§ 4.3, 4.4 Joint distributions, independence

Def Let X and Y be random variables with respective ranges S and T. Then their joint probability mass function is given by $m(x,y) = P(X = x, Y = y)$ for each $x \in S$, $y \in T$.

Example Suppose $m(x,y) = P(X = x, Y = y) = cxy$ where $x = 1, 2$ and $y = 1, 2, 3$. Represent m in a table, find c, and find the (marginal) probability mass functions for X and Y.

	$y=1$	$y=2$	$y=3$	row sums
$x=1$	c	$2c$	$3c$	$6c$
$x=2$	$2c$	$4c$	$6c$	$12c$
col. sums	$3c$	$6c$	$9c$	

"marginal" probabilities of X

"marginal" probabilities of Y

table of $m(x,y)$ values

Note that $1 = \sum_{x=1}^{2} \sum_{y=1}^{3} P(X = x, Y = y) = 18c$, so $c = \frac{1}{18}$.

And $P(X = 1) = \sum_{y=1}^{3} P(X = 1, Y = y)$ by Law of Total Prob.

$$= c + 2c + 3c$$

$$= 6c = \frac{1}{3}$$

Similarly, $P(X=2) = 12c = \frac{2}{3}$.

The marginal probability mass function of X is

$$P(X=k) = \begin{cases} 1/3 & k=1 \\ 2/3 & k=2 \end{cases}$$

and the marginal prob. mass function of Y is

$$P(Y=k) = \begin{cases} 3/18 & k=1 \\ 6/18 & k=2 \\ 9/18 & k=3 \end{cases}$$

Def If X and Y have joint pmf with ranges S and T respectively, their marginal pmf's are

$$P(X=k) = \sum_{y \in T} P(X=k, Y=y) \quad \text{for each } k \in S$$

and

$$P(Y=k) = \sum_{x \in S} P(X=x, Y=k) \quad \text{for each } k \in T.$$

Example A bag contains 4 red, 2 blue, 3 white balls and we draw a sample of size 2 without replacement.

Let R be the number of red balls in the draw, and B the number of blue balls. Find the joint pmf of R and B, their marginal pmfs, $E[RB]$, and $E[R]$ and $E[B]$.

$$P(R=r, B=b) = \frac{\binom{4}{r}\binom{2}{b}\binom{3}{2-(r+b)}}{\binom{9}{2}} \quad \text{where } 0 \le r+b \le 2$$

	$b=0$	$b=1$	$b=2$
$r=0$	$3/36$	$6/36$	$1/36$
$r=1$	$12/36$	$8/36$	0
$r=2$	$6/36$	0	0

$$P(R=r) = \begin{cases} 10/36 & r=0 \\ 20/36 & r=1 \\ 6/36 & r=2 \end{cases} \qquad P(B=b) = \begin{cases} 21/36 & b=0 \\ 14/36 & b=1 \\ 1/36 & b=2 \end{cases}$$

$$E[R+B] = \sum_{r=0}^{2} \sum_{b=0}^{2} (r+b)\, P(R=r, B=b)$$

$$= (0+0)\,P(R=0, B=0) + (0+1)\,P(R=0, B=1)$$

$$+ (0+2)\,P(R=0, B=2) + (1+0)\,P(R=1, B=0)$$

$$+ (1+1)\,P(R=1, B=1) + (2+0)\,P(R=2, B=0)$$

$$= 6/36 + 2(1/36) + 12/36 + 2(8/36) + 2(6/36)$$

$$= 48/36 = 4/3$$

$$E[RB] = \sum_{r=0}^{2} \sum_{b=0}^{2} (rb)\, P(R=r, B=b)$$

$$= \frac{8}{36}$$

$$E[R] = \sum_{r=0}^{2} r\, P(R=r) = \frac{20}{36} + 2\left(\frac{6}{36}\right) = \frac{32}{36} = \frac{8}{9}$$

$$E[B] = \sum_{b=0}^{2} b\, P(B=b) = \frac{14}{36} + 2\left(\frac{1}{36}\right) = \frac{16}{36} = \frac{4}{9}$$

Notice $E[R+B] = E[R] + E[B]$ but $E[RB] \neq E[R]\,E[B]$

Theorem For any random variables X, Y with ranges S, T respectively $E[g(X,Y)] = \sum_{x \in S} \sum_{y \in T} g(x,y) P(X=x, Y=y)$

Theorem If X and Y are independent, then

$$E[XY] = E[X]E[Y]$$

Theorem For any random variables $E[X+Y] = E[X] + E[Y]$.

Proof to come.

Problem 1. Suppose you have the following data on pet ownership in a town with 1000 households.

	Has 0 cats	Has 1 cat	Has 2 cats	total
Has 0 dogs	400	150	150	700
Has 1 dog	50	5	40	95
Has 2 dogs	150	15	40	205
total	600	170	230	1000

Suppose a household is chosen at random, and we let X denote the number of dogs in the household and let Y denote the number of cats in the household.

a. Make a table for the joint probability mass function of X and Y.

b. Find the marginal probability mass functions of X and Y.

c. Find $E[X]$ and $E[Y]$.

(a)

	$y=0$	$y=1$	$y=2$	totals
$x=0$	0.4	0.15	0.15	0.7
$x=1$	0.05	0.005	0.04	0.095
$x=2$	0.15	0.015	0.04	0.205

(b)

$$P(X=x) = \begin{cases} 0.7 & x=0 \\ 0.095 & x=1 \\ 0.205 & x=2 \end{cases}, \quad P(Y=y) = \begin{cases} 0.6 & y=0 \\ 0.17 & y=1 \\ 0.23 & y=2 \end{cases}$$

(c)

$$E[X] = 0.095 + 2(0.205) = 0.505$$

$$E[Y] = 0.17 + 2(0.23) = 0.63$$

Problem 2. Suppose we draw two numbers, one at a time without replacement, from the set $\{1, 2, 3, 4\}$. Let X denote the first number drawn, and let Y denote the second number drawn.

a. For each value of x and y, compute $P(X = x, Y = y)$ by computing $P(Y = y \mid X = x)P(X = x)$. Write your answers in a table.

b. Find the marginal probability mass functions of X and Y.

c. Find $E[X]$ and $E[Y]$.

d. Compute $E[XY]$. How does it compare to $E[X]E[Y]$? Is this surprising?

e. How does $E[X + Y]$ compare to $E[X] + E[Y]$?

(a)

	1	2	3	4
1	0	1/12	1/12	1/12
2	1/12	0	1/12	1/12
3	1/12	1/12	0	1/12
4	1/12	1/12	1/12	0

(b) $P(X = x) = \frac{1}{4}$, $x = 1, 2, 3, 4$

$P(Y = y) = \frac{1}{4}$, $y = 1, 2, 3, 4$

(c) $E[X] = 2.5 = E[Y]$

(d) $E[XY] = \sum_{x=1}^{4} \sum_{y=1}^{4} xy \, P(X = x, Y = y)$

$= \frac{1}{12}(2 + 3 + 4$

$+ 2 + 6 + 8$

$+ 3 + 6 + 12$

$+ 4 + 8 + 12)$

$= \frac{1}{12}(9 + 16 + 21 + 24)$

$= \frac{70}{12} = 5.833..$

$E[XY] \neq E[X]E[Y]$ which is not surprising since X and Y are not independent

(e) $E[X + Y] = E[X] + E[Y]$ regardless of lack of independence.