STAT 346/446 Lecture 8

Asymptotic evaluations of point estimators

Section 10.1

1. Consistency
   - Consistency of MLEs

10.1.1

2. Asmyptotic efficiency
3. Asymptotic relative efficiency

Note: We skip Section 10.1.4 (Bootstrap)

Asymptotics

- $n =$ sample size ← CLT kicks in!
- What happens when $n \to \infty$?
- Why do we care?
  - Often calculations simplify so we can find approximate inference procedures for large sample sizes
  - Useful evaluation/comparison tools

Consistency

Definition: Consistency

A sequence of estimators $W_n = W_n(X_1, X_2, \ldots, X_n)$ is a **consistent sequence of estimators** of the parameter θ if for every $\epsilon > 0$ and every $\theta \in \Theta$

or: $\lim_{n\to\infty} P(|W_n - \theta| \geq \epsilon) = 0$

$$\lim_{n\to\infty} P\left(|W_n - \theta| < \epsilon\right) = 1$$

We also say that W_n is a **consistent estimator**

- That is, $W_n \xrightarrow{p} \theta$ for all θ converges in probability
- A consistent estimator will be arbitrarily close to the parameter with high probability as sample size increases.

Example

$\to \bar{X}_n \sim N(\theta, 1/n)$

- Let $X_1, X_2, \ldots, X_n$ be iid. $N(\theta, 1)$. Show that $\overline{X}_n$ is a consistent sequence of estimators for θ

show $P(|\bar{X}_n - \theta| < \varepsilon) \to 1$ as $n \to \infty$

Actually, by WLLN $\bar{X}_n \xrightarrow{P} \theta$

$$P(|\bar{X}_n - \theta| < \varepsilon) = P(-\varepsilon < \bar{X}_n - \theta < \varepsilon)$$

$$= P\left(-\frac{\varepsilon}{\sqrt{1/n}} < \frac{\bar{X}_n - \theta}{\sqrt{1/n}} < \frac{\varepsilon}{\sqrt{1/n}}\right)$$

$$= \Phi(\sqrt{n}\,\varepsilon) - \Phi(-\sqrt{n}\,\varepsilon)$$

where Φ is the standard normal cdf.

$$\to 1 - 0 = 1 \text{ as } n \to \infty$$

by properties of a cdf.

$\lim_{x \to \infty} F(x) = 1$

$\lim_{x \to -\infty} F(x) = 0$

Note: The sample mean is always a consistent estimator of the population mean (by WLLN)

Consistent estimators

Theorem

If W_n is a sequence of estimators of a parameter θ satisfying

(i) $\lim_{n\to\infty} Var(W_n) = 0$

(ii) $\lim_{n\to\infty} \mathrm{bias}(W_n) = 0$

then W_n is a consistent sequence of estimators of θ

Proof : Chebychev

Recall: Chebychev: $P(g(X) \geq r) \leq \frac{E(g(X))}{r}$

$g(\cdot)$ is a non-negative function

Setting $g(x) = (x-\theta)^2$ gives

$$P(|X-\theta| \geq \varepsilon) \leq \frac{E((X-\theta)^2)}{\varepsilon^2}$$

proof: show: $\lim_{n\to\infty} P(|W_n - \theta| \geq \varepsilon) = 0$ *

By Chebyshev:

$$P(|W_n - \theta| \geq \varepsilon) \leq \frac{E((W_n-\theta)^2)}{\varepsilon^2}$$

$\Rightarrow$ If $\lim_{n\to\infty} E((W_n-\theta)^2) = 0$ then * is true

$E((W_n-\theta)^2) = MSE(W_n) = Var(W_n) + bias(W_n)^2$

$\to 0$ as $n \to \infty$

if $Var(W_n) \to 0$ as $n \to \infty$

and $bias(W_n) \to 0$ as $n \to \infty$

Consistency of sample mean

- When $X_1, X_2, \ldots, X_n$ are iid. $N(\mu, \sigma^2)$ then

$$Var(\overline{X_n}) = \frac{\sigma^2}{n} \to 0 \qquad \text{and} \qquad \mathrm{bias}(\overline{X}_n) = 0$$

as $n \to \infty$

so $\overline{X}_n$ is a consistent estimator of μ

- In fact, by WLLN for any random sample with a finite variance we have

$$\overline{X}_n \xrightarrow{p} E(X)$$

So $\overline{X}_n$ is a consistent estimator of $E(X)$

Example

- Let $X_1, X_2, \ldots, X_n$ be a random sample from $\mathrm{Uniform}(0, \theta)$, $\theta > 0$.
- We found before (Lecture 4 ~~5~~) that the MLE for θ is $X_{(n)}$. Is $X_{(n)}$ a consistent estimator of θ?

MLEs are (generally) consistent

Theorem

Let $X_1, X_2, \dots, X_n$ be a random sample from $f(x \mid \theta)$ and let $\hat{\theta}$ be the MLE of θ. Under some regularity assumptions on $f(x \mid \theta)$

- $\hat{\theta}$ is a consistent estimator of θ
- $\tau(\hat{\theta})$ is a consistent estimator of $\tau(\theta)$, if τ is a continuous function

- One of the conditions: The support of the distribution cannot depend on θ

If $\hat{\theta}$ is the MLE of θ then $\tau(\hat{\theta})$ is the MLE of $\tau(\theta)$

→ First general guarantee that MLE's are good estimators.

end of material for mid term.

Asymptotic variance

Skip the rest of this Lecture

- For a good estimator we usually have $\lim_{n \to \infty} V(W_n) = 0$
- Need a different set-up to compare asymptotic behavior of estimators

Definition

Suppose that for an estimator T_n we have

$$k_n \left(T - \tau(\theta)\right) \xrightarrow{d} N(0, \sigma^2)$$

Then σ^2 is called the **asymptotic variance** of the limit distribution of T_n

- Usually $k_n = \sqrt{n}$

Asymptotic variance - Example

- Let $X_1, \ldots, X_n$ be a random sample from a pdf with mean μ and variance $\sigma^2 < \infty$
- We know that $\mathrm{Var}(\overline{X}_n) = \sigma^2/n$
- We also know that

$$\sqrt{n}(\overline{X}_n - \mu) \xrightarrow{d} N(0, \sigma^2)$$

So: σ^2 is the asymptotic variance of $\overline{X}_n$

Asymptotically efficient

- Recall: Estimator is *efficient* if it reaches its CRLB

Definition

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

$$\sqrt{n}\left(W_n - \tau(\theta)\right) \xrightarrow{d} N(0, v(\theta))$$

and

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

- That is, the asymptotic variance of W_n achieves the Cramér-Rao lower bound (for $n = 1$)

MLEs are (generally) asymptotically efficient

Theorem

Let $X_1, X_2, \ldots, X_n$ be a random sample from $f(x \mid \theta)$ and let $\hat{\theta}$ be the MLE of θ. Under some regularity conditions on $f(x \mid \theta)$

- $\hat{\theta}$ is an asymptotically efficient estimator of θ
- $\tau(\hat{\theta})$ is an asymptotically efficient estimator of $\tau(\theta)$, if τ is a continuous function

- So the MLE method of finding estimators has some proven optimality characteristics.
- Asymptotically efficient = asymptotically best unbiased estimator of its expected value. That is, when n is large

$$E(W_n) \approx \tau(\theta) \qquad \text{and} \qquad Var(W_n) \approx \frac{[\tau'(\theta)]^2}{nI_1(\theta)}$$

More on MLEs

- Let $\hat{\theta}$ be the MLE of θ and let $\tau(\theta)$ be a continuous function. Theorem says: Under some regularity conditions

$$\sqrt{n}\,(\tau(\hat{\theta}) - \tau(\theta)) \xrightarrow{d} N\left(0,\ \frac{[\tau'(\theta)]^2}{I_1(\theta)}\right)$$

- This can be used to find the approximate distribution for an MLE (for fixed n):

$$\tau(\hat{\theta}) \overset{\text{approx}}{\sim} N\left(\tau(\theta),\ \frac{[\tau'(\theta)]^2}{nI_1(\theta)}\right)$$

  where

$$I_1(\theta) = E\left(\left(\frac{\partial}{\partial\theta}\log(f(X \mid \theta))\right)^2\right) = -E\left(\frac{\partial^2}{\partial\theta^2}\log(f(X \mid \theta))\right)$$

Example

- Let $X_1, X_2, \ldots, X_n$ be a random sample from $\mathrm{Gamma}(1, 1/\theta)$

$$f(x \mid \theta) = \theta e^{-\theta x} \qquad x > 0,\ \theta > 0$$

- MLE for θ is $\hat{\theta} = \frac{1}{\overline{X}_n}$. What is the asymptotic variance of $\hat{\theta}$?

Example

- Let $X_1, X_2, \ldots, X_n$ be a random sample from $\mathrm{Beta}(\theta, 1)$

$$f(x \mid \theta) = \theta x^{\theta-1} I_{(0,1)}(x) \qquad \theta > 0$$

- MLE for θ is

$$\hat{\theta} = \frac{n}{-\sum_{i=1}^{n} \log(X_i)}$$

What is the asymptotic variance of $\hat{\theta}$?

Asymptotic relative efficiency

Definition

If two estimators W_n and V_n satisfy

$$\sqrt{n}\,(W_n - \tau(\theta)) \xrightarrow{d} N(0, \sigma_W^2)$$
$$\text{and} \quad \sqrt{n}\,(V_n - \tau(\theta)) \xrightarrow{d} N(0, \sigma_V^2)$$

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

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

- If $\mathrm{ARE}(V_n, W_n) > 1$ then V_n is preferred
- If $\mathrm{ARE}(V_n, W_n) < 1$ then W_n is preferred

Example

- Let $X_1, X_2, \ldots, X_n$ be a random sample from $\mathrm{Poisson}(\lambda)$.
- Want to estimate $\tau(\lambda) = e^{-\lambda} = P(X = 0)$
- Compare the MLE and a "naive" approach: proportion of observations that are equal to zero.

ARE curve for example