

# Biostatistics 602 - Statistical Inference Lecture 17 Asymptotic Evaluation of Point Estimators

Hyun Min Kang

March 19th, 2013

## Bayes Estimator based on absolute error loss

Suppose that  $L(\theta, \hat{\theta}) = |\theta - \hat{\theta}|$ . The posterior expected loss is

$$\begin{aligned} E[L(\theta, \hat{\theta}(\mathbf{x}))] &= \int_{\Omega} |\theta - \hat{\theta}(\mathbf{x})| \pi(\theta|\mathbf{x}) d\theta \\ &= E[|\theta - \hat{\theta}||\mathbf{X} = \mathbf{x}] \\ &= \int_{-\infty}^{\hat{\theta}} -(\theta - \hat{\theta})\pi(\theta|\mathbf{x}) d\theta + \int_{\hat{\theta}}^{\infty} (\theta - \hat{\theta})\pi(\theta|\mathbf{x}) d\theta \end{aligned}$$

$$\frac{\partial}{\partial \hat{\theta}} E[L(\theta, \hat{\theta}(\mathbf{x}))] = \int_{-\infty}^{\hat{\theta}} \pi(\theta|\mathbf{x}) d\theta - \int_{\hat{\theta}}^{\infty} \pi(\theta|\mathbf{x}) d\theta = 0$$

Therefore,  $\hat{\theta}$  is posterior median.

## Last Lecture

- What is a Bayes Risk?
- What is the Bayes rule Estimator minimizing squared error loss?
- What is the Bayes rule Estimator minimizing absolute error loss?
- What are the tools for proving a point estimator is consistent?
- Can a biased estimator be consistent?

## Asymptotic Evaluation of Point Estimators

When the sample size  $n$  approaches infinity, the behaviors of an estimator are unknown as its *asymptotic* properties.

### Definition - Consistency

Let  $W_n = W_n(X_1, \dots, X_n) = W_n(\mathbf{X})$  be a sequence of estimators for  $\tau(\theta)$ . We say  $W_n$  is consistent for estimating  $\tau(\theta)$  if  $W_n \xrightarrow{P} \tau(\theta)$  under  $P_{\theta}$  for every  $\theta \in \Omega$ .

$W_n \xrightarrow{P} \tau(\theta)$  (converges in probability to  $\tau(\theta)$ ) means that, given any  $\epsilon > 0$ ,

$$\begin{aligned} \lim_{n \rightarrow \infty} \Pr(|W_n - \tau(\theta)| \geq \epsilon) &= 0 \\ \lim_{n \rightarrow \infty} \Pr(|W_n - \tau(\theta)| < \epsilon) &= 1 \end{aligned}$$

When  $|W_n - \tau(\theta)| < \epsilon$  can also be represented that  $W_n$  is close to  $\tau(\theta)$ . Consistency implies that the probability of  $W_n$  close to  $\tau(\theta)$  approaches to 1 as  $n$  goes to  $\infty$ .

## Tools for proving consistency

- Use definition (complicated)
- Chebychev's Inequality

$$\begin{aligned} \Pr(|W_n - \tau(\theta)| \geq \epsilon) &= \Pr((W_n - \tau(\theta))^2 \geq \epsilon^2) \\ &\leq \frac{E[W_n - \tau(\theta)]^2}{\epsilon^2} \\ &= \frac{\text{MSE}(W_n)}{\epsilon^2} = \frac{\text{Bias}^2(W_n) + \text{Var}(W_n)}{\epsilon^2} \end{aligned}$$

Need to show that both  $\text{Bias}(W_n)$  and  $\text{Var}(W_n)$  converges to zero

## Weak Law of Large Numbers

### Theorem 5.5.2

Let  $X_1, \dots, X_n$  be iid random variables with  $E(X) = \mu$  and  $\text{Var}(X) = \sigma^2 < \infty$ . Then  $\bar{X}_n$  converges in probability to  $\mu$ .  
i.e.  $\bar{X}_n \xrightarrow{P} \mu$ .

## Theorem for consistency

### Theorem 10.1.3

If  $W_n$  is a sequence of estimators of  $\tau(\theta)$  satisfying

- $\lim_{n \rightarrow \infty} \text{Bias}(W_n) = 0$ .
- $\lim_{n \rightarrow \infty} \text{Var}(W_n) = 0$ .

for all  $\theta$ , then  $W_n$  is consistent for  $\tau(\theta)$

## Consistent sequence of estimators

### Theorem 10.1.5

Let  $W_n$  is a consistent sequence of estimators of  $\tau(\theta)$ . Let  $a_n, b_n$  be sequences of constants satisfying

- $\lim_{n \rightarrow \infty} a_n = 1$
- $\lim_{n \rightarrow \infty} b_n = 0$ .

Then  $U_n = a_n W_n + b_n$  is also a consistent sequence of estimators of  $\tau(\theta)$ .

### Continuous Map Theorem

If  $W_n$  is consistent for  $\theta$  and  $g$  is a continuous function, then  $g(W_n)$  is consistent for  $g(\theta)$ .

## Example - Exponential Family

### Problem

Suppose  $X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} \text{Exponential}(\beta)$ .

- ① Propose a consistent estimator of the median.
- ② Propose a consistent estimator of  $\Pr(X \leq c)$  where  $c$  is constant.

## Consistent estimator of $\Pr(X \leq c)$ - Alternative Method

Define  $Y_i = I(X_i \leq c)$ . Then  $Y_i \stackrel{\text{i.i.d.}}{\sim} \text{Bernoulli}(p)$  where  $p = \Pr(X \leq c)$ .

$$\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i = \frac{1}{n} \sum_{i=1}^n I(X_i \leq c)$$

is consistent for  $p$  by Law of Large Numbers.

## Consistent estimator of $\Pr(X \leq c)$

$$\begin{aligned} \Pr(X \leq c) &= \int_0^c \frac{1}{\beta} e^{-x/\beta} dx \\ &= 1 - e^{-c/\beta} \end{aligned}$$

As  $\bar{X}$  is consistent for  $\beta$ ,  $1 - e^{-c/\beta}$  is continuous function of  $\beta$ .  
 By continuous mapping Theorem,  $g(\bar{X}) = 1 - e^{-c/\bar{X}}$  is consistent for  $\Pr(X \leq c) = 1 - e^{-c/\beta} = g(\beta)$

## Consistency of MLEs

### Theorem 10.1.6 - Consistency of MLEs

Suppose  $X_i \stackrel{\text{i.i.d.}}{\sim} f(x|\theta)$ . Let  $\hat{\theta}$  be the MLE of  $\theta$ , and  $\tau(\theta)$  be a continuous function of  $\theta$ . Then under "regularity conditions" on  $f(x|\theta)$ , the MLE of  $\tau(\theta)$  (i.e.  $\tau(\hat{\theta})$ ) is consistent for  $\tau(\theta)$ .

## Asymptotic Normality

### Definition: Asymptotic Normality

A statistic (or an estimator)  $W_n(\mathbf{X})$  is *asymptotically normal* if

$$\sqrt{n}(W_n - \tau(\theta)) \xrightarrow{d} \mathcal{N}(0, \nu(\theta))$$

for all  $\theta$   
 where  $\xrightarrow{d}$  stands for "converge in distribution"

- $\tau(\theta)$  : "asymptotic mean"
- $\nu(\theta)$  : "asymptotic variance"

We denote  $W_n \sim \mathcal{AN}\left(\tau(\theta), \frac{\nu(\theta)}{n}\right)$ .

## Example - Estimator of $\Pr(X \leq c)$

Define  $Y_i = I(X_i \leq c)$ . Then  $Y_i \stackrel{\text{i.i.d.}}{\sim} \text{Bernoulli}(p)$  where  $p = \Pr(X \leq c)$ .

$$\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i = \frac{1}{n} \sum_{i=1}^n I(X_i \leq c)$$

is consistent for  $p$ . Therefore,

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n I(X_i \leq c) &\sim \mathcal{AN}\left(\mathbb{E}(Y), \frac{\text{Var}(Y)}{n}\right) \\ &= \mathcal{AN}\left(p, \frac{p(1-p)}{n}\right) \end{aligned}$$

## Central Limit Theorem

### Central Limit Theorem

Assume  $X_i \stackrel{\text{i.i.d.}}{\sim} f(x|\theta)$  with finite mean  $\mu(\theta)$  and variance  $\sigma^2(\theta)$ .

$$\bar{X} \sim \mathcal{AN}\left(\mu(\theta), \frac{\sigma^2(\theta)}{n}\right)$$

$$\Leftrightarrow \sqrt{n}(\bar{X} - \mu(\theta)) \xrightarrow{d} \mathcal{N}(0, \sigma^2(\theta))$$

### Theorem 5.5.17 - Slutsky's Theorem

If  $X_n \xrightarrow{d} X$ ,  $Y_n \xrightarrow{P} a$ , where  $a$  is a constant,

- 1  $Y_n \cdot X_n \xrightarrow{d} aX$
- 2  $X_n + Y_n \xrightarrow{d} X + a$

## Example

Let  $X_1, \dots, X_n$  be iid samples with finite mean  $\mu$  and variance  $\sigma^2$ . Define

$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

By Central Limit Theorem,

$$\begin{aligned} \bar{X}_n &\sim \mathcal{AN}\left(\mu, \frac{\sigma^2}{n}\right) \\ \Leftrightarrow \sqrt{n}(\bar{X} - \mu) &\xrightarrow{d} \mathcal{N}(0, \sigma^2) \\ \Leftrightarrow \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} &\xrightarrow{d} \mathcal{N}(0, 1) \end{aligned}$$

## Example (cont'd)

$$\frac{\sqrt{n}(\bar{X} - \mu)}{S_n} = \frac{\sigma}{S_n} \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma}$$

We showed previously  $S_n^2 \xrightarrow{P} \sigma^2 \Rightarrow S_n \xrightarrow{P} \sigma \Rightarrow \sigma/S_n \xrightarrow{P} 1$ .  
 Therefore, By Slutsky's Theorem  $\frac{\sqrt{n}(\bar{X} - \mu)}{S_n} \xrightarrow{P} \mathcal{N}(0, 1)$ .

## Delta Method - Example

$X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} \text{Bernoulli}(p)$  where  $p \neq \frac{1}{2}$ , we want to know the asymptotic distribution of  $\bar{X}(1 - \bar{X})$ . By central limit Theorem,

$$\frac{\sqrt{n}(\bar{X}_n - p)}{\sqrt{p(1-p)}} \xrightarrow{d} \mathcal{N}(0, 1)$$

$$\Leftrightarrow \bar{X}_n \sim \mathcal{N}\left(p, \frac{p(1-p)}{n}\right)$$

Define  $g(y) = y(1 - y)$ , then  $\bar{X}(1 - \bar{X}) = g(\bar{X})$ .

$$g'(y) = (y - y^2)' = 1 - 2y$$

By Delta Method,

$$g(\bar{X}) = \bar{X}(1 - \bar{X}) \sim \mathcal{N}\left(g(p), [g'(p)]^2 \frac{p(1-p)}{n}\right)$$

$$= \mathcal{N}\left(p(1-p), (1-2p)^2 \frac{p(1-p)}{n}\right)$$

## Delta Method

### Theorem 5.5.24 - Delta Method

Assume  $W_n \sim \mathcal{N}\left(\theta, \frac{\nu(\theta)}{n}\right)$ . If a function  $g$  satisfies  $g'(\theta) \neq 0$ , then

$$g(W_n) \sim \mathcal{N}\left(g(\theta), [g'(\theta)]^2 \frac{\nu(\theta)}{n}\right)$$

## Asymptotic Normality

Given a statistic  $W_n(\mathbf{X})$ , for example  $\bar{X}, s_{\mathbf{X}}^2, e^{-\bar{X}}$

$$\sqrt{n}(W_n - \tau(\theta)) \xrightarrow{d} \mathcal{N}(0, \nu(\theta)) \quad \text{for all } \theta$$

$$\Leftrightarrow W_n \sim \mathcal{N}\left(\tau(\theta), \frac{\nu(\theta)}{n}\right)$$

Tools to show asymptotic normality

- ① Central Limit Theorem
- ② Slutsky Theorem
- ③ Delta Method (Theorem 5.5.24)

## Using Central Limit Theorem

$$\bar{X} \sim \mathcal{N}\left(\mu(\theta), \frac{\sigma^2(\theta)}{n}\right)$$

where  $\mu(\theta) = E(X)$ , and  $\sigma^2(\theta) = \text{Var}(X)$ .

For example, in order to get the asymptotic distribution of  $\frac{1}{n} \sum_{i=1}^n X_i^2$ , define  $Y_i = X_i^2$ , then

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n X_i^2 &= \frac{1}{n} \sum_{i=1}^n Y_i = \bar{Y} \\ &\sim \mathcal{N}\left(EY, \frac{\text{Var}(Y)}{n}\right) \\ &\sim \mathcal{N}\left(EX^2, \frac{\text{Var}(X^2)}{n}\right) \end{aligned}$$

## Using Slutsky Theorem

When  $X_n \xrightarrow{d} X$ ,  $Y_n \xrightarrow{P} a$ , then

- ①  $Y_n X_n \xrightarrow{d} aX$
- ②  $X_n + Y_n \xrightarrow{d} X + a$ .

## Using Delta Method (Theorem 5.5.24)

Assume  $W_n \sim \mathcal{N}\left(\theta, \frac{\nu(\theta)}{n}\right)$ . If a function  $g$  satisfies  $g'(\theta) \neq 0$ , then

$$g(W_n) \sim \mathcal{N}\left(g(\theta), [g'(\theta)]^2 \frac{\nu(\theta)}{n}\right)$$

## Example

### Problem

$$X_1, \dots, X_n \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \sigma^2) \quad \mu \neq 0$$

Find the asymptotic distribution of MLE of  $\mu^2$ .

### Solution

- ① It can be easily shown that MLE of  $\mu$  is  $\bar{X}$ .
- ② By the invariance property of MLE, MLE of  $\mu^2$  is  $\bar{X}^2$ .
- ③ By central limit theorem, we know that

$$\bar{X} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$$

## Solution (cont'd)

4 Define  $g(y) = y^2$ , and apply Delta Method.

$$g'(y) = 2y$$

$$\bar{X}^2 \sim \mathcal{AN}\left(g(\mu), [g'(\mu)]^2 \frac{\sigma^2}{n}\right)$$

$$\sim \mathcal{AN}\left(\mu^2, (2\mu)^2 \frac{\sigma^2}{n}\right)$$

## Example

### Problem

Let  $X_i \stackrel{\text{i.i.d.}}{\sim} \text{Poisson}(\lambda)$ . consider estimating  $\Pr(X = 0) = e^{-\lambda}$

Our estimators are

$$W_n = \frac{1}{n} \sum_{i=1}^n I(X_i = 0)$$

$$V_n = e^{-\bar{X}}$$

Determine which one is more asymptotically efficient estimator.

## Asymptotic Relative Efficiency (ARE)

If both estimators are consistent and asymptotic normal, we can compare their asymptotic variance.

### Definition 10.1.16 : Asymptotic Relative Efficiency

If two estimators  $W_n$  and  $V_n$  satisfy

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

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

The asymptotic relative efficiency (ARE) of  $V_n$  with respect to  $W_n$  is

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

If  $\text{ARE}(V_n, W_n) \geq 1$  for every  $\theta \in \Omega$ , then  $V_n$  is asymptotically more efficient than  $W_n$ .

## Solution - Asymptotic Distribution of $V_n$

$V_n(\mathbf{X}) = e^{-\bar{X}}$ , by CLT,

$$\bar{X} \sim \mathcal{AN}(EX, \text{Var}X/n) \sim \mathcal{AN}(\lambda, \lambda/n)$$

Define  $g(y) = e^{-y}$ , then  $V_n = g(\bar{X})$  and  $g'(y) = -e^{-y}$ . By Delta Method

$$\begin{aligned} V_n = e^{-\bar{X}} &\sim \mathcal{AN}\left(g(\lambda), [g'(\lambda)]^2 \frac{\lambda}{n}\right) \\ &\sim \mathcal{AN}\left(e^{-\lambda}, e^{-2\lambda} \frac{\lambda}{n}\right) \end{aligned}$$

## Solution - Asymptotic Distribution of $W_n$

Define  $Z_i = I(X_i = 0)$

$$W_n = \frac{1}{n} \sum_{i=1}^n I(X_i = 0) = \bar{Z}_n$$

$$Z_i \sim \text{Bernoulli}(E(Z))$$

$$E(Z) = \Pr(X = 0) = e^{-\lambda}$$

$$\text{Var}(Z) = e^{-\lambda}(1 - e^{-\lambda})$$

By CLT,

$$\begin{aligned} W_n = \bar{Z}_n &\sim \mathcal{AN}(E(Z), \text{Var}(Z)/n) \\ &\sim \mathcal{AN}\left(e^{-\lambda}, \frac{e^{-\lambda}(1 - e^{-\lambda})}{n}\right) \end{aligned}$$

## Asymptotic Efficiency

### Definition : Asymptotic Efficiency for iid samples

A sequence of estimators  $W_n$  is asymptotically efficient for  $\tau(\theta)$  if for all  $\theta \in \Omega$ ,

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

$$\iff W_n \sim \mathcal{AN}\left(\tau(\theta), \frac{[\tau'(\theta)]^2}{nI(\theta)}\right)$$

$$I(\theta) = E\left[\left\{\frac{\partial}{\partial\theta} \log f(X|\theta)\right\}^2 \mid \theta\right]$$

$$= -E\left[\frac{\partial^2}{\partial\theta^2} \log f(X|\theta) \mid \theta\right] \text{ (if interchangeability holds)}$$

Note:  $\frac{[\tau'(\theta)]^2}{nI(\theta)}$  is the C-R bound for unbiased estimators of  $\tau(\theta)$ .

## Solution - Calculating ARE

$$\begin{aligned} \text{ARE}(W_n, V_n) &= \frac{e^{-2\lambda}\lambda/n}{e^{-\lambda}(1 - e^{-\lambda})/n} \\ &= \frac{\lambda}{e^{\lambda}(1 - e^{-\lambda})} \\ &= \frac{\lambda}{e^{\lambda} - 1} \\ &= \frac{\lambda}{\left(1 + \lambda + \frac{\lambda^2}{2} + \frac{\lambda^3}{3!} + \dots\right) - 1} \\ &\leq 1 \quad (\forall \lambda \geq 0) \end{aligned}$$

Therefore  $W_n = \frac{1}{n} \sum I(X_i = 0)$  is less efficient than  $V_n$  (MLE), and ARE attains maximum at  $\lambda = 0$ .

## Asymptotic Efficiency of MLEs

### Theorem 10.1.12

Let  $X_1, \dots, X_n$  be iid samples from  $f(x|\theta)$ . Let  $\hat{\theta}$  denote the MLE of  $\theta$ . Under same regularity conditions,  $\hat{\theta}$  is consistent and asymptotically normal for  $\theta$ , i.e.

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} \mathcal{N}\left(0, \frac{1}{I(\theta)}\right) \text{ for every } \theta \in \Omega$$

And if  $\tau(\theta)$  is continuous and differentiable in  $\theta$ , then

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

$$\implies \tau(\hat{\theta}) \sim \mathcal{AN}\left(\tau(\theta), \frac{[\tau'(\theta)]^2}{nI(\theta)}\right)$$

Again, note that the asymptotic variance of  $\tau(\hat{\theta})$  is Cramer-Rao lower bound for unbiased estimators of  $\tau(\theta)$ .



## Summary

### Today

- Central Limit Theorem
- Slutsky Theorem
- Delta Method
- Asymptotic Relative Efficiency

### Next Lecture

- Hypothesis Testing