

SOME NEW THEOREMS ON LIMITS OF VARIATION*

BY R. L. PEEK, JR.

1. *Introduction.* The application of statistical methods to engineering and inspection data makes it possible to describe the variability of such data in quantitative terms. It is often impractical to attempt a complete description of the variability, such as would be given by the frequency distribution of the quantity in question, and it suffices in many cases to state limits to the variations from the arithmetic mean such that variations exceeding these limits occur with a frequency less than some specified amount. In a practical methodology the evaluation of such virtual upper limits to the variations requires the use of expressions which are independent of the type of frequency distribution involved, as the form of the latter will often be unknown.

The best known expression for such an upper limit is that given by Tchebycheff's theorem. While the limits given by this theorem are as close as any that can be stated with perfect generality in terms of the standard deviation alone, they are much larger than those applying to most frequency distributions of common occurrence in engineering data. There is presented below a new theorem, as general in its application as Tchebycheff's, by means of which much closer limits can be placed on the relative frequency of results differing from the arithmetic mean by more than a specified amount. This more stringent inequality involves the use of both the standard deviation and the average deviation.

2. *Folded Distributions.* It will be convenient in the following discussion to employ the concept of *folding* a frequency distribution. By considering the original distribution as referred to the arithmetic mean as origin, any member of the distribution will be denoted by its deviation from the mean, x , and its relative frequency by p . The absolute values of these deviations $x_1 = |x|$, constitute a new frequency distribution for which

* Presented to the Society, June 22, 1933.

$p(x_1) = p(x) + p(-x)$. This derived distribution may be described as the *folded* distribution, for it may be obtained by folding a graphical representation of the original distribution about its arithmetic mean, and adding the superimposed ordinates. The arithmetic mean, \bar{x}_1 , of the *folded* distribution is, by definition, the average deviation of the original distribution.

Similarly, the *folded* distribution can itself be *folded*, yielding a second derived distribution, of which the members are denoted by $x_2 = |x_1 - \bar{x}_1|$, and which may be described as the second folded distribution. The standard deviation of the original distribution will be written as σ , that of the first folded distribution as σ_1 , and so on.

The inequalities presented below follow from a consideration of the limits of variation in these folded distributions. There is required a preliminary theorem, which may be proved by direct substitution in the expression for σ_1 of the expressions for σ and \bar{x}_1 . This theorem states that:

If σ is the standard deviation of any set of numbers (the original distribution), \bar{x}_1 their average deviation, and σ_1 the standard deviation of their absolute deviations from the mean (the folded distribution), then

$$(1) \quad \sigma^2 - \bar{x}_1^2 = \sigma_1^2 .$$

3. *The New Inequality.* With the aid of the preceding theorem, it can be shown, by a proof similar to that for Tchebycheff's theorem, that the following is true.

If σ is the standard deviation of any set of numbers, and $\rho = \bar{x}_1/\sigma$, where \bar{x}_1 is the average deviation, then the relative frequency, $1 - P_{t\rho}$, of numbers differing from the mean by more than $t\rho$ obeys (for $t > \rho$) the following inequality:

$$(2) \quad 1 - P_{t\rho} \leq \frac{1 - \rho^2}{t^2 - 2t\rho + 1} .$$

Considering the folded distribution, we see readily that σ_1^2 must exceed $(1 - P_{t\rho})(t\rho - \bar{x}_1)^2$, this being the least contribution to σ_1^2 that can be made by the fraction, $1 - P_{t\rho}$, of the numbers which exceed $t\rho$. Of the remaining fraction, $P_{t\rho}$, of the numbers, expressed as deviations from the mean, the absolute

values of some at least must be less than \bar{x}_1 . It will next be shown that the least contribution which the fraction $P_{t\sigma}$ of the numbers can make to σ_1 is that made in the special case in which all these numbers have the absolute value z_1 , where $P_{t\sigma}(\bar{x}_1 - z_1) = (1 - P_{t\sigma})(t\sigma - \bar{x}_1)$. This leaves the value of \bar{x}_1 unchanged, if all the numbers included in the fraction $1 - P_{t\sigma}$ have the absolute value $t\sigma$, and it is evident that if some of the numbers included in the fraction $1 - P_{t\sigma}$ have absolute values greater than $t\sigma$, z_1 must be smaller for \bar{x}_1 to be unchanged, and the contribution of the fraction $P_{t\sigma}$ to σ_1 will thereby be increased. It remains to be shown that if some of the numbers included in the fraction $P_{t\sigma}$ have absolute values other than z_1 , σ_1 will thereby be increased.

Suppose a group of numbers having a relative frequency q' to be changed in absolute value from z_1 to some other value z_1' . For \bar{x}_1 to be unchanged in value, another group of numbers having the frequency q'' must be changed to z_1'' , where

$$(3) \quad q'(\bar{x}_1 - z_1') + q''(\bar{x}_1 - z_1'') = (q' + q'')(\bar{x}_1 - z_1).$$

Now before the change, these numbers contributed to σ_1^2 an amount $(q' + q'')(\bar{x}_1 - z_1)^2$, while after the change they contribute an amount $q'(\bar{x}_1 - z_1')^2 + q''(\bar{x}_1 - z_1'')^2$, which, from (3), exceeds the former amount by $q'q''(z_1' - z_1'')^2 / (q' + q'')^2$. Hence any change from the special case cited results in an increase to the contribution to σ_1^2 made by the fraction $P_{t\sigma}$ of the numbers. It follows that

$$\sigma_1^2 \geq (1 - P_{t\sigma})(t\sigma - \bar{x}_1)^2 + P_{t\sigma} \left(\frac{(1 - P_{t\sigma})(t\sigma - \bar{x}_1)^2}{P_{t\sigma}} \right).$$

Substituting in this the value of σ_1 given by (1), and writing ρ for \bar{x}_1/σ , gives (2) above.

4. *A Corollary of the Preceding Theorem.* The preceding theorem has been obtained by determining the minimum value of σ_1 , the standard deviation of the folded distribution, composed of the absolute deviations from the mean of the original distribution. The same treatment may be applied to each of the distributions formed by taking successive folds. Thus the ap-

plication to the second folded distribution of the preceding argument gives the inequality

$$(4) \quad 1 - P_{t\sigma} \leq \frac{1 - \rho^2 - \rho'^2}{t^2 - 2t(\rho + \rho') + 2\rho\rho' + 1},$$

where $\rho' = \bar{x}_2/\sigma$. The argument may be readily generalized to apply to the distribution of the n th fold, and if ρ_n is written for \bar{x}_n/σ , a general expression for $1 - P_{t\sigma}$ may be obtained from (4) by writing the sum of the ρ 's for $\rho + \rho'$, the sum of their squares for $\rho^2 + \rho'^2$, and the sum of their paired products for $\rho\rho'$.

5. *Limits to Variation in a Continuous Distribution.* The preceding theorem is primarily an application to the quantity $\sigma^2 - \bar{x}_1^2$ of the argument leading to Tchybecheff's theorem. Similarly, an inequality applying to continuous frequency distributions may be obtained by applying to this same quantity the argument employed in developing the Camp-Meidell inequality. There may thus be obtained the following theorem.

For a frequency distribution such that the probability P of a result differing from the arithmetic mean by less than x_1 is a continuous function for which d^2P/dx_1^2 is always negative for $x_1 > \bar{x}_1$ (the average deviation), and for which σ is the standard deviation and $\rho = \bar{x}_1/\sigma$, we have

$$(5) \quad 1 - P_{t\sigma} < \frac{4(1 - \rho^2)}{9(t - \rho)^2}.$$

This is most readily proved by reference to a geometrical representation of the relation between P and x_1 , such as that shown in Fig. 1. As the slope of this curve is, by definition, the sum of the ordinates at x_1 and $-x_1$ of the original frequency distribution curve, σ_1^2 may be written

$$\sigma_1^2 = \int_0^1 (x_1 - \bar{x}_1)^2 dP.$$

Referring to Fig. 1, it is evident that the value of σ_1^2 will exceed the value of this last integral over the (shaded) area beneath the tangent to P at $t\sigma$. Writing the slope of this tangent in terms of the intercept P_0 on $x_1 = \bar{x}_1$, we have

$$(6) \quad \sigma^2 - \bar{x}_1^2 > \int_{P_0}^1 \frac{(t\sigma - \bar{x}_1)^2}{(P_{t\sigma} - P_0)^2} (P - P_0)^2 dP \\ = \frac{(t\sigma - \bar{x}_1)^2 (1 - P_0)^3}{3(P_{t\sigma} - P_0)^2}.$$

Now for any distribution meeting the conditions stated in the theorem, the above inequality will hold if P_0 is taken as having that value for which the integral of (6) is a minimum. This value of P_0 is, of course, that for which the partial deriva-

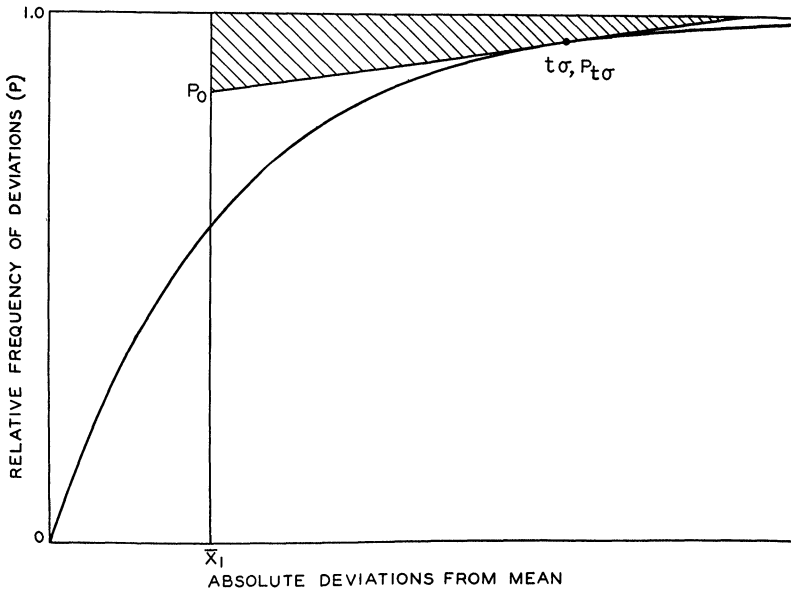


FIG. 1

tive of the right hand side of (6) with respect to P_0 is zero, and is thus found to equal $3P_{t\sigma} - 2$. Substituting this latter quantity in (6) gives the minimum value for σ_1 for any distribution meeting the conditions stated above, and thus leads directly to (5).

6. *Discussion.* The closeness of the limits given by (2) depends upon the value of ρ for the frequency distribution in question. This quantity can have any value from zero to unity. Values for several familiar types of distribution are listed in Table I.

TABLE I

Distribution	ρ	$1 - \rho^2$
Two numbers of Equal Frequency	1	0
Normal Law	$(2/\pi)^{1/2} = 0.798$	0.363
3 Term Gram Charlier Series	$(2/\pi)^{1/2} \left(1 - \frac{\beta_2 - 3}{24} \right)$	$\frac{\pi - 2}{\pi} + \frac{\beta_2 - 3}{12\pi} \left(2 - \frac{\beta_2 - 3}{24} \right)$
Rectangular	$3/2^{1/2} = 0.867$	0.250
Right Triangular	$(16/27) \cdot 2^{1/2} = 0.838$	0.298
Distribution for which (in above notation) $P = kx_1^n$	$(2n + 1)^{1/2} / (n + 1)$	$[n / (n + 1)]^2$

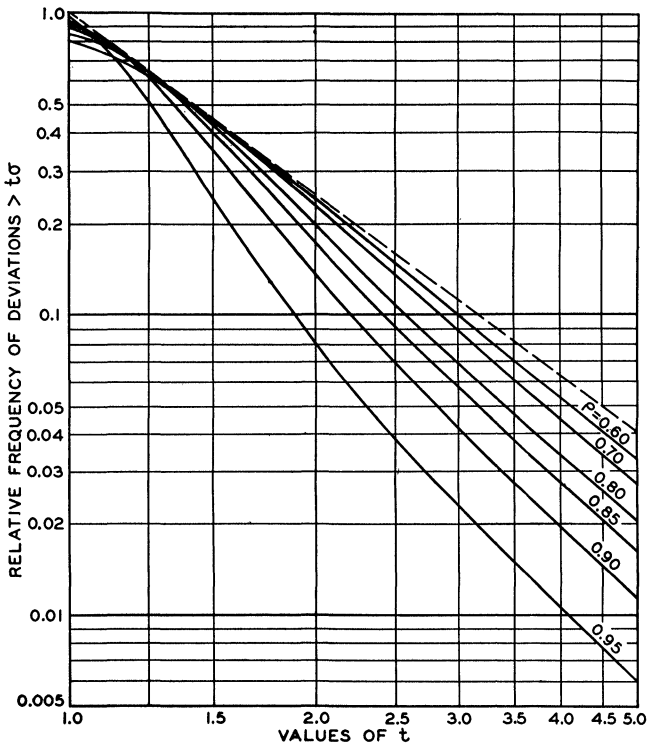


FIG. 2

Comparison of (2) with Tchebycheff's theorem shows that ρ is really a measure of the flatness of a distribution, and this is illustrated by the values of ρ listed in Table I. To afford a com-

parison of the limits given by (2) and by Tchebycheff's theorem, Fig. 2 has been prepared. This shows the values of $1 - P_t$ plotted against t , the several curves corresponding to various values of ρ as indicated. The dotted line gives the limits from Tchebycheff's theorem.

BELL TELEPHONE LABORATORIES

CHARACTERISTICS OF MULTIPLE CURVES AND THEIR RESIDUALS*

BY T. R. HOLLCROFT

Salmon† obtained formulas relating the characteristics of two curves which together form the complete intersection of two algebraic surfaces when one of the curves is double on one of the surfaces. In this paper, by a generalization of Salmon's method, the relations between the characteristics of two such curves are found when one of the curves is of given multiplicity on each of the two surfaces. Such a formula is useful in studying a system of surfaces with a multiple basis curve. It was this need for it that led to its derivation.

Consider two algebraic surfaces f_1 and f_2 of orders μ_1 and μ_2 , respectively, whose complete intersection consists of two curves C_1, C_2 of orders n_1, n_2 ; ranks r_1, r_2 ; genera p_1, p_2 ; and with h_1, h_2 apparent double points, respectively. Assume that C_1 is of multiplicity i_1 on f_1 and i_2 on f_2 and also that C_1 itself is the complete intersection curve of two surfaces. C_1 (counted simply) and C_2 have t actual intersections and $n_1 n_2 - t$ apparent intersections.

Consider a third surface f_3 of order μ_3 passing simply through C_1 but not through C_2 . The equivalence E of C_1 on the three surfaces f_1, f_2, f_3 is‡

$$E = n_1(i_2\mu_1 + i_1\mu_2 + i_1i_2\mu_3 - 2i_1i_2) - i_1i_2r_1.$$

*Presented to the Society, April 14, 1933.

† Salmon, *Geometry of Three Dimensions*, 4th ed., 1882, p. 322.

‡ M. Noether, *Sulle curve multiple di superficie algebriche*, *Annali di Matematica*, (2), vol. 5 (1871), p. 166.