Lecture 18

Topics:

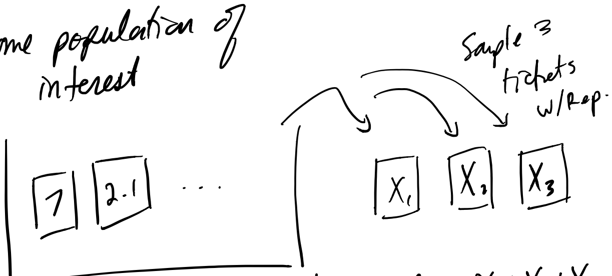
I) Estimation continued.

The setup: we want to estimate some population parameter θ from samples $X_1, \dots, X_n$ with replacement from the population.

The $MSE(\hat{\theta})$ can be used to choose between two estimators.

e.g.

some population of interest



You want to estimate the average of these ticket numbers, denoted μ

$$\hat{\mu}_1 = \frac{X_1+X_2+X_3}{3}$$

$$\hat{\mu}_2 = \frac{X_1+X_2+X_3}{3-1}$$

which is better?

$$bias(\hat{\mu}_1) = E(\hat{\mu}_1 - \mu)$$
$$= E(\bar{X} - \mu) \overset{MF1}{=} 0$$

$$bias(\hat{\mu}_2) = E(\hat{\mu}_2 - \mu)$$
$$= E\left(\frac{X_1+X_2+X_3}{3-1} - \mu\right)$$
$$\overset{MF1}{=} -\mu + \frac{E(X_1)+E(X_2)+E(X_3)}{3-1}$$
$$= -\mu + \frac{3}{2}\mu = +\frac{1}{2}\mu$$

$\hat{\mu}_2$ tends to be too small

$$var(\hat{\mu}_1) = var(\bar{X})$$
$$\overset{MF2}{=} \frac{\sigma^2}{3}$$

σ^2 ← the variance of each draw

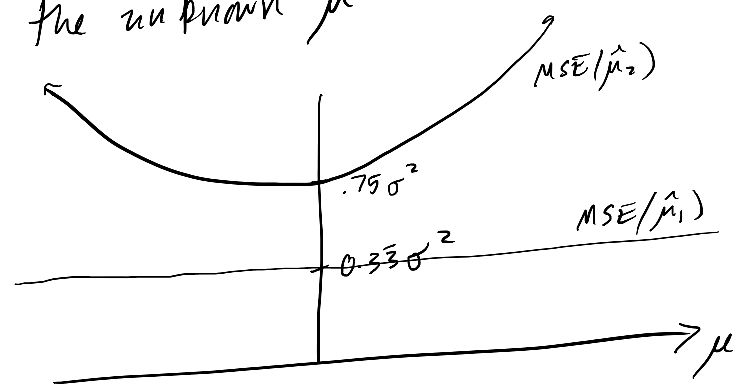
$$var(\hat{\mu}_2) = var\left(\frac{X_1+X_2+X_3}{3-1}\right)$$
$$\overset{MF2}{=} \frac{1}{(3-1)^2}\left[var(X_1)+var(X_2)+var(X_3)\right]$$
$$= \frac{3\sigma^2}{(3-1)^2} = \frac{3}{4}\sigma^2$$

$\therefore$

$$MSE(\hat{\mu}_1) = \frac{\sigma^2}{3}$$

$$MSE(\hat{\mu}_2) = \frac{3}{4}\sigma^2 + \left(+\frac{1}{2}\mu\right)^2$$
$$= \frac{3}{4}\sigma^2 + \frac{\mu^2}{4}$$

Plot these two as a function of the unknown μ:



So, no matter what μ is, $\hat{\mu}_1$ has a smaller typical estimation error.

what about the estimate

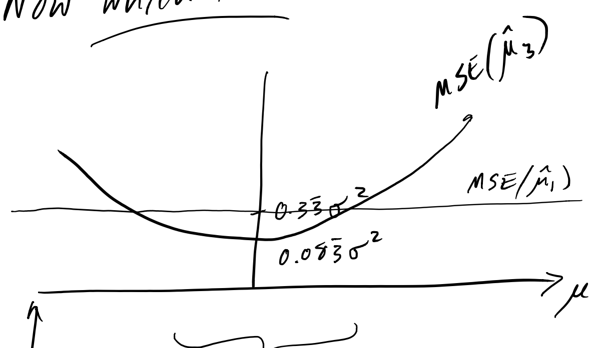
$$\hat{\mu}_3 = \frac{X_1+X_2+X_3}{6} \quad ?$$

$$\text{bias}(\hat{\mu}_3) = E\left(\frac{X_1+X_2+X_3}{6} - \mu\right) \overset{MF1}{=} -\mu + \frac{3\mu}{6}$$

$$= -\frac{1}{2}\mu$$ ← $\hat{\mu}_3$ tends to be too big

$$\text{var}(\hat{\mu}_3) \overset{MF2}{=} \frac{3}{6^2}\sigma^2 = 0.08\bar{3}\,\sigma^2$$

Now which is better?



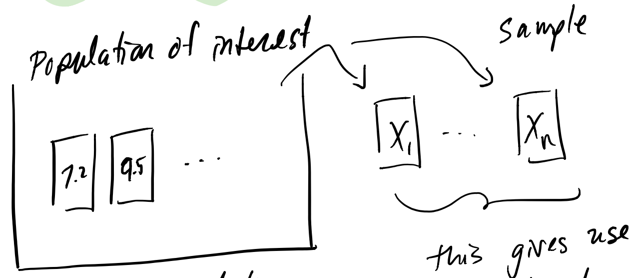
but if μ happens to be far away from 0 then $\hat{\mu}_3$ will be way off

if μ happens to be near 0 then $MSE(\hat{\mu}_3)$ will have smaller typical error

The safe bet: $\hat{\mu}_1$ is better if you have no idea where μ is

Prior information: the only time you might use $\hat{\mu}_3$ is if you have some info that μ is near 0.

Geometric interpretation of the MSE decomposition



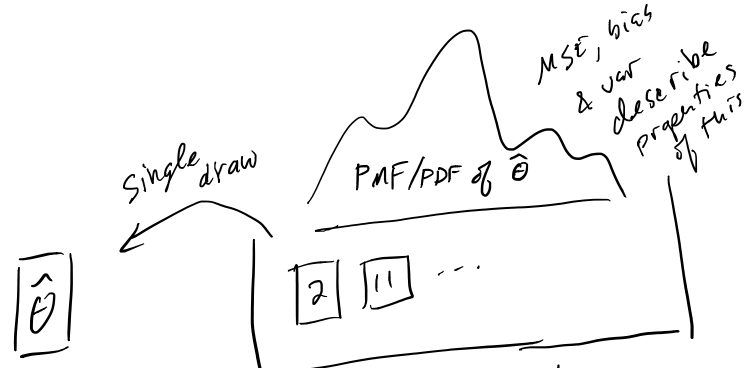
θ is a population parameter we want to estimate

this gives use an estimate $\hat{\theta}$

$MSE(\hat{\theta})$, $bias(\hat{\theta})$, $var(\hat{\theta})$ all describe the sampling variability of $\hat{\theta}$, i.e. the ensemble behavior of $\hat{\theta}$ if you did the experiment over & over again.

i.e. treating $\hat{\theta}$ as a R.V. & studying its PMF or PDF.

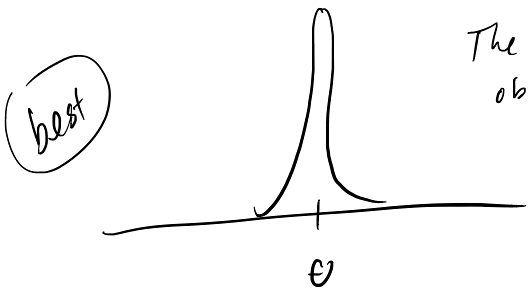
Since $\hat{\theta}$ is itself a R.V. there exists a box model description of its sampling variability



Box model for the Randomness in $\hat{\theta}$

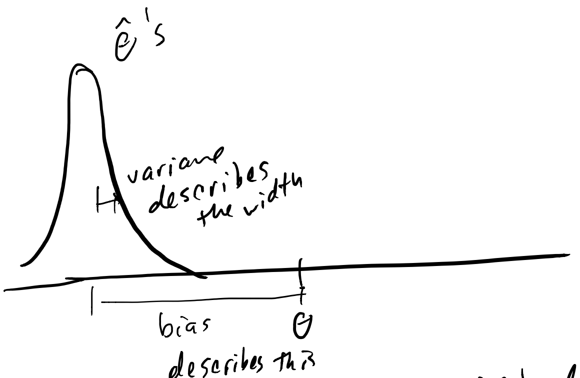
If $\text{bias}(\hat{\theta}) = \text{small}$ & $\text{var}(\hat{\theta}) = \text{small}$

then the PDF/PMF looks like

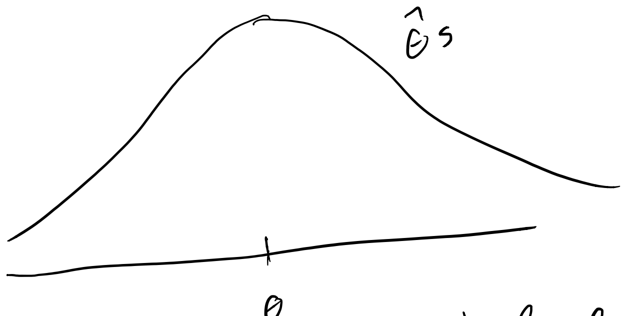


The $\hat{\theta}$ you observe is a draw from this

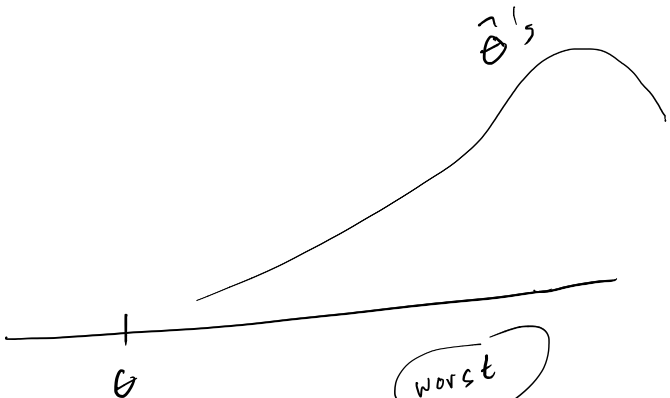
If $\text{bias}(\hat{\theta}) = \text{large}$, $\text{var}(\hat{\theta}) = \text{small}$



If $\text{bias}(\hat{\theta}) = \text{small}$, $\text{var}(\hat{\theta}) = \text{large}$



If $\text{bias}(\hat{\theta}) = \text{large}$, $\text{var}(\hat{\theta}) = \text{large}$



So knowing the "width" of the $\hat{\theta}$'s & the "offset" you can compute the typical estimation error.

Inverse variance weighted averaging

e.g. Suppose our data is X_1, X_2, X_3 where

$$X_1 \sim N(\mu, 4)$$
$$X_2 \sim N(\mu, 16)$$
$$X_3 \sim N(\mu, 8)$$

and we want to estimate μ.

Since X_2 & X_3 have larger variance we probably want to "downweight" them.

Consider a "weighted average" estimate

$$\hat{\mu} = w_1 X_1 + w_2 X_2 + w_3 X_3$$

with weights $w_i \geq 0$ s.t. $w_1 + w_2 + w_3 = 1$.

for example $\bar{X}$ has the above form since

$$\bar{X} = \frac{X_1 + X_2 + X_3}{3} = \underbrace{\tfrac{1}{3}}_{w_1} X_1 + \underbrace{\tfrac{1}{3}}_{w_2} X_2 + \underbrace{\tfrac{1}{3}}_{w_3} X_3$$

$$w_1 + w_2 + w_3 = 1$$

Question: what are the best weights?

e.g. is $\hat{\mu} = \frac{3}{6} X_1 + \frac{1}{6} X_2 + \frac{2}{6} X_3$ better than $\bar{X}$?

down weighting X_2 since it is more volatile.

Answer: To find the "best" w_i we will find the weights which makes $MSE(\hat{\mu})$ smallest. (7)

$$bias(\hat{\mu}) = E(\hat{\mu} - \mu)$$

$$= -E\left(\mu - (w_1 X_1 + w_2 X_2 + w_3 X_3)\right)$$

$$\overset{MF1}{=} -\mu + (w_1 \mu + w_2 \mu + w_3 \mu)$$

$$= -\mu + \underbrace{(w_1 + w_2 + w_3)}_{=1}\mu$$

$$= 0$$

so all weighted average estimates are unbiased.

$$\therefore MSE(\hat{\mu}) = var(\hat{\mu}) + 0^2$$

$$= var(w_1 X_1 + w_2 X_2 + w_3 X_3)$$

$$\overset{MF2}{=} w_1^2 4 + w_2^2 16 + w_3^2 8$$

To minimize this w.r.t w_1, w_2, w_3 you can use calculus to show the "best" w_i's are simply proportional to $\frac{1}{variance(X_i)}$

$$w_1 = \frac{c}{4}, \quad w_2 = \frac{c}{16}, \quad w_3 = \frac{c}{8}$$

where $c = 1 \Big/ \left(\frac{1}{4} + \frac{1}{16} + \frac{1}{8}\right) = 1 \Big/ \left(\frac{4+1+2}{16}\right)$

$$= \frac{16}{7}$$

$\therefore$ The best estimate $\hat{\mu}$ is

$$\hat{\mu} = \frac{16}{7}\left(\frac{1}{4}X_1 + \frac{1}{16}X_2 + \frac{1}{8}X_3\right)$$

downweights X_2 proportional to $\frac{1}{var(X_2)}$

Lets finish by seeing how much better it does than $\bar{X}$. (8)

$$var(\bar{X}) \overset{MF2}{=} \frac{4}{9} + \frac{16}{9} + \frac{8}{9} = 3.111$$

$$var(best\ \hat{\mu}) = \left(\frac{16}{7}\right)^2\left(\frac{1}{4^2} \cdot 4 + \frac{1}{16^2} 16 + \frac{1}{8^2} 8\right)$$

$$= \left(\frac{16}{7}\right)^2\left(\frac{1}{4} + \frac{1}{16} + \frac{1}{8}\right)$$

$$= \left(\frac{16}{7}\right)^2\left(\frac{7}{16}\right) = \frac{16}{7} = 2.286$$