STAT 346/446 Lecture 5

Methods of Evaluating Point Estimators

CB Sections 7.3.1- 7.3.2, DS Section 7.6

Statistical Inference

- **Model:** Distribution of the population can be described with a distribution function (pmf or pdf) of a known form but with unknown parameters

$$f(x \mid \theta_1, \ldots, \theta_k)$$

  - So if we know the values of the parameters, we know all there is to know about the population.
- **Inference:** Have a sample $X_1, X_2, \ldots, X_n$ from $f(x \mid \theta)$ and want to use it to learn about the value of θ
- Point estimator: Any function of $X_1, X_2, \ldots, X_n$
  - Used to estimate θ
  - Some estimators are better than others

What is a good estimator?

- What is a good estimator of a parameter θ?
- θ is an unknown number
  - Has some unknown "true" value
- An estimator
$$W = W(X_1, X_2, \ldots, X_n)$$
is a random variable
  - W has a distribution (= sampling distribution)
  - **We evaluate W based on the properties of this distribution**
- Good properties:
  - $E(W) = \theta$ (we are correct on average)
  - $V(W)$ is small (W is an accurate estimator)

Unbiased estimators

Definition: Bias

Let W be a point estimator of a parameter θ. The **bias** of W is

$$\mathrm{bias}(W) = E(W) - \theta$$

- Book notation: $E_\theta(W)$

Definition: Unbiased estimator

Let W be a point estimator of a parameter θ. Then W is called an **unbiased estimator** if

$$E(W) = \theta \qquad \text{i.e. } \mathrm{bias}(W) = 0$$

Mean squared error

Mean squared error

Let W be a point estimator of a parameter θ.
The **mean squared error** of W is

$$\mathrm{MSE}(W) = \mathrm{E}\left((W - \theta)^2\right)$$

- Alternative evaluation criteria: mean absolute error $\mathrm{MAE}(W) = \mathrm{E}\left(|W - \theta|\right)$

Mean squared error

bias(W) = E(W) − θ

- MSE can be written as

$$\mathrm{MSE}(W) = \mathrm{E}\left((W - \theta)^2\right) = \mathrm{Var}(W) + (\mathrm{bias}(W))^2$$

= E(W) if W is unbiased

⇓

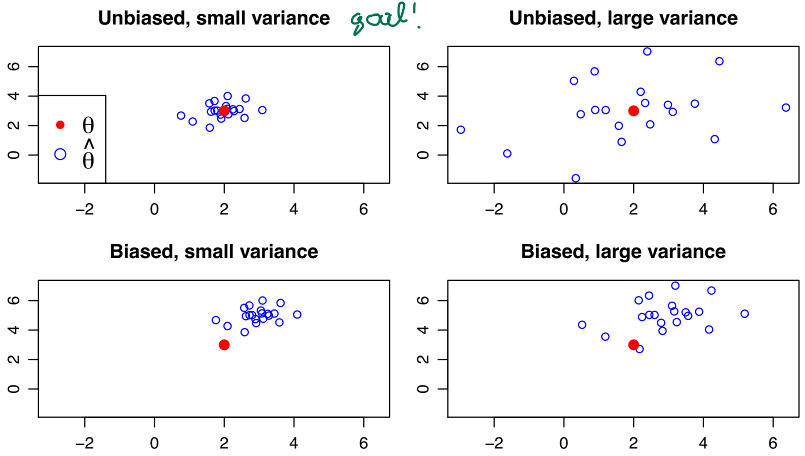
bias(W) = 0

⇒ If W is an unbiased estimator of θ then

MSE(W) = V(W)

Small MSE

- Want small variance and small bias



Example 1: Normal model

- Let $X_1, X_2, \ldots, X_n$ be a random sample from $N(\mu, \sigma^2)$ where μ and σ^2 are unknown.
- Find the MSE of $\overline{X}$ and S^2 as points estimators of μ and σ^2 respectively.

Better estimators?

- Are there other estimators of σ^2 or μ that have smaller MSE?
- What about the MLE for σ^2:

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 \quad = \frac{n-1}{n}S^2$$

Example 2: Binomial model

- Let $Y \sim \mathrm{Binomial}(n, p)$ where p is unknown. The Bayes and MLE estimators of p are:

$$\hat{p}^B = \frac{Y + \alpha}{\alpha + \beta + n} \qquad \text{and} \qquad \hat{p} = \frac{Y}{n}$$

- Which estimator is better?

More about MSE

- Notice that both bias and variance (and therefore the MSE) sometimes depend on the very same unknown parameter we are estimating.
  - E.g. for a normal random sample we have

$$\mathrm{MSE}(S^2) = \frac{2\sigma^{\cancel{2}4}}{n-1}$$

- The actual value of the MSE of an estimator is not as important as comparing the MSE of two estimators
  - E.g. we found that for a normal random sample we have

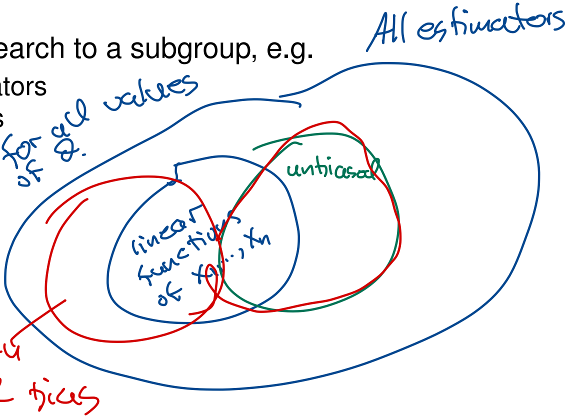
$$\frac{(2n-1)\sigma^4}{n^2} = \mathrm{MSE}(\hat{\sigma}^2) < \mathrm{MSE}(S^2)$$

    so in terms of MSE $\hat{\sigma}^2$ is a better estimator of σ^2 than our usual estimator S^2
  - But $\hat{\sigma}^2$ is biased - something that is traditionally frowned upon

Best Estimator?

- Can we find the best estimator in terms of MSE?
- Out of all possible functions of $X_1, \ldots, X_n$?
  - Too large space to search
- Typically restrict our search to a subgroup, e.g.
  - All unbiased estimators
  - All linear estimators

Sometimes we can find the best (e.g. lowest MSE) in a subgroup → for all values of θ.



Best Unbiased estimators

- Let W be an *unbiased* estimator of a parameter θ
  - Then $\mathrm{MSE}(W) = \mathrm{Var}(W)$
- In general, we want our unbiased estimators to have low variance (and hence low MSE)

Best unbiased estimator

= **Uniform minimum variance unbiased estimator (UMVUE)**

- The estimator that has the smallest variance in the set of all unbiased estimators
- Section 7.3 is mostly about different theoretical tricks to see if we have a best unbiased estimator

We didn't cover "equivariant" estimators so you can skip Example 7.3.6 and the text right above it

Best unbiased estimator (UMVUE)

Best unbiased estimator

An estimator W is called a **best unbiased estimator** of θ if

(i) $E(W) = \theta$ for all θ → i.e. W is unbiased.

(ii) For any other estimator U with $E(U) = \theta$ we have

$$\mathrm{Var}(W) \leq \mathrm{Var}(U)$$

And: W is a best unbiased estimator of $\tau(\theta)$ if $E(W) = \tau(\theta)$ and for any estimator U with $E(U) = \tau(\theta)$ we have $\mathrm{Var}(W) \leq \mathrm{Var}(U)$

- Can generalize to *same-bias estimators*
  - Only need to compare variances
  - If $\mathrm{bias}(W) = \mathrm{bias}(U)$ then

i.e. only need to compare variances

$$\mathrm{MSE}(W) - \mathrm{MSE}(U) = \mathrm{Var}(W) - \mathrm{Var}(U)$$

Best unbiased estimator

- Usually Impossible to find the estimator with the smallest MSE.
- Difficult even when restricting search to a group of estimators
- Sometimes we can come up with a *lower bound* for the variance of estimators in some group
- Argument we sometimes can use to show that we have found the lowest variance estimator:

Say we have a lower bound LB for the variance of all estimators in a group, i.e. (all unbiased)

$$Var(U) \geq LB \quad \text{for all } U \text{ in a group}$$

Say also that we have found an estimator W in same group with $Var(W) = LB$, then W is a best (minimum variance) estimator in the group.

Cramer-Rao Inequality

Theorem 7.3.9: Cramer-Rao Inequality

Let $X_1, X_2, \ldots, X_n$ be a sample with joint pdf $f(\mathbf{x} \mid \theta)$, and let $W(\mathbf{X}) = W(X_1, X_2, \ldots, X_n)$ be an estimator that has $\mathrm{Var}(W(\mathbf{X})) < \infty$ and satisfies

$$\frac{d}{d\theta} E\left(W(\mathbf{X})\right) = \int_{\mathcal{X}} \frac{d}{d\theta} W(\mathbf{x}) f(\mathbf{x} \mid \theta) d\mathbf{x} \qquad *$$

Then

$$\mathrm{Var}(W(\mathbf{X})) \geq \frac{\left(\frac{d}{d\theta} E(W(\mathbf{X}))\right)^2}{E\left(\left(\frac{d}{d\theta} \log(f(\mathbf{X} \mid \theta))\right)^2\right)}$$

= LB or CRLB = Cramer-Rao Lower Bound.

- Here $X_1, X_2, \ldots, X_n$ do not have to be independent

$$\frac{d}{d\theta} E\left(W(\mathbf{X})\right) = \int_{\mathcal{X}} \frac{d}{d\theta} W(\mathbf{x}) f(\mathbf{x} \mid \theta) d\mathbf{x} \qquad *$$

def. of E

$$\frac{d}{d\theta} E(W(\underline{X})) = \frac{d}{d\theta} \int \cdots \int w(\underline{x}) f(\underline{x} \mid \theta)\, dx_1 \cdots dx_n$$

$$= \frac{d}{d\theta} \int_{\mathcal{X}} w(\underline{x}) f(\underline{x} \mid \theta)\, d\underline{x}$$

n-dim integral.

condition $*$

$$= \int_{\mathcal{X}} \frac{\partial}{\partial\theta} w(\underline{x}) f(\underline{x} \mid \theta)\, d\underline{x}$$

This is true for all exponential families of distribution and most other distr. in this course.

$$\mathrm{Var}(W(\mathbf{X})) \geq \frac{\left(\frac{d}{d\theta} E(W(\mathbf{X}))\right)^2}{E\left(\left(\frac{d}{d\theta}\log(f(\mathbf{X} \mid \theta))\right)^2\right)}$$

numerator: $E(W(\underline{X}))$ is a function of θ

e.g. if $E(W(\underline{X})) = \theta$ we get

$$\left(\frac{d}{d\theta} E(W(\underline{X}))\right)^2 = \left(\frac{d}{d\theta}\theta\right)^2 = 1^2 = 1$$

denominator: joint pdf as a function of a rand. variable

$$\underset{\substack{\text{w.r.t.} \\ f(\underline{X}\mid\theta)}}{E_{\theta}}\left[\underbrace{\left(\frac{d}{d\theta}\underbrace{\log(f(\underline{X}\mid\theta))}_{\text{log-likelihood}}\right)^2}_{\text{function of } \underline{X}}\right]$$

$$\mathrm{Var}(W(\mathbf{X})) \geq \frac{\left(\frac{d}{d\theta} E(W(\mathbf{X}))\right)^2}{E\left(\left(\frac{d}{d\theta}\log(f(\mathbf{X} \mid \theta))\right)^2\right)}$$ ***

section 4.7

proof outline: Based on the Cauchy - Schwarz Inequality that implies that for any random variables X and Y:

$$\left(\mathrm{Cov}(X,Y)\right)^2 \leq \mathrm{Var}(X)\,\mathrm{Var}(Y)$$

$$\Rightarrow \quad \mathrm{Var}(X) \geq \frac{\mathrm{Cov}(X,Y)^2}{\mathrm{Var}(Y)} \quad **$$

set $X = W(\underline{X})$ (the estimator)

and $Y = \frac{\partial}{\partial\theta} \log f(\underline{X} \mid \theta)$, puts into ** to get ***

Cramer-Rao Inequality – iid case

Theorem 7.3.10: Cramer-Rao Inequality – iid case

Same conditions as before, but now suppose $X_1, X_2, \ldots, X_n$ are independent. Then

$$\mathrm{Var}(W(\mathbf{X})) \geq \frac{\left(\frac{d}{d\theta} E(W(\mathbf{X}))\right)^2}{nE\left(\left(\frac{d}{d\theta}\log(f(X \mid \theta))\right)^2\right)} \leftarrow \text{Fisher information}$$

- Here we don't need the **joint** pdf

i.e. $E\left(\left[\frac{d}{d\theta}\log\left(\prod_{i=1}^{n} f(x_i \mid \theta)\right)\right]^2\right) = E\left(\left[\sum_{i=1}^{n} \frac{d}{d\theta}\log f(x_i \mid \theta)\right]^2\right)$

$= \ldots = n E\left(\frac{d}{d\theta}\log f(X \mid \theta)\right)$

Fisher information

- The denominator in the Cramer-Rao lower bound is called the **information number** or **Fisher Information** $I(\theta)$.
- Larger information number $\Rightarrow$ smaller lower bound on variance
- Short-cuts to calculate Fisher information

$$I(\theta) = E\left(\left(\frac{d}{d\theta}\log(f(\mathbf{X} \mid \theta))\right)^2\right)$$

$$= nE\left(\left(\frac{d}{d\theta}\log(f(X \mid \theta))\right)^2\right) \quad \text{if iid}$$

$$= -nE\left(\frac{d^2}{d\theta^2}\log(f(X \mid \theta))\right) \quad \text{if exponential family}$$

$\frac{\delta^2}{\delta\theta^2} f$ instead of $\left(\frac{\delta}{\delta\theta} f\right)^2$

Fisher Information

Lemma

If $f(x \mid \theta)$ satisfies

$$\frac{d}{d\theta} E\left(\frac{d}{d\theta}\log(f(X \mid \theta))\right) = \int_{\mathcal{X}} \frac{d}{d\theta}\left[\left(\frac{d}{d\theta}\log f(x \mid \theta)\right) f(x \mid \theta)\right] dx \quad (1)$$

then

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

- The condition in (1) holds for all exponential families

Example 1

- Let $X_1, X_2, \ldots, X_n$ be iid $\mathrm{Gamma}(1, \theta)$
  1. Find the Fisher information
  2. Find the CRLB for $W = \overline{X}$ as an estimator of θ
  3. Find the CRLB for $U = \frac{n}{n+1}\overline{X}^2$ as an estimator of θ^2

Saw last time: (details on the white board)

2.: $W = \overline{X}$ attains the CRLB, i.e.

$Var(\overline{X}) = CRLB$ ← efficient est.

$\Rightarrow$ $\overline{X}$ is the UMVUE of θ

3.: $U = \frac{n}{n+1}\overline{X}^2$ does not attain the CRLB ← not efficient estimator

i.e. $Var(\frac{n}{n+1}\overline{X}^2) > CRLB$

could not use CRLB to show that U is the UMVUE of θ^2

Note: Read the Poisson example in the Book - Example 7.3.8

Cramer-Rao Lower bound

- Puts a lower bound on the variance of all estimators - given that some "nicety" conditions hold
- Our estimators can't have a smaller variance than the Cramer-Rao lower bound (CRLB)
- So if our unbiased estimator has a variance that is equal to CRLB we know that it *is* the best unbiased estimator

Efficient estimators

An estimator W is called and **efficient estimator** if it has a variance that is equal to its CRLB.

- Note that a best unbiased estimator is not necessarily efficient
- The CRLB may not be obtainable

Example 2

- Let $X_1, X_2, \ldots, X_n$ be iid $\mathrm{N}(\mu, \sigma^2)$ where μ is known.
- Is S^2 an efficient estimator of σ^2?

$$\mathrm{Var}(S^2) = \frac{2\sigma^4}{n-1}$$

$$\mathrm{CRLB}(S^2) = \frac{2\sigma^2}{n} < \frac{2\sigma^2}{n-1} = \mathrm{Var}(S^2)$$

on the board

$\Rightarrow$ S^2 is <u>not</u> an efficient estimator of σ^2

Attaining the CBLB

Corollary 7.3.15

Assume same conditions as in the CR Theorem.

An unbiased estimator W of $\tau(\theta)$ attains the CRLB if and only if there exists a function $a(\theta)$ such that

→ i.e. is efficient

$$a(\theta)\,(W(\mathbf{x}) - \tau(\theta)) = \frac{\partial}{\partial\theta} \log L(\theta \mid \mathbf{x})$$

"

$f(\underline{x} \mid \theta)$

joint pdf.

Use:

- If condition does not hold then W is not efficient
- Can be used to find a UMVUE

Example: Back to the S^2, $X_1, \dots, X_n$ iid $N(\mu, \theta)$

$\tau(\theta) = \theta \qquad W(\underline{X}) = S^2$

$$\log(L(\theta)) = \log\left((2\pi)^{-n/2} \theta^{-n/2} \exp\left(-\frac{1}{2\theta} \sum_{i=1}^{n} (x_i - \mu)^2 \right) \right)$$

$$= -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log\theta - \frac{1}{2\theta}\sum_{i=1}^{n}(x_i-\mu)^2$$

$$\frac{d}{d\theta}\log(L(\theta)) = -\frac{n}{2\theta} + \frac{1}{2\theta^2}\sum_{i=1}^{n}(x_i-\mu)^2$$

$$= \frac{1}{2\theta^2}\left(\sum_{i=1}^{n}(x_i-\mu)^2 - \theta \right) \quad *$$

$\Rightarrow \sum_{i=1}^{n}(x_i-\mu)^2$ is the UMVUE of $\theta = \sigma^2$ but it can't be calculated unless μ is known

* Cannot be written as $a(\theta)(S^2 - \theta)$

$\Rightarrow S^2$ is not efficient

Note: S^2 <u>is</u> the UMVUE for σ^2, just need other tools.

Back to Gamma example:

$X_1, \dots, X_n$ iid Gamma$(1, \theta)$ = Expo(θ)

$$\log(f(\underline{x} \mid \theta)) = \log\left(\theta^{-n} e^{-n\bar{x}/\theta}\right) = -n \log\theta - \frac{n\bar{x}}{\theta}$$

$$\frac{d}{d\theta} \log(f(\underline{x} \mid \theta)) = -\frac{n}{\theta} + \frac{n\bar{x}}{\theta^2} = \frac{n}{\theta^2}\left(\bar{x} - \theta\right) \; *$$

$\Rightarrow$ $\bar{X}$ is efficient est. of θ

$\Rightarrow$ $\bar{X}$ is the UMVUE of θ

$*$ cannot be written as

$$a(\theta)\left(\frac{n}{n+1}\bar{x}^2 - \theta^2\right)$$

$\Rightarrow$ $\frac{n}{n+1}\bar{X}^2$ is not an efficient est. of θ^2

(it is the UMVUE, need other tools)

Note: did not have to find the variance of the statistic.