and by induction on k we may show that for all $k = 1, 2, \cdots$,

$$\frac{(n+1)(n+3)\cdots(n+2k-1)}{n(n+2)\cdots(n+2k-2)} \leq 1 + k/n, \tag{29}$$

where the equality holds only for $k = 1$. Hence

$$F((n+1)/2, n/2, \lambda w) < 1 + \sum_{k=1}^{\infty} (1 + k/n)\cdot(\lambda w)^k/k! = e^{\lambda w}(1 + \lambda w/n), \tag{30}$$

$$R(t) < e^{-\lambda(1-w)} \cdot (1 + \lambda w/n) < e^{-\lambda(1-w)} \cdot e^{\lambda w/n} = e^{-\lambda[1-w(1+1/n)]}. \tag{31}$$

Hence $R(t) < 1$ if $w < n/(n+1)$, which is equivalent to (26).

REFERENCES

[1] H. Cramér, *Mathematical Methods of Statistics*, Princeton University Press, Princeton, 1946.

[2] P. C. Tang, "The power function of the analysis of variance tests with tables and illustrations of their use," *Stat. Res. Memoirs*, Vol. 2 (1938), pp. 127–149.

[3] Emma Lehmer, "Inverse tables of probabilities of errors of the second kind," *Annals of Math. Stat.*, Vol. 15 (1944), pp. 388–398.

A DISTRIBUTION-FREE CONFIDENCE INTERVAL FOR THE MEAN

By Louis Guttman

Cornell University

1. Summary. Consider a random sample of N observations $x_1, x_2, \cdots, x_N$, from a universe of mean μ and variance σ^2. Let m and s^2 be the sample mean and variance respectively:

$$m = \frac{1}{N}\sum_{i=1}^{N} x_i, \qquad s^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - m)^2. \tag{1}$$

It is shown that the following conservative confidence interval holds for μ:

$$\text{Prob}\,\{(m-\mu)^2 \leqq s^2/(N-1) + \lambda\sigma^2\sqrt{2/N(N-1)}\} > 1 - \lambda^{-2}, \tag{2}$$

where λ is any positive constant. Inequality (2) also holds if, in the braces, λ is replaced by $\sqrt{\lambda^2 - 1}$, with $\lambda \geqq 1$.

Inequality (2) is much more efficient on the average than Tchebychef's inequality for the mean, namely,

$$\text{Prob}\,\{(m-\mu)^2 \leqq \lambda^2\sigma^2/N\} > 1 - \lambda^{-2}, \tag{3}$$

yet (2) and (3) are both distribution-free, requiring only knowledge about σ^2. At the $1 - \lambda^{-2} = .99$ level of confidence, the expected value of the right member in the braces of (2) is only about 1/6 the corresponding member of (3); at the .999 level of confidence the ratio is about 1/20.

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

A more general inequality than (2) is developed, also involving only the single parameter σ^2.

2. Derivation. Consider the function

$$u = (m - \mu)^2 - s^2/(N - 1) - c\sigma^2, \tag{4}$$

where c is an arbitrary constant. It is easily verified that $Eu = -c\sigma^2$, and that

$$\underline{E}u^2 = \sigma^4[2/N(N - 1) + c^2]. \tag{5}$$

A basic feature of (5) is that the only population parameter in the right member is σ^2. Contrary to what might have been surmised, the fourth moment of x about μ is not involved, and indeed need not exist.

According to Tchebychef's inequality,

$$\text{Prob}\,\{-\lambda\sqrt{\underline{E}u^2} \leqq u \leqq \lambda\sqrt{\underline{E}u^2}\} > 1 - \lambda^{-2}, \tag{6}$$

where λ is an arbitrary positive number. Using (4) and (5), it is possible to write (6) as:

$$\begin{aligned}\text{Prob}\,\{s^2/(N - 1) + c\sigma^2 - \lambda\sigma^2\sqrt{2/N(N - 1) + c^2} &\leqq (m - \mu)^2 \\ \leqq s^2/(N - 1) + \sigma^2[c + \lambda\sqrt{2/N(N - 1) + c^2}]\} &> 1 - \lambda^{-2}.\end{aligned} \tag{7}$$

In the braces of (7), if the left member is negative, there is no harm in replacing it by zero; if it is positive, then replacing it by zero may only increase the probability of the braces. Regardless of the value of this left member, it is true that

$$\begin{aligned}\text{Prob}\,\{(m - \mu)^2 \leqq s^2/(N - 1) \\ + \sigma^2[c + \lambda\sqrt{2/N(N - 1) + c^2}]\} > 1 - \lambda^{-2}.\end{aligned} \tag{8}$$

If we set $c = 0$, we have inequality (2). Some improvement over (2) is obtained by determining c to minimize the right member in the braces of (8), yielding as the shortest confidence interval:

$$\text{Prob}\,\{(m - \mu)^2 \leqq s^2/(N - 1) + \sigma^2\sqrt{2(\lambda^2 - 1)/N(N - 1)}\} > 1 - \lambda^{-2}. \tag{9}$$

Inequality (9) differs from (2) only by replacing λ in the braces by $\sqrt{\lambda^2 - 1}$.

3. Comparison with Tchebychef's inequality. The expected value of the right member of the braces in (2) is

$$\sigma^2[1/N + \lambda\sqrt{2/N(N - 1)}]. \tag{10}$$

The ratio of (10) to the corresponding value of Tchebychef's inequality (3), namely $\lambda^2\sigma^2/N$, is

$$[1 + \lambda\sqrt{2N/(N - 1)}]/\lambda^2. \tag{11}$$

Since (11) decreases as λ increases, the efficiency of inequality (2) increases compared with that of Tchebychef as the level of confidence $1 - \lambda^{-2}$ increases. The squared interval of (2) involves only the first power of λ, while that of (3) involves the second power.

4. Approach to normality. If the fourth moment of the universe's distribution exists, then it is well known that the ratio of $\underline{E}(m - \mu)^4$ to σ^4/N^2 must approach 3—the ratio for the normal distribution—as N increases. That is, if $\alpha^2 + 1$ is the ratio, then $\lim_{N\to\infty} \alpha^2 = 2$. It is known[1] that Tchebychef's inequality can be replaced by one involving both α^2 and σ^2, and that

$$\text{Prob}\ \{(m - \mu)^2 \leqq \sigma^2(1 + \lambda\alpha)/N\} > 1 - \lambda^{-2}. \tag{12}$$

If $\alpha^2 = 2$, then the right member in the braces of (12) becomes $\sigma^2(1 + \lambda\sqrt{2})/N$. This is virtually the same as (10), the expected value from (2). In a sense, then, (2) implicitly takes account of the fact that the distribution of sample means approaches that of the normal distribution with respect to the fourth moment. A striking feature, however, is that (2) holds for any $N > 1$ and does not even presume the fourth moment of the universe to exist, whereas to set $\alpha = \sqrt{2}$ in (12) in general requires a large N and finite universe fourth moment.

5. Further possibilities. Confidence interval (2) is derived from but one of a series of general intervals, each of which depends only on σ^2. It may be possible to derive from this series even more efficient intervals, according to the method now to be outlined.

One way of arriving at (2) is to consider all products of the form $(x_i - \mu)(x_j - \mu)$, where $i > j$ and $i, j = 1, 2, \cdots, N$. Let p_2 be the mean of these $N(N - 1)/2$ products. It can easily be seen that $p_2 = u$ in (4) with $c = 0$, so that p_2 is a second degree polynomial in $m - \mu$, the coefficients being sample statistics. A more general quadratic would be $u_2 = p_2 + c_1p_1 + c_0$, where c_1 and c_0 are arbitrary constants and p_1 is the mean of the N values $(x_i - \mu)$ or $p_1 = m - \mu$. It is easily seen that $\underline{E}p_1 = \underline{E}p_2 = \underline{E}p_1p_2 = 0$, and that the only universe parameter involved in $\underline{E}p_1^2$ and $\underline{E}p_2^2$ is σ^2. Hence the only universe parameter upon which u_2^2 depends is also σ^2.

Higher degree polynomials in $m - \mu$ can be defined, possessing the same properties as u_2. Let p_3 be the mean of the $N(N - 1)(N - 2)/3!$ products of the form $(x_i - \mu)(x_j - \mu)(x_k - \mu)$, where $i > j > k$ and $i, j, k = 1, 2, \cdots, N$; etc.; and let $p_N = (x_1 - \mu)(x_2 - \mu) \cdots (x_N - \mu)$. Set $p_0 = 1$, and let

$$u_n = \sum_{a=0}^{n} c_{an}\, p_a \qquad (n = 1, 2, \cdots, N), \tag{13}$$

where the c_{an} are arbitrary constants. It is easily seen that $\underline{E}p_a = 0$ $(a > 0)$, $\underline{E}p_ap_b = 0$ $(a \neq b)$, and that each $\underline{E}p_a^2$ depends on only the parameter σ^2 as far

[1] See, for example, Louis Guttman, "An inequality for kurtosis," *Annals of Math. Stat.*, Vol. 19 (1948), pp. 277–278.

as the universe is concerned. Hence $\underline{E}u_n^2$ depends only on σ^2. Furthermore, by writing $x_i - \mu$ as $(x_i - m) + (m - \mu)$, it is seen that p_a is a polynomial of degree a in $m - \mu$, the coefficients being sample statistics. From (13), then, u_n is a polynomial of degree n in $m - \mu$ with statistics as coefficients.

According to Tchebychef's inequality,

$$\text{Prob}\ \{u_n^2 \leqq \lambda^2 \underline{E}u_n^2\} > 1 - \lambda^{-2}. \tag{14}$$

The interval for u_n^2 in the braces can be expressed in two statements:

$$f_n(m - \mu) = u_n - \lambda\sqrt{\underline{E}u_n^2} \leqq 0, \tag{15}$$

$$g_n(m - \mu) = u_n + \lambda\sqrt{\underline{E}u_n^2} \geqq 0. \tag{16}$$

Both f_n and g_n are polynomials of degree n in $m - \mu$, g_n exceeding f_n always by the additive constant $2\lambda\sqrt{\underline{E}u_n^2}$. Let q_n and Q_n be the smallest and largest real zeros respectively of f_n, and let r_n and R_n be the smallest and largest real zeros respectively of g_n.

For convenience, we can suppose that c_{nn}—the coefficient of $(m - \mu)^n$ in u_n—is positive. If n is even, then f_n is positive for $m - \mu > Q_n$ and for $m - \mu < q_n$. Hence the interval $q_n \leqq m - \mu \leqq Q_n$ contains all the points included in (15) and possibly more. Since the probability of (15) is not less than the probability of (14), we can write the following confidence interval:

$$\text{Prob}\ \{q_n \leqq m - \mu \leqq Q_n\} > 1 - \lambda^{-2} \qquad (n \text{ even}). \tag{17}$$

The problem remains to determine the c_{an} so as to minimize the expected value of $Q_n - q_n$. Inequality (9) provides the minimum for the case $n = 2$. This can be verified by adding the term c_1p_1 to u in (4) and finding that the minimum requires $c_1 = 0$.

If n is odd, we again may set $c_{nn} > 0$. Then $f_n > 0$ for $m - \mu > Q_n$, and $g_n < 0$ for $m - \mu < r_n$. The interval $r_n \leqq m - \mu \leqq Q_n$ thus contains at least all the points found jointly in (15) and (16) and hence forms a conservative confidence interval:

$$\text{Prob}\ \{r_n \leqq m - \mu \leqq Q_n\} > 1 - \lambda^{-2} \qquad (n \text{ odd}). \tag{18}$$

Again, the problem is to determine the c_{an} so as to minimize the expected value of $Q_n - r_n$. Tchebychef's inequality (3) does this for the case $n = 1$.

Although the only *population* parameter involved throughout is σ^2, the *sample* moments up to the nth order are present in (15) and (16). It thus seems plausible that improvement over inequality (9) should be possible for $n > 2$. To obtain such an improvement requires developing a distribution-free theory of the zeros of f_n and g_n beyond the quadratic case.