Characterizations of distributions via order statistics with random exponential shifts

M. Ahsanullah   V.B. Nevzorov   G.P. Yanev

Abstract

We characterize probability distributions via equalities in law between two order statistics shifted by independent exponential variables. An explicit formula for the quintile function of the identified family of distributions is obtained. The results extend some known characterizations of exponential and logistic distributions.

Keywords characterizations, order statistics, exponential distribution, generalized logistic distribution.

1 Introduction

Distributional (in law) relations between order statistics are useful and elegant tools for characterizing probability distributions. More specifically, there is a large number of publications on characterizations based on recurrences involving both order statistics and standard exponential variables. An excellent review and discussion of the available results on this subject can be found in [3].

AlZaid and Ahsanullah [2] showed that a non-negative random variable X with an absolutely continuous cumulative distribution function (cdf) has standard exponential distribution iff for a fixed k, such that $1 \leq k \leq n-1$,

$$X_{k,n} + \frac{\xi}{n-k} \stackrel{d}{=} X_{k+1,n}, \tag{1}$$

where $X_{1,n}, X_{2,n} \ldots, X_{n,n}$ are the order statistics in a sample with parent X and ξ is standard exponential and independent from $X_{k,n}$. Please, see also [9] and [7] for an alternative proof and some extensions in the context of random contractions (cf. [6], p.14). On the other hand, it was shown in Ahsanullah et al. [1], under some additional regularity assumptions, that a random variable X with absolutely continuous cdf is standard logistic if and only if for a fixed k such that $1 \leq k \leq n-1$,

$$X_{k,n} + \frac{\xi_1}{n-k} \stackrel{d}{=} X_{k+1,n} - \frac{\xi_2}{k}, \tag{2}$$

where ξ_1 and ξ_2 are independent standard exponential variables, which are also independent from $X_{k,n}$ and $X_{k+1,n}$, respectively.

1

First, we will study some extensions of the distributional relations (1) and (2) involving order statistics shifted by independent exponential random variables. What can be said about the distribution of the parent variable X if the following more general distributional equality holds

$$X_{k,n} + a\xi_1 \stackrel{d}{=} X_{r,n} - b\xi_2, \qquad a \geq 0, \quad b \geq 0, \quad 1 \leq k < r \leq n, \qquad (3)$$

where ξ_1 and ξ_2 are independent standard exponential variables, which are also independent from $X_{k,n}$ and $X_{r,n}$, respectively?

Our first result answers the above question in the case of adjacent order statistics, i.e., when $r = k + 1$. Throughout this paper we will use the term $Q(y)$ to refer to the pseudo-inverse (quintile) function of $F(x)$, i.e., $Q(y) = \inf\{x : F(x) \geq y\}$ for $y \in (0, 1)$.

Theorem 1 *Let k be a fixed integer such that $1 \leq k \leq n - 1$ and $a \geq 0$ and $b \geq 0$ be two fixed real numbers such that $(a, b) \neq (0, 0)$. Assume X is a random variable with continuous cdf $F(x)$ and ξ_1 and ξ_2 are independent standard exponential variables, which are also independent from $X_{k,n}$ and $X_{k+1,n}$, respectively. Then*

$$X_{k,n} + a\xi_1 \stackrel{d}{=} X_{k+1,n} - b\xi_2 \qquad (4)$$

holds if and only if the quintile function $Q(x)$ satisfies

$$Q(x) = \log\left\{cx^{bk}(1-x)^{-a(n-k)}\right\}, \qquad 0 < x < 1, \qquad (5)$$

where $c > 0$ is an arbitrary constant.

Remarks. The following particular cases of (4) and (5) might be of independent interest. (i) Setting $a = 1/(n - k)$ and $b = 0$ in (5), we obtain $F(x) = 1 - e^{-(x-c)}$ for $x \geq c$, i.e., an exponential cdf. (ii) If $a = 0$ and $b = 1$ we have $F(x) = \exp\{x - c\}$ for $-\infty < x \leq c$. (iii) If $bk = a(n - k)$, then (5) yields the logistic cdf (cf. [5], p.114)

$$F(x) = \frac{c_1}{c_1 + \exp\{-x\}}, \qquad -\infty < x < \infty, \quad c_1 > 0.$$

(iv) If $n = 2k$ and $a = b \neq 0$, then Theorem 1 implies

$$X_{k,2k} + a\xi_1 \stackrel{d}{=} X_{k+1,2k} - a\xi_2 \quad \text{iff} \quad F(x) = \frac{1}{1 + c_2 \exp\{-x/k\}}, \quad c_2 > 0.$$

The next result addresses the case when (3) involves non-adjacent (two spacings away) order statistics.

Theorem 2 *Let k be a fixed integer such that $1 \leq k \leq n - 2$ and $a \geq 0$ and $b \geq 0$ be two fixed real numbers such that $(a, b) \neq (0, 0)$. Assume X is a random variable with continuous cdf $F(x)$ and ξ_1 and ξ_2 are independent standard exponential variables, which are also independent from $X_{k,n}$ and $X_{k+2,n}$, respectively. Then*

$$X_{k,n} + a\xi_1 \stackrel{d}{=} X_{k+2,n} - b\xi_2 \qquad (6)$$

holds if and only if the quintile function $Q(x)$ satisfies

$$Q(x) = \log\left\{cx^{bk}(1-x)^{-a(n-k-1)}\right\} - W(x;k,n), \qquad 0 < x < 1, \qquad (7)$$

where $c > 0$ is an arbitrary constant and

$$W(x;k,n) = \begin{cases} \dfrac{d\log\{(n-2k+1)x+k+1\}}{n-2k+1}, & if \quad n \neq 2k+1; \\ \\ (a+b)kx, & if \quad n = 2k+1, \end{cases} \qquad (8)$$

where $d = bk(n-k) + a(n-k-1)(k+1)$.

Remark. Note that if $n = 2k+1$ and $a = b = 1/k$, then Theorem 2 implies

$$X_{k,2k+1} + \frac{\xi_1}{k} \stackrel{d}{=} X_{k+2,2k+1} - \frac{\xi_2}{k} \quad \text{iff} \quad Q(x) = \log\left\{\frac{cx}{1-x}\right\} - 2x.$$

A number of characterizations of the logistic distribution use distributional relations between X and order statistics with positive and negative exponential random shifts. George and Mudholkar [4] (see also [8]) proved that X is standard logistic if and only if

$$X \stackrel{d}{=} X_{1,2} + \xi \qquad \text{or} \qquad X \stackrel{d}{=} X_{2,2} - \xi.$$

More generally, Lin and Hu [8] established that, under some smoothness conditions, X is standard logistic if and only if

$$X \stackrel{d}{=} X_{1,n} + \sum_{j=1}^{n-1} \frac{\xi_j}{j},$$

where ξ_j for $j = 1, 2, \ldots, n-1$ are independent standard exponential variables, which are also independent from $X_{1,n}$. Ahsanullah et al. [1] proved that the standard logistic distribution is characterized by the following distributional equality holding for a fixed k, such that $1 \leq k \leq n-1$,

$$X \stackrel{d}{=} X_{1,n} + \sum_{j=1}^{n-k} \frac{\xi_1(j)}{j} - \sum_{j=1}^{k-1} \frac{\xi_2(j)}{j},$$

where $X_{1,n}$, $\xi_1(j)$, and $\xi_2(j)$ for $j = 1, 2, \ldots, n-1$ are mutually independent and all ξ's are standard exponential. Finally, Zykov and Nevzorov [10] obtained characterizations based on either

$$X \stackrel{d}{=} X_{n,n} - \xi \qquad \text{or} \qquad X + \xi \stackrel{d}{=} X_{n,n}. \qquad (9)$$

Exploring the so-called F^{α}-scheme (cf. [6], Lecture 25), we will study characterizations of logistic and related distributions based on one extension of (9).

3

Theorem 3 *Let $a \geq 0$ and $b \geq 0$ be two fixed real numbers such that $(a, b) \neq (0, 0)$ and $F(x)$ be a continuous distribution function. Suppose Y_1 and Y_2 are independent random variables with distribution functions $F^\alpha(x)$ and $F^\beta(x)$, respectively for $\alpha > 0$ and $\beta > 0$. Furthermore, let ξ_1 and ξ_2 be independent standard exponential variables, which are also independent from Y_1 and $\max\{Y_1, Y_2\}$, respectively. Then*

$$Y_1 + a\xi_1 \stackrel{d}{=} \max\{Y_1, Y_2\} - b\xi_2 \tag{10}$$

holds if and only if the quintile function $Q(x)$ satisfies

$$Q(x) = \log\left\{cx^{b\alpha}(1 - x^\beta)^{-d}\right\}, \qquad 0 < x < 1, \tag{11}$$

where $c > 0$ is an arbitrary constant and $d = [a(\alpha + \beta) + b\alpha]/\beta$.

Remarks. (i) Setting $a = 1$ and $b = 0$ in (10), yields the cdf

$$F(x) = \left(1 - \exp\left\{-\frac{\beta}{\alpha + \beta}(x - c)\right\}\right)^{1/\beta}, \qquad c \leq x < \infty.$$

(ii) If $b\alpha = \beta$ and $d = 1$, then (11) is the quintile function of the Type I generalized logistic cdf (cf. [5], p. 140)

$$F(x) = \left(\frac{c_2}{c_2 + \exp\{-x\}}\right)^{1/\beta}, \qquad -\infty < x < \infty, \quad c_2 > 0.$$

The special case of Theorem 3 when $\alpha = 1$ and $\beta = n - 1$ is particularly noteworthy.

Corollary *Let $a \geq 0$ and $b \geq 0$ be two real numbers such that $(a, b) \neq (0, 0)$. Assume X is a random variable with continuous cdf $F(x)$ and ξ_1 and ξ_2 are independent standard exponential variables, which are also independent from X and $X_{n,n}$, respectively. Then*

$$X + a\xi_1 \stackrel{d}{=} X_{n,n} - b\xi_2, \qquad n \geq 2, \tag{12}$$

holds if and only if the quintile function $Q(x)$ satisfies

$$Q(x) = \log\left\{cx^b(1 - x^{n-1})^{-d}\right\}, \qquad 0 < x < 1, \tag{13}$$

where $c > 0$ is an arbitrary constant and $d = (b + na)/(n - 1)$.

The rest of the paper is organized as follows. In Section 2, we present an auxiliary result before proving Theorem 1. Section 3 deals with the case of non-adjacent order statistics. The proof of Theorem 3 is given in Section 4. We summarize the findings and discuss potential future work in the last section.

2   Adjacent order statistics

Let Z_1 and Z_2 be two random variables with continuous distribution functions G and H, respectively. Let ξ_1 and ξ_2 be independent standard exponential variables, which are also independent from Z_1 and Z_2, respectively.

Lemma *Let a and b be two real numbers such that $(a, b) \neq (0, 0)$. Then*

$$Z_1 + a\xi_1 \overset{d}{=} Z_2 - b\xi_2, \tag{14}$$

if and only if

$$bG'(x) + aH'(x) = G(x) - H(x). \tag{15}$$

Proof. Let us first prove that (14) implies (15). Assume both $a \neq 0$ and $b \neq 0$. It is not difficult to see that (14) is equivalent to

$$\frac{1}{a} \int_{-\infty}^{x} G(u) \exp\left\{\frac{u - x}{a}\right\} du = \frac{1}{b} \int_{x}^{\infty} H(u) \exp\left\{\frac{x - u}{b}\right\} du,$$

which, in turn, implies

$$b \int_{-\infty}^{x} G(u) \exp\left\{\frac{u}{a}\right\} du = a \exp\left\{\frac{(a + b)x}{ab}\right\} \int_{x}^{\infty} H(u) \exp\left\{-\frac{u}{b}\right\} du.$$

Since G and H are continuous, we can differentiate with respect to x both sides of the above equation and obtain

$$\exp\left\{-\frac{x}{b}\right\} [bG(x) + aH(x)] = \frac{a + b}{b} \int_{x}^{\infty} H(u) \exp\left\{-\frac{u}{b}\right\} du.$$

Differentiating with respect to x again (using the continuity of $H(x)$) we obtain (15). The proof when exactly one of a and b equals zero is similar and is omitted.

To prove that (15) implies (14), one needs to follow the steps in the above proof in reverse order. The proof of the lemma is complete.

Further on we denote $F_{i,n}(x) = P(X_{i,n} \leq x)$ for $1 \leq i \leq n$. Recall that

$$F_{i,n}(x) = \sum_{j=i}^{n} \binom{n}{j} F^j(x)(1 - F(x))^{n-j}, \qquad 1 \leq i \leq n. \tag{16}$$

Proof of Theorem 1. First, we will prove that (4) implies (5). Assume both $a \neq 0$ and $b \neq 0$. It follows from Lemma 1 with $G(x) = F_{k,n}(x)$ and $H(x) = F_{k+1,n}(x)$, making use of (16), that

$$
\begin{aligned}
bF'_{k,n}(x) + aF'_{k+1,n}(x) &= F_{k,n}(x) - F_{k+1,n}(x) \tag{17} \\
&= \binom{n}{k} F^k(x)(1 - F(x))^{n-k}.
\end{aligned}
$$

On the other hand, for the left-hand side of (17), using again (16), we obtain

$$bF'_{k,n}(x) + aF'_{k+1,n}(x) \tag{18}$$

$$= (a+b)\frac{d}{dx}\left[\sum_{j=k+1}^{n}\binom{n}{j}F^j(x)(1-F(x))^{n-j}\right] + b\frac{d}{dx}\left[\binom{n}{k}F^k(x)(1-F(x))^{n-k}\right]$$

$$= (a+b)n\binom{n-1}{k}F^k(x)(1-F(x))^{n-k-1}F'(x)$$

$$+bnF^{k-1}(x)(1-F(x))^{n-k-1}F'(x)\left[\binom{n-1}{k-1}(1-F(x)) - \binom{n-1}{k}F(x)\right]$$

$$= an\binom{n-1}{k}F^k(x)(1-F(x))^{n-k-1}F'(x) + bn\binom{n-1}{k-1}F^{k-1}(x)(1-F(x))^{n-k}F'(x)$$

$$= \binom{n}{k}F^{k-1}(x)(1-F(x))^{n-k-1}F'(x)[a(n-k)F(x) + bk(1-F(x))].$$

Equating the right-hand sides of (17) and (18), and dividing throughout by $\binom{n}{k}F^k(x)(1-F(x))^{n-k} \neq 0$, we obtain

$$bk\frac{F'(x)}{F(x)} + a(n-k)\frac{F'(x)}{1-F(x)} = 1,$$

which, upon integration with respect to x, yields

$$\log\left\{F^{bk}(x)(1-F(x))^{-a(n-k)}\right\} = x - c, \tag{19}$$

for an arbitrary constant $c > 0$. Finally, replacing x with $Q(x)$ in the last equation, we obtain (5). The proof for the cases when either a or b is zero is similar and is omitted here.

To prove the "if" part of the theorem, i.e., (5) implies (4), one needs to start with (19) and repeat the steps of the above proof going backwards to (17).

3 Non-adjacent order statistics

In the beginning we will prove that (6) yields (7). Assume both $a \neq 0$ and $b \neq 0$. Referring to (16), we have

$$F_{k,n}(x) - F_{k+2,n}(x) \tag{20}$$

$$= \binom{n}{k}F^k(x)(1-F(x))^{n-k} + \binom{n}{k+1}F^{k+1}(x)(1-F(x))^{n-k-1}.$$

On the other hand, similarly to (18), we obtain

$$bF'_{k,n}(x) + aF'_{k+2,n}(x) = bk\binom{n}{k}F^{k-1}(x)(1-F(x))^{n-k}F'(x) \tag{21}$$

$$+ \quad a(n-k-1)\binom{n}{k+1}F^{k+1}(x)(1-F(x))^{n-k-2}F'(x).$$

6

Lemma 1 with $G(x) = F_{k,n}(x)$ and $H(x) = F_{k+2,n}(x)$ implies that the left-hand sides of (20) and (21) are equal. Equating the right-hand sides of these equations, it is not difficult to see that

$$F'(x) = \frac{(k+1)F(x)(1-F(x))^2 + (n-k)F^2(x)(1-F(x))}{bk(k+1)(1-F(x))^2 + a(n-k-1)(n-k)F^2(x)}.$$

Replacing x with $Q(x)$ and using that $F'(Q(x)) = 1/Q'(x)$, we obtain

$$Q'(x) = \frac{a(n-k-1)(n-k)x^2 + bk(k+1)(1-x)^2}{x(1-x)(k+1+(n-2k-1)x)}. \tag{22}$$

If $n \neq 2k+1$, then (22) can be written as

$$Q'(x) = \frac{a(n-k-1)}{1-x} + \frac{bk}{x} - \frac{a(n-k-1)(k+1) + bk(n-k)}{k+1+(n-2k-1)x}. \tag{23}$$

After integrating with respect to x we obtain (7) with the upper branch of (8). In the case $n = 2k+1$, (22) becomes

$$Q'(x) = \frac{ak}{1-x} + \frac{bk}{x} - (a+b)k \tag{24}$$

and (7) follows after integrating with respect to x again. The proof when exactly one of a and b equals zero is omitted.

Starting with (23) and (24) and following the steps of the above proof going backwards, one can verify that (7) yields (6).

4 F^α scheme

Let us first prove that (10) implies (11). Assume $a \neq 0$ and $b \neq 0$. Furthermore, assume $\beta = 1$. It follows from Lemma 1 with $G(x) = F^\alpha(x)$ and $H(x) = F^{\alpha+1}(x)$ that

$$b\alpha F^{\alpha-1}(x)F'(x) + a(\alpha+1)F^\alpha(x)F'(x) = F^\alpha(x) - F^{\alpha+1}(x),$$

which, upon division by $F^{\alpha-1}(x) \neq 0$, becomes

$$F'(x)\left[b\alpha + a(\alpha+1)F(x)\right] = F(x)(1 - F(x)).$$

This, for $F(x)(1 - F(x)) \neq 0$, can be written as

$$\frac{1}{F'(x)} = \frac{b\alpha}{F(x)} + \frac{(a+b)\alpha + a}{1 - F(x)}.$$

Replacing x with the quintile function $Q(x)$, we obtain

$$Q'(x) = \frac{b\alpha}{x} + \frac{(a+b)\alpha + a}{1 - x}.$$

Integrating both sides with respect to x, we have

$$Q(x) = \log\left\{cx^{b\alpha}(1-x)^{-(a+b)\alpha-a}\right\},\tag{25}$$

where $c > 0$ is an arbitrary constant. This completes the proof when $\beta = 1$. To extend this proof to the general case $\beta > 0$, consider the independent random variables Y_1' and Y_2' with cdf's $(H(x))^{\alpha/\beta}$ and $H(x)$, respectively, where $H(x^\beta) = F(x)$. It follows from (25) that the quintile function $H^{-1}(x)$ satisfies

$$H^{-1}(x) = \log\left\{cx^{b\alpha/\beta}(1-x)^{-(a+b)\alpha/\beta-a}\right\}.\tag{26}$$

Now, taking into account that $Q(x) = H^{-1}(x^\beta)$, we see that (26) yields (11). The proof of the cases when either a and b is zero is omitted.

It is not difficult to verify that (11) implies (10), following the steps of the above proof in reverse order.

5 Concluding remarks

We obtained characterizations based on distributional equalities between two order statistics plus or minus multiples of independent standard exponential variables. The resulting family of distributions includes as its members exponential, logistic, and generalized logistic. In case of adjacent order statistics, the quintile function of the underlying distribution takes on a compact explicit form. If the order statistics are two spacings away, the expression for the quintile function includes an additional additive term, which is linear if the order statistics are on both sides of the median. Further calculations show that if higher-order spacings are involved then the resulting quintile function will have more additive terms.

Exploring the so called F^α-scheme, we studied the distribution of a continuous X for which X plus a multiple of a standard exponential variable equals the maximum order statistic minus a multiple of another standard exponential variable. The obtained results generalize those of Zykov and Nevzorov (2010).

One area of future work will be to study in more detail the distributions with quintile functions (5), (7), and (11). It also remains to be seen what the distribution of X is if the corresponding order statistics differ by a linear combination of standard exponential variables. Some results in this direction are given in Ahsanullah et al. [1].

References

[1] Ahsanullah, M., Yanev, G.P., Onica, C. Characterizations of logistic distribution based on distributional properties of order statistics, submitted.

[2] AlZaid, A.A., Ahsanullah, M. (2003). A characterization of the Gumbel distribution based on record values. Commun Stat. - Theory and Methods. 32, 11:2101-2108.

[3] Gather, U., Kamps, U., Schweitzer, N. (1998). Characterizations via identically distributed functions of order statistics. In: Balakrishnan, N., Rao, C.R. (Eds.) Handbook of Statistics, Vol. 16, 257-290, Elsevier, Amsterdam.

[4] George, E.O., Mudholkar, G.S. (1981). Some relationships between the logistic and the exponential distributions. In: C. Taillie et al. (Eds.) Statistical Distributions in Scientific Work, Vol 4, Reidel, Dordrecht, 401-409.

[5] Johnson, N.L., Kotz, S., Balakrishnan, N. (1995). Continuous Univariate Distributions, Vol. 2, Second Edition, John Willey & Sons, New York.

[6] Nevzorov, V.B. (2001). Records: Mathematical Theory, American Math. Soc., Providence, Rhode Island.

[7] Nevzorov, V.B., Akhundov, I. (2008). On an extension of a family of distributions introduced by Weselowski and Ahsanullah. J. Stat. Theory Appl., 7, 4:379-384.

[8] Lin, G.D., Hu, C.-Y. (2008). On characterizations of the logistic distribution. J. Statistical Planning and Inference, 138, 1147-1156.

[9] Wesołowski, J., Ahsanullah, M. (2004). Switching order statistics through random power contractions. Aust. N.Z. J. Stat., 42, 2:297-303.

[10] Zykov, V.O., Nevzorov, V.B. (2010). On some characterizations of families of distributions, including logistic or exponential, by properties of order statistics. Notes of Sci . Semin. POMI, v.384, 176-181 (in Russian).