Asymptotic Relative Efficiency (ARE)

Consider two estimators which are asymptotically unbiased. If one of them has a smaller variance than the other, then the former is more precise, more efficient.

Suppose $T_n^{(1)}$ and $T_n^{(2)}$ are two estimators of $q(\theta)$ such that

$$\sqrt{n}\left(T_n^{(1)} - q(\theta)\right) \xrightarrow[n\to\infty]{d} N(0, \sigma_1^2(\theta)) \text{ and } \sqrt{n}\left(T_n^{(2)} - q(\theta)\right) \xrightarrow[n\to\infty]{d} N(0, \sigma_2^2(\theta)).$$

Then the asymptotic relative efficiency of $T_n^{(1)}$ w.r.t. $T_n^{(2)}$ is defined to be

$$\mathrm{e}\left(\theta, T_n^{(1)}, T_n^{(2)}\right) = \frac{\sigma_2^2(\theta)}{\sigma_1^2(\theta)}.$$

Notice $T_n^{(1)}$ is better if $\sigma_2^2 > \sigma_1^2$.

Example. $X_1, \ldots, X_n$ i.i.d Bernoulli(p). Let $S_n = \sum_{i=1}^n X_i$. Then MLE of p is $T_n^{(1)} = \frac{S_n}{n}$. Consider the Bayes estimate of p under the Beta(a, b) prior for p, $a > 0$, $b > 0$. Since

$$p|X_1, \ldots, X_n \sim \text{Beta}(S_n + a, n - S_n + b),$$

we obtain the Bayes estimate to be

$$T_n^{(2)} = E(p|X_1, \ldots, X_n) = \frac{S_n + a}{S_n + a + n - S_n + b} = \frac{S_n + a}{n + a + b}.$$

We know already that

$$\sqrt{n}(T_n^{(1)} - p) \xrightarrow[n\to\infty]{d} N(0, p(1-p)).$$

Since

$$\begin{aligned}
T_n^{(2)} &= \frac{S_n + a}{n + a + b} = \frac{S_n}{n}\frac{n}{n + a + b} + \frac{a}{n + a + b} \\
&= \frac{S_n}{n}\left(1 - \frac{a + b}{n + a + b}\right) + \frac{a}{n + a + b} \\
&= \frac{S_n}{n} - \frac{(a + b)S_n}{n(n + a + b)} + \frac{a}{n + a + b},
\end{aligned}$$

$$\sqrt{n}(T_n^{(2)} - p) = \sqrt{n}(T_n^{(1)} - p) - \frac{(a + b)S_n}{\sqrt{n}(n + a + b)} + \frac{\sqrt{n}a}{n + a + b}.$$

Now, note that, since a and b are fixed, and $\frac{S_n}{n} \xrightarrow[n\to\infty]{P} p$,

$$\frac{\sqrt{n}a}{n+a+b} \xrightarrow[n\to\infty]{} 0, \text{ and } \frac{(a+b)S_n}{\sqrt{n}(n+a+b)} \xrightarrow[n\to\infty]{P} 0.$$

Therefore $T_n^{(2)}$ has the same asymptotic distribution as $T_n^{(1)}$. i.e.,

$$\sqrt{n}(T_n^{(2)} - p) \xrightarrow[n\to\infty]{d} N(0, p(1-p)).$$

Thus the two estimators have the same asymptotic relative efficiency, or

$$\mathrm{e}(p, T_n^{(1)}, T_n^{(2)}) = \frac{p(1-p)}{p(1-p)} = 1.$$

Example. Let $X_1, X_2, \ldots$ be i.i.d $N(0, \sigma^2)$. Consider the following two estimators for σ:

$$\hat{\sigma}_1 = \sqrt{\frac{1}{n}\sum_{i=1}^n X_i^2}, \qquad \hat{\sigma}_2 = \sqrt{\frac{\pi}{2}}\frac{1}{n}\sum_{i=1}^n |X_i|.$$

Since $X_i^2 \sim \sigma^2\chi_1^2$, $E(X_i^2) = \sigma^2$, $Var(X_i^2) = 2\sigma^4$, by CLT,

$$\sqrt{n}\left(\hat{\sigma}_1^2 - \sigma^2\right) \xrightarrow[n\to\infty]{d} N(0, 2\sigma^4).$$

Taking $h(x) = \sqrt{x}$ for $x > 0$, we have $h'(x) = \frac{1}{2\sqrt{x}}$, and so

$$\sqrt{n}\left(\hat{\sigma}_1 - \sigma\right) \xrightarrow[n\to\infty]{d} N\left(0, \left(\frac{1}{2\sigma}\right)^2 2\sigma^4\right) = N\left(0, \frac{\sigma^2}{2}\right).$$

Since $Z_i = X_i/\sigma \sim N(0,1)$, and

$$\begin{aligned} E(|Z_i|) &= \int_{-\infty}^{\infty} |z| \frac{1}{\sqrt{2\pi}} \exp(-z^2/2)\, dz \\ &= 2\int_0^{\infty} z \frac{1}{\sqrt{2\pi}} \exp(-z^2/2)\, dz \\ &= \sqrt{\frac{2}{\pi}} \int_0^{\infty} \exp(-u)\, du = \sqrt{\frac{2}{\pi}}, \end{aligned}$$

we get $E(|X_i|) = \sqrt{\frac{2}{\pi}}\sigma$ and $Var(|X_i|) = E(X_i^2) - (E(|X_i|))^2 = \sigma^2 - \frac{2}{\pi}\sigma^2 = \left(1 - \frac{2}{\pi}\right)\sigma^2$. Therefore, $E(\sqrt{\frac{\pi}{2}}|X_i|) = \sigma$ and $Var(\sqrt{\frac{\pi}{2}}|X_i|) = \frac{\pi}{2}(1 - \frac{2}{\pi})\sigma^2 = (\frac{\pi}{2} - 1)\sigma^2$. Hence

$$\sqrt{n}\left(\hat{\sigma}_2 - \sigma\right) \xrightarrow[n\to\infty]{d} N\left(0, \left(\frac{\pi}{2} - 1\right)\sigma^2\right).$$

Thus we have the ARE:

$$\mathrm{e}\left(\sigma^2, \hat{\sigma}_1, \hat{\sigma}_2\right) = \frac{\left(\frac{\pi}{2}-1\right)\sigma^2}{\frac{1}{2}\sigma^2} = 2\left(\frac{\pi}{2}-1\right) = \pi - 2 > 1.$$

What can be done when CLT cannot be used to obtain the asymptotic distribution?

Example. Consider i.i.d observations $X_1, X_2, \ldots$ from a location family with $\theta =$ the median. i.e., the density is $f(x|\theta) = f_0(x-\theta)$ and the cdf $F_\theta(x) = F_0(x-\theta)$, and further, $F_\theta(\theta) = F_0(0) = 1/2$ and $f_0(0) > 0$. Then we have the following result.

Result.

$$\sqrt{n}\left(\mathrm{median}(X_1, X_2, \ldots, X_n) - \theta\right) \xrightarrow[n\to\infty]{d} N\left(0, \frac{1}{4f_0^2(0)}\right).$$

Remark. Note that there are no conditions on the moments of X.

Proof. Let $n = 2m-1$, an odd integer, so that the median of $X_1, X_2, \ldots, X_n$ is $X_{(m)}$. Also, let $Y = X - \theta$. Then $X_{(m)} - \theta = Y_{(m)}$. Fix a and consider

$$\begin{aligned} F_n(a) &= P_\theta\left(\sqrt{n}(X_{(m)} - \theta) \le a\right) \\ &= P\left(\sqrt{n}Y_{(m)} \le a\right) = P\left(Y_{(m)} \le \frac{a}{\sqrt{n}}\right). \end{aligned}$$

Let $S_n =$ number of Y's that exceed $\frac{a}{\sqrt{n}}$. Then

$$S_n \sim \mathrm{Binomial}\left(n, p_n = 1 - F_0(\frac{a}{\sqrt{n}})\right).$$

Also $Y_{(m)} \le \frac{a}{\sqrt{n}}$ iff number of Y_i that exceed $\frac{a}{\sqrt{n}}$ is less than or equal to $m-1$ iff $S_n \le m-1 = \frac{n-1}{2}$. Therefore,

$$\begin{aligned} P\left(Y_{(m)} \le \frac{a}{\sqrt{n}}\right) &= P\left(S_n \le \frac{n-1}{2}\right) \\ &= P\left(\frac{S_n - np_n}{\sqrt{np_n(1-p_n)}} \le \frac{\frac{n-1}{2} - np_n}{\sqrt{np_n(1-p_n)}}\right). \end{aligned}$$

From CLT for S_n, we obtain

$$P\left(\frac{S_n - np_n}{\sqrt{np_n(1-p_n)}} \le \frac{\frac{n-1}{2} - np_n}{\sqrt{np_n(1-p_n)}}\right) - \Phi\left(\frac{\frac{n-1}{2} - np_n}{\sqrt{np_n(1-p_n)}}\right) \xrightarrow[n\to\infty]{} 0.$$

Now let

$$x_n = \frac{\frac{n-1}{2} - np_n}{\sqrt{np_n(1-p_n)}} = \frac{\sqrt{n}(\frac{1}{2} - p_n) - \frac{1}{2\sqrt{n}}}{\sqrt{F_0(\frac{a}{\sqrt{n}})(1 - F_0(\frac{a}{\sqrt{n}}))}}.$$

Then

$$\begin{aligned}
\lim_{n\to\infty} x_n &= \frac{\lim_{n\to\infty} \sqrt{n}(\frac{1}{2} - p_n) - 0}{\sqrt{F_0(0)(1 - F_0(0))}} \\
&= 2 \lim_{n\to\infty} \sqrt{n} \left(\frac{1}{2} - (1 - F_0(\frac{a}{\sqrt{n}}) \right) \\
&= 2a \lim_{n\to\infty} \frac{F_0(a/\sqrt{n}) - F_0(0)}{a/\sqrt{n}} \\
&= 2aF_0'(0) = 2af_0(0).
\end{aligned}$$

Therefore,

$$F_n(a) = P_\theta \left(\sqrt{n}(X_{(m)} - \theta) \le a \right) \underset{n\to\infty}{\longrightarrow} \Phi\left(2f_0(0)a\right).$$

Hence, the density of the asymptotic distribution is

$$\begin{aligned}
\frac{d}{da}\Phi\left(2f_0(0)a\right) &= 2f_0(0)\phi\left(2f_0(0)a\right) \\
&= 2f_0(0)\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(2f_0(0))^2 a^2 \right) \\
&= \frac{1}{\sqrt{2\pi}} \frac{1}{(2f_0(0))^{-1}} \exp\left(-\frac{1}{2} \frac{a^2}{[(2f_0(0))^{-1}]^2} \right),
\end{aligned}$$

which is that of $N\left(0, (2f_0(0))^{-2}\right)$.