where ψ is an arbitrary function.

From (5) it is evident that if A_{ij} are defined as the components of the curl of covariant vector, then (2) are necessarily satisfied; but (2) is not a sufficient condition. That this condition is not sufficient was overlooked by me in a recent paper,* and my conclusions in § 5 are not correct. In fact, the skew-symmetric tensor there defined by S_{ij} is given by

$$S_{ij} = \frac{\partial \Gamma^{\alpha}_{\alpha j}}{\partial x^i} - \frac{\partial \Gamma^{\alpha}_{\alpha i}}{\partial x^j},$$

and the functions $\Gamma^{\alpha}_{\alpha i}$ and $\Gamma'^{\alpha}_{\alpha i}$ in two sets of coordinates are in the relation

$$\Gamma'^{\alpha}_{\alpha i} = \Gamma^{\alpha}_{\alpha j} \frac{\partial x^j}{\partial x'^i} + \frac{\partial}{\partial x'^i} \log \Delta,$$

where Δ is the Jacobian of the transformation.

PRINCETON UNIVERSITY

A NEW GENERALIZATION OF TCHEBYCHEFF'S STATISTICAL INEQUALITY

BY B. H. CAMP

1. *Introduction.* If $f(x)$ is any frequency distribution, and s its standard deviation, the symbol $P(\lambda s)$ may be used to represent the probability that a datum drawn from this distribution will differ from the mean value by as much as λs, numerically. For the solution of various statistical problems it is desirable to have a formula which will measure $P(\lambda s)$ when $f(x)$ is only partially known. A case of practical importance occurs when $f(x)$ represents the distribution of values of a statistical constant determined by sampling from a known distribution, such a constant as, for example, a mean value, or a coefficient of correlation. In such cases it is usually difficult or impossible to find the complete distribution $f(x)$, but quite feasible to find its lower moments. Tchebycheff's well known inequality is: $P(\lambda s) \leqq 1/\lambda^2$. It has been general-

* PROCEEDINGS OF THE NATIONAL ACADEMY, vol. 8 (1922), p. 236.

ized in a formula first discovered by K. Pearson:*

$$P(\lambda s) \leqq \frac{\beta_{2r-2}}{\lambda^{2r}}, \qquad (r = 1, 2, \cdots),$$

where $\beta_{2r-2} = m_{2r}/s^{2r}$, m_{2r} being the $2r$th moment of $f(x)$ about its mean. Pearson's view is that, although in most cases this is a closer inequality than Tchebycheff's, it is usually not close enough to be of practical assistance. However, it is the best formula so far obtained which logically can be used with distributions whose larger even β's depart considerably from their Gaussian values. It is proposed to exhibit here a method by which the right side of Pearson's inequality may usually be decreased by about fifty per cent. The theory will be explained in § 2, and illustrated in § 3.

2. *Theory.* LEMMA. *Let $Q(t) \geqq 0$, and be monotonic, decreasing, and let $d^2Q/dt^2 \geqq 0$, in the interval $1 \leqq t \leqq k$. Let the straight line $y = a + bt$ pass through the points of Q for which $t = 1$ and $t = p$. Then*

$$\mathrm{I} \equiv \int_1^k t^{2r-1}Q(t)dt \geqq \int_1^k t^{2r-1}y dt \equiv \mathrm{II},$$

if $r \geqq 1$, $k \geqq 2$, $p = 2rk/(2r+1)$.

It will be observed that, by subtracting a constant from each side of the inequality, the proposition can be reduced to the case where $Q(k) = 0$. This being supposed done, let $h = Q(1)/Q(p)$, and note that by definition $p > 1$, $h \geqq 1$. After determination of the values of a and b in accordance with the conditions imposed on y in the hypothesis, II becomes

$$\text{(1)} \qquad \mathrm{II} = Q(p)\int_1^k t^{2r-1}\frac{hp - 1 + (1-h)t}{p-1}dt.$$

Consider now the function represented by the line tangent to the curve Q at $t = p$, that is,

$$\text{(2)} \qquad z = Q(p) + Q'(p)(t - p).$$

Since Q is concave upwards it never crosses this tangent, and so $Q(t) \geqq z(t)$, and

$$\text{(3)} \qquad \mathrm{I} \equiv \int_1^k t^{2r-1}Q(t)dt \geqq \int_1^k t^{2r-1}z(t)dt \equiv \mathrm{III}.$$

* BIOMETRIKA, vol. 12 (1919), p. 284.

In order to prove that I $\geqq$ II it is by (3) sufficient to show that III $\geqq$ II, and the lemma will now be established in this manner. First, fix arbitrarily all the parameters that occur in III, viz.: p, r, k, $Q(p)$, and $Q'(p)$, thus fixing the value of III. The value of II is also thus fixed, except so far as variation due to h is concerned. But, from (1),

$$\frac{d(\mathrm{II})}{dh} = \frac{Q(p)(1-k)}{(p-1)(2r+1)} \leqq 0,$$

and therefore, so far as variation with h is concerned, II is greatest when h is least, that is, when $h = 1$. But, then, $Q(1) = Q(p)$, and, since $Q(t)$ is monotonic, $Q(t) = Q(p)$ throughout the interval $1 \leqq t \leqq p$, and so $Q'(p) = 0$. Substituting this value for $Q'(p)$ in (2) and (3), and $h = 1$ in (1), we learn finally that, for this special value of h which makes II greatest, II = III. So, in general, II $\leqq$ III.

THEOREM: (*a*) *Let $f(x)$ be any frequency distribution, and suppose the origin and units to be so chosen that zero is its mean value and*

$$P(x) = \int_x^\infty f(x)dx.$$

(*b*) *Let $f(x)$ be a monotonic, decreasing function of $|x|$ when $|x| \geqq cs$, $c \geqq 0$.* This is a real restriction on $f(x)$, which varies in its severity according to the value of c. It will be discussed later. Its general effect is to except distributions of more than one mode, notably the "U" distributions.

(*c*) *Let $f(x)$ be symmetrical with respect to the centroid ordinate.* This may be assumed without loss of generality, for, if $g(x)$ is a skew distribution whose mean is at $x = 0$, it is clear that $f(x) = \frac{1}{2}[g(x) + g(-x)]$ has the same mean, even moments, and the same values of $P(\lambda s)$ as $g(x)$. *Then*

$$P(\lambda s) \leqq \frac{\beta_{2r-2}}{\left(\lambda \frac{2r+1}{2r}\right)^{2r}} \cdot \frac{1}{1+\phi} + \theta, \tag{4}$$

where

$$(5) \qquad \theta = \frac{P(cs)}{1 + \frac{1}{\phi}}, \qquad \phi = \frac{\left(\frac{c}{\lambda} \cdot \frac{2r}{2r+1}\right)^{2r}}{(2r+1)\left(\frac{\lambda}{c} - 1\right)},$$

quantities which can be tabulated and are usually very small.

Since $c \geqq 0$, we may write

$$\frac{\beta_{2r-2}}{c^{2r}} \geqq \int_{cs}^{\infty} \left(\frac{x}{cs}\right)^{2r} f(x)dx = -\left[\left(\frac{x}{cs}\right)^{2r} \int_{x}^{\infty} f(t)dt\right]_{cs}^{\infty} + \int_{cs}^{\infty} \frac{2rx^{2r-1}}{(cs)^{2r}} dx \int_{x}^{\infty} f(t)dt$$

$$\geqq P(cs) + \frac{2r}{cs}\int_{cs}^{kcs} \left(\frac{x}{cs}\right)^{2r-1} P(x)dx = P(cs) + 2r\int_{1}^{k} t^{2r-1}Q(t)dt,$$

where $x = tcs$, $Q(t) = P(tcs)$. Therefore, by the lemma, and by use of the relation $hQ(p) = Q(1)$,

$$\frac{\beta_{2r-2}}{c^{2r}} \geqq P(cs) + \frac{2rQ(p)}{p-1}\int_{1}^{k} t^{2r-1}[(hp-1) + (1-h)t]dt$$

$$= \frac{P(cs)}{1 + 2r - 2rk} + P(pcs)\left[\frac{(2r+1)(1-k^{2r}) + 2r(k^{2r+1}-1)}{1 + 2r - 2rk}\right].$$

After the substitutions $k = p(2r+1)/2r$, $p = \lambda/c$, and some reductions, the result (4) follows from the above inequality. The assumption (b) will now be discussed.

CASE I: ($c = 0$). It was stated under assumption (c) that $f(x)$ might be assumed symmetrical, because a symmetrical function $f(x)$ having the same essential characteristics could be constructed from any skew function $g(x)$ that might arise in practice. It should be noticed, however, that if $g(x)$ is skew, the resulting $f(x)$ may be saddle shaped, with a geometrical minimum at its mean and two symmetrically situated maxima on either side. In such a case, the value of c would equal the distance from the mean to either maximum, s being the unit of measurement. In other words, if skewness be defined as (mean $-$ mode)$/s$, *c may* be as large as the absolute value of the skewness. Now, skew functions are usually the ones that occur in practice, and so one could not usually assert, a priori, that $c = 0$ for a given distribution. This embarrassment might be avoided by choosing the origin at the mode of the skew function instead of at the mean, and by defining the moments and $P(\lambda s)$ with reference to such an origin, in which case, it could be shown, the above theorem would apply, and, for unimodal functions, c would equal zero.* But this is often impracticable, and certainly not customary. So it will be necessary to consider larger values of c. The following table for the case $c = 0$ is of value, however, since it is involved in the computation for the cases where $c > 0$. When $c = 0$, $\phi = \theta = 0$, and formula (4) is exactly Pearson's, except for the factor $[(2r + 1)/2r]^{2r}$ in the denominator. The values of this factor are as follows:

Formula	r	Factor	Formula	r	Factor
β_0	1	2.25	β_6	4	2.56
β_2	2	2.44	β_8	5	2.59
β_4	3	2.52	β_{10}	6	2.60

CASE II: ($c = 1$). If functions of more than one mode be excepted, a skewness greater than unity is very rare. Therefore, this may be regarded as a strong case, and Case III below, where $c = 2$, as an extreme case. To compute formula (4) we may use the table under Case I for the first fraction, and the tables below for ϕ and θ. The values given for θ are really the extreme values which θ has when $P(cs = s) = 1$. It is unnecessary to estimate them more closely. In this case, the theorem will be found of value only when $\lambda > 1$, but this is not a serious misfortune, since in practice it is most needed for about the range $2 < \lambda < 4$.

r	θ			ϕ		
	$\lambda = 2.5$	$\lambda = 3.0$	$\lambda = 3.5$	$\lambda = 2.5$	$\lambda = 3.0$	$\lambda = 3.5$
1	.015 6	.008 13	.004 28	.015 8	.008 23	.004 84
2	.001 40	.000 51	.000 22	ϕ equals θ to the number of places indicated		
3	.000 15	.000 039	.000 012			
4	$1.9 \cdot 10^{-5}$	$3.8 \cdot 10^{-6}$	$7.7 \cdot 10^{-7}$			

* While this paper was in press a paper by B. Meidel appeared, COMPTES RENDUS, vol. 175, p. 806 (Nov. 6, 1922), giving the result of this paper for the case $c = 0$, and using the mode as origin.

CASE III: ($c = 2$). The computation in this case will again be performed by use of the table under Case I, and of the subjoined table for ϕ and for $\theta/P(cs = 2s)$. Then Pearson's inequality should be used to estimate $P(2s)$. In this case the theorem will be found of value only when $\lambda > 2$.

r	$\theta/P(2s)$			ϕ		
	$\lambda = 2.5$	$\lambda = 3.0$	$\lambda = 3.5$	$\lambda = 2.5$	$\lambda = 3.0$	$\lambda = 3.5$
1	.275	.116	.060 6	.379	.132	.065
2	.118	.031 4	.011 5	.134	.032	.012
3	.056 1	.009 88	.002 63	.059	.010	.003
4	.028 3	.003 38	.006 58	.029	.003	.001

3. *Example.* In BIOMETRIKA * the frequency distributions of a number of coefficients of correlation are actually computed. One may apply the above inequalities to these distributions (using the β's as if they only, not the entire frequency distributions, had been calculated), and then compare the results obtained with the true values of $P(\lambda s)$ to be found by partially summing the frequencies there tabulated. Take $\lambda = 3$ in the distribution for which $\rho = 0.8$, $n = 100$, on page 403. If the distribution be regarded as a set of rectangles, the true value of $P(3s)$ is about 0.009. The β_2 formula shows that $P(3s) \leqq 0.018$, and the β_4 formula that $P(3s) \leqq 0.012$. Case II has been used on the theory that the practical statistician would be confident that $c \leqq 1$, but the use of Case III would increase the better of the two results given to only 0.014. Here, although β_2 departs from its Gaussian value by only 0.42, it is not safe to use the Gaussian table, which would make $P(3s) = 0.0027$. The reason is that the higher β's depart much more widely from their Gaussian values. They are, roughly, $\beta_2 = 3.42$, $\beta_4 = 22.7$, $\beta_6 = 290$, $\beta_8 = 5000$, $\beta_{10} = 100\,000$; instead of 3, 15, 105, 945, and 10 395, respectively. It is a common, but unwarranted and sometimes very faulty practice, to use the Gaussian where β_2 only is near its Gaussian value.

WESLEYAN UNIVERSITY

* Vol. 11 (1915–17), pp. 379–404.