# ON THE DISTRIBUTION OF A STATISTIC BASED ON ORDERED UNIFORM CHANCE VARIABLES

By Shanti S. Gupta[1] and Milton Sobel

*Bell Telephone Laboratories*

**1. Summary.** The exact distribution of a statistic based on the $r$ smallest of $n$ independent observations from a unit uniform distribution is derived. In life-testing terminology, this statistic includes as special cases (i) the sum of the $r$ earliest failure times, (ii) the total observed life up to the $r$th failure, and (iii) the sum of all $n$ failure times. The density, cumulative distribution function (c.d.f.) and first four moments of the general statistic are summarized in Sec. 2. Section 3 gives the derivation of the density and c.d.f. The moments are obtained from the moment generating function in Sec. 4. Asymptotic normality under certain conditions is proved in Sec. 5 and illustrations of the rapidity of approach to normality are given in Sec. 6.

**2. Introduction and statement of results.** We shall consider the statistic

$$T_{r,m}^{(n)} = t_1 + t_2 + \cdots + t_r + (m - r)t_r, \tag{2.1}$$

where $t_i = t_i^{(n)}$ is the $i$th smallest of $n$ independent observations and $m$ is greater than $r - 1$ but is not necessarily an integer. For $m = n$ this statistic can be interpreted as the total observed life in a life-testing experiment without replacement. When the underlying distribution of the unordered $t$'s is exponential, i.e., $f(t) = (1/\theta)e^{-t/\theta}$, then it is known [3] that $2T_{r,n}^{(n)}/\theta$ is distributed as chi-square $\chi_{2r}^2$ with $2r$ degrees of freedom.

Before stating further results let us introduce for $0 \leqq t \leqq m$ and non-negative integers $p$, $q$, $n$

$$A_{p,m}^{(q,n)}(t) = \frac{n}{p!}\left\{\binom{p}{0}\frac{t^{n-1}}{m^{q-p}} - \binom{p}{1}\frac{(t-1)^{n-1}}{(m-1)^{q-p}} + \binom{p}{2}\frac{(t-2)^{n-1}}{(m-2)^{q-p}} - \cdots\right\}, \tag{2.2}$$

where $m > p$, $n \geqq 1$ and the summation is continued as long as the arguments $t, t-1, t-2, \cdots$ are positive. It is understood that the binomial coefficient $\binom{p}{j} = 0$ for $p < j$ so that there are at most $(p + 1)$ terms in the above summation.

It is clear from (2.1) that $T_{n,n}^{(n)}$ is the sum of all the $n$ observations. When the underlying distribution is unit uniform, then the density of $T_{n,n}^{(n)}$ is given on p. 246 of [2] by

$$f_{n,n}^{(n)}(t) = \frac{1}{(n-1)!}\left\{\binom{n}{0}t^{n-1} - \binom{n}{1}(t-1)^{n-1} + \cdots\right\} = A_{n,m}^{(n,n)}(t). \tag{2.3}$$

Received March 5, 1957, revised May 16, 1957.

[1] Now at the University of Alberta.

(We have removed the superscripts and subscripts from the chance variables and put them on $f$ and on $F$ below which are the symbols for the density and c.d.f., respectively.)

Using the symmetry of the above density about $t = n/2$, we can replace $t$ by $n - t$ in (2.3) obtaining

$$f_{n,n}^{(n)}(t) = \frac{n}{(n-1)!}\left\{\binom{n-1}{0}\frac{(n-t)^{n-1}}{n} - \binom{n-1}{1}\frac{(n-1-t)^{n-1}}{(n-1)} + \cdots\right\} = A_{n-1,n}^{(n,n)}(n-t), \tag{2.4}$$

where $0 \leqq t \leqq n$. The form (2.4) is more comparable with the results derived here. It is shown below that the density and c.d.f. of $T_{r,m}^{(n)}$ are given by the comparable results

$$f_{r,m}^{(n)}(t) = A_{r-1,m}^{(n,n)}(m - t) \tag{2.5}$$

and

$$F_{r,m}^{(n)}(t) = 1 - \frac{1}{(n+1)} A_{r-1,m}^{(n,n+1)}(m - t), \tag{2.6}$$

from which we get as special cases the densities and c.d.f.'s of (i) $T_{r,r}^{(n)}$, (ii) $T_{r,n}^{(n}$ and (iii) $T_{n,n}^{(n)}$.

Barton and David [1] have derived another equivalent formula for the density $f_{r,r}^{(n)}(t)$, i.e., in the special case (i). Their result, with two typographical corrections taken into account, is

$$f_{r,r}^{(n)}(t) = \frac{n}{r!}\sum_{i=1}^{r}(-1)^{r-i}i^{r-n}\binom{r}{i}\left[\frac{i - t + |i - t|}{2}\right]^{n-1}. \tag{2.7}$$

The total life statistic arises as an optimum statistic under exponential distribution assumptions in [3]. In the present paper we give the distribution of this statistic when the exponential distribution assumption is replaced by the uniform distribution. Hence these results can be used to study the robustness of the tests based on the total life statistic. The results on asymptotic normality are also of interest in this connection since under the exponential assumption the distribution of $2T_{r,n}^{(n)}/\theta$ is that of $\chi_{2r}^2$ which, for large $r$, also is close to that of a normal distribution. It is felt that the model of a uniform distribution from 0 to $\theta$, $\theta > 0$ and unknown, and the results of this paper may prove to be useful in some life-testing problems.

**3. Derivation of results.** Let $u = t_1 + t_2 + \cdots + t_{r-1}$, $v = t_r$, $w = u/v$ and $y = T_{r,m}^{(n)} = u + (m - r + 1)v$, where $t_i = t_i^{(n)}$ is the $i$th smallest of $n$ independent chance variables uniformly distributed from zero to one. The conditional distribution of $w$ given $v$ is exactly that of a sum of $r - 1$ independent uniform chance variables and is given by (2.3) with $n$ replaced by $r - 1$. Hence the joint density of $v$ and $w$ is given by

$$g(w, v) = \frac{n!}{(r-1)!(n-r)!} v^{r-1}(1-v)^{n-r} \frac{1}{(r-2)!} \cdot \left\{\binom{r-1}{0} w^{r-2} - \binom{r-1}{1}(w-1)^{r-2} + \cdots\right\}, \tag{3.1}$$

where $0 \leqq w \leqq r - 1$ and $0 < v \leqq 1$ and the joint density of $u$ and $v$ is given by

$$h(u, v) = \frac{n!}{(r-2)!(r-1)!(n-r)!} \cdot \left\{\binom{r-1}{0} u^{r-2} - \binom{r-1}{1}(u-v)^{r-2} + \cdots\right\}(1-v)^{n-r}, \tag{3.2}$$

where $0 \leqq u \leqq (r-1)v$ and $v \leqq 1$.

If we now derive the density of $y$, then the full range of $y$ from 0 to $m$ is broken into $r$ parts. For $0 \leqq y \leqq m - r + 1$, the density of $y$ becomes

$$\begin{aligned} f_{r,m}^{(n)}(y) = {} & \frac{n!}{(r-2)!(r-1)!(n-r)!} \\ & \cdot \int_{y/m}^{y/(m-r+1)} [y-(m-r+1)v]^{r-2}(1-v)^{n-r}\,dv \\ & - \binom{r-1}{1} \int_{y/m}^{y/(m-r+2)} [y-(m-r+2)v]^{r-2}(1-v)^{n-r}\,dv \\ & + \cdots + (-1)^{r-2}\binom{r-1}{r-2} \int_{y/m}^{y/(m-1)} y-(m-1)v]^{r-2}(1-v)^{n-r}\,dv. \end{aligned} \tag{3.3}$$

Using the finite difference operators $\mathcal{E}$, $\Delta$ (with $\mathcal{E} = 1 + \Delta$),

$$\begin{aligned} f_{r,m}^{(n)}(y) = {} & \frac{r}{(r-2)!}\binom{n}{r} \sum_{\alpha=0}^{r-1} \binom{r-1}{\alpha} \\ & \cdot (-\mathcal{E})^{\alpha} \left\{ \int_{y/m}^{y/(m-r+1+x)} [y-(m-r+1+x)v]^{r-2}(1-v)^{n-r}\,dv \right\}, \end{aligned} \tag{3.4}$$

where $\mathcal{E}$ operates on $x$ and it is understood that $x$ is then to be set equal to 0. Using the relation between $\mathcal{E}$ and $\Delta$,

$$\begin{aligned} f_{r,m}^{(n)}(y) = {} & \frac{r}{(r-2)!}\binom{n}{r} \\ & \cdot \left[ (-\Delta)^{r-1} \left\{ \int_{y/m}^{y/(m-r+1+x)} [y-(m-r+1+x)v]^{r-2}(1-v)^{n-r}\,dv \right\} \right]_{x=0}. \end{aligned} \tag{3.5}$$

If we now integrate by parts, the first term vanishes at the upper limit and also at the lower limit because of the operator $\Delta^{r-1}$. After $r - 1$ such integrations we obtain

$$f_{r,m}^{(n)}(y) = \frac{n}{(r-1)!} \left[ \Delta^{r-1} \left\{ \frac{(m-r+1+x-y)^{n-1}}{(m-r+1+x)^{n-r+1}} \right\} \right]_{x=0}. \tag{3.6}$$

Using $\Delta = \mathcal{E} - 1$ we obtain

$$f_{r,m}^{(n)}(y)$$

$$(3.7)\qquad = \frac{n}{(r-1)!}\left\{\binom{r-1}{0}\frac{(m-y)^{n-1}}{m^{n-r+1}} - \binom{r-1}{1}\frac{m-1-y)^{n-1}}{(m-1)^{n-r+1}} + \cdots\right\}$$
$$= A_{r-1,m}^{(n,n)}(m-y),$$

where $0 \leqq y \leqq m - r + 1$ and $A_{p,m}^{(q,n)}$ is defined in (2.2).

We shall now show that the expression (3.7) gives the result for all $y(0 \leqq y \leqq m)$. For $m - r + i \leqq y \leqq m - r + 1 + i$ $(i = 1, 2, \cdots, r - 1)$ the only difference is that the first $i$ upper limits of integration in (3.3) are all changed to unity. For the $j$th integral $(j = 1, 2, \cdots, i)$ we have to add to the complete set of $r$ terms in (3.7) the quantity

$$(3.8)\qquad (-1)^{j+1}\binom{r-1}{j-1}\frac{n!}{(r-2)!(r-1)!(n-r)!} \cdot \int_{y/m-r+j)}^{1} [y - (m - r + j)v]^{r-2}(1 - v)^{n-r}\,dv$$
$$= (-1)^{j+r-1}\frac{n}{(r-1)!}\binom{r-1}{j-1}\frac{(m-r+j-y)^{n-1}}{(m-r+j)^{n-r+1}}.$$

For each $j(1 \leqq j \leqq i)$ the quantity on the right in (3.8) cancels the $j$th term from the end of the complete expression with $r$ terms in (3.7). Hence for

$$m - r + i \leqq y \leqq m - r + i + 1$$

the density is given by the first $r - i$ terms of (3.7) which are precisely those terms with positive arguments. This proves that the expression $A_{r-1,m}^{(n,n)}(m - y)$ of (3.7) gives the result for all $y(0 \leqq y \leqq m)$.

The c.d.f. $F_{r,m}^{(n)}(y)$ of $y$ is easily obtained by integrating (3.7) between the limits 0 and $y$ and is given by

$$(3.9)\qquad F_{r,m}^{(n)}(y) = 1 - \frac{1}{(n+1)} A_{r-1,m}^{(n,n+1)}(m - y).$$

**4. Moments of** $y = T_{r,m}^{(n)}$. Using the expression for the density it can be shown that the moment generating function $M_\theta(y)$ of $y = T_{r,m}^{(n)}$ is given by

$$(4.1)\qquad M_\theta(y) = \left[\frac{n}{(r-1)!}\sum_{\alpha=0}^{r-1}(-1)^\alpha\binom{r-1}{\alpha} \cdot \mathcal{E}^\alpha\left\{\frac{1}{(m-x)^{n-r+1}}\int_0^{m-x} e^{\theta y}(m - x - y)^{n-1}\,dy\right\}\right]_{x=0}$$

$$(4.2)\qquad = \frac{n!}{(r-1)!}\sum_{j=0}^{\infty}\frac{(-\theta)^j}{(j+n)!}[\Delta^{r-1}(x - m)^{j+r-1}]_{x=0}.$$

Thus we have for the $j$th moment

$$E(y^j) = \frac{(-1)^j j!n!}{(r-1)!(n+j)!} [\Delta^{r-1}\{(x-m)^{r-1+j}\}]_{x=0} \tag{4.3}$$

$$= \frac{(-1)^j j!n!}{(r-1)!(n+j)!} \sum_{\beta=0}^{j} (-1)^\beta m^\beta \binom{r-1+j}{\beta} [\Delta^{r-1} x^{r-1+j-\beta}]_{x=0}. \tag{4.4}$$

It can be shown that for $j \geqq 0$ and $r \geqq 1$

$$[\Delta^{r-1} x^{r+j}]_{x=0} = \sum_{\alpha=1}^{r-1} \frac{(r-1)!}{(\alpha-1)!} [\Delta^\alpha x^{\alpha+j}]_{x=0}. \tag{4.5}$$

The results for various values of $j$ in (4.5) are known and are given, for example, in [4], p. 127. Using these we have from (4.4)

$$E(y) = \frac{r(2m-r+1)}{(2n+1)}, \tag{4.6}$$

$$E(y^2) = \frac{r(r+1)}{12(n+1)(n+2)}\left[12m^2 - 12m(r-1) + (r-1)(3r-2)\right], \tag{4.7}$$

$$\sigma^2(y) = \frac{r(n-r+1)(2m-r+1)^2}{4(n+1)^2(n+2)} + \frac{r(r+1)(r-1)}{12(n+1)(n+2)}, \tag{4.8}$$

$$\begin{aligned} E(y^3) &= \frac{r(r+1)(r+2)}{8(n+1)(n+2)(n+3)} \\ &\quad \cdot \left[8m^3 - 12m^2(r-1) + 2m(r-1)(3r-2) - r(r-1)^2\right], \end{aligned} \tag{4.9}$$

$$\begin{aligned} E(y^4) &= \frac{r(r+1)(r+2)(r+3)}{2(n+1)(n+2)(n+3)(n+4)} \\ &\quad \cdot \left[2m^4 - 4m^3(r-1) + m^2(r-1)(3r-2) - m(r-1)^2 r \right. \\ &\qquad \left. + \frac{(r-1)(15r^3 - 15r^2 - 10r + 8)}{120}\right]. \end{aligned} \tag{4.10}$$

Since the computation of cumulants leads to no simplification, they have not been given here; they can be obtained by the usual formulae. It should be mentioned that the above expressions for the moments can also be obtained directly by using the moments of the order statistics.

**5. Asymptotic normality of $y = T_{r,m}^{(n)}$.** We shall randomize the order of the chance variables $t_1, t_2, \cdots, t_{r-1}$ and thus define new unordered *equi-correlated* and identically distributed chance variables $u_1, u_2, \cdots, u_{r-1}$. Furthermore, if we consider the conditional joint distribution of the $u_i$ given $v$ $(=t_r)$, then we have *independent* chance variables which are uniformly distributed from 0 to $v$.

Let

$$y^* = \frac{y - E(y)}{\sigma(y)} \text{ and } y_v^* = \frac{y - E(y \mid v)}{\sigma(y \mid v)}, \tag{5.1}$$

where $y = T_{r,m}^{(n)} = u_1 + u_2 + \cdots + u_{r-1} + (m - r + 1)v$.
The characteristic function of $y^*$ is given by

$$\begin{aligned} \varphi_y^*(t) &= \int_0^1 \int_0^v \cdots \int_0^v \exp\left\{it\left(\frac{y - E(y)}{\sigma(y)}\right)\right\}\left[\prod_{i=1}^{r-1} \frac{du_i}{v}\right] g(v)\,dv \\ &= \int_0^1 \int_0^v \cdots \int_0^v \exp\left\{it\,\frac{\sigma(y \mid v)}{\sigma(y)}\left[\frac{y - E(y \mid v)}{\sigma(y \mid v)}\right]\right. \\ &\quad \left. + it\,\frac{\sigma(v)}{\sigma(y)}\left[\frac{E(y \mid v) - E(y)}{\sigma(v)}\right]\right\}\left[\prod_{i=1}^{r-1} \frac{du_i}{v}\right] g(v)\,dv, \end{aligned} \tag{5.2}$$

where

$$E(y \mid v) = (r - 1)\frac{v}{2} + (m - r + 1)v = (2m - r + 1)\frac{v}{2}, \tag{5.3}$$

$$\sigma(y \mid v) = v\sqrt{\frac{r-1}{12}}; \qquad \sigma(v) = \frac{1}{(n+1)}\sqrt{\frac{r(n - r + 1)}{n + 2}}, \tag{5.4}$$

and

$$g(v) = \frac{n!}{(r - 1)!(n - r)!} v^{r-1}(1 - v)^{n-r}. \tag{5.5}$$

Letting

$$t' = t\,\frac{\sigma(y \mid v)}{\sigma(y)}, \qquad t'' = t\,\frac{\sigma(v)}{\sigma(y)}\left[m - \left(\frac{r - 1}{2}\right)\right], \tag{5.6}$$

and $x_i = u_i/v (i = 1, 2, \cdots, r - 1)$, we obtain

$$\varphi_y^*(t) = \int_0^1 \left[\int_0^1 e^{it'\sqrt{\frac{12}{r-1}}(x-\frac{1}{2})}\,dx\right]^{r-1} e^{it''\left[\frac{v-E(v)}{\sigma(v)}\right]} g(v)\,dv \tag{5.7}$$

$$= \int_0^1 \left[\frac{\sin\left(t'\sqrt{\frac{3}{r-1}}\right)}{t'\sqrt{\frac{3}{r-1}}}\right]^{r-1} e^{it''\left[\frac{v-E(v)}{\sigma(v)}\right]} g(v)\,dv. \tag{5.8}$$

Since for $r = \lambda n$ and $n \to \infty$ we have

$$E(v) = \frac{r}{n+1} \to \lambda \quad \text{and} \quad \sigma(v) = \frac{1}{n+1}\sqrt{\frac{r(n - r + 1)}{n + 2}} = O\left(\frac{1}{\sqrt{n}}\right), \tag{5.9}$$

then we shall write $v = \lambda + O(1/\sqrt{n})$ in the expression (5.6) for $t'$ which is needed for the first part of the integrand in (5.8). For $m = \gamma n$ we obtain the two asymptotic relations

$$\frac{\sigma(y|v)}{\sigma(y)} \cong \frac{v}{\sqrt{3(1-\lambda)(2\gamma-\lambda)^2+\lambda^2}} = \frac{\sqrt{\lambda^2}}{\sqrt{3(1-\lambda)(2\gamma-\lambda)^2+\lambda^2}} + O\left(\frac{1}{\sqrt{n}}\right) \tag{5.10}$$

and

$$\frac{\sigma(v)}{\sigma(y)}\left[m - \left(\frac{r-1}{2}\right)\right] \cong \frac{\sqrt{3(1-\lambda)(2\gamma-\lambda)^2}}{\sqrt{3(1-\lambda)(2\gamma-\lambda)^2+\lambda^2}}, \tag{5.11}$$

so that if we denote the first term in the right hand members of (5.10) and (5.11) by $a$ and $b$ respectively, then $a^2 + b^2 = 1$. Taking the limit in (5.8) as $n \to \infty$ with $r = \lambda n$, $m = \gamma n$ and using the Lebesgue theorem, we can bring the limit operator under the integral sign. Then, using (5.10), we obtain

$$\varphi_y^*(t) \cong \int_0^1 \lim\left[1 - \frac{a^2t^2}{2(r-1)} + O\left(\frac{1}{n^{3/2}}\right)\right]^{r-1} \lim e^{it^*\left[\frac{v-E(v)}{\sigma(v)}\right]} g(v)\, dv \tag{5.12}$$

$$= e^{-\frac{a^2t^2}{2}} \int_0^1 \lim e^{it^*\left[\frac{v-E(v)}{\sigma(v)}\right]} g(v)\, dv. \tag{5.13}$$

Using the same Lebesgue theorem the limit operator can be taken outside the integral sign. Then, using a result on the asymptotic normality of quantiles given on page 369 in Cramér [2], we obtain

$$\varphi_y^*(t) \cong e^{-\frac{a^2t^2}{2}} e^{-\frac{b^2t^2}{2}} = e^{-\frac{t^2}{2}} \tag{5.14}$$

since $a^2 + b^2 = 1$. This proves the asymptotic normality of $y$ for $r = \lambda n$, $m = \gamma n$ ($\gamma$ and $\lambda$ fixed with $0 < \lambda \leqq 1$ and $\lambda \leqq \gamma < \infty$) and $n \to \infty$.

It should be noted that the above proof holds no matter how fast $m$ tends to infinity. If $m/n \to \infty$ then $a = 0$ and $b = 1$ and (5.14) still holds.

**6. Illustration of rapidity of approach to normality.** To illustrate the rapidity of approach to normality of the statistics, we shall use the Edgeworth series expansion

$$F_{r,m}^{(n)}(x) = \{\Phi(x)\} - \left\{\frac{1}{3!}\frac{\mu_3}{\sigma^3}\Phi^{(3)}(x)\right\} + \left\{\frac{1}{4!}\left(\frac{\mu_4}{\sigma^4} - 3\right)\Phi^{(4)}(x) + \frac{10}{6!}\left(\frac{\mu_3}{\sigma^3}\right)^2\Phi^{(6)}(x)\right\} + \cdots, \tag{6.1}$$

where $\Phi(x)$ is the standard normal c.d.f., $\Phi^{(r)}(x)$ is its $r$th derivative and $x$ denotes the standardized variate corresponding to $t$. We wish to compute one, two, and three terms of (6.1) as indicated by the braces for the two special cases of $T_{r,m}^{(n)}$; viz., (i) $m = r$ and (ii) $m = n$. These have been computed for $n = 10$, $r = 5$ and the results are compared in Table I below with the exact values computed from (2.6).

TABLE I

*Comparison of exact probability $P[T_{r.m}^{(n)} \leqq t]$ and Edgeworth approximations*

| Case | $t$ | $x$ | Approximations | | | Exact Probability |
|---|---|---|---|---|---|---|
| | | | 1 term | 2 terms | 3 terms | |
| (i) | 1.5 | 0.26656 | .6051 | .6340 | .6318 | .6327 |
| $r = 5$ | 2.0 | 1.24393 | .8932 | .8851 | .8849 | .8839 |
| $m = 5$ | 2.5 | 2.22131 | .9868 | .9761 | .9780 | .9769 |
| $n = 10$ | 3.0 | 3.19868 | .9993 | .9975 | .9969 | .9973 |
| (ii) | 4.0 | 0.30754 | .6208 | .6312 | .6261 | .6259 |
| $r = 5$ | 5.0 | 1.15329 | .8756 | .8736 | .8681 | .8671 |
| $m = 10$ | 6.0 | 1.99902 | .9772 | .9723 | .9741 | .9739 |
| $n = 10$ | 7.0 | 2.84475 | .9978 | .9963 | .9976 | .9979 |

**7. Acknowledgment.** The authors wish to thank Prof. J. W. Tukey for his helpful comments and suggestions.

## REFERENCES

[1] D. E. BARTON AND F. N. DAVID, "Sums of ordered intervals and distances," *Mathematika*, Vol. 2 (1955), pp. 150–159.

[2] H. CRAMÉR, *Mathematical Methods of Statistics*, Princeton University Press, 1946.

[3] B. EPSTEIN AND M. SOBEL, "Some theorems relevant to life testing from an exponential distribution," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 373–381.

[4] L. M. MILNE-THOMSON, *The Calculus of Finite Differences*, MacMillan, London, 1933.