The Annals of Statistics
1973, Vol. 1, No. 5, 958-963

LIMITING DISTRIBUTIONS OF STATISTICS SIMILAR TO STUDENT'S t

BY Z. W. BIRNBAUM[1] AND I. VINCZE[2]

University of Washington and Mathematical Institute of the Hungarian Academy of Sciences

A class of statistics is considered, which are based on order-statistics and have properties analogous to Student's t. They can be used to estimate a required quantile of a random variable or to test hypotheses about a quantile; they are simple to compute, and can be calculated when not all sample values are available, e.g. for censored samples. The limiting distributions of these statistics are derived and shown to be independent of the distribution of the underlying random variable. A numerical tabulation of the limiting distribution is included, for the special case when the quantile considered is the median.

1. Definition of a statistic and some of its properties.

1.1. Let X be a random variable with continuous distribution function $F(x) = P\{X \leqq x\}$, and $X_{(1)} \leqq X_{(2)} \leqq \cdots \leqq X_{(n)}$ an ordered sample of X. The q-quantile of X will be defined by

$$\mu_q = \inf\{x : F(x) = q\} \quad \text{for given} \quad 0 < q < 1 , \tag{1.1.1}$$

and the corresponding sample-quantile is the order-statistic

$$X_{(k)} \quad \text{such that} \quad \left|\frac{k}{n} - q\right| \leqq \frac{1}{2n} . \tag{1.1.2}$$

For given integers r, s such that $0 < r < k$, $0 < s < n + 1 - k$ we consider the statistic

$$S_{n,k,r,s} = \frac{X_{(k)} - \mu_q}{X_{(k+s)} - X_{(k-r)}} . \tag{1.1.3}$$

The special case $q = \frac{1}{2}$, $n = 2m + 1$, $r = s$ yields the statistic

$$S = \frac{X_{(m+1)} - \mu_{\frac{1}{2}}}{X_{(m+1+r)} - X_{(m+1-r)}} \tag{1.1.4}$$

which was considered in [1], while similar statistics were previously studied in [3]. Furthermore, the authors are indebted to the referee for calling their attention to paper [4] by M. M. Siddiqui in which a general class of related

Received July 1971; revised January 1973.

[1] Research by this author was supported in part by the U. S. Office of Naval Research and by the National Science Foundation.

[2] During his stay at the Catholic University of America, Washington D. C., this author's research was supported in part by the National Science Foundation.

Key words and phrases. Order-statistics, nonparametric statistics, distribution-free statistics, quantiles, censored samples, Student's t.

statistics is considered and, as a special case, a statement is obtained which is equivalent with the theorem in Section 3.2.

1.2. The structure of S in (1.1.4) is somewhat similar to that of Student's t: (i) the numerator is the difference between a location-parameter (the population median $\mu_{\frac{1}{2}}$) and its estimate (sample median $X_{(m+1)}$), and the denominator is an estimate (sample interquantile range) of a scale parameter (population interquantile range); (ii) the statistic $S_{n,k,r,s}$, and as a special case the statistic S, is invariant under linear transformations and hence, for given distribution function $F(\cdot)$, the probability distribution of $S_{n,k,r,s}$ is independent of location- and scale-parameters, i.e. is the same for all random variables with distribution functions $F((x-a)/b)$, a real, $b > 0$. In particular, if X has normal distribution, the statistic S can be used in a manner analogous to that in which one uses the t-statistic, with the additional practical advantage that it can be computed when no more than three order statistics are available, and part or all of the remaining sample has been "censored".

1.3. In practical situations the values of r and s are often given, and not within the experimenter's control. This is, for example, the case when $X_{(1)} \leqq \cdots \leqq X_{(n)}$ is an ordered sample of diameters of a mass-produced item, which is processed as follows: first each item is put through a go–no–go gauge which rejects all items with diameters $< D_1$ or $> D_2$, $D_1 < D_2$; then the remaining items are carefully measured. The resulting data consist of (1) the number $a-1$ of diameters $< D_1$, (2) the number $b-1$ of diameters $> D_2$, and (3) the actually observed diameters $X_{(a)} \leqq X_{(a+1)} \leqq \cdots \leqq X_{(b)}$ contained in the interval $[D_1, D_2]$. This determines the subscripts a and b of the extreme order-statistics which can be used as $X_{(k-r)}$ and $X_{(k+s)}$ in (1.1.3) and, if the gauge is narrow, leads to values r and s which are not large, although n is large.

2. Exact distribution of $S_{n,k,r,s}$.

2.1. We assume in the following the existence everywhere of the probability density $f(x) = F'(x) > 0$ for $0 < F(x) < 1$. From the joint probability density of the order statistics

$$U = X_{(k-r)}, \qquad V = X_{(k)}, \qquad W = X_{(k+s)} \tag{2.1.1}$$

we obtain for $S_{n,k,r,s}$ defined in (1.1.3) (assuming without loss of generality $\mu_q = 0$):

$$\begin{aligned} &P\{S_{n,k,r,s} > \lambda\} \\ &\quad = P(\lambda) = \frac{n!}{(k-r-1)!\,(r-1)!\,(s-1)!\,(n-k-s)!} \\ &\qquad \times \int_{v=0}^{+\infty} \int_{u=(1-1/\lambda)v}^{v} \int_{w=v}^{u+v/\lambda} f(u)f(v)f(w)F(u)^{k-r-1} \\ &\qquad \times [F(v)-F(u)]^{r-1}[F(w)-F(v)]^{s-1}[1-F(w)]^{n-k-s}\, dw\, du\, dv \end{aligned} \tag{2.1.2}$$

for $\lambda > 0$.

2.2. For every two-parameter family of distributions with a location- and a scale-parameter, determined by a given $F(\cdot)$, expression (2.1.2) and a similar expression for $\lambda < 0$ could be used to compute numerically the exact values of all probabilities needed for practical use. These computations are being prepared for the special case of (1.1.4) and the family of normal random variables. Monte Carlo estimates of the probabilities (2.1.2) for this special case have been obtained by J. Tague [5], for $m = 1\ (1)\ 10$, $r = 1\ (1)\ m$, $\lambda = 0.0\ (0.1)\ 5$, and some selected larger values of λ.

3. Limiting distributions.

3.1. The following intuitive argument makes it reasonable to expect that for n large and r/n and s/n small the probability distribution of $S_{n,k,r,s}$ will be practically independent of the given distribution $F(\cdot)$; for such n, r, s all three order statistics (2.1.1) fall with probability close to 1 very close to μ_q, where $F(x)$ is approximately equal to $q + f(\mu_q) \cdot (x - \mu_q)$. Therefore S is approximately distributed as if the sample were obtained from a random variable with uniform distribution on $[\mu_q - q/f(\mu_q), \mu_q + (1 - q)/f(\mu_q)]$, provided $f(\mu_q) > 0$, and since a change of location and of scale does not affect the probability distribution of $S_{n,k,r,s}$ it is approximately distributed as if the sample came from a random variable distributed uniformly on $[-\frac{1}{2}, \frac{1}{2}]$. This intuitive argument is born out and given a specific form in the following theorem.

3.2. THEOREM. *Let $F(x)$ be a probability distribution function with q-quantile μ_q, such that $f(x) = F'(x)$ exists for $0 < F(x) < 1$,*[3] *and*

$$f(x) > 0 \quad \text{for} \quad 0 < F(x) < 1\,; \tag{3.2.1}$$

(3.2.2) *the derivative $f'(x)$ exists and is continuous in an interval $(\mu_q - \delta, \mu_q + \delta)$ for some $\delta > 0$.*

Then, for fixed integers r, s, one has

$$\lim_{n\to\infty} P\left\{\frac{1}{n^{\frac{1}{2}}} S_{n,k,r,s} \leqq \lambda\right\} = \frac{1}{(r+s-1)!} \int_0^\infty \phi[\lambda u/(q(1-q))^{\frac{1}{2}}] u^{s+r-1} e^{-u}\, du \tag{3.2.3}$$

where $\phi(\cdot)$ is the standardized normal distribution function.

PROOF. We assume, again without loss of generality,

$$\mu_q = 0 \tag{3.2.4}$$

so that $S_{n,k,r,s} = V/(W - U)$.

The joint probability density of

$$\begin{aligned} x &= n[F(v) - F(u)]\,, \\ y &= n^{\frac{1}{2}}[F(v) - q]\,, \\ z &= n[F(w) - F(v)] \end{aligned} \tag{3.2.5}$$

[3] The assumption that the probability density $f(x)$ exists exerywhere is made for convenience in the proof. It can be replaced by a weaker assumption.

is

$$p_{n,k,r,s}(x, y, z) = \frac{n!}{(k-r-1)!\,(r-1)!\,(s-1)!\,(n-k-s)!} n^{-\frac{3}{2}} \tag{3.2.6}$$
$$\times (q + n^{-\frac{1}{2}}y - n^{-1}x)^{k-r-1}$$
$$\times (n^{-1}x)^{r-1}(n^{-1}z)^{s-1}(1 - q - n^{-\frac{1}{2}}y - n^{-1}z)^{n-k-s}$$

for

$$0 < x < nq + n^{\frac{1}{2}}y$$
$$-qn^{\frac{1}{2}} < y < (1-q)n^{\frac{1}{2}} \tag{3.2.6.1}$$
$$0 < z < n(1-q) - n^{\frac{1}{2}}y$$

and

$$p_{n,k,r,s}(x, y, z) = 0 \quad \text{elsewhere.}$$

For $n \to \infty$ one obtains by Stirling's formula and (1.1.2) for x, y, z satisfying (3.2.6.1)

$$p_{n,k,r,s}(x, y, z) \sim \frac{1}{(2\pi q(1-q))^{\frac{1}{2}}} \frac{1}{(r-1)!\,(s-1)!} x^{r-1}z^{s-1}$$
$$\times (1 + n^{-\frac{1}{2}}y/q)^{k-r-1}[1 - x/n(q + n^{-\frac{1}{2}}y)]^{k-r-1}$$
$$\times [1 - n^{-\frac{1}{2}}y/(1-q)]^{n-k-s}[1 - z/n(1 - q - n^{-\frac{1}{2}}y)]^{n-k-s} .$$

Since $k \sim nq$, and r and s are fixed, we readily find that

$$\lim_{n\to\infty} p_{n,k,r,s}(x, y, z) = p_{r,s}(x, y, z) \tag{3.2.7}$$

for all x, y, z, where the right-hand side is defined by

$$p_{r,s}(x, y, z) = \frac{1}{(r-1)!} e^{-x}x^{r-1} \cdot \frac{1}{(s-1)!} e^{-z}z^{s-1} \tag{3.2.8}$$
$$\times [2\pi q(1-q)]^{-\frac{1}{2}}e^{-y^2/[2q(1-q)]}$$

for $x > 0$, $z > 0$, $-\infty < y < +\infty$, and $p_{r,s}(x, y, z) = 0$ elsewhere.

For the inverse function of $F(x)$ one has

$$F^{(-1)}(q + h) = F^{(-1)}(q) + \frac{1}{f(\mu_q)} h - \frac{1}{2} \frac{f'(\mu_q)}{f^3(\mu_q)} h^2 + o(h^2) .$$

Hence in view of (3.2.4) and (3.2.5)

$$u = F^{(-1)}(q + n^{-\frac{1}{2}}y - n^{-1}x)$$
$$= \frac{1}{f(\mu_q)} (n^{-\frac{1}{2}}y - n^{-1}x)$$
$$- \frac{1}{2} \frac{f'(\mu_q)}{f^3(\mu_q)} (n^{-\frac{1}{2}}y - n^{-1}x)^2 + o(n^{-1}) ,$$

and similarly for v and w, so that

$$S_{n,k,r,s} = \frac{V}{W - U} = \frac{n^{\frac{1}{2}}Y - \frac{1}{2}[f'(\mu_q)/f^2(\mu_q)]Y^2 + o(1)}{X + Z + o(1)} .$$

Using this and (3.2.8), and verifying the validity of the passages to the limit, one obtains for $n \to \infty$ and $\lambda > 0$

$$P\{n^{-\frac{1}{2}}S_{n,k,r,s} \leqq \lambda\} = P\left\{\frac{Y + o_p(n^{-\frac{1}{2}})}{X + Z + o_p(1)} \leqq \lambda\right\}$$
$$\to \iiint_{y/(x+z)\leqq\lambda} P_{r,s}(x, y, z)\, dx\, dy\, dz$$
$$= \int_{y=-\infty}^{+\infty} \frac{1}{(2\pi q(1-q))^{\frac{1}{2}}} e^{-y^2/2q(1-q)} \iint_{x+z\geqq\min(0,y/\lambda)} \frac{1}{(r-1)!} e^{-x}x^{r-1}$$
$$\times \frac{1}{(s-1)!} e^{-z}z^{s-1}\, dx\, dz\, dy$$
$$= \int_{t=0}^{+\infty} \frac{1}{(r+s+1)!} e^{-t}t^{r+s-1} \frac{1}{(2\pi)^{\frac{1}{2}}} \int_{u=-\infty}^{\lambda t/(q(1-q))^{\frac{1}{2}}} e^{-u^2/2}\, du\, dt$$

and a similar argument yields the same expression for $\lambda < 0$, which concludes the proof of (3.2.3).

3.3. Writing

$$P_{r,s}(\lambda) = \frac{1}{(r+s-1)!} \int_0^\infty \phi[\lambda u/(q(1-q))^{\frac{1}{2}}]u^{s+r-1}e^{-u}\, du \tag{3.3.1}$$

we have for fixed ξ

$$\lim_{r+s\to\infty} P_{r,s}\left(\frac{\xi}{r+s-1}\right) = \phi[\xi/(q(1-q))^{\frac{1}{2}}], \tag{3.3.2}$$

so that for large $r + s$ a table of $\phi(\cdot)$ can be used to approximate values of $P_{r,s}(\lambda)$.

PROOF.

$$P_{r,s}\left(\frac{\xi}{r+s-1}\right) = \frac{1}{(r+s-1)!} \int_0^\infty \phi\left[\frac{\xi}{r+s-1} u/(q(1-q))^{\frac{1}{2}}\right] u^{r+s-1}e^{-u}\, du$$
$$= \frac{(r+s-1)^{r+s}}{(r+s-1)!} \int_0^\infty \phi[\xi z/(q(1-q))^{\frac{1}{2}}]z^{r+s-1}e^{-(r+s-1)z}\, dz$$
$$= \int_0^\infty \phi[\xi z/(q(1-q))^{\frac{1}{2}}]T_{r,s}(z)\, dz$$

where

$$T_{r,s}(z) = \frac{(r+s-1)^{r+s}}{(r+s-1)!} z^{r+s-1}e^{-(r+s-1)z}.$$

For every $\varepsilon > 0$ we have, as $r + s \to \infty$,

$$T_{r,s}(z) \to 0 \qquad \text{uniformly for} \quad |z - 1| \geqq \varepsilon$$

and

$$\int_{1-\varepsilon}^{1+\varepsilon} T_{r,s}(z)\, dz \to 1,$$

and (3.3.2) follows.

4. Special case $q = \frac{1}{2}$, $n = 2m + 1$, $r = s$; some numerical tabulations.

4.1. For the special case when $\mu_q = \mu$ is the population median, the sample size $n = 2m + 1$ is odd, $X_{(m+1)}$ is the sample median and $r = s$, our statistic (1.1.3) becomes the statistic S of (1.1.4), and (3.2.3) may be written

$$\lim_{m\to\infty} P\{(2/m)^{\frac{1}{2}}S \leqq s\} = \frac{1}{(2r-1)!} \int_0^\infty \phi(su)u^{2r-1}e^{-u}\,du = P_r(s)\,. \tag{4.1.1}$$

Similarly, when $P_r(\cdot)$ is written instead of $P_{r,r}(\cdot)$, (3.3.2) becomes

$$\lim_{r\to\infty} P_r\left(\frac{\xi}{2r-1}\right) = \phi(\xi)\,. \tag{4.1.2}$$

4.2. Critical values for the limiting distribution (4.1.1) are presented in Table 1. This table has been abstracted from a more detailed tabulation prepared by Dr. G. F. Steck of Sandia Laboratories. The authors wish to express their gratitude to Dr. Steck for the permission to use his results.

TABLE 1

Table of asymptotic critical values for S, i.e., of $s_{r,\alpha}$ such that according to (4.1.1.)
$\lim_{m\to\infty} P\{(2/m)^{\frac{1}{2}}S < s_{r,\alpha}\} = 1 - \alpha = P_r(s_{r,\alpha})$

r	α				
	.10	.05	.025	.01	.005
1	1.0086	1.6778	2.6128	4.4575	6.5318
2	.3987	.5826	.7935	1.1272	1.4318
3	.2465	.3452	.4501	.6029	.7321
4	.1782	.2442	.3115	.4052	.4812
5	.1395	.1887	.2375	.3035	.3556
6	.1146	.1537	.1917	.2421	.2811
7	.0972	.1296	.1606	.2011	.2319
8	.0844	.1120	.1381	.1718	.1972
9	.0746	.0986	.1212	.1500	.1715
10	.0668	.0881	.1079	.1330	.1516

REFERENCES

[1] BIRNBAUM, Z. W. (1970). On a statistic similar to Student's *t*. *Nonparametric Techniques in Statistical Inference*. Cambridge Univ. Press. 427–433.

[2] BIRNBAUM, Z. W. (1972). Asymptotically distribution-free statistics similar to Student's *t*. *Proc. Sixth Berkeley Symp. Math. Statist. Prob.* 325–329. Univ. of California Press.

[3] DAVID, F. N. and JOHNSON, N. L. (1956). Some tests of significance with ordered variables. *J. Roy. Statist. Soc.* **18** 1–20.

[4] SIDDIQUI, M. M. (1960). Distribution of quantiles in samples from a bivariate population. *J. Res. Nat. Bur. Standards Sect. B* No. 3 145–150.

[5] TAGUE, JEAN (1969). Monte Carlo tables for the *S*-statistic. Memorial Univ. of Newfoundland. Unpublished.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF WASHINGTON
SEATTLE, WASHINGTON 98195