Math 3070, Applied Statistics

Section 1

October 2, 2019

Lecture Outline, 10/2

Section 5.1

- Joint Probability Mass Functions
- Marginal PMFs
- Independence
- Conditional PMFs

Joint Probability Mass Function

Let X and Y be discrete random variables. Their **joint probability mass function** (joint pmf) is

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

Need $f(x, y) \geq 0$ and $\sum_{x,y} f(x, y) = 1$

Example: An insurance agency serves both an automobile policy and a homeowner's policy. The possible amounts for a deductible are \$100 and \$250 for an automobile policy, and \$0, \$100, and \$200 for a homeowner's policy. If a customer is selected at random, and X is their auto deductible and Y the deductible on the homeowner's policy, then the joint pmf of X and Y is:

$f(x, y)$	$y = 0$	$y = 100$	$y = 200$
$x = 100$	.20	.10	.20
$x = 250$	.05	.15	.30

Example, Joint PMF

Suppose we roll two fair 6-sided dice, and let X be the smaller of the two rolls, and let Y be the larger. Find the joint probability mass function $f(x,y)$ of X and Y.

$f(x,y)$	$y=1$	$y=2$	$y=3$	$y=4$	$y=5$	$y=6$
$x=1$	$1/36$	$2/36$	$2/36$	$2/36$	$2/36$	$2/36$
$x=2$	0	$1/36$	$2/36$	$2/36$	$2/36$	$2/36$
$x=3$	0	0	$1/36$	$2/36$	$2/36$	$2/36$
$x=4$	0	0	0	$1/36$	$2/36$	$2/36$
$x=5$	0	0	0	0	$1/36$	$2/36$
$x=6$	0	0	0	0	0	$1/36$

Note: $f(2,3)=2/36$, but $f(3,2)=0$.

Marginal Probability Mass Function

Let X and Y be discrete random variables with joint probability mass function $f(x, y)$. The **marginal probability mass function** of X is defined as

$$f_X(x) = \sum_y f(x, y)$$

where the sum is taken over all possible values of Y. Similarly, the marginal probability mass function of Y is

$$f_Y(y) = \sum_x f(x, y)$$

where the sum is taken over all possible values of X.

Note: f_X is simply the probability mass function of X considered as a random variable on its own: $f_X(x) = P(X = x)$.

Example, Marginal PMF

Recall the joint pmf for the auto insurance deductible X and homeowner's insurance deductible Y from earlier:

$f(x,y)$	$y=0$	$y=100$	$y=200$
$x=100$	.20	.10	.20
$x=250$	.05	.15	.30

The marginal pmf for the auto insurance deductible X is given by

$$f_X(100) = .50$$
$$f_X(250) = .50$$

The marginal pmf for the homeowner's deductible Y is given by

$$f_Y(0) = .25$$
$$f_Y(100) = .25$$
$$f_Y(200) = .50$$

Independence of Random Variables

Discrete random variables X and Y are **independent** if their joint probability mass function $f(x, y)$ is the product of the marginal probability mass functions $f_X(x)$ and $f_Y(y)$:

$$f(x, y) = f_X(x)f_Y(y)$$

Their events are also independent,
$P(a < X < b \cap c < Y < d) = P(a < X < b)P(c < Y < d)$.

In the insurance example,

$f(x, y)$	$y = 0$	$y = 100$	$y = 200$
$x = 100$	.20	.10	.20
$x = 250$	.05	.15	.30

the random variables X and Y are dependent, because, e.g.,

$$f(100, 0) = .20 \neq (.50)(.25) = f_X(100)f_Y(0)$$

Example, Independence

If X and Y are independent geometric random variables with parameter p and q respectively, find the joint pmf of X and Y.

Solution: The marginal pmf of X is

$$f_X(x) = p(1-p)^x$$

while the marginal pmf of Y is

$$f_Y(y) = q(1-q)^y$$

Since X and Y are independent, their joint pmf is the product of their marginal pmfs:

$$f(x, y) = f_X(x) f_Y(y) = p(1-p)^x q(1-q)^y$$

Conditional Probability Mass Function

Let X and Y be random variables with joint pmf $f(x, y)$. The **conditional probability mass function** of Y given $X = x$ is

$$f_{Y|X}(y \mid x) = \frac{f(x, y)}{f_X(x)}$$

The conditional pmf may be expressed in terms of conditional probabilities:

$$f_{Y|X}(y \mid x) = P(Y = y \mid X = x)$$

If X and Y are independent, then the conditional pmf of Y given X is simply the marginal pmf of Y:

$$f_{Y|X}(y \mid x) = \frac{f(x, y)}{f_X(x)} = \frac{f_X(x) f_Y(y)}{f_X(x)} = f_Y(y)$$

Example, Conditional PMF

In the insurance example, the auto deductible X and homeowner's deductible Y had joint pmf given by

$f(x,y)$	$y=0$	$y=100$	$y=200$
$x=100$	.20	.10	.20
$x=250$	.05	.15	.30

Find the conditional pmf of homeowner's deductible Y, given the auto deductible X is \$100:

Solution:

$$f_{Y|X}(0 \mid 100) = \frac{f(100,0)}{f_X(100)} = \frac{.20}{.50} = .40$$
$$f_{Y|X}(100 \mid 100) = \frac{f(100,100)}{f_X(100)} = \frac{.10}{.50} = .20$$
$$f_{Y|X}(200 \mid 100) = \frac{f(100,200)}{f_X(100)} = \frac{.20}{.50} = .40$$

PMF of Several Random Variables

If $X_1, \ldots, X_n$ are random variables, their joint pmf is defined as

$$f(x_1, \ldots, x_n) = P(X_1 = x_1, \ldots, X_n = x_n)$$

The marginal pmf of X_1 is defined as

$$f_{X_1}(x) = \sum_{x_2} \sum_{x_3} \cdots \sum_{x_n} P(X_1 = x, X_2 = x_2, \ldots, X_n = x_n)$$

The random variables $X_1, \ldots, X_n$ are independent if

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

Conditioning can be applied on any set of variables. Questions in this course will not cover conditioning with more than two variables.

Joint PMF, Summary

- Joint PMF: $f(x, y) = P(X = x \cap Y = y)$
- Marginal PDF: $f_X(x) = \sum_y f(x, y)$. Is a univariate PMF.
- If X and Y are independent, $f(x, y) = f_X(x) f_Y(y)$.
  Their events are also independent
  $P(a < X < b \cap c < Y < d) = P(a < X < b)P(c < Y < d)$
- Conditional PMF:

  $$f_{Y|X}(y \mid x) = \frac{f(x, y)}{f_X(x)}$$

  Note: the conditioned varible X is fixed, $X = x$. Think of this as a function of y.
- Can define these with more variables than two.

Example, Independence?

There are two cashiers. The number of customers that visit each is modeled by a Poisson distribution with a mean of 5. The probability that each cashier sees one customer in an hour less than 85%. Is the number of customers at one cashier independent from the other?

$X =$ number customer at one cashier. $X \sim Poisson(5)$

$Y =$ number of customers at the other. $Y \sim Poisson(5)$

Assume independence $f(x,y) = f_X(x)f_Y(y)$ and see if there is a contradiction

$$\begin{aligned} P(X > 0 \cap Y > 0) &= P(X > 0)P(Y > 0) \\ &= (1 - P(X = 0))(1 - P(Y = 0)) \\ &= (1 - 5^0 e^{-5}/0!)(1 - 5^0 e^{-5}/0!) \\ &= (1 - e^{-5})^2 \approx 0.98656950593 \end{aligned}$$

No, independence assumption yields a probability that is too high.

Example, Reconstructing from Marginals (for fun)

Suppose that X and Y are Bernoulli random variables with the following marginal distributions. Can one calculate the joint PMF?

$$f_X(x) = \begin{cases} 0.25, & x = 1 \\ 0.75, & x = 0 \end{cases} \quad f_Y(y) = \begin{cases} 0.75, & y = 1 \\ 0.25, & y = 0 \end{cases}$$

$f(x, y)$	$y = 0$	$y = 1$	$f_X(x)$
$x = 0$	$a = P(X = 0 \cap Y = 0)$	$b = P(X = 0 \cap Y = 1)$	0.75
$x = 1$	$c = P(X = 1 \cap Y = 0)$	$d = P(X = 1 \cap Y = 1)$	0.25
$f_Y(y)$	0.25	0.75	

$$a + b = 0.75 \qquad c + d = 0.25$$

$$a + c = 0.25 \qquad b + d = 0.75$$

Example, Reconstructing from Marginals (for fun)

Suppose that X and Y are Bernoulli random variables with the following marginal distributions. Can one calculate the joint PMF?

$$f_X(x) = \begin{cases} 0.25, & x = 1 \\ 0.75, & x = 0 \end{cases} \quad f_Y(y) = \begin{cases} 0.75, & y = 1 \\ 0.25, & y = 0 \end{cases}$$

$$M \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} = \begin{bmatrix} 0.75 \\ 0.25 \\ 0.75 \\ 0.25 \end{bmatrix}$$

$det(M) = 0 \implies$ no unique solution.
The PMF cannot be determined.