LEC. 21 10/18/99

PROBABILITY III.

RANDOM VARIABLE IS A FUNCTION FROM S TO $\mathbb{R}$

SAMPLE SPACE

REAL NUMBERS

NOTATION $X(t)\big|_{t \in S}$, $X(s)\big|_{s \in S}$

EXAMPLE: A COIN IS FLIPPED n TIMES

$X(t)$ = THE NUMER OF HEADS IN t

$n = 3$

$X(TTT) = 0$

$X(HTT) = X(THT) = X(TTH) = 1$

$X(HHT) = X(HTH) = X(THH) = 2$

$X(HHH) = 3$

EXAMPLE: $X(t)$ = the SUM OF NUMBERS THAT APPEAR WHEN A PAIR OF DICE IS ROLLED

$X(1,1) = 2$

$X(2,1) = X(1,2) = 3$

- - - - - -

$X(5,6) = X(6,5) = 11$

$X(6,6) = 12$

NOTE: RANDOM VARIABLE IS NEITHER VARIABLE NOR RANDOM

THE EXPECTED VALUE (OR EXPECTATION) of X

$$E(X) = \sum_{s \in S} p(s) \cdot X(s)$$

EXAMPLE: $X(s)$ = NUMBER OF HEADS IN s

$n=3$ $E(X) = \frac{1}{8}\cdot 0 + \frac{1}{8}\cdot 1 + \frac{1}{8}\cdot 1 + \frac{1}{8}\cdot 1 + \frac{1}{8}\cdot 2 + \frac{1}{8}\cdot 2 + \frac{1}{8}\cdot 2 + \frac{1}{8}\cdot 3$

$$= 0\cdot\frac{1}{8} + 1\cdot\frac{3}{8} + 2\cdot\frac{3}{8} + 3\cdot\frac{1}{8} = \frac{3}{2}$$

A PRACTICAL FORMULA: $E(X) = \sum_{r \in X(S)} p(X=r)\cdot r$

COUNTING $E(X)$ BY LARGER PORTIONS

EXAMPLE: $X(s)$ = THE SUM ON A PAIR OF DICE

$$E(X) = 2\cdot p(X=2) + 3\cdot p(X=3) + 4\cdot p(X=4) + \cdots + 12\cdot p(X=12)$$

$$= 2\cdot\frac{1}{36} + 3\cdot\frac{2}{36} + 4\cdot\frac{3}{36} + 5\cdot\frac{4}{36} + 6\cdot\frac{5}{36} + 7\cdot\frac{6}{36} +$$

$$+ 8\cdot\frac{5}{36} + 9\cdot\frac{4}{36} + 10\cdot\frac{3}{36} + 11\cdot\frac{2}{36} + 12\cdot\frac{1}{36} = 7$$

EXAMPLE: n BERNOULLI TRIALS

$$E(X) = \sum_{k=1}^{n} k\cdot p(X=k) = \sum_{k=1}^{n} k\cdot C(n,k)\cdot p^k q^{n-k}$$

NOTE: $k\cdot C(n,k) = \frac{k\cdot n!}{k!\,(n-k)!} = \frac{n!}{(k-1)!\,(n-k)!} = \frac{n\cdot(n-1)!}{(k-1)!\,(n-k)!} = n\,C(n-$

So, $E(X) = \sum_{k=1}^{n} n \cdot C(n-1, k-1) \cdot p^k \cdot q^{n-k} =$

$$= np \cdot \sum_{k=1}^{n} C(n-1, k-1)\, p^{k-1} \cdot q^{(n-1)-(k-1)} =$$

$$= np \cdot \underbrace{\sum_{j=0}^{n-1} C(n-1, j) \cdot p^j \cdot q^{(n-1)-j}}_{\text{BY THE BINOMIAL FORMULA } = (p+q)^{n-1} = 1^{n-1} = 1} = np$$

<u>Th</u> X, Y ARE RANDOM VARIABLES ON A SPACE S.

$$E(X+Y) = E(X) + E(Y)$$

$$E(aX) = a \cdot E(X)$$

<u>PROOF</u>. $E(X+Y) = \sum p(s)\big(X(s) + Y(s)\big) = \sum p(s) X(s) +$

$+ \sum p(s) \cdot Y(s) = E(X) + E(Y)$

$E(aX) = \sum p(s) \cdot a \cdot X(s) = a \sum p(s) \cdot X(s) = a E(X)$

COR. $E(X_1 + \dots + X_n) = E(X_1) + \dots + E(X_n)$

$E(aX+b) = aE(X) + b$

$$E(aX+b) = E(aX) + E(b) = aE(X) + E(b) = aE(X) + b$$

$$E(b) = \sum p(s) \cdot b = b \cdot \sum p(s) = b \cdot 1 = b$$

INDEPENDENT RANDOM VARIABLES IF

$$p(X=a \text{ AND } Y=b) = p(X=a) \cdot p(Y=b)$$

i.e. THE EVENTS $X=a$ AND $Y=b$ ARE INDEPENDENT FOR ALL a, b.

EXAMPLE $X_1(s)$ = THE NUMBER ON THE FIRST DIE
$X_2(s)$ = — " — SECOND — " —

$$X_1(i,j) = i \qquad X_2(i,j) = j$$

$$\left.\begin{array}{l} p(X_1=i \text{ AND } X_2=j) = p(i,j) = \frac{1}{36} \\ p(X_1=i) = \frac{1}{6} \;;\; p(X_2=j) = \frac{1}{6} \end{array}\right\} \Rightarrow X_1, X_2 \text{ ARE INDEPENDENT}$$

X_1 AND $X_3 = X_1 + X_2$ ARE NOT INDEPENDENT

INDEED $p(X_1=1 \text{ AND } X_3=10) = 0$

$$p(X_1=1) = \frac{1}{6}; \quad p(X_3=10) = \frac{3}{36} = \frac{1}{12} \neq 0$$

Th. IF X, Y ARE INDEPENDENT THEN $E(XY) = E(X) \cdot E(Y)$

PROOF.
$$E(XY) = \sum_{\substack{a \in X(S) \\ b \in Y(S)}} a \cdot b \cdot p(X=a \text{ AND } Y=b) =$$
$$= \sum a \cdot b \cdot p(X=a) \cdot p(Y=b) =$$
$$= \left[\sum a \cdot p(X=a)\right] \cdot \left[\sum b \cdot p(Y=b)\right] =$$
$$= E(X) \cdot E(Y)$$

THE VARIANCE OF X IS A MEASURE OF HOW WIDELY X IS DISTRIBUTED AROUND ITS EXPECTED VALUE E(X). THE NAIVE TRY:

+'S AND −'S AROUND E(X) ARE BALANCED! $\Leftarrow$ $V = E(X - E(X))$ FAILS SINCE $E(X - E(X)) = E(X) - E(X) = 0$

$$V = E(X - E(X))^2 = \sum_{s \in S} [X(s) - E(X)]^2 \cdot p(s)$$

STANDARD DEVIATION $\sigma(X) = \sqrt{V(X)}$

th. $V = E(X^2) - [E(X)]^2$

PROOF $V = E(X - E(X))^2 = E[X^2 - 2X\,E(X) + [E(X)]^2] =$

$= E(X^2) - E(2 \cdot X \cdot E(X)) + [E(X)]^2 =$

$= E(X^2) - 2 \cdot E(X) \cdot E(X) + [E(X)]^2 =$

$= E(X^2) - [E(X)]^2$

EXAMPLE. ONE BERNOULLI TRIAL $X(t) = \begin{cases} 1, & \text{IF } t = \text{SUC} \\ 0, & \text{IF } t = \text{FAIL} \end{cases}$

NOTE, THAT $X^2(t) = X(t)$. $E(X) = p = E(X^2)$

$V(X) = E(X^2) - [E(X)]^2 = p - p^2 = p(1-p) = p \cdot q$

Th. IF $X_1, \dots, X_n$ ARE PAIRWISE INDEPENDENT THEN $V(X_1 + \dots + X_n) = V(X_1) + \dots + V(X_n)$

PROOF. $n=2$ $V(X+Y) = E((X+Y)^2) - [E(X+Y)]^2 =$

$$= E(X^2 + 2XY + Y^2) - [E(X) + E(Y)]^2 =$$

$$= E(X^2) + 2E(XY) + E(Y^2) - [E(X)]^2 - 2E(X)\cdot E(Y) - [E(Y)]^2 =$$

$$= \underbrace{E(X^2) - [E(X)]^2}_{} + \underbrace{E(Y^2) - [E(Y)]^2}_{} + 2\underbrace{[E(X\cdot Y) - E(X)\cdot E(Y)]}_{}$$

$$= V(X) + V(Y)$$

0, SINCE $E(X\cdot Y) = E(X) E(Y)$

EXAMPLE: TWO DICE $X_1(ij) = i$; $X_2(ij) = j$

$X = X_1 + X_2$ INDEPENDENT !

$$V(X) = V(X_1 + X_2) = V(X_1) + V(X_2)$$

$$V(X_1) = E(X_1^2) - [E(X_1)]^2 = [1^2 + 2^2 + \dots + 6^2]\cdot\frac{1}{6} - \left[(1+2+\dots+6)\cdot\frac{1}{6}\right]^2 =$$

$= \frac{35}{12} = V(X_2)$. THEREFORE $V(X) = \frac{35}{12} + \frac{35}{12} = \frac{35}{6}$

$$\sigma(X) = \sqrt{35/6}$$

AVERAGE-CASE COMPUTATIONAL COMPLEXITY

$$\xrightarrow[s\in S]{\text{INPUT}} \boxed{\mathfrak{A}}$$

$X(s)$ = THE NUMBER OF OPERATIONS USED BY $\mathfrak{A}$ ON AN INPUT s

1. ASSIGN A PROVABILITY $p(s)$ TO EACH POSSIBLE INPUT VALUE s // USUALLY THE MOST CONTROVERSIAL PART !

2. $E(X)$ IS THE AVERAGE-CASE COMPLEXITY

EXAMPLE. THE AVERAGE CASE COMPLEXITY OF THE LINEAR SEARCH ALGORITHM.

n NUMBERS, A SAMPLE NUMBER x

$2i+1$ COMPARISONS IF x is i-th in the list

p = PROBABILITY THAT x IS IN THE LIST

$q = 1-p$ — " — " — NOT — " —

p/n = PROBABILITY THAT x IS i-th IN THE LIST

$$E = 3\cdot p/n + 5\cdot p/n + \cdots + (2n+1)\cdot p/n + (2n+2)\cdot q =$$

$$= \frac{p}{n}(3+5+\cdots+2n+1) + (2n+2)q = \frac{p}{n}\left((n+1)^2-1\right) + (2n+2)\cdot q =$$

$$= p(n+2) + (2n+2)\cdot q.$$ IF $p=1$, $E = n+2$ (VS. $2n+1$ IN THE WORST-CASE).