5.1 1/2 CH 5 JOINT PROBABILITY DISTRIBUTIONS (FOCUS DISCRETE / SKIP CONTINUOUS)

5.1 TWO DISCRETE RVs.

RECALL FOR A DISCRETE RV X, THE pmf IS $f_X(x) = P_r(X=x)$.

CONSIDER 2 RVs, X, Y, AND LOOK AT THEIR JOINT pmf,

$$f_{XY}(x,y) = P_r(X=x \overset{\cap}{\text{AND}} Y=y) = P_r(X=x, Y=y).$$

(READ AS AN "AND")

EX SUPPOSE YOU TOSS A COIN 3 TIMES

$S = \{HHH, TTH, HHT, THT, HTH, HTT, THH, TTT\}$

$X = \begin{cases} 1, & \text{IF LAST ONE IS } H \\ 0, & \text{OTHERWISE} \end{cases}$

Y = NUMBER OF H^s IN 3 TOSSES $= \{0,1,2,3\}$

FIND THE JOINT MASS FUNCTION OF X AND Y.

$X \sim \text{Bern}(\frac{1}{2}), \quad Y \sim \text{Bin}(3, .5)$

$P_r(X=0, Y=0)$ = PROBABILITY TAILS ON LAST ONE, AND NO HEADS AT ALL (3 TAILS) $= \frac{1}{8}$

$P_r(X=0, Y=1)$ = TAILS LAST, 1 HEAD (2 TAILS) $= \frac{2}{8}$

$P_r(X=0, Y=2)$ = TAILS LAST, 2 HEADS (1 TAIL) $= \frac{1}{8}$

$P_r(X=0, Y=3)$ = TAILS LAST, 3 HEADS $= 0$

$f_{XY}(x,y)$:

x \ Y	0	1	2	3
0	1/8	2/8	1/8	0
1	0	1/8	2/8	1/8

OR

x	Y	$f_{XY}(x,y) = P_r(X=x, Y=y)$
0	0	1/8
0	1	2/8
0	2	1/8
0	3	0
1	0	0
1	1	1/8
1	2	2/8
1	3	1/8
		= 1

$f_{XY}(x,y) \geq 0$

$\sum_x \sum_y f_{XY}(x,y) = 1$

2/2

MARGINAL PROBABILITY DISTRIBUTIONS

X \ Y	0	1	2	3	$f_X(x)$
0	1/8	2/8	1/8	0	1/2
1	0	1/8	2/8	1/8	1/2
$f_Y(y)$	1/8	3/8	3/8	1/8	1

SUM ROWS AND COLUMNS

$Pr(Y=1) = \frac{3}{8}$ = MARGINAL PROBABILITY

FROM MARGINAL PROBABILITY MASS FUNCTIONS

$$f_X(x) = Pr(X=x) = \sum_y f_{XY}(x,y)$$
$$f_Y(y) = Pr(Y=y) = \sum_x f_{XY}(x,y)$$

(SKIPPING CONDITIONAL DISTRIBUTIONS)

5.5 1/4

5.5 Covariance and Correlation

Ex Suppose the joint pmf of RV's X, Y is

X	Y	$f_{XY}(x,y)$
-1	1	.06
-1	2	.14
0	1	.18
0	2	.42
1	1	.06
1	2	.14

$f_{XY}(x,y)$	$Y=1$	$Y=2$	$f_X(x)$
X = -1	.06	.14	.2
0	.18	.42	.6
1	.06	.14	.2
$f_Y(y)$	.3	.7	1

Find $E(X \cdot Y)$, $E(X)$, $E(Y)$, $Var(X)$, $Var(Y)$, $Cov(X,Y)$.

μ_{XY} a) $E(XY) = \sum_x \sum_y xy f_{XY}(x,y)$

$= (-1)(1)(.06) + (-1)(2)(.14) + (0)(1)(.18) + (0)(2)(.4) + (1)(1)(.06) + (1)(2)(.14) = 0.$

μ_X b) $E(X) = \sum_x \sum_y x f_{XY}(x,y) = \sum_x x f_X(x) =$

$= (-1)(.2) + (0)(.6) + (1)(.2) = 0$

μ_Y c) $E(Y) = \sum_x \sum_y y f_{XY}(x,y) = \sum_y y f_Y(y)$

$= (1)(.3) + (2)(.7) = 1.7$

σ_X^2 d) $Var(X) = E(X^2) - (E(X))^2$

$E(X^2) = \sum_x x^2 f_X(x) = (-1)^2(.2) + 0^2(.6) + 1^2(.2) = .4$

$Var(X) = E(X^2) - (E(X))^2 = .4 - 0^2 = .4.$

σ_Y^2 e) $Var(Y) = E(Y^2) - (E(Y))^2$

$E(Y^2) = \sum_y y^2 f_Y(y) = 1^2(.3) + 2^2(.7) = 3.1$

$Var(Y) = E(Y^2) - (E(Y))^2 = 3.1 - (1.7)^2 = 0.21$

σ_{XY} f) Covariance of XY $= Cov(X,Y) = E((X-\mu_X)(Y-\mu_Y))$ where $\mu_X = E(X)$, $\mu_Y = E(Y)$

$= E(XY) - E(X)E(Y)$ {Note: $Var(X) = Cov(X,X)$}.

$= 0 - 0(1.7) = 0.$ So X and Y not linearly related.

2/4

CORRELATION COEFFICIENT ρ

$$\rho = \frac{Cov(X,Y)}{\sqrt{VAR(X)\,VAR(Y)}} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

MEASURE OF LINEAR RELATIONSHIP BETWEEN TWO VARIABLES.

EX LET XY HAVE PMF $f_{XY}(x,y)$ GIVEN BY

x	y	$f_{XY}(x,y)$
-1	-2	1/8
0	-1	1/4
0	1	1/2
1	2	1/8
		1

FIND CORRELATION COEFFICIENT OF XY.

NEED: $f_X(x), f_Y(y), E(X), E(Y), E(XY)$
$E(X^2), E(Y^2) \Rightarrow Cov(X,Y), Var(X), Var(Y)$.

MARGINALS

1)

x	$f_X(x)$
-1	1/8
0	3/4
1	1/8

y	$f_Y(y)$
-2	1/8
-1	1/4
1	1/2
2	1/8

2) EXPECTATIONS

$\mu_X = E(X) = \sum_x x f_X(x) = (-1)(1/8) + 0(3/4) + 1(1/8) = 0$

$\mu_Y = E(Y) = \sum_y y f_Y(y) = (-2)1/8 + (-1)(1/4) + 1(1/2) + 2(1/8) = 1/4$

$= -1/4 - \frac{1}{4} + 1/2 + 1/4 = 1/4.$

$\mu_{XY} = E(XY) = \sum_x \sum_y xy f_{XY}(x,y) = (-1)(-2)1/8 + (0)(-1)1/4 + (0)(1)(1/2) + (1)(2)1/8$

$= \frac{1}{4} + 0 + 0 + \frac{1}{4} = \frac{1}{2}.$

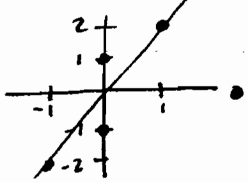
$Cov(X,Y) = E(XY) - E(X)E(Y) = \frac{1}{2} - 0(\frac{1}{4}) = \frac{1}{2}.$

$E(X^2) = \sum_x x^2 f_X(x) = (-1)^2(1/8) + 0^2(3/4) + 1^2(1/8) = 1/4$

$E(Y^2) = \sum_y y^2 f_Y(y) = (-2)^2(1/8) + (-1)^2(1/4) + 1^2(1/2) + 2^2(1/8) = \frac{1}{2} + \frac{1}{4} + \frac{1}{2} + \frac{1}{2} = \frac{7}{4}$

$Var(X) = E(X^2) - (E(X))^2 = \frac{1}{4} - 0^2 = \frac{1}{4}$

$Var(Y) = E(Y^2) - (E(Y))^2 = \frac{7}{4} - (\frac{1}{4})^2 = \frac{27}{16}$



CORR COEF $\rho = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}} = \frac{1/2}{\sqrt{\frac{1}{4}\cdot\frac{27}{16}}} = .7698$

3/4

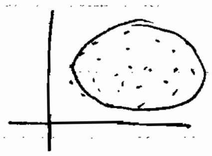
Properties of Correlation Coefficient, ρ

1. Only gives information about <u>linear</u> relationships.

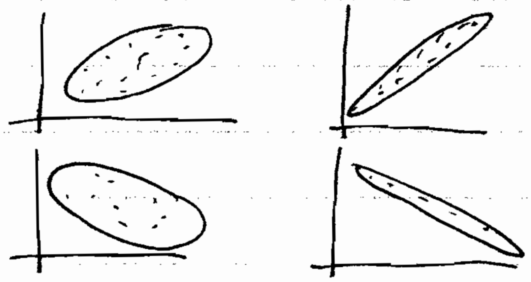
$y = x^2$

$\rho = 0$ even though a perfect relationship, not linear.

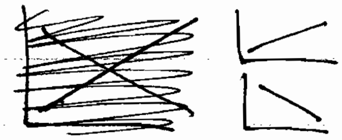
2. $-1 \le \rho \le 1$



$\rho = 0$
No linear relationship



weak relationship — strong



Perfect relationship.
$\rho = \pm 1$

Independence

1. $f_{XY}(x,y) = f_X(x)\, f_Y(y) \iff X, Y$ are independent RVs.

2. If X, Y are independent then $Cov(X,Y) = 0$.

~~[illegible]~~

$Cov(X,Y) = 0$ <u>does not</u> imply that X, Y independent.

<u>Ex</u> Recall

x \ Y	0	1	2	3	
0	1/8	1/4	1/8	0	1/2
1	0	1/8	1/4	1/8	1/2
	1/8	3/8	3/8	1/8	

check if $f_{XY}(0,0) = f_X(0)\, f_Y(0)$

$$1/8 = (1/2)(1/8) = \frac{1}{16} \Rightarrow X, Y \text{ not independent.}$$

4/4

EX

X	Y	$f_{XY}(x,y)$
-2	4	1/5
-1	1	1/5
0	0	1/5
1	1	1/5
2	4	1/5

X	$f_X(x)$
-2	1/5
-1	1/5
0	1/5
1	1/5
2	1/5

Y	$f_Y(y)$
0	1/5
1	2/5
4	2/5

FIND ρ. ARE X, Y INDEPENDENT?

$$\rho = \frac{Cov(X,Y)}{\sqrt{Var(x)Var(Y)}}, \quad Cov(X,Y) = E(XY) - E(X)E(Y).$$

$$E(XY) = (-2)(4)(1/5) + (-1)(1)(1/5) + (0)(0)(1/5) + (1)(1)(1/5) + (2)(4)(1/5).$$
$$= 0$$

$$E(X) = (-2)(1/5) + (-1)(1/5) + (0)(1/5) + (1)(1/5) + (2)(1/5)$$
$$= 0$$

$$E(Y) = (0)(1/5) + (1)(2/5) + (4)(2/5)$$
$$= 2$$

$$Cov(X,Y) = 0 - (0)(2) = 0 \Rightarrow \rho = 0$$

NO LINEAR RELATIONSHIP.

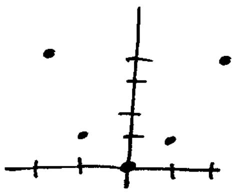
ARE X, Y INDEP?

$f_{XY}(x,y) = f_X(x) f_Y(y)$? FOR EVERY X, Y PAIR?

$$f_{XY}(2,4) = 1/5 \neq f_X(2) f_Y(4) = (1/5)(2/5) = 2/25$$

NOT EQUAL, SO NOT INDEPENDENT.

DEFINITE RELATIONSHIP, EVEN THOUGH $\rho = 0$.



IF $Cov(X,Y) \neq 0 \Rightarrow$ X, Y NOT INDEPENDENT

(CONTRAPOSITIVE OF)

IF X, Y INDEPENDENT $\Rightarrow Cov(X,Y) = 0$.

5.7 1/4

5.7 Linear Combinations of RV^s (Sums)

Typically we look at a collection of RV^s $\{X_1, X_2, \ldots, X_n\}$. If they are independent then

$$f_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = f_{X_1}(x_1)\, f_{X_2}(x_2) \cdots f_{X_n}(x_n).$$ useful.

Recall: For X a RV and c a constant,

$$E(cX) = c\,E(X)$$
$$E(X+c) = E(X)+c$$
$$Var(cX) = c^2\,Var(X)$$
$$Var(X+c) = Var(X)$$

The linear combination of 2 RV^s X, Y is $aX+bY$, a,b are real.

Suppose X,Y discrete RV^s with joint PMF $f_{XY}(x,y)$.

Find mean and variance of $aX+bY$.

$$
\begin{aligned}
E(aX+bY) &= \sum_x \sum_y (aX+bY)\, f_{XY}(x,y) = \sum_x \sum_y aX\, f_{XY}(x,y) + \sum_x \sum_y bY\, f_{XY}(x,y) \\
&= a \sum_x \sum_y x f_{XY}(x,y) + b \sum_x \sum_y y\, f_{XY}(x,y) \\
&= a \sum_x x \underbrace{\sum_y f_{XY}(x,y)}_{f_X(x)} + b \sum_y y \underbrace{\sum_x f_{XY}(x,y)}_{f_Y(y)} \\
&= a \sum_x x\, f_X(x) + b \sum_y y\, f_Y(y) \\
&= a\,E(X) + b\,E(Y). \\
&= a\mu_X + b\mu_Y.
\end{aligned}
$$

2/4

$$\text{Var}(aX+bY) = E((aX+bY)^2) - (E(aX+bY))^2$$
$$= E((aX+bY)^2) - (a\mu_x + b\mu_y)^2$$
$$= E((aX)^2 + 2(aX)(bY) + (bY)^2) - ((a\mu_x)^2 + 2(a\mu_x)(b\mu_y) + (b\mu_y)^2)$$
$$= E(a^2X^2) - a^2\mu_x^2 + E(b^2Y^2) - b^2\mu_y^2 + E(2aXbY) - 2a\mu_x b\mu_y$$
$$= a^2(E(X^2) - \mu_x^2) + b^2(E(Y^2) - \mu_y^2) + 2ab(E(XY) - \mu_x\mu_y)$$
$$= a^2\,\text{Var}(X) + b^2\,\text{Var}(Y) + 2ab\,\text{Cov}(XY).$$

$$E(aX+bY) = aE(X) + bE(Y)$$
$$\text{Var}(aX+bY) = a^2\,\text{Var}(X) + b^2\,\text{Var}(Y) + 2ab\,\text{Cov}(XY).$$

NOTE: IF X, Y INDEPENDENT, $\text{Cov}(X,Y) = 0$ SO $\text{Var}(aX+bY) = a^2\text{Var}(X) + b^2\text{Var}(Y)$.

Ex SUPPOSE JOINT PMF OF X, Y IS

		Y = −1	Y = 0	
X	−1	.2	.3	.5
	0	.3	.2	.5
		.5	.5	

FIND $E(3X+2Y)$, $\text{Var}(3X+2Y)$.

$$E(X) = (-1)(.5) + (0)(.5) = -.5$$
$$E(Y) = (-1)(.5) + (0)(.5) = -.5$$
$$E(X^2) = (-1)^2(.5) + 0^2(.5) = .5$$
$$E(Y^2) = (-1)^2(.5) + 0^2(.5) = .5$$
$$\text{Var}(X) = .5 - (-.5)^2 = .25$$
$$\text{Var}(Y) = .5 - (-.5)^2 = .25$$
$$E(XY) = (-1)(-1)(.2) + (-1)(0)(.3) + (0)(-1)(.3) + (0)(0)(.2) = .2$$
$$\text{Cov}(X,Y) = .2 - (-.5)(-.5) = .2 - .25 = -.05.$$

$$E(3X+2Y) = 3E(X) + 2E(Y) = 3(-.5) + 2(-.5) = -2.5$$
$$\text{Var}(3X+2Y) = 3^2\text{Var}(X) + 2^2\text{Var}(Y) + 2(3)(2)\text{Cov}(X,Y) = 9(.25) + 4(.25) + 12(-.05)$$
$$= 2.65$$

3/4 THE LINEAR COMBINATION OF RVS $X_1, X_2, \ldots, X_n$ IS
$a_1 X_1 + a_2 X_2 + \cdots + a_n X_n$, WHERE $a_1, a_2, \ldots, a_n$ ARE REAL NUMBERS.

$$E(a_1 X_1 + a_2 X_2 + \cdots + a_n X_n) = a_1 E(X_1) + a_2 E(X_2) + \cdots + a_n E(X_n).$$

$$\begin{aligned} VAR(a_1 X_1 + a_2 X_2 + \cdots + a_n X_n) &= a_1^2 VAR(X_1) + a_2^2 VAR(X_2) + \cdots + a_n^2 VAR(X_n) \\ &\quad + 2a_1 a_2 COV(X_1, X_2) + 2a_1 a_3 COV(X_1, X_3) + \cdots \\ &\quad + 2a_2 a_3 COV(X_2, X_3) + 2a_2 a_4 COV(X_2, X_4) + \cdots + 2a_{n-1} a_n COV(X_{n-1}, X_n). \end{aligned}$$

HAPPILY, IF $X_1, X_2, \ldots, X_n$ ARE INDEPENDENT,

$$VAR(a_1 X_1 + a_2 X_2 + \cdots + a_n X_n) = a_1^2 VAR(X_1) + a_2^2 VAR(X_2) + \cdots + a_n^2 VAR(X_n).$$

EX SUPPOSE $X_1, X_2, \ldots, X_n$ ARE INDEPENDENT RVS.
FIND MEAN AND VARIANCE OF $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{1}{n}(X_1 + X_2 + \cdots + X_n)$

AND SUPPOSE $E(X_1) = \cdots = E(X_n) = \mu$
AND $VAR(X_1) = \cdots = VAR(X_n) = \sigma^2$

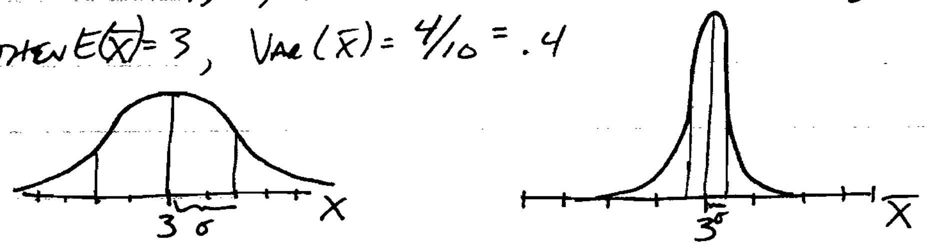
$$\begin{aligned} E(\bar{X}) &= E\left(\tfrac{1}{n}X_1 + \tfrac{1}{n}X_2 + \cdots + \tfrac{1}{n}X_n\right) \\ &= \tfrac{1}{n}E(X_1) + \tfrac{1}{n}E(X_2) + \cdots + \tfrac{1}{n}E(X_n) \\ &= \tfrac{1}{n}\mu + \tfrac{1}{n}\mu + \cdots + \tfrac{1}{n}\mu \\ &= n\left(\tfrac{1}{n}\mu\right) = \mu. \end{aligned}$$

$$\begin{aligned} VAR(\bar{X}) &= VAR\left(\tfrac{1}{n}X_1 + \cdots + \tfrac{1}{n}X_n\right) \\ &= \left(\tfrac{1}{n}\right)^2 VAR(X_1) + \cdots + \left(\tfrac{1}{n}\right)^2 VAR(X_n) \\ &= \left(\tfrac{1}{n}\right)^2 \sigma^2 + \cdots + \left(\tfrac{1}{n}\right)^2 \sigma^2 \\ &= n\left(\left(\tfrac{1}{n}\right)^2 \sigma^2\right) = \tfrac{1}{n}\sigma^2 \end{aligned}$$

IF $X_1, X_2, \ldots, X_n$ ARE INDEPENDENT WITH MEAN μ AND VARIANCE σ^2,
THEN $E(\bar{X}) = \mu$ AND $VAR(\bar{X}) = \sigma^2/n$.

4/4 EX SUPPOSE $X_1, \ldots, X_{10}$ ARE EACH $\sim N(3,4) \Rightarrow \mu=3,\ \sigma^2=4$.

THEN $E(\bar{X})=3$, $Var(\bar{X}) = 4/10 = .4$



AVERAGES VARY LESS THAN INDIVIDUAL OBSERVATIONS.

IF $X_1, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$ THEN $\bar{X} \sim N(\mu, \sigma^2/n)$

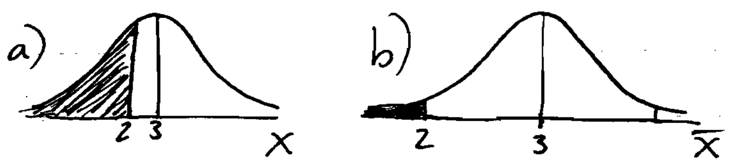
IID = INDEPENDENT AND IDENTICALLY DISTRIBUTED

EX $X_1, \ldots, X_{10} \overset{iid}{\sim} N(3,4)$ THEN $\bar{X} \sim N(3, 4/10) = N(3, .4)$.

FIND a) $Pr(X<2)$ AND b) $Pr(\bar{X}<2)$.

a) $Pr(X<2) = Pr\left(\frac{X-\mu}{\sigma} < \frac{2-3}{2}\right) = Pr(Z < -.5) = .308538$

b) $Pr(\bar{X}<2) = Pr\left(\frac{\bar{X}-\mu}{\sigma} < \frac{2-3}{\sqrt{.4}}\right) = Pr(Z < -1.58) = .057053$



Exam 2 1/1

Exam 2 topics

Ch 3 Discrete Distributions: Mean, Variance, Probabilities (PMF)

- Discrete Uniform
- Bernoulli
- Binomial ☆ ☆ = Recognize when to use
- Geometric ☆
- Negative Binomial ☆
- Hypergeometric ☆
- Poisson

Ch 4 Continuous Distributions: Mean, Variance, Probabilities (PDF), CDF.

- Continuous Uniform
- Exponential
- Normal ** use tables (provided)

Find $E(x)$, $Var(x)$, $E(h(x))$, cdfs, Probabilities for General Continuous Distributions

Ch 5 Joint Distributions of Discrete RVs.

Find marginals $(f_X(x), f_Y(y))$, means, variances, $Covar(X,Y)$, Probs.

Independence $\iff f_{XY}(x,y) = f_X(x) f_Y(y)$.

$\implies Cov(X,Y) = 0$.

Correlation measures Linear Association between X and Y.

Linear Combinations of RVs: $a_1 X_1 + a_2 X_2 + \cdots + a_n X_n$

Find $E(\cdot)$, $Var(\cdot)$ when NOT independent, and when independent.

Mean, Var of $\bar{X}$ if $X_1, X_2, \ldots, X_n$ are IID, ie, $E(X_i) = \mu$, $Var(X_i) = \sigma^2$ for all i.

Mean, Var and Distribution of $\bar{X}$ if $X_1, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$.