HOW DEVIANT CAN YOU BE?

PAUL A. SAMUELSON
Massachusetts Institute of Technology

For a finite universe of N items, it is proved no one can lie more than $\sqrt{N-1}$ standard deviations away from the mean. This is an improvement over the result given by Tchebycheff's inequality: and a similar improvement is possible when speaking of how far from the mean any odd-number r out of N observations can lie. However, the relative inefficiency of Tchebycheff's inequality as applied to a finite universe does go to zero as N goes to infinity.

1. To assert that a man is a million standard deviations above the mean is to assert a falsehood (there being less than 4 billion men extant). Where does error leave off and truth begin?

Let the number of men be finite, $N<\infty$. Intuitively one realizes that the greatest relative deviant will occur when all men but one are bunched together. (Proof: putting 2 measurements at their common mean reduces the standard deviation while leaving the mean invariant; so continue doing so.) For $(X_1, X_2, \cdots, X_n)=(a, b, \cdots, b)$, it is a harmless convention to set $(a, b)=(1, 0)$ by proper choice of arbitrary scale and origin constants: hence,

$$m = E[X] = \frac{1}{N}, \qquad \sigma^2 = \overset{n}{\underset{1}{E}}\,[(X-m)^2] = \frac{N-1}{N^2}.$$

The distance of X_1 from the mean in units of standard deviations is

$$\frac{X_1 - m}{\sigma} = \frac{\left(1 - \frac{1}{N}\right)N}{\sqrt{N-1}} = \sqrt{N-1}.$$

Theorem. If there are N items in the universe (say $36\times 10^9+1$ men), no one can be more than $\sqrt{N-1}\,(=60{,}000)$ standard deviations away from the mean in any measured attribute.

Rigorous proof of the theorem requires demonstration of the following inequality:

$$\phi(x_1, \cdots, x_n) = \underset{(X_i)}{\mathrm{Max}}\left\{ \left| X_i - \sum X_j/N \right| - (N-1)^{\frac{1}{2}}\left(\sum [X_k - \sum X_j/N]^2/N\right)^{\frac{1}{2}}\right\} \leq 0 \quad (1)$$

To maximize ϕ is a problem in (non-smooth) concave programming, which by a variety of artifices can be converted into a problem in the standard calculus with maximum solution $\phi(a, b, \cdots, b)=0$.

The above constructive solution shows that the inequality cannot, in general, be improved upon. However, with special restrictions, sharper inequalities can be found. Thus, if the probability distribution is known to be symmetric, the greatest relevant deviant will be found where all but two of the observations are clustered half way between the remaining two. Thus, for $(-1, 0, \cdots, 0, +1)$, $m=0$, $\sigma^2=2/N$, and for symmetric distribution the above theorem can have $\sqrt{N-1}$ replaced by $\sqrt{N/2}$, a definite improvement whenever $N>2$.

2. By use of Tchebycheff's inequality (T.I.), we can derive a result almost as good. That inequality states that, for all $t>0$,

$$P\left\{\left|\frac{X-m}{\sigma}\right| \geq t\right\} \leq \frac{1}{t^2} \tag{2}$$

Applied to a universe of N items, $1/t^2$ is equated to $1/N$ to give

$$P\left\{\left|\frac{X-m}{\sigma}\right| \geq \sqrt{N}\right\} \leq \frac{1}{N} \tag{3}$$

This says literally that at most 1 observation can lie as much as $\sqrt{N}$ standard deviations from the mean. For $\sqrt{N-1}=60{,}000$, $\sqrt{N}=60{,}000.000017$, not much of a difference. However, for $N=2$, T.I. merely says "not more than 1 observation can lie more than 1.4 . . . standard deviations away from the mean," whereas the present inequality yields the sharpest statement possible, namely that no observation can lie more than 1 standard deviation away from the mean.

Knowledge that a probability distribution is symmetric adds nothing to the sharpness of Tchebycheff's inequality. To see this, from any cumulative distribution $P\{X\leq x\}=F[x-m]$ construct a symmetric distribution $\frac{1}{2}\{F[x-m]+1-F[(m-x)-0]\}$, with identical mean m, identical standard deviation σ *and* identical $P\{|X-m|/\sigma\geq t\}=p(t)$ function.

3. The literal interpretation of T. I.'s (3) can be improved on, if we consider a valid variant form for (2)'s Techebycheff's inequality, namely

$$P\left\{\left|\frac{x-m}{\sigma}\right| > t\right\} < \frac{1}{t^2} \tag{4}$$

This differs from the ordinary* T.I. in the omission of *both* inequalities from (2). Since any observation from a finite universe must have probability at least equal to $1/N$, we can now replace (3)'s literal rendering "At most 1 item can be $\sqrt{N}\sigma$'s from the mean" by (4)'s implication "No item can be more than $\sqrt{N}\sigma$'s from the mean". This follows from

$$P\left\{\left|\frac{x-m}{\sigma}\right| > \sqrt{N}\right\} < \frac{1}{N}\cdot \tag{5}$$

4. A generalization of the present problem involves answering the question: What is the minimum function $F_r(N)$ for which it is valid to say, "Of N observations, no r can be more than $F_r(N)\sigma$'s away from the mean?".

* I am grateful to a referee for pointing out that (4) is a well-known result, both for Tchebycheff's inequality and the more general Markoff inequality. A first casual sampling of introductory and intermediate texts failed to turn it up; but an elapse of time improved my luck and now I find it with about the same frequency that I encounter the spelling Chebychev. (Empirical rule: the newer and more advanced the book, the more likely that it is a translation from the Russian, then the more likely that the name is spelled with an initial "c".) Neither form of T.I. is stronger than the other; neither logically implies the other directly; both are true. The whole subject seems worth pursuing elsewhere, particularly to show that the alleged greater generality of the Markoff over the Tchebycheff inequality is not valid: i.e., from T. I. we can deduce M.I., and not just vice versa.

The solution is fairly intuitive. For r even, say $2R$, the Tchebycheff inequality cannot be improved upon since the probability distribution $Nf(1)=Nf(-1)=R$, $Nf(0)=N-2R$ will achieve its equality. Hence $F_{2R}(N)=\sqrt{N/2R}$. For r odd, say $2R+1$, we can approximate the even-r case by concentrating r of our observations as equally as possible far away from the mean. This suggests—what can be proved by quadratic programming or other methods—that the most stringent case to consider is where

$Nf(-1) = R$, $Nf(1) = R + 1$, $Nf(0) = N - r$. This yields

$$m = \frac{1}{N}, \; \sigma^2 = \frac{2R+1}{N} - \left(\frac{1}{N}\right)^2 = \frac{r}{N} - \frac{1}{N^2}$$

$$F_r(N) = \frac{1-m}{\sigma} = \frac{1 - \frac{1}{N}}{\sqrt{\frac{Nr-1}{N^2}}} = \frac{N-1}{\sqrt{Nr-1}}$$

For $r/N=x$ and N large, the T.I. result is asymptotically relatively "efficient." The T.I. result is at its relative worst for N even and $r=N-1$. We have already seen the case $r=1$, $N=2$. The next worst case is $r=3$, $N=4$. Here T.I. gives $\sqrt{4/3}=1.16$, while $F_3(4)=3/\sqrt{11}=.90$.

We may summarize as follows.

Theorem. Of N observations, no r can be more than $F_r(N)$ standard deviations away from the mean where

$$F_r(N) = \sqrt{\frac{N}{r}} \qquad \text{for } r \text{ an even number}$$

$$= \frac{N-1}{\sqrt{Nr-1}} \qquad \text{for } r \text{ an odd number}$$

and

$$\lim_{N\to\infty} F_{xN}(N) = \sqrt{\frac{1}{x}}, \quad 0 < x < 1.$$

6. It is natural to reverse the question and ask: By how many σ's from the mean must we encounter at least 1 observation? At least 2? . . . At least r? Some answers are trivial. By 1 σ you must meet at least one observation (and may meet them all). For $N=2$, you must have met all observations by $\sigma=1$. For any N, if you have met an observation before $\sigma=1$, you cannot have met them all by $\sigma=1$.

7. Instead of using mean and σ, one can use median and mean-absolute-deviation, or still other measures. Replacing our first theorem would then be the following:

Theorem. No one of N observations can be more than N mean-absolute-deviations away from the median.

Since the case $(1, 0, \cdots, 0)$ leads to one deviation exactly N m.a.d.'s from the median, this inequality cannot be improved upon. That it is true follows from the consideration that moving 2 discordant observations to the median lowers the m.a.d. while leaving the median unchanged: hence, $(1, 0, \cdots, 0)$ is indeed optimal.

Further generalization will occur to the reader.

8. *Summary*. Although Tchebycheff's inequality cannot, in general, be improved upon, for universes (or samples) known to consist of a finite number of items N, an improvement on Tchebycheff's inequality is possible when dealing with r of N items, r being odd, but with the relative amount of improvement $\to 0$ as $N \to \infty$.