STATISTICAL PREDICTION WITH SPECIAL REFERENCE TO THE PROBLEM OF TOLERANCE LIMITS[1]

By S. S. Wilks

Princeton University

1. Introduction. Statistical methodology is becoming recognized in industry as an effective tool for dealing with certain problems of inspection and quality control in mass production. Quality control experts have found statistical methods useful in detecting excessive variation in a given quality characteristic of a product from a series of observations on the given quality characteristic, and in isolating the causes of such variations back in the materials or operations involved in manufacturing the product. By a process of successive detection and elimination of causes of variability, a *controlled state of quality* is established. A practical statistical procedure for establishing a controlled state of quality has been developed by Shewhart.[2] More recently, manuals for routine application of this procedure have been issued by the American Standards Association.[3]

In this paper we do not propose to go into a discussion of the application of the well known Shewhart procedure. The reader may refer to the literature mentioned in footnotes 2 and 3 for such discussion. It is sufficient to remark that experience shows that the application of this procedure leads to a controlled state of quality. Such a state of control provides a basis for making statistical predictions about measurements on the given quality characteristic in future production.

More specifically, suppose a given quality characteristic of a given product is measured by a variable X, such that X has a specific value for each individual product-piece. For example, the product may be a given type of fuse and X may be the blowing time in seconds. A product-piece would be a single fuse, and X would take on a value for each fuse. Thus, for a sequence of n fuses taken from the production line, there would be a corresponding sequence of values of X, say X_1, X_2, $\cdots X_n$. If a state of control has been established with respect to blowing time as measured by X, then the sequence of values of X will "behave like a random sequence." By this we mean that the sequence will be such that we can safely assume that it can be described mathematically by regarding X as a *continuous random variable*, i.e., such that there exists some

---

[1] An expository paper presented at a joint session of the American Mathematical Society and the Institute of Mathematical Statistics at Poughkeepsie, September 9, 1942.

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

[3] *Guide for Quality Control and Control Chart Method of Analyzing Data* (*1941*), and *Control Chart Method of Controlling Quality During Production* (*1942*), American Standards Association, New York.

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *The Annals of Mathematical Statistics.* www.jstor.org

probability function $f(x)$ which describes the distribution of values of X, such that $\int_a^b f(x)\,dx$ is the probability that $a < X < b$ for any two real numbers a and b. Now, suppose we consider a sequence or *sample* S_1 of n values of X, and let X_1 and X_n be the smallest and largest values of X in the sequence. The types of questions with which we are concerned are the following: If a further sample, say S_2 of N values of X is taken, what is the probability P that at least N_0 of the values will lie between X_1 and X_2 as determined by S_1? If we choose a given probability α, at least what proportion of values of X in an indefinitely large sample S_2 will fall between X_1 and X_2 of S_1 with probability α? What is the probability P' that at least N_0 of the values of S_2 will exceed X_1 of S_1? At least what proportion of values of X in an indefinitely large sample S_2 will exceed X_1 with probability α? These questions suggest several of a more general nature which can be treated by methods similar to those which will be discussed. For example, instead of taking X_1 and X_n, i.e. the smallest and largest items in S_1 as *tolerance limits* we could use X_m and X_{n-m+1}. More generally, we may define *$100R_\alpha\%$ tolerance limits $L_1(x_1, x_2, \cdots x_n)$ and $L_2(x_1, x_2, \cdots, x_n)$ for probability level α* of a sample S_1 of size n from a population with distribution $f(x)\,dx$ as two functions of the X's in S_1 such that the probability is α that at least $100R_\alpha\%$ of the X's of a further indefinitely large sample S_2 (i.e. the population) will lie between L_1 and L_2. Or more briefly

$$P\left(\int_{L_1}^{L_2} f(x)\,dx \geq R_\alpha\right) = \alpha.$$

The same notion clearly applies if S_2 is a finite sample of size N, rather than an indefinitely large one. In this case we would be interested in the largest integer N_α such that the probability is at least α that at least $100\bar{R}_\alpha\%$ $\left(\bar{R}_\alpha = \frac{N_\alpha}{N}\right)$ of the X's in S_2 would lie between L_1 and L_2. In most practical situations we are able to assume nothing more about $f(x)$ than it is a probability density function. We make only this assumption here. The only functions of the values of X in S_1 that we shall consider here in setting tolerance limits are *order statistics*, i.e. the ordered values of X, because the results will then be fairly simple and independent of $f(x)$.

2. A General Probability Formula. It will be convenient perhaps to derive a general probability formula at this stage from which we can derive certain special cases as we need them.

Let $X_1, X_2, \cdots, X_n$ be the n values of X in S_1 arranged in order of increasing magnitude. Let $r_1, r_2, \cdots, r_k$ be integers such that $1 \leq r_1 < r_2 < \cdots < r_k \leq n$. Let $x_{r_1}, x_{r_2}, \cdots, x_{r_k}$ be k real numbers. Let

$$\int_{-\infty}^{x_{r_1}} f(x)\,dx = p_1, \int_{x_{r_1}}^{x_{r_2}} f(x)\,dx = p_2, \cdots, \int_{x_{r_k}}^{\infty} f(x)\,dx = p_{k+1},$$

from which

$$f(x_{r_1})\,dx_{r_1} = dp_1, \qquad f(x_{r_2})\,dx_{r_2} = dp_2, \cdots, f(x_{r_k})\,dx_{r_k} = dp_k.$$

Then assuming $X_1, X_2, \cdots, X_n$ to be a random sample (ordered) from a population with probability element $f(x)\,dx$ it follows from the multinomial distribution law[4] that the probability of $x_{r_i} < X_{r_i} < x_{r_i} + dx_{r_i}$ $(i = 1, 2, \cdots, k)$ is given by

$$(1) \qquad \frac{n!}{r_1 - 1!\, r_2 - r_1 - 1! \cdots r_k - r_{k-1} - 1!\, n - r_k!}\, p_1^{r_1-1} p_2^{r_2-r_1-1} \cdots p_k^{r_k-r_{k-1}-1} p_{k+1}^{n-r_k}\, dp_1\, dp_2 \cdots dp_k$$

except for terms of order higher than $(dp_1 dp_2 \cdots dp_k)$. Given that $X_{r_1} = x_{r_1}, \cdots, X_{r_k} = x_{r_k}$ in S_1, the conditional probability that $N_1, N_2, \cdots, N_{k+1}$ $\left(\sum_1^{k+1} N_i = N\right)$ of the values of X in S_2 will fall in the intervals $(-\infty, x_{r_1})$, $(x_{r_1}, x_{r_2}), \cdots, (x_{r_k}, \infty)$ respectively is by the multinomial law

$$(2) \qquad \frac{N!}{N_1!\, N_2! \cdots N_{k+1}!}\, p_1^{N_1} p_2^{N_2} \cdots p_{k+1}^{N_{k+1}}.$$

The joint probability law of $X_{r_1}, X_{r_2}, \cdots X_{r_k}$ and $N_1, N_2, \cdots, N_{k+1}$ $\left(\sum_1^{k+1} N_i = N\right)$ is given by the product of (1) and (2). Integrating this product with respect to the x's (i.e. the p's) we find the probability law of the N's to be

$$(3) \qquad \frac{N!\, n!\, N_1 + r_1 - 1!\, N_2 + r_2 - r_1 - 1! \cdots N_k + r_k - r_{k-1} - 1!\, N_{k+1} + n - r_k!}{r_1 - 1!\, r_2 - r_1 - 1! \cdots r_k - r_{k-1} - 1!\, n - r_k!\, N + n!\, N_1!\, N_2! \cdots N_{k+1}!}$$

which is clearly independent of $f(x)$. This result can be derived by direct combinatorial methods but the present derivation provides a simple proof that the result is independent of $f(x)$.

3. The Problem of One Tolerance Limit. There are problems in quality control in which it is important to consider only one tolerance limit. For example, in testing breaking strength of steel wire the most significant tolerance limit is the lower one. The problem of prediction in this case is as follows:

[4] Which states that if a trial results in one and only one of the mutually exclusive events $E_1, E_2, \cdots, E_k$, the probability P that in a total of n trials n_1 will result in E_1, n_2 in $E_2, \cdots, n_k$ in E_k $\left(\sum_1^k n_i = n\right)$, is given by

$$P = \frac{n!}{n_1!\, n_2! \cdots n_k!}\, p_1^{n_1} p_2^{n_2} \cdots p_k^{n_k}$$

where $p_1, p_2, \cdots, p_k$, $\left(\sum_1^k p_i = 1\right)$ are the probabilities of a single trial resulting in E_1, $E_2, \cdots, E_k$ respectively.

Suppose the given quality characteristic, as measured by X, is in a state of statistical control, and that a sequence of n measurements on X have been made. Let X_1 be the smallest of the n values. What is the probability that at least N_0 of N further measurements on X will exceed the value X_1 as determined by the initial sample? Instead of considering the smallest value of X as the lower tolerance limit we could just as easily choose the second smallest, or any other small *order statistic* but the case of the smallest value is perhaps of greater practical interest than any other case. The problem of an upper tolerance limit is entirely similar to that of a lower tolerance limit.

TABLE I

Values of N_α and $\bar{R}_\alpha$ for $\alpha = 0.99$ and 0.95 for several combinations of values of N and n, and for the problem of one tolerance limit. (For $N = \infty$, $\bar{R}_\alpha$ is denoted by R_α)

n	N	$\alpha = 0.99$		$\alpha = 0.95$	
		$N_{.99}$	$\bar{R}_{.99}$	$N_{.95}$	$\bar{R}_{.95}$
10	10	5	.500	7	.700
10	20	11	.550	14	.700
10	∞	—	.631	—	.741
50	50	44	.880	46	.920
50	100	90	.900	93	.930
50	∞	—	.912	—	.942
100	100	94	.940	96	.960
100	200	189	.945	193	.965
100	∞	—	.955	—	.970
500	500	494	.988	496	.992
500	1000	989	.989	993	.993
500	∞	—	.991	—	.994

The probability $P_1(N_0)$ that N_0 of the N further measurements will exceed the smallest value of X in an initially drawn sample of size n is given by (3) for $k = 1$, $r_1 = 1$, $N_2 = N_0$, $N_1 = N - N_0$, i.e.

$$P_1(N_0) = n \frac{N!\,\overline{N_0 + n - 1}!}{N_0!\,\overline{N + n}!}. \tag{4}$$

Values of $P_1(N_0)$ can be easily calculated by using the recursion formula

$$P_1(N_0 - 1) = \frac{N_0}{N_0 + n - 1} P_1(N_0). \tag{5}$$

For given values of N, n and α we are interested in the largest integer N_α for which

$$\sum_{N_0=N_\alpha}^{N} P_1(N_0) \geq \alpha. \tag{6}$$

If we set $\frac{N_\alpha}{N} = \bar{R}_\alpha$ and set $\lim_{N\to\infty} \bar{R}_\alpha = R_\alpha$ it can be verified that the value of R_α is given by solving the following equation for R_α

$$n \int_{R_\alpha}^{1} \xi^{n-1}\, d\xi = \alpha. \tag{7}$$

It will be observed that $n\xi^{n-1}\, d\xi$ is to within terms of order $d\xi$ the probability that $\xi < \int_{X_1}^{\infty} f(x)\, dx < \xi + d\xi$ in samples of size n from a distribution with probability element $f(x)\, dx$, where X_1 is the smallest value of X in the sample. The statistical interpretation of (7) is simply this: *The probability is α that the proportion of values of X exceeding X_1 in a further indefinitely large sample is at least R_α.*

Choosing $\alpha = 0.99$ and 0.95 Table I shows values of N_α and $\bar{R}_\alpha$ for various combinations of values of n and N for the case of one tolerance limit. The table indicates the degree of precision with which predictions about a single tolerance limit can be made from a sample of size n about a further sample of size N for a few important values of n and N. It should be noted that each prediction is made concerning a pair of samples, i.e. an initial sample of size n and a further sample of size N and that the prediction holds for any function $f(x)$. Thus as a typical entry we may state that if a sample of 100 is drawn and also a sample of 200, then the probability is 0.99 (approx.) that the X's of at least 189 (or 94.5%) of the cases in the second sample will exceed the smallest X in the first sample.

4. The Problem of Two Tolerance Limits. Again, suppose the given quality characteristic as measured by X is in a state of statistical control and that a sequence of n measurements are made on X. Let X_1 and X_n be the smallest and largest values of X respectively. The question to be considered now is the following: What is the probability that at least N_0 of N further measurements on X will lie between the values X_1 and X_n, as determined by the initial sample?

We proceed by considering the special case of (3) for which $k = 2$, $r_1 = 1$ $r_2 = n$, $N_2 = N_0$, $N_3 = N - N_0 - N_1$. We find for the joint distribution of N_1 and N_0

$$P(N_1, N_0) = \frac{N!\, n!\, \overline{N_0 + n - 2}!}{\overline{n-2}!\, N_0!\, \overline{N+n}!}. \tag{8}$$

To obtain the distribution of N_0, we simply sum (8) with respect to N_1 from 0 to $N - N_0$, thus obtaining

$$P_2(N_0) = n(n-1)(N - N_0 + 1)\frac{N!\,\overline{N_0 + n - 2}!}{N_0!\,\overline{N + n}!}. \tag{9}$$

A convenient recursion formula for computation purposes is

$$P_2(N_0 - 1) = \frac{N_0(N - N_0 + 2)}{(N - N_0 + 1)(N_0 + n - 2)} P_2(N_0). \tag{10}$$

For given values of N, n and α we require the largest value of N_α for which

$$\sum_{N_0 = N_\alpha}^{N} P_2(N_0) \geq \alpha. \tag{11}$$

Setting $\frac{N_\alpha}{N} = \bar{R}_\alpha$ and $\lim_{N\to\infty} \bar{R}_\alpha = R_\alpha$ one finds that R_α is given by solving the equation[5] for R_α

$$n(n-1)\int_{R_\alpha}^{1} \xi^{n-2}(1-\xi)\,d\xi = \alpha. \tag{12}$$

It can be verified that $n(n-1)\xi^{n-2}(1-\xi)\,d\xi$ is to within terms of order $d\xi$ the probability that $\xi < \int_{X_1}^{X_n} f(x)\,dx < \xi + d\xi$, thus showing that (12) is the probability that the proportion of an indefinitely large number of further values of X lying between X_1 and X_n is at least R_α.

Table II gives, for the case of two tolerance limits, values of N_α and R_α for several important combinations of n and N, including limiting values R_α of $\bar{R}_\alpha$ for indefinitely large N.

It should be noted that the problem of two tolerance limits can be immediately extended to the case where the lower and upper tolerance limits may be any two of the order statistics in S_1.

5. The Problem of Tolerance Limits for Two Quality Characteristics. We have thus far devoted our discussion to the problem of tolerance limits for a single quality characteristic. The problem of two or more quality characteristics can be treated by methods similar to those already used. The simplest case is that in which each product-piece under consideration is measured on two independent quality characteristics. Suppose the two characteristics are measured by X and Y. Let a sample of n product-pieces be taken, assuming a state of statistical control has been established, and let X_1 be the smallest of the X values and Y_1 the smallest of the Y values. The question with which we are

[5] This limiting case in the problem of tolerance limits as well as that expressed in (7) and other similar limiting cases have been considered by the author in an earlier paper: "Determination of Sample Sizes for Setting Tolerance Limits," *Annals of Math. Stat.* Vol. XII (1941) pp. 91-96.

concerned here is the following: If N further product-pieces are measured on X and Y, what is the probability that $X > X_1$ and $Y > Y_1$ for N_0 of the pieces? Let X and Y be statistically independent and let $f(x)$ and $g(y)$ be the probability functions of X and Y respectively. Let $\int_{-\infty}^{X_1} f(x)\,dx = p$ and $\int_{-\infty}^{Y_1} g(y)\,dy = q$. The probability law of p and q is

$$(13) \qquad n^2(1-p)^{n-1}(1-q)^{n-1}\,dp\,dq.$$

TABLE II

Values of N_α and $\bar{R}_\alpha$ for $\alpha = .99$ and $.95$ for several combinations of values of N and n and for the problem of two tolerance limits. (For $N = \infty$, $\bar{R}_\alpha$ is denoted by R_α)

n	N	$\alpha = 0.99$		$\alpha = 0.95$	
		$N_{.99}$	$\bar{R}_{.99}$	$N_{.95}$	$\bar{R}_{.95}$
10	10	4	.400	5	.500
10	20	8	.400	11	.550
10	∞	—	.496	—	.606
50	50	42	.840	44	.880
50	100	85	.850	90	.900
50	∞	—	.874	—	.909
100	100	89	.890	92	.920
100	200	184	.920	188	.940
100	∞	—	.935	—	.953
500	500	491	.982	494	.988
500	1000	985	.985	989	.989
500	∞	—	.987	—	.991

In a further sample of size N the probability that for N_0 of the cases, $X > X_1$ and $Y > Y_1$, X_1 and Y_1 being determined by the first sample, is

$$(14) \qquad \frac{N!}{N_0!\,N - N_0!}[(1-p)(1-q)]^{N_0}[1-(1-p)(1-q)]^{N-N_0}.$$

The joint probability law of N_0, p and q is given by the product of (13) and (14). Integrating this product with respect to p and q we obtain as the probability law of N_0,

$$(15) \qquad H_3(N_0) = n^2\binom{N}{N_0}\sum_{i=0}^{N-N_0}\binom{N-N_0}{i}\frac{(-1)^i}{(n+N_0+i)^2}.$$

For given values of N, n and α it is important, as before, to determine N_α as the largest integer for which

$$\sum_{N_0=N_\alpha}^{N} P_3(N_0) \geq \alpha. \tag{16}$$

Setting $\frac{N_\alpha}{N} = \bar{R}_\alpha$ and $\lim_{N\to\infty} \bar{R}_\alpha = R_\alpha$ one finds R_α to be given by solving the following equation for R_α

$$-n^2 \int_{R_\alpha}^{1} \xi^{n-1} \log \xi \, d\xi = \alpha. \tag{17}$$

The expression $-n^2\xi^{n-1} \log \xi \, d\xi$ is simply the probability that $\xi < \left(\int_{X_1}^{\infty} f(x)\, dx\right)\left(\int_{Y_1}^{\infty} g(y)\, dy\right) < \xi + d\xi$ to within terms of order $d\xi$, which is the proportion of the population pairs (X, Y) for which $X > X_1$ and $Y > Y_1$.

In the problem of two tolerance limits for each quality characteristic, as determined by an initial sample of size n, we calculate the probability that N_0 members of a further sample of size N will fall within the two sets of tolerance limits, with respect to the two characteristics. The problem is similar to that for one tolerance limit for each of two quality characteristics. For this case, we find corresponding to (15), (16), (17), respectively, the following:

$$P_4(N_0) = n^2(n-1)^2 \binom{N}{N_0} \sum_{i=0}^{N-N_0} \binom{N-N_0}{i} \frac{(-1)^i}{(N_0+n-1+i)^2(N_0+n+1)^2}, \tag{18}$$

and

$$\sum_{N_0=N_\alpha}^{N} P_4(N_0) \geq \alpha \tag{19}$$

and

$$n^2(n-1)^2 \int_{R_\alpha}^{1} \xi^{n-2}[2(\xi - 1) - (\xi + 1) \log \xi] \, d\xi = \alpha. \tag{20}$$

The derivations of results analogous to (15), (16), (17), (18), (19), (20) for tolerance limits defined by other order statistics than least and greatest and also for more than two independent[6] quality characteristics are straightforward.

6. Further Remarks and Discussion. For a given set of tolerance limits on a random variable X as determined by an initial sample of size n, we have discussed the problem of predicting, with a given degree of probability, at least what proportion of values of x in a further sample (finite or indefinitely large) will lie between these tolerance limits. We have obtained theoretical results

[6] In a paper to appear in a forthcoming issue of the *Annals of Math. Stat.*, A. Wald has shown how to set up tolerance limits for the case of two or more statistically *dependent* variables.

which depend only on the assumption that X is a continuous random variable with some probability element $f(x)\,dx$, where $f(x)$ is not assumed known.

It should be emphasized that the concept of a random variable is very broad in the sense that X may be a random variable determined as a result of calculations on other random variables. For example, X may be the difference, product, or ratio of two random variables, or the average or any other "reasonable" function of several random variables which may be of interest in any given situation. Thus, on the basis of an initial sample of differences of two random variables, we may set up tolerance limits of differences and make predictions, for a given probability level as to how many differences in a further sample of differences will lie between these tolerance limits. Similarly for products, ratios, and other functions of random variables.

From the point of view of practical application, we should again note that the mathematical assumption that X is a random variable means that a state of statistical control as described in §1 must exist in the measurements to which the tolerance limit prediction theory is to be applied. In practice X is often a discrete variable, i.e. one which can take on only certain isolated values. For example, if X is the number of defective product-pieces in a drawing of one product-piece, X is either 0 or 1, depending on whether the piece was non-defective or defective. Our theory would not be applicable to such a case. However, if we take as a new variable the average value of X for several product-pieces, we then obtain a variable that is continuous enough for the tolerance limit theory to be applicable for all practical purposes.

Finally, we remark that although we have used, as concrete examples, situations in mass production engineering, the notions of tolerance limits and predictions within tolerance limits which have been discussed apply equally well to situations in any branch of applied science where measurements are made and used as a basis for predictions concerning future measurements.

7. Summary. After a state of statistical control has been established with respect to a quality characteristic of product-pieces in mass production by the standard statistical quality control methods developed and refined by Shewhart and others, there remains the problem of determining the accuracy of predictions as to how many future product-pieces will fall within tolerance limits specified by measurements on product-pieces already produced under the given state of control. This problem and some of its extensions are discussed in the present paper.

More specifically, suppose an initial sample of n product-pieces, manufactured under a given state of statistical control, are measured with respect to a given quality characteristic. Let X be a variable which measures the given characteristic, so that X has a definite value for each product-piece. Let X_1 be the smallest and X_n the largest value of X which occurs in the initial sample. Now consider a further sample of size N. The following problems of prediction relating to the second sample from information yielded by the initial sample are

considered: (1) What is the probability that at least N_0 values of X in the second sample will exceed the *tolerance limit* X_1 set by the first sample? (2) What is the probability that at least N_0 values of X in the second sample will lie between the two *tolerance limits* X_1 and X_n set by the first sample? (3) For given values of n and N and α (e.g., .99 or .95), what is the largest integer N_α such that the probability is at least α that $N_0 \geq N_\alpha$? (4) What is the limiting value of $\frac{N_\alpha}{N} = \bar{R}_\alpha$ as N increases indefinitely? Tables of values of N_α and R_α are given for each of the two problems (1) and (2), for several important combinations of values of n and N and for $\alpha = .99$ and .95.

Problems similar to (1), (2) and (3) are discussed for the case in which tolerance limits are placed on two or more quality characteristics simultaneously.

The generality of the theory of tolerance limits and how it applies to differences, products and ratios and other functions of two or more random variables are briefly discussed.