ON A CONNECTION BETWEEN CONFIDENCE AND TOLERANCE INTERVALS

By Gottfried E. Noether

Boston University

The purpose of this note is to point out the close connection which exists between confidence intervals for the parameter p of a binomial distribution and tolerance intervals.

Let k be the number of successes in a random sample of size n from a binomial population with probability p of success in a single trial. Then it is well known that a confidence interval with confidence coefficient at least $1 - \alpha_1 - \alpha_2$ for the parameter p is given by

$$p_1(k) < p < p_2(k), \tag{1}$$

where $p_1(k)$ and $p_2(k)$ are determined by $I_{p_1(k)}(k, n - k + 1) = \alpha_1$ and $I_{1-p_2(k)}(n - k, k + 1) = 1 - I_{p_2(k)}(k + 1, n - k) = \alpha_2$, respectively, $I_x(a, b) = [\Gamma(a + b)/(\Gamma(a)\Gamma(b))] \int_0^x u^{a-1}(1 - u)^{b-1}\, du$ being the incomplete B-function.

Let $X_1, \cdots, X_n$ represent a random sample of size n from a population having continuous cdf $F(x)$. For simplicity assume that the X's are already arranged in increasing order of size and define $X_0 = -\infty$, $X_{n+1} = +\infty$. The coverage provided by the interval (X_i, X_{i+1}), $i = 0, 1, \cdots, n$, is called an elementary coverage.[1] If we then let U_r stand for the sum of r elementary coverages, $U_r > U_r(\alpha)$ unless an event of probability α has occurred, where $U_r(\alpha)$ is defined by $\alpha = [\Gamma(n + 1)/(\Gamma(r)\Gamma(n - r + 1))] \int_0^{U_r(\alpha)} u^{r-1}(1 - u)^{n-r}\, du = I_{U_r(\alpha)}(r, n - r + 1)$.

In this notation (1) becomes

$$U_k(\alpha_1) < p < U_{k+1}(1 - \alpha_2).$$

Thus the lower end point of a confidence interval for p on the basis of k observed successes is determined by the corresponding lower limit for the sum of k elementary coverages, while the upper end point is determined by the corresponding upper limit of the sum of $(k + 1)$ elementary coverages. The reason for this becomes obvious if we look at the k successes as the observations $X_1, \cdots, X_k$ which are smaller than the p-quantile q_p of $F(x)$, so that the coverage U_k of the chance interval (X_0, X_k) provides an "inner" estimate of p, while the coverage U_{k+1} of the chance interval (X_0, X_{k+1}) provides an "outer" estimate.

We may ask what kind of a confidence interval we obtain if we consider as successes the k observations belonging to an arbitrary interval I for which $\int_I dF(x) = p$, as long as I does not coincide with either $(-\infty, q_p)$ or $(q_{1-p}, +\infty)$.

[1] For rigorous definitions and formulas see, e.g., Wilks [1], p. 13.

It is easily seen that an "outer" estimate of p is still given by U_{k+1}. However, an "inner" estimate is now given by U_{k-1}, leading to a lower end point of the confidence interval which is unnecessarily small.

The method of obtaining a confidence interval for p discussed in this note is in a certain sense the reverse of the method discussed in an earlier paper of the author [2]. There it was shown how confidence intervals for p can be used to obtain confidence intervals for quantiles, which then can be used to obtain tolerance intervals.

REFERENCES

[1] S. S. WILKS, "Order statistics," *Bull. Am. Math. Soc.*, Vol. 54 (1948), pp. 6–50.
[2] G. E. NOETHER, "On confidence limits for quantiles," *Annals of Math. Stat.*, Vol. 19 (1948), pp. 416–419.

ABSTRACTS OF PAPERS

(Abstracts of papers presented at the Minneapolis meeting of the Institute, September 4–7, 1951)

1. On Stieltjes Integral Equations of Stochastic Processes. MARIA CASTELLANI, University of Kansas City.

This paper considers two methods of solving certain S-integral equations.

a. *A Fredholm-Stieltjes integral equation of generating functions.* We give the F-S integral equation $\int_E A(s, x)\, dg(x) = f(s)$, where $A(s, x) = \Sigma_{k=0}^{\infty} \alpha_k(x)s^{-k}$ and $f(s) = \Sigma a_k s^k$ for $s \to \varphi(s)$ and $a_0 = 0$ if $k = 0$. Let us assume that $u(x)$ and $v(x)$ are respectively solutions of $\int_E A(s, x)\cdot A(-s_1, x)\, du(x) = 1/(S - S_1)$ and $\int_E A(s, x)\, dv(x) = 0$. If we consider

$$\int_E A(s, x)A(-s_1, x)f(s_1)\, du(x) = f(s_1)/(S - S_1)$$

and if $\gamma(x)$ is the coefficient of $-1/S_1$ in the serial expansion of $A(-s, x)f(s_1)$, then under fairly general conditions the required solutions are given, almost everywhere, by $g(x) = \text{const.} \int^x dv(x) + \int^x \gamma(x)\, du(x)$. The proof is based on a Murphy D'Arcais linear operator and on the ρ operation of S-integrals.

b. *A Volterra-Stieltjes integral of recurrent random functions.* Let us have over a time interval (τ, t) an unknown rf$\delta(t - \tau)$ satisfying the following recursive equation: $\delta(t - \tau) = \delta(\tau) - \int_\tau^t \delta(x - \tau)\rho(x)\, dF(x)$ where $F(x)$ is a df and $\rho(x)$ is bounded. We assume the interval divided into n parts and also that the set of the n discrete values of δ satisfy the following relation: $\delta(t - \tau)/\delta(\tau) = \Pi_{s=\tau}^{s=t}(1 - \rho(s)\Delta F(s))$. If $F = F_1 + F_2$, where the F_1 is a continuous function and F_2 is a jump function over a set S of points, then by a generalized method of Cantelli, taking finer and finer partitions, we obtain as a limit $\delta(t - \tau)/\delta(\tau) = \left[\exp\left(-\int_\tau^t \rho(x)\, dF_1(x)\right)\right] \Pi_{s\varepsilon S}(1 - \rho(s)\, dF_2(s))$. This gives almost everywhere the required solutions.