Lecture 8: Comparison of Estimators

MATH 667-01
Statistical Inference
University of Louisville

October 3, 2017
Last modified: 10/4/2017

Introduction

- We define loss and risk functions for comparing methods for finding estimators of the unknown parameter(s) in a model which are discussed in Sections 7.3 of Casella and Berger (2001)[1].
- We also define the bias of an estimator, and show the bias-variance decomposition for the mean squared error of an estimator.
- We present a couple of examples comparing the sample mean and sample median, and compare the performance of these estimators by simulation studies using R code.

[1] Casella, G. and Berger, R. (2001). Statistical Inference, second edition. Duxbury Press.

Introduction

- We have discussed several methods for obtaining estimators of unknown parameters $\boldsymbol{\theta}$ based on an observed random sample $X_1, \ldots, X_n$ from a population with pdf/pmf $f(x|\boldsymbol{\theta})$.
- We would like to know if an estimator is likely to give "good" estimates of the parameters based on observed data $x_1, \ldots, x_n$. Or given competing estimators, we want to know which one is likely to perform "best" at estimating the parameters.
- So, we need to define what we mean by "good" or "best" and describe a mathematical framework for evaluating estimators.

Loss and Risk Functions

- *Definition L8.1* (p.348): Suppose $\boldsymbol{X}$ is a random vector with pmf/pdf $f(\boldsymbol{x}|\boldsymbol{\theta})$ where $\boldsymbol{\theta}$ is in the parameter space Θ. Define $\mathcal{A}$ to be the *action space* giving the set of allowable decisions that can be made regarding $\boldsymbol{\theta}$. Then a *loss function* $L(\boldsymbol{\theta}, \boldsymbol{a})$ is a real-valued function that assigns a number in the interval $[0, \infty)$ to each $\boldsymbol{\theta} \in \Theta$ and $\boldsymbol{a} \in \mathcal{A}$.
- Some common loss functions for $\Theta \subset \mathbb{R}$ are
  - squared error loss: $L(\theta, a) = (\theta - a)^2$
  - absolute error loss: $L(\theta, a) = |\theta - a|$.
- *Definition L8.2* (p.349): The *risk function* of a rule $\delta(\boldsymbol{x})$ for estimating θ based on observed data $\boldsymbol{x}$ is

$$R(\theta, \delta) = \mathsf{E}_\theta \left[L(\theta, \delta(\boldsymbol{X})) \right].$$

Mean Squared Error and Bias

- The risk function of an estimator under squared-error loss has another name.
- *Definition L8.3* (Def 7.3.1 on p.330): The *mean squared error* (MSE) of any estimator W of a parameter θ is the function of θ defined by $\mathsf{E}_\theta\left[(W-\theta)^2\right]$.
- *Definition L8.4* (Def 7.3.2 on p.330): The *bias* of a point estimator W, of a parameter θ, is the difference between the expected value of W and θ. That is, $\mathsf{Bias}_\theta[W] = \mathsf{E}_\theta\left[W-\theta\right]$. An estimator is called *unbiased* if it satisfies $\mathsf{E}_\theta[W]=\theta$ for all θ.

Mean Squared Error and Bias

- The MSE can be split into two parts, one that measures the variability (precision) and one that measures the bias (accuracy).
- *Theorem L8.1*: If $\mathsf{Var}[W]$ exists, then

$$\mathsf{E}_\theta\left[(W-\theta)^2\right] = \mathsf{Var}_\theta[W] + (\mathsf{Bias}_\theta[W])^2.$$

- *Proof of Theorem L8.1*:

$$\begin{aligned}\mathsf{E}_\theta\left[(W-\theta)^2\right] &= \mathsf{E}_\theta\left[((W-\mathsf{E}_\theta[W]) + (\mathsf{E}_\theta[W]-\theta))^2\right] \\ &= \mathsf{E}_\theta[(W-\mathsf{E}_\theta[W])^2] + 2\mathsf{E}_\theta\left[W-\mathsf{E}_\theta[W]\right](\mathsf{E}_\theta[W]-\theta) + \\ &\quad (\mathsf{E}_\theta[W]-\theta)^2 \\ &= \mathsf{E}_\theta\left[(W-\mathsf{E}_\theta[W])^2\right] + (\mathsf{E}_\theta[W]-\theta)^2 \\ &= \mathsf{Var}_\theta[W] + (\mathsf{Bias}_\theta[W])^2\end{aligned}$$

Simulation: Comparison of the Sample Mean and Median

- Suppose $X_1, \ldots, X_9$ are iid Normal$(0, 100)$ random variables and we want to estimate the population mean (equivalently, the population median).
- In R, the data can be simulated using the following code:

```
> set.seed(57487)
> x=rnorm(9,sd=10)
> round(x)
[1]   2   1  12   2 -10  -2  -7  18  -7
```

- The estimates (sample mean and sample median) can be computed as follows.

```
> mean(x)
[1] 0.9751672
> median(x)
[1] 1.381941
```

so for the sample $\boldsymbol{x} = (x_1, \ldots, x_9)$, the squared error losses are $L_2(\mu, \bar{x}) = (0 - 0.9751672)^2 \approx 0.95$ and $L_2(\mu, x_{(5)}) = (0 - 1.381941)^2 \approx 1.91$, and the absolute error losses are $L_1(\mu, \bar{x}) = |0 - 0.9751672| \approx 0.975$ and $L_1(\mu, x_{(5)}) = |0 - 1.381941| \approx 1.38$.

Simulation: Comparison of Mean and Median

- To see that what happens on average, we need to repeat this simulation many times.
- The following R code repeats this calculation 10 000 times.

```
> set.seed(93873)
> R=1e4
> sample.mean=rep(0,R)
> sample.median=rep(0,R)
> for (i in 1:R){
+ x=rnorm(9,sd=10)
+ sample.mean[i]=mean(x)
+ sample.median[i]=median(x)
+ cat(round(x)," mean=",round(mean(x),2)," median=",
+ round(median(x),2),"\n")
+ }
```

Simulation: Comparison of Mean and Median

Here is some output from the 10 000 simulated data sets.

```
-5 -7 13 9 15 3 -2 12 -2  mean= 3.8  median= 2.72
-15 -2 11 -13 2 -3 5 0 2  mean= -1.5  median= -0.4
-2 1 -11 -2 -13 -9 -6 -3 7  mean= -4.08  median= -2.83
24 -2 10 4 -4 -4 14 -8 11  mean= 4.88  median= 4.27
-5 0 18 15 5 13 1 7 3  mean= 6.51  median= 5.43
...
8 -8 -15 -12 7 6 22 -23 -9  mean= -2.62  median= -8.41
1 11 -9 -7 1 12 13 -2 16  mean= 3.95  median= 0.94
-4 3 2 15 6 5 4 13 16  mean= 6.76  median= 4.72
20 -6 -2 8 -1 -12 17 -5 -14  mean= 0.52  median= -1.6
-1 -4 -15 -4 24 -18 -10 17 -6  mean= -1.99  median= -3.67
```

Simulation: Comparison of Mean and Median

- Now, we can use the 10 000 simulations to estimate the risk under each of the loss functions.

```
> mean((0-sample.mean)^2)
[1] 11.37269
> mean((0-sample.median)^2)
[1] 16.65672
> mean(abs(0-sample.mean))
[1] 2.699447
> mean(abs(0-sample.median))
[1] 3.2542
```

- So, in this simulation from the Normal$(0, 100)$ distribution under squared error loss, the risk of the sample mean is estimated to be $\hat{R}(\mu, \bar{x}) = 11.37$ which is smaller than the estimated risk of the sample median $\hat{R}(\mu, x_{(5)}) = 16.66$.

Simulation: Comparison of the Sample Mean and Median

- Instead, suppose $X_1, \ldots, X_9$ are iid Cauchy$(0,1)$ random variables and we want to estimate the population median.
- We perform the simulation study replacing the sample from the standard normal distribution with one from a Cauchy distribution.

```
> set.seed(626913)
> R=1e4
> sample.mean=rep(0,R)
> sample.median=rep(0,R)
> for (i in 1:R){
+ x=rt(9,df=1)
+ sample.mean[i]=mean(x)
+ sample.median[i]=median(x)
+ }
```

- ```
  > mean((0-sample.mean)^2)
  [1] 228501.2
  > mean((0-sample.median)^2)
  [1] 0.4044862
  > mean(abs(0-sample.mean))
  [1] 10.4965
  > mean(abs(0-sample.median))
  [1] 0.4696931
  ```
- So, in this simulation from the Cauchy$(0, 1)$ distribution under squared error loss, the risk of the sample mean is estimated to be $\hat{R}(\mu, \bar{x}) = 228501$ which is smaller than the estimated risk of the sample median $\hat{R}(\mu, x_{(5)}) = 0.4045$.
  ```