A.R.E. of Sample mean to Sample Median

(Asymptotic distribution of sample quantiles)

Suppose X has cdf $F(x)$ and pdf $f(x)$. Let θ be the γth quantile of X, i.e., $F(\theta) = \gamma$. Let $\widehat{\theta}$ be the γth sample quantile based on iid $X_1, \cdots, X_n$ from $F(x)$. Then

$$\sqrt{n}(\widehat{\theta} - \theta) \xrightarrow{\mathcal{D}} N\left(0, \frac{\gamma(1-\gamma)}{f^2(\theta)}\right)$$

(Asymptotic relative efficiency of sample median to sample mean)

For the median θ, it is $\gamma = \frac{1}{2}$. Let $\widetilde{X}_n$ denote the sample median, i.e.,

$$\widetilde{X}_n = \begin{cases} \frac{1}{2}(n+1)^{th} \text{ order statistic} & \text{if } n \text{ is odd} \\ \text{average of the } \frac{n}{2}\text{th and } (\frac{n}{2}+1)\text{th order statistics} & \text{if } n \text{ is even} \end{cases}$$

Then

$$\sqrt{n}(\widetilde{X}_n - \theta) \xrightarrow{\mathcal{D}} N(0, \frac{1}{4f^2(\theta)}).$$

If $T^{(1)} = \{T_n^{(1)}\}$ and $T^{(2)} = \{T_n^{(2)}\}$ are two sequences of estimators for θ such that

$$\sqrt{n}(T_n^{(i)} - \theta) \xrightarrow{\mathcal{D}} N(0, \sigma_i^2) \quad \text{for } i = 1, 2.$$

Then

$$\text{ARE}(T^{(1)}, T^{(2)}) = \frac{\sigma_2^2}{\sigma_1^2}.$$

Example 1. Let $X_1, \cdots, X_n$ be iid from $N(\theta, 1)$ distribution. Since $N(\theta, 1)$ distribution is symmetric about θ, both $\widetilde{X}_n$ and $\overline{X}_n$ are reasonable estimators of θ.

Both asymtotically have normal distributions. Since

$$f(\theta) = (\sqrt{2\pi})^{-1} \exp[-(1/2)(\theta - \theta)^2] = (\sqrt{2\pi})^{-1}$$

So, the asymptotic variance of $\widetilde{X}_n$ is

$$\sigma_1^2 = \frac{1}{4f^2(\theta)} = \frac{1}{4(\sqrt{2\pi})^{-2}} = \frac{\pi}{2}$$

and the the asymptotic variance of $\overline{X}_n$ is $\sigma_2^2 = 1$. Thus

$$\text{ARE}(\widetilde{X}_n, \overline{X}_n) == \frac{\text{AVar}(\overline{X}_n)}{\text{AVar}(\widetilde{X}_n)} = \frac{1}{(\pi/2)} = \frac{2}{\pi}$$

which is less than 1, implying that for the normal distribution using sample median is asymtotically less efficient than using sample mean for estimating the mean θ.

Example 2. Let $X_1, \cdots, X_n$ be iid from a double exponential distribution with pdf

$$f(x) = \frac{1}{2} e^{-|x-\theta|}, \quad -\infty < x < \infty$$

Since this distribution is also symmetric about θ, both $\widetilde{X}_n$ and $\overline{X}_n$ are competitive estimators for θ.

Again, since both estimators are asymtotically normal and

$$f(\theta) = \frac{1}{2}.$$

The asymptotic variance of $\widetilde{X}_n$ is

$$\sigma_1^2 = \frac{1}{4f^2(\theta)} = \frac{1}{4(1/2)^2} = 1$$

and the the asymptotic variance of $\overline{X}_n$ is

$$\sigma_2^2 = \frac{1}{2}\int_{-\infty}^{\infty} (x-\theta)^2 e^{-|x-\theta|} dx = \frac{1}{2}\int_0^{\infty} 2x^2 e^{-x} dx = \Gamma(3) = 2.$$

Therefore,

$$ARE(\widetilde{X}_n, \overline{X}_n) = \frac{\sigma_2^2}{\sigma_1^2} = \frac{2}{1} = 2.$$

Thus, using the sample median is asymptotically twice more efficent than using the sample mean for estimating the center point θ under the double exponential distribution model.