Stat210A: Theoretical Statistics	**Lecture Date: October 14, 2014**

Asymptotic Relative Efficiency

Lecturer: Michael I. Jordan *Scribe: Horia Mania*

1 Asymptotic Distribution of the Order Statistics

Last time we saw that

$$\sqrt{n}(\tilde{X}_n - \theta) \Longrightarrow N\left(0, \frac{1}{4[F'(\theta)]^2}\right),$$

where $\tilde{X}_n$ is the median of the random variables $X_1, X_2, ..., X_n$ i.i.d. with cumulative distribution function F (we assumed that there exists a unique θ such that $F(\theta) = 1/2$, otherwise $F'(\theta) = 0$). This result can be extended to any order statistic.

Theorem 1 (keener's, thm 8.18)**.** *Let $(X_i)_{i\geqslant 1}$ be i.i.d. random variables with common cumulative distribution function F, let $\gamma \in (0,1)$, and let $\tilde{\theta}_n$ be the $\lfloor \gamma n \rfloor$th order statistic for $X_1, X_2, ..., X_n$ (or a weighted average of the $\lfloor \gamma n \rfloor$th and $\lceil \gamma n \rceil$th order statistics). If $F(\theta) = \gamma$, and if $F'(\theta)$ exists and is finite and positive, then*

$$\sqrt{n}(\tilde{\theta}_n - \theta) \Longrightarrow N\left(0, \frac{\gamma(1-\gamma)}{[F'(\theta)]^2}\right),$$

as $n \to \infty$.

Proof. The proof is similar to the proof given for the case $\gamma = 1/2$. The idea is to express the distribution of $\sqrt{n}(\tilde{\theta}_n - \theta)$ in terms of a binomial distribution and then apply the CLT. □

Remark 2. The n-th order statistic $X_{(n)} = \max_{1\leqslant i\leqslant n} X_i$ does not satisfy this property. A different scaling is needed to find a limiting distribution for $X_{(n)}$.

2 Asymptotic Relative Efficiency

In this section we compare the asymptotic behavior of $\tilde{X}_n$ and $\overline{X}_n$, the median and the mean of $X_1, X_2, ..., X_n$ i.i.d. with distribution F, for different choices of the cumulative distribution F. Such a comparison makes sense only if both the median and the mean estimate the same parameter. This is the case when F has density $f(x-\theta)$, with f symmetric about zero.

By the CLT we know that

$$\sqrt{n}(\overline{X}_n - \theta) \Longrightarrow N(0, \overline{\sigma}^2),$$

where

$$\overline{\sigma}^2 = \int (x-\theta)^2 f(x-\theta)dx = \int x^2 f(x)dx.$$

Also, by what we discussed in the previous section, we know

$$\sqrt{n}(\tilde{X}_n - \theta) \Longrightarrow N\left(0, \tilde{\sigma}^2\right),$$

where $\tilde{\sigma}^2 = \frac{1}{4[F'(\theta)]^2} = \frac{1}{4f(0)^2}$.

Example 3. Take $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$, the density of the standard normal. Then $\overline{\sigma}^2 = 1$ and $\tilde{\sigma}^2 = \frac{\pi}{2}$. Since the limiting variance of the median is larger than that of the mean, we expect that the mean if more "efficient" than the median. This is clearer if we look at the number of data points needed by the median to yield an estimate with the same asymptotic variance as the mean. We make this explicit. Let $m = \lfloor \pi n/2 \rfloor$, and observe that $\sqrt{n/m} \to \sqrt{2/\pi}$ as $n \to \infty$. By Slutzki's theorem,

$$\sqrt{n}(\tilde{X}_m - \theta) = \sqrt{\frac{n}{m}} \sqrt{m}(\tilde{X}_m - \theta) \Longrightarrow N(0,1).$$

This suggests that the estimate produced by the median using $\pi/2$ times the data used by the mean behaves as good as the estimate produced by the mean. Although $\overline{X}_n$ seems to behave better than $\tilde{X}_n$, sometimes we might choose the latter estimator because it is more robust to outliers.

The previous example motivates us to define the *asymptotic relative efficiency* (ARE) of two estimators $\overline{\theta}_n$ and $\tilde{\theta}_n$ as follows. Suppose

$$\sqrt{n}(\overline{\theta}_n - \theta) \Longrightarrow N\left(0, \sigma^2_{\overline{\theta}}\right)$$
$$\sqrt{n}(\tilde{\theta}_n - \theta) \Longrightarrow N\left(0, \sigma^2_{\tilde{\theta}}\right),$$

then the ARE of $\overline{\theta}_n$ with respect to $\tilde{\theta}_n$ is $\frac{\sigma^2_{\tilde{\theta}}}{\sigma^2_{\overline{\theta}}}$.

Example 4. Let $f(x) = \frac{1}{2} e^{-|x|}$. Then

$$\overline{\sigma}^2 = \int_{-\infty}^{\infty} x^2 \frac{1}{2} e^{-|x|} dx = \int_0^{\infty} x^2 e^{-x} dx = \Gamma(3) = 2, \text{ and}$$
$$\tilde{\sigma}^2 = \frac{1}{4f(0)^2} = 1.$$

Hence the ARE of $\tilde{X}_n$ with respect to $\overline{X}_n$ is 2.

Example 5. Let $X_1, X_2, ..., X_n$ be i.i.d. samples from $N(\theta, 1)$. We are interested in estimating

$$p = P_\theta(X_i \leqslant a) = \Phi(a - \theta).$$

Two natural choices of estimators are:

$$\hat{p} = \Phi(a - \overline{X}), \text{ where } \overline{X} = \frac{1}{n}\sum_{i=1}^n X_i \text{ (this is the MLE estimator)}$$
$$\tilde{p} = \frac{1}{n}\sum_{i=1}^n 1_{X_i \leqslant a}.$$

The CLT immediately implies that $\sqrt{n}(\tilde{p} - p) \Longrightarrow N(0, \tilde{\sigma}^2)$, where

$$\tilde{\sigma}^2 = \mathrm{Var}(1_{X_i \leqslant a}) = \Phi(a-\theta)(1 - \Phi(a-\theta)).$$

On the other hand, $\hat{p}$ is a function of $\overline{X}$, so the delta method implies that

$$\sqrt{n}(\hat{p} - p) \Longrightarrow N(0, \hat{\sigma}^2),$$

where

$$\hat{\sigma}^2 = \left[\frac{\partial}{\partial x} \Phi(a-x) \bigg|_{\theta} \right]^2 = \phi^2(a-\theta).$$

Therefore, the ARE of $\hat{p}$ with respect to $\tilde{p}$ is

$$ARE = \frac{\Phi(a-\theta)(1-\Phi(a-\theta))}{\phi^2(a-\theta)}.$$

We note that the ARE depends on θ. When $\theta = a$ the ARE is $\pi/2$, and as $|\theta - a|$ increases the ARE increases without bound. Nonetheless, the estimator $\tilde{p}$ is sensible when the model is wrong, while $\hat{p}$ is a good estimator only when the model is correct.

3 Limiting distribution of the MLE

Recall that we derived the limiting distribution of the MLE of an exponential family.

$$\sqrt{n}(\hat{\eta} - \eta) \Longrightarrow N\left(0, \frac{1}{A''(\eta)}\right),$$

where $\hat{\eta} = \psi(\overline{T})$. This result can be extended to other families of distributions. The following theorem makes this explicit.

Theorem 6 (keener's, thm 9.14)**.** *Assume:*

1. *Variables $X, X_1, X_2, ...$ are i.i.d. with common density f_θ, $\theta \in \Omega \subset \mathbf{R}$.*
2. *The set $A = \{x \colon f_\theta(x) > 0\}$ is independent of θ.*
3. *For every $x \in A$, $\partial^2 f_\theta(x)/\partial\theta^2$ exists and is continuous in θ.*
4. *Let $W(\theta) = \log f_\theta(X)$. The Fisher information $I(\theta)$ from a single observation exists, is finite and can be found using either*

$$I(\theta) = E_\theta W'(\theta)^2 \text{ or } I(\theta) = -E_\theta W''(\theta).$$

Also, $E_\theta W'(\theta) = 0$.

5. *For every θ in the interior of Ω there exists $\epsilon > 0$ such that*

$$E_\theta \| 1_{[\theta-\epsilon, \theta+\epsilon]} W'' \|_\infty < \infty.$$

6. *The maximum likelihood estimator $\hat{\theta}_n$ is consistent.*

Then for any θ in the interior of Ω,

$$\sqrt{n}(\hat{\theta}_n - \theta) \Longrightarrow N\left(0, \frac{1}{I(\theta)}\right),$$

as $n \to \infty$.

Proof. We just sketch the main idea. Let $\overline{W_n}(\omega) = \frac{1}{n}\sum_{i=1}^n \log f_\omega(X_i)$. Then by the mean value theorem we know that

$$\overline{W}_n'(\hat{\theta}_n) = \overline{W}_n'(\theta) + \overline{W}_n''(\tilde{\theta}_n)(\hat{\theta}_n - \theta),$$

for some $\tilde{\theta}_n \in [\theta, \hat{\theta}_n]$. Since $\hat{\theta}_n$ maximizes $\overline{W}_n$, $\overline{W}'_n(\hat{\theta}_n) = 0$. Hence

$$\sqrt{n}(\hat{\theta}_n - \theta) = \frac{\sqrt{n}\overline{W}'_n(\theta)}{-\overline{W}''_n(\tilde{\theta}_n)},$$

and we will analyze the limiting distribution of this ratio in the following lectures using empirical process theory. $\square$