Gertrude M. Cox

THE "INEFFICIENCY" OF THE SAMPLE MEDIAN FOR FOR MANY FAMILIAR SYMMETRIC DISTRIBUTIONS[1]

by

John T. Chu
Department of Statistics
University of North Carolina, Chapel Hill

Special report to the Office of Naval Research of work at Chapel Hill under Contract NR 042 031, Project N7-onr-284(02), for research in probability and statistics.

Institute of Statistics
Mimeograph Series No. 125
February 23, 1955

THE "INEFFICIENCY" OF THE SAMPLE MEDIAN FOR MANY FAMILIAR SYMMETRIC DISTRIBUTIONS[1]

By J.T.Chu, University of North Carolina

1. A lower bound. If the reciprocal of the (asymptotic) variance of an estimate is taken as a measure of its (asymptotic) efficiency, the sample median $\tilde{x}$ is often (asymptotically) less efficient than the sample mean $\bar{x}$, for many symmetric distributions familiar to statisticians. In fact, for a symmetric distribution having its maximum frequency at the point of symmetry, if $\tilde{x}$ is asymptotically less efficient than $\bar{x}$, then quite often $\tilde{x}$ is not so efficient as $\bar{x}$ at all, with the possible exceptions of very small samples. To show these facts, we derive a very simple, yet sharp, lower bound for the variance of the sample median.

Suppose that $F(x)$ and $f(x)$ are the cdf (cumulative distribution function) and pdf (probability density function) of a certain continuous distribution, and $f(x)$ is symmetric with respect to $x = \xi$, and $f(\xi) \geqq f(x)$ for all x. Let $\tilde{x}$ be the sample median of a sample of size $2n + 1$, then, where $C_n = (2n + 1)! / n!\, n!$,

$$\begin{aligned}(1)\qquad \operatorname{var} \tilde{x} &= \int_{-\infty}^{\infty} (x - \xi)^2 C_n [F(x)]^n [1 - F(x)]^n f(x)\, dx \\ &= \int_0^1 (x - \xi)^2 C_n F^n (1 - F)^n\, dF \\ &\geqq [f(\xi)]^{-2} \int_0^1 (F - 1/2)^2 C_n F^n (1 - F)^n\, dF \\ &= \left\{ 4 [f(\xi)]^2 (2n + 3) \right\}^{-1} .\end{aligned}$$

Equality holds for a rectangular distribution.

1 Sponsored by the Office of Naval Research under Contract NR 042 031

2. Examples. It is well known that for a normal or rectangular distribution $\bar{x}$ is more efficient than $\tilde{x}$. We shall show that this is also true for many familiar symmetric distributions.

(1) Triangular distribution: $f(x) = 1 - |x|$, $|x| \leqq 1$. Following (1), $\tilde{x}$ is less efficient than $\bar{x}$ for samples of sizes $2n + 1$ where $n > 1$. Direct computation shows that this is also true if $n = 1$.

(2) t-distribution. For a t-distribution with k degrees of freedom, $\tilde{x}$ is less efficient than $\bar{x}$ if (not necessarily only if)

$$\pi (k - 2) \Gamma^2(k/2) / 4 \Gamma^2((k + 1)/2) > (2n + 3)/(2n + 1).$$

For both $k= 2m$ and $k = 2m + 1$, the LHS (left hand side) of the above inequality is an increasing function of m. Computation shows, e.g., that the inequality holds if $k \geqq 5$ and $n \geqq 25$.

(3) Symmetric β-distribution. The pdf is given by [2, p. 244]

$$f(x) = \Gamma(2p)\Gamma^{-2}(p)\, x^{p - 1} (1 - x)^{p - 1},$$

$$0 < x < 1;\ p > 0.$$

$\tilde{x}$ is less efficient than $\bar{x}$ if

$$2^{4p - 4} (2p + 1) \Gamma^4(p) / \Gamma^2(2p) > (2n + 3)/(2n + 1) .$$

The LHS becomes smaller if p is replaced by p + 1, and tends to $\pi/2$ as p tends to ∞. So it has a lower bound $\pi/2$. Hence the inequality holds for every $p > 0$ and $n \geqq 2$.

(4) Cauchy type distribution. It is defined to be one with a pdf of the type $f(x) = C_\alpha / (1 + |x|^\alpha)$, $-\infty < x < \infty$; $\alpha > 1$. If $\alpha = 2$, we obtain the well known Cauchy distribution for which $\tilde{x}$ is infinitely more efficient than $\bar{x}$. It would be interesting to examine whether or not $\bar{x}$ becomes more efficient as α increases. Now $\bar{x}$ has finite variance only if $\alpha > 3$. C_α and var $\bar{x}$ can be obtained by using contour integration [3, p. 118]. It follows that $\tilde{x}$ is less efficient than $\bar{x}$ if

$$x^2 \sin 3x / \sin^3 x > (2n + 3)/ (2n + 1),$$

where $x = \pi/\alpha$. The LHS is a decreasing function of x, so an increasing function of α. The least α's for which the LHS is equal to 5/3 and 1, the maximum and minimum of the RHS, are found to be 4.65 and 3.75 approximately.

3. Remarks.

(1) Not for all symmetric distributions is $\bar{x}$ more efficient than $\tilde{x}$. When the parent population has a Laplace distribution, e.g., $\tilde{x}$ is more efficient for all samples of odd sizes [1].

(2) If f(x) satisfies certain continuity conditions, $\tilde{x}$ has an asymptotically normal distribution and the asymptotic variance is $\{ 4 [f(\xi)]^2 (2n + 1) \}^{-1}$. Therefore if the sample size is not too small, the asymptotic variance is for all practical purposes a lower bound for var $\tilde{x}$. And if $\tilde{x}$ is asymptotically less efficient than $\bar{x}$, then $\tilde{x}$ is less efficient than $\bar{x}$ for all samples whose sizes are not too small.

References

[1] Chu, J.T., and Hotelling, H., The moments of the sample median, to be published.

[2] Cramér, H., Mathematical Methods of Statistics, Princeton University Press, 1946.

[3] Whittaker, E.T., and Watson, G.N., Modern Analysis, Cambridge University Press, 1952.