SETTING OF TOLERANCE LIMITS WHEN THE SAMPLE IS LARGE

By Abraham Wald

Columbia University

1. Introduction. Let $f(x_1, \cdots, x_p, \theta_1, \cdots, \theta_k)$ be the joint probability density function of the variates $x_1, \cdots, x_p$ involving k unknown parameters $\theta_1, \cdots, \theta_k$. A sample of size n is drawn from this population. Denote by $x_{i\alpha}(i = 1, \cdots, p; \alpha = 1, \cdots, n)$ the α-th observation on x_i. We will deal here with the following two problems of setting tolerance limits, which are of importance in the mass production of a product:

Problem 1. *For any two positive numbers $\beta < 1$ and $\gamma < 1$ we have to construct p pairs of functions of the observations $L_i(x_{11}, \cdots, x_{pn})$ and $U_i(x_{11}, \cdots, x_{pn})$ $(i = 1, \cdots, p)$ such that*

$$(1) \quad P\left\{\int_{L_p}^{U_p} \cdots \int_{L_1}^{U_1} f(x_1, \cdots, x_p, \theta_1, \cdots, \theta_k)\, dx_1 \cdots dx_p \geq \gamma \mid \theta_1, \cdots, \theta_k\right\} = \beta,$$

where for any relation R, $P(R \mid \theta_1, \cdots, \theta_k)$ denotes the probability that R holds, calculated under the assumption that $\theta_1, \cdots, \theta_k$ are the true values of the parameters.

Problem 2. *For any positive numbers $\beta < 1, \lambda < 1$ and for any positive integer N we have to construct p pairs of functions of the observations $L_i(x_{11}, \cdots, x_{pn})$ and $U_i(x_{11}, \cdots, x_{pn})$ with the following property: Let $y_{i\alpha}(i = 1, \cdots, p;\ \alpha = 1, \cdots, N)$ be the α-th observation on the variate x_i in a second sample of size N drawn from the same population as the first sample has been drawn. Denote by M the number of different values of α for which the p inequalities*

$$L_i(x_{11}, \cdots, x_{pn}) \leq y_{i\alpha} \leq U_i(x_{11}, \cdots, x_{pn}) \quad (i = 1, \cdots, p),$$

are fulfilled. Then

$$(2) \qquad P(M \geq \lambda N \mid \theta_1, \cdots, \theta_k) = \beta,$$

where $\theta_1, \cdots, \theta_k$ denote the unknown parameter values of the population from which the observations $x_{i\alpha}$ and $y_{i\alpha}$ have been drawn.

The functions L_i and U_i are called the tolerance limits for the variate x_i. We will say that L_i is the lower, and U_i the upper tolerance limit of x_i. In general, there exist infinitely many tolerance limits L_i and U_i which are solutions of Problem 1 or Problem 2. It is clear that the tolerance limits L_i and U_i are the more favorable the smaller the difference $U_i - L_i$. Hence if there exist several solutions for the tolerance limits L_i and U_i we should select that one for which the difference $U_i - L_i$ becomes a minimum in some sense.

S. S. Wilks[1] gave a solution of Problems 1 and 2 in the univariate case, i.e.

[1] S. S. Wilks, "Determination of sample sizes for setting tolerance limits," *Annals of Math. Stat.*, Vol. 12 (1941). See also his paper on the same subject presented at the meeting of the Institute of Mathematical Statistics in Poughkeepsie, September, 1942.