Chapter 6: Point Estimation

Motivation: Given a r.v. X with known distribution type but unknown parameters.

Sample X many times to get IID $X_1, X_2, \ldots, X_n$

Want to use observed values $\left\{\begin{array}{l} X_1 = x_1 \\ X_2 = x_2 \end{array} \text{ etc}\right\}$ to estimate the parameters.

"i.e. guess"

Main example. What are $\left\{\begin{array}{l} \mu = E[X] \\ \sigma^2 = Var[X] \end{array}\right.$???

Guess: $\bar{x} = \frac{1}{n}(x_1 + \cdots + x_n) \approx \mu$

$s^2 = \frac{1}{n-1}\left[(x_1-\bar{x})^2 + \cdots + (x_n-\bar{x})^2\right] \approx \sigma^2$

To distinguish $\bar{x}$ and s^2 from μ and σ^2 we will say

$\bar{x}$ = "sample mean"
s^2 = "sample variance"

μ = "population mean"
σ^2 = "population variance"

§6.1 General Concepts

Notation: X is a random variable (the "population")

$X_1, X_2, \ldots, X_n$ are IID r.v. (the "sample")

(In Ch 5 we considered functions of random variables $Y = h(X_1, X_2, \ldots, X_n)$
We called this type of random variable a "statistic")

Def: A statistic which is intended to estimate a parameter θ for a distribution is called a "point estimator" for θ

We will usually label point estimators with ^

Example: Point estimator for θ is $\hat{\theta}$

Point estimator for μ is $\hat{\mu}$

etc.

Def: The numeric value of a point estimator at a given set of observations $X_1 = x_1$, $X_2 = x_2$, etc is called the "point estimate"

Technically, we should probably try to use capital letters for point estimators (Random Variable) and lower case for point estimates (Number)

Parameter	(Random Variable) Point Estimator	(Observed Value) Point Estimate	
θ	$\hat{\Theta}$	$\hat{\theta}$	$\approx \theta$
$\mu = E[X]$	$\hat{M} = \bar{X}$	$\hat{\mu} = \bar{x}$	$\approx \mu$

That's the idea... but we won't always follow it because sometimes it looks too silly...
Also I am not good at drawing "capital θ" $\rightsquigarrow$ Θ Θ Θ Θ Θ Θ Θ
ugh...

Example: Suppose X is Binomial(n, p).

Then $\hat{P} = X/n$ is point estimator for p.

If $x = 25$ and $n = 40$ then

$\hat{p} = 25/40 = 5/8$ is point estimate of p

Example: Suppose X is Normal(μ, σ). Sample 20 times.

Then $\bar{X} = \frac{1}{20}(X_1 + \cdots + X_{20})$ is point estimator for μ.

If we observe values $x_1, x_2, \ldots, x_n$ so that $\bar{x} = 25$ then

$\bar{x} = 25$ is point estimate of μ

Note: Point estimators are random variables, so they each have their own distribution.

We will call the standard deviation of an estimator its "standard error"

$\boxed{\bar{X}}$

- $E[\bar{X}] = E[\frac{1}{n}(X_1 + \cdots + X_n)] = \frac{1}{n} \cdot n E[X]$
  $= \mu$
- $\text{Var}[\bar{X}] = \text{Var}[\frac{1}{n}(X_1 + \cdots + X_n)] = \frac{1}{n^2} \cdot n \text{Var}[X]$
  $= \sigma^2/n$

"standard error" $\sigma_{\bar{X}} = \sigma/\sqrt{n}$

$\boxed{\hat{P}}$

- $E[\hat{P}] = E[X/n] = 1/n\, E[X] = 1/n\,(np)$
  $= p$
- $\text{Var}[\hat{P}] = \text{Var}[X/n] = 1/n^2\, \text{Var}[X] = \frac{1}{n^2} \cdot npq$
  $= pq/n$

"standard error" $\sigma_{\hat{P}} = \sqrt{pq/n}$

Note: In this example, $X \sim$ Binomial(n, p)
Recall that we write $q = (1-p)$ in this case

Often there are many possible estimators for θ

Possible estimators for $\mu = E[X]$:

$\bar{X}$ = "sample mean" $\qquad \bar{X}_e$ = avg. of max & min

$\tilde{X}$ = "sample median" $\qquad \bar{X}_{tr}$ = trimmed mean

Remove extreme values first

Possible estimators for $\sigma^2 = \text{Var}[X]$:

$S^2 = \frac{1}{n-1} \sum (X_k - \bar{X})^2 \qquad \hat{\sigma}^2 = \frac{1}{n} \sum (X_k - \bar{X})^2$

How do we choose which point estimator to use?

Bias $E[\hat{\Theta}]$

A good estimator should have the correct expected value.

Def: A point estimator is **unbiased** if

$$E[\hat{\Theta}] = \theta$$

Otherwise the value $E[\hat{\Theta}] - \theta$ is the "bias"

Example: S^2 or $\hat{\sigma}^2$ for estimating $\sigma^2 = Var[X]$.

Recall: Two formulas for $\hat{\sigma}^2$ — same as for $Var[X]$

$$Var[X] = E[(X-\bar{X})^2] = E[X^2] - (E[X])^2$$

$$\hat{\sigma}^2 = \frac{1}{n}\sum_1^n (X_k - \bar{X})^2 = \underbrace{\frac{1}{n}\sum_1^n X_k^2 - \left(\frac{1}{n}\sum_1^n X_k\right)^2}$$

Use second formula to compute $E[\hat{\sigma}^2]$ along with one extra trick: $Var[Y] = E[Y^2] - (E[Y])^2$

rearranging $\hookrightarrow$ $E[Y^2] = Var[Y] + (E[Y])^2$ (*)

$$E[\hat{\sigma}^2] = E\left[\frac{1}{n}\sum_1^n X_k^2\right] - E\left[\left(\frac{1}{n}\sum_1^n X_k\right)^2\right]$$

$$= \frac{1}{n}\cdot n\, E[X^2] - \frac{1}{n^2} E\left[\left(\sum_1^n X_k\right)^2\right] \quad \text{use (*) twice}$$

$$= (\sigma^2 + \mu^2) - \frac{1}{n^2}\left(n\sigma^2 + (n\mu)^2\right) = \sigma^2(1 - 1/n)$$

So $\hat{\sigma}^2$ is **biased**!!!

On the other hand $S^2 = \frac{n}{n-1}\hat{\sigma}^2$ has

$$E[S^2] = E\left[\frac{n}{n-1}\hat{\sigma}^2\right]$$

$$= \frac{n}{n-1} E[\hat{\sigma}^2]$$

$$= \frac{n}{n-1}\sigma^2(1 - 1/n) = \sigma^2$$

Unbiased!

Example: Suppose $X_1, \dots, X_n$ are samples of $X \sim Uniform[0, B]$

Try: $\max(X_k)$ as estimator of B.

Note: This is definitely biased because

sometimes $\max(X_k) < B$

but never $\max(X_k) > B$

pdf of $\max(X_k) = \frac{d}{dx}$ cdf of $\max(X_k)$

cdf is $P(X_1, \dots, X_n \le x) = \frac{x^n}{B^n}$

pdf is $\frac{d}{dx}\frac{x^n}{B^n} = \frac{n}{B^n}x^{n-1}$

$$E[\max(X_k)] = \int_0^B x \cdot \frac{n}{B^n} x^{n-1}\, dx = \boxed{\frac{n}{n+1} B}$$

Biased!!

Remark: To remove bias of $\max(X_k)$ as estimator of B for $X \sim \text{Uniform}[0, B]$, multiply by $\frac{n+1}{n}$:

$$E\left[\frac{n+1}{n}\max(X_k)\right] = \frac{n+1}{n} E[\max(X_k)]$$

$$= \frac{n+1}{n} \cdot \frac{n}{n+1} B = B$$

Some other unbiased estimators:

- $\bar{X}$ If X is any distribution with $\mu = E[X]$

  then

$$E[\bar{X}] = E\left[\tfrac{1}{n}(X_1 + \cdots + X_n)\right]$$

$$= \tfrac{1}{n}\left[E(X_1) + \cdots + E(X_n)\right]$$

$$= \tfrac{1}{n} \cdot n\mu = \mu$$

- $\hat{p}$ If $X \sim \text{Binomial}(n, p)$, let $\hat{p} = X/n$

  then

$$E[\hat{p}] = E[X/n]$$

$$= 1/n \, E[X]$$

$$= 1/n \cdot np = p$$

- $\hat{X}$ sample median
- $\bar{X}_{tr}$ trimmed mean
- $\bar{X}_e$ avg. of max & min

Also Unbiased...

How do we decide which unbiased estimator to use?

(i.e. How to pick between $\bar{X}, \hat{X}, \bar{X}_e, \bar{X}_{tr}$)

Variance $\text{Var}[\hat{\Theta}]$

A good estimator should have small variance.

> **Thm:** If $X \sim$ Normal then $\bar{X}$ has smallest variance among all possible unbiased estimators for μ.

Note: If X is not Normal then $\bar{X}$ may not be the best. This is a difficult computation.... Usually people will just use $\bar{X}_{tr}$ in cases where X is not Normal

Even if $\bar{X}_{tr}$ is not "the best", it is almost always "pretty good".

Example: Variance of Sample Mean.

We've done this computation before...

$$Var[\bar{X}] = Var[\tfrac{1}{n}(X_1 + \cdots + X_n)]$$

$$= \tfrac{1}{n^2} Var[X_1 + \cdots + X_n]$$

$$= \tfrac{1}{n^2} \cdot n\, Var[X] = \sigma^2/n$$

Example: Variance of Sample Proportion

Let $X \sim \text{Binomial}(n, p)$ & $\hat{p} = X/n$

$$Var[\hat{p}] = Var[X/n]$$

$$= \frac{1}{n^2} Var[X]$$

$$= \frac{1}{n^2} \cdot npq = pq/n$$

In general computing variance of an estimator is very difficult and usually relies on the precise distribution of X.

↳ See extra page for example.

Note: We can estimate the variance of an estimator using a (computationally intensive) method called "Boot-Strapping"

Boot-Strapping

Given a point estimate $\hat{\theta}_0$ for parameter θ

→ Use a computer to generate the same amount of random data with parameter $\hat{\theta}_0$

Plug data into point estimator to get new point estimate $\hat{\theta}_1$ for parameter θ

→ Use a computer to generate the same amount of random data with parameter $\hat{\theta}_1$

Repeat this "many" times (at least 100)

to get $\hat{\theta}_0, \hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_b$

Use sample variance of these results:

$$\hat{\sigma}_{\hat{\theta}}^2 = \frac{1}{b} \sum_{k} (\hat{\theta}_k - \bar{\theta})^2$$

← there are b+1 elements in sum, so sample var uses $1/b$

Extra Page

Example: If $X \sim \text{Uniform}[0, B]$ then

$$\frac{n+1}{n}\max(X_k) \quad \& \quad 2\cdot\bar{X}$$

are both unbiased! How to choose?

$$\text{Var}\left[\frac{n+1}{n}\max(X_k)\right] = \left(\frac{n+1}{n}\right)^2\left[E[\max(X_k)^2] - (E[\max(X_k)])^2\right]$$

Note to self: Let's skip these computations during lecture...

For $E[\max(X_k)^2]$ we will compute pdf.

$$P(X < x) = x/B \Longrightarrow P(X^2 < x) = P(X < \sqrt{x})$$
$$= \sqrt{x}/B$$

cdf:

$$P(X_1^2, X_2^2, \ldots, X_n^2 < x) = \left(\sqrt{x}/B\right)^n$$
$$= x^{n/2}/B^n$$

pdf:

$$\frac{d}{dx}\frac{x^{n/2}}{B^n} = \frac{n}{2B^n}\cdot x^{\frac{n-2}{2}}$$

$$E[\max(X_k)^2] = \int_0^{B^2} x\cdot\frac{n}{2B^n}\cdot x^{\frac{n-2}{2}}\,dx$$

$$= \frac{n}{2B^n}\cdot\frac{2}{n+2}\,x^{\frac{n+2}{2}}\Big|_0^{B^2} = \frac{n}{n+2}B^2$$

For $(E[\max(X_k)])^2$ recall that

$$E[\max(X_k)] = \frac{n}{n+1}B$$

$$\text{Var}[\max(X_k)] = \frac{n}{n+2}B^2 - \frac{n^2}{(n+1)^2}B^2 = \frac{n}{(n+2)(n+1)^2}B^2$$

So

$$\text{Var}\left[\frac{n+1}{n}\max(X_k)\right] = \frac{(n+1)^2}{n^2}\cdot\frac{n}{(n+2)(n+1)^2}B^2$$
$$= \frac{1}{n(n+2)}B^2$$

What about $\text{Var}[2\cdot\bar{X}]$?

$$\text{Var}[X] = E[X^2] - (E[X])^2$$

$$E[X] = \int_0^B x\cdot\frac{1}{B}\,dx = B/2$$
$$E[X^2] = \int_0^B x^2\cdot 1/B\,dx = B^2/3$$

$$\Rightarrow = B^2/3 - (B/2)^2 = B^2/12$$

So $\text{Var}[2\bar{X}] = 4\,\text{Var}[\bar{X}]$

$$= 4/n^2\,\text{Var}[X] = B^2/3n^2$$

$$\boxed{\text{Var}\left[\frac{n+1}{n}\max(X_k)\right] = \frac{1}{n(n+2)}B^2}$$

$$\text{Var}[2\cdot\bar{X}] = \frac{1}{3n^2}B^2$$

Smaller!!

(as long as $n > 1$)