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.