THE EFFECT ON A DISTRIBUTION FUNCTION OF SMALL CHANGES IN THE POPULATION FUNCTION

By Burton H. Camp

Wesleyan University

1. Summary. It is generally assumed in the application of distribution theory that, if the actual population function is not very different from the one used in the theory, then the true sampling distribution of a statistic will not be very different from the one obtained in the theory. But elsewhere in mathematics we do not assert that a conclusion will be only slightly modified by a small deviation in the hypothesis. This paper presents some theorems which are useful in determining the maximum effect on a sampling distribution of certain kinds of small changes in the population function. In particular, if the population is denoted by the function $\phi(t)$, if a sample of n independent measurements $(t_1, \cdots, t_n)$ is taken from this population, if a statistic $x = g(t_1, \cdots, t_n)$ is formed from the sample, and if $D(x)$ denotes the distribution of this statistic; then, when $\phi(t)$ is changed by a small proportionate amount to $\phi_1(t)$, $D(x)$ will be changed to $D_1(x)$, and the relation between D and D_1 will be subject to the inequality:

$$\left| \int_a^b (D - D_1)dx \right| \leqq \epsilon \int_a^b D(x)dx,$$

where

$$\epsilon = (1 + \delta)^n - 1, \qquad \text{and} \qquad | \phi_1/\phi - 1 | < \delta.$$

2. It is generally assumed in the application of distribution theory that, if the actual population function is not very different from the one used in the theory, then the true sampling distribution of a statistic will be not very different from the one obtained in the theory. For example, we commonly apply to practical problems the distribution theory that has been obtained on the hypothesis that the population is normally distributed even though we know that our actual populations are only approximately normal in form, and we commonly assume that our results are approximately correct. But elsewhere in mathematics we do not assert that a conclusion will be only slightly modified if we only slightly modify the hypothesis. An example of our unwillingness to do this in other branches of mathematics is illustrated in the following example.

Example 1. Let $y = \phi(t)$ have the derivative $y' = \phi'(t)$. Let $\phi(t)$ be replaced by $\phi_1(t)$, where $\phi_1 - \phi = s(t)\phi(t)$, and $| s(t) | \leq \epsilon$, ϵ being small. We have thus chosen to make $(\phi_1 - \phi)$ small relative to ϕ rather than small absolutely so that this example may be useful in another connection. The derivative of ϕ_1 may of course differ very greatly from $\phi'(t)$, as for example in some of the approximations made by a few terms of a Fourier series; and it would be a major error to assume that the two derivatives are approximately equal. How can we be sure that, in the process of finding a distribution function, we are not making an error of the same[1] sort?

The following theorems partly answer this question. The theorems will first be stated and proved in great generality. Then we shall return to the functions in Example 1 as a special case. We shall be concerned with a sample consisting of a single observation of n measurements $(t_1, \cdots, t_n)$ drawn from the multivariate universe $\psi(t_1, \cdots, t_n)$, or, more briefly, with the vector T as a sample from the n-way universe $\psi(T)$. Throughout this paper ψ and ψ_1 shall be functions which are non-negative and whose integrals over the entire spaces of their definition are unity. Let the statistics $(x_1, \cdots, x_m)$, or more briefly the vector X, be constructed from T thus:

$$x_1 = g_1(T), \cdots, x_m = g_m(T). \tag{1}$$

If now p represents any measurable point set in X space and if dX is used for $(dx_1 \cdots dx_m)$ and dT for $(dt_1 \cdots dt_n)$, a fundamental theorem [1] of distribution theory asserts that, if q is the point set in T space for which X is in p, then the distribution $D(X)$ is determined by the equation,

$$\int_p D(X)\,dX = \int_q \psi(T)dT, \text{ if these integrals exist.} \tag{2}$$

THEOREM 1. *Using the foregoing notation, let $\psi(T)$ be replaced by $\psi_1(T)$ and let $\psi_1(T) - \psi(T) = \psi(T)S(T)$, where $| S | \leq \epsilon$, and as a consequence let $D(X)$ be replaced by $D_1(X)$; then*

$$\left| \int_p D_1(X)dX - \int_p D(X)dX \right| \leq \epsilon \int_p D(X)dX \leq \epsilon. \tag{3}$$

To prove these inequalities we merely need to notice that the point set q depends on the g's but not on the universe, and that therefore we may use the same p and q as in (2) in the following equation which determines D_1:

$$\int_p D_1(X)dX = \int_q \psi_1(T)dT. \tag{4}$$

Subtracting (2) from (4) we obtain

$$\left| \int_p D_1\,dX - \int_p D dX \right| = \left| \int_p (D_1 - D)dX \right| = \left| \int_q (\psi_1 - \psi)dT \right| \tag{5}$$

$$= \left| \int_q \psi S dT \right| \leq \epsilon \left| \int_q \psi dT \right| = \epsilon \left| \int_p D dX \right| \leq \epsilon,$$

[1] The general question being raised here has been approached heretofore from different points of view. In particular, other exact population functions besides the normal have been studied, and in some cases the distribution theory has not been greatly disturbed as a result. Also, the effects of slight changes in the parameters of a population function have been studied.

since ψ is never negative, and the integral of D is never greater than unity. It should be noticed that the final inequality of (5) is independent of the g's, although this is not true of the preceding inequalities, which do depend on the g's because they involve p and q.

COROLLARY. *In particular*[2] *let* $\psi = \phi(t_1) \cdots \phi(t_n)$, *where* $\phi(t)$ *defines a one-way universe function, and* $t_1, \cdots, t_n$ *are independent samples from it. Let* $x = g(t_1, \cdots, t_n)$. *Then, if* $\phi(t)$ *is replaced by* $\phi_1(t)$, *and if* $\phi_1 - \phi = s(t)\phi(t)$, *and if* $| s(t) | \leq \delta$, *and if* $D(x)$ *is the distribution of* x *before the replacement, and* $D_1(x)$ *is the corresponding distribution after the replacement,*

$$\left| \int_a^b (D_1 - D)dx \right| \leq \epsilon \left| \int_a^b Ddx \right| \leq \epsilon,$$

where

$$\epsilon = (1 + \delta)^n - 1, \text{ and } -\infty \leq a < b \leq \infty.$$

This corollary follows from the theorem because of the universe,

$$\psi(t_1, \cdots, t_n) = \phi(t_1) \cdots \phi(t_n),$$

and

$$\psi_1(t_1, \cdots, t_n) = \phi(t_1) \cdots \phi(t_n)[1 + s(t_1)] \cdots [1 + s(t_n)],$$

so that, in the notation of the theorem,

$$\psi_1(T) = \psi(T) + \psi(T)S(T),$$

where

$$S(T) = [s(t_1) + \cdots + s(t_n)] + [s(t_1)s(t_2) + \cdots + s(t_{n-1})s(t_n)] + \cdots + [s(t_1) \cdots s(t_n)].$$

Hence

$$\left| S \right| \leq \left| n\delta + \frac{n!}{2!(n-2)!}\delta^2 + \cdots \delta^n \right| = (1 + \delta)^n - 1 = \epsilon.$$

The interval (a, b) now replaces the point set p of the theorem.

This theorem and its corollary are powerful in that they may be applied to all statistics, but they are weak because of the restrictions on $S(T)$ and $s(t)$. It is to be noted also that the corollary is ineffective when n is large, a difficulty which seems to the author to be implicit in the sampling process. The restrictions on $s(t)$ make it impracticable to apply the corollary to the following example since, as will be observed, if $| t | > c$, $\phi_1 - \phi = -\phi$, and so then $| s | = 1$; and when $\delta = 1$, $\epsilon = 2^n - 1$.

Example 2. Let $\phi(t) = (2\pi)^{-1/2}e^{-t^2/2}$ in $(-\infty, \infty)$, and let $\phi_1(t) = A(2\pi)^{-1/2} e^{-t^2/2}$ in $(-c, c)$ and let $\phi_1(t) = 0$ if $| t | > c$, where c is not infinite and A is so chosen that the integral of ϕ_1 over $(-\infty, \infty)$ is unity.

This type of example is important because, in the attempt to apply the theory of normal distributions to practical matters, the first discrepancy that appears

[2] One could as well use $\phi^1(t_1) \cdots \phi^{(n)}(t_n)$, but we choose the simpler case on account of its importance.

is that in the theory the given distribution is infinite in extent while in practice it is finite. The following theorem generalizes the preceding one so as to permit it to apply to this example.

THEOREM 2. *Let all of T-space be divisible into two parts, Q_0 and Q_1, satisfying the following conditions. In Q_0 let $\psi_1(T) - \psi(T) = S(T)\psi(T)$, and let $| S(T) | \leq \epsilon$. In Q_1 let $\psi_1(T) = 0$, and let*

$$\int_{Q_1} \psi(T)d(T) \leq \epsilon_1 .$$

Then

$$\left| \int_p D_1 dX - \int_p DdX \right| \leq \epsilon \int_p DdX + \epsilon_1 \leq \epsilon + \epsilon_1 .$$

It is not required that Q_0 or Q_1 be the totality of points for which its attendant conditions are true.

PROOF. As before, if the integrals exist,

$$\int_p D_1 dX = \int_q \psi_1 dT, \quad \text{and} \quad \int_p DdX = \int_q \psi dT.$$

Hence

$$\int_p D_1 dX - \int_p DdX = \int_q (\psi_1 - \psi)dT = \int_{q_0} (\psi_1 - \psi)dT + \int_{q_1} (\psi_1 - \psi)dT,$$

where q_0 is that part of q which is in Q_0, and q_1 is that part of q which is in Q_1.

$$(6) \qquad \left| \int_p D_1 dX - \int_p DdX \right| \leq \left| \int_{q_0} (\psi_1 - \psi)dT \right| + \left| \int_{q_1} (\psi_1 - \psi)dT \right| .$$

$$(7) \quad \left| \int_{q_0} (\psi_1 - \psi)dT \right| = \left| \int_{q_0} S\psi dT \right| \leq \epsilon \int_{q_0} \psi dT$$

$$\leq \epsilon \int_q \psi dT = \epsilon \int_p DdX, \text{ because } \psi \geq 0.$$

$$(8) \qquad \left| \int_{q_1} (\psi_1 - \psi)dT \right| = \left| \int_{q_1} \psi dT \right| \leq \left| \int_{Q_1} \psi dT \right| \leq \epsilon_1 ,$$

because $\psi_1 = 0$ in q_1. The inequalities (7) and (8), when substituted in (6), prove the theorem.

COROLLARY. *In particular, let ψ, and x be defined as in the corollary to Theorem 1, and let $\phi_1(t)$ be so defined that, if $| t | \leq c$, $\phi_1(t) - \phi(t) = s(t)\phi(t)$, where as before $| s(t) | \leq \delta$, and $\epsilon = (1 + \delta)^n - 1$; and, if $| t | > c$, let $\phi_1(t) = 0$. Also let $\int_{Q_1} \phi(t) \cdots \phi(t_n)dT \leq \epsilon_1$ where Q_1 is the set where $| t_i | > c$ for at least one value of i. Then*

$$\left| \int_a^b D_1(x)dx - \int_a^b D(x)dx \right| \leq \epsilon \int_a^b D(x)dx + \epsilon_1 \leq \epsilon + \epsilon_1 ,$$

provided these integrals exist.

PROOF. This corollary is implied in the theorem if we let $\psi(T) = \phi(t_1) \cdots \phi(t_n)$ and $\psi_1(T) = \phi_1(t_1) \cdots \phi_1(t_n)$, and then let Q_0 be the point set in T-space where $|t_i| \leq c$ for all values of i, and Q_1 be the point set where $|t_i| > c$ for at least one value of i. As in the corollary to Theorem 1, p becomes the interval (a, b).

Example 3. Let ϕ and ϕ_1 be as in Example 2, and choose $c = 3$. Then $A = 1/0.9973 = 1.0027$, and

$$\int_{Q_1} \phi(t_1) \cdots \phi(t_n) dT = 1 - (.9973)^n.$$

This quantity may be taken as ϵ_1. Also

$$|(\phi_1 - \phi)/\phi| = |A - 1| = 0.0027.$$

This quantity may be taken as δ. Then $\epsilon = (1.0027)^n - 1$. Hence

$$\left| \int_a^b D_1(x)dx - \int_a^b D(x) \right| dx \leq \epsilon \int_a^b D(x)dx + \epsilon_1 .$$

If n is not large, an approximate value for both ϵ and ϵ_1 is $0.003n$. This quantity is not particularly small unless n is small, but it could not be expected to be very small since the corollary pertains to all statistics of the form $x = g(t_1, \cdots, t_n)$.

Example 4. In one of the author's earlier papers [2] he found the distribution of the geometric mean, $x = (t_1 \cdots t_n)^{1/n}$, of n observations chosen from the universe described by the so-called curve of equal facility, whose equation is

$$y = \frac{1}{tc\sqrt{2\pi}} e^{-(1/2c^2)(\log t/G)^2}$$

The author stated that there was about as good justification for assuming that the distribution of statures was given by that universe as for assuming that it was normal. After one more theorem we shall now be able to state that, if one wishes to cling to the assumption that the distribution of statures is normal, then the distribution of the geometric mean is close to the distribution found in that earlier paper. We do need another theorem for this because we should be dealing with two distributions, $\phi_1(t)$ and $\phi(t)$, which do not obey the requirements of the corollary of Theorem 1, because they approach zero at different rates as t becomes infinite, and do not obey the requirements of the corollary of Theorem 2 because neither vanishes throughout the infinite intervals for which $|t| > c$. But the following theorem and corollary will take care of this and of similar cases. It will be observed that Theorem 3 includes Theorem 2 as a special case.

THEOREM 3. *Using the foregoing notation, let all of T-space be divisible into two parts Q_0 and Q_1 satisfying the following conditions. In Q_0 let $\psi_1(T) - \psi(T) = S(T)\psi(T)$, and let $|S(T)| \leq \epsilon$. Let $T = Q_0 + Q_1$ and*

$$\int_{Q_1} \psi_1(T)dT + \int_{Q_1} \psi(T)dT \leq \epsilon_1 .$$

Then

$$\left| \int_p D_1(X)dX - \int_p D(X)dX \right| \leq \epsilon \int_p D(X)dX + \epsilon_1 \leq \epsilon + \epsilon_1 .$$

PROOF. As before,

$$\int_p D_1 dX - \int_p DdX = \int_q (\psi_1 - \psi)dT = \int_{q_0} (\psi_1 - \psi)dT + \int_{q_1} (\psi_1 - \psi)dT$$

$$\left| \int_p D_1 \, dX - \int_p D \, dX \right| \leq \left| \int_{q_0} (\psi_1 - \psi) \, dT \right| + \left| \int_{q_1} (\psi_1 - \psi) \, dT \right| = \text{I} + \text{II}.$$

$$\text{I} \leq \epsilon \int_p D \, dX \leq \epsilon.$$

$$\text{II} \leq \int_{q_1} \psi_1 \, dT + \int_{q_1} \psi \, dT \leq \int_{Q_1} \psi_1 \, dT + \int_{Q_1} \psi \, dT \leq \epsilon_1 .$$

These inequalities together prove the theorem.

COROLLARY. *In particular, let ψ, ϕ_1, and x be as in the corollary of Theorem 2, except that now, instead of requiring $\phi_1(t)$ to vanish when $|t| > c$ we shall let Q_1 and ϵ_1 be so chosen that*

$$\int_{Q_1} \phi_1(t_1) \cdots \phi_1(t_n)dT + \int_{Q_1} \phi(t_1) \cdots \phi(t_n)dT \leq \epsilon_1 .$$

Then

$$\left| \int_a^b D_1(x)dx - \int_a^b D(x)dx \right| \leq \epsilon \int_a^b D(x)dx + \epsilon_1 \leq \epsilon + \epsilon_1 .$$

As before stated, the inequalities of this paper apply to all statistics for which the integrals involved exist. It seems probable that closer inequalities could be devised by placing appropriate restrictions on the g functions which define these statistics.

REFERENCES

[1] BURTON H. CAMP, "Methods of obtaining probability distributions," *Annals of Math. Stat.*, Vol. 8 (1937), pp. 90, 91.

[2] BURTON H. CAMP, "Notes on the distribution of the geometric mean," *Annals of Math. Stat.*, Vol. 9 (1938), pp. 221–226.