STA 522, Spring 2021
Introduction to Theoretical Statistics II

Lecture 13

Department of Biostatistics
University at Buffalo

26 April, 2021

AGENDA

- Asymptotic Evaluations

Consistent Estimators

Definition: A sequence of estimators $W_n = W_n(\underline{X})$ is a (weakly) **consistent sequence of estimators** of the parameter θ if and only if for every $\theta \in \Theta$, $W_n \xrightarrow{P} \theta$ i.e., for every $\varepsilon > 0$,

$$\lim_{n \to \infty} P_\theta \left(|W_n - \theta| \geq \varepsilon \right) = 0,$$

or, equivalently,

$$\lim_{n \to \infty} P_\theta \left(|W_n - \theta| < \varepsilon \right) = 1.$$

Definition: A sequence of estimators $W_n = W_n(\underline{X})$ is a strongly **consistent sequence of estimators** of the parameter θ if and only if $W_n \xrightarrow{a.s.} \theta$, for every $\theta \in \Theta$, i.e.,

$$P\left(\lim_{n \to \infty} W_n = \theta \right) = 1.$$

Example: Let $X_1, X_2, \ldots, X_n \sim$ iid $\mathrm{N}(\theta, 1)$. Then $\overline{X}_n$ is a consistent sequence of estimators of θ.

Recall that $\overline{X}_n \sim \mathrm{N}(\theta, 1/n)$. So,

$$
\begin{aligned}
P_\theta(|\overline{X}_n - \theta| < \varepsilon) &= \int_{\theta-\varepsilon}^{\theta+\varepsilon} \left(\frac{n}{2\pi}\right)^{1/2} e^{-(n/2)(\overline{x}_n-\theta)^2}\ d\overline{x}_n \\
&= \int_{-\varepsilon}^{\varepsilon} \left(\frac{n}{2\pi}\right)^{1/2} e^{-(n/2)y^2}\ dy && (y = \overline{x}_n - \theta) \\
&= \int_{-\varepsilon\sqrt{n}}^{\varepsilon\sqrt{n}} \left(\frac{1}{2\pi}\right)^{1/2} e^{-(1/2)t^2}\ dt && (t = y\sqrt{n}) \\
&= P(-\varepsilon\sqrt{n} < Z < \varepsilon\sqrt{n}) && (Z \sim \mathrm{N}(0,1)) \\
&\to 1 \quad \text{as } n \to \infty
\end{aligned}
$$

This shows $\overline{X}_n \xrightarrow{P} \theta$.

How to Verify Consistency for a Sequence of Estimators

Theorem 10.1.3

Let W_n be a sequence of estimators of a parameter θ satisfying

(a) $\lim_{n\to\infty} \text{Var}_\theta(W_n) = 0$ and

(b) $\lim_{n\to\infty} \text{Bias}_\theta(W_n) = 0$

for every $\theta \in \Theta$.

Then W_n is a consistent sequence of estimators of θ.

Example (contd.): $X_1, X_2, \dots, X_n \sim$ iid $\text{N}(\theta, 1)$, consider the estimator $\overline{X}_n$ of θ.

We have $\text{E}_\theta(\overline{X}_n) = \theta$ for all θ, i.e., $\text{Bias}_\theta(\overline{X}_n) = 0$, and $\text{Var}_\theta(\overline{X}_n) = \frac{1}{n} \to 0$. Hence, from the above theorem, it follows that $\overline{X}_n$ is consistent for θ.

Theorem 10.1.6 (Consistency of MLEs)

Let $X_1, X_2, \ldots, X_n \sim \text{iid } f(x \mid \theta)$.

Let $L(\theta \mid \underline{x}) = \prod_{i=1}^n f(x_i \mid \theta)$ be the likelihood function.

Let $\hat{\theta}$ denote the MLE of θ.

Let $\tau(\theta)$ be a continuous function of θ.

Under certain regularity conditions on $f(x \mid \theta)$ (see Miscellanea 10.6.2; these hold, e.g., for the regular exponential family) and, hence, $L(\theta \mid \underline{x})$, for every $\varepsilon > 0$ and every $\theta \in \Theta$,

$$\lim_{n \to \infty} P_\theta \left(|\tau(\hat{\theta}) - \tau(\theta)| \geq \varepsilon \right) = 0.$$

In other words, $\tau(\hat{\theta})$ is a consistent estimator of $\tau(\theta)$, i.e., MLEs are weakly consistent (converge in probability).

Asymptotic Variance & Efficiency

Definition: For an estimator T_n, suppose that $k_n(T_n - \tau(\theta)) \xrightarrow{d} \mathrm{N}(0, \sigma^2)$, where $\{k_n\}$ is a sequence of constants. The parameter σ^2 is called the **asymptotic variance** or **variance of the limit distribution** of T_n.

Definition: A sequence of estimators W_n is **asymptotically efficient** for a parameter $\tau(\theta)$ if and only if

$$\sqrt{n}(W_n - \tau(\theta)) \xrightarrow{d} W \sim \mathrm{N}(0, v(\theta))$$

and

$$v(\theta) = \frac{[\tau'(\theta)]^2}{I_1(\theta)} = \frac{[\tau'(\theta)]^2}{\mathrm{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\log f(X \mid \theta)\right)^2\right]};$$

that is, the asymptotic variance of W_n (almost) achieves the Cramér–Rao Lower Bound. $I_1(\theta)$ is the *unit* Fisher information (i.e., Fisher information computed for *one* random observation).

Notes on Asymptotic Efficiency

(a) There is no n in the denominator of $v(\theta)$, as it got moved to the left-hand side.

(b) This definition does not mean

$$W_n \xrightarrow{d} \mathsf{N}(\tau(\theta), \mathrm{CRLB}),$$

since

$$\mathrm{CRLB} = \frac{[\tau'(\theta)]^2}{n\, \mathsf{E}_\theta \left[\left(\frac{\partial}{\partial\theta} \log f(X \mid \theta) \right)^2 \right]}.$$

$v(\theta)$ is written without the n because we are letting $n \to \infty$

(c) In practical terms, what this does mean is that W_n is approximately normally distributed with mean $\tau(\theta)$ and variance equal to the CRLB, evaluated at $\hat{\theta}$. This is often denoted as $W_n \overset{a}{\sim} \mathsf{N}\left(\tau(\theta), CRLB(\hat{\theta})\right)$.

Asymptotic distribution of a function of a sequence of random variables

Theorem 5.5.24 (Delta Method)

Suppose Y_n is a sequence of random variables that satisfies $\sqrt{n}(Y_n - \theta) \xrightarrow{d} \mathrm{N}(0, \sigma^2)$. For a given function g and a specific value of θ, suppose that $g'(\theta)$ exists and is not 0. Then

$$\sqrt{n}[g(Y_n) - g(\theta)] \xrightarrow{d} \mathrm{N}(0, \sigma^2[g'(\theta)]^2)$$

Heuristic Proof: The Taylor expansion of $g(Yn)$ around $Y_n = \theta$ is

$$g(Y_n) = g(\theta) + g'(\theta)(Y_n - \theta) + R_n$$

where R_n is such that $\sqrt{n}R_n \xrightarrow{P} 0$ as $n \to \infty$. Therefore,

$$\sqrt{n}\ [g(Y_n) - g(\theta)] = g'(\theta)\sqrt{n}\ (Y_n - \theta) + \sqrt{n}\ R_n \xrightarrow{d} \mathrm{N}(0, \sigma^2[g'(\theta)]^2)$$

by Slutsky's theorem.

Theorem 10.1.12 (Asymptotic Efficiency of MLEs)

Let $X_1, X_2, \ldots, X_n \sim \text{iid } f(x \mid \theta)$. Let $\hat{\theta}$ denote the MLE of θ. Under certain regularity conditions (see Miscellanea 10.6.2) on $f(x \mid \theta)$ and, hence, on $L(\theta \mid \underline{x})$,

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} \mathrm{N}\left(0, \frac{1}{I_1(\theta)}\right),$$

where $I_1(\theta)$ is the unit Fisher information.

Corollary

Suppose $\tau(\theta)$ is a differentiable function of theta. If $\hat{\theta}$ denotes the MLE of θ, then $\tau(\hat{\theta})$ is the MLE of $\tau(\theta)$. The asymptotic distribution of $\tau(\hat{\theta})$ is obtained using the delta method:

$$\sqrt{n}(\tau(\hat{\theta}) - \tau(\theta)) \xrightarrow{d} \mathrm{N}(0, v(\theta))$$

In other words, $\tau(\hat{\theta})$ is a consistent and asymptotically efficient estimator of $\tau(\theta)$.

Example: Suppose $X_1, X_2, \ldots, X_n \sim$ iid *Bernoulli*(p). The MLE of p is $\hat{p} = \frac{1}{n}\sum_{i=1}^{n} X_i$. We have seen an asymptotic distribution of $\overline{X}_n$ via de Moivre-Laplace CLT.

To apply the asymptotic normality of the MLE, note that here $I_1(\theta) = \frac{1}{p(1-p)}$.

Hence

$$\sqrt{n}\,(\hat{p} - p) \xrightarrow{d} \mathrm{N}(0, p(1-p)) \implies \frac{\sqrt{n}\,(\hat{p} - p)}{\sqrt{p(1-p)}} \xrightarrow{d} \mathrm{N}(0,1)$$

The standard deviation $\sqrt{p(1-p)}$ has MLE $\sqrt{\hat{p}(1-\hat{p})}$, and due to consistency of $\hat{p}$

$$\sqrt{\hat{p}(1-\hat{p})} \xrightarrow{P} \sqrt{p(1-p)}$$

Therefore using Slutsky's theorem

$$\frac{\sqrt{n}\,(\hat{p} - p)}{\sqrt{\hat{p}(1-\hat{p})}} \xrightarrow{d} \mathrm{N}(0,1)$$

In practical terms this means $\hat{p} \overset{a}{\sim} \mathrm{N}\left(p, \frac{p(1-p)}{n}\right)$.

Asymptotic Relative Efficiency (ARE)

If two estimators W_n and V_n satisfy

$$\sqrt{n}\,[W_n - \tau(\theta)] \xrightarrow{d} \mathsf{N}(0, \sigma_W^2)$$
$$\sqrt{n}\,[V_n - \tau(\theta)] \xrightarrow{d} \mathsf{N}(0, \sigma_V^2)$$

then the asymptotic relative efficiency (ARE) of V_n with respect to W_n is

$$\mathsf{ARE}(V_n, W_n) = \frac{\sigma_W^2}{\sigma_V^2}$$

Example (AREs of Poisson Estimators): Suppose that $X_1, X_2, \ldots$ are iid Poisson(λ), and we are interested in estimating $\tau(\lambda) = P_\lambda(X = 0) = e^{-\lambda}$. Consider two estimators of λ, viz.,

(a) $V_n = \frac{1}{n}\sum_{i=1}^{n} I(X_i = 0) = \frac{1}{n} Y_i =: \overline{Y}_n$ and

(b) $W_n = \tau(\overline{X}) = e^{-\overline{X}_n}$. Note that $\overline{X}_n$ is the MLE of λ.

and find the ARE of V_n relative to W_n.

For (a), note that $Y_i = I(X_i = 0) \sim$ iid Binomial$(1, \tau(\lambda))$. Therefore,

$$\mathrm{E}_\lambda(V_n) = e^{-\lambda} \quad \text{and} \quad \mathrm{Var}_\lambda(V_n) = \frac{e^{-\lambda}(1 - e^{-\lambda})}{n}$$

For (b), using delta method approximation we have (HW)

$$\mathrm{E}_\lambda(W_n) \approx e^{-\lambda} \quad \text{and} \quad \mathrm{Var}_\lambda(W_n) = \frac{\lambda e^{-2\lambda}}{n}$$

Since

$$\sqrt{n}\ \left[V_n - e^{-\lambda}\right] \xrightarrow{d} \mathrm{N}\left(0, e^{-\lambda}(1 - e^{-\lambda})\right)$$
$$\sqrt{n}\ \left[W_n - e^{-\lambda}\right] \xrightarrow{d} \mathrm{N}\left(0, \lambda e^{-2\lambda}\right)$$

Therefore

$$\mathrm{ARE}(V_n, W_n) = \frac{\lambda e^{-2\lambda}}{e^{-\lambda}(1 - e^{-\lambda})} = \frac{\lambda}{1 - e^{-\lambda})}$$

The ARE is strictly decreasing with a maximum of 1 attained at $\lambda = 0$.

Note that the UMVUE for $e^{-\lambda}$ is
$U_n = \left(\frac{n-1}{n}\right)^{\sum_{i=1}^{n} X_i} = \left[\left(1 - \frac{1}{n}\right)^n\right]^{\overline{X}_n} \approx e^{-\overline{X}_n} = V_n$ for large n.

Asymptotic Distribution of LRT

Suppose $X_1, X_2, \ldots, X_n \sim$ iid poisson(λ), and we want to construct a level α test of

$$H_0 : \lambda = \lambda_0$$

$$H_1 : \lambda \neq \lambda_0.$$

A level α test is obtained using rejection region

$$R = \{\underline{x} \,:\, -2 \log \lambda(\underline{x}) > \chi^2_{1,\alpha}\},$$

where $\chi^2_{1,\alpha}$ is the χ^2_1 value with area α to its right.

Example (Poisson Testing): Suppose that $X_1, X_2, \ldots$ are iid Poisson(λ), and we are interested in testing $H_0 : \lambda = \lambda_0$ vs. $H_1 : \lambda \neq \lambda_0$.

We have

$$-2 \log \lambda(\underline{x}) = -2 \log \left(\frac{e^{-n\lambda_0} \;\; \lambda_0^{\sum_{i=1}^n x_i}}{e^{-n\hat{\lambda}} \;\; \hat{\lambda}^{\sum_{i=1}^n x_i}} \right) = 2n \left[(\lambda_0 - \hat{\lambda}) - \hat{\lambda} \log(\lambda_0/\hat{\lambda}) \right]$$

where $\hat{\lambda} = \overline{x}$ is the MLE of λ.

The asymptotic theory based test would be to reject H_0 at level α if $-2 \log \lambda(\underline{x}) > \lambda_{1,\alpha}$

Asymptotic normality based tests

Suppose $X_1, X_2, \ldots, X_n$ is a random sample from some population $f_\theta(x)$.

Wald test

Let $\hat{\theta}_n = \hat{\theta}_n(X_1, X_2, \ldots, X_n)$ be the MLE of θ. Then using the asymptotic normality of $\hat{\theta}_n$ (holds under certain regularity conditions):

$$\sqrt{n}\left(\hat{\theta}_n - \theta\right) \xrightarrow{d} \mathrm{N}\left(0, \frac{1}{I_1(\theta)}\right)$$

one can perform tests of hypotheses about the real valued parameter θ.

Score test

The *score statistic* is defined as

$$S(\theta) = \frac{\partial}{\partial\theta} \log f(\underline{X} \mid \theta) = \frac{\partial}{\partial\theta} \log L(\theta \mid \underline{X})$$

We know that $\mathrm{E}_\theta(S(\theta)) = 0$ for all θ. Furthermore

$$\mathrm{Var}_\theta(S(\theta)) = \mathrm{E}_\theta\left[\left(\frac{\partial}{\partial\theta} \log L(\theta \mid \underline{X})\right)^2\right] = -E_\theta\left(\frac{\partial^2}{\partial\theta^2} \log L(\theta \mid \underline{X})\right) = I_n(\theta)$$

where $I_1(\theta)$ is the Fisher information obtained from one random observations. Tests of hypothesis such as $H_0 : \theta = \theta_0$ vs. $H_1 : \theta \neq \theta_0$ can be performed using the asymptotic normality of the score statistic:

$$\frac{S(\theta)}{\sqrt{I_n(\theta)}} \xrightarrow{d} \mathrm{N}(0, 1)$$

Homework

- Read p. 467 – 481, 488 – 495.
- Exercises: TBA.