WEEK 3 (24 JAN 2012)

REVIEW OF SOME STATISTICAL CONCEPTS (CONT.)

- CONDITIONAL PROB. DENSITY FUNCTION

$$f(x, y) = P(X = x \mid Y = y)$$

IS KNOWN AS THE CONDITIONAL PDF.

READ: THE PROBABILITY THAT X TAKES ON THE VALUE OF x GIVEN THAT Y HAS ASSUMED THE VALUE OF y

IN THE SAME MANNER, $f(y \mid x) = P(Y = y \mid X = x)$.

THE CONDITIONAL PDF CAN BE DERIVED BY:

$$f(x \mid y) = \frac{f(x, y)}{f(y)}$$ JOINT PDF; MARGINAL PDF OF Y $\Rightarrow$ CONDITIONAL PDF OF X

$$f(y \mid x) = \frac{f(x, y)}{f(x)}$$ $\Rightarrow$ CONDITIONAL PDF OF Y

EX: $$f(X = -2 \mid Y = 3) = \frac{f(X = -2, Y = 3)}{f(Y = 3)} = \frac{0.27}{0.51} = 0.53,$$

$$f(X = 2 \mid Y = 6) = \frac{f(X = 2, Y = 6)}{f(Y = 6)} = \frac{0.10}{0.49} = 0.20$$

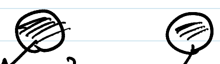
JOINT PDF OF X; MARGINAL PDF OF Y

- STATISTICAL INDEPENDENCE

> "TWO RANDOM VARIABLES ARE STATISTICALLY INDEPENDENT, OR INDEPENDENTLY DISTRIBUTED
> IF KNOWING THE VALUE THAT ONE WILL TAKE DOES NOT REVEAL ANYTHING ABOUT WHAT VALUE MAY TAKE.

EXAMPLE TOSSING TWO COINS SIMULTANEOUSLY

1ST 2ND



$\{H, T\}$ $\{H, T\}$

IF X = NUMBER OF HEADS SHOWING ON A TOSS OF THE 1ST COIN

Y = NUMBER OF HEADS OCCURING ON THE SECOND COIN.

SO X AND Y ARE INDEPENDENT RVs.

EXAMPLE

MR. A AND MR B

EXPENDITURE ON ICE-CREAM

EXPENDITURE ON ICE-CREAM

" INDEPENDENT RANDOM VARIABLES "

EXAMPLE GDP AND π (INFLATION)

" DEPENDENT RANDOM VARIABLES "

- WHEN RVs ARE STATISTICALLY INDEPENDENT, THEIR JOINT PDF IS EQUAL TO THE PRODUCT OF THEIR PDFs, AND VICE VERSA.

$\underbrace{\text{i.e.}}_{}$, THAT IS

$$f(x, y) = f(x) \cdot f(y)$$

FOR EACH AND EVERY PAIR OF VALUES x AND y.

NOTE THAT: THE CONVERSE IS ALSO TRUE.

CHARACTERISTICS OF PROBABILITY DISTRIBUTIONS

A PROBABILITY DISTRIBUTION ⇒

- MEAN OR EXPECTED VALUE
- VARIANCE.

TWO IMPORTANT CHARACTERISTICS OF ANY DISTRIBUTION.

MEAN OR EXPECTED VALUE

THE EXPECTED VALUE OF A DISCRETE RV X DENOTED BY $E(X)$ IS DEFINED AS FOLLOWS:

$$E(X) = \underbrace{\sum_{i=1}^{n} x_i f(x_i)}$$

WEIGHTED AVERAGE OF THE VALUES OF THE RD X, WITH THE WEIGHTS BEING THE PROBABILITIES ATTACHED TO EACH VALUE

= MEAN VALUE OF A RV IN INFINITE NUMBER OF REPETITIONS OF THE EXPERIMENT (REPEATED SAMPLES)

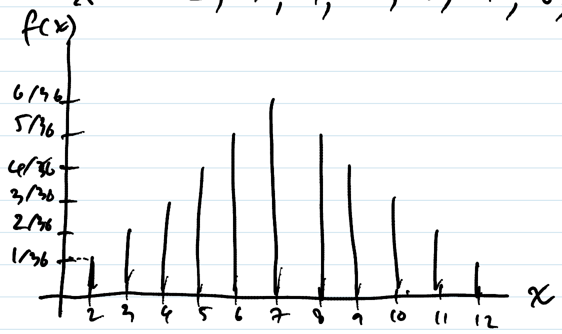
$$= x_1 f(x_1) + x_2 f(x_2) + \dots + x_n f(x_n)$$

$$= \sum x f(x)$$

EX: X = THE SUM OF THE NUMBERS APPEARED ON THE TWO DICES

THAT HAS BEEN THROWN SIMULTANEOUSLY

$X = 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12$



$$E(X) = \sum x f(x)$$

$$= 2\left(\tfrac{1}{36}\right) + 3\left(\tfrac{2}{36}\right) + 4\left(\tfrac{3}{36}\right) + \cdots + 12\left(\tfrac{1}{36}\right).$$

$$E(X) = 7$$

WHICH IS THE MEAN OR AVERAGE VALUE OF THE SUM OF THE NUMBERS OBSERVED IN THE EXPERIMENT.

PROPERTIES OF EXPECTED VALUES

1. $E(b) = b$ WHERE b IS A CONSTANT.
2. $E(aX + b) = a\,E(X) + b$

GENERAL FORM: $E(a_1 X_1 + a_2 X_2 + \cdots a_N X_N + b) =$

$$a_1 E(X_1) + a_2 E(X_2) + \cdots + a_N E(X_N) + b.$$

* 3. IF X AND Y ARE INDEPENDENT RVs,

$$E(X \cdot Y) = E(X) \cdot E(Y)$$

BUT $$E\left(\frac{X}{Y}\right) \neq \frac{E(X)}{E(Y)}$$

IN ALL CASES.
(EVEN IF X AND Y ARE INDEPENDENT.)

4. IF X IS A RV WITH PDF $f(x)$ AND $g(X)$ IS ANY FUNCTION OF X,

THEN
$$\boxed{E[g(X)] = \sum_x g(x) f(x) \quad \text{IF } X \text{ IS DISCRETE}}$$

AND
$$\boxed{= \int_{-\infty}^{\infty} g(x) f(x)\,dx \quad \text{IF } X \text{ IS CONTINUOUS.}}$$

EXAMPLE: LOOK AT THIS PDF:

IF X IS CONTINUOUS,

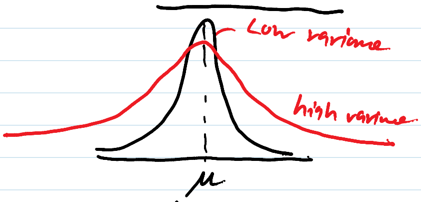
EXAMPLE: LOOK AT THIS PDF:

x	-2	1	2
$f(x)$	$\frac{5}{8}$	$\frac{1}{8}$	$\frac{2}{8}$

$$E(X) = -2\left(\frac{5}{8}\right) + 1\left(\frac{1}{8}\right) + 2\left(\frac{2}{8}\right)$$
$$= -\frac{5}{8}$$

AND
$$E(X^2) = (-2)^2\left(\frac{5}{8}\right) + (1)^2\left(\frac{1}{8}\right) + (2)^2\left(\frac{2}{8}\right)$$
$$= \frac{29}{8}.$$

VARIANCE: SPREAD OR DISPERSION OF THE X VALUES AROUND THE EXPECTED VALUE.



LET X BE A RANDOM VARIABLE
AND LET $E(X) = \mu$.

THE DISPERSION OR SPREAD OF THE X VALUES "AROUND" THE EXPECTED VALUE IS MEASURED BY THE VARIANCE:

$$VAR(X) = \sigma_x^2 = \underbrace{E[X - E(X)]^2}$$

READ: SIGMA SQUARED

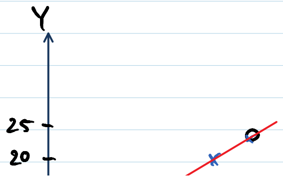
"AVERAGE" SQUARED DIFFERENCE BETWEEN THE RV X AND ITS EXPECTED VALUE.

THE LARGER THE VARIANCE OF A RANDOM VARIABLE, THE MORE "SPREAD OUT" THE VALUE OF THE RANDOM VARIABLE AROUND THEIR MEAN VALUE.

TWO - VARIABLE REGRESSION ANALYSIS (GUJARATI, CH. 2)

HYPOTHETICAL OBSERVATION: WHERE X = OUTPUT, Y=FERTILIZER

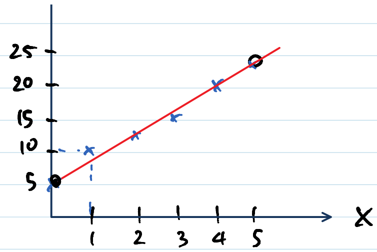
X	Y
0	5
1	10



$\therefore Y_i = a + bX_i$

a, b = parameters

x	y
0	5
1	10
2	12
3	15
4	20
5	23



a, b = parameters
$\hat{a}, \hat{b}$ = estimators.

IF WE JUST TAKE TWO OBSERVATIONS?

1^{st} OBS: $5 = a + b0$
5^{th} OBS: $23 = a + b.5$

$\hat{a} = 5$
$\hat{b} = 3.6$

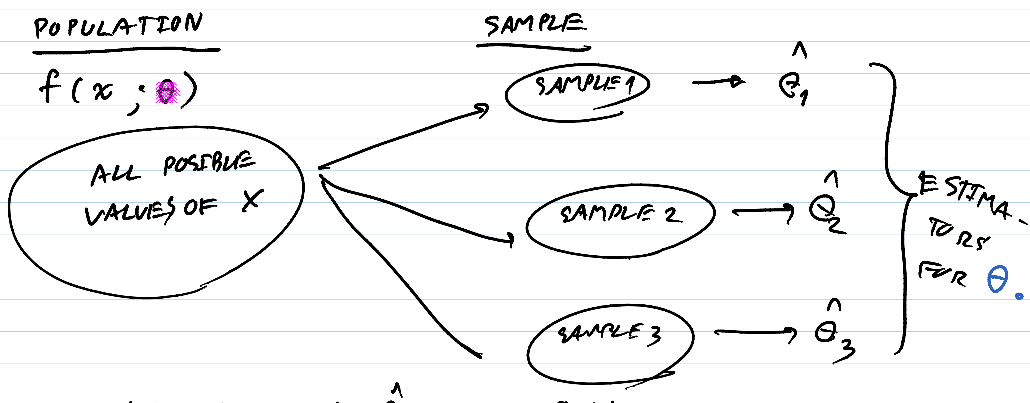
CRITICISM: (1) THEN, IF YOU PICK UP 2^{nd} AND 5^{TH} OBS, YOU WOULD GET ANOTHER ESTIMATED LINE.

Q: WHICH LINE IS THE BEST?

(2) NOT ALL OBSERVATIONS ARE USED TO ESTIMATE THE RELATIONSHIP.

ESTIMATORS WITH DESIRABLE PROPERTIES

- BASIC CHARACTERISTICS OF A DISTRIBUTION ⇒ WE LOOK AT MEAN AND VARIANCE.



IF WE USE ONLY $\hat{\theta}_1$, WE WOULD ENCOUNTER W/ "SAMPLING ERROR" ($\because$ ONLY CONTAIN A SUBSET OF POPULATION)

SO, WE NEED THE LEAST SAMPLING ERROR → HOW?

LET $\hat{\theta}$ BE AN ESTIMATOR FOR THE TRUE PARAMETER θ

(1) MEAN OF $\hat{\theta}$ = $E(\hat{\theta})$

(2) VARIANCE OF $\hat{\theta}$ = $\underbrace{E\left[\hat{\theta} - E(\hat{\theta})\right]^2}_{\text{AVERAGE SQUARED DIFFERENCE BET. } \hat{\theta} \text{ AND ITS EXPECTED VALUE.}}$.

(3) SAMPLING ERROR = $\hat{\theta} - \theta$

(4) BIAS = $E(\hat{\theta}) - \theta$

EXPECTED VALUE OF THE ESTIMATOR

THE TRUE PARAMETER

2

④ BIAS $= E(\hat{\theta}) - \theta$ → THE TRUE PARAMETER

THE ESTIMATOR

⑤ MSE $= E[\hat{\theta} - \theta]^2$

"MEAN SQUARE ERROR"

MEASURE DISPERSION OF $\hat{\theta}$ AROUND THE TRUE VALUE.

WE WANT AN ESTIMATOR THAT GIVES THE LEAST MSE.

$$MSE = E[\hat{\theta} - \theta]^2$$
$$= E[\hat{\theta} - E(\hat{\theta}) + E(\hat{\theta}) - \theta]^2$$
$$= E[(\hat{\theta} - E(\hat{\theta})) + (E(\hat{\theta}) - \theta)]^2$$
$$= E\left\{[\hat{\theta} - E(\hat{\theta})]^2 + [E(\hat{\theta}) - \theta]^2 + 2[(\hat{\theta} - E(\hat{\theta}))(E(\hat{\theta}) - \theta)]\right\}$$
$$= E[\hat{\theta} - E(\hat{\theta})]^2 + E[E(\hat{\theta}) - \theta]^2$$
$$+ \underbrace{2E[(\hat{\theta} - E(\hat{\theta}))(E(\hat{\theta}) - \theta)]}_{=0}.$$

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

PROOF:

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

LET $E(\hat{\theta}) = A$ (ANY CONSTANT)

$$= 2E[\hat{\theta}E(\hat{\theta}) - \hat{\theta}\theta - E(\hat{\theta})E(\hat{\theta}) + E(\hat{\theta})\theta]$$
$$= 2[E(\hat{\theta})E(\hat{\theta}) - E(\hat{\theta})\theta - E(\hat{\theta})E(\hat{\theta}) + E(\hat{\theta})\theta]$$
$$= 2[\cancel{A \cdot A} - \cancel{A \cdot \theta} - \cancel{A \cdot A} + \cancel{A \cdot \theta}]$$
$$= 0$$

END OF THE PROOF.

$$\boxed{MSE = E[\hat{\theta} - E(\hat{\theta})]^2 + E[\underbrace{E(\hat{\theta}) - \theta}_{BIAS}]^2}$$

$$= \underbrace{VAR(\hat{\theta})} + \underbrace{(BIAS\ \hat{\theta})^2}$$

IF THE VARIANCE IS SMALL, "THE ESTIMATOR IS NOT SO MUCH AFFECTED BY THE RANDOMNESS OF DATA

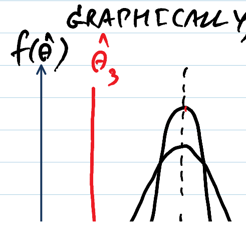
IF BIAS IS SMALL THEN $\hat{\theta}$ HAS A DISTRIBUTION AROUND MEAN

(NOTE: $E(BIAS) = BIAS$)

IF AN ESTIMATOR HAS BIAS = 0 AND HAS THE MINIMUM VARIANCE, THEN MSE WILL BE MINIMUM!

SUMMARY TO GET LEAST MSE, YOU NEED < 0 BIAS & MIN VARIANCE

GRAPHICALLY, ...



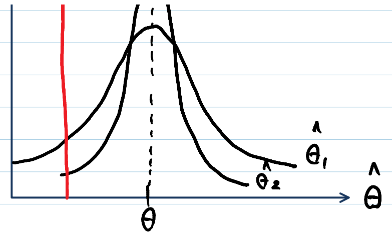
$E(\hat{\theta}_1) = \theta$ OR $E(\hat{\theta}_1) - \theta = 0$

$E(\hat{\theta}_2) = \theta$ OR $E(\hat{\theta}_2) - \theta = 0$

NO BIAS

$\hat{\theta}_2$ IS BETTER THAN $\hat{\theta}_1$

BECAUSE IT HAS LOWER VARIANCE



NO BIAS

$\hat{\theta}_2$ IS BETTER THAN $\hat{\theta}_1$
BECAUSE IT HAS LOWER VARIANCE.

FOR $\hat{\theta}_3$, WE WILL NEVER CHOOSE IT B/C IT CONTAINS "BIASNESS"

(IF IT'S BIASED → NO CHANCE TO HIT THE TRUE PARAMETER !)

" BIASNESS " COMES FIRST WHEN CHOOSING AN ESTIMATOR THAT HAS THE LEAST MSE.

SMALL SAMPLE PROPERTIES

(WHEN YOU HAVE TO DEAL WITH DATA WITH SMALL SAMPLE SIZE)

① UNBIASEDNESS : BIAS $= E(\hat{\theta}) - \theta = 0$.

$\hat{\theta}$ IS AN UNBIASED ESTIMATOR OF θ IF
$E(\hat{\theta}) = \theta$.

② MINIMUM VARIANCE : IF $VAR(\hat{\theta}) \leq VAR(\hat{\hat{\theta}})$

WHERE $\hat{\hat{\theta}}$ IS ANY OTHER UNBIASED ESTIMATOR, THEN WE CAN SAY THAT $VAR(\hat{\theta})$ IS MINIMUM.

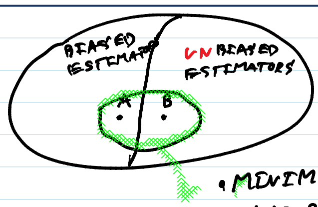
SUMMARY

$\hat{\theta}$ IS AN EFFICIENT ESTIMATOR FOR θ

IFF
(IF AND ONLY IF)

① $\hat{\theta}$ IS AN UNBIASED ESTIMATOR FOR θ,

② $\hat{\theta}$ HAS THE MINIMUM VARIANCE AMONG THE CLASS OF UNBIASED ESTIMATORS.



ALL POSSIBLE ESTIMATORS

• MINIMUM VARIANCE
• VAR B > VAR A

CHOOSE B → B IS EFFICIENT B/C IT IS UNBIASED.