The Annals of Statistics
1993, Vol. 21, No. 3, 1640–1647

ON A CHARACTERIZATION OF THE EXPONENTIAL DISTRIBUTION BASED ON A TYPE 2 RIGHT CENSORED SAMPLE[1]

BY JULIAN LESLIE[2] AND CONSTANCE VAN EEDEN

Macquarie University, and
University of British Columbia and Université du Québec à Montréal

Dufour gives a conjecture concerning a characterization of the exponential distribution based on type 2 right censored samples. This conjecture, if true, generalizes the characterization based on complete samples of Seshadri, Csörgő and Stephens (1969) and Dufour, Maag and van Eeden (1984). In this paper it is shown that Dufour's conjecture is true if the number of censored observations is no larger than $(1/3)n - 1$, where n is the sample size. The result has implications for testing fit of censored data to the exponential distribution.

1. Introduction. Dufour, in his 1982 Ph.D. dissertation, presents the following conjecture concerning a characterization of the exponential distribution. Let $X_1, X_2, \dots, X_n$ be independent, identically distributed, nonnegative random variables and let r be an integer satisfying $2 \le r < n$. Write $Y_{1,n} \le Y_{2,n} \le \cdots \le Y_{n,n}$ for the order statistics of $X_1, X_2, \dots, X_n$, with $Y_{0,n} = 0$ and define

$$\begin{aligned} D_{i,n} &= (n-i+1)(Y_{i,n} - Y_{i-1,n}) \qquad i = 1, 2, \dots, n, \\ S_{i,n} &= \sum_{j=1}^{i} D_{j,n} \qquad i = 1, 2, \dots, n. \end{aligned} \tag{1.1}$$

The conjecture then states that, if

$$W_{r,n} = \left(\frac{S_{1,n}}{S_{r,n}}, \frac{S_{2,n}}{S_{r,n}}, \dots, \frac{S_{r-1,n}}{S_{r,n}} \right) \tag{1.2}$$

is distributed as the vector of order statistics of a sample of size $r - 1$ from a $U(0,1)$-distribution [i.e., a uniform distribution on the interval $(0, 1)$], then X_1 has an exponential distribution. That this result, if true, characterizes the exponential distribution follows from the fact that $W_{r,n}$ has this uniform-order-statistics distribution when X_1 is exponentially distributed.

For the uncensored case $(n = r)$ a proof of the characterization for the case where $n \ge 3$ was given by Dufour, Maag and van Eeden (1984) [see also Seshadri, Csörgő and Stephens (1969)]; Menon and Seshadri (1975) show that,

Received June 1991; revised June 1992.

[1]Research partially supported by the Natural Sciences and Engineering Research Council of Canada, by F.C.A.R. of Québec and by Birkbeck College, London.

[2]The major component of this author's contribution was carried out while at Birkbeck College, University of London, U.K.

AMS 1991 *subject classifications*. Primary 62E10; secondary 62E15.
Key words and phrases. Characterization, exponential distribution, goodness-of-fit.

for the case $n = 2$, the uniform distribution of $W_{2,2}$ does not characterize the exponential distribution.

The problem of whether the uniform-order-statistics distribution of $W_{r,n}$ characterizes the exponential distribution arises, for example, when one bases a test of the hypothesis that X_1 is exponentially distributed on the statistic $W_{r,n}$. If the conjecture is true, then the hypothesis that X_1 is exponential is equivalent to the hypothesis that $W_{r,n}$ is distributed as the vector of order statistics of a sample of size $r - 1$ from a $U(0,1)$-distribution. There are of course numerous tests of uniformity available and this result offers a natural way of testing the exponentiality of a sequence of ordered observations before the complete set of observations has been collected. However, if the conjecture is false, then there exists at least one alternative distribution for X_1 for which the power of the test equals the size.

In this paper it will be shown that Dufour's conjecture is true if $r \geq (2/3)n + 1$. We do not know whether it is true for the case where $n \geq 3$ and $r < (2/3)n + 1$.

Section 2 contains the main result and its proof. Some lemmas, needed for the proof in Section 2, are given in Section 3. The vector of order statistics of a sample of size j from a $U(0,1)$-distribution will be denoted by $U_{(\cdot)}(j) = (U_{(1)}, U_{(2)}, \dots, U_{(j)})$ and $Z_1 \sim Z_2$ will be used to denote that the random vectors Z_1 and Z_2 have the same distribution. Finally, $\bar{F}(x) = 1 - F(x)$, $-\infty < x < \infty$, where F is the distribution function of X_1.

Under the assumption that the density of X_1 exists, the proof of the main theorem is relatively straightforward. In avoiding that assumption the proof has had to become somewhat more complicated.

2. The main result. This section contains the proof of the following theorem.

THEOREM 2.1. *Let $r \geq (2/3)n + 1$. Then X_1 is exponentially distributed if and only if*

$$W_{r,n} \sim U_{(\cdot)}(r-1).$$

PROOF. As already mentioned above, $W_{r,n} \sim U_{(\cdot)}(r-1)$ when X_1 is exponentially distributed. We show that the reverse also holds, namely

$$W_{r,n} \sim U_{(\cdot)}(r-1) \quad \Rightarrow \quad X_1 \text{ is exponentially distributed.} \tag{2.1}$$

By Lemma 3.9, $W_{r,n} \sim U_{(\cdot)}(r-1)$ implies

$$P\left(\frac{Z_1}{Z} > s_1, \frac{Z_2}{Z} > s_2\right) = \frac{1}{1 + s_1 + s_2}, \qquad s_1, s_2 \geq 0, \tag{2.2}$$

where Z, Z_1 and Z_2 are independent random variables each distributed as

$$\min(X_1, X_2, \dots, X_{n-r+1}).$$

We apply the following key theorem from Kotlarski (1967).

THEOREM A (Kotlarski). *Let W_1, W_2, W_3 be independent positive random variables, and set $Y_1 = W_1/W_3$, $Y_2 = W_2/W_3$. The necessary and sufficient condition for W_k, $k = 1, 2, 3$, to be identically and exponentially distributed is that the joint distribution of (Y_1, Y_2) has density*

$$g(y_1, y_2) = 2(1 + y_1 + y_2)^{-3}, \qquad y_1 > 0, y_2 > 0.$$

From (2.2) and Theorem A it follows that Z_1 has an exponential distribution and from

$$P(Z_1 > t) = \bar{F}^{n-r+1}(t) \qquad t \geq 0 \tag{2.3}$$

it then follows that X_1 has an exponential distribution. □

REMARK 2.2. Lemma 3.9 is essentially Lemma 3.8 extended to include s_1 and/or s_2 greater than 1. The argument justifying this extension only works so long as the number of X's in each minimum term, and in the definition of the distribution of Z, are equal. In order for this to be true it is necessary that $r \geq (2/3)n + 1$. Although Lemma 3.7 tells us more about the joint distributions of the ratios X_i/Z than does Lemma 3.8, where we only look at two minima among these ratios, we do not see how to exploit this extra information to obtain a more general result than the one we have.

3. Some lemmas. In this section we develop a sequence of lemmas which lead ultimately to a proof of Lemma 3.9.

LEMMA 3.1 [Dufour (1982), pages 146–151]. *For $2 \leq j \leq n$, let*

$$V_{1,j,n} = Y_{1,n}/Y_{j,n}, \qquad V_{2,j,n} = Y_{2,n}/Y_{j,n}, \dots, V_{j-1,j,n} = Y_{j-1,n}/Y_{j,n}$$

then $W_{j,n} \sim U_{(\cdot)}(j-1)$ if and only if $(V_{1,j,n}, V_{2,j,n}, \dots, V_{j-1,j,n})$ has density

$$\begin{cases} \dfrac{n!(j-1)!}{(n-j)!}\left(\sum_{i=1}^{j-1} v_i + n - j + 1\right)^{-j}, & 0 < v_1 \leq v_2 \leq \cdots \leq v_{j-1} < 1 \\ 0, & \textit{otherwise}. \end{cases} \tag{3.1}$$

REMARK 3.2. Note that if $(V_{1,j,n}, V_{2,j,n}, \dots, V_{j-1,j,n})$ has a density, then F must be continuous. For if F has a jump at x_0 (say) of size p, then $(V_{1,j,n}, V_{2,j,n}, \dots, V_{j-1,j,n})$ can take the value $(1, 1, \dots, 1)$ with probability at least p^n.

For $2 \leq j \leq n$ let $X_{1,j,n}, X_{2,j,n}, \dots, X_{j-1,j,n}$ be random variables which are, conditionally on $Y_{j,n} = y$, independent and identically distributed with distribution function $F(x)/F(y)$, $0 < x \leq y$. (Note that the $X_{i,j,n}$ should, in fact, carry an extra subscript, namely $Y_{j,n}$).

Let

$$T_{i,j,n} = X_{i,j,n}/Y_{j,n}, \qquad i = 1, 2, \dots, j-1$$

and set

$$T_{j,n} = \left(T_{1,j,n}, T_{2,j,n}, \dots, T_{j-1,j,n}\right).$$

The following is an immediate consequence of Lemma 3.1.

LEMMA 3.3. *For* $2 \le j \le n$, $W_{j,n} \sim U_{(\cdot)}(j-1)$ *if and only if* $T_{j,n}$ *has density*

$$(3.2)\quad [n!/(n-j)!]\left(\sum_{i=1}^{j-1} t_i + n - j + 1\right)^{-j}, \qquad 0 < t_i < 1, i = 1, 2, \dots, j-1.$$

We can further simplify the joint distribution of $T_{j,n}$ by representing it in terms of the original independent X_i's.

LEMMA 3.4. *For* $2 \le j \le n$ *and* $\mathbf{t} = (t_1, t_2, \dots, t_{j-1})$, $\mathbf{t} \in [0,1]^{j-1}$,

$$P(T_{j,n} \le \mathbf{t}) = \binom{n}{j-1} P\left(X_1 \le t_1 Z, X_2 \le t_2 Z, \dots, X_{j-1} \le t_{j-1} Z\right),$$

where Z *is independent of* $X_1, X_2, \dots, X_{j-1}$ *and distributed as*

$$\min\left(X_j, X_{j+1}, \dots, X_n\right).$$

PROOF. From the definition of $T_{j,n}$,

$$\begin{aligned}
P\{T_{j,n} \le \mathbf{t}\}
&= P\{X_{1,j,n} \le t_1 Y_{j,n}, X_{2,j,n} \le t_2 Y_{j,n}, \dots, X_{j-1,j,n} \le t_{j-1} Y_{j,n}\} \\
&= \int_0^\infty \prod_{i=1}^{j-1} \left[\frac{F(xt_i)}{F(x)}\right] \frac{n!}{(j-1)!(n-j)!} [F(x)]^{j-1} [\bar{F}(x)]^{n-j} \, dF(x) \\
&= \binom{n}{j-1} \int_0^\infty \left[\prod_{i=1}^{j-1} F(xt_i)\right] (n-j+1) [\bar{F}(x)]^{n-j} \, dF(x).
\end{aligned}$$

The second equality follows from the fact that, for $j = 1, 2, \dots, n$,

$$F_{Y_{j,n}}(x) = \binom{n}{j-1}(n-j+1)\int_0^{F(x)} t^{j-1}(1-t)^{n-j}\, dt$$

and that F is uniformly continuous, so that for any measurable function h,

$$(3.3)\quad \begin{aligned}
&\int h \, dF_{Y_{j,n}}(x) \\
&\quad = \int h \cdot \binom{n}{j-1}(n-j+1)[F(x)]^{j-1}[\bar{F}(x)]^{n-j} \, dF(x).
\end{aligned}$$

The lemma now follows. □

The next result will be useful in an induction argument needed later on.

LEMMA 3.5. *For* $0 \le k \le r - 2 \le n - 2$,

$$\text{(3.4)} \qquad \text{if } W_{r,n} \sim U_{(\cdot)}(r-1), \quad \text{then } W_{r-k,n-k} \sim U_{(\cdot)}(r-k-1).$$

PROOF. The result is trivial when $r = 2$, so assume $r > 2$.

Using standard properties of uniform order statistics we have that for $2 \le j \le r$,

$$\text{(3.5)} \qquad W_{r,n} \sim U_{(\cdot)}(r-1) \quad \Rightarrow \quad W_{j,n} \sim U_{(\cdot)}(j-1).$$

Further we have that for any $A \in \mathscr{B}_{j-2}$, the σ-algebra of Borel sets of $[0,1]^{j-2}$,

$$\text{(3.6)} \qquad \begin{aligned} P(T_{j-1,n-1} \in A) &= (n-j+1)n^{-1}P(T_{j-1,n} \in A) \\ &\quad + (j-1)n^{-1}P\big(T_{j,n} \in A \times [0,1]\big). \end{aligned}$$

This can be seen as follows. From Lemma 3.4 one obtains

$$\begin{aligned} P(T_{j-1,n-1} \in A) &= \binom{n-1}{j-2} P\big((X_1, X_2, \dots, X_{j-2}) \in AZ,\ X_{j-1} > Z\big) \\ &\quad + \binom{n-1}{j-2} P\big((X_1, X_2, \dots, X_{j-2}) \in AZ,\ X_{j-1} \le Z\big), \end{aligned}$$

where $Z \sim \min(X_j, X_{j+1}, \dots, X_n)$ and independent of $(X_1, X_2, \dots, X_{j-1})$. Using (3.3) and noting that $Z \sim Y_{1,n-j+1}$

$$\begin{aligned} &P\big((X_1, X_2, \dots, X_{j-2}) \in AZ,\ X_{j-1} > Z\big) \\ &\quad = E_Z\big\{\big[P\big((X_1, X_2, \dots, X_{j-2}) \in AZ|Z\big)\big]\bar{F}(Z)\big\} \\ &\quad = \{(n-j+1)/(n-j+2)\}P\big((X_1, X_2, \dots, X_{j-2}) \in AZ^*\big), \end{aligned}$$

where $Z^* \sim Y_{1,n-j+2}$ and is independent of $(X_1, X_2, \dots, X_{j-2})$, and E_Z indicates expectation with respect to Z. A second application of Lemma 3.4 yields the first term on the right of (3.6). The second term follows since

$$\binom{n-1}{j-2} = \left(\frac{j-1}{n}\right)\binom{n}{j-1} \text{ and}$$

$$\begin{aligned} &P\big((X_1, X_2, \dots, X_{j-2}) \in AZ,\ X_{j-1} \le Z\big) \\ &\quad = P\big((X_1, X_2, \dots, X_{j-1}) \in (A \times [0,1])Z\big). \end{aligned}$$

From (3.5) with $j = r - 1$, (3.6) with $j = r$ and Lemma 3.3 with $j = r$ and with $j = r - 1$, we have that $W_{r,n} \sim U_{(\cdot)}(r-1)$ implies that for all $A \in \mathscr{B}_{r-2}$

and $\mathbf{t} = (t_1, t_2, \dots, t_{r-2})$,

$$
\begin{aligned}
P(T_{r-1,n-1} \in A) & \\
&= \frac{n-r+1}{n} \int \cdots \int_{\mathbf{t} \in A} \frac{n!}{(n-r+1)!} \left(\sum_{i=1}^{r-2} t_i + n - r + 2 \right)^{-r+1} \\
&\qquad \times dt_1 \cdots dt_{r-2} \\
&\quad + \frac{r-1}{n} \int \cdots \int_{\mathbf{t} \in A} \int_{t_{r-1} \in [0,1]} \frac{n!}{(n-r)!} \left(\sum_{i=1}^{r-1} t_i + n - r + 1 \right)^{-r} \\
&\qquad \times dt_1 \cdots dt_{r-1} \\
&= \int \cdots \int_{\mathbf{t} \in A} \frac{(n-1)!}{(n-r)!} \left(\sum_{i=1}^{r-2} t_i + n - r + 1 \right)^{-r+1} dt_1 \cdots dt_{r-2}.
\end{aligned}
\tag{3.7}
$$

From Lemma 3.3 with $j = r - 1$ and n replaced by $n - 1$, we have that $W_{r-1,n-1} \sim U_{(\cdot)}(r-2)$. This establishes the case $k = 1$. The other cases follow by repeated application of the $k = 1$ case. □

From Lemma 3.3 and Lemma 3.5 one immediately obtains the following:

LEMMA 3.6. *For* $2 \le j \le r \le n$, $W_{r,n} \sim U_{(\cdot)}(r-1)$ *implies that* $T_{j,n-r+j}$ *has density*

$$
\frac{(n-r+j)!}{(n-r)!} \left(\sum_{i=1}^{j-1} t_i + n - r + 1 \right)^{-j}, \qquad 0 < t_i < 1,\ i = 1, 2, \dots, j-1. \tag{3.8}
$$

The following is an immediate consequence of Lemmas 3.4 and 3.6.

LEMMA 3.7. *For* $2 \le j \le r \le n$, $W_{r,n} \sim U_{(\cdot)}(r-1)$ *implies*

$$
\begin{aligned}
&P(X_1 \le s_1 Z, X_2 \le s_2 Z, \dots, X_{j-1} \le s_{j-1} Z) \\
&\quad = (n-r+1)(j-1)! \int_0^{s_1} \cdots \int_0^{s_{j-1}} \left(\sum_{i=1}^{j-1} t_i + n - r + 1 \right)^{-j} dt_1 \cdots dt_{j-1}, \\
&\qquad\qquad 0 \le s_i \le 1,\ i = 1, 2, \dots, j-1,
\end{aligned}
$$

where $Z \sim \min(X_1, X_2, \dots, X_{n-r+1})$ *and independent of* $(X_1, X_2, \dots, X_{j-1})$.

We assume from now on that

$$
Z \sim \min(X_1, X_2, \dots, X_{n-r+1}) \quad \text{independently of } (X_1, X_2, \dots, X_{j-1}).
$$

We are in a position to state the key lemma on which Lemma 3.9 depends and hence on which the main result of the paper rests.

LEMMA 3.8. *For* $2 \le j \le r \le n$, $0 \le l \le j-2$,

$$P\Big(\min_{1\le i\le l} X_i > s_1 Z, \min_{l+1\le i\le j-1} X_i > s_2 Z\Big) = \frac{n-r+1}{n-r+1+ls_1+(j-l-1)s_2}, \qquad 0 \le s_i \le 1, i = 1,2. \tag{3.9}$$

PROOF. Using induction arguments [the full details of which can be found in Leslie and van Eeden (1991) which is available from the authors on request] expressions for the following probabilities can be derived in sequence. The expressions involve integrals having a form similar to the one in Lemma 3.7.

(i) For $2 \le j \le r \le n$ and $0 \le l \le j-1$, with $s, s_i \in [0,1]$, $i = 1, 2, \ldots, j-l-1$,

$$P\Big(X_i \le s_i Z, i = 1,2,\ldots, j-l-1, \min_{j-l\le i\le j-1} X_i > sZ\Big).$$

(ii) For $3 \le j \le r \le n$, $l \ge 0$, $k \ge 0$, $l+k \le j-2$, $\alpha = j-l-k$ and $s, s', s_i \in [0,1]$, $i = 1,2,\ldots,\alpha-1$,

$$P\Big(X_i \le s_i Z, i = 1,2,\ldots,\alpha-1, \min_{\alpha\le i\le j-l-1} X_i > sZ, \min_{j-l\le i\le j-1} X_i > s'Z\Big).$$

(iii) For $3 \le j \le r \le n$, $0 \le l \le j-2$ and $0 \le s_i \le 1$, $i = 1, 2$,

$$P\Big(X_1 \le s_1 Z, \min_{2\le i\le l+1} X_i > s_2 Z, \min_{l+2\le i\le j-1} X_i > s_1 Z\Big).$$

Using the expression for (iii) we derive the result of the lemma. □

LEMMA 3.9. *If* $r \ge (2/3)n + 1$, *then* $W_{r,n} \sim U_{(\cdot)}(r-1)$ *implies*

$$P(Z_1 > s_1 Z, Z_2 > s_2 Z) = (s_1 + s_2 + 1)^{-1}, \qquad s_1, s_2 \ge 0,$$

where Z_1, Z_2, Z *are independent and identically distributed as*

$$\min(X_1, X_2, \ldots, X_{n-r+1}).$$

PROOF. In Lemma 3.8 we make sure that the *number* of variables involved in the two minimum terms as well as in the minimum expression defining the distribution of Z, are all equal. We thus set $l = j-l-1 = n-r+1$. Clearly $j = 2l+1$ and as $j \le r$ so $2l+1 \le r$. But l also equals $n-r+1$ so we need $2(n-r+1)+1 \le r$. This is equivalent to $r \ge (2/3)n+1$. Lemma 3.8 then states that

$$P(Z_1 > s_1 Z, Z_2 > s_2 Z) = (s_1 + s_2 + 1)^{-1}, \qquad 0 \le s_i \le 1, i = 1,2, \tag{3.10}$$

where Z, Z_1, Z_2 are i.i.d. as $\min(X_1, X_2, \ldots, X_{n-r+1})$.

We now show that (3.10) holds for all $s_1, s_2 \geq 0$. Let $V_1 = Z_1/Z$, $V_2 = Z_2/Z$ and let $g_{V_1, V_2}(v_1, v_2)$ denote the density of (V_1, V_2) on $[0, 1]^2$. Then, by (3.10),

$$g_{V_1, V_2}(v_1, v_2) = 2(1 + v_1 + v_2)^{-3}, \quad 0 \leq v_i \leq 1, i = 1, 2. \tag{3.11}$$

Now set $W_1 = V_1^{-1}$, $W_2 = V_2/V_1$, then $W_1 = Z/Z_1$, $W_2 = Z_2/Z_1$ and from (3.11) it follows that the density of (W_1, W_2) is

$$\begin{aligned} h_{W_1 W_2}(w_1, w_2) &= 2w_1^{-3}\left(1 + w_1^{-1} + w_2 w_1^{-1}\right)^{-3} \\ &= 2(1 + w_1 + w_2)^{-3}, \qquad w_1 > 1, 0 < w_2 < w_1. \end{aligned}$$

The lemma then follows from the fact that W_1 and W_2 are exchangeable and that V_1, V_2 have the same distribution. □

Acknowledgments. The authors wish to thank a referee who made a number of valuable suggestions. J. Leslie would also like to thank M. A. Stephens for having drawn his attention to the problem considered in the paper.

REFERENCES

DUFOUR, R. (1982). Tests d'ajustement pour des échantillons tronqués ou censurés, Ph.D. dissertation, 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.

KOTLARSKI, I. (1967). On characterizing the gamma and the normal distribution. *Pacific J. Math.* **20** 69–76.

LESLIE, J. R. and VAN EEDEN, C. (1991). On a characterization of the exponential distribution based on a type 2 right censored sample. Technical Report 107, Dept. Statistics, Univ. British Columbia.

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. ISI, 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.

SCHOOL OF ECONOMIC
AND FINANCIAL STUDIES
MACQUARIE UNIVERSITY
SYDNEY 2109
AUSTRALIA

MOERLAND 19
1151 BH
BROEK IN WATERLAND
THE NETHERLANDS