NSE ice cream social
today 3:30-4:30
Singer

02: PROPERTIES OF ESTIMATORS I

Larsen & Marx 5.4

Prof Amanda Luby

We've seen two methods of estimating parameters: the MLE and the MoM. Both give very reasonable criteria to identify estimators for unknown parameters, but they do not always yield the same answer.

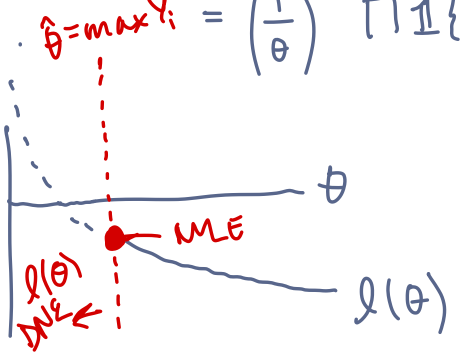
For example, on your homework, you showed that the ~~MLE~~ MoM estimator for θ in a (continuous) Unif$(0, \theta)$ distribution is $\hat{\theta} = 2\bar{Y}$.

Example: Find the MLE for $Y_1, Y_2, ..., Y_n \sim$ Unif$(0, \theta)$. (Recall $f_y(y) = \begin{cases} \frac{1}{\theta}, 0 \leq y \leq \theta \\ 0 \quad y > \theta, y < 0 \end{cases}$)

* similar to German Tank problem

$$L(\theta) = \prod_{i=1}^{n} \frac{1}{\theta} \mathbb{1}\{y_i \leq \theta\}$$

$$= \left(\frac{1}{\theta}\right)^n \prod \mathbb{1}\{y_i \leq \theta\} = \begin{cases} \left(\frac{1}{\theta}\right)^n & \underline{\text{all}}\ Y_i \leq \theta \\ 0 & \text{any } Y_i > \theta \end{cases}$$



$$\ell(\theta) = \begin{cases} -n \ln(\theta) & \text{all } Y_i \leq \theta \\ \text{DNE} & \text{any } Y_i > \theta \end{cases}$$

$\max Y_i > \theta$

$\rightarrow \theta < \max Y_i$

Implicit in the two estimators for the same parameters is the obvious question: which one should we use?

$\hat{\theta}_1 = 2\bar{Y}$

$\hat{\theta}_2 = Y_{max}$

There are actually an infinite number of estimators for any given parameter, and this requires that we have a principled way of evaluating the statistical properties associated with any given estimator. What qualities should a "good" estimator have? Is it possible to find a "best" $\hat{\theta}$? This set of notes, and the second unit of the course, is going to begin to address these questions.

Note: Every estimator is a function of a set of random variables (ie $\hat{\theta} = g(Y_1, Y_2, ..., Y_n)$) and is itself a random variable.

$\Longrightarrow \hat{\theta}$ has a pdf, expected value, variance

θ is our true parameter $\underline{\text{value}}$

Notation for $\hat{\theta}$

u is any value of θ

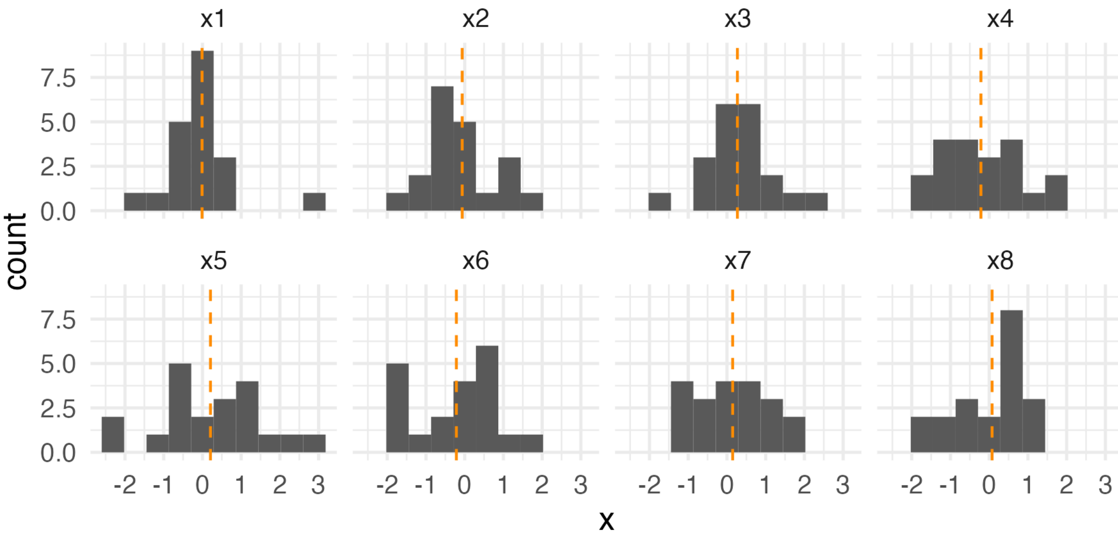
pdf: $\begin{cases} f_{\hat{\theta}}(u) & \text{continuous} \\ p_{\hat{\theta}}(u) & \text{discrete} \end{cases}$

$E(\hat{\theta}) = \text{mean} = \mu_{\hat{\theta}}$

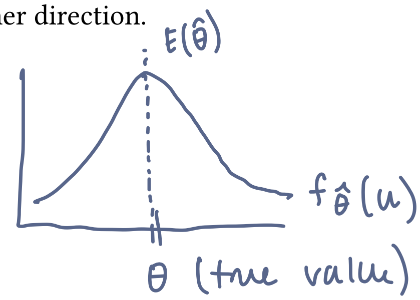
$V(\hat{\theta}) = \text{variance} = \sigma^2_{\hat{\theta}}$

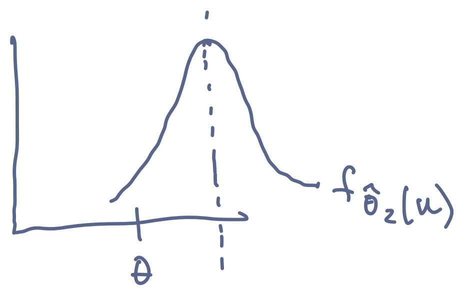
1 Unbiasedness / Biasedness / Bias

Lab01 ended with the idea of *random samples*, and noting that every sample is going to give a slightly different estimate for θ. Here are eight random samples of size 20 of $N(0, 1)$ random variables, with the sample mean overlaid on each facet.



Ideally, we want the overestimates to "balance out" the underestimates: $\hat{\theta}$ should not systematically err in either direction.





"balances out"
$\Rightarrow$ unbiased
$E(\hat{\theta}_1) = \theta$

consistently over estimating θ
$\Rightarrow$ biased
$E(\hat{\theta}_2) > \theta$

When the mean of the estimator $\hat{\theta}$ is equal to the true parameter θ, we say not that the estimator is **unbiased**.

> **Definition**
>
> Let $W_1, W_2, \ldots, W_n$ be a random sample from $f_w(w, \theta)$. An estimator $\hat{\theta} = g(W_1, W_2, \ldots, W_n)$ is said to be **unbiased for** θ if $E(\hat{\theta}) = \theta$ for all θ

$E(X) = \frac{\theta}{2}$

Example: is the MoM for the Unif$(0,\theta)$ distribution, $\hat{\theta}_1 = 2\bar{X}$ unbiased? AKA is $E(\hat{\theta}_1) = \theta$?

$$E(\hat{\theta}_1) = E(2\bar{X}) = E\left(\frac{2}{n}\sum X_i\right)$$

properties of expectation

$$= \frac{2}{n}\sum E(X_i)$$

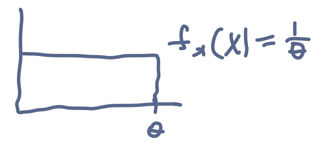
$E(X_i) = \frac{\theta}{2}$

$$= \frac{2}{n}\sum \frac{\theta}{2}$$

algebra

$$= \frac{2}{n}\cdot n\,\frac{\theta}{2}$$

$$= \theta \quad \checkmark \text{unbiased}$$



Example: is the MLE $\hat{\theta}_2 = \max(X_i)$ unbiased?

hint: $X \sim \text{unif}(0,\theta)$

$f_{x_{max}}(u) = n\cdot\left(\frac{1}{\theta}\right)\left(\frac{u}{\theta}\right)^{n-1} \quad 0 \le u \le \theta$

order statistic

$$E(\hat{\theta}_2) = \int_0^\theta u\cdot\frac{n}{\theta}\left(\frac{u}{\theta}\right)^{n-1} du$$

$$= \int_0^\theta \frac{n}{\theta^n} u^n \, du$$

$$= \left[\frac{n}{\theta^n(n+1)} u^{n+1}\right]_0^\theta$$

$$= \frac{n}{(n+1)\theta^n}\theta^{n+1} - 0$$

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

$\neq \theta \Rightarrow$ biased estimator

unpack this:

$n=3 \quad \frac{3}{4}\theta \rightarrow$ bad

$n=100 \quad \frac{100}{101}\theta \rightarrow$ meh?

$n=100{,}000 \quad \frac{100{,}000}{100{,}001}\theta \rightarrow$ does it matter?

asymptotically unbiased

Example: Construct an estimator, $\hat{\theta}_3$ based on $\max(X_i)$ that is unbiased.

$$\hat{\theta}_3 = \frac{n+1}{n} X_{max}$$

$$E(\hat{\theta}_3) = E\left(\frac{n+1}{n} X_{max}\right)$$

$$= \frac{n+1}{n} E(X_{max})$$

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

$$= \theta \rightarrow \text{unbiased estimator}$$

* biasedness is usually 'fixable'

HW 2 is due on Wed

Mon : 11:30 – 12:20

Wed : 2:30 – 4

bit.ly/prof-luby-oh (incl. appts)

Quiz 1 next <u>Wed</u> - Formula sheet coming

Recap:

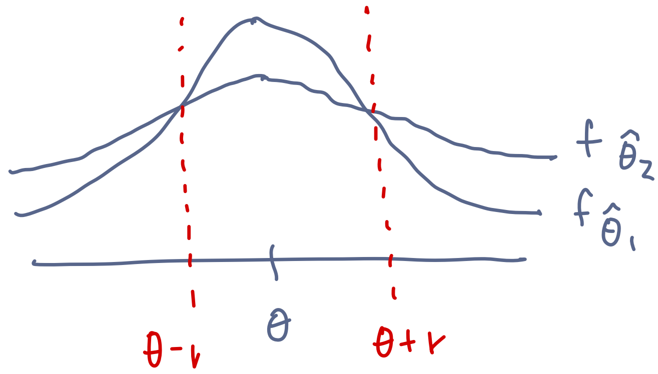
- Estimation : finding & evaluating $\hat{\theta}$ for θ
- MLE, MOM
- Unbiasedness : $E(\hat{\theta}) = \theta$ } evaluating estimators
- Efficiency / $V(\hat{\theta})$

2 Efficiency

$2\bar{X}$ $\frac{n+1}{n} X_{max}$

We now have two estimators, $\hat{\theta}_1$ and $\hat{\theta}_3$, that are unbiased estimators for θ. Does it matter which one we choose?

Idea:



$P(\theta - r < \hat{\theta}_i < \theta + r)$

$\hat{\theta}_1$ is more likely to be in the interval than $\hat{\theta}_2$ $\Rightarrow$ more likely to be "close" to the true value

$\hat{\theta}_1$ has smaller variance than $\hat{\theta}_2$

> **Definition**
>
> Let $\hat{\theta}_1$ and $\hat{\theta}_2$ be two unbiased estimators for a parameter θ. If $\mathrm{Var}(\hat{\theta}_1) < \mathrm{Var}(\hat{\theta}_2)$, we say that $\hat{\theta}_1$ is more efficient than $\hat{\theta}_2$.
>
> Relative Efficiency of $\hat{\theta}_1$ WRT $\hat{\theta}_2$ is the ratio $\frac{V(\hat{\theta}_2)}{V(\hat{\theta}_1)}$

Sometimes in reverse - not picky about it, as long as the interpretation is correct

Example: Let $Y_1, Y_2, Y_3 \sim N(\mu, \sigma^2)$. Which is a more efficient estimator for mu:

$$\hat{\mu_1} = \frac{1}{4}Y_1 + \frac{1}{2}Y_2 + \frac{1}{4}Y_3$$

or

Note: both are unbiased

$$\hat{\mu_2} = \frac{1}{3}Y_1 + \frac{1}{3}Y_2 + \frac{1}{3}Y_3? = \bar{Y}$$

$\mathrm{var}(\hat{\mu}_1) = V\left(\frac{1}{4}Y_1 + \frac{1}{2}Y_2 + \frac{1}{4}Y_3\right)$
$= \frac{1}{16}V(Y_1) + \frac{1}{4}V(Y_2) + \frac{1}{16}V(Y_3)$
$= \frac{6}{16}\sigma^2 = \frac{3}{8}\sigma^2$

(A) $V(\hat{\mu}_2) = V\left(\frac{1}{3}Y_1 + \frac{1}{3}Y_2 + \frac{1}{3}Y_3\right)$
$= \left(\frac{1}{9} + \frac{1}{9} + \frac{1}{9}\right)\sigma^2$
$= \frac{3}{9}\sigma^2$

(B) CLT
$\bar{Y} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$
$V(\bar{Y}) = \frac{1}{3}\sigma^2$

$3/9\,\sigma^2 < 3/8\,\sigma^2 \Rightarrow \hat{\mu}_2$ is more efficient

relative efficiency of $\hat{\mu}_2$ to $\hat{\mu}_1$: $\frac{3/8\,\sigma^2}{3/9\,\sigma^2} = 9/8$

$\hat{\mu}_2$ is 1.125 times more efficient than $\hat{\mu}_1$

$\hat{\mu}_1$ is 88.9% as efficient as $\hat{\mu}_2$

$X_1, \ldots, X_n \sim \text{unif}(0, \theta) \quad f_x = 1/\theta \quad 0 \le x \le \theta$

$\boxed{\hat{\theta}_1 = 2\bar{X} \quad \text{MOM}}$

$\hat{\theta}_2 = X_{max}$ MLE, but this was biased $\Rightarrow$ $\boxed{\hat{\theta}_3 = \frac{n+1}{n} X_{max}}$ was unbiased

Exercise: Which of our two unbiased estimators for the uniform distribution is more efficient?

Helpful Hints:

$E(X) = \frac{\theta}{2}$

$V(X) = E(X^2) - E(X)^2$

$f_{X_{max}} = \frac{n}{\theta}\left(\frac{y}{\theta}\right)^{n-1}$

$E(X_{max}) = \frac{n}{n+1}\theta$

ANSWER:

$V(\hat{\theta}_1) = \frac{\theta^2}{3n}$

$V(\hat{\theta}_3) = \frac{\theta^2}{n(n+2)}$

$$\frac{\theta^2/3n}{\theta^2/n(n+2)} = \frac{n+2}{3}$$

$\rightarrow$ If $n \ge 2$, $\hat{\theta}_3$ is more efficient

1. $V(\hat{\theta}_1) = V(2\bar{X}) = V\left(\frac{2}{n}\sum X_i\right) = \frac{4}{n^2}\sum V(X_i)$

$\rightarrow$ need to find $V(X)$

$V(X) = E(X^2) - E(X)^2 = E(X^2) - \left(\frac{\theta}{2}\right)^2$

$\rightarrow$ need to find $E(X^2)$

$\hookrightarrow E(X^2) = \int_{-\infty}^{\infty} x^2 f_x(x)\,dx = \int_0^{\theta} x^2 \cdot \frac{1}{\theta}\,dx = \theta^2/3$

So $V(X) = \theta^2/3 - \theta^2/4 = \theta^2/12$

Then, $V(\hat{\theta}_1) = \frac{4}{n^2}\sum\left(\frac{\theta^2}{12}\right) = \frac{4n}{n^2}\frac{\theta^2}{12} = \frac{4\theta^2}{12n} = \boxed{\frac{\theta^2}{3n}}$

2. $V(\hat{\theta}_3) = V\left(\frac{n+1}{n}X_{max}\right) = \left(\frac{n+1}{n}\right)^2 V(X_{max})$

$\rightarrow$ need to find $V(X_{max})$

$\hookrightarrow$ need to find $E(X_{max}^2)$

$$E(X_{max}^2) = \int_0^{\theta} x^2 \cdot \frac{n}{\theta}\left(\frac{x}{\theta}\right)^{n-1} dx$$

$$= \frac{n}{\theta^2}\int_0^{\theta} x^{n+1}\,dx = \frac{n}{\theta^2}\left[\frac{1}{n+2}x^{n+2}\right]_0^{\theta}$$

$$= \frac{n}{\theta^2} \cdot \frac{\theta^{n+2}}{n+2} = \frac{n}{n+2}\theta^2$$

$$V(\hat{\theta}_3) = V\left(\frac{n+1}{n}X_{max}\right) = \left(\frac{n+1}{n}\right)^2 V(X_{max})$$

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

$$= \left(\frac{n+1}{n}\right)^2\left[\frac{n}{n+2}\theta^2 - \left(\frac{n}{n+1}\right)^2\theta^2\right]$$

$$= \left(\frac{n+1}{n}\right)^2\left[\frac{n(n+1)^2 - n^2(n+2)}{(n+2)(n+1)^2}\right]\theta^2$$

$$= \left(\frac{n+1}{n}\right)^2\left[\frac{n^3 + 2n^2 + n - n^3 - 2n^2}{(n+2)(n+1)^2}\right]\theta^2$$

$$= \left(\frac{n+1}{n}\right)^2\left[\frac{n}{(n+2)(n+1)^2}\right]\theta^2$$

$$\boxed{= \frac{\theta^2}{n(n+2)}}$$

3 The Bias-Variance Tradeoff

Mean Square Error (MSE)

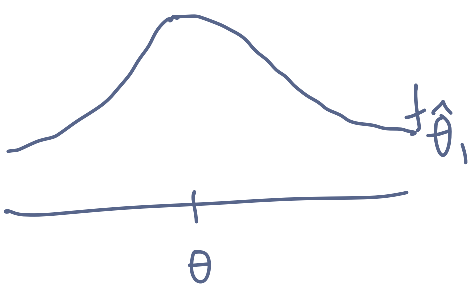
$$MSE(\hat{\theta}) = E(\hat{\theta} - \theta)^2 = V(\hat{\theta}) + \text{bias}^2$$

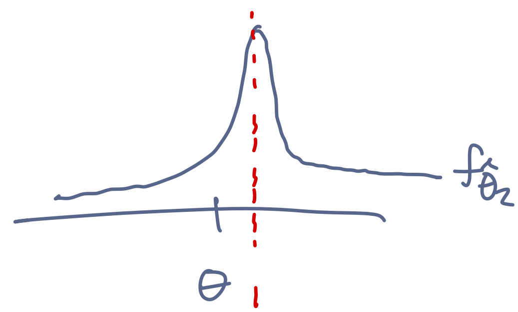
Note: We'll come back to this, and show that it is true, later in the course

Idea:

$MSE = E[(\hat{\theta} - \theta)^2]$ is the "total error" of our estimator. Has some minum value that can't be improved upon.

$MSE = V(\hat{\theta}) + \text{bias}^2$ bias: $E(\hat{\theta}) - \theta$





$\Longrightarrow$ Sometimes, MSE is smaller for a biased estimator than an unbiased one