ON THE TCHEBYCHEF INEQUALITY OF BERNSTEIN

By

CECIL C. CRAIG[1]

From Tchebychef's inequality we know that if $x_1, x_2, \cdots, x_n$ are a set of independent statistical variables with

$$m_{x_1} = m_{x_2} = \cdots\cdots = m_{x_n} = 0,$$

and

$$\sigma^2 = \sigma_{x_1}^2 + \sigma_{x_2}^2 + \cdots\cdots + \sigma_{x_n}^2,$$

then the probability P that

$$-t\sigma \leq x_1 + x_2 + \cdots + x_n \leq t\sigma$$

satisfies the inequality,

$$P \geq 1 - \frac{1}{t^2} \cdot$$

This gives a lower limit for P which is often unsatisfactory. Improvement of this result requires further hypotheses. As is well-known, Pearson, Camp, Guldberg, Meidel, Narumi,[2] and Smith[3] have attacked this problem with considerable success. Another interesting and important attempt in this direction due to S. Bernstein seems to have generally escaped attention in the English-speaking world, at least, since it has been published only in Russian.[4] Because of the latter fact, it seems necessary to give

---

[1]This paper was written in substantially its present form during the author's tenure of a National Research Fellowship at Stanford University.

[2]For references to all these papers except Smith's and a brief discussion see Rietz, H. L., Mathematical Statistics, (Open Court Publishing Company, Chicago, 1927), pp. 140-144.

[3]Smith, C. D., On Generalized Tchebychef Inequalities in Mathematical Statistics, American Journal of Mathematics, Vol. 52, (1930), pp. 109-126.

[4]Bernstein, S., Theory of Probability, (Moscow, 1927), pp. 159-165. The present account of this work of Bernstein is taken from a lecture of Professor J. V. Uspensky.

a brief account of this work of Bernstein's preliminary to the remarks based on it the writer wishes to make.

Bernstein imposed the condition in addition that

$$(1)\qquad E(|x_i|^k) \leq \frac{\sigma_{x_i}^2}{2} k!\, h^{k-2};\ k \geq 2,\quad i = 1, 2, \cdots\cdots, n,$$

($E(x)$ is read "the mathematical expectation of x.") in which h is arbitrary. (This condition is satisfied, e.g., if the x_i's are bounded.) and used the following lemma due to Tchebychef. Let the statistical variable u be always > 0. If $E(u) = A$, then the probability Q that $u \geq At^2$ satisfies the inequality, $Q \leq \frac{1}{t^2}$

Then taking,

$$u = e^{\varepsilon(x_1 + x_2 + \cdots\cdots + x_n)},$$

$$= e^{\varepsilon x_1} e^{\varepsilon x_2} \cdots\cdots e^{\varepsilon x_n},$$

in which ε is arbitrary,

$$E(u) = E(e^{\varepsilon x_1}) E(e^{\varepsilon x_2}) \cdots\cdots E(e^{\varepsilon x_n}).$$

Now

$$e^{\varepsilon x_i} = 1 + \varepsilon x_i + \frac{\varepsilon^2 x_i^2}{2!} + \frac{\varepsilon^3 x_i^3}{3!} + \cdots\cdots,$$

and under the condition (1),

$$E(e^{\varepsilon x_i}) \leq 1 + \frac{\varepsilon^2 \sigma_{x_i}^2}{2} + \frac{\varepsilon^3 \sigma_{x_i}^2 h}{2} + \frac{\varepsilon^4 \sigma_{x_i}^2 h^2}{2} + \cdots\cdots$$

If it is assumed that

$$|\varepsilon| h \leq c < 1$$

then

$$E(e^{\varepsilon x_i}) \leq 1 + \frac{\varepsilon^2 \sigma_{x_i}^2}{2(1-c)} < e^{\frac{\varepsilon^2 \sigma_{x_i}^2}{2(1-c)}},$$

and thus

$$(2)\qquad E(u) < e^{\frac{\varepsilon^2 \sigma^2}{2(1-c)}}$$

If in the inequality, $u \geq At^2$, a greater quantity is substituted for A, then certainly $Q \leq \frac{1}{t^2}$. Therefore the probability Q of

$$u \geq e^{\frac{\varepsilon^2\sigma^2}{2(1-c)}} e^{\tau^2}$$

satisfies the inequality

$$Q \leq e^{-\tau^2}$$

Now

$$u = e^{\varepsilon(x_1 + x_2 + \cdots x_n)} \geq e^{\tau^2 + \frac{\varepsilon^2\sigma^2}{2(1-c)}}$$

implies for $\varepsilon > 0$,

$$x_1 + x_2 + \cdots + x_n \geq \frac{\tau^2}{\varepsilon} + \frac{\varepsilon\sigma^2}{2(1-c)}.$$

The value of ε is next chosen so as to make Q a minimum, i.e., so as to make $\frac{\tau^2}{\varepsilon} + \frac{\varepsilon\sigma^2}{2(1-c)}$ a minimum. Thus

$$\varepsilon^2 = \frac{2(1-c)\tau^2}{\sigma^2}.$$

Then the probability Q that

$$x_1 + x_2 + \cdots + x_n \geq \tau\sigma\left(\frac{2}{1-c}\right)^{\frac{1}{2}}$$

satisfies the inequality,

$$Q \leq e^{-\tau^2}$$

if $\varepsilon^2 = \frac{2(1-c)\tau^2}{\sigma^2}$; $\varepsilon \leq \frac{c}{h}$ with $c < 1$.

To get the corresponding result for the lower limit of the sum $x_1 + x_2 + \cdots\cdots + x_n$, it is only necessary to choose $\varepsilon < 0$ and as before, the probability, Q', that

$$x_1 + x_2 + \cdots + x_n \leq -\tau\sigma\left(\frac{2}{1-c}\right)^{\frac{1}{2}}$$

satisfies the inequality,

$$Q' \leq e^{-\tau^2}$$

if also $\varepsilon^2 = \frac{2(1-c)}{\sigma^2}\tau^2$ and $|\varepsilon| \leq \frac{c}{h}$ with $c < 1$.

Combining these two results, if P is the probability of

$$-\tau\sigma\left(\frac{2}{1-c}\right)^{\frac{1}{2}} \leq x_1 + x_2 + \cdots + x_n \leq \tau\sigma\left(\frac{2}{1-c}\right)^{\frac{1}{2}}$$

then since

$$P + Q + Q' = 1,$$

$$P \geq 1 - 2e^{-\tau^2}.$$

But setting

$$\tau\sigma\left(\frac{2}{1-c}\right)^{\frac{1}{2}} = \omega,$$

and also,

$$\varepsilon^2 \leq \frac{c^2}{h^2},$$

the condition

$$\frac{2(1-c)}{\sigma^2}\tau^2 \leq \frac{c^2}{h^2}$$

(Bernstein set $\varepsilon^2 = \frac{c^2}{h^2}$, using merely the equality sign in this condition. The value of c as here given is necessary in the author's developments below.) must be satisfied, or what is the same thing,

$$\frac{2(1-c)^2\omega^2}{2\sigma^4} \leq \frac{c^2}{h^2},$$

from which

$$c \geq \frac{h\omega}{\sigma^2 + h\omega}.$$

This last quantity on the right is positive and < 1 as required so that the constants can actually be chosen as specified.

This gives

$$\tau = \omega\left[2(\sigma^2 + h\omega)\right]^{-\frac{1}{2}},$$

and finally the probability, P, that

$$-\omega \leq x_1 + x_2 + \cdots\cdots x_n \leq \omega$$

is such that

$$P \geq 1 - 2e^{-\frac{\omega^2}{2\sigma^2 + 2h\omega}},$$

or setting $\omega = t\sigma$

$$(3) \qquad P \geq 1 - 2e^{-\frac{t^2}{2 + \frac{2ht}{\sigma}}}.$$

It is to be observed that generally the quantity $\frac{2ht}{\sigma}$ rapidly decreases with increasing n.

This is the inequality reached by Bernstein under the condition (1).

If all the x_i's are bounded, if, say, always

$$|x_i| \leq b, \quad i = 1, 2, \cdots\cdots, n,$$

one may take $h = \frac{b}{3}$.

It is the purpose of the author's remarks to discuss less severe conditions than (1) under which the inequality (3) can be obtained. These more general conditions are obtained, however, at the expense of assuming quite generally satisfied regularity conditions with regard to the "tails" of the frequency distribution of x, which needs not necessarily to be regarded as the sum of n component variables, $x_1, x_2, \cdots\cdots, x_n$.

If we now take

$$(4) \qquad u = e^{\varepsilon x}$$

we have

$$E(u) = \int_{-\infty}^{\infty} dF(x) e^{\varepsilon x} \quad (F(x) \text{ is the probability function of } x).$$

$$= \int_{-\infty}^{\infty} dF(x)\left(1 + \varepsilon x + \frac{\varepsilon^2 x^2}{2!} + \frac{\varepsilon^3 x^3}{3!} + \cdots\right).$$

The condition (1) insures that the series under the sign of integration may be integrated over the interval $(-\infty, \infty)$. But the series can also be integrated over the same interval if it converges uniformly in any fixed finite interval, which it does, and if the series $\sum_{n=0}^{\infty} g_n(y)$, where

$$g_n(y) = \int_{-y}^{y} dF(x) \frac{\varepsilon^n x^n}{n!},$$

converges uniformly in the interval $(-\infty, \infty)$.

Formally, at least,

$$E(u) = 1 + \mu_2 \frac{\varepsilon^2}{2!} + \mu_3 \frac{\varepsilon^3}{3!} + \cdots \quad , \tag{5}$$

in which μ_k is the k-th moment about the mean of x . If

$$|\mu_k| \leq \frac{k!}{2} \mu_2 h^{k-2}, \; k \geq 2, \tag{6}$$

for some $h > 0$, then for $h\,|\varepsilon| \leq c < 1$ the right hand side of (5) is convergent and is $\leq 1 + \frac{\varepsilon^2 \sigma^2}{2(1-c)}$ as before. Now let us suppose that the condition (6) is satisfied not only for moments taken over the whole interval $(-\infty, \infty)$ but also for moments taken over any interval which includes the interval $(-b, b)$ in which b is an arbitrarily large though finite number. This is the *first regularity condition*, mentioned above, which we shall impose on the tails of the frequency function of x .

Then it is obvious, from the remark above, that

$$\sum_{n=0}^{\infty} g_{n(y)}$$

is uniformly convergent in the interval for $|y| \leq b$ for $\frac{b}{3}|\mathcal{E}| \leq c < 1$
And for $|y| > b$ it is also obvious that for $h|\mathcal{E}| \leq c < 1$,

$$\sum_{n=0}^{\infty} g_n(y)$$

is uniformly convergent if our first regularity condition is satisfied. And since $|\mathcal{E}|$ may be taken arbitrarily small, the inequality (3) follows as before.

It is evident that if our first regularity condition holds, that the condition (6) is more general than the condition (1). And it is easily seen that this first regularity condition holds for a very wide class of frequency functions. For, in order for it to hold, it is sufficient that the frequency curve (continuous or not) outside some finite interval $(-b, b)$ about the mean as center, be never increasing as $|x|$ increases and that if $f(x)$ be the ordinate of the frequency curve at the abscissa x, always $f(x) \geq f(-x)$ or else always $f(x) \leq f(-x)$ for $x > b$.

But if the first regularity condition be satisfied, then for all intervals which include $(-b, b)$ the corresponding moments have upper limits in absolute value. And if this be so for all such intervals, the semi-invariants (of Thiele) will also have upper limits for their absolute values. If λ_k is the k-th semi-invariant, we will take for our *second regularity condition* on the tails of the frequency distribution of x, that

$$(7) \qquad |\lambda_k| \leq \frac{k!}{2} \lambda_2 h^{k-2} \quad k \geq 2 \quad (\lambda_2 = \mu_2)$$

for some $h > 0$ if λ_k is taken for any interval which includes the arbitrarily large, though finite, interval $(-b, b)$.

If this second regularity condition holds, it is again easy to show that (5) is an equality if $h|\mathcal{E}| \leq c < 1$. The right member of (5) is still uniformly convergent in the interval $(-b, b)$ for $\frac{b}{3} |\varepsilon| \leq c < 1$. For all intervals which include $(-b, b)$ we use the formal identity which defines the semi-invariants of Thiele:

$$(8) \quad e^{\lambda_2 \frac{\varepsilon^2}{2!} + \frac{1}{3!}\lambda_3 \frac{\varepsilon^3}{3!} + \cdots}$$

$$= 1 + \mu_2 \frac{\varepsilon^2}{2!} + \mu_3 \frac{\varepsilon^3}{3!} + \cdots = e^{\phi(\varepsilon)}$$

Under the condition (7), $\phi(\varepsilon)$ is uniformly convergent over the intervals in question for $h|\varepsilon| \leq c < 1$ and for these values of ε, (8) becomes an equality since its second member is only the first arranged in powers of ε. Moreover, on account of (7) the right member must be uniformly convergent for all intervals which include $(-b, b)$.

At least one important class of frequency distributions satisfies our second regularity condition. The distributions of characteristics in samples of N have finite ranges as long as N is finite and they commonly have semi-invariants which are rapidly decreasing with increasing N. If such distributions approach normality their semi-invariants of order above the second approach zero, in particular they may become in absolute value less than or equal to the corresponding semi-invariants of a Pearson's Type III distribution which are given by

$$\frac{\lambda_k}{\lambda_2^{\frac{k}{2}}} = \frac{(k-1)!}{a^{k-2}} \qquad k \geq 2$$

in which $a = \dfrac{2\lambda_2^{\frac{3}{2}}}{\lambda_3}$, or

$$\lambda_k = (k-1)!\,\lambda_2 \left(\frac{\sigma}{a}\right)^{k-2}$$

Taking $h = \left|\frac{\sigma}{a}\right|$ it is easy to see that such distributions satisfy our second regularity condition. The smaller the skewness of the Type III distribution, the smaller h may be taken. Thus in such cases we can give a lower limit for $P(|x|\le t\sigma)$, the probability that $|x|\le t\sigma$, which is improved with decreasing skewness of the Type III distribution. By the use of the first regularity condition we could only take $h=\frac{\sigma}{2}$ as the distribution approaches normality.

As a second application, let us suppose that $x=x_1+x_2$ in which x_1 and x_2 are independent, and in which the semi-invariants of the distribution of x_1 are $\ell_2\,(=\sigma_1^2),\ \ell_3,\ \ell_4\cdots$, and the semi-invariants of the distribution of x_2 are $\gamma_2\,(=\sigma_2^2)\,\gamma_3,\ \gamma_4,\cdots$. Then the distribution of x has for semi-invariants

$$\lambda_2=\ell_2+\gamma(=\sigma^2),\quad \lambda_3=\ell_3+\gamma_3,\quad \lambda_4=\ell_4+\gamma_4,\ \cdots\cdots.$$

Further let it be assumed that $\frac{\sigma_2}{\sigma_1}<1$, and that the distribution of x_2 satisfies our second regularity condition.

Then

$$P(|x|\le t\sigma)>P(|x_1|\le t\sigma_1)\,P(|x_2|\le t(\sigma-\sigma_1))$$

But

$$P(|x_2|\le t(\sigma-\sigma_1))=P\left(|x_2|\le \frac{t(\sigma-\sigma_1)}{\sigma_2}\sigma_2\right)$$

$$>1-2e^{-\frac{\frac{t^2(\sigma-\sigma_1)^2}{\sigma_2^2}}{2+2\frac{h}{\sigma_2}\frac{t(\sigma-\sigma_1)}{\sigma_2}}}$$

Now

$$\frac{\sigma-\sigma_1}{\sigma_2}=\frac{(\sigma_1^2+\sigma_2^2)^{\frac{1}{2}}-\sigma_1}{\sigma_2}<1 \qquad \left\{\begin{array}{l}(1+x^2)^{\frac{1}{2}}<1+x\\ \text{if } 0<x<1\end{array}\right.$$

so that we get

$$P(|x_2|\le t(\sigma-\sigma_1))>1-2e^{-\frac{t^2}{2+2h\frac{t}{\sigma_2}}}.$$

This gives finally in such cases

$$P(|x|\le t\sigma)>P(|x_1|\le t\sigma)-2e^{-\frac{t^2}{2+2h\frac{t}{\sigma_2}}}$$

Cecil C. Craig