Sirindhorn International Institute of Technology
Thammasat University
School of Information, Computer and Communication Technology

ECS315 2017/1 Part V Dr.Prapun

11 Multiple Random Variables

One is often interested not only in individual random variables, but also in relationships between two or more random variables. Furthermore, one often wishes to make inferences about one random variable on the basis of observations of other random variables.

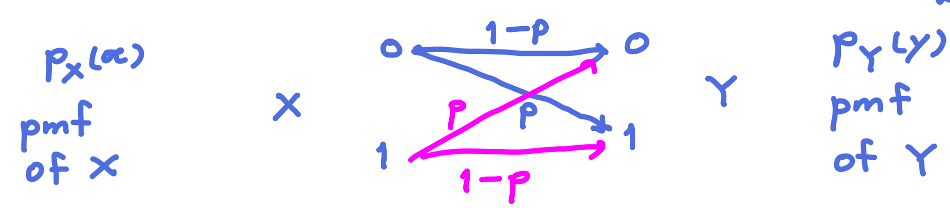
Example 11.1. If the experiment is the testing of a new medicine, the researcher might be interested in cholesterol level, blood pressure, and the glucose level of a test person.

11.1 A Pair of Discrete Random Variables

In this section, we consider two discrete random variables, say X and Y, simultaneously.

11.2. The analysis are different from Section 9.2 in two main aspects. First, there may be *no* deterministic relationship (such as $Y = g(X)$) between the two random variables. Second, we want to look at both random variables as a whole, not just X alone or Y alone.

Example 11.3. Communication engineers may be interested in the input X and output Y of a communication channel. $\boxed{\text{BSC}}$



11.4. Recall that, in probability, "," means "and". For example,

$$P[X = x, Y = y] = P[X = x \text{ and } Y = y]$$

$$P[X=3, Y=4] = P[X=3 \text{ and } Y=4]$$

and

$$\begin{aligned} P[3 \le X < 4, Y < 1] &= P[3 \le X < 4 \text{ and } Y < 1] \\ &= P[X \in [3,4) \text{ and } Y \in (-\infty, 1)]. \end{aligned}$$

In general, the event

$$[\text{"Some condition(s) on } X\text{"}, \text{"Some condition(s) on } Y\text{"}]$$

is the same as the intersection of two events:

$$[\text{"Some condition(s) on } X\text{"}] \cap [\text{"Some condition(s) on } Y\text{"}]$$

which simply means both statements happen.

More technically,

$$[X \in B, Y \in C] = [X \in B \text{ and } Y \in C] = [X \in B] \cap [Y \in C]$$

and

$$\begin{aligned} P[X \in B, Y \in C] &= P[X \in B \text{ and } Y \in C] \\ &= P([X \in B] \cap [Y \in C]). \end{aligned}$$

Remark: Linking back to the original sample space, this shorthand actually says

$$\begin{aligned} [X \in B, Y \in C] &= [X \in B \text{ and } Y \in C] \\ &= \{\omega \in \Omega : X(\omega) \in B \text{ and } Y(\omega) \in C\} \\ &= \{\omega \in \Omega : X(\omega) \in B\} \cap \{\omega \in \Omega : Y(\omega) \in C\} \\ &= [X \in B] \cap [Y \in C]. \end{aligned}$$

11.5. The concept of conditional probability can be straightforwardly applied to discrete random variables. For example,

$$P\left[\text{“Some condition(s) on } X\text{”} \mid \text{“Some condition(s) on } Y\text{”}\right] \tag{26}$$

is the conditional probability $P(A|B)$ where

$$\begin{aligned} A &= [\text{“Some condition(s) on } X\text{”] and} \\ B &= [\text{“Some condition(s) on } Y\text{”]}. \end{aligned}$$

Recall that $P(A|B) = P(A \cap B)/P(B)$. Therefore,

$$\overset{P(A|B)}{P\left[X = x|\,Y = y\right]} = \frac{\overset{P(A\cap B)}{P\left[X = x \text{ and } Y = y\right]}}{\underset{P(B)}{P\left[Y = y\right]}},$$

and

$$P\left[3 \le X < 4|\,Y < 1\right] = \frac{P\left[3 \le X < 4 \text{ and } Y < 1\right]}{P\left[Y < 1\right]}$$

More generally, (26) is

$$\begin{aligned} &= \frac{P\left([\text{“Some condition(s) on } X\text{”}] \cap [\text{“Some condition(s) on } Y\text{”}]\right)}{P\left([\text{“Some condition(s) on } Y\text{”}]\right)} \\ &= \frac{P\left([\text{“Some condition(s) on } X\text{”}, \text{“Some condition(s) on } Y\text{”}]\right)}{P\left([\text{“Some condition(s) on } Y\text{”}]\right)} \\ &= \frac{P[\text{“Some condition(s) on } X\text{”}, \text{“Some condition(s) on } Y\text{”}]}{P[\text{“Some condition(s) on } Y\text{”}]} \end{aligned}$$

More technically,

$$\begin{aligned} P\left[X \in B|\,Y \in C\right] = P\left([X \in B]\,|[Y \in C]\right) &= \frac{P\left([X \in B] \cap [Y \in C]\right)}{P\left([Y \in C]\right)} \\ &= \frac{P\left[X \in B, Y \in C\right]}{P\left[Y \in C\right]}. \end{aligned}$$

$p_X(x) = P[X = x]$

Definition 11.6. ***Joint pmf***: If X and Y are two discrete random variables (defined on a same sample space with probability measure P), the function $p_{X,Y}(x,y)$ defined by

Note that

$$p_{X,Y}(x,y) = P\left[X = x, Y = y\right] \qquad \sum_x \sum_y p_{X,Y}(x,y) = 1$$

is called the ***joint probability mass function*** of X and Y.

(a) We can visualize the joint pmf via stem plot. See Figure 32.

(b) To evaluate the probability for a statement that involves both X and Y random variables:

Ex. $P[XY > 1]$

$P[X + Y = 7]$

$P[\text{condition(s) on } X \text{ and } Y]$

We first find all pairs (x,y) that satisfy the condition(s) in the statement, and then add up all the corresponding values from the joint pmf .

More technically, we can then evaluate $P\left[(X,Y) \in R\right]$ by

$$P\left[(X,Y) \in R\right] = \sum_{(x,y):(x,y)\in R} p_{X,Y}(x,y).$$

Example 11.7 (F2011). Consider random variables X and Y whose joint pmf is given by

$$p_{X,Y}(x,y) = \begin{cases} c\,(x+y), & x \in \{1,3\} \text{ and } y \in \{2,4\}, \\ 0, & \text{otherwise.} \end{cases}$$

x	y
1	2
1	4
3	2
3	4

x	y	$p_{X,Y}(x,y)$	x^2+y^2
1	2	$3c$	$\boxed{5}$
1	4	$5c$	$\boxed{17}$
3	2	$5c$	$\boxed{13}$
3	4	$7c$	25

(a) Check that $c = 1/20$.

$\sum_{(x,y)}$

$\sum_x \sum_y p_{X,Y}(x,y) = 1$

$20c = 1 \Rightarrow c = \frac{1}{20}$

(b) Find $P\left[X^2 + Y^2 = 13\right]$.

$= 5c = 5 \times \frac{1}{20} = \frac{1}{4}$

(c) $P\left[X^2 + Y^2 < 20\right] = 3c + 5c + 5c$

$= 13c = 13/20$

In most situation, it is much more convenient to focus on the "important" part of the joint pmf. To do this, we usually present the joint pmf (and the conditional pmf) in their matrix forms:

$$\mathbf{P}_{X,Y} = \begin{array}{c|cc} x \backslash y & 2 & 4 \\ \hline 1 & \boxed{3c} & \boxed{5c} \\ 3 & \boxed{5c} & 7c \end{array} \qquad \begin{array}{c|cc} x^2+y^2 & & \\ x \backslash y & 2 & 4 \\ \hline 1 & \boxed{5} & \boxed{17} \\ 3 & \boxed{13} & 25 \end{array}$$

Definition 11.8. When both X and Y take finitely many values (both have finite supports), say $S_X = \{x_1, \ldots, x_m\}$ and $S_Y = \{y_1, \ldots, y_n\}$, respectively, we can arrange the probabilities $p_{X,Y}(x_i, y_j)$ in an $m \times n$ matrix

$$\begin{array}{cc} x \backslash y & \begin{array}{cccc} y_1 & y_2 & & y_n \end{array} \\ \begin{array}{c} x_1 \\ x_2 \\ \vdots \\ x_m \end{array} & \begin{bmatrix} p_{X,Y}(x_1, y_1) & p_{X,Y}(x_1, y_2) & \ldots & p_{X,Y}(x_1, y_n) \\ p_{X,Y}(x_2, y_1) & p_{X,Y}(x_2, y_2) & \ldots & p_{X,Y}(x_2, y_n) \\ \vdots & \vdots & \ddots & \vdots \\ p_{X,Y}(x_m, y_1) & p_{X,Y}(x_m, y_2) & \ldots & p_{X,Y}(x_m, y_n) \end{bmatrix} \end{array}. \qquad (27)$$

- We shall call this matrix the ***joint pmf matrix***.
- The sum of all the entries in the matrix is one.

$$\sum_x \sum_y p_{X,Y}(x,y) = 1 \qquad \sum_{i=1}^{m} \sum_{j=1}^{n} (P_{X,Y})_{i,j}$$

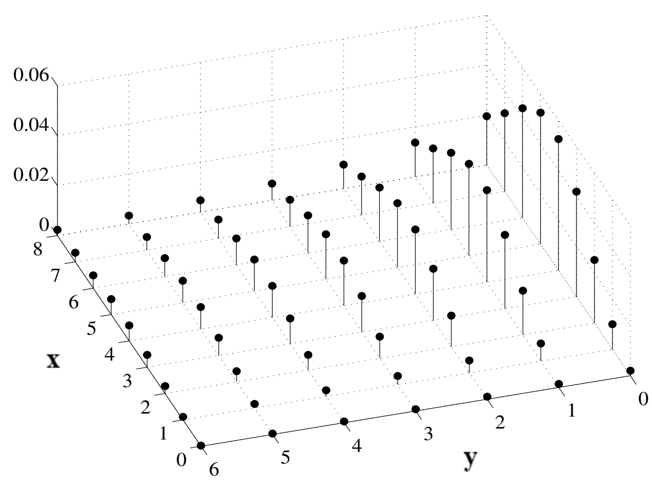


Figure 32: Example of the plot of a joint pmf. [9, Fig. 2.8]

- $p_{X,Y}(x, y) = 0$ if[52] $x \notin S_X$ or $y \notin S_Y$. In other words, we don't have to consider the x and y outside the supports of X and Y, respectively.

[52]To see this, note that $p_{X,Y}(x, y)$ can not exceed $p_X(x)$ because $P(A \cap B) \le P(A)$. Now, suppose at $x = a$, we have $p_X(a) = 0$. Then $p_{X,Y}(a, y)$ must also $= 0$ for any y because it can not exceed $p_X(a) = 0$. Similarly, suppose at $y = a$, we have $p_Y(a) = 0$. Then $p_{X,Y}(x, a) = 0$ for any x.

11.9. From the joint pmf, we can find $p_X(x)$ and $p_Y(y)$ by

$$p_X(x) = \sum_y p_{X,Y}(x,y) \tag{28}$$

$$p_Y(y) = \sum_x p_{X,Y}(x,y) \tag{29}$$

In this setting, $p_X(x)$ and $p_Y(y)$ are call the ***marginal pmfs*** (to distinguish them from the joint one).

(a) Suppose we have the joint pmf matrix in (27). Then, the sum of the entries in the ith row is[53] $p_X(x_i)$, and
the sum of the entries in the jth column is $p_Y(y_j)$:

$$p_X(x_i) = \sum_{j=1}^{n} p_{X,Y}(x_i, y_j) \quad \text{and} \quad p_Y(y_j) = \sum_{i=1}^{m} p_{X,Y}(x_i, y_j)$$

(b) In MATLAB, suppose we save the joint pmf matrix as `P_XY`, then the marginal pmf (row) vectors `p_X` and `p_Y` can be found by

```
p_X = (sum(P_XY,2))'
p_Y = (sum(P_XY,1))
```

Example 11.10. Consider the following joint pmf matrix

$$\begin{array}{c|c} x \backslash y & \begin{array}{cccc} 0 & 1 & 2 & 3 \end{array} \\ \begin{array}{c} 0 \\ 1 \\ 2 \end{array} & \begin{bmatrix} 0.1 & 0 & 0.2 & 0 \\ 0 & 0.5 & 0 & 0 \\ 0 & 0.1 & 0.1 & 0 \end{bmatrix} \end{array} \begin{array}{l} \xrightarrow{\Sigma} 0.3 \\ \xrightarrow{\Sigma} 0.5 \\ \xrightarrow{\Sigma} 0.2 \end{array}$$

$$\Sigma \downarrow \quad 0.1 \qquad 0.6 \qquad 0.3 \qquad 0$$

$$p_X(x) = \begin{cases} 0.3, & x = 0, \\ 0.5, & x = 1, \\ 0.2, & x = 2, \\ 0, & \text{otherwise.} \end{cases}$$

$$p_Y(y) = \begin{cases} 0.1, & y = 0, \\ 0.6, & y = 1, \\ 0.3, & y = 2, \\ 0, & \text{otherwise.} \end{cases}$$

(a) $p_{X,Y}(1,1) = 0.5$
$= P[X = 1, Y = 1]$

(b) $P[XY > 2] = 0.1 + 0 + 0 = 0.1$

$$xy: \quad \begin{array}{c|c} x \backslash y & \begin{array}{cccc} 0 & 1 & 2 & 3 \end{array} \\ \begin{array}{c} 0 \\ 1 \\ 2 \end{array} & \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 2 & 3 \\ 0 & 2 & 4 & 6 \end{bmatrix} \end{array}$$

(c) $p_X(2) = P[X = 2]$
$= 0 + 0.1 + 0.1 + 0 = 0.2$

[53] To see this, we consider $A = [X = x_i]$ and a collection defined by $B_j = [Y = y_j]$ and $B_0 = [Y \notin S_Y]$. Note that the collection $B_0, B_1, \ldots, B_n$ partitions Ω. So, $P(A) = \sum_{j=0}^{n} P(A \cap B_j)$. Of course, because the support of Y is S_Y, we have $P(A \cap B_0) = 0$. Hence, the sum can start at $j = 1$ instead of $j = 0$.

Definition 11.11. The ***conditional pmf*** of X given Y is defined as

$$p_{X|Y}(x|y) = P[\overbrace{X = x}^{A}|\overbrace{Y = y}^{B}] = P(A|B) = \frac{P(A\cap B)}{P(B)}$$

$$= \frac{P[X=x, Y=y]}{P[Y=y]} = \frac{p_{X,Y}(x,y)}{p_Y(y)}$$

which gives

$$p_{X,Y}(x, y) = p_{X|Y}(x|y)p_Y(y) = p_{Y|X}(y|x)p_X(x). \tag{30}$$

11.12. Equation (30) is quite important in practice. In most cases, systems are naturally defined/given/studied in terms of their conditional probabilities, say $p_{Y|X}(y|x)$. Therefore, it is important the we know how to construct the joint pmf from the conditional pmf.

Example 11.13. Consider a binary symmetric channel. Suppose the input X to the channel is Bernoulli(0.3). At the output Y of this channel, the crossover (bit-flipped) probability is 0.1. Find the joint pmf $p_{X,Y}(x, y)$ of X and Y.

Setup:

Input: $p_X(x) = \begin{cases} 0.3, & x = 1, \\ 0.7, & x = 0, \\ 0, & \text{otherwise} \end{cases}$

Channel:

$p_{Y|X}(0|1) \equiv P[Y=0|X=1] = 0.1$

$$p_{Y|X}(y|x) = \begin{cases} 0.1, & (x=0, y=1), (x=1, y=0), \\ 0.9, & (x=0, y=0), (x=1, y=1), \\ 0, & \text{otherwise.} \end{cases}$$

$$p_{X,Y}(1,0) \equiv P[\overbrace{X=1}^{A}, \overbrace{Y=0}^{B}] = P(A\cap B)$$

$$P(B|A) = \frac{P(A\cap B)}{P(A)}$$

$$P(A\cap B) = P(B|A)P(A)$$

$$= P[Y=0|X=1]\,P[X=1]$$

$$= p_{Y|X}(0|1)\,p_X(1)$$

$$= 0.1 \times 0.3 = 0.03$$

$$p_{X,Y}(x,y) = p_{Y|X}(y|x)p_X(x)$$

$$p_{Y|X} \quad \begin{array}{c|cc} x\backslash y & 0 & 1 \\ \hline 0 & 0.9 & 0.1 \\ 1 & 0.1 & 0.9 \end{array} \quad \begin{array}{l} \xrightarrow{\times p_X(0) = 0.7} \\ \xrightarrow[\times p_X(1) = 0.3]{} \end{array} \quad p_{X,Y} \quad \begin{array}{c|cc} x\backslash y & 0 & 1 \\ \hline 0 & 0.63 & 0.07 \\ 1 & 0.03 & 0.27 \end{array}$$

$$p_{X,Y}(x,y) = \begin{cases} 0.63, & x=0, y=0, \\ 0.07, & x=0, y=1, \\ 0.03, & x=1, y=0, \\ 0.27, & x=1, y=1, \\ 0, & \text{otherwise.} \end{cases}$$

Exercise 11.14 (F2011). Continue from Example 11.7. Random variables X and Y have the following joint pmf

$$p_{X,Y}(x,y) = \begin{cases} c(x+y), & x \in \{1,3\} \text{ and } y \in \{2,4\}, \\ 0, & \text{otherwise.} \end{cases}$$

(a) Find $p_X(x)$.

(b) Find $\mathbb{E}X$.

(c) Find $p_{Y|X}(y|1)$. Note that your answer should be of the form

$$p_{Y|X}(y|1) = \begin{cases} ?, & y = 2, \\ ?, & y = 4, \\ 0, & \text{otherwise.} \end{cases}$$

(d) Find $p_{Y|X}(y|3)$.

Definition 11.15. The ***joint cdf*** of X and Y is defined by

$$F_{X,Y}(x,y) = P[X \le x, Y \le y].$$

Definition 11.16. Two random variables X and Y are said to be ***identically distributed*** if, for every B, $P[X \in B] = P[Y \in B]$.

In words, for any probability statement about X (and only X), if we replace X by Y, we get the same probability.

Example 11.17. Roll a dice twice. Let X be the result from the first roll. Let Y be the result from the second roll.

- X and Y are not the same. (Most of the time, they will be different. By chance, they occasionally take the same value.)
- $P[X > 3] = P[Y > 3]$
  $P[|X| = 1] = P[|Y| = 1]$
  $P[X = 8] = P[Y = 8]$

Example 11.18. Let $X \sim$ Bernoulli(1/2). Let $Y = X$ and $Z = 1 - X$. Then, all of these random variables are identically distributed.

11.19. The following statements are equivalent:

(a) Random variables X and Y are ***identically distributed***.

(b) For every B, $P[\underbrace{X \in B}_{\text{event about } X}] = P[Y \in B]$ ← Powerful property. Difficult to check.

(c) $p_X(c) = p_Y(c)$ for all c ← Easier to check.

(d) $F_X(c) = F_Y(c)$ for all c

Definition 11.20. Two random variables X and Y are said to be ***independent*** if the events $[X \in B]$ and $[Y \in C]$ are independent for all sets B and C.

11.21. The following statements are equivalent:

(a) Random variables X and Y are ***independent***.

(b) $[X \in B] \perp\!\!\!\perp [Y \in C]$ for all B, C. ← Powerful property. Difficult to check.

(c) $P[X \in B, Y \in C] = P[X \in B] \times P[Y \in C]$ for all B, C.

(d) $p_{X,Y}(x, y) = p_X(x) \times p_Y(y)$ for all x, y. ← Easier to check.

(e) $F_{X,Y}(x, y) = F_X(x) \times F_Y(y)$ for all x, y.

Definition 11.22. Two random variables X and Y are said to be ***independent and identically distributed (i.i.d.)*** if X and Y are both independent and identically distributed.

11.23. Being identically distributed does not imply independence. Similarly, being independent, does not imply being identically distributed.

Example 11.24. Roll a dice. Let X be the result. Set $Y = X$.

Example 11.25. Suppose the pmf of a random variable X is given by

$$p_X(x) = \begin{cases} 1/4, & x = 3, \\ \alpha, & x = 4, \\ 0, & \text{otherwise.} \end{cases}$$

Let Y be another random variable. Assume that X and Y are i.i.d.

Find

(a) α, $\sum_x p_X(x) = 1 \Rightarrow \frac{1}{4} + \alpha + 0 = 1 \Rightarrow \alpha = \frac{3}{4}$

(b) the pmf of Y, and $p_Y(y) = \begin{cases} 1/4, & y = 3, \\ 3/4, & y = 4, \\ 0, & \text{otherwise.} \end{cases}$

(c) the joint pmf of X and Y.

$p_{X,Y}(x,y) = p_X(x)\, p_Y(y)$ for all x, y

$$P_{X,Y} \quad \begin{array}{c|cc} x \backslash y & 3 & 4 \\ \hline 3 & \frac{1}{4}\times\frac{1}{4} & \frac{1}{4}\times\frac{3}{4} \\ 4 & \frac{3}{4}\times\frac{1}{4} & \frac{3}{4}\times\frac{3}{4} \end{array} = \begin{array}{c|cc} x \backslash y & 3 & 4 \\ \hline 3 & 1/16 & 3/16 \\ 4 & 3/16 & 9/16 \end{array}$$

$$p_{X,Y}(x,y) = \begin{cases} 1/16, & x=3, y=3, \\ 3/16, & (x=3, y=4) \text{ or } (x=4, y=3) \\ 9/16, & x=4, y=4, \\ 0, & \text{otherwise.} \end{cases}$$

Example 11.26. Consider a pair of random variables X and Y whose joint pmf is given by

$$p_{X,Y}(x,y) = \begin{cases} 1/15, & x = 3, y = 1, \\ 2/15, & x = 4, y = 1, \\ 4/15, & x = 3, y = 3, \\ \beta, \ 8/15 & x = 4, y = 3, \\ 0, & \text{otherwise.} \end{cases}$$

(a) Are X and Y identically distributed? No. (Different possible values)

(b) Are X and Y independent?

$$p_{X,Y} \quad \begin{array}{c|cc} x \backslash y & 1 & 3 \\ \hline 3 & 1/15 & 4/15 \\ 4 & 2/15 & 8/15 \end{array} \quad \begin{array}{l} \xrightarrow{\Sigma} 1/3 \\ \xrightarrow{\Sigma} 2/3 \end{array}$$

$$\begin{array}{cc} \Sigma\downarrow & \Sigma\downarrow \\ 1/5 & 4/5 \end{array}$$

$$p_X(x) = \begin{cases} 1/3, & x = 3, \\ 2/3, & x = 4, \\ 0, & \text{otherwise.} \end{cases}$$

$$p_Y(y) = \begin{cases} 1/5, & y = 1, \\ 4/5, & y = 3, \\ 0, & \text{otherwise.} \end{cases}$$

11.2 Extending the Definitions to Multiple RVs

Definition 11.27. Joint pmf:

$$p_{X_1,X_2,\ldots,X_n}(x_1,x_2,\ldots,x_n) = P\left[X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n\right].$$

Joint cdf:

$$F_{X_1,X_2,\ldots,X_n}(x_1,x_2,\ldots,x_n) = P\left[X_1 \le x_1, X_2 \le x_2, \ldots, X_n \le x_n\right].$$

11.28. Marginal pmf:

$$p_X(x) = \sum_y \sum_z p_{X,Y,Z}(x,y,z)$$

Example 11.29. Consider three random variables X, Y, and Z whose joint pmf is given by

$$p_{X,Y,Z}(x,y,z) = \begin{cases} 1/7, & (x,y,z) \in \left\{(0,1,0), (1,1,1)\right\}, \\ 2/7, & (x,y,z) = (0,0,1), \\ 3/7, & (x,y,z) = (0,1,1), \\ 0, & \text{otherwise.} \end{cases}$$

Then,

$$\begin{aligned} p_X(0) &\equiv P\left[X = 0\right] = \\ p_X(1) &\equiv P\left[X = 1\right] = \end{aligned}$$

Therefore,

$$p_X(x) = \begin{cases} \quad, & x = 0, \\ \quad, & x = 1, \\ 0, & \text{otherwise.} \end{cases}$$

Definition 11.30. ***Identically distributed*** random variables: The following statements are equivalent.

(a) Random variables $X_1, X_2, \ldots$ are ***identically distributed***

(b) For every B, $P\left[X_j \in B\right]$ does not depend on j.

(c) $p_{X_i}(c) = p_{X_j}(c)$ for all c, i, j.

(d) $F_{X_i}(c) = F_{X_j}(c)$ for all c, i, j.

Definition 11.31. ***Independence*** among finite number of random variables: The following statements are equivalent.

(a) $X_1, X_2, \ldots, X_n$ are ***independent***

(b) $[X_1 \in B_1], [X_2 \in B_2], \ldots, [X_n \in B_n]$ are independent, for all $B_1, B_2, \ldots, B_n$.

(c) $P[X_i \in B_i, \forall i] = \prod_{i=1}^{n} P[X_i \in B_i]$, for all $B_1, B_2, \ldots, B_n$.

(d) $p_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} p_{X_i}(x_i)$ for all $x_1, x_2, \ldots, x_n$.

(e) $F_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} F_{X_i}(x_i)$ for all $x_1, x_2, \ldots, x_n$.

Example 11.32. Toss a coin n times. For the ith toss, let

$$X_i = \begin{cases} 1, & \text{if H happens on the } i\text{th toss,} \\ 0, & \text{if T happens on the } i\text{th toss.} \end{cases}$$

We then have a collection of i.i.d. random variables $X_1, X_2, X_3, \ldots, X_n$.

11.33. Fact: For i.i.d. $X_i \sim$ Bernoulli(p), $Y = X_1 + X_2 + \cdots + X_n$ is $\mathcal{B}(n, p)$.

To see this, consider n (independent) Bernoulli trials (as in Example 11.32). Let

$$X_i = \begin{cases} 1, & \text{if success happens on the } i\text{th trial,} \\ 0, & \text{if failure happens on the } i\text{th trial.} \end{cases}$$

Then, Y is simply counting the number of successes in the n trials. From Definition 8.32 of Binomial RV, we conclude that Y is binomial.

Example 11.34. Roll a dice n times. Let N_i be the result of the ith roll. We then have another collection of i.i.d. random variables $N_1, N_2, N_3, \ldots, N_n$.

Example 11.35. Let X_1 be the result of tossing a biased coin. Set $X_2 = X_3 = \cdots = X_n = X_1$.

11.36. If $X_1, X_2, \ldots, X_n$ are independent, then so is any subcollection of them.

Definition 11.37. A ***pairwise independent*** collection of random variables is a collection of random variables any two of which are independent.

(a) Any collection of (mutually) independent random variables is pairwise independent

(b) Some pairwise independent collections are not independent. See Example (11.38).

Example 11.38. Let suppose X, Y, and Z have the following joint probability distribution: $p_{X,Y,Z}(x,y,z) = \frac{1}{4}$ for $(x,y,z) \in \{(0,0,0),(0,1,1),(1,0,1),(1,1,0)\}$. This, for example, can be constructed by starting with independent X and Y that are Bernoulli-$\frac{1}{2}$. Then set $Z = X \oplus Y = X + Y \mod 2$.

(a) X, Y, Z are pairwise independent.

(b) X, Y, Z are not independent.

11.3 Expectation of Function of Discrete Random Variables

11.39. Recall that the expected value of "any" function g of a discrete random variable X can be calculated from

$$\mathbb{E}[g(X)] = \sum_x g(x) p_X(x).$$

Similarly[54], the expected value of "any" function g of two discrete random variables X and Y can be calculated from

$$\mathbb{E}[g(X,Y)] = \sum_x \sum_y g(x,y) p_{X,Y}(x,y).$$

[54]Again, these are called the **law/rule of the lazy statistician** (LOTUS) [22, Thm 3.6 p 48],[9, p. 149] because it is so much easier to use the above formula than to first find the pmf of $g(X)$ or $g(X,Y)$. It is also called **substitution rule** [21, p 271].

	Discrete
$P[X \in B]$	$\sum_{x \in B} p_X(x)$
$P[(X,Y) \in R]$	$\sum_{(x,y):(x,y) \in R} p_{X,Y}(x,y)$
Joint to Marginal: (Law of Total Prob.)	$p_X(x) = \sum_y p_{X,Y}(x,y)$ $p_Y(y) = \sum_x p_{X,Y}(x,y)$
$P[X > Y]$	$\sum_x \sum_{y:\, y<x} p_{X,Y}(x,y)$ $= \sum_y \sum_{x:\, x>y} p_{X,Y}(x,y)$
$P[X = Y]$	$\sum_x p_{X,Y}(x,x)$
$X \perp\!\!\!\perp Y$	$p_{X,Y}(x,y) = p_X(x)p_Y(y)$
Conditional	$p_{X\|Y}(x\|y) = \frac{p_{X,Y}(x,y)}{p_Y(y)}$
$\mathbb{E}[g(X,Y)]$	$\sum_x \sum_y g(x,y) p_{X,Y}(x,y)$

Table 8: Joint pmf: A Summary

11.40. $\mathbb{E}[\cdot]$ is a **linear** operator: $\mathbb{E}[aX + bY] = a\mathbb{E}X + b\mathbb{E}Y$.

$\mathbb{E}[3X + 5Y] = 3\mathbb{E}X + 5\mathbb{E}Y$

(a) Homogeneous: $\mathbb{E}[cX] = c\mathbb{E}X$

(b) Additive: $\mathbb{E}[X + Y] = \mathbb{E}X + \mathbb{E}Y$

(c) Extension: $\mathbb{E}\left[\sum_{i=1}^n c_i g_i(X_i)\right] = \sum_{i=1}^n c_i \mathbb{E}[g_i(X_i)]$.

Ex. $\mathbb{E}[3X^2 + 8\sqrt{Y}] = 3\mathbb{E}[X^2] + 8\mathbb{E}[\sqrt{Y}]$

Example 11.41. Recall from 11.33 that when i.i.d. $X_i \sim \text{Bernoulli}(p)$, $Y = X_1 + X_2 + \cdots + X_n$ is $\mathcal{B}(n,p)$. Also, from Example 9.4, we have $\mathbb{E}X_i = p$. Hence,

$$\mathbb{E}Y = \mathbb{E}\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n \mathbb{E}[X_i] = \sum_{i=1}^n p = np.$$

Therefore, the expectation of a binomial random variable with parameters n and p is np.

Example 11.42. A binary communication link has bit-error probability p. What is the expected number of bit errors in a transmission of n bits?

Theorem 11.43 (Expectation and Independence)**.** Two random variables X and Y are independent if and only if

$$\mathbb{E}\left[h(X)g(Y)\right] = \mathbb{E}\left[h(X)\right]\mathbb{E}\left[g(Y)\right]$$

for "all" functions h and g.

Ex. If $X \perp\!\!\!\perp Y$, then

$$\mathbb{E}\left[(X^2-3)^4(\sin Y)\right] = \mathbb{E}\left[(X^2-3)^4\right] \times \mathbb{E}\left[\sin Y\right]$$

- In other words, X and Y are independent if and only if for every pair of functions h and g, the expectation of the product $h(X)g(Y)$ is equal to the product of the individual expectations.

- One special case is that

$$X \perp\!\!\!\perp Y \quad \text{implies} \quad \mathbb{E}\left[XY\right] = \mathbb{E}X \times \mathbb{E}Y. \tag{31}$$

  However, independence means more than this property. In other words, having $\mathbb{E}\left[XY\right] = (\mathbb{E}X)(\mathbb{E}Y)$ does not necessarily imply $X \perp\!\!\!\perp Y$. See Example 11.54.

11.44. Let's combined what we have just learned about independence into the definition/equivalent statements that we already have in 11.~~3~~21.

The following statements are equivalent:

(a) Random variables X and Y are ***independent***.

(b) $[X \in B] \perp\!\!\!\perp [Y \in C]$ for all B, C.

(c) $P\left[X \in B, Y \in C\right] = P\left[X \in B\right] \times P\left[Y \in C\right]$ for all B, C.

(d) $p_{X,Y}(x, y) = p_X(x) \times p_Y(y)$ for all x, y.

(e) $F_{X,Y}(x, y) = F_X(x) \times F_Y(y)$ for all x, y.

(f) $\mathbb{E}\left[h(X)g(Y)\right] = \mathbb{E}\left[h(X)\right]\mathbb{E}\left[g(Y)\right]$ for all h, g.

(g) $I(X;Y) = 0$

Exercise 11.45 (F2011)**.** Suppose X and Y are i.i.d. with $\mathbb{E}X = \mathbb{E}Y = 1$ and $\operatorname{Var} X = \operatorname{Var} Y = 2$. Find $\operatorname{Var}[XY]$.

11.46. To quantify the amount of ***dependence*** between two random variables, we may calculate their ***mutual information***. This quantity is crucial in the study of digital communications and information theory. However, in introductory probability class (and introductory communication class), it is traditionally omitted.

11.4 Linear Dependence

Definition 11.47. Given two random variables X and Y, we may calculate the following quantities:

(a) **Correlation**: $\mathbb{E}[XY]$. $= \sum_x \sum_y xy\, p_{X,Y}(x,y) = \mathbb{E}[YX]$

(b) **Covariance**: $\operatorname{Cov}[X,Y] = \mathbb{E}[(X - \mathbb{E}X)(Y - \mathbb{E}Y)]$.

(c) **Correlation coefficient**: $\rho_{X,Y} = \frac{\operatorname{Cov}[X,Y]}{\sigma_X \sigma_Y}$

Exercise 11.48 (F2011)**.** Continue from Exercise 11.7.

(a) Find $\mathbb{E}[XY]$.

(b) Check that $\operatorname{Cov}[X,Y] = -\frac{1}{25}$.

$\operatorname{Cov}[X,X] = \mathbb{E}[XX] - \mathbb{E}X\,\mathbb{E}X$
$= \mathbb{E}[X^2] - (\mathbb{E}X)^2$

11.49. $\operatorname{Cov}[X,Y] = \mathbb{E}[(X - \underbrace{\mathbb{E}X}_{m_X})(Y - \underbrace{\mathbb{E}Y}_{m_Y})] = \mathbb{E}[XY] - \mathbb{E}X\mathbb{E}Y$

$= \mathbb{E}[(X - m_X)(Y - m_Y)] = \mathbb{E}[XY - Ym_X - Xm_Y + m_Xm_Y]$
$= \mathbb{E}[XY] - m_X\mathbb{E}Y - \cancel{m_Y\mathbb{E}X} + \cancel{m_Xm_Y}$

$\operatorname{Cov}[X,X]$
$= \mathbb{E}[(X-\mathbb{E}X)(X-\mathbb{E}X)]$
$= \mathbb{E}[(X-\mathbb{E}X)^2]$
$= \operatorname{Var} X$

- Note that $\operatorname{Var} X = \operatorname{Cov}[X,X]$.

11.50. $\operatorname{Var}[\underbrace{X+Y}_{Z}] = \operatorname{Var} X + \operatorname{Var} Y + 2\operatorname{Cov}[X,Y]$

$= \mathbb{E}[(Z - \mathbb{E}Z)^2] = \mathbb{E}[(X + Y - (\mathbb{E}X + \mathbb{E}Y))^2]$

$= \mathbb{E}[(\underbrace{(X - \mathbb{E}X)}_{A} + \underbrace{(Y - \mathbb{E}Y)}_{B})^2] = \mathbb{E}[(A+B)^2]$

$= \mathbb{E}[A^2 + B^2 + 2AB]$

Definition 11.51. X and Y are said to be ***uncorrelated*** if and only if $\operatorname{Cov}[X, Y] = 0$.

11.52. The following statements are equivalent:

(a) X and Y are ***uncorrelated***.

(b) $\operatorname{Cov}[X, Y] = 0$.

(c) $\mathbb{E}[XY] = \mathbb{E}X\mathbb{E}Y$.

(d) $\mathbb{E}[(X-\mathbb{E}X)(Y-\mathbb{E}Y)] = 0$

11.53. Independence implies uncorrelatedness; that is if $X \perp\!\!\!\perp Y$, then $\operatorname{Cov}[X, Y] = 0$.

The converse is not true. Uncorrelatedness does not imply independence. See Example 11.54.

Example 11.54. Let X be uniform on $\{\pm 1, \pm 2\}$ and $Y = |X|$.

11.55. The variance of the sum of uncorrelated (or independent) random variables is the sum of their variances.

$$\operatorname{Var}[X+Y] = \operatorname{Var} X + \operatorname{Var} Y + \cancel{2\operatorname{Cov}[X,Y]} \to 0$$

Exercise 11.56. Suppose two fair dice are tossed. Denote by the random variable V_1 the number appearing on the first dice and by the random variable V_2 the number appearing on the second dice. Let $X = V_1 + V_2$ and $Y = V_1 - V_2$.

(a) Show that X and Y are not independent.

(b) Show that $\mathbb{E}[XY] = \mathbb{E}X\mathbb{E}Y$.

11.57. $\text{Cov}[aX + b, cY + d] = ac\text{Cov}[X, Y]$

$$\begin{aligned}
\text{Cov}[aX + b, cY + d] &= \mathbb{E}[((aX + b) - \mathbb{E}[aX + b])((cY + d) - \mathbb{E}[cY + d])] \\
&= \mathbb{E}[((aX + b) - (a\mathbb{E}X + b))((cY + d) - (c\mathbb{E}Y + d))] \\
&= \mathbb{E}[(aX - a\mathbb{E}X)(cY - c\mathbb{E}Y)] \\
&= ac\mathbb{E}[(X - \mathbb{E}X)(Y - \mathbb{E}Y)] \\
&= ac\text{Cov}[X, Y].
\end{aligned}$$

Definition 11.58. ***Correlation coefficient***:

$$\begin{aligned}
\rho_{X,Y} &= \frac{\text{Cov}[X, Y]}{\sigma_X \sigma_Y} \\
&= \mathbb{E}\left[\left(\frac{X - \mathbb{E}X}{\sigma_X}\right)\left(\frac{Y - \mathbb{E}Y}{\sigma_Y}\right)\right] = \frac{\mathbb{E}[XY] - \mathbb{E}X\mathbb{E}Y}{\sigma_X \sigma_Y}.
\end{aligned}$$

- $\rho_{X,Y}$ is dimensionless
- $\rho_{X,X} = 1$
- $\rho_{X,Y} = 0$ if and only if X and Y are uncorrelated.
- ***Cauchy-Schwartz Inequality***[55]:

$$|\rho_{X,Y}| \le 1.$$

In other words, $\rho_{XY} \in [-1, 1]$.

[55]Cauchy-Schwartz inequality shows up in many areas of Mathematics. A general form of this inequality can be stated in any inner product space:

$$|\langle a, b \rangle|^2 \le \langle a, a \rangle \langle b, b \rangle.$$

Here, the inner product is defined by $\langle X, Y \rangle = \mathbb{E}[XY]$. The Cauchy-Schwartz inequality then gives

$$|\mathbb{E}[XY]|^2 \le \mathbb{E}[X^2]\mathbb{E}[Y^2].$$

11.59. Linear Dependence and Cauchy-Schwartz Inequality

(a) If $Y = aX + b$, then $\rho_{X,Y} = \operatorname{sign}(a) = \begin{cases} 1, & a > 0 \\ -1, & a < 0. \end{cases}$

- To be rigorous, we should also require that $\sigma_X > 0$ and $a \neq 0$.

(b) When $\sigma_Y, \sigma_X > 0$, equality occurs in the Cauchy-Schwartz inequality if and only if the following conditions holds

$\equiv \exists a \neq 0$ such that $(X - \mathbb{E}X) = a(Y - \mathbb{E}Y)$
$\equiv \exists a \neq 0$ and $b \in \mathbb{R}$ such that $X = aY + b$
$\equiv \exists c \neq 0$ and $d \in \mathbb{R}$ such that $Y = cX + d$
$\equiv |\rho_{XY}| = 1$

In which case, $|a| = \frac{\sigma_X}{\sigma_Y}$ and $\rho_{XY} = \frac{a}{|a|} = \operatorname{sgn} a$. Hence, ρ_{XY} is used to quantify **linear dependence** between X and Y. The closer $|\rho_{XY}|$ to 1, the higher degree of linear dependence between X and Y.

Example 11.60. [21, Section 5.2.3] Consider an important fact that *investment experience* supports: spreading investments over a variety of funds (diversification) diminishes risk. To illustrate, imagine that the random variable X is the return on every invested dollar in a local fund, and random variable Y is the return on every invested dollar in a foreign fund. Assume that random variables X and Y are i.i.d. with expected value 0.15 and standard deviation 0.12.

If you invest all of your money, say c, in either the local or the foreign fund, your return R would be cX or cY.

- The expected return is $\mathbb{E}R = c\mathbb{E}X = c\mathbb{E}Y = 0.15c$.
- The standard deviation is $c\sigma_X = c\sigma_Y = 0.12c$

Now imagine that your money is equally distributed over the two funds. Then, the return R is $\frac{1}{2}cX + \frac{1}{2}cY$. The expected return is $\mathbb{E}R = \frac{1}{2}c\mathbb{E}X + \frac{1}{2}c\mathbb{E}Y = 0.15c$. Hence, the expected return remains at 15%. However,

$$\operatorname{Var} R = \operatorname{Var}\left[\frac{c}{2}(X+Y)\right] = \frac{c^2}{4}\operatorname{Var} X + \frac{c^2}{4}\operatorname{Var} Y = \frac{c^2}{2} \times 0.12^2.$$

So, the standard deviation is $\frac{0.12}{\sqrt{2}}c \approx 0.0849c$.

In comparison with the distributions of X and Y, the pmf of $\frac{1}{2}(X+Y)$ is concentrated more around the expected value. The centralization of the distribution as random variables are averaged together is a manifestation of the central limit theorem.

11.61. [21, Section 5.2.3] Example 11.60 is based on the assumption that return rates X and Y are independent from each other. In the world of investment, however, risks are more commonly reduced by combining negatively correlated funds (two funds are negatively correlated when one tends to go up as the other falls).

This becomes clear when one considers the following hypothetical situation. Suppose that two stock market outcomes ω_1 and ω_2 are possible, and that each outcome will occur with a probability of $\frac{1}{2}$ Assume that domestic and foreign fund returns X and Y are determined by $X(\omega_1) = Y(\omega_2) = 0.25$ and $X(\omega_2) = Y(\omega_1) = -0.10$. Each of the two funds then has an expected return of 7.5%, with equal probability for actual returns of 25% and -10%. The random variable $Z = \frac{1}{2}(X+Y)$ satisfies $Z(\omega_1) = Z(\omega_2) = 0.075$. In other words, Z is equal to 0.075 with certainty. This means that an investment that is equally divided between the domestic and foreign funds has a guaranteed return of 7.5%.