Two Dimensional Random Variables (Continued)

Definition 3.3

1. (X, Y) is a two-dimensional **discrete** random variable if the possible values of $(X(s), Y(s))$ are finite or countable infinite.

   i.e. the possible values of $(X(s), Y(s))$ may be represented as
   $$(x_i, y_j), i = 1, 2, 3, \cdots; j = 1, 2, 3, \cdots$$

2. (X, Y) is a two-dimensional **continuous** random variable if the possible values of $(X(s), Y(s))$ can assume all values in some region of the Euclidean plane $\mathbb{R}^2$.

To judge whether a two dimensional random vector (X, Y) is discrete or continuous, we can view X and Y separately.

- ✓ If both X and Y are discrete random variables, we say (X, Y) is a discrete random vector.
- ✓ Likewise, if both X and Y are continuous random variables, we say (X, Y) is a continuous random vector.
- ✓ Certainly, there are other cases. For example, X is discrete, but Y is continuous, or Y is neither a discrete nor a continuous random variable. But these are not the main focus of this module.

An example: Consider toss a coin twice
The sample space = {(H,H), (H,T), (T,H), (T,T)}
Let X = number of heads in two tosses and
Y = number of head in the first toss

s	(H,H)	(H,T)	(T,H)	(T,T)
probability	1/4	1/4	1/4	1/4
x	2	1	1	0
y	1	1	0	0
(x,y)	(2,1)	(1,1)	(1,0)	(0,0)

Note: (x,y) does not take values (0,1) and (2,0)

3.2.1 Joint Probability Function for Discrete RVs

Definition 3.4

- Let (X, Y) be a 2-dimensional **discrete** random variable defined on the sample space of an experiment. With each possible value (x_i, y_j), we associate a number $f_{X,Y}(x_i, y_j)$ representing $\Pr(X = x_i, Y = y_j)$ and satisfying the following conditions:

1. $f_{X,Y}(x_i, y_j) \geq 0$ for all $(x_i, y_j) \in R_{X,Y}$.
2. $\sum_{i=1}^{\infty} \sum_{j=1}^{\infty} f_{X,Y}(x_i, y_j) = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \Pr(X = x_i, Y = y_j) = 1$ (3.1)

Equation (3.1) on this page of the lecture slide essentially requires that the summation over all $f(x_i, y_j) > 0$ equals 1. It can be equivalently written as

$$\sum_{(x_i,y_j):f_{X,Y}(x_i,y_j)>0} f_{X,Y}(x_i, y_j) = 1.$$

Note that in this case, $f_{X,Y}(x_i, y_j)$ may not be defined for some x_i and y_j; see the distribution given on page 3-20. So, in this case, if you would like to add $i = 0, 1, 2, 3$ and $j = 0, 1, 2, 3$ freely, you need use 0 to replace those $f_{X,Y}(x, y)$ who does not have a point mass on (x, y).

Solution to Example 3 (Continued)

The above p.f. are given explicitly in the following table.

x	*y*				Row Total
	0	**1**	**2**	**3**	
0	0	3/84	6/84	1/84	10/84
1	4/84	24/84	12/84	0	40/84
2	12/84	18/84	0	0	30/84
3	4/84	0	0	0	4/84
Column Total	20/84	45/84	18/84	1/84	1

Joint pdf for Continuous RVs (Continued)

1. $f_{X,Y}(x,y) \geq 0$ for all $(x,y) \in R_{X,Y}$.
2.

$$\iint_{(x,y)\in R_{X,Y}} f_{X,Y}(x,y)dx\,dy = 1$$

or

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

ST2334 Probability and Statistics CYM 2-dim RVs & cond prob dist 3-22

✓ In most cases of this module, when we do the bivariate integration, the integration region is a rectangular; therefore, the variables x and y can be integrated separately; and the order of which is integrated first does not matter. See examples 3-24, 3-28, and 3-29.

✓ However, we need to bear in mind that there are cases under which the integration region is NOT a rectangular, so that x and y can not move freely for a unified expression of $f_{X,Y}(x,y)$. See the example given on pages 3-25, 3-26, and 3-27 of the lecture slides: the region is defined by straight lines such as a triangle or a trapezium.

Note: when we integrate a two dimensional function in a region which is not a rectangular, we need to take care that x and y may not move freely! Based on mathematical theory, integrating which variable first won't change the outcome of the integration; however, a right choice of integration order may make the computation easier; read pages 3-25 to 3-26 carefully for such an example.

Marginal Distributions (Continued)

- For **discrete** case,

$$f_X(x) = \sum_y f_{X,Y}(x,y) \quad \text{and} \quad f_Y(y) = \sum_x f_{X,Y}(x,y)$$

- For **continuous** case,

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy$$

and

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dx$$

The practical interpretation of the marginal distribution for X is: focusing on viewing the distribution of X by ignoring the presence of Y. **Note that**

★ **$f_X(x)$ should NOT involve y; and**

★ **it is a pdf/pmf; so it must have all the properties of a pdf/pmf.**

If (X, Y) is discrete, then the marginals are also discrete; likewise, if (X, Y) is continuous, the marginals are also continuous.

The meaning of the formulae for $f_X(x)$ is that "for each given x, integrate (or sum) over all the value of y such that $f_{X,Y}(x, y) > 0$." So, similar to the discussion of page 4 above, we need to take care of the region of y for each x.

Conditional Distribution (Continued)

Definition 3.7 (Continued)

- Then **the conditional distribution of Y given that $X = x$** is given by

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)}, \qquad \text{if } f_X(x) > 0,$$

for each x within the range of X.

✓ The condition distribution is similar in meaning to the condition probability. It is the distribution of the random variable of Y when the random X is fixed at a certain value x.

✓ It is important to take note that it is a distribution for y, so it must satisfies all the properties of a pdf/pmf in terms of the argument y for every x that it is defined.

✓ It may or may not be a function of x. But it is defined only when x satisfies $f_X(x) > 0$. If it does not depend on x, then we have X and Y independent.

✓ It is not a pdf/pmf for x. So there is NO requirement that $\int_{-\infty}^{\infty} f_{Y|X}(y|x)dx = 1$ when Y is continuous or $\sum_x f_{Y|X}(y|x) = 1$, when Y is discrete.

✓ Can you find $f_{Y|X}(y|x)$ for the example given on page 5?

Pages 3-36 & 3-47

Example 1 (Continued)

- $f_{X,Y}(x,y)$, $f_X(x)$ and $f_Y(y)$ are displayed in the following table

y	x						$f_Y(y)$
	0	1	2	3	4	5	
0	0	0.01	0.02	0.05	0.06	0.08	0.22
1	0.01	0.03	0.04	0.05	0.05	0.07	0.25
2	0.02	0.03	0.05	0.06	0.06	0.07	0.29
3	0.02	0.04	0.03	0.04	0.06	0.05	0.24
$f_X(x)$	0.05	0.11	0.14	0.20	0.23	0.27	1

Example 1 (Continued)

Outcome	HHH	THH	HTH	HHT	TTH	THT	HTT	TTT
(x,y)	(1,3)	(1,2)	(1,2)	(0,2)	(1,1)	(0,1)	(0,1)	(0,0)
$f_{X,Y}(x,y)$	1/8	1/8	1/8	1/8	1/8	1/8	1/8	1/8

- The joint probability distribution of (X, Y) is given in the following table:

x	y				$f_X(x)$
	0	1	2	3	
0	1/8	1/4	1/8	0	1/2
1	0	1/8	1/4	1/8	1/2
$f_Y(y)$	1/8	3/8	3/8	1/8	1

For a discrete random vector (X, Y). The two-dimensional tables as shown in these slides are particularly useful to help us understand the joint, marginal, and the conditional distributions.

Example 4 (Continued)

(c) Given that the drive-up facility is busy 80% of the time, what is the probability that the walk-in facility is busy at most half the time?

i.e. Find $\Pr(Y \le 1/2 \mid X = 4/5)$.

(d) Given that the drive-up facility is busy 80% of the time, what is the expected proportion of time that the walk-in facility is busy?

i.e. Find $E(Y \mid X = 4/5)$.

✓ Both conditional probability and conditional expectation are established on the conditional distribution. In particular, if (X, Y) is a continuous random vector, for any x and y,

$$\begin{aligned} P(Y \le y|X = x) &= \int_{-\infty}^{y} f_{Y|X}(t|x)dt \\ E(Y|X = x) &= \int_{-\infty}^{\infty} y f_{Y|X}(y|x)dy, \end{aligned}$$

where the former depends on both y and x, but the latter depends only on x. If (X, Y) is a discrete random vector, the integration is replaced with summation.

✓ The key to evaluate these quantities is to find the conditional pdf/pmf $f_{Y|X}(y|x)$. See also the example on Page 3-120

✓ In occasions, you may also see

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

which is a function of the random variable X.

✓ Think of the following questions:

- ★ What is the meaning of $E(E(Y|X))$? How to evaluate it?
- ★ What is $E(Y)$? How to evaluate it?
- ★ How to compute $E(a_1 g_1(X) + a_2 g_2(Y))$ and $E(g_1(X) g_2(Y))$, where a_1, a_2 are real numbers, and $g_1(\cdot)$ and $g_2(\cdot)$ are given functions.

Example 5

Let X and Y be **uniformly distributed** over the triangle with the boundaries: $0 \le x \le y, 0 \le y \le 2$.

(a) Find the joint p.d.f. of (X, Y),

(b) Find $f_X(x)$ and $f_Y(y)$.

(c) Find $f_{Y|X}(y|x)$ and $f_{X|Y}(x|y)$.

(d) Find $\Pr(X \le 1/2 \mid Y = 1)$

(e) Find $\Pr(X \le 1, Y \le 1)$.

✓ (X, Y) is uniformly distributed if its pdf/pmf is in the form

$$f_{X,Y}(x, y) = \begin{cases} c & (x, y) \in A \\ 0 & \text{elsewhere} \end{cases},$$

where c is a real number not depending on x and y. In fact, if (X, Y) is continuous, $c = 1/\text{area}(A)$; if (X, Y) is discrete, $c = 1/\#A$.

✓ (X, Y) is uniform does not imply X or/and Y is uniform. Likewise, "both X and Y are uniformly distributed" does not imply that "(X, Y) is uniformly distributed." The example given on the lecture slide above illustrates this idea.

✓ But if A is a product space, then both X and Y are uniformly distributed; and vice versa. In this case, X and Y are independent. For example, $f(x, y) = 1, x \in [0, 1], y \in [0, 1]$, and $= 0$ otherwise.

3.4 Independent Random Variables

3.4.1 Definition of Independent RVs

Definition

- Random variables X and Y are **independent** if and only if

$$f_{X,Y}(x,y) = f_X(x)\, f_Y(y), \quad \textbf{for all } \boldsymbol{x}, \boldsymbol{y}.$$

Extension:

- Random variables $X_1, X_2, \cdots, X_n$ are independent if and only if

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

for all $\boldsymbol{x_i}, \boldsymbol{i} = \mathbf{1}, \cdots, \boldsymbol{n}$.

ST2334 Probability and Statistics CYM 2-dim RVs & cond prob dist 3-84

✓ Independence of random variables is a very important concept in statistics and probability. Similar to the independence of probability events, it practically means that the value of one random variable is not related to that of the other.

✓ In the definition "f" could be pdf or pmf. Therefore this definition is applicable no matter whether X and Y are discrete or continuous. It continue to be applicable when one is discrete but the other is continuous.

✓ There are several equivalent ways to define/check the independence.

★ Random variables X and Y are independent if and only if for ARBITRARY sets $A, B \subset \mathbb{R}$,

$$Pr(X \in A; Y \in B) = Pr(X \in A)Pr(Y \in B).$$

★ Random variables X and Y are independent if and only if for any $x, y \in \mathbb{R}$,

$$Pr(X \leq x; Y \leq y) = Pr(X \leq x)Pr(Y \leq y),$$

which can also be written as $F_{X,Y}(x, y) = F_X(x)F_Y(y)$.

★ Random variables X and Y are independent, if and only if for any functions $g_1(\cdot)$ and $g_2(\cdot)$, $E(g_1(X)g_2(Y)) = E(g_1(X))E(g_2(Y))$.

★ These statements can also be extended to multiple random variables.

Remark

- The product of 2 positive functions $f_X(x)$ and $f_Y(y)$ means a function which is positive on a **product space.**
- That is, if

$$f_X(x) > 0, \quad \text{for } x \in A_1 \quad \text{and}$$
$$f_Y(y) > 0, \quad \text{for } y \in A_2$$

then $\quad f_X(x)f_Y(y) > 0, \text{ for } (x,y) \in A_1 \times A_2.$

"$f_{X,Y}(x,y)$ is positive in a product space" is a necessary (but not sufficient) condition so that two random variables are independent. It can be used to assert that two random variables are not independent.

✓ If X and Y are continuous random variables, for them to be independent, we need that $A = \left\{(x,y)\middle| f_{X,Y}(x,y) > 0\right\}$ can be written in the form $(\cup_{i=1}^{\infty}[a_i, b_i]) \times \left(\cup_{j=1}^{\infty}[c_j, d_j]\right)$. An even quicker view is that at least it must be a union of a countable number of rectangles.

✓ If X and Y are discrete random variables, for them to be independent, we need that for every $x \in A_1, y \in A_2$, $f_{X,Y}(x,y) > 0$. One example that this is not satisfied can be found on page 3-35 of the lecture slides.

Example 1 (Continued)

- $f_{X,Y}(x,y)$, $f_X(x)$ and $f_Y(y)$ are displayed in the following table

y	x						$f_Y(y)$
	0	1	2	3	4	5	
0	0	0.01	0.02	0.05	0.06	0.08	0.22
1	0.01	0.03	0.04	0.05	0.05	0.07	0.25
2	0.02	0.03	0.05	0.06	0.06	0.07	0.29
3	0.02	0.04	0.03	0.04	0.06	0.05	0.24
$f_X(x)$	0.05	0.11	0.14	0.20	0.23	0.27	1

ST2334 Probability and Statistics CYM 2-dim RVs & cond prob dist 3-36

To check this easily, we only need to view that $f_{X,Y}(x,y) > 0$ for $x \in A_1$ and $y \in A_2$ with A_1 and A_2 two subsets of real numbers not depending on x and y.

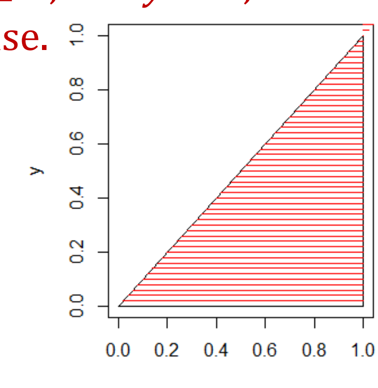
Can you use the discussion above to quickly conclude that X and Y are not independent in the following example?

Example 4

- Given that

$$f_{X,Y}(x,y) = \begin{cases} 2\,(x+y), & \text{for } 0 \le x \le 1, 0 < y < x, \\ 0, & \text{otherwise.} \end{cases}$$

- are X and Y independent?



ST2334 Probability and Statistics CYM 2-dim RVs & cond prob dist 3-94

Example 2

- Refer to example 1 in Section 3.2.1 on p3-12.

$$f_{X,Y}(x,y) = \frac{xy}{36}$$

for $x = 1, 2, 3$, and $y = 1, 2, 3$.

- Are X and Y independent?

A handy way to check independence in applications is to check that

✓ the joint density function is positive on a product space; see the discussion on page 6; and

✓ for the positive part of the joint density, we have $f_{X,Y}(x,y) = C \cdot g_1(x)g_2(y)$, i.e., it can be factorized as the product of two functions g_1 and g_2, where the former depends on x only, the latter depends on y only, and C is a constant not depending on x and y.

Here, we note that $g_1(x)$ and $g_2(y)$ on their own are not necessarily pdf/pmfs.

To illustrate, consider the example in the slide.

✓ $A_1 = \{1, 2, 3\}$ and $A_2 = \{1, 2, 3\}$, so the joint density is positive in product space.

✓ $f_{X,Y}(x,y) = \frac{1}{36}(x) \cdot (y)$, which is the multiplication of two functions: one depends on x only, the other depends on y only.

So we conclude that X and Y are independent.

Furthermore, we can also get the marginal distributions of X and Y easily by standardizing $g_1(\cdot)$ and $g_2(\cdot)$ to ensure that they satisfy the definition of the pdf/pmf; we use X to illustrate:

✓ If X is a discrete random variable, its pmf is give by $f_X(x) = \frac{g_1(x)}{\sum_{t \in A_1} g_1(t)}$.

✓ If X is a continuous random variable, its pdf is given by $f_X(x) = \frac{g_1(x)}{\int_{t \in A_1} g_1(t)dt}$.

Again, we use the example given on the slide to illustrate. Here X is a discrete random variable, so its pmf is given by

$$f_X(x) = \frac{x}{\sum_{x=1}^{3} x} = \frac{x}{6}.$$

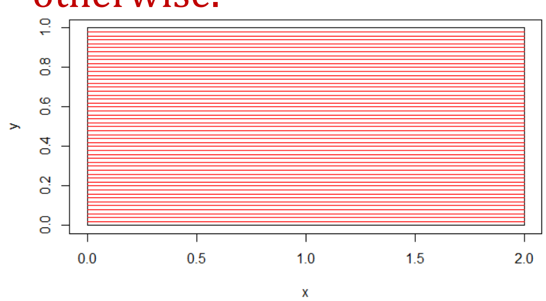
With the discussion above, try to figure out the solution for the following example: check that X and Y are independent and find the pdf of X and Y.

Example 5

- Given that

$$f_{X,Y}(x,y) = \begin{cases} \frac{1}{3}\, x(1+y), & \text{for } 0 < x < 2, 0 < y < 1, \\ 0, & \text{otherwise.} \end{cases}$$

- are X and Y independent?



ST2334 Probability and Statistics CYM 2-dim RVs & cond prob dist 3-100

Remarks

1. $Cov(X, Y) = E(XY) - \mu_X \mu_Y$.
2. If X and Y are independent, then $Cov(X, Y) = 0$. However $Cov(X, Y) = 0$ does not imply X and Y are independent.
3. $Cov(aX + b, cY + d) = ac\, Cov(X, Y)$
4. $V(aX + bY) = a^2V(X) + b^2V(Y) + 2ab\, Cov(X, Y)$

ST2334 Probability and Statistics CYM 2-dim RVs & cond prob dist 3-111

✓ The second point in the slide that X and Y independent leads to $Cov(X, Y) = 0$ can be easily derived based on the argument given on page 5. Try this derivation on your own.

Also try to compare $Cov(X, Y) = 0$ with that equivalent definition of X and Y given on page 5 to view that $Cov(X, Y) = 0$ is not enough to conclude the independence of X and Y.

This is an advanced information: there is a specific situation that $Cov(X, Y) = 0$ is equivalent to the independence of X and Y: (X, Y) follows a bivariate normal distribution.

✓ $Cov(aX+b, cY+d) = ac\, Cov(X, Y)$ given by point 3 in the slide can be viewed/remembered/derived as three formulae: for arbitrary random variables X and Y,

★ $Cov(X, Y) = Cov(Y, X)$;

★ $Cov(X + b, Y) = Cov(X, Y)$ for any real number b;

★ $Cov(aX, Y) = aCov(X, Y)$ for any real number a.

Check these three formulae on your own and figure out how they lead to the covariance formula given by point 3 in the slide.

✓ With point 3 and the formula we gave Chapter 2: $V(aX) = a^2V(X)$ for any real number a, Point 4 in the slide can be simplified to be $V(X+Y) = V(X) + V(Y) + 2Cov(X,Y)$.

✓ $V(X+Y) = V(X) + V(Y) + 2Cov(X,Y)$ can be extended to multiple random variables, namely $V(X_1 + X_2 + \ldots + X_n)$. This leads to the sum of n variance terms and $\binom{n}{2}$ covariance terms. However, with independence/uncorrelated assumption, this formula can be greatly simplified, as based on Point 2, all the covariance terms disappear; so we have if $X_1, X_2, \ldots, X_n$ are pairwise independent/uncorrelated,

$$V(X_1 \pm X_2 \pm \ldots \pm X_n) = V(X_1) + V(X_2) + \ldots + V(X_n).$$

Note: "±" on the left and "+" on the right are not typos. Think about why. Keep this formula in mind; it is very useful in our later on development.