Note Title 11/23/2009

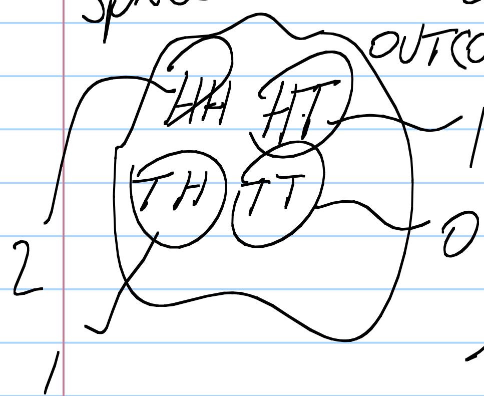
# STT 200 11-23-09

Ch 15 + 16 NOTION OF "EXPECTATION OR AVERAGE"

NEED NOTION OF RANDOM VARIABLE.

r.v. = "JUST A NUMERICAL FUNCTION ON THE OUTCOMES OF A PROBABILITY EXPERIMENT."

SPACE OF POSSIBLE OUTCOMES




r.v. $X$ = "# OF H IN TWO TOSSES"

$$P(X=1) \overset{DEF}{=} P(HT \cup TH) = 2/4$$

$$P(X=2) = P(HH)$$

r.v. $2X$, $X^3$, $\sin(X)$, $1/X$ UNDEFINED AT OUTCOME TT $1/0$.

eg ALSO – r.v. $X$ = \$ AMT OF COIN

\$.10 \$1.63 \$87.42 CLASS $Y$ = \$ " " PAPER

r.v. $X+Y$

$X+Y$ is ALSO A r.v. \$ MONEY

DEFINE THE EXPECTATION OF r.v. $X$

$$E(X+Y) = \sum_{\text{STUDENTS}} \big(X(\text{STUDENT}) + Y(\text{STUDENT})\big)\, P(\text{STUDENT})$$

(WEIGHTED AVG)

$$= \sum_{\text{STUDENTS}} X(\text{STUDENT})\, P(\text{STUDENT}) + \sum_{\text{STUDENTS}} Y(\text{STUDENT})\, P(\text{STUDENT}) = EX + EY$$

**ALWAYS** $E(X+Y) = EX + EY$ IN PARTICULAR WE DO NOT REQUIRE **INDEPENDENCE**

ALSO, $E\,C = \sum_{\text{POINTS OF SAMPLE SPACE}} C\; P(\text{POINT}) = C \underbrace{\sum_{\text{POINTS}} P(\text{POINT})}_{1} = C$

CONSTANT r.v.

ALSO $E\,cX = \sum_{\text{POINTS}} c\,X(\text{POINT})\,P(\text{POINT})$

$= c \sum_{\text{POINTS}} X(\text{POINT})\,P(\text{POINT}) = c\,EX$

PUT TOGETHER: $E(aX + bY + c) = aEX + bEY + c$

---

EX AS ~ THE TIME AVG OF SAMPLES

$X_1, X_2, \ldots, X_n$ WITH-REPL STUDENTS SAMPLED ~~(EQ PR)~~

$n \to \infty$ $\quad \bar{X} = \dfrac{X_1 + \cdots + X_n}{n} = \dfrac{X_2 \;\; X_3 \;\; \cdots}{n}$

\$2.50

\$2.50

$$\bar{X} \sim \sum X(\text{STUDENT})\, \underbrace{P(\text{STUDENT})}_{\approx (\#\text{TIMES THAT STUDENT IS SAMPLED AMONG } n)/n} = EX$$

SAMPLE MEAN $\bar{X}$ "SHOULD" ACCORDING TO "LAW OF AVGS" BE $\approx$ SAME AS $EX$.

MODEL BASED.

EXAMPLE 1.

| | |
|---|---|
| *H1 H2 | *H1 T2 |
| T1 H2 | T1 T2 |

* IF $X_1 = 1$

$X_1 = 1$ H1
$\quad\;\; 0$ NOT

$EX_1 = \sum$ "VALUE × PROB" (OVER ENTIRE SPACE OF OUTCOMES)

$= 1\, P(H1\,H2) + 1\, P(H1\,T2) + 0\, P(T1\,H2) + 0\, P(T1\,T2) = 1/2$

LIKEWISE $E X_2 = \frac{1}{2}$ $\quad X_2 = \begin{cases} 1 & H2 \\ 0 & \text{NOT} \end{cases}$

SO $E(\text{\# HEADS IN TWO TOSSES}) = E(X_1 + X_2)$

$= \frac{1}{2} + \frac{1}{2} = 1$

---

EXAMPLE 2:

$2^2$ — HH

HT TH — EACH $1^2$

TT — $0^2$

$E(\#H)^2$

$E(\#H)^2 = \sum$ VALUE × PROBY =

(WAY 1) DIRECT OFF MODEL

$= 2^2\left(\frac{1}{4}\right) + 1^2\left(\frac{1}{4}\right) + 1^2\left(\frac{1}{4}\right) + 0^2\left(\frac{1}{4}\right)$

NOTICE - THIS RESULT IS SAME IF WE GROUP BY VALUES OF $(\#H)^2$.

$$E(\#H)^2 = 2^2\left(\tfrac{1}{4}\right) + 1^2\left(\tfrac{1}{2}\right) + 0^2\left(\tfrac{1}{4}\right)$$

DISTINCT PROBY'S

$$\left[\begin{array}{cccc} (\#H)^2 & 2 & 1 & 0 \\ \text{PROBY} & \tfrac{1}{4} & \tfrac{1}{2} & \tfrac{1}{4} \end{array}\right] \Bigg\} \text{ PROBABILITY DISTRIBUTION}$$

SO $EX$ IS $\displaystyle\sum_{\text{ELEMENTS OF POP}} X(\text{ELEMENT})\, P(\text{ELEMENT})$

A<u>LSO</u> SAME AS $\displaystyle\sum_{x \text{ A POSSIBLE VALUE OF r.v.}} x \; P(X=x)$ } DONE FROM THE PROBY DIST'N

SO. SUPPOSE BUSINESS WAS e-"HITS"

LET $X_1$ = \$ PURCHASE FROM FIRST HIT

$EX_1$ SO $E(X_1 + \cdots + X_{5000}) = 5000\, EX_1$

LIKEWISE CAN ESTIMATE $EX_1 \cong \dfrac{X_1 + \cdots + X_n}{n}$.

QUESTION: HOW MUCH VARIABILITY?

$(a+b)^2 = a^2 + 2ab + b^2$

CONCEPT OF VAR $X$, SD $X$.

NUMBER

DEF $\operatorname{Var} X \overset{DEF}{=} E(X - EX)^2 = E(X^2 - 2(EX)X + (EX)^2)$

$= E(X^2) + E(-2(EX)X) + E(EX)^2$

$= E(X^2) - 2(EX)(EX) + (EX)^2$

SO $\operatorname{Var} X = E(X^2) - (EX)^2$

PROPERTIES OF $\operatorname{Var} X$ ?

$$\operatorname{Var}(cX) = E(cX - EcX)^2 = c^2 E(X - EX)^2 = c^2 \operatorname{Var} X$$

$$\operatorname{Var}(X + \cancel{c}) = E((X + \cancel{c}) - (EX + \cancel{c}))^2 = \operatorname{Var} X$$

RELATION OF $\operatorname{Var} X$ (AND $SD\, X \overset{DEF}{=} \sqrt{\operatorname{Var} X}$) TO INDEPENDENCE.

NOTION OF INDEPENDENT r.v. $X, Y$:

$$P(Y = y \mid X = x) = P(Y = y) \text{ ALL } x, y.$$

BIG

RECALL EVENTS $A, B$ ARE INDEP IF

$$P(B \mid A) = P(B)$$

EQUIV TO

$$P(A \cap B) = P(A)P(B)$$

IF r.v. $X, Y$ ARE INDEP.

$$E(XY) = \sum_{x,y} xy\, P(X=x \text{ AND } Y=y)$$

$$\overset{\text{INDEP}}{=} \sum_{x,y} xy\, P(X=x)P(Y=y)$$

$$= \left(\sum_x x\, P(X=x)\right)\left(\sum_y y\, P(X=x)\right) = (EX)(EY)$$

SO $E(XY) \underset{\text{INDEP}}{\overset{\text{IF}}{=}} (EX)(EY)$

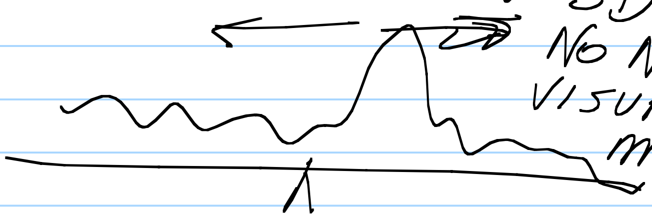
CONSEQUENCE: $\mathrm{Var}(X+Y) \underset{\text{INDEP}}{\overset{\text{IF}}{=}} \mathrm{Var}\, X + \mathrm{Var}\, Y$

App'l to Casino Game:

Suppose each play returns r.v. $X$

$EX = \$.2 \qquad \mathrm{Var}\,X = \cancel{\$1.50}\ \$2.25 \qquad SD\,X = \$1.5$

$n = 10{,}000$ independent plays.



1.5 SD.

NO NICE VISUAL MEANING

$x$

$EX = \$.2$ (ONE PLAY)

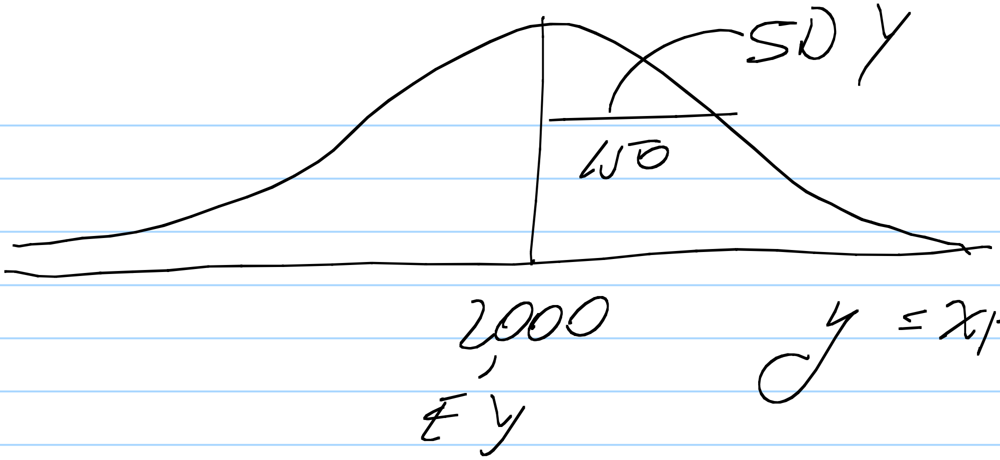
$$X_1 + \cdots + X_{10{,}000} = Y$$

$$EY = 10{,}000(.2) = 2000$$

$$SD\,Y = \sqrt{\textstyle\sum_i \mathrm{Var}\,X_i} = \sqrt{10{,}000 \cdot 1.5^2} = \$150$$

ERROR FIXED

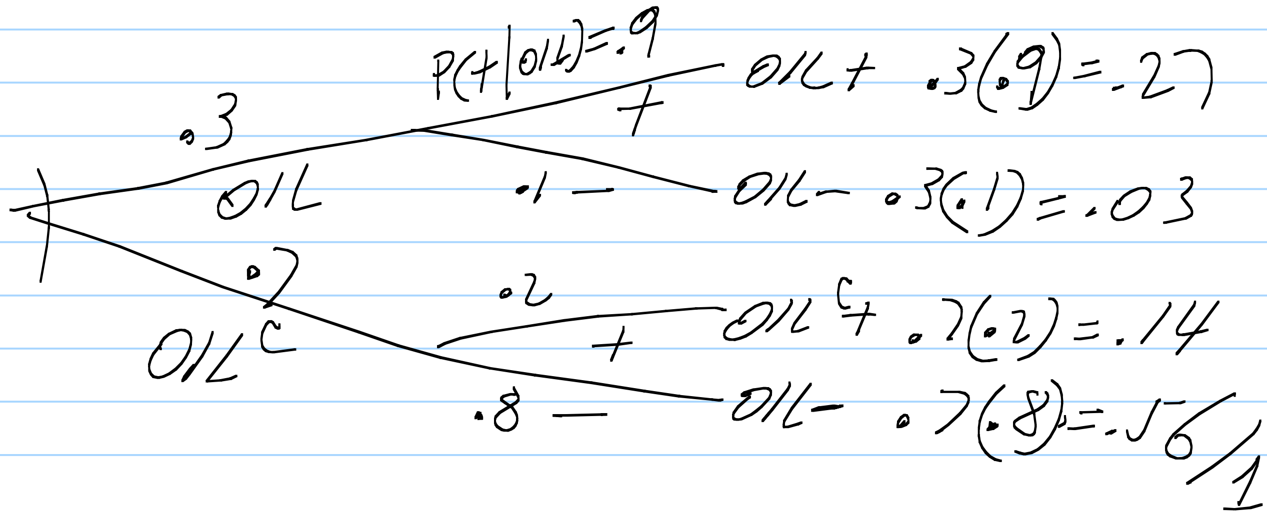
≈ CENTRAL LIMIT THEOREM




95% INTERVAL IS $2000-2(150)$, $2000+2(150)$
$=[1700, 2300]$.

$y = x_1 + \dots + x_{10000}$

CONNECTED W/ TREE DIAGRAM



SUPPOSE COSTS 200 TO DRILL. RETURN FROM OIL
" 40 TO TEST IS 800.

POLICY I: JUST DRILL – NO TEST.

$$E(\underset{I}{NET}) = .3(\underset{DRILL}{800 - 200}) + .7(0 - 200)$$

POLICY II: TEST, BUT ONLY DRILL IF TEST IS +.

$$E(\underset{II}{NET}) = \underset{OIL+}{.27}(-200 - 40 + 800) + \underset{OIL-}{.3(.1)}(-0 - 40 + 0)$$

$$+ \underset{OIL^{C}+}{.3(.2)}(-200 - 40 + 0) + \underset{OIL-}{.3(.8)}(-0 - 40 + 0)$$

SO POLICY I JUST DRILL

$$E(\text{NET I}) = .3(600) - .7(200) = 40$$

$$E(\text{NET II}) = .27(560) - .03(40) - .06(240) - .24(40)$$

$$= 151.2 - 1.2 - 14.4 - 9.6 = 126.$$

N