar{F}^{n-k-3}(y)f(y)\,dy= \\ =\frac{\bar{F}^{n-k-1}(x)(x+b)}{(n-k-1)(n-k-2)}-\frac{\bar{F}^{n-k}(x)}{(n-k)(n-k-1)(n-k-2)f(x)}.$$

Since the left-hand side of (5) is differentiable then f' exists in (μ, ν). Differentiating (5) and simplifying we have

$$(6)\qquad \int_x^\nu y\bar{F}^{n-k-3}(y)f(y)\,dy = \frac{\bar{F}^{n-k-2}(x)(x+b)}{(n-k-2)} + \\ -\frac{2\bar{F}^{n-k-1}(x)}{(n-k-1)(n-k-2)f(x)} - \frac{\bar{F}^{n-k}(x)f'(x)}{(n-k)(n-k-1)(n-k-2)f^3(x)}.$$

Thus f'' exists. If we differentiate both sides of (6) and simplify, then we obtain the following equation

$$(7)\qquad cf^5(x) - 3\bar{F}(x)f^4(x)(n-k)(n-k-1) - 3\bar{F}^2(x)f'(x)f^2(x)(n-k) + \\ +\bar{F}^3(x)f''(x)f(x) - 3\bar{F}^3(x)f'^2(x) = 0$$

where $c = b(n-k)(n-k-1)(n-k-2)$. Now denoting $y = \bar{F}$ (i.e. $f = -y', f' = -y'', f'' = -y'''$) we get by (7) a third order differential equation in (μ, ν)

$$(8)\quad cy'^5(x) - 3(n-k)(n-k-1)yy'^4 + 3(n-k) + y^2y'^2y'' + y^3y'y''' - 3y^3y''^2 = 0.$$

Substituting $u(y) = y'$ (i.e. $y'' = u'u, y''' = u''u^2 + u'^2u$) in (8) and dividing both sides by u^2 ($y' = 0$ is impossible since it yields $f = 0$) we have

$$(9)\quad y^3uu'' - 2y^3u'^2 + 3(n-k)y^2uu' - 3(n-k)(n-k-1)yu^2 - cu^3 = 0.$$

Without losing generality we can assume $k_2 > k_1$. Consequently we have two differential equations of the third order

(10)

$$\begin{cases} y^3uu'' - 2y^3u'^2 + 3(n-k_1)y^2uu' - 3(n-k_1)(n-k_1-1)yu^2 - c_1u^3 = 0 \\ y^3uu'' - 2y^3u'^2 + 3(n-k_2)y^2uu' - 3(n-k_2)(n-k_2-1)yu^2 - c_2u^3 = 0 \end{cases}$$

where $c_i = b_i(n-k_i)(n-k_i-1)(n-k_i-2)$ for $i = 1, 2$.

Subtracting them and then dividing both sides by uy^2 ($y = 0$ is impossible in (μ, ν)) we get

$$(11)\qquad (k_2-k_1)u' - 3g\frac{1}{y}u = (c_1-c_2)\frac{1}{y^2}u^2$$

where $g = (n-k_1)(n-k_1-1) - (n-k_2)(n-k_2-1)$.

The equation (11) is of a Bernoulli type. Using the routine technique we get the solution

$$(12)\qquad y' = \frac{y^{\frac{g}{k_2-k_1}}}{\frac{1}{3}\frac{c_2-c_1}{g-(k_2-k_1)}y^{\frac{g}{k_2-k_1}-1} + D}$$

which can be rewritten as

$$(13)\qquad f(x)=\frac{\bar{F}^{\frac{g}{k_2-k_1}}(x)}{-\frac{1}{3}\frac{c_2-c_1}{g-(k_2-k_1)}\bar{F}^{\frac{g}{k_2-k_1}-1}(x)-D}.$$

The left-hand side of (13) is always non-negative in (μ,ν). Upon taking limit for $x\to\nu^-$ the right-hand side has the same sign as $-D$ (since $\bar{F}(x)\to 0$). Thus $-D\geq 0$.

Solving (12) we get

$$(14)\qquad \frac{1}{3}\frac{c_2-c_1}{g-(k_2-k_1)}ln(y)+\frac{D}{1-\frac{g}{k_2-k_1}}y^{1-\frac{g}{k_2-k_1}}=x+E.$$

Let us consider two possible cases:

First case: $D=0$

$$\ln\bar{F}(x)=3\frac{g-(k_2-k_1)}{c_2-c1}(x+E)$$

$$\bar{F}(x)=\exp\left[3\frac{g-(k_2-k_1)}{c_2-c1}(x+E)\right].$$

Taking limit $x\to\mu^+$ we get

$$1=\exp\left[3\frac{g-(k_2-k_1)}{c_2-c1}(\mu+E)\right].$$

Thus $E=-\mu$ and

$$(15)\qquad \bar{F}(x)=\exp\left[3\frac{g-(k_2-k_1)}{c_2-c_1}(x-\mu)\right]\quad\text{for}\quad x\in(\mu,\nu).$$

Taking limit $x\to\nu^-$ in (15) we observe that its left-hand side equals zero, while the right-hand side remains positive. Hence $\nu=\infty$ and we get the exponential distribution.

Second case: $D<0$

Inserting (14) in (10) upon performing suitable differentations we get

$$\frac{1}{3}\frac{c_2-c_1}{g-(k_2-k_1)}[-2-3(n-k_1)(n-k_1-2)]-c_1+$$

$$Dy^{1-\frac{g}{k_2-k_1}}\left[\frac{-g}{k_2-k_1}(1+\frac{g}{k_2-k_1})+3(n-k_1)\frac{g}{k_2-k_1}-3(n-k_1)(n-k_1-1)\right]=0$$

which is a contradiction since y can not be constant on (μ,ν). Hence $D<0$ is impossible and the only solution is the exponential distribution. ∎

Remark. Theorem 1 can be easily extended to the case

$$E(X_{k_i+3:n_i}|X_{k_i:n_i})=X_{k_i:n_i}+b_i,\quad i=1,2$$

for some $1\leq k_i\leq n_i-3$ and $n_1-k_1\neq n_2-k_2$.

3. Special triplets with spacings > 3

The case of $r > 3$ is much more complicated. Applying here the method used in Section 2 we arrive at an $(r-2)$ order differential equation and we do not know how to solve it in general. Here we present a solution only for special choices of n_i and k_i: $n_1 - k_1 = n_2 - k_2 + 1$.

THEOREM 2. *Let k_1, k_2, n_1, n_2 be natural numbers such that $1 \le k_i \le n_i - r$ for $i = 1, 2$ and $r \ge 2$. Assume that F is absolutely continuous and $E(|X_{k_2+r:n_2}|) < \infty$. If $E(X_{k_i+r:n_i}|X_{k_i:n_i}) = X_{k_i:n_i} + b_i$, $i = 1, 2$, where b_1, b_2 are some real numbers and $n_1 - k_1 = n_2 - k_2 + 1$, then F is an exponential df.*

Proof. Let (μ, ν) be the support of F. Then

$$E(X_{k_1+r:n_1}|X_{k_1:n_1} = x) = x + b_1 =$$
$$\frac{(n_1-k_1)!}{(r-1)!(n_1-k_1-r)!}\,\frac{1}{\bar{F}^{n_1-k_1}(x)}\int_x^\nu y[\bar{F}(x) - \bar{F}(y)]^{r-1}\bar{F}^{n_1-k_1-r}(y)f(y)\,dy.$$

Hence

$$\frac{(r-1)!(n_2-k_2-r+1)!}{(n_2-k_2+1)!}\bar{F}^{n_2-k_2+1}(x)(x+b_1) = \tag{16}$$
$$= \int_x^\nu y[\bar{F}(x) - \bar{F}(y)]^{r-1}\bar{F}^{n_2-k_2-r+1}(y)f(y)\,dy.$$

Differentiating both sides of (16) we get

$$-\frac{(r-1)!(n_2-k_2-r+1)!}{(n_2-k_2)!}\bar{F}^{n_2-k_2}(x)f(x)(x+b_1) + \tag{17}$$
$$\frac{(r-1)!(n_2-k_2-r+1)!}{(n_2-k_2+1)!}\bar{F}^{n_2-k_2+1}(x) =$$
$$= (r-1)f(x)\int_x^\nu y[\bar{F}(x) - \bar{F}(y)]^{r-2}[-\bar{F}(y)]\bar{F}^{n_2-k_2-r}(y)f(y)\,dy.$$

Substituting $-\bar{F}(y) = [\bar{F}(x) - \bar{F}(y)] - \bar{F}(x)$ in (17) and dividing both sides by $(r-1)f(x)$ we get

$$-\frac{(r-2)!(n_2-k_2-r+1)!}{(n_2-k_2)!}\bar{F}^{n_2-k_2}(x)(x+b_1) + \tag{18}$$
$$+\frac{(r-2)!(n_2-k_2-r+1)!}{(n_2-k_2+1)!}\,\frac{\bar{F}^{n_2-k_2+1}(x)}{f(x)} =$$

$$= \int_x^\nu y[\bar F(x) - \bar F(y)]^{r-1} \bar F^{n_2-k_2-r}(y) f(y)\,dy +$$

$$-\bar F(x) \int_x^\nu y[\bar F(x) - \bar F(y)]^{r-2} \bar F^{n_2-k_2-r}(y) f(y)\,dy.$$

Upon assumptions we can write

$$\frac{(r-1)!(n_2-k_2-r)!}{(n_2-k_2)!} \bar F^{n_2-k_2}(x)(x+b_2) = \tag{19}$$

$$= \int_x^\nu y[\bar F(x) - \bar F(y)]^{r-1} \bar F^{n_2-k_2-r}(y) f(y)\,dy.$$

Differentiating both sides of (19) and dividing by $(r-1)f(x)$ we have

$$\tfrac{(r-2)!(n_2-k_2-r)!}{(n_2-k_2-1)!} \bar F^{n_2-k_2-1}(x)(x+b_2) - \tfrac{(r-2)!(n_2-k_2-r)!}{(n_2-k_2)!} \tfrac{\bar F^{n_2-k_2}(x)}{f(x)} = \tag{20}$$

$$= \int_x^\nu y[\bar F(x) - \bar F(y)]^{r-2} \bar F^{n_2-k_2-r}(y) f(y)\,dy.$$

Substituting (19) and (20) to (18) and simplyfing we arrive at

$$(n_2-k_2-r+1)(n_2-k_2+1)f(x)(b_2-b_1) = r\bar F(x). \tag{21}$$

Equation (21) is easy to solve if we substitute $y = \bar F(x)$ (i.e. $-y' = f(x)$).The only solution turns out to be the df of the exponential distribution: $\bar F(x) = \exp[\delta(x-\mu)]$ for $x \in (\mu, \infty)$ where $\delta = \frac{-r}{(n_2-k_2-r+1)(n_2-k_2+1)(b_2-b_1)}$. ■

References

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

[2] T. A. Azlarov, N. A. Volodin, *Characterization Problems Associated with the Exponential Distribution*, Springer, New York, 1986.

[3] M. I. Beg, K. Balasubramanian, *Distributions determined by conditioning on a single order statistic*, Metrika 37 (1990), 37–43.

[4] M. M. M. El-din, M. A. W. Mahmoud, S. A. E. Youssef, *Moments of order statistics from parabolic and skewed distributions and a characterization of Weibull distribution*, Comm. Statist. Simulation Comput. 20 (1991), 639–645.

[5] T. S. Ferguson, *On characterizing distributions by properties of order statistics*, Sankhya, Ser A 29 (1967), 265–278.

[6] M. Fisz, *Characterization of some probability distributions*, Skandinavisk Aktuarietidskrift 41 (1958), 65–70.

[7] J. Galambos, S. Kotz, *Characterizations of Probability Distributions*, Lecture Notes in Mathematics 675, Springer, Berlin, 1978.

[8] A. G. Pakes, M. E. Fakhry, M. R. Mahmoud, A. A. Ahmad, *Characterizations by regressions of order statistics*, J. Appl. Stat. Sci. 3(1) (1996), 11–23.

[9] G. S. Rogers, *An alternative proof of the characterization of the density Ax^B*, Amer. Math. Monthly 70 (1963), 857–858.
[10] D. Roy, S. P. Mukherjee, *Revival of a life-time distribution*, Economic Quality Control 6 (1991), 164–173.
[11] J. W. H. Swanepoel, *A note on the characterization of the exponential distribution by the covariance of spacings*, South African Statist. J. 25 (1991), 73–78.
[12] E. Tanis, *Linear forms in the order statistics from an exponential distribution*, Ann. Math. Statist. 35 (1964), 270–276.
[13] J. Wesołowski, M. Ahsanullah, *On characterizing distributions via linearity of regression for order statistics*, Austral. J. Statist. 39 (1997), 69–78.

INSTITUTE OF MATHEMATICS
WARSAW UNIVERSITY OF TECHNOLOGY
Plac Politechniki 1
00-661 WARSAW, POLAND

Received July 23, 1997.