Introduction to the Theory of Statistics Part 2

PM522b

Meredith Franklin

Division of Biostatistics, University of Southern California

Slides 4, 2019

Topics covered

- Evaluating Estimators:
  - Bias
  - Mean Squared Error (MSE)
  - Minimum Variance Unbiased Estimation (MVUE)
    - Cramer-Rao Inequality
    - Fisher Information and the Cramer Rao Lower Bound
  - Sufficiency and unbiasedness: Rao-Blackwell Theorem, Lehmann-Scheffé Theorem

Estimators

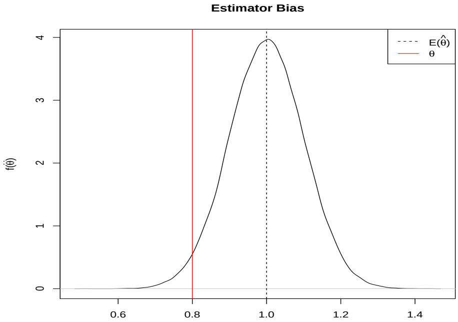
- An estimator is a rule that tells us how to calculate the value of an estimate based on the observations in our random sample.
- A point estimator of an unknown population parameter θ is denoted $\hat{\theta}$.
- An estimator is expressed as a formula, for example the sample mean $\bar{X}$ is a point estimator of the population mean μ
- Some estimators are good and some are bad; what defines a good estimator of our target parameter?
- Bias, Mean Square Error, Minimum Variance Unbiased Estimator, Sufficient Estimator

Evaluating Estimators: Bias

- Some estimators are good and some are bad; what defines a good estimator of our target parameter?
- An estimator $\hat{\theta}$ is *unbiased* if $\mathrm{E}(\hat{\theta}) = \theta$
- The difference is the bias of the estimator.

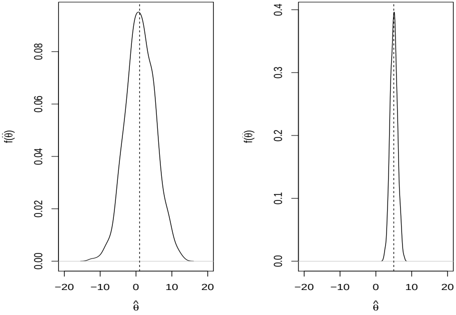
Bias

$$\text{Bias}{=}\mathrm{E}(\hat{\theta}) - \theta$$



Evaluating Estimators: Bias

- We prefer that our estimator is unbiased and that the spread of the distribution around $\mathrm{E}(\hat{\theta})$ is small
- If $\mathrm{Var}(\hat{\theta})$ is small we are more certain that $\hat{\theta}$ will be "close" to θ



Evaluating Estimators: Bias

- Example of an unbiased estimator is $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, since $\mathrm{E}(\bar{X}) = \mu$
- However, $\mathrm{Var}(\bar{X})$ is a *biased* estimator of the variance since $\mathrm{Var}(\bar{X}) = \sigma^2/n$
- The variance of the sample mean (i.e. the sample variance) is often denoted by:

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

- We can show that it is biased because:

Evaluating Estimators: Bias

- Since S^{*2} is a biased estimate, the following modified statistic is used as it is unbiased:

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

Unbiased estimator of the sample variance

$$\begin{aligned}
S^2 &= \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 \\
&= \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 \\
&= \frac{1}{n-1}\sum_{i=1}^{n}(X_i^2 - 2X_i\bar{X} + \bar{X}^2) \\
&= \frac{1}{n-1}\sum_{i=1}^{n}X_i^2 - \frac{n}{n-1}\bar{X}^2
\end{aligned}$$

Evaluating Estimators: Bias

Con't: Unbiased estimator of the sample variance

Showing unbiasedness by assuming S^2 is our estimate of σ^2

$$\begin{aligned}
E(S^2) &= \frac{1}{n-1}\sum_{i=1}^{n} E(X_i^2) - \frac{n}{n-1}E(\bar{X}^2) \\
&= \frac{1}{n-1}\sum_{i=1}^{n}\{V(X_i) + [E(X_i)]^2\} - \frac{n}{n-1}\{V(\bar{X}) + [E(\bar{X}_i)]^2\} \\
&= \frac{1}{n-1}n(\sigma^2 + \mu^2) - \frac{n}{n-1}(\frac{\sigma^2}{n} + \mu^2) \\
&= \sigma^2
\end{aligned}$$

Evaluating Estimators: Bias

From the previous result,

$$S^{*2} = \frac{n-1}{n} S^2$$

So it follows that

$$E(S^{*2}) = \frac{n-1}{n}\ E(S^2) = \frac{n-1}{n}\sigma^2$$

So its bias is,

$$\text{Bias}(S^{*2}) = E(S^{*2}) - \sigma^2 = -\frac{\sigma^2}{n}$$

Evaluating Estimators: Mean Squared Error

The deviation of $\hat{\theta}$ from the true value of θ by $|\hat{\theta} - \theta|$ (absolute error) or $(\hat{\theta} - \theta)^2$ gives a measure of the quality of the estimator. Since $\hat{\theta}$ is a random variable, we take an average. This gives us the mean squared error (MSE), defined as the expected value of the squared difference between the estimator and the parameter:

$$\text{MSE}(\hat{\theta}) = \text{E}[(\hat{\theta} - \theta)^2]$$

It is a function of both the variance and the bias of the estimator:

$$\begin{aligned} MSE(\hat{\theta}) &= E[(\hat{\theta} - \theta)^2] \\ &= E[(\hat{\theta} - E(\hat{\theta}) + E(\hat{\theta}) - \theta)^2] \\ &= E[(\hat{\theta} - E(\hat{\theta}))^2] + 2(E(\hat{\theta}) - \theta)E[\hat{\theta} - E(\hat{\theta})] + [E(\hat{\theta}) - \theta]^2 \\ &= E[(\hat{\theta} - E(\hat{\theta}))^2] + [E(\hat{\theta}) - \theta]^2 \\ &= Var(\hat{\theta}) + [B(\hat{\theta})]^2 \end{aligned}$$

Since $\text{E}[\hat{\theta} - \text{E}(\hat{\theta})] = 0$

Evaluating Estimators: Mean Squared Error

- Other terms for $(\hat{\theta} - \theta)^2$ are quadratic loss and risk function.
- While the mean absolute error $E|\hat{\theta} - \theta|$ is a reasonable alternative, the MSE has the advantage that it can be broken down into the variance of the estimator plus its squared bias.
- You will often hear of the saying "bias-variance tradeoff" and the interpretation of MSE in terms of both variance and bias of an estimator allows us to see how this tradeoff works.
- If we want to minimize the MSE, we can try to minimize the bias. However the only way to make the bias zero is to make the variance large.
- To get the smallest MSE you need some bias to reduce the variance.

Evaluating Estimators: Mean Squared Error

- If we calculate the MSE for our estimate of the sample variance:

$$\mathrm{Var}(\bar{X}) = S^{*2} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

- We need Var(S^{*2}) and B(S^{*2}) =E(S^{*2}) $- \sigma^2$ (which we solved previously)

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

- The MSE is variance + bias^2:

$$\begin{aligned} E(S^{*2} - \sigma^2)^2 &= \frac{2(n-1)\sigma^4}{n^2} + (\frac{n-1}{n}\sigma^2 - \sigma^2)^2 \\ &= \frac{2(n-1)\sigma^4}{n^2} + \frac{\sigma^4}{n^2} \\ &= \frac{(2n-1)}{n^2}\sigma^4 \end{aligned}$$

Evaluating Estimators: Mean Squared Error

- If we calculate the MSE for our *unbiased* estimate of the sample variance:

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

- We need $\text{Var}(S^2)$, and the MSE is variance $+$ bias^2
- From Theorem 5.3.1 (CB) we derive $\text{Var}(S^2) = \frac{2\sigma^4}{n-1}$ so:

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

- So $\text{MSE}(\text{S}^{*2}) < \text{MSE}(\text{S}^2)$
- Our unbiased estimator has larger MSE than our biased one
- This shows that trading off variance for bias improved the MSE.

Evaluating Estimators: Best Estimators

- We want to narrow down the set of all possible estimators to those that are best. One approach is to only choose those that are unbiased.
- If $\hat{\theta}_1$ and $\hat{\theta}_2$ are both unbiased for θ, $E(\hat{\theta}_1) = \theta$ and $E(\hat{\theta}_2) = \theta$
- Their MSEs are therefore equal to their variances.
- In choosing between the two, we should choose the estimator with the smallest variance.

Example in class.

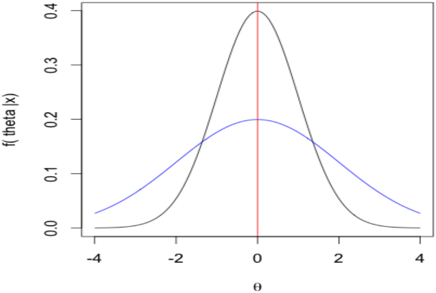
Evaluating Estimators: Minimum Variance Unbiased Estimators (MVUE)

- In the event that we have an unbiased estimator of θ, the MSE is equal to Var($\hat{\theta}$).
- It is for this reason that we get the term "standard error"; it is the standard deviation of the MSE of the unbiased estimate, i.e. $\sqrt{(Var(\hat{\theta})}$.
- Our preference in estimators are such that they have the smallest MSE combined with the requirement that it is unbiased.
- If we have two unbiased estimators W_1 and W_2 for a parameter θ then we will choose the estimator with the smallest variance.
- This leads us to searching for the "minimum variance unbiased estimators" (MVUE). These are considered the best unbiased estimators.
- We have certain restrictions: the variance of an unbiased estimator $\hat{\theta}$ cannot be less than a lower bound.
- This lower bound is determined by the Cramer-Rao inequality.

Evaluating Estimators: MVUE

Best Unbiased Estimators

An estimator W* is a best unbiased estimator of $\tau(\theta)$ if it satisfies $E(W^*) = \tau(\theta)$ for all θ and for any other estimator W with $E(W) = \tau(\theta)$, $Var(W^*) \leq Var(W)$ for all θ. W* is also called the uniform minimum variance unbiased estimator (UMVUE) of $\tau(\theta)$.



Evaluating Estimators: MVUE

- The sharpness of the pdf/pmf determines how accurately we can estimate a parameter.
- If we think about a likelihood plot (parameter value vs likelihood) in terms of how sharp or wide it is, we can relate this to how sure we are about a parameter.
- The second derivative is a measure of curvature:
  - high curvature means narrow spread of values which also means we are more sure about a parameter estimate.
- The variance of an estimated parameter is proportional to 1/curvature.

Evaluating Estimators: MVUE

It can be challenging to find the best unbiased estimator, so the process of finding the lower bound of the variance helps. Basically for estimating a parameter of a distribution we can specify a lower bound on the variance ($B(\theta)$ in CB) of any unbiased estimator of the parameter. If we can find that an unbiased estimator W^* satisfying $Var(W^*) = B(\theta)$ we have the best unbiased estimator (the UMVUE). This approach is formalized with the Camer-Rao Lower Bound.

Cramer-Rao Inequality

Given $X_1, ..., X_n$ iid with pdf $f(x|\theta)$, for an unbiased estimator $\hat{\theta}$ of θ we define:

$$Var(\hat{\theta}) \geq \frac{[\frac{d}{d\theta} E(\hat{\theta})]^2}{E[(\frac{\partial}{\partial\theta} \log f(\mathbf{X}|\theta))^2]}$$

(Proof and other forms in class)

Evaluating Estimators: MVUE

The partial derivative (with respect to θ) of the log likelihood is called the score. The first moment of the score is the expected value, the second moment (expected value of the second derivative, or variance of the score) is the Fisher Information, $I(\theta)$.

$$\begin{aligned} I(\theta) &= E[(\frac{\partial}{\partial\theta} \log f(X|\theta))^2] \\ &= E[(\ell^{'}(\theta))^2] \\ &= \int (\ell^{'}(\theta))^2 f(x|\theta) dx \end{aligned}$$

The information gives a bound on the variance of the best unbiased estimator (MVUE) of θ. As the information increases, we know more about the parameter and the bound on the variance gets smaller through:

$$Var(\hat{\theta}) \geq \frac{[\frac{d}{d\theta} E(\hat{\theta})]^2}{nI(\theta)}$$

This is known as the Cramer-Rao lower bound (CRLB).

Evaluating Estimators: MVUE

The other forms of $I(\theta)$ (proven in class) are:

$$\begin{aligned} I(\theta) &= -E[(\ell^{''}(\theta))] \\ &= \int (\ell^{''}(\theta)) f(x|\theta) dx \end{aligned}$$

and

$$I(\theta) = Var[\ell^{'}(\theta)]$$

We see that the information, which is in the denominator of the CRLB, is the expectation of the second derivative. As we mentioned above, the second derivative relates to curvature, and the variance of an estimator can be thought of as being related to 1/curvature. The higher the curvature, the smaller the variance, and the more sure we are about the parameter estimate.

Evaluating Estimators: MVUE

Example: Finding the Fisher Information

Consider a coin toss where $X_1, X_2, ..., X_n$ are Bernoulli
$f(x|p) = p^x(1-p)^{1-x}; x = 0, 1, 0 < p < 1$
Find the Fisher information. Recall $E(X) = p, V(X) = p(1-p)$

We saw before that $E(\bar{X})$ is and unbiased estimator of p
$E(\bar{X}) = E(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n}np = p$
And $\text{Var}(\bar{X}) = \text{Var}(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n^2}\sum_{i=1}^{n} \text{Var}(X_i) = \frac{1}{n^2}np(1-p) = \frac{p(1-p)}{n}$
Taking the log of the pdf,

$$\ell(p) = x\log p + (1-x)\log(1-p)$$
$$\ell^{'}(p) = \frac{x}{p} - \frac{1-x}{1-p}$$
$$\ell^{''}(p) = -\frac{x}{p^2} - \frac{1-x}{(1-p)^2}$$

Evaluating Estimators: MVUE

Example: Finding Fisher Information, con't

Since $\hat{p} = X$ we have an unbiased estimator, so

$$
\begin{aligned}
I(\theta) &= -E[\ell^{''}(\theta)] \\
&= \frac{E(x)}{p^2} + \frac{1 - E(x)}{(1-p)^2} \\
&= \frac{1}{p} + \frac{1}{1-p} \\
&= \frac{1}{p(1-p)}
\end{aligned}
$$

And if we have a random sample $X_1, ..., X_n$, then the Fisher information is $I_n(p) = nI(p) = \frac{n}{p(1-p)}$. The CRLB is $\frac{1}{I_n(\theta)}$, so for this example it is $p(1-p)/n$. Comparing $V(\bar{X})$ to $I_n(p)$, we see they are the same, so $\bar{X}$ is the UMVUE for p.

Attainment

In class discussion.

Sufficiency and Unbiasedness

The CRLB tells us that if $\hat{\theta}$ is an unbiased estimator for θ then its variance cannot be less than $[nI(\theta)]^{-1}$, where $I(\theta)$ is the Fisher Information.
It is possible that:

1. It isn't possible to find an exact formula for the variance of an unbiased estimator
2. The smallest possible variance of an unbiased estimator does not attain the CRLB (i.e. it is strictly larger and the lower bound is not always attainable).

In such cases we do not know if an estimator is the best unbiased estimator. We can use the connection between sufficient statistics and our best estimators as stated through the Rao-Blackwell theorem.

Sufficiency and Unbiasedness

If we want an estimator with small variance we can search estimators which are functions of sufficient statistics.

Rao-Blackwell Theorem

Let $\hat{\theta}$ be an unbiased estimator for θ such that $Var(\hat{\theta}) < \infty$ for all θ. Let T be a sufficient statistic for θ and define $\phi(T) = E[\hat{\theta}|T = t]$, then $E[\phi(T)] = \theta$ and $Var(\phi(T)) \leq V(\hat{\theta})$. $\phi(T)$ is a uniformly better unbiased estimator of θ

In words, given an estimator for the parameter θ, the conditional expectation of the estimator given a sufficient statistic is a better estimator for θ (it is never worse). Thus, it is possible that we can improve upon an unbiased estimator through the Rao-Blackwell theorem, however it may be challenging to find the conditional distribution. The theorem is useful in that if we restrict our attention to functions of a sufficient statistic we can improve our search for MVUEs. (Proof and Example in class)

Uniqueness

The Rao Blackwell theorem tells us that for an unbiased estimator $\hat{\theta}$ of θ, conditioning $\hat{\theta}$ on a sufficient statistic T for θ results in uniformly better unbiased estimator of θ.

A few more things:

- $\phi(T) = E(\hat{\theta}|T)$ is a function only of the sample (i.e. it is independent of θ).
- CB Example 7.3.18 illustrates that conditioning on an insufficient statistic results in $\phi(T)$ being dependent on θ.
- All of this indicates that we should look at estimators based on sufficient statistics to find the best unbiased estimators of θ.

Suppose $E(\hat{\theta}|T) = \theta$ and ϕ is based on a sufficient statistic. How do we know it is best unbiased? If ϕ attains the CRLB then it is.

Uniqueness

There is only one best unbiased estimator (MVUE). So, of all unbiased estimators, those with minimum variance are "best". We want to find the minimum variance unbiased estimator of θ. The MVUE is unique.

Lehmann-Scheffé Theorem

This theorem ties together sufficiency, completeness and uniqueness.
Lehmann-Scheffé states that uniased estimators based on complete sufficient statistics are unique.
Let T be a complete sufficient statistic for a parameter θ, and let $\phi(T)$ be any estimator based only on T. Then $\phi(T)$ is the unique best unbiased estimator of its expected value.

Proof: T is a complete sufficient statistic and $\phi(T)$ is an estimator with $E[\phi(T)] = \theta$. From the theorem above we know that $\phi(T)$ is the best unbiased estimator of θ and we have shown that best unbiased estimators are unique.