DETERMINATION OF SAMPLE SIZES FOR SETTING TOLERANCE LIMITS

By S. S. Wilks

Princeton University, Princeton, N. J.

1. Introduction. In the mass production of a given product or apparatus piece-part, Shewhart[1] has discussed a practical procedure for detecting the existence of assignable causes of variation in a given quality characteristic of the product as measured by a variable x. For example, x may be the thickness in inches of a washer or the tensile strength in pounds of a small aluminum casting made according to a given set of specifications; x varies in value from washer to washer or from casting to casting. Now suppose assignable causes of variability in x have been detected by Shewhart's procedure and have been sufficiently well eliminated by making appropriate refinements in the manufacturing process so that for all practical purposes the remaining variability may be considered "random," thus allowing us to assume that we have a statistical universe U in which x is a random variable with some distribution law $f(x)$. $f(x)$ is, in general, unknown and cannot be determined until long after the refined manufacturing operation has been under way. Two types of situations arise in practice, one in which x is a discrete variable taking on only certain isolated values as for example 1, 2, 3, $\cdots$, etc. with corresponding probabilities $p(1)$, $p(2)$, $\cdots$, the other being that in which x is essentially a continuous variable over some range with a corresponding probability density function $f(x)$. In this paper we shall consider the latter type of variable.

The problem now arises as to how we should calculate a tolerance range (L_1, L_2) for x from a sample, and how large the sample should be in order for the tolerance range to have a given degree of stability. More specifically, *for a given method of calculating tolerance limits, how large should our sample be in order that the proportion P of the universe included between L_1 and L_2 have an average value a, and will be such that the probability is at least p that P will lie between two given numbers, say b and c?* For example, if a tolerance range is obtained by using a *truncated sample range*, that is by letting L_1 be the greatest of the r smallest values in a sample and L_2 the smallest of the r largest values, r being chosen so that $E(P) = .99$, how large should the sample size, say n, be in order for the probability to be .9 that P would lie between .985 and .995? A similar question can be asked when the setting of only one tolerance limit is under consideration.

---

[1] W. A. Shewhart, *Economic Control of Quality of Manufactured Product*, D. Van Nostrand Company, New York, 1931.

2. Tolerance ranges from truncated sample ranges. Suppose that nothing is known about the distribution function $f(x)$ except enough to enable us to assume that it is continuous. Let a be the average value which P is to have, and suppose a sample of size n is drawn from the universe U so that $[(1 - a)(n + 1)]/2 = r$, say, is a positive integer. Let $x_1, x_2, \cdots, x_n$ be the sample values of x arranged in order of increasing magnitude. Let $L_1 = x_r$ and $L_2 = x_{n-r+1}$. The distribution law, say $g(P)$ of P the proportion of the universe included between these values of L_1 and L_2 is given by

$$(1)\quad g(P)\,dP = \frac{\Gamma(n+1)}{\Gamma[a(n+1)]\Gamma[(1-a)(n+1)]} P^{a(n+1)-1}(1-P)^{(1-a)(n+1)-1}\,dP.$$

This follows at once from the joint distribution law of x_n and x_{n-r+1} which can be derived as follows: Consider the x axis as being divided into k mutually exclusive intervals $I_1, I_2, \cdots, I_k$ with $p_1, p_2, \cdots, p_k$ as the associated probabilities $\left(\sum_1^k p_i = 1\right)$. In a sample of size n the probability that $n_1, n_2, \cdots, n_k$ $\left(\sum_1^k n_i = n\right)$ values of x will fall into $I_1, I_2, \cdots, I_k$ respectively is given by the well-known multinomial distribution law

$$(2)\qquad \frac{n!}{n_1!n_2!\cdots n_k!}\,p_1^{n_1}p_2^{n_2}\cdots p_k^{n_k}.$$

To get the distribution of x_r and x_{n-r+1} we take $k = 5$ and for $I_1, I_2, \cdots, I_5$ we take the intervals $(-\infty, x_r)$, $(x_r, x_r + dx_r)$, $(x_r + dx_r, x_{n-r+1})$, $(x_{n-r+1}, x_{n-r+1} + dx_{n-r+1})$, $(x_{n-r+1} + dx_{n-r+1}, \infty)$ respectively. The values of $p_1, p_2, \cdots, p_5$ are the integrals of $f(x)\,dx$ over these five intervals respectively and the values of $n_1, n_2, \cdots, n_5$ are $r - 1, 1, n - 2r, 1, r - 1$ respectively. By substituting these values of the p's and n's in (2) and neglecting terms of order higher than $dx_r\,dx_{n-r+1}$ the probability element for x_r and x_{n-r+1} is found at once to be[2]

$$(3)\qquad \frac{n!}{[(r-1)!]^2(n-2r)!}\left(\int_{-\infty}^{x_r} f(x)\,dx\right)^{r-1}\left(\int_{x_{n-r+1}}^{\infty} f(x)\,dx\right)^{r-1} \cdot\left(\int_{x_r}^{x_{n-r+1}} f(x)\,dx\right)^{n-2r} f(x_r)f(x_{n-r+1})\,dx_r\,dx_{n-r+1}.$$

Now let $\int_{-\infty}^{x_r} f(x)\,dx = u$, $\int_{x_{n-r+1}}^{\infty} f(x)\,dx = v$, then since $du = f(x_r)\,dx_r$ and $dv = -f(x_{n-r+1})\,dx_{n-r+1}$, the probability element of u and v may be written as

$$(4)\qquad \frac{\Gamma(n+1)}{\Gamma^2(r)\Gamma(n-2r+1)}\,u^{r-1}v^{r-1}(1-u-v)^{n-2r}\,du\,dv,$$

[2] For a discussion and a rather complete bibliography of the probability theory of "extreme values" such as x_r and x_{n-r+1} see E. J. Gumbel, "Les valeurs extrêmes des distributions statistiques," *Annales de l'Institut H. Poincaré* (1935).

the region of u and v of non-zero probability being the triangle bounded by the u and v axes and the line $u + v = 1$. Making the change of variables $1 - u - v = P$ and $u = Q$, integrating with respect to Q, and setting $r = (1/2)(1 - a)(n + 1)$ we find the distribution of P, the proportion of the universe included between x_r and x_{n-r+1} to be (1). It should be remarked that even if L_1 and L_2 are obtained by asymmetrical truncation by taking $L_1 = x_s$, $L_2 = x_t$ where $t - s = n - 2r + 1$, the distribution of $P = \int_{x_s}^{x_t} f(x)\,dx$ remains unchanged. Thus for a given p, by taking $L_1 = x_s$ and $L_2 = x_t$ where $t - s = n - 2r + 1 = a(n + 1)$, and choosing the smallest value of n for which $\int_b^c g(P)\,dP \geq p$ and such that $(1 - a)(n + 1)$ is a positive integer we have provided the answer to the italicized question for one method of calculating L_1 and L_2; a method which is valid for any unknown continuous distribution $f(x)$.

As an example, suppose we take $a = .99$, $b = .985$, $c = .995$ and $p = .99$. The size of sample required is found to be 1000 (999 to be exact). In fact in this case the probability of P being between .985 and .995 is .992. In this example, we may therefore make the statement that if x is a continuous variable under statistical control, and if samples of size 1000 are taken, the tolerance limits L_1 and L_2 taken as the fifth smallest and fifth largest values of x in the sample respectively, will, on the average, include 99% of the universe between them and furthermore, the tolerance limits calculated in this way for samples of size 1000 will, in about 99.2% of the samples, include between 98.5% and 99.5% of the universe between them.

If L_1 and L_2 are taken as the smallest and largest values of x in the sample respectively (corresponding to $r = 1$, i.e. sample range with no truncation), then in samples of size 1000, these tolerance limits will, on the average include 99.8% of the universe between them and the probability is .996 that L_1 and L_2 will include at least 99.5% of the universe between them. If the largest and smallest values of x in samples are used as tolerance limits and if we wish to state that the probability is .99 that such tolerance limits will include at least 99% of the universe, the size of sample required is 660. If the probability is lowered to .95 of including at least 99% of the universe, with such tolerance limits, the size of sample required is 130. Engineering statisticians[3] have pointed out on basis of practical experience the need of using samples of 100 to 1000 on even more cases in order to set tolerance limits which will include at least 99% of the universe with a satisfactorily high degree of certainty. The examples we have given based on sizes 1000, 660 and 130 will indicate the degree of stability to be expected for tolerance ranges for samples in this range of sizes. The degree of stability of the tolerance limits for samples of the size range 500 to 1000 appears to be of about the order of that demanded by the engineering statistician.

---

[3]Cf. W. A. Shewhart, *Statistical Methods from the Point of View of Quality Control*, The Gaduate School of the J.S. Department of Agriculttre, Washington (1939). P. 63.

In some cases it may be desirable to determine the size of samples so as to control the tolerance limits L_1 and L_2 individually, that is so that the probability is at least p that the proportions of the universe contained in the tails of the distribution cut off by L_1 and L_2 are in both cases between two given numbers, say d and e. In this case we would determine the least value of n so that

$$\int_d^e \int_d^e h(u, v)\, du\, dv \geq p \tag{5}$$

where $h(u, v)\, du\, dv$ denotes the function given by (4). For example, suppose $p = .99$, $d = 0$, $e = .005$. $r = 1$. The size of the sample needed is 1060. Thus in samples of size 1060, the probability is .99 that L_1 and L_2 taken as the smallest and the largest values in the sample respectively will cut off tails of the universe such that each tail will include not more than 0.5% of the universe.

If it is desired to set only one tolerance limit, say L_1; then the distribution of u would be used. This can be found by integrating (4) with respect to v. The distribution is

$$\frac{\Gamma(n+1)}{\Gamma(r)\Gamma(n-r+1)} u^{r-1}(1-u)^{n-r}\, du. \tag{6}$$

The probability p that the proportion of the universe in the tail which will be cut off by L_1 is between d and e is given by integrating the expression (6) from d to e. The value of n required to obtain any given value of p can then be determined. For example, in the case where $p = .99$, $d = 0$, $e = .005$, $r = 1$, the size of the sample needed is 920.

3. Tolerance range for a normal universe. The method of setting tolerance limits discussed in Section 2 assumes nothing about the distribution $f(x)$ except that it is continuous. If $f(x)$ can be assumed to have a given functional form involving unknown parameters, methods based on the theory of statistical estimation and having greater efficiency than those already discussed could be used for setting tolerance limits. We shall not go into a general discussion of such methods here although it does appear desirable to consider one very important example of the application of the methods. Suppose $f(x)$ can be assumed to be a normal distribution function with unknown mean m and variance σ^2. In a sample of size n let $\bar{x}$ be the sample mean and let $s^2 = \sum_1^n (x_i - \bar{x})^2/(n-1)$. Let us consider as tolerance limits L_1' and L_2' the quantities $\bar{x} \pm ks$. The proportion P' of the universe included between these limits is

$$P' = \frac{1}{\sqrt{2\pi}\,\sigma} \int_{\bar{x}-ks}^{\bar{x}+ks} e^{-\frac{1}{2}(x-m)^2/\sigma^2}\, dx. \tag{7}$$

We wish to determine k so that $E(P') = a$. It can be verified by straightforward analysis that $E(P')$, defined by $\int_{-\infty}^{\infty}\int_0^{\infty} P' f(\bar{x}, s)\, ds\, d\bar{x}$, has the value

$$\frac{\Gamma(n/2)}{\sqrt{\pi(n-1)}\,\Gamma((n-1)/2)} \int_{-t}^{t} \frac{dx}{(1 + x^2/(n-1))^{n/2}}, \qquad t = k\sqrt{\frac{n}{n+1}} \tag{8}$$

where $f(\bar{x}, s)$ is the well-known distribution of $\bar{x}$ and s given by

$$\frac{\sqrt{n}(n-1)^{(n-1)/2}s^{n-2}}{2^{n/2-1}\sigma^n\sqrt{\pi}\,\Gamma((n-1)/2)}\,e^{-\frac{1}{2}[n(\bar{x}-m)^2+(n-1)s^2]/\sigma^2}. \tag{9}$$

Therefore the tolerance limits L_1' and L_2' which will include, on the average, a proportion a of the universe between them are

$$\bar{x} \pm t_a\sqrt{(n+1)/n}\cdot s \tag{10}$$

where t_a is the value of t for which the integral in (8) has the value a. The value of t_a can be found from Fisher's t-table for $n - 1$ degrees of freedom, and for certain values of a including .99, .95, etc. and for values of n up to 30. Although the tolerance limits (10) will include, on the average, the proportion a of the universe between them, we must now investigate the size of sample needed to obtain a given degree of stability of P'. The exact distribution of P' seems to be too complicated to be of any practical value. It is not difficult to verify that to within terms of order $1/n$, the variance of P' is given by

$$\sigma_{P'}^2 = t_a^2 e^{-t_a^2}/(\pi n). \tag{11}$$

The variance of P, the proportion of the universe included between x_r and x_{n-r+1}, to within terms of order $1/n$ is given by

$$\sigma_P^2 = a(1-a)/n. \tag{12}$$

For a large sample of a given size, say $n = 100$ or more, a simple comparison of the stabilities of the two tolerance ranges (x_r, x_{n-r+1}) and $(\bar{x} \pm t_a\sqrt{(n+1)/n}\cdot s)$ can be made by comparing σ_P^2 and $\sigma_{P'}^2$. For $a = .99$, the efficiency ratio $\sigma_{P'}^2/\sigma_P^2$ is .28 indicating that for large n and when the universe is normal, samples of size $.28n$ have the same degree of stability in setting tolerance ranges (10) as a sample of size n has when (x_r, x_{n-r+1}) is taken as the tolerance range. The same thing may be viewed in another way: The fact that the range of values of P' is 0 to 1 suggests that we may be able to get a fairly close approximation to the true distribution of P' by fitting a Pearson Type I function of the form

$$\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}P'^{\alpha-1}(1-P')^{\beta-1}, \tag{13}$$

determining α and β by equating the mean and variance of the distribution (13) to the mean and variance of P' respectively. Accordingly we find

$$\begin{aligned} \alpha &= [a^2(1-a) - a\sigma_{P'}^2]/\sigma_{P'}^2 \\ \beta &= [a(1-a)^2 - (1-a)\sigma_{P'}^2]/\sigma_{P'}^2. \end{aligned} \tag{14}$$

Thus it will be seen from (14) that in order for the fitted distribution (13) to be identical with the distribution (1) a sample of only $\dfrac{t_a^2 e^{-t_a^2}}{\pi a(1-a)}(n+2)$ cases is needed.

In case only one tolerance limit is to be set, e.g. $\bar{x} - t_a\sqrt{(n+1)/n}\cdot s$, the proportion, say u', of the universe which will be included in the tail has mean value $(1 - a)/2$ and variance $\frac{(2 + t_a^2)}{4\pi n} e^{-t_a^2}$ (approximately) for large n. The ratio of this variance to that of u, which is approximately $(1 - a^2)/4n$ for large n, gives the efficiency of using x_r for the lower tolerance limit in case of a normal universe. For example, if $a = .99$, the efficiency is .18.

It is perhaps appropriate here to point out the distinction between confidence limits and tolerance limits. It is well-known that in a sample from a normal universe with mean m the probability is a that the confidence limits $\bar{x} \pm t_a s$ will include the population mean m between them. The tolerance limits $\bar{x} \pm t_a\sqrt{(n + 1)/n}\cdot s$, on the other hand are used to estimate the middle $100a\%$ of the universe. Although the tolerance limits $\bar{x} \pm t_a\sqrt{(n + 1)/n}\cdot s$ are much more stable for a given sample size than those given by x_r and x_{n-r+1}, in case of a normal distribution, it should be emphasized that in case of even slight non-normality, particularly when skewness is present, the former pair of limits are apt to give very erroneous results with reference to the proportion of the universe included in the tails. Confidence limits estimating m are probably much less sensitive to skewness than tolerance limits estimating the middle $100a\%$ of the universe, particularly when a is nearly unity.

Another important aspect of the problem of setting tolerance limits is the following: Suppose small samples of a given size are taken from a universe under statistical control. How many of these small samples should be taken as a basis for determining tolerance limits L_1 and L_2 of some function, say g, of the samples (e.g. the sum of the measurements in each sample) so that the proportion of samples in the universe of such samples having values of g between L_1 and L_2 will have a given mean with a given degree of stability? One obvious approach to this question is to consider a universe of samples in the same manner in which we have considered a universe of individuals throughout the present paper. This approach, however, does not make very efficient use of the observations, but we shall not enter into a treatment of the problem here. This problem and various related problems in the statistical methods of mass production remain to be studied.

4. Summary. A method based on truncated sample ranges for determining size of sample required for setting tolerance limits on a random variable x having any unknown continuous distribution $f(x)$ and having a given degree of stability is given. A method for setting tolerance limits corresponding to a given degree of stability in case $f(x)$ is normal is discussed and a comparison of the stabilities of the tolerance limits set by the two methods in the normal case is made. Illustrative examples of the methods are given.