COWLES COMMISSION DISCUSSION PAPER: STATISTICS: 335

July 29, 1949

A Note on the Use of Tolerance Intervals as Test Regions

by

A. Marshall

In my thesis[1] I proposed the use of tolerance regions as a test of the hypothesis of unchanged structure for linear stochastic equations that are members of a larger system. All questions of the validity of the normality assumptions aside, Prof. K. J. Arrow raised the question of the "studentization" of the test procedure. It can be seen now that on the one hand the answer is very easy and that on the other hand the test does not have certain asymptotic properties that have been proposed as desirable. This is a general failing of tolerance limit procedures although the quality control people are likely to be unimpressed.

The Two Parameter Case.

The problem may be formulated in this way: it is well known that if the null hypothesis H:0 is that X is distributed $F(\theta_1, \theta_2)$ and one wishes to test alternatives of the type $\theta_1 < \hat{\theta}_1$ then test region ω will not have desirable properties if

$$\int_{\omega} F(x, \theta_1, \theta_2)dx$$

is a function of θ_2. Thus the problem of "studentization" is the problem of finding regions ω similar with respect to θ_2, i.e.

$$\int_{\omega} F(x, \theta_1, \theta_2)dx = \alpha$$

for all values of θ_2. Now it is easily seen that test procedure I used requires no studentization, insofar as this would be possible at all, just as test procedures pertaining to Poisson distributions require none. The region R depends upon estimates of all the parameters, as it should. Thus

1. A. Marshall, A Test of Klein's Model III for Changes of Structure, the University of Chicago, 1949 (unpublished).

$$R_1 = \bar{x} \pm k_{\alpha,\beta} \left(s^2 + \frac{s^2}{T} + \operatorname{tr} \hat{\Lambda} \hat{\Omega} + \frac{s^2}{T} z_t^{*} \hat{M}^{-1}_{z^{*}z^{*}} z_t^{\prime *} + z_t^{o} \hat{\Pi}^{\prime **} \hat{\Lambda} \hat{\Pi}^{**} z_t^{o\prime}\right)^{1/2}$$

Th above indicates also to what extent it is possible to treat equations from larger systems separately since the test procedure involves the parameters of the other equations. But let us turn to a simpler problem as no generality is lost if we approach this problem from the point of view of the two parameter case. Let us then consider the case where our null hypothesis is

H:0 x is distributed $N(\mu, \sigma)$,

$R = \bar{x} \pm k_{\alpha,\beta}\, s$ such that

$$\frac{1}{\sqrt{2\pi}\,\sigma} \int_R e^{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}}\, dx \geq \alpha$$

with probability β. We are testing against an alternative of the form (say)

$\mu_1 \neq \mu$ and $\sigma_1 \geq \sigma$

The test procedure is to accept H:0 on the basis of one observation if the observation falls within R and to reject it otherwise. Both parameters μ and σ are under test so that the problem of "studentization" does not arise.

The Power of the Test

Some results on the power of such test regions are easily obtained although I have not seen them in the literature. Thus the probability of type I errors can be seen to be

$$\begin{aligned} P(\text{Type I error}) &\leq \beta(1-\alpha) + (1-\beta) \\ &\leq 1 - \alpha\beta \end{aligned}$$

This is probably not the best result obtainable but for $\beta \geq .9$ refinements will not sharpen the inequality greatly. And for the normal distribution

$$P(\text{Type II error}) \leq (1-\beta)\,\gamma(\mu-\mu_1, \sigma/\sigma_1) + \beta\; 1-\beta$$

where $\gamma(\mu-\mu_1, \sigma/\sigma_1)$ can be evaluated on the basis of the method of constructing tolerance limits for normal distributions and the relations between the null hypothesis and alternative distribution.

Exact results for the above problems depend upon the evaluation of the following integral for various values of N, δ, and λ:

$$\int_{-\infty}^{\infty}\int_{0}^{\infty}\frac{1}{\sqrt{2\pi}\,\lambda\sigma}\int_{\bar{x}-k_{\alpha,\beta}s}^{\bar{x}+k_{\alpha,\beta}s} e^{-\frac{(x-\mu+\delta)^2}{2\lambda^2\sigma^2}}\,dt\;dG(s)\,dF(\bar{x})$$

$$\delta,\lambda \geq 0$$

It is easily seen from the above results that the test procedure and all similar procedures do not have certain properties that have been advocated as being desirable. In particular the test is not consistent in the sense of Wald and Wolfowitz for the probability of type II error does not approach zero as the sample size become large. Also, unless $\sigma_1 \geq \sigma$ the test is biased, but these are perhaps not important alternatives. A modification of the test procedure will correct the first deficiency but for moderate sized samples this would be only a gesture of good faith toward asymptotic virtues for in actual procedure the test procedure would remain identically the same.

The truth and relevance of Arrow's remark should be made more explicit. Arrow is very right that the parameters of the other equations are nuisance parameters if we consider our null hypothesis as being concerned with the equation under test. On the other hand this is not the case if our null hypothesis is that the equation comes from the original structure of the equation system as a whole.