5 Joint Probability Distributions and Random Samples

5.1 Jointly Distributed Random Variables

Two Discrete Random Variables

Two Discrete Random Variables

The probability mass function (pmf) of a single discrete rv X specifies how much probability mass is placed on each possible X value.

The joint pmf of two discrete rv's X and Y describes how much probability mass is placed on each possible pair of values (x, y).

Definition
Let X and Y be two discrete rv's defined on the sample space $\mathcal{S}$ of an experiment. The **joint probability mass function** $p(x, y)$ is defined for each pair of numbers (x, y) by

$$p(x, y) = P(X = x \text{ and } Y = y)$$

Two Discrete Random Variables

It must be the case that $p(x, y) \geq 0$ and $\sum_x \sum_y p(x, y) = 1$.

Now let A be any set consisting of pairs of (x, y) values (e.g., $A = \{(x, y): x + y = 5\}$ or $\{(x, y): \max(x, y) \leq 3\}$).

Then the probability $P[(X, Y) \in A]$ is obtained by summing the joint pmf over pairs in A:

$$P[(X, Y) \in A] = \sum_{(x, y)} \sum_{\in A} p(x, y)$$

5

Example 1

A large insurance agency services a number of customers who have purchased both a homeowner's policy and an automobile policy from the agency. For each type of policy, a deductible amount must be specified.

For an automobile policy, the choices are \$100 and \$250, whereas for a homeowner's policy, the choices are 0, \$100, and \$200.

Suppose an individual with both types of policy is selected at random from the agency's files. Let X = the deductible amount on the auto policy and Y = the deductible amount on the homeowner's policy.

6

Example 1 cont'd

Possible (X, Y) pairs are then (100, 0), (100, 100), (100, 200), (250, 0), (250, 100), and (250, 200); the joint pmf specifies the probability associated with each one of these pairs, with any other pair having probability zero.

Suppose the joint pmf is given in the accompanying **joint probability table:**

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

7

Example 1 cont'd

Then $p(100, 100) = P(X = 100 \text{ and } Y = 100) = P(\$100$ deductible on both policies) = .10.

The probability $P(Y \geq 100)$ is computed by summing probabilities of all (x, y) pairs for which $y \geq 100$:

$$P(Y \geq 100) = p(100, 100) + p(250, 100) + p(100, 200) + p(250, 200)$$

$$= .75$$

8

Two Discrete Random Variables

Once the joint pmf of the two variables X and Y is available, it is in principle straightforward to obtain the distribution of just one of these variables.

As an example, let X and Y be the number of statistics and mathematics courses, respectively, currently being taken by a randomly selected statistics major.

Suppose that we wish the distribution of X, and that when $X = 2$, the only possible values of Y are 0, 1, and 2.

9

Two Discrete Random Variables

Then

$$p_X(2) = P(X = 2) = P[(X, Y) = (2, 0) \text{ or } (2, 1) \text{ or } (2, 2)]$$

$$= p(2, 0) + p(2, 1) + p(2, 2)$$

That is, the joint pmf is summed over all pairs of the form $(2, y)$. More generally, for any possible value x of X, the probability $p_X(x)$ results from holding x fixed and summing the joint pmf $p(x, y)$ over all y for which the pair (x, y) has positive probability mass.

The same strategy applies to obtaining the distribution of Y by itself.

10

Two Discrete Random Variables

Definition

The **marginal probability mass function of *X*,** denoted by $p_X(x)$, is given by

$$p_X(x) = \sum_{y:\, p(x,y)>0} p(x, y) \text{ for each possible value } x$$

Similarly, the **marginal probability mass function of *Y*** is

$$p_Y(y) = \sum_{x:\, p(x,y)>0} p(x, y) \text{ for each possible value } y.$$

11

Example 2

Example 1 continued…

The possible X values are $x = 100$ and $x = 250$, so computing row totals in the joint probability table yields

$$p_X(100) = p(100, 0) + p(100, 100) + p(100, 200) = .50$$

and

$$p_X(250) = p(250, 0) + p(250, 100) + p(250, 200) = .50$$

The marginal pmf of X is then

$$p_X(x) = \begin{cases} .5 & x = 100, 250 \\ 0 & \text{otherwise} \end{cases}$$

12

Example 2

cont'd

Similarly, the marginal pmf of Y is obtained from column totals as

$$p_Y(y) = \begin{cases} .25 & y = 0, 100 \\ .50 & y = 200 \\ 0 & \text{otherwise} \end{cases}$$

so $P(Y \geq 100) = p_Y(100) + p_Y(200) = .75$ as before.

Two Continuous Random Variables

Two Continuous Random Variables

The probability that the observed value of a continuous rv X lies in a one-dimensional set A (such as an interval) is obtained by integrating the pdf $f(x)$ over the set A.

Similarly, the probability that the pair (X, Y) of continuous rv's falls in a two-dimensional set A (such as a rectangle) is obtained by integrating a function called the *joint density function*.

Two Continuous Random Variables

Definition

Let X and Y be continuous rv's. **A joint probability density function** $f(x, y)$ for these two variables is a function satisfying $f(x, y) \geq 0$ and

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)\, dx\, dy = 1.$$

Then for any two-dimensional set A

$$P[(X, Y) \in A] = \iint_A f(x, y)\, dx\, dy$$

Two Continuous Random Variables

In particular, if A is the two-dimensional rectangle $\{(x, y): a \le x \le b,\ c \le y \le d\}$, then

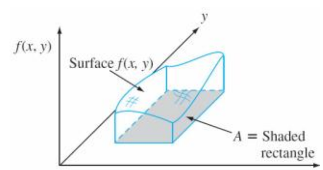
$$P[(X,Y) \in A] = P(a \le X \le b, c \le Y \le d) = \int_a^b \int_c^d f(x, y)\, dy\, dx$$

We can think of $f(x, y)$ as specifying a surface at height $f(x, y)$ above the point (x, y) in a three-dimensional coordinate system.

Then $P[(X, Y) \in A]$ is the volume underneath this surface and above the region A, analogous to the area under a curve in the case of a single rv.

Two Continuous Random Variables

This is illustrated in Figure 5.1.



$P[(X, Y) \in A]$ = volume under density surface above A

Figure 5.1

Example 3

A bank operates both a drive-up facility and a walk-up window. On a randomly selected day, let X = the proportion of time that the drive-up facility is in use (at least one customer is being served or waiting to be served) and Y = the proportion of time that the walk-up window is in use.

Then the set of possible values for (X, Y) is the rectangle $D = \{(x, y): 0 \le x \le 1, 0 \le y \le 1\}$.

Example 3

cont'd

Suppose the joint pdf of (X, Y) is given by

$$f(x, y) = \begin{cases} \frac{6}{5}(x + y^2) & 0 \le x \le 1, 0 \le y \le 1 \\ 0 & \text{otherwise} \end{cases}$$

To verify that this is a legitimate pdf, note that $f(x, y) \ge 0$ and

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)\, dx\, dy = \int_0^1 \int_0^1 \frac{6}{5}(x + y^2)\, dx\, dy$$

$$= \int_0^1 \int_0^1 \frac{6}{5} x\, dx\, dy + \int_0^1 \int_0^1 \frac{6}{5} y^2\, dx\, dy$$

Example 3

cont'd

$$= \int_0^1 \frac{6}{5} x\, dx + \int_0^1 \frac{6}{5} y^2\, dy = \frac{6}{10} + \frac{6}{15} = 1$$

$$= \frac{6}{10} + \frac{6}{15}$$

$$= 1$$

The probability that neither facility is busy more than one-quarter of the time is

$$P\left(0 \le X \le \frac{1}{4}, 0 \le Y \le \frac{1}{4}\right) = \int_0^{1/4} \int_0^{1/4} \frac{6}{5}(x + y^2)\, dx\, dy$$

$$= \frac{6}{5} \int_0^{1/4} \int_0^{1/4} x\, dx\, dy + \frac{6}{5} \int_0^{1/4} \int_0^{1/4} y^2\, dx\, dy$$

21

Example 3

cont'd

$$= \frac{6}{20} \cdot \frac{x^2}{2}\Big|_{x=0}^{x=1/4} + \frac{6}{20} \cdot \frac{y^3}{3}\Big|_{y=0}^{y=1/4}$$

$$= \frac{7}{640}$$

$$= .0109$$

22

Two Continuous Random Variables

The marginal pdf of each variable can be obtained in a manner analogous to what we did in the case of two discrete variables.

The marginal pdf of X at the value x results from holding x fixed in the pair (x, y) and *integrating* the joint pdf over y. Integrating the joint pdf with respect to x gives the marginal pdf of Y.

23

Two Continuous Random Variables

Definition

The **marginal probability density functions** of X and Y, denoted by $f_X(x)$ and $f_Y(y)$, respectively, are given by

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y)\, dy \qquad \text{for } -\infty < x < \infty$$

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y)\, dx \qquad \text{for } -\infty < y < \infty$$

24

Independent Random Variables

Independent Random Variables

In many situations, information about the observed value of one of the two variables X and Y gives information about the value of the other variable.

In Example 1, the marginal probability of X at $x = 250$ was .5, as was the probability that $X = 100$. If, however, we are told that the selected individual had $Y = 0$, then $X = 100$ is four times as likely as $X = 250$.

Thus there is a dependence between the two variables. Earlier, we pointed out that one way of defining independence of two events is via the condition $P(A \cap B) = P(A) \cdot P(B)$.

Independent Random Variables

Here is an analogous definition for the independence of two rv's.

Definition

Two random variables X and Y are said to be **independent** if for every pair of x and y values

$p(x, y) = p_X(x) \cdot p_Y(y)$ when X and Y are discrete

or **(5.1)**

$f(x, y) = f_X(x) \cdot f_Y(y)$ when X and Y are continuous

If (5.1) is not satisfied for all (x, y), then X and Y are said to be **dependent.**

Independent Random Variables

The definition says that two variables are independent if their joint pmf or pdf is the product of the two marginal pmf's or pdf's.

Intuitively, independence says that knowing the value of one of the variables does not provide additional information about what the value of the other variable might be.

Example 6

In the insurance situation of Examples 1 and 2,

$$p(100, 100) = .10 \neq (.5)(.25) = p_X(100) \cdot p_Y(100)$$

so X and Y are not independent.

Independence of X and Y requires that *every* entry in the joint probability table be the product of the corresponding row and column marginal probabilities.

29

Independent Random Variables

Independence of two random variables is most useful when the description of the experiment under study suggests that X and Y have no effect on one another.

Then once the marginal pmf's or pdf's have been specified, the joint pmf or pdf is simply the product of the two marginal functions. It follows that

$$P(a \le X \le b, c \le Y \le d) = P(a \le X \le b) \cdot P(c \le Y \le d)$$

30

More Than Two Random Variables

31

More Than Two Random Variables

To model the joint behavior of more than two random variables, we extend the concept of a joint distribution of two variables.

Definition

If $X_1, X_2, \ldots, X_n$ are all discrete random variables, the joint pmf of the variables is the function

$$p(x_1, x_2, \ldots, x_n) = P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)$$

32

More Than Two Random Variables

If the variables are continuous, the joint pdf of $X_1, \ldots, X_n$ is the function $f(x_1, x_2, \ldots, x_n)$ such that for any n intervals $[a_1, b_1], \ldots, [a_n, b_n]$,

$$P(a_1 \le X_1 \le b_1, \ldots, a_n \le X_n \le b_n) = \int_{a_1}^{b_1} \cdots \int_{a_n}^{b_n} f(x_1, \ldots, x_n)\, dx_n \ldots dx_1$$

In a binomial experiment, each trial could result in one of only two possible outcomes.

Consider now an experiment consisting of n independent and identical trials, in which each trial can result in any one of r possible outcomes.

33

More Than Two Random Variables

Let $p_i = P$(outcome i on any particular trial), and define random variables by X_i = the number of trials resulting in outcome i ($i = 1, \ldots, r$).

Such an experiment is called a **multinomial experiment,** and the joint pmf of $X_1, \ldots, X_r$ is called the **multinomial distribution.**

34

More Than Two Random Variables

By using a counting argument analogous to the one used in deriving the binomial distribution, the joint pmf of $X_1, \ldots, X_r$ can be shown to be

$$p(x_1, \ldots, x_r) = \begin{cases} \dfrac{n!}{(x_1!)(x_2!) \cdot \cdots \cdot (x_r!)} p_1^{x_1} \cdot \ldots \cdot p_r^{x_r} & x_i = 0, 1, 2, \ldots, \text{with } x_1 + \cdots + x_r = n \\ 0 & \text{otherwise} \end{cases}$$

The case $r = 2$ gives the binomial distribution, with

X_1 = number of successes and
$X_2 = n - X_1$ = number of failures.

35

Example 9

If the allele of each of ten independently obtained pea sections is determined and $p_1 = P(\text{AA})$, $p_2 = P(\text{Aa})$, $p_3 = P(\text{aa})$, X_1 = number of AAs, X_2 = number of Aas, and X_3 = number of aa's, then the multinomial pmf for these X_i's is

$$p(x_1, x_2, x_3) = \frac{10!}{(x_1!)(x_2!)(x_3!)} p_1^{x_1} p_2^{x_2} p_3^{x_3} \quad x_i = 0, 1, \ldots \quad \text{and } x_1 + x_2 + x_3 = 10$$

With $p_1 = p_3 = .25$, $p_2 = .5$,

$P(X_1 = 2, X_2 = 5, X_3 = 3) = p(2, 5, 3)$

$$= \frac{10!}{2!\,5!\,3!}(.25)^2(.5)^5(.25)^3 = .0769$$

36

More Than Two Random Variables

The notion of independence of more than two random variables is similar to the notion of independence of more than two events.

Definition

The random variables $X_1, X_2, \ldots, X_n$ are said to be **independent** if for *every* subset $X_{i_1}, X_{i_2}, \ldots, X_{i_k}$ of the variables (each pair, each triple, and so on), the joint pmf or pdf of the subset is equal to the product of the marginal pmf's or pdf's.

37

Conditional Distributions

38

Conditional Distributions

Suppose X = the number of major defects in a randomly selected new automobile and Y = the number of minor defects in that same auto.

If we learn that the selected car has one major defect, what now is the probability that the car has at most three minor defects—that is, what is $P(Y \leq 3 \mid X = 1)$?

39

Conditional Distributions

Similarly, if X and Y denote the lifetimes of the front and rear tires on a motorcycle, and it happens that X = 10,000 miles, what now is the probability that Y is at most 15,000 miles, and what is the expected lifetime of the rear tire "conditional on" this value of X?

Questions of this sort can be answered by studying conditional probability distributions.

40

Conditional Distributions

Definition

Let X and Y be two continuous rv's with joint pdf $f(x, y)$ and marginal X pdf $f_X(x)$. Then for any X value x for which $f_X(x) > 0$, the **conditional probability density function of *Y* given that *X* = *x*** is

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

If X and Y are discrete, replacing pdf's by pmf's in this definition gives the **conditional probability mass function of *Y* when *X* = *x*.**

41

Conditional Distributions

Notice that the definition of $f_{Y|X}(y \mid x)$ parallels that of $P(B \mid A)$, the conditional probability that B will occur, given that A has occurred.

Once the conditional pdf or pmf has been determined, questions of the type posed at the outset of this subsection can be answered by integrating or summing over an appropriate set of Y values.

42

Example 12

Reconsider the situation of example 3 and 4 involving X = the proportion of time that a bank's drive-up facility is busy and Y = the analogous proportion for the walk-up window.

The conditional pdf of Y given that X = .8 is

$$f_{Y|X}(y \mid .8) = \frac{f(.8, y)}{f_X(.8)} = \frac{1.2(.8 + y^2)}{1.2(.8) + .4} = \frac{1}{34}(24 + 30y^2) \quad 0 < y < 1$$

43

Example 12

The probability that the walk-up facility is busy at most half the time given that X = .8 is then

$$P(Y \le .5 \mid X = .8) = \int_{-\infty}^{.5} f_{Y|X}(y \mid .8)\, dy$$

$$= \int_0^{.5} \frac{1}{34}(24 + 30y^2)\, dy$$

$$= .390$$

44

Example 12

cont'd

Using the marginal pdf of *Y* gives $P(Y \leq .5) = .350$. Also $E(Y) = .6$, whereas the expected proportion of time that the walk-up facility is busy given that $X = .8$ (a *conditional* expectation) is

$$E(Y|X = .8) = \int_{-\infty}^{\infty} y \cdot f_{Y|X}(y|.8)\, dy$$

$$= \frac{1}{34}\int_0^1 y(24 + 30y^2)\, dy$$

$$= .574$$

45