where $\phi(z)$ is the ordinary m.g.f. of a non-negative random variable. Likewise a necessary and sufficient condition for $\omega(z)$ to be the f.m.g.f. of a generalized Poisson distribution is that it be of the form

$$(2) \qquad \omega_2(z) = e^{\alpha(\Omega(z)-1)}, \qquad \alpha > 0,$$

where $\Omega(z)$ is the f.m.g.f. of an arbitrary distribution function $F(x)$. If we choose $\phi(z) = e^{\alpha(e^{cz}-1)}$ and $\Omega(z) = e^{cz}$, then $\omega_1(z) = \omega_2(z)$, and the distribution whose f.m.g.f. is $\omega_1(z)$ (the Neyman contagious distribution of Type A) is simultaneously a compound and a generalized Poisson distribution (cf. Feller [2]). We now show that there is an infinite class of distributions with this property.

First note that if $\phi(z)$ is the m.g.f. of an arbitrary distribution, then exp $\{\alpha(\phi(z) - 1)\}$ is also the m.g.f. of a d.f., and in fact is the m.g.f. of the generalized Poisson distribution associated with the distribution whose m.g.f. is $\phi(z)$. Now let $\phi(z)$ be the m.g.f. of an arbitrary non-negative random variable, and define

$$(3) \qquad \omega(z) = \exp\{\alpha(\phi(z) - 1)\} \qquad \alpha > 0.$$

Then $\omega(z)$ is simultaneously of the forms (1) and (2), since $\phi(z)$ is, by (1), also the f.m.g.f. of a distribution function, i.e. the compound Poisson distribution associated with the distribution whose m.g.f. is $\phi(z)$. However, not every distribution which is both a compound and a generalized Poisson distribution can be generated in this manner. For example, the Polya-Eggenberger distribution is easily shown to be both a generalized and a compound Poisson distribution, yet its f.m.g.f.

$$\omega(z) = (1 - dz)^{-h/d}, \qquad d > 0, h > 0,$$

manifestly is not of the form (3), since this would imply $\phi(iz) = 1 - \dfrac{h}{\alpha d}$ log $(1 - diz)$ is a characteristic function. But $|\phi(iz)|$ is unbounded as $z \to \pm\infty$ and thus is not the characteristic function of a distribution.

REFERENCES

[1] H. Cramér, "Problems in probability theory," *Annals of Math. Stat.*, Vol. 18 (1947), pp. 165–193.

[2] W. Feller, "On a general class of contagious distributions," *Annals of Math. Stat.*, Vol. 14 (1943), pp. 389–400.

[3] P. Hartman and A. Wintner, "On the infinitesimal generators of integral convolutions," *Am. Jour. of Math.*, Vol. 64 (1942), pp. 272–279.

ON CONFIDENCE LIMITS FOR QUANTILES

By Gottfried E. Noether

Columbia University

In finding confidence limits for quantiles it is usual to determine two order statistics Z_i and Z_j which with a given probability contain the unknown quantile between them. The values of i and j corresponding to a given confidence coefficient can be determined with the help of the distribution laws of order statistics as is shown, e.g., in Wilks [1]. The purpose of this note is to determine i and j with the help of a confidence band for the unknown cumulative distribution function.

In what follows we shall always denote the cumulative distribution function (cdf) by $F(x)$, i.e., $F(x) = P\{X \leq x\}$. Then the quantile q_p is determined by

$$F(q_p - 0) \leq p \leq F(q_p) \tag{1}$$

which reduces to

$$F(q_p) = p \tag{1'}$$

if $F(x)$ is continuous. Given a sample of size n we can construct the sample cdf $F_n(x)$ defined by $F_n(x) = 1/n$ (number of observations $\leq x$). Confidence coefficients will always be denoted by $1 - \alpha$.

Assume that we can construct two step functions $L(x)$ and $U(x)$ parallel to $F_n(x)$ such that for any fixed value x

$$P\{L(x) < F(x) < U(x)\} = 1 - \alpha. \tag{2}$$

We do not require that the confidence band determined by $L(x)$ and $U(x)$ cover the graph of the unknown cdf $F(x)$ with probability $1 - \alpha$, but only that for any arbitrarily chosen value x (2) is true.

Let

$$L(x) = \eta_k, \qquad U(x) = \theta_k$$

for $z_k \leq x < z_{k+1}$, $k = 0, 1, \cdots, n$ where z_k is the value taken by the order statistic Z_k and $z_0 = -\infty$, $z_{n+1} = +\infty$. Then if $F(x)$ is continuous it follows from (2) that a confidence interval with confidence coefficient $1 - \alpha$ for q_p is given by

$$Z_i \leq q_p < Z_j \tag{3}$$

where i and j are determined by

$$\theta_{i-1} \leq p, \qquad \theta_i > p \tag{4}$$

$$\eta_{j-1} < p, \qquad \eta_j \geq p. \tag{5}$$

It will be noted that (3) represents a half-open interval. However as long as we only admit continuous cdf's the confidence coefficient is not changed if we use

$$Z_i < q_p < Z_j \tag{3'}$$

or

$$Z_i \leq q_p \leq Z_j \tag{3''}$$

instead. This is no longer true if we also admit discontinuous cdf's. Then the confidence coefficient connected with (3') is $\leq 1 - \alpha$, while that connected with

(3″) is $\geq 1 - \alpha$, as follows immediately from consideration of the possible outcomes when (1) is true. This is the same result as that obtained by Scheffé and Tukey [2].

We shall now indicate how η_k and θ_k can be obtained and find their values in a particular case. For any arbitrary value x we can consider $F_n(x)$ as the sample estimate of the unknown parameter $p = F(x)$ of a binomial distribution. Clopper and Pearson [3] have discussed how confidence intervals for the unknown parameter of a binomial variate can be found. Thus we can determine η_k and θ_k correspondingly, but as is well known (2) cannot be achieved with probability exactly equal to $1 - \alpha$. We shall have to be satisfied with probability $\geq 1 - \alpha$. Consequently the same will hold true for the confidence coefficient connected with the confidence interval for q_p.

In many cases central confidence intervals seem to be more desirable, at least intuitively, than others. Our method produces such central confidence intervals for the unknown quantile if we use central confidence intervals in the construction of the confidence band. In that case η_k and θ_k are determined by

$$\frac{\alpha}{2} = I_{\eta_k}(k, n - k + 1) \tag{6}$$

$$k = 0, 1, \cdots, n$$

$$\frac{\alpha}{2} = I_{1-\theta_k}(n - k, k + 1), \tag{7}$$

except that $\eta_0 = 0$, $\theta_n = 1$ by definition, where

$$I_x(p, q) = \int_0^x t^{p-1}(1 - t)^{q-1}\, dt \Big/ \int_0^1 t^{p-1}(1 - t)^{q-1}\, dt$$

is the incomplete beta function. Scheffé [4] has pointed out how the tables of percentage points of the incomplete beta function by C. M. Thompson, etc. [5] can be used to find η_k and θ_k.

We shall show now that in the case of the median M the solution based on (3)–(7) leads to the same confidence interval as that suggested originally by W. R. Thompson [6]. Thompson found that for $k < n + \frac{1}{2}$

$$P\{Z_k < M < Z_{n-k+1}\} = 1 - 2I_{\frac{1}{2}}(n - k + 1, k) \tag{8}$$

provided the unknown distribution had a continuous cdf. (8) can be used to maximize k under the condition that the righthand side is $\geq 1 - \alpha$.

We shall first show that our method leads to the same kind of a confidence interval, i.e., one with $i = l, j = n - l + 1$. This follows immediately from the fact that by (6) and (7)

$$1 - \theta_l = \eta_{n-l}. \tag{9}$$

For let

$$\theta_{l-1} \leq \tfrac{1}{2} \text{ and } \theta_l > \tfrac{1}{2}, \tag{10}$$

then by (9) $\eta_{n-l} < \frac{1}{2}$ and $\eta_{n-l+1} \geq \frac{1}{2}$.

It remains to be shown that k as determined by (8) equals l. This will be so if we can show that

$$I_{\frac{1}{2}}(n - l + 1, l) \leq \frac{\alpha}{2} < I_{\frac{1}{2}}(n - l, l + 1). \tag{11}$$

Remembering that $I_x(p, q)$ is a monotonically increasing function of x we get with the help of (7) and (10)

$$\frac{\alpha}{2} = I_{1-\theta_{l-1}}(n - l + 1, l) \geq I_{\frac{1}{2}}(n - l + 1, l)$$

and

$$\frac{\alpha}{2} = I_{1-\theta_l}(n - l, l + 1) < I_{\frac{1}{2}}(n - l, l + 1)$$

which proves (11).

In conclusion it may be worth while pointing out that the formula

$$P\{Z_i < q_p < Z_j\} = I_p(i, n - i + 1) - I_p(j, n - j + 1)$$

given, e.g, in Wilks [1] for the continuous case can be obtained by a slight modification of (6).

REFERENCES

[1] S. S. Wilks, "Order statistics," *Am. Math. Soc. Bull.*, Vol. 54 (1948), pp. 6–50.
[2] H. Scheffé and J. W. Tukey, "Non-parametric estimation. I. Validation of order statistics," *Annals of Math. Stat.*, Vol. 16 (1945), pp. 187–192.
[3] C. J. Clopper and E. S. Pearson, "The use of confidence or fiducial limits illustrated in the case of the binomial," *Biometrika*, Vol. 26 (1934), pp. 404–413.
[4] H. Scheffé, "Note on the use of the tables of percentage points of the incomplete beta function to calculate small sample confidence intervals for binomial p," *Biometrika*, Vol. 33 (1944), p. 181.
[5] C. M. Thompson, E. S. Pearson, L. J. Comrie, and H. O. Hartley, "Tables of percentage points of the incomplete beta function," *Biometrika*, Vol. 32 (1941), pp. 151–181.
[6] W. R. Thompson, "On confidence ranges for the median and other expectation distributions for populations of unknown distribution form," *Annals of Math. Stat.*. Vol. 7 (1936), pp. 122–128.

A LOWER BOUND FOR THE EXPECTED TRAVEL AMONG m RANDOM POINTS

By Eli S. Marks

Bureau of the Census

In connection with cost determinations in sampling problems, it is frequently necessary to determine the amount of travel among m random sample points in an area. A lower bound for the expected value of this distance is found to be:

$$\sqrt{\frac{A}{2}}\frac{m - 1}{\sqrt{m}},$$