

Joint Distributions, Independence
Class 7, 18.05, Spring 2014
Jeremy Orloff and Jonathan Bloom

1 Learning Goals

1. Understand what is meant by a *joint* pmf, pdf and cdf of two random variables.
2. Be able to compute probabilities and marginals from a joint pmf or pdf.
3. Be able to test whether two random variables are independent.

2 Introduction

In science and in real life, we are often interested in two (or more) random variables at the same time. For example, we might measure the height and weight of giraffes, or the IQ and birthweight of children, or the frequency of exercise and the rate of heart disease in adults, or the level of air pollution and rate of respiratory illness in cities, or the number of Facebook friends and the age of Facebook members.

Think: What relationship would you expect in each of the five examples above? Why?

In such situations the random variables have a *joint distribution* that allows us to compute probabilities of events involving both variables and understand the relationship between the variables. This is simplest when the variables are *independent*. When they are not, we use *covariance* and *correlation* to measure of the nature of the dependence between them.

3 Joint Distribution

3.1 Discrete case

Suppose X and Y are two discrete random variables and that X takes values $\{x_1, x_2, \dots, x_n\}$ and Y takes values $\{y_1, y_2, \dots, y_m\}$. The ordered pair (X, Y) take values in the product $\{(x_1, y_1), (x_1, y_2), \dots, (x_n, y_m)\}$. The *joint probability mass function* (joint pmf) of X and Y is the function $p(x_i, y_j)$ giving the probability of the joint outcome $X = x_i, Y = y_j$.

We organize this in a *joint probability table* as shown:

$X \setminus Y$	y_1	y_2	\dots	y_j	\dots	y_m
x_1	$p(x_1, y_1)$	$p(x_1, y_2)$	\dots	$p(x_1, y_j)$	\dots	$p(x_1, y_m)$
x_2	$p(x_2, y_1)$	$p(x_2, y_2)$	\dots	$p(x_2, y_j)$	\dots	$p(x_2, y_m)$
\dots	\dots	\dots	\dots	\dots	\dots	\dots
\dots	\dots	\dots	\dots	\dots	\dots	\dots
x_i	$p(x_i, y_1)$	$p(x_i, y_2)$	\dots	$p(x_i, y_j)$	\dots	$p(x_i, y_m)$
\dots	\dots	\dots	\dots	\dots	\dots	\dots
x_n	$p(x_n, y_1)$	$p(x_n, y_2)$	\dots	$p(x_n, y_j)$	\dots	$p(x_n, y_m)$

Example 1. Roll two dice. Let X be the value on the first die and let Y be the value on the second die. Then both X and Y take values 1 to 6 and the joint pmf is $p(i, j) = 1/36$ for all i and j between 1 and 6. Here is the joint probability table:

$X \setminus Y$	1	2	3	4	5	6
1	1/36	1/36	1/36	1/36	1/36	1/36
2	1/36	1/36	1/36	1/36	1/36	1/36
3	1/36	1/36	1/36	1/36	1/36	1/36
4	1/36	1/36	1/36	1/36	1/36	1/36
5	1/36	1/36	1/36	1/36	1/36	1/36
6	1/36	1/36	1/36	1/36	1/36	1/36

Example 2. Roll two dice. Let X be the value on the first die and let T be the total on both dice. Here is the joint probability table:

$X \setminus T$	2	3	4	5	6	7	8	9	10	11	12
1	1/36	1/36	1/36	1/36	1/36	1/36	0	0	0	0	0
2	0	1/36	1/36	1/36	1/36	1/36	1/36	0	0	0	0
3	0	0	1/36	1/36	1/36	1/36	1/36	1/36	0	0	0
4	0	0	0	1/36	1/36	1/36	1/36	1/36	1/36	0	0
5	0	0	0	0	1/36	1/36	1/36	1/36	1/36	1/36	0
6	0	0	0	0	0	1/36	1/36	1/36	1/36	1/36	1/36

A joint probability mass function must satisfy two properties:

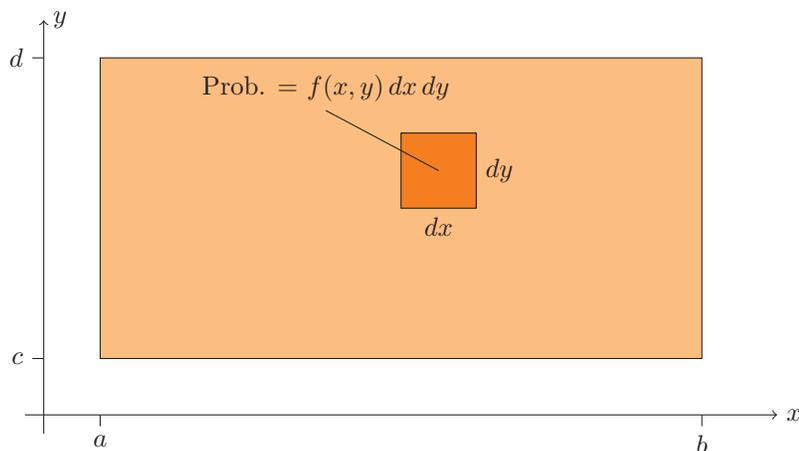
- $0 \leq p(x_i, y_j) \leq 1$
- The total probability is 1. We can express this as a *double sum*:

$$\sum_{i=1}^n \sum_{j=1}^m p(x_i, y_j) = 1$$

3.2 Continuous case

The continuous case is essentially the same as the discrete case: we just replace discrete sets of values by continuous intervals, the joint pmf by a joint pdf, and sums by integrals.

If X takes values in $[a, b]$ and Y takes values in $[c, d]$ then the pair (X, Y) takes values in the product $[a, b] \times [c, d]$. The *joint probability density function* (joint pdf) of X and Y is a function $f(x, y)$ giving the probability density at (x, y) . That is, the probability that (X, Y) is in a small rectangle of width dx and height dy around (x, y) is $f(x, y) dx dy$.



A joint probability density function must satisfy two properties:

1. $0 \leq f(x, y)$
2. The total probability is 1. We now express this as a *double integral*:

$$\int_c^d \int_a^b f(x, y) dx dy = 1$$

Note: as with the pdf of a single random variable, the joint pdf $f(x, y)$ can take values greater than 1; it is a probability density, *not* a probability.

In 18.05 we won't expect you to be experts at double integration. Here's what we will expect.

- You should understand double integrals conceptually as double sums.
- You should be able to compute double integrals over rectangles.
- For a non-rectangular region, when $f(x, y) = c$ is constant, you should know that the double integral is the same as the $c \times$ (the area of the region).

3.3 Events

Random variables are useful for describing events. Recall that an event is a set of outcomes and that random variables assign numbers to outcomes. For example, the event ' $X > 1$ ' is the set of all outcomes for which X is greater than 1. These concepts readily extend to pairs of random variables and joint outcomes.

Example 3. In Example 1, describe the event $B = \{Y - X \geq 2\}$ and find its probability.

answer: We can describe B as a set of (X, Y) pairs:

$$B = \{(1, 3), (1, 4), (1, 5), (1, 6), (2, 4), (2, 5), (2, 6), (3, 5), (3, 6), (4, 6)\}.$$

We can also describe it visually

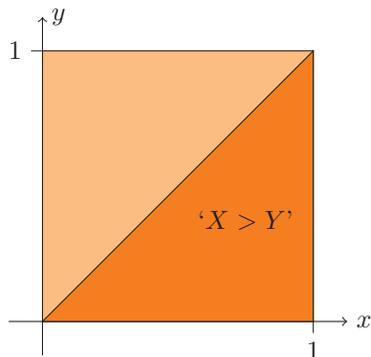
$X \backslash Y$	1	2	3	4	5	6
1	1/36	1/36	1/36	1/36	1/36	1/36
2	1/36	1/36	1/36	1/36	1/36	1/36
3	1/36	1/36	1/36	1/36	1/36	1/36
4	1/36	1/36	1/36	1/36	1/36	1/36
5	1/36	1/36	1/36	1/36	1/36	1/36
6	1/36	1/36	1/36	1/36	1/36	1/36

The event B consists of the outcomes in the shaded squares.

The probability of B is the sum of the probabilities in the orange shaded squares, so $P(B) = 10/36$.

Example 4. Suppose X and Y both take values in $[0, 1]$ with uniform density $f(x, y) = 1$. Visualize the event ' $X > Y$ ' and find its probability.

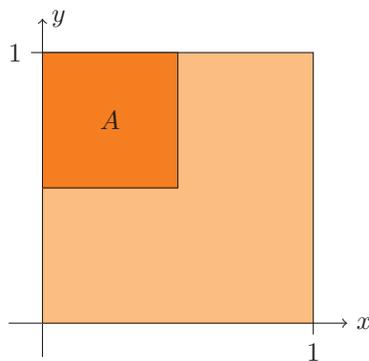
answer: Jointly X and Y take values in the unit square. The event ' $X > Y$ ' corresponds to the shaded lower-right triangle below. Since the density is constant, the probability is just the fraction of the total area taken up by the event. In this case, it is clearly .5.



The event ' $X > Y$ ' in the unit square.

Example 5. Suppose X and Y both take values in $[0, 1]$ with density $f(x, y) = 4xy$. Show $f(x, y)$ is a valid joint pdf, visualize the event $A = \{X < .5 \text{ and } Y > .5\}$ and find its probability.

answer: Jointly X and Y take values in the unit square.

The event A in the unit square.

To show $f(x, y)$ is a valid joint pdf we must check that it is positive (which it clearly is) and that the total probability is 1.

$$\text{Total probability} = \int_0^1 \int_0^1 4xy \, dx \, dy = \int_0^1 [2x^2y]_0^1 \, dy = \int_0^1 2y \, dy = 1. \quad \text{QED}$$

The event A is just the upper-left-hand quadrant. Because the density is not constant we must compute an integral to find the probability.

$$P(A) = \int_0^{.5} \int_{.5}^1 4xy \, dy \, dx = \int_0^{.5} [2xy^2]_{.5}^1 \, dx = \int_0^{.5} \frac{3x}{2} \, dx = \boxed{\frac{3}{16}}.$$

3.4 Joint cumulative distribution function

Suppose X and Y are jointly-distributed random variables. We will use the notation ' $X \leq x, Y \leq y$ ' to mean the event ' $X \leq x$ and $Y \leq y$ '. The *joint cumulative distribution function* (joint cdf) is defined as

$$F(x, y) = P(X \leq x, Y \leq y)$$

Continuous case: If X and Y are continuous random variables with joint density $f(x, y)$ over the range $[a, b] \times [c, d]$ then the joint cdf is given by the double integral

$$F(x, y) = \int_c^y \int_a^x f(u, v) \, du \, dv.$$

To recover the joint pdf, we differentiate the joint cdf. Because there are two variables we need to use partial derivatives:

$$f(x, y) = \frac{\partial^2 F}{\partial x \partial y}(x, y).$$

Discrete case: If X and Y are discrete random variables with joint pmf $p(x_i, y_j)$ then the joint cdf is given by the double sum

$$F(x, y) = \sum_{x_i \leq x} \sum_{y_j \leq y} p(x_i, y_j).$$

3.5 Properties of the joint cdf

The joint cdf $F(x, y)$ of X and Y must satisfy several properties:

1. $F(x, y)$ is non-decreasing: if x or y increase then $F(x, y)$ must stay constant or increase.

2. $F(x, y) = 0$ at the lower-left of the joint range.

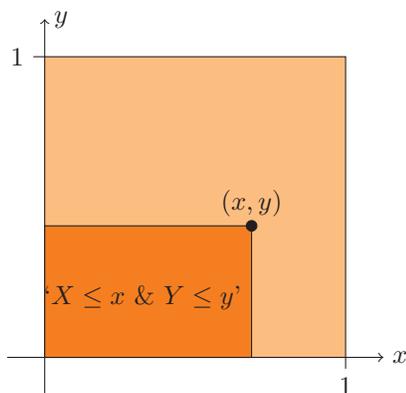
If the lower left is $(-\infty, -\infty)$ then this means $\lim_{(x,y) \rightarrow (-\infty, -\infty)} F(x, y) = 0$.

3. $F(x, y) = 1$ at the upper-right of the joint range.

If the upper-right is (∞, ∞) then this means $\lim_{(x,y) \rightarrow (\infty, \infty)} F(x, y) = 1$.

Example 6. Find the joint cdf for the random variables in Example 5.

answer: The event ' $X \leq x$ and $Y \leq y$ ' is a rectangle in the unit square.



To find the cdf $F(x, y)$ we compute a double integral:

$$F(x, y) = \int_0^y \int_0^x 4uv \, du \, dv = \boxed{x^2 y^2}.$$

Example 7. In Example 1, compute $F(3.5, 4)$.

answer: We redraw the joint probability table. Notice how similar the picture is to the one in the previous example.

$F(3.5, 4)$ is the probability of the event ' $X \leq 3.5$ and $Y \leq 4$ '. We can visualize this event as the shaded rectangles in the table:

$X \setminus Y$	1	2	3	4	5	6
1	1/36	1/36	1/36	1/36	1/36	1/36
2	1/36	1/36	1/36	1/36	1/36	1/36
3	1/36	1/36	1/36	1/36	1/36	1/36
4	1/36	1/36	1/36	1/36	1/36	1/36
5	1/36	1/36	1/36	1/36	1/36	1/36
6	1/36	1/36	1/36	1/36	1/36	1/36

The event ' $X \leq 3.5$ and $Y \leq 4$ '.

Adding up the probability in the shaded squares we get $F(3.5, 4) = 12/36 = 1/3$.

3.6 Marginal distributions

When X and Y are jointly-distributed random variables, we may want to consider only one of them, say X . In that case we need to find the pmf (or pdf or cdf) of X without Y . This is called a *marginal pmf* (or pdf or cdf). The next example illustrates the way to compute this and the reason for the term ‘marginal’.

3.7 Marginal pmf

Example 8. In Example 2 we rolled two dice and let X be the value on the first die and T be the total on both dice. Compute the marginal pmf of X and of T .

answer: In the table each row represents a single value of X . So the event ‘ $X = 3$ ’ is the third row of the table. To find $P(X = 3)$ we simply have to sum up the probabilities in this row. We put the sum in the right-hand *margin* of the table. Likewise $P(T = 5)$ is just the sum of the column with $T = 5$. We put the sum in the bottom *margin* of the table.

$X \setminus T$	2	3	4	5	6	7	8	9	10	11	12	$p(x_i)$
1	1/36	1/36	1/36	1/36	1/36	1/36	0	0	0	0	0	1/6
2	0	1/36	1/36	1/36	1/36	1/36	1/36	0	0	0	0	1/6
3	0	0	1/36	1/36	1/36	1/36	1/36	1/36	0	0	0	1/6
4	0	0	0	1/36	1/36	1/36	1/36	1/36	1/36	0	0	1/6
5	0	0	0	0	1/36	1/36	1/36	1/36	1/36	1/36	0	1/6
6	0	0	0	0	0	1/36	1/36	1/36	1/36	1/36	1/36	1/6
$p(t_j)$	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36	1

Computing the marginal probabilities $P(X = 3) = 1/6$ and $P(T = 5) = 4/36$.

Note: Of course in this case we already knew the pmf of X and of T . It is good to see that our computation here is in agreement.

As motivated by this example, marginal pmf’s are obtained from the joint pmf by summing:

$$p_X(x_i) = \sum_j p(x_i, y_j), \quad p_Y(y_j) = \sum_i p(x_i, y_j)$$

The term marginal refers to the fact that the values are written in the margins of the table.

3.8 Marginal pdf

For a continuous joint density $f(x, y)$ with range $[a, b] \times [c, d]$, the marginal pdf’s are:

$$f_X(x) = \int_c^d f(x, y) dy, \quad f_Y(y) = \int_a^b f(x, y) dx.$$

Compare these with the marginal pmf's above; as usual the sums are replaced by integrals. We say that to obtain the marginal for X , we *integrate out* Y from the joint pdf and vice versa.

Example 9. Suppose (X, Y) takes values on the square $[0, 1] \times [1, 2]$ with joint pdf $f(x, y) = \frac{8}{3}x^3y$. Find the marginal pdf's $f_X(x)$ and $f_Y(y)$.

answer:

$$f_X(x) = \int_1^2 \frac{8}{3}x^3y \, dy = \left[\frac{4}{3}x^3y^2 \right]_1^2 = \boxed{4x^3}$$

$$f_Y(y) = \int_0^1 \frac{8}{3}x^3y \, dx = \left[\frac{2}{3}x^4y \right]_0^1 = \boxed{\frac{2}{3}y}.$$

Example 10. Suppose (X, Y) takes values on the unit square $[0, 1] \times [0, 1]$ with joint pdf $f(x, y) = \frac{3}{2}(x^2 + y^2)$. Find the marginal pdf $f_X(x)$ and use it to find $P(X < .5)$.

answer:

$$f_X(x) = \int_0^1 \frac{3}{2}(x^2 + y^2) \, dy = \left[\frac{3}{2}x^2y + \frac{y^3}{2} \right]_0^1 = \boxed{\frac{3}{2}x^2 + \frac{1}{2}}.$$

$$P(X < .5) = \int_0^{.5} f_X(x) \, dx = \int_0^{.5} \left(\frac{3}{2}x^2 + \frac{1}{2} \right) \, dx = \left[\frac{1}{2}x^3 + \frac{1}{2}x \right]_0^{.5} = \boxed{\frac{5}{16}}.$$

3.9 Marginal cdf

Finding the marginal cdf from the joint cdf is easy. If X and Y jointly take values on $[a, b] \times [c, d]$ then

$$F_X(x) = F(x, d), \quad F_Y(y) = F(b, y).$$

If d is ∞ then this becomes a limit $F_X(x) = \lim_{y \rightarrow \infty} F(x, y)$. Likewise for $F_Y(y)$.

Example 11. The joint cdf in the last example was $F(x, y) = \frac{1}{2}(x^3y + xy^3)$ on $[0, 1] \times [0, 1]$. Find the marginal cdf's and use $F_X(x)$ to compute $P(X < .5)$.

answer: We have $F_X(x) = F(x, 1) = \frac{1}{2}(x^3 + x)$ and $F_Y(y) = F(1, y) = \frac{1}{2}(y + y^3)$. So $P(X < .5) = F_X(.5) = \frac{1}{2}(.5^3 + .5) = \frac{5}{16}$ as before.

3.10 3D visualization

We visualized $P(a < X < b)$ as the area under the pdf $f(x)$ over the interval $[a, b]$. Since the range of values of (X, Y) is already a two dimensional region in the plane, the graph of $f(x, y)$ is a surface over that region. We can then visualize probability as *volume* under the surface.

Think: Summoning your inner artist, sketch the graph of the joint pdf $f(x, y) = 4xy$ and visualize the probability $P(A)$ as a volume for Example 5.

$X \setminus T$	2	3	4	5	6	7	8	9	10	11	12	$p(x_i)$
1	1/36	1/36	1/36	1/36	1/36	1/36	0	0	0	0	0	1/6
2	0	1/36	1/36	1/36	1/36	1/36	1/36	0	0	0	0	1/6
3	0	0	1/36	1/36	1/36	1/36	1/36	1/36	0	0	0	1/6
4	0	0	0	1/36	1/36	1/36	1/36	1/36	1/36	0	0	1/6
5	0	0	0	0	1/36	1/36	1/36	1/36	1/36	1/36	0	1/6
6	0	0	0	0	0	1/36	1/36	1/36	1/36	1/36	1/36	1/6
$p(y_j)$	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36	1

Example 13. For **continuous variables** independence means you can factor the joint pdf or cdf as the product of a function of x and a function of y .

i) If $f(x, y) = 4x^2y^3$ then X and Y are independent. The marginal densities are $f_X(x) = ax^2$ and $f_Y(y) = by^3$ for some constants a and b with $ab = 4$.

ii) If $f(x, y) = x^2 + y^2$ then X and Y are not independent because there is no way to factor $f(x, y)$ into a product $f_X(x)f_Y(y)$.

iii) If $F(x, y) = \frac{1}{2}(x^3y + xy^3)$ then X and Y are not independent because the cdf does not factor into a product $F_X(x)F_Y(y)$.

MIT OpenCourseWare
<http://ocw.mit.edu>

18.05 Introduction to Probability and Statistics
Spring 2014

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.