The Annals of Statistics
1995, Vol. 23, No. 3, 769–773

A NOTE ON A CHARACTERIZATION OF THE EXPONENTIAL DISTRIBUTION BASED ON A TYPE II CENSORED SAMPLE

BY JIAN-LUN XU AND GRACE L. YANG[1]

National Cancer Institute and University of Maryland

Let $X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}$ be the order statistics of a random sample of n lifetimes. The total-time-on-test statistic at $X_{(i)}$ is defined by $S_{i,n} = \sum_{j=1}^{i}(n-j+1)(X_{(j)} - X_{(j-1)})$, $1 \le i \le n$. A type II censored sample is composed of the r smallest observations and the remaining $n-r$ lifetimes which are known only to be at least as large as $X_{(r)}$. Dufour conjectured that if the vector of proportions $(S_{1,n}/S_{r,n}, \dots, S_{r-1,n}/S_{r,n})$ has the distribution of the order statistics of $r-1$ uniform(0, 1) random variables, then X_1 has an exponential distribution. Leslie and van Eeden proved the conjecture provided $n-r$ is no larger than $(1/3)n - 1$. It is shown in this note that the conjecture is true in general for $n \ge r \ge 5$. If the random variable under consideration has either NBU or NWU distribution, then it is true for $n \ge r \ge 2$, $n \ge 3$. The lower bounds obtained here do not depend on the sample size.

1. Introduction. Let $X_{(1)} \le X_{(2)} \le \cdots \le X_{(n)}$ be the order statistics of n independent, identically distributed (iid), nonnegative random variables $X_1, \dots, X_n$. Let r be an integer satisfying $2 \le r \le n$, $n \ge 3$. Set $X_{(0)} = 0$. Define

$$(1)\qquad \begin{aligned} S_{i,n} &= \sum_{j=1}^{i}(n-j+1)\big(X_{(j)} - X_{(j-1)}\big), \qquad i = 1, \dots, n, \\ W_{r,n} &= \big(S_{1,n}/S_{r,n}, S_{2,n}/S_{r,n}, \dots, S_{r-1,n}/S_{r,n}\big). \end{aligned}$$

A conjecture stated by Dufour (1982) is that if $W_{r,n}$ is distributed as the random vector $U_{(\cdot)}(r-1) = (U_{(1)}, \dots, U_{(r-1)})$, where $U_{(1)} \le U_{(2)} \le \cdots \le U_{(r-1)}$ are the order statistics of $r-1$ iid random variables with a uniform distribution on the interval $(0,1)$, then X_1 has an exponential distribution. This result, if true, will characterize the exponential distribution.

The characterization problem for the complete sample, that is, when $r = n$, has been studied by Seshadri, Csörgő and Stephens (1969) and Dufour, Maag and van Eeden (1984). Dufour's conjecture has been partially proved by Leslie and van Eeden (1993), who showed that the conjecture is true for $r \ge (2/3)n + 1$, but the case $r < (2/3)n + 1$ has not been determined. The principal

Received May 24, 1994; revised November 14, 1994

[1]Supported in part by Office of Naval Research, N00014-93-J-1097.

AMS 1991 *subject classifications*. Primary 62E10; secondary 62E15.

Key words and phrases. Characterization, exponential distribution, type II right-censored data.

application of the number r is in type II sampling of lifetimes. In that setting, n brand-new items are put on test and the experiment is terminated upon observation of a predetermined number r of complete lifetimes, and $S_{r,n}$ is the observed total time on test.

It is shown in Section 2 that the conjecture is true when $r \geq 5$. Here the lower bound is independent of the sample size n, whereas in Leslie and van Eeden (1993) the lower bound increases with n. The cases $r = 2$, 3 and 4 are still not determined. It is known, however, that for $r = 2 = n$ the conjecture is false [see Menon and Seshadri (1975)]. Thus we assume $n \geq 3$. If, however, the distribution of X_1 is restricted to either the class of "new better than used" (NBU) or "new worse than used" (NWU) distributions, then Dufour's conjecture is true when $r \geq 2$.

2. Characterization. Following Leslie and van Eeden (1993), we use $Z_1 \sim Z_2$ to denote that the random vectors Z_1 and Z_2 have the same distribution. Let $F(x) = P(X_1 \leq x)$ be the cumulative distribution function (cdf) of X_1. Let $\bar{F}(x) = 1 - F(x)$, $0 \leq x < \infty$, and $F^{-1}(u) = \inf\{x: F(x) \geq u\}$, for $0 < u < 1$.

The following lemma is used in the proof of Theorem 1.

Write $R = n - r + 1$. Let Z be a random variable that is independent of the sample $\{X_1, \dots, X_n\}$ and has the same distribution as $\min\{X_1, \dots, X_R\}$. Denote by G_R the cdf of Z.

LEMMA 1. *Let r and n be fixed. For every j satisfying $2 \leq j \leq r \leq n$, $W_{r,n} \sim U_{(\cdot)}(r-1)$ implies that*

$$P(X_i > s_i Z,\ i=1,\dots,j-1) = \int_0^\infty \prod_{i=1}^{j-1} \bar{F}(s_i z)\, dG_R(z) = R/(R + s_1 + \cdots + s_{j-1}), \tag{2}$$

for $0 \leq s_i \leq 1$, $1 \leq i \leq j-1$.

PROOF. We prove the result by induction on j. The case $j = 2$ is Lemma 3.7 of Leslie and van Eeden (1993) for $j = 2$. For $j \geq 3$, it is convenient to write $P(X_i \leq s_i Z, 1 \leq i \leq j-1)$ as $L_{j-1}(s_1, \dots, s_{j-1})$. Integrating out t_{j-1} in Lemma 3.7 of Leslie and van Eeden (1993) yields a recurrence formula:

$$L_{j-1}(s_1, \dots, s_{j-1}) = L_{j-2}(s_1, \dots, s_{j-2}) - \frac{1}{a_{j-1}} L_{j-2}\left(\frac{s_1}{a_{j-1}}, \dots, \frac{s_{j-2}}{a_{j-1}}\right), \tag{3}$$

where $a_{j-1} = 1 + (s_{j-1}/\mathrm{R})$ and L_0 is set equal to 1. Suppose that (2) holds for $j = k$. To show it holds for $j = k+1$, we let $A_i(t) = \{X_i \leq tZ\}$ and let $A_i^c(t)$ be its complement. Then

$$P\left(\bigcap_{i=1}^{k} A_i^c(s_i)\right) = P\left(\bigcap_{i=1}^{k-1} A_i^c(s_i)\right) - P(A_k(s_k)) + P\left(\bigcup_{i=1}^{k-1} A_i(s_i) A_k(s_k)\right). \tag{4}$$

Writing out the last term on the right-hand side of (4) using the inclusion–exclusion formula, we obtain, in terms of the notation L_j, that

$$P\left(\bigcup_{i=1}^{k-1} A_i(s_i)A_k(s_k)\right) = \sum_{i=1}^{k-1} L_2(s_i, s_k) - \sum_{i<l} L_3(s_i, s_l, s_k) + \sum_{i<l<m} L_4(s_i, s_l, s_m, s_k) - \cdots + (-1)^k L_k(s_1, \dots, s_k). \tag{5}$$

Applying (3) to each L-term in (5) reduces the number of A-events in the intersections by 1. For instance,

$$L_3(s_i, s_l, s_k) = P\{A_i(s_i)A_l(s_l)A_k(s_k)\} = P\{A_i(s_i)A_l(s_l)\} - \frac{1}{a_k}P\left\{A_i\left(\frac{s_i}{a_k}\right)A_l\left(\frac{s_l}{a_k}\right)\right\},$$

and event $A_k(s_k)$ is eliminated. Applying the inclusion–exclusion formula to this new expression for the right-hand side of (5), we conclude immediately that

$$P\left(\bigcup_{i=1}^{k-1} A_i(s_i)A_k(s_k)\right) = P\left(\bigcup_{i=1}^{k-1} A_i(s_i)\right) - \frac{1}{a_k}P\left(\bigcup_{i=1}^{k-1} A_i\left(\frac{s_i}{a_k}\right)\right).$$

Substituting this into (4) and using the induction hypothesis yield that

$$P\left(\bigcap_{j=1}^{k} A_j^c(s_j)\right) = P(A_k^c(s_k)) - \frac{1}{a_k}\left[1 - P\left(\bigcap_{i=1}^{k-1} A_i^c\left(\frac{s_i}{a_k}\right)\right)\right] = \frac{R}{R + s_1 + \cdots + s_k}. \qquad \square$$

THEOREM 1. *Suppose that $n \ge r \ge 5$. Then X_1 has an exponential distribution if and only if $W_{r,n} \sim U_{(\cdot)}(r-1)$.*

PROOF. If X_1 is exponentially distributed, it is trivially true that $W_{r,n} \sim U_{(\cdot)}(r-1)$. We prove the converse. Applying (2), it is easy to see that

$$\int_0^\infty \left[\bar{F}(s_1 x)\bar{F}(s_2 x) - \bar{F}(sx)\right]^2 dG_R(x) \equiv 0, \tag{6}$$

for all $0 \le s, s_1, s_2 \le 1$ provided $s = s_1 + s_2$. Set $s = 1$ in (6). Introduce a new cdf, $H(x) = 1 - \bar{F}(s_1 x)\bar{F}((1-s_1)x)$. As shown by Leslie and van Eeden (1993), $W_{r,n} \sim U_{(\cdot)}(r-1)$ implies that F is continuous, and hence H is continuous. By the change of variable $u = F(x)$ and using the fact $G_R(x) = 1 - [\bar{F}(x)]^R$, (6) becomes

$$R\int_0^1 \left[H(F^{-1}(u)) - u\right]^2 (1-u)^{R-1}\, du = 0. \tag{7}$$

Therefore, $H(F^{-1}(u)) = u$ almost everywhere with respect to Lebesgue measure on (0, 1). Since F and H are continuous cdf's, we conclude that $H = F$, that is,

$$\bar{F}(s_1 x)\bar{F}((1 - s_1)x) = \bar{F}(x), \tag{8}$$

for all $0 \le s_1 \le 1$ and $x \in [0, \infty)$. A more general treatment of an equation similar to (7) may be found in the lemma in Lin [(1990), Section 2]. Setting $s_1 = t_1/(t_1 + t_2)$ and $x = t_1 + t_2$, equation (8) becomes the usual form,

$$\bar{F}(t_1)\bar{F}(t_2) = \bar{F}(t_1 + t_2),$$

for $0 \le t_1, t_2 < \infty$. Therefore F is exponential. This proves the theorem. □

REMARK. Theorem 1 may be stated alternatively as follows. If $n \ge 5$, X_1 has an exponential distribution if and only if $W_{5,n} \sim U_{(\cdot)}(4)$. The general case is an easy consequence of this. This statement would emphasize the fact that the sample size $n = 5$ would be large enough to characterize the exponentiality of the sample.

A distribution function F of a nonnegative random variable X is said to be *new better than used* (NBU) if

$$\bar{F}(x + y) \le \bar{F}(x)\bar{F}(y) \quad \text{for all } x \ge 0, y \ge 0.$$

It is said to be *new worse than used* (NWU) if the inequality is reversed.

THEOREM 2. *Suppose that the cdf F of X_1 is either NBU or NWU and that $\{F^{-1}(u): 0 < u < 1\} = (0, \infty)$. Let r for $2 \le r \le n$, $n \ge 3$, be fixed. Then X_1 is exponentially distributed if and only if $W_{r,n} \sim U_{(\cdot)}(r - 1)$.*

PROOF. Similar to Theorem 1, it suffices to show that $W_{r,n} \sim U_{(\cdot)}(r - 1)$ implies that X_1 is exponentially distributed. By Lemma 3.5 of Leslie and van Eeden (1993), $W_{r,n} \sim U_{(\cdot)}(r - 1)$ implies that $W_{r-k,n-k} \sim U_{(\cdot)}(r - k - 1)$ for any $0 \le k \le r - 2$. Thus it suffices to show that the theorem holds for $r = 2$. In this case, $S_{1,n}/S_{2,n}$ has the uniform distribution, $U(0, 1)$. Let $G^*(z) = 1 - \bar{F}^{n-1}(z)$. Applying Lemma 3.7 of Leslie and van Eeden (1993), we obtain

$$\int_0^\infty \bar{F}(sz)\, dG^*(z) = \frac{n - 1}{n - 1 + s}, \tag{9}$$

where $0 \le s \le 1$. On the other hand, it is always true that

$$\int_0^\infty \bar{F}^s(z)\, dG^*(z) = \frac{n - 1}{n - 1 + s}. \tag{10}$$

If F is NBU, then $\bar{F}^{1/k}(z) \le \bar{F}(z/k)$ for every positive integer k and all $z \in [0, \infty)$. Setting $s = 1/k$ in both (9) and (10) and subtracting (10) from (9), we obtain

$$\begin{aligned}
&\int_0^\infty \left[\bar{F}\left(\frac{z}{k}\right) - \bar{F}^{1/k}(z)\right] dG^*(z) \\
&\quad = (n - 1)\int_0^1 \left[\bar{F}\left(\frac{F^{-1}(u)}{k}\right) - \bar{F}^{1/k}(F^{-1}(u))\right](1 - u)^{n-2}\, du \\
&\quad = 0.
\end{aligned}$$

Since $\bar{F}(F^{-1}(u)/k) - \bar{F}^{1/k}(F^{-1}(u)) \geq 0$, $\bar{F}(F^{-1}(u)/k) = \bar{F}^{1/k}(F^{-1}(u))$ almost everywhere with respect to Lebesgue measure on $(0,1)$ and for all positive integers $k \geq 1$; that is, $\bar{F}^{1/k}(z) = \bar{F}(z/k)$ for all $z \in [0,\infty)$ and for all integers $k \geq 1$. The result follows immediately from that of Desu (1971).

The same proof applies to an F which is NWU; we only need to reverse the subtraction, that is, to subtract (9) from (10). □

REMARK. The result of Ahsanullah (1977) could be applied to prove Theorem 2. However, there seems to be a gap in his proof; namely, it does not follow directly from his equation (2.5) that F is an exponential distribution. Equation (2.5) only implies that the solution

$$\bar{F}(x) = \exp(-xk^{C(\log_k(x))}),$$

where $C(\cdot)$ is an arbitrary periodic function with period 1. See Azlarov and Volodin [(1986), Lemma 8.1].

Acknowledgments. The authors wish to thank the referees and an Associate Editor for their careful reading and helpful suggestions.

REFERENCES

AHSANULLAH, M. (1977). A characterization property of the exponential distribution. *Ann. Statist.* **5** 580–582.

AZLAROV, T. A. and VOLODIN, N. A. (1986). *Characterization Problems Associated with the Exponential Distribution.* Springer, New York.

DESU, M. M. (1971). A characterization of the exponential distribution by order statistics. *Ann. Math. Statist.* **42** 837–838.

DUFOUR, R. (1982). Tests d'adjustement pour des échantillons tronqués censurés. Ph.D. dissertation, Univ. Montréal.

DUFOUR, R., MAAG, U. R. and VAN EEDEN, C. (1984). Correcting a proof of a characterization of the exponential distribution. *J. Roy. Statist. Soc. Ser. B* **46** 238–241.

LESLIE, J. and VAN EEDEN, C. (1993). On a characterization of the exponential distribution based on a type 2 right censored sample. *Ann. Statist.* **21** 1640–1647.

LIN, G. W. (1990). Characterizations of continuous distributions via expected values of two functions of order statistics. *Sankhyā Ser. A* **52** 84–90.

MENON, M. V. and SESHADRI, V. (1975). A characterization theorem useful in hypothesis testing. In *Contributed Papers, 40th Session of the International Statistical Institute* 586–590. Internat. Statist Inst., Voorburg.

SESHADRI, V., CSÖRGŐ, M. and STEPHENS, M. A. (1969). Tests for the exponential distribution using Kolmogorov-type statistics. *J. Roy. Statist. Soc. Ser. B* **31** 499–509.

BIOMETRY BRANCH
NATIONAL CANCER INSTITUTE
BETHESDA, MARYLAND 20892

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF MARYLAND
COLLEGE PARK, MARYLAND 20742