GENERALIZATION TO N DIMENSIONS OF INEQUALITIES OF THE TCHEBYCHEFF TYPE

By Burton H. Camp

Wesleyan University

1. Summary. The Tchebycheff statistical inequality and its generalizations are further generalized so as to apply equally well to n-dimensional probability distributions. Comparisons may be made with other generalizations [1], [2] that have been developed recently for the two-dimensional case. The inequalities given in this paper are generally as close as the most favorable corresponding inequalities that exist for the one-dimensional case and in many simple cases they are closer than those that have been given heretofore for two dimensions. In a special case the upper bound of our inequality is actually attained. The theory contains also a less important generalization in one dimension.

2. Introduction. It is necessary to introduce a new kind of moment, to be called a "contour" moment, which is a generalization of the usual one-dimensional moment. If we consider first a simple two-dimensional frequency surface, $y = f(t_1, t_2)$, we may think of y as a function of a single variable, x, where x is the area of the contour on that surface at the y level. This function may be defined so that it is monotonic decreasing and has other simple characteristics. Then we define the rth contour moment as

$$\hat{\mu}_r = \int_0^\infty x^r y \, dx,$$

and then the generalization of the Tchebycheff-type inequalities follows easily. This theory can be applied equally well to almost any single-valued function of n variables which is limited and integrable in the sense of Lebesgue. Therefore the theory will be enunciated initially in a very general form. The reasons for the initial statements will be indicated only briefly because a detailed discussion of quite similar ideas has been given by this author in another paper [3], where he applied the same general principle to obtain generalizations of certain theorems in integration theory.

3. Preliminary theory. Let $f(t_1, \cdots, t_n)$ be a probability distribution with limited upper bound L and defined at all points of infinite n-space, which is to be denoted by T, dT being the Lebesgue measure of a differential element. We thus assume that: $0 \leqq f(t, \cdots, t_n) \leqq L$, f has a Lebesgue integral in T, and

$$\int_T f \, dT = 1.$$

Let Q_λ denote the set of points in T where $f > \lambda$, $(0 \leqq \lambda \leqq L)$, and let x_λ be the measure of Q_λ, for Q_λ is known to be measurable. Therefore $x_L = 0$, $x_0 \leqq \infty$, and for each λ there exists a unique Q_λ and therefore a unique x_λ. This means that x is a single-valued function of λ and that it exists (or is positive infinite) for every value of λ in the interval $(0 \leqq \lambda \leqq L)$. If $\lambda' > \lambda$, $x_{\lambda'} \leqq x_\lambda$. This means that x is a monotonic decreasing function of λ. It need not be continuous; that is, it may be asymptotic to the line $\lambda = 0$, and it may have finite discontinuities or "jumps". Also there may be an enumerably infinite number of λ intervals in which x is constant. It follows that λ is a monotonic decreasing function of x in the interval $(0 \leqq x \leqq x_0 \leqq \infty)$, but it may not exist (in intervals where x has jumps), and it may be multiple valued (at points where x is constant). We now let $y(x) = \lambda_x$, except that: if λ is multiple valued at any point x we let y have the minimum value of λ at that point. Any other value would do equally well because the total measure of such points is zero and they can be left out of the integrals that follow. If λ does not exist in an x interval, we let y have in that interval the value which it has at the beginning of the interval. This is a λ point where x has a jump. We have thus defined y as a single valued monotonic decreasing function of x in the interval $(0 \leqq x \leqq x_0 \leqq \infty)$ and $0 \leqq y \leqq L$. It follows from Lebesgue's theory that:

$$\int_0^{x_\lambda} y(x)\,dx = \int_{Q_\lambda} f\,dT, \ (0 < \lambda \leqq L); \quad \int_0^{x_0} y(x)\,dx = \int_T f\,dT = 1.$$

Finally we restrict our function f so that there shall be at most a finite number of points x where λ is multiple valued (intervals of λ over which x is constant), and hence the number of discontinuities of y will be finite. This restriction may not be necessary but it is convenient and not embarrassing in applications.

4. Contour moments. The rth contour moment is denoted by $\hat{\mu}_r$. The contour standard deviation is denoted by $\hat{\sigma}$. We define

$$\hat{\mu}_r = \int_0^{x_0} x^r y\,dx.$$

It follows that $\mu_0 = 1$, and that

$$\hat{\mu}_2 = \hat{\sigma}^2 = \int_0^{x_0} x^2 y\,dx.$$

We shall also let $\hat{\alpha}_{2r} = \hat{\mu}_{2r}/\hat{\sigma}^{2r}$. We now assume that r is either zero or a positive integer, but in much of what follows this assumption is not necessary.

EXAMPLE 1. Let $f(t_1, t_2) = (2\pi)^{-1}e^{-(t_1^2+t_2^2)/2}$. The equation, $f(t_1, t_2) = \lambda$, defines a circular contour whose area is $x = \pi(t_1^2 + t_2^2) = -2\pi \log 2\pi\lambda$. Hence $y = \lambda = (2\pi)^{-1}e^{-x/2\pi}$, and

$$\hat{\mu}_r = \int_0^\infty x^r y\,dx = (2\pi)^r r!, \ \hat{\sigma}^2 = 8\pi^2, \ \hat{\alpha}_{2r} = (2r)!/2^r.$$

5. Contour moments and one-dimensional moments. If $n = 1$ and if $f(t_1) = f(-t_1)$, then

$$\hat{\mu}_{2r} = \int_0^{x_0} x^{2r} y\,dx = 2\int_0^{(x_0/2)} (2t)^{2r} f(t)\,dt = \mu_{2r} \cdot 2^{2r},$$

where μ_{2r} is an ordinary moment. Hence also $\hat{\sigma} = 2\sigma$, $\hat{\alpha}_{2r} = \hat{\mu}_{2r}/\hat{\sigma}^{2r} = \mu_{2r} \cdot 2^{2r}/\sigma^{2r} \cdot 2^{2r} = \alpha_{2r}$. It is to be noticed that, although $\hat{\alpha}_{2r} = \alpha_{2r}$, $\hat{\mu}_{2r} \neq \mu_{2r}$. One could alter the definition so that these two moments would be equal by inserting into the definition of contour moments the factor 2^n, using $x/2^n$ in place of x, but this would introduce a slight complication for a doubtful advantage. Although it would seem to be desirable to define the even contour moments $\hat{\mu}_{2r}$ so that they would become the ordinary moments μ_{2r} in the symmetrical one-dimensional case, such a definition would not make the two corresponding odd moments equal, and it would not make the two even moments equal in the non-symmetrical one-dimensional case. So it seems better not to introduce this factor 2^n, but to take note of the relationships that hold in the one-dimensional case.

THEOREM. *Let*

$$P_\delta = \int_{Q_\lambda} f \, dT,$$

where λ *is such that* $x_\lambda = \delta\hat{\sigma}$. *Then*

$$1 - P_\delta \leqq \hat{\alpha}_{2r} \Big/ \left(\delta \cdot \frac{2r+1}{2r}\right)^{2r}.$$

COROLLARY 1. *In particular* $1 - P_\delta \leqq \hat{\alpha}_{2r}/\delta^{2r}$.

COROLLARY 2. *If* $r = 1$, $1 - P_\delta \leqq 4/9\delta^2$. This theorem and these two corollaries are minor generalizations even of the corresponding one-dimensional inequalities, for it is no longer assumed that the probability distribution $f(t)$ has but one mode.

PROOF OF THEOREM. Let $g(x) = y(x)$ if $0 \leqq x \leqq x_0 \leqq \infty$, let $g(x) = y(-x)$ if $-\infty \leqq -x_0 \leqq x \leqq 0$, and let $g(x) = 0$ elsewhere in $(-\infty, \infty)$. Then $g(x)$ has all the properties explicitly required of $f(x)$ in a former paper by this author [4] in which this theorem was proved for the one-dimensional case. That is: $g(x)$ is a frequency function whose mean is zero, and

$$\int_{-\infty}^{\infty} g(x) \, dx = 2, \quad \text{and} \quad \int_{\delta\sigma}^{\infty} g(x) \, dx$$

is the probability that $|x| > \delta\sigma$; $g(x)$ is a monotonic decreasing function of $|x|$ for all values of x; and is symmetrical with respect to the central ordinate. Therefore, transforming the symbols of that paper to our present notation, we have

$$\int_{\delta\sigma}^{\infty} g(x) \, dx \leqq \alpha_{2r} \Big/ \left(\delta \cdot \frac{2r+1}{2r}\right)^{2r},$$

where

$$\sigma^2 = \int_0^{\infty} x^2 g \, dx = \int_0^{\infty} x^2 y \, dx = \hat{\sigma}^2.$$

Similarly $\mu_{2r} = \hat{\mu}_{2r}$, $\alpha_{2r} = \hat{\alpha}_{2r}$, and finally

$$1 - P_\delta = 1 - \int_0^{\delta\hat{\sigma}} y\,dx = 1 - \int_0^{\delta\sigma} g\,dx = \int_{\delta\sigma}^{\infty} g\,dx$$
$$\leqq \alpha_{2r} \Big/ \left(\delta \cdot \frac{2r+1}{2r}\right)^{2r} = \hat{\alpha}_{2r} \Big/ \left(\delta \cdot \frac{2r+1}{2r}\right)^{2r}.$$

This proves the theorem except that there is one exceptional case that requires attention. In the proof of the theorem in the paper just referred to the author assumed that the function corresponding to our present $g(x)$ was continuous. At that time a "frequency" function was often thought of as determined by a smooth curve approximating a histogram and implied even the existence of derivatives, and so continuity was not added to the explicit requirement that the function be a "frequency" function, but this condition was explicitly introduced in the lemma on which the proof of the theorem was based, and so we do now have to consider separately the case where y, and hence g, may have a finite number of jumps. It is quite easy to handle this case as the limiting form of a continuous case. In that lemma it was also required that d^2Q/dt^2 should exist and be non-negative, which would imply that we now have to make the requirement that y (corresponding to dQ/dt) shall have a non-negative first derivative. On examination of the proof, however, it will be observed that this is not necessary, since y is monotonic decreasing and continuous. That is, in the lemma the only use made of the condition, $d^2Q/dt^2 \geqq 0$, was that the function $Q(t)$ should determine a curve which would be never concave down. But for this it is sufficient that dQ/dt be continuous and monotonic increasing, and these conditions are now satisfied by the function which plays the rôle of Q in the present discussion. This function will now be defined as

$$\int_x^\infty \gamma(x)\,dx.$$

Let $\gamma(x)$ be a continuous function defined as equal to $g(x)$ except in the neighborhood of the points of finite discontinuity. Near such points it is to be so defined that it shall have all the properties just required of $g(x)$, and in addition so that, for any prescribed $R > 1$ and $\epsilon > 0$,

$$\int_0^\infty x^{2r}\gamma(x)\,dx = \int_0^\infty x^{2r} g(x)\,dx + \eta_r, \qquad (1 \leqq r \leqq R);$$
$$\int_x^\infty \gamma(x)\,dx = \int_x^\infty g(x)\,dx + \eta,$$

where $|\eta, \eta_r| < \epsilon$. It is obvious that such a definition of γ may be made in many ways, and one of them is by making use of a linear function in the neighborhood of each point of discontinuity. Since $\gamma(x)$ now satisfies all the conditions of the author's earlier paper the corresponding inequality is true:

$$\left(\int_{\delta\sigma_1}^\infty \gamma\,dx\right)\left(\delta \cdot \frac{2r+1}{2r}\right)^{2r}\left(\int_0^\infty x^2\gamma\,dx\right)^r \leqq \int_0^\infty x^{2r}\gamma\,dx,$$

where

$$\sigma_1^2 = \int_0^\infty x^2 \gamma \, dx.$$

Hence

$$\left(\int_{\delta\sigma_1}^\infty g \, dx - \eta\right)\left(\delta \cdot \frac{2r+1}{2r}\right)^{2r} (\hat{\sigma}^2 - \eta_1)^r \leqq \hat{\mu}_{2r} - \eta_r.$$

Let ϵ approach zero and we have, as desired:

$$1 - P_\delta \leqq \hat{\alpha}_{2r} \bigg/ \left(\delta \cdot \frac{2r+1}{2r}\right)^{2r}.$$

EXAMPLE 2. Let

$$f(t_1, \cdots, t_n) = A \exp\left\{-\frac{1}{2}\left(\frac{t_1^2}{\sigma_1^2} + \cdots + \frac{t_n^2}{\sigma_n^2}\right)\right\}, \qquad A = (2\pi)^{-n/2}(\sigma_1 \cdots \sigma_n)^{-1}.$$

This is a form into which the general correlation solid may be put by means of a linear transformation. Since P_δ is a ratio between two parts of such a solid and since this ratio is preserved under a linear transformation, the more general case may be transformed into this one, or even, as will appear shortly, into the simpler one where all the standard deviations are unity. If $f = \lambda$ the contour is the ellipsoid,

$$\frac{t_1^2}{\sigma_1^2} + \cdots + \frac{t_n^2}{\sigma_n^2} = -2 \log \frac{\lambda}{A}.$$

The volume of this ellipsoid is

$$x = h(-2 \log \lambda/A)^{n/2}, \qquad h = V_0 \sigma_1 \cdots \sigma_n, \; V_0 = \frac{2\pi^{n/2}}{n\Gamma(n/2)}.$$

Hence $y = Ae^{-\frac{1}{2}(x/h)^{2/n}}$, $\quad \hat{\mu}_r = \int_0^\infty x^r y \, dx$

$$= nAh^{r+1} 2^{n/2(r+1)-1} \Gamma\left(\frac{nr+n}{2}\right) = \left(\frac{\pi^{n/2} 2^{n/2+1} \sigma_1 \cdots \sigma_n}{n}\right)^r \frac{\Gamma\left(\frac{nr+n}{2}\right)}{[\Gamma(n/2)]^{r+1}}.$$

Putting $r = 2$ we obtain

$$\hat{\sigma}^2 = \frac{\pi^n 2^{n+2} (\sigma_1 \cdots \sigma_n)^2}{n^2} \frac{\Gamma(3n/2)}{[\Gamma(n/2)]^3},$$

and then

$$\hat{\alpha}_{2r} = \frac{\hat{\mu}_{2r}}{\hat{\sigma}^{2r}} = \frac{\Gamma\left(\frac{2rn+n}{2}\right)}{\Gamma(n/2)} \left[\frac{\Gamma(n/2)}{\Gamma(3n/2)}\right]^r.$$

Our inequality becomes: $1 - P_\delta \leqq J$, where

$$J = \frac{\hat{\alpha}_{2r}}{\left(\hat{\delta} \cdot \frac{2r+1}{2r}\right)^{2r}}, \text{ or 1, whichever is smaller.}$$

Typical numerical values of $\hat{\alpha}_{2r}$ and of J are given in Tables I and II.

TABLE I

Values of $\hat{\alpha}_{2r}$

n	$\hat{\alpha}_{2r}$	$\hat{\alpha}_2$	$\hat{\alpha}_4$	$\hat{\alpha}_6$
1	$1 \cdot 3 \cdots (2r-1)$	1	3	15
2	$(2r)!/2^r$	1	6	90
3	$3 \cdot 5 \cdot 7 \cdots (6r+1)/(3 \cdot 5 \cdot 7)^r$	1	12.26	566
4	$(4r+1)!/(5!)^r$	1	25.20	3604

TABLE II

Values of J

δ	n	r	J
1	1	1	0.444
		2	1.000
1	2	1	0.444
		2	1.000
2	1	1	0.111
		2	0.077
		3	0.093
3	1	1	0.049
		2	0.015
		3	0.008
		4	0.006
		5	0.006
3	2	1	0.049
		2	0.030
		3	0.049
3	3	1	0.049
		2	0.062
		3	0.308

Let us now compare J with the true value of $(1 - P_\delta)$ in one of these cases, *viz.*, when $\delta = 3$ and $n = 3$. The true value is given by

$$1 - P_3 = 1 - A \int_0^{3\hat{\sigma}} e^{-\frac{1}{2}(x/h)^{2/3}}\, dx,$$

where now $\hat{\sigma} = 4\pi\sqrt{105}(\sigma_1\sigma_2\sigma_3)/3$, $h = 4\pi(\sigma_1\sigma_2\sigma_3)/3$. The integral may be evaluated by means of the transformation, $t = (x/h)^{1/3}$ and a table of the integral of $(2\pi)^{-1/2}e^{-t^2/2}(t^2 - 1)$. We obtain: $1 - P_3 = 0.0205$. This is the true value to be compared with the approximation, $J = 0.049$. The closeness of this approximation is similar to that which may be obtained for the normal law by using the corresponding inequalities for one dimension. To illustrate this we find from the usual tables that, if for the normal law $1 - P_\delta = 0.0205$, $\delta = 2.32$. Hence the corresponding inequality is (for $r = 2$): $1 - P_\delta \leqq 0.042$.

We shall now show that the upper bound of our inequality is actually attained in a special case. Let $f(t_1, \cdots, t_n) = 2^{-n}$ in the region $(-1 \leqq t_1, \cdots, t_n \leqq 1)$, and let $f = 0$ elsewhere. For this case we shall have $x = 0$ when $\lambda = 2^{-n}$, and $x = 2^n$ when $0 \leqq \lambda < 2^{-n}$. Therefore $y = 2^{-n}$ if $0 \leqq x < 2^n$, and $y = 0$ if $2^n \leqq x$. Hence $\hat{\sigma} = 2^n/\sqrt{3}$, $\mu_0 = 1$, and the true value of $(1 - P_\delta)$ is $1 - \delta/\sqrt{3}$; and when $\delta = 2/\sqrt{3}$, this true value is $1/3$. The appropriate inequality is: $1 - P_\delta \leqq 4/9\ \delta^2$, and when $\delta = 2/\sqrt{3}$ the right hand side of this inequality is also equal to $1/3$. These relationships are true for all values of n.

REFERENCES

[1] P. O. Berge, "A note on a form of Tchebycheff's inequality for two variables," *Biometrika*, Vol. 29 (1937), pp. 405–406.

[2] Z. W. Birnbaum, J. Raymond, and H. S. Zuckerman," A generalization of Tchebycheff's inequality to two dimensions," *Annals of Math. Stat.*, Vol. 18 (1947), pp. 70–79.

[3] B. H. Camp, "A method of extending to multiple integrals properties of simple integrals," *Math. Ann.*, Vol. 75 (1914), pp. 274–289.

[4] B. H. Camp, "A new generalization of Tchebycheff's statistical inequality," *Amer. Math. Soc. Bull.*, Vol. 28 (1922), pp. 427–432. See also: "A note on Narumi's paper," *Biometrika*, Vol. 15 (1923), pp. 421–423.