Chapter 7: Point Estimation

MATH 450

September 21st, 2017

Where are we?

Week 1	Chapter 1: Descriptive statistics
Week 2	Chapter 6: Statistics and Sampling Distributions
Week 4	**Chapter 7: Point Estimation**
Week 7	Chapter 8: Confidence Intervals
Week 10	Chapter 9: Test of Hypothesis
Week 13	Two-sample inference, ANOVA, regression

Overview

7.1 Point estimate

- unbiased estimator
- mean squared error
- bootstrap

7.2 Methods of point estimation

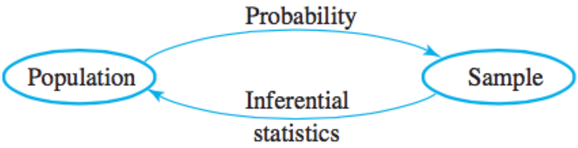
- method of moments
- method of maximum likelihood.

7.3 Sufficient statistic

7.4 Information and Efficiency

- Large sample properties of the maximum likelihood estimator

Point estimate



Definition

A point estimate $\hat{\theta}$ of a parameter θ is a single number that can be regarded as a sensible value for θ.

$$\begin{array}{ccccc} \text{population parameter} & \Longrightarrow & \textit{sample} & \Longrightarrow & \textit{estimate} \\ \theta & \Longrightarrow & X_1, X_2, \ldots, X_n & \Longrightarrow & \hat{\theta} \end{array}$$

Mean Squared Error

- Measuring error of estimation

$$|\hat{\theta} - \theta| \qquad \text{or} \qquad (\hat{\theta} - \theta)^2$$

- The error of estimation is random

Definition

The mean squared error of an estimator $\hat{\theta}$ is

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

Theorem

$$MSE(\hat{\theta}) = E[(\hat{\theta} - \theta)^2] = V(\hat{\theta}) + \left(E(\hat{\theta}) - \theta\right)^2$$

Bias-variance decomposition

$$\text{Mean squared error} = \text{variance of estimator} + (bias)^2$$

Unbiased estimators

Definition

A point estimator $\hat{\theta}$ is said to be an unbiased estimator of θ if

$$E(\hat{\theta}) = \theta$$

for every possible value of θ.

Unbiased estimator

$\Leftrightarrow$ Bias $= 0$

$\Leftrightarrow$ Mean squared error $=$ variance of estimator

Proposition

If $X_1, X_2, \ldots, X_n$ is a random sample from a distribution with mean μ, then $\bar{X}$ is an unbiased estimator of μ.

Proof: $E(\bar{X}) = \mu$.

- Let
$$T = a_1 X_1 + a_2 X_2 + \ldots + a_n X_n,$$
then the mean and of T can be computed by
$$E(T) = a_1 E(X_1) + a_2 E(X_2) + \ldots + a_n E(X_n)$$

Sample variance as an unbiased estimator

Theorem

The sample variance

$$S^2 = \frac{1}{n-1}\left[\left(\sum X_i^2\right) - \frac{1}{n}\left(\sum X_i\right)^2\right]$$

is an unbiased estimator of the population variance σ^2.

Ideas:

- $V(X) = E[X^2] - (EX)^2$
- $Var[\bar{X}] = \frac{\sigma^2}{n}$, $E[\bar{X}] = \mu$

Sample proportion

- A test is done with probability of success p
- n independent tests are done, denote by Y the number of successes
- Denote by X_i the result of test i^{th}, where $X_i = 1$ when the test success and $X_i = 0$ if not, then
  - Each X_i is distributed by

x	0	1
p(x)	1-p	p

  - $E[X] =?$
  - Moreover,

$$Y = \sum_{i=1}^{n} X_i$$

  - $E[Y] =?$

Sample proportion

- A test is done with probability of success p
- n independent tests are done, denote by Y the number of successes
- Let
$$\hat{p} = \frac{Y}{n}$$
the $E[\hat{p}] = p$, i.e., $\hat{p}$ is an unbiased estimator

Example 1

- A test is done with probability of success p
- n independent tests are done, denote by Y the number of successes
- Let
$$\hat{p} = \frac{Y}{n}$$
the $E[\hat{p}] = p$, i.e., $\hat{p}$ is an unbiased estimator
- Crazy idea: How about using
$$\tilde{p} = \frac{Y+2}{n+4}$$
- What is the bias of $\tilde{p}$?

Example 2

Problem

Suppose a certain type of fertilizer has an expected yield per acre of μ_1 with variance σ^2, whereas the expected yield for a second type of fertilizer is μ_2 with the same variance σ^2. Let S_1^2 and S_2^2 denote the sample variances of yields based on sample sizes n_1 and n_2, respectively, of the two fertilizers.
Show that the pooled (combined) estimator

$$\hat{\sigma}^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

is an unbiased estimator of σ^2.

Example 3

Problem

Consider a random sample $X_1, \ldots, X_n$ from the pdf

$$f(x) = \frac{1 + \theta x}{2} \qquad -1 \leq x \leq 1$$

Show that $\hat{\theta} = 3\bar{X}$ is an unbiased estimator of θ.

- $V(X) = E[(X - EX)^2] = E[X^2] - (EX)^2$
- $V(cX) = c^2 V(X)$
- $V(X + c) = V(X)$
- If $X_1, X_2, \ldots, X_n$ are independent, the

$$V(X_1 + X_2 + \ldots + X_n) = V(X_1) + V(X_2 + \ldots + V(X_n)$$

Warm-up: Sample proportion

- A test is done with probability of success p
- n independent tests are done, denote by Y the number of successes
- Denote by X_i the result of test i^{th}, where $X_i = 1$ when the test success and $X_i = 0$ if not, then
  - Each X_i is distributed by

x	0	1
p(x)	1-p	p

  - $V[X] = ?$
  - Moreover,

$$Y = \sum_{i=1}^{n} X_i$$

  - $V[Y] = ?$

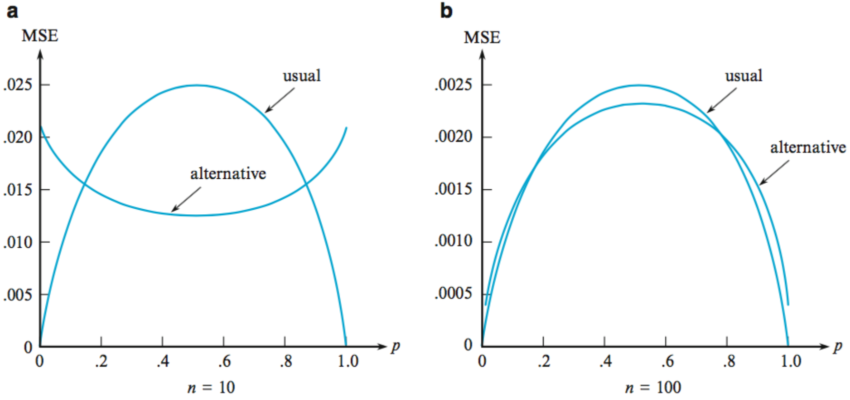
Example 4

- A test is done with probability of success p
- n independent tests are done, denote by Y the number of successes
- Crazy idea: How about using

$$\tilde{p} = \frac{Y + \sqrt{n/4}}{n + \sqrt{n}}$$

- What is the bias of $\tilde{p}$?
- Compute $V(\tilde{p})$.
- Compute MSE$(\tilde{p})$?

Example 7.1 and 7.4



Minimum variance unbiased estimator (MVUE)

Definition

Among all estimators of θ that are unbiased, choose the one that has minimum variance. The resulting $\hat{\theta}$ is called the minimum variance unbiased estimator (MVUE) of θ.

Recall:

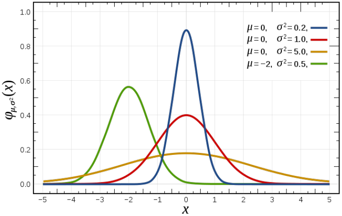
- Mean squared error $=$ variance of estimator $+$ $(bias)^2$
- unbiased estimator $\Rightarrow$ bias $=0$

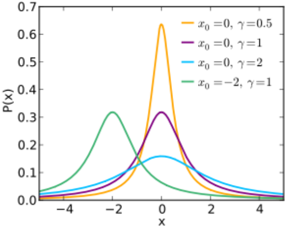
$\Rightarrow$ MVUE has minimum mean squared error among unbiased estimators

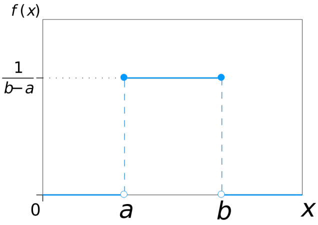
MVUE of normal distributions

Theorem

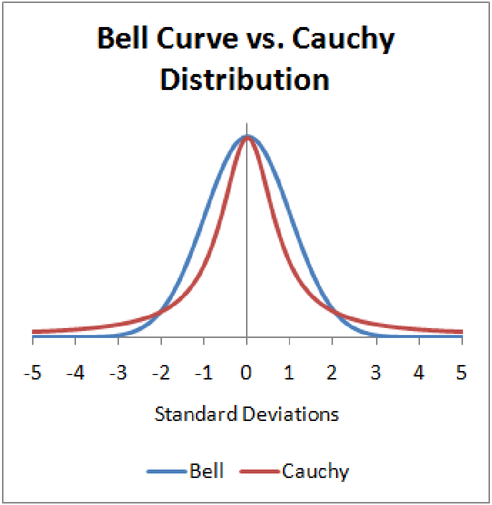
Let $X_1, \ldots, X_n$ be a random sample from a normal distribution with parameters μ and σ. Then the estimator $\hat{\mu} = \bar{X}$ is the MVUE for μ.


$\mu=0, \sigma^2=0.2$,
$\mu=0, \sigma^2=1.0$,
$\mu=0, \sigma^2=5.0$,
$\mu=-2, \sigma^2=0.5$,
$\varphi_{\mu,\sigma^2}(x)$
x


$x_0=0, \gamma=0.5$
$x_0=0, \gamma=1$
$x_0=0, \gamma=2$
$x_0=-2, \gamma=1$
P(x)
x


$f(x)$
$\frac{1}{b-a}$
0
a
b
x

Normal vs. Cauchy



What is the best estimator of the mean?

Question: Let $X_1, \ldots, X_n$ be a random sample from a normal distribution with parameters μ and σ. What is the best estimator of the mean μ?

Answer: It depends.

- Normal distribution $\rightarrow$ reasonable tails $\rightarrow$ sample mean $\hat{X}$
- Cauchy distribution $\rightarrow$ heavy tails, symmetric $\rightarrow$ sample median $\tilde{X}$
- Uniform distribution $\rightarrow$ no tails, uniform

$$\hat{X}_e = \frac{\text{largest number} + \text{smaller number}}{2}$$

- In all cases, 10% trimmed mean performs pretty well

Reporting a point estimate: the standard error

Definition

$$\text{standard error} = \sigma_{\hat{\theta}} = \sqrt{V(\hat{\theta})}$$

If the standard error itself involves unknown parameters whose values can be estimated, substitution of these estimates into $\sigma_{\hat{\theta}}$ yields the estimated standard error of the estimator, denoted by $s_{\hat{\theta}}$.

How to compute standard error?

$$\begin{array}{ccccc} \text{population parameter} & \Longrightarrow & \textit{sample} & \Longrightarrow & \textit{estimate} \\ \theta & \Longrightarrow & X_1, X_2, \ldots, X_n & \Longrightarrow & \hat{\theta} \end{array}$$

- We now that
$$\sigma_{\bar{X}} = \frac{\sigma_X}{\sqrt{n}}$$
- ...but computing that is quite difficult
- What if the formula of $\hat{\theta}$ is very complicated?

Parametric model

- Suppose that the population pdf is $f(x;\theta)$ (which means that $X_1, X_2, \ldots, X_n$ are sampled from a distribution with pdf $f(x;\theta)$)
- data $x_1, x_2, \ldots, x_n$ are collected $\rightarrow$ point estimate $\hat{\theta}$
- if we have time/money, we can do the experiment again, collect new set of data, and get $\hat{\theta}_1$
- do the experiment again, get $\hat{\theta}_2$

  $\ldots$
- do the experiment again for the B^{th} time, get $\hat{\theta}_B$

$$\sigma_{\hat{\theta}} = \sqrt{\frac{1}{B-1}\sum (\hat{\theta}_i - \bar{\theta})^2}, \qquad \bar{\theta} = \frac{\hat{\theta}_1 + \hat{\theta}_2 + \ldots + \hat{\theta}_B}{B}$$

Bootstrap

boot·strap

/ˈbo͞otˌstrap/

noun

1. a loop at the back of a boot, used to pull it on.
2. COMPUTING
   a technique of loading a program into a computer by means of a few initial instructions that enable the introduction of the rest of the program from an input device.

verb

1. get (oneself or something) into or out of a situation using existing resources.
   "the company is bootstrapping itself out of a marred financial past"
2. COMPUTING
   fuller form of boot[1] (sense 3 of the verb).

adjective

1. (of a person or project) using one's own resources rather than external help.
   "a bootstrap capitalist's trip up the entrepreneurial ladder"

Parametric bootstrap

- Suppose that the population pdf is $f(x;\theta)$ (which means that $X_1, X_2, \ldots, X_n$ are sampled from a distribution with pdf $f(x;\theta)$)
- data $x_1, x_2, \ldots, x_n$ are collected $\rightarrow$ point estimate $\hat{\theta}$

Bootstrapping:

- plug $\hat{\theta}$ into the formula of $f(x,\theta) \rightarrow$ density function $f(x,\hat{\theta})$
- *simulate* new sample $x_1, x_2, \ldots, x_n$ from $f(x,\hat{\theta})$

Parametric bootstrap

- plug $\hat{\theta}$ into the formula of $f(x, \theta)$
- *simulate* new sample $x_1^*, x_2^*, \ldots, x_n^*$ from $f(x, \hat{\theta})$
  - First bootstrap sample: $x_1^*, x_2^*, \ldots, x_n^* \to$ get $\hat{\theta}_1$
  - Second bootstrap sample $\to \hat{\theta}_2$

    $\ldots$
  - B^{th} bootstrap sample $\to \hat{\theta}_B$

Bootstrapping estimate:

$$\sigma_{\hat{\theta}} = \sqrt{\frac{1}{B-1} \sum (\hat{\theta}_i - \bar{\theta})^2}, \qquad \bar{\theta} = \frac{\hat{\theta}_1 + \hat{\theta}_2 + \ldots + \hat{\theta}_B}{B}$$