

## Independence of random variables

We say that random variables  $X$  and  $Y$  are **independent** if

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

that is, if we can **factor** the joint pmf into a pmf that depends only on  $X$  and a pmf that depends only on  $Y$ .

This is the same as saying that the *events*  $\{X = x\}$  and  $\{Y = y\}$  are independent for all  $x, y$ .

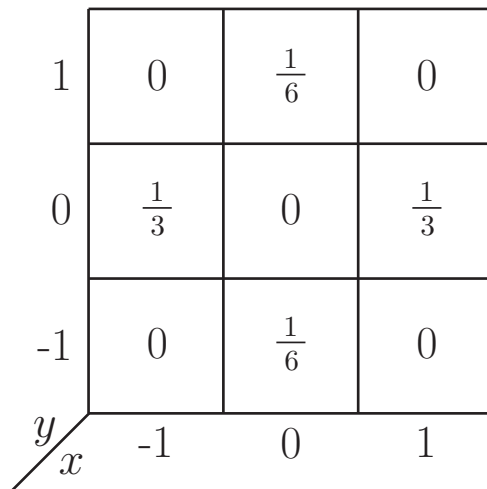
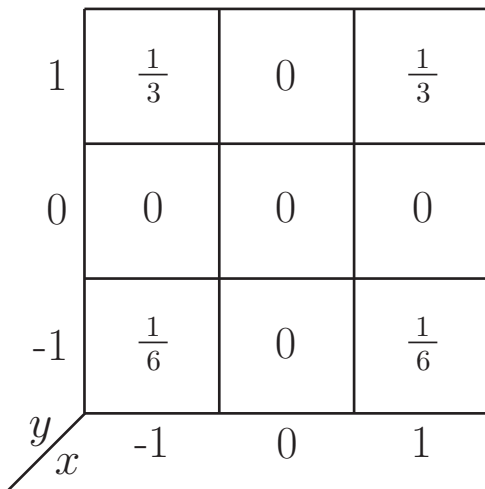
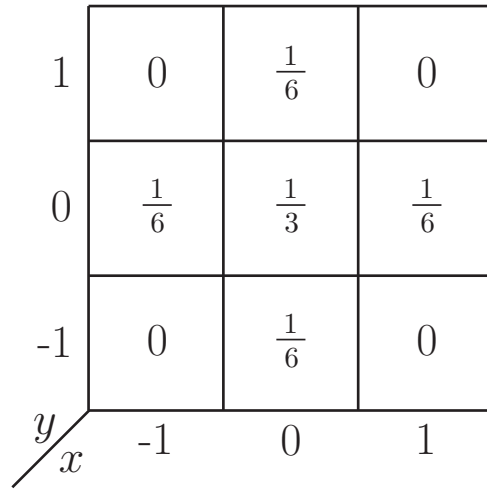
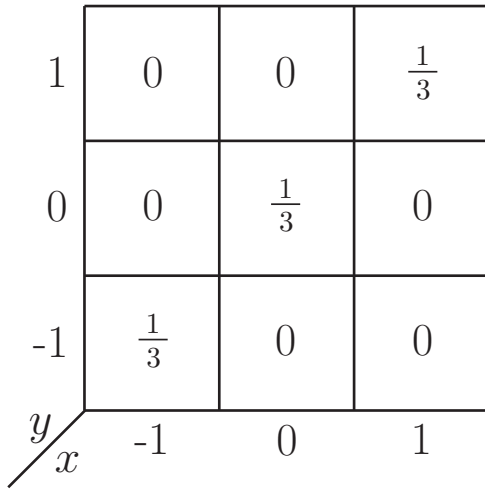
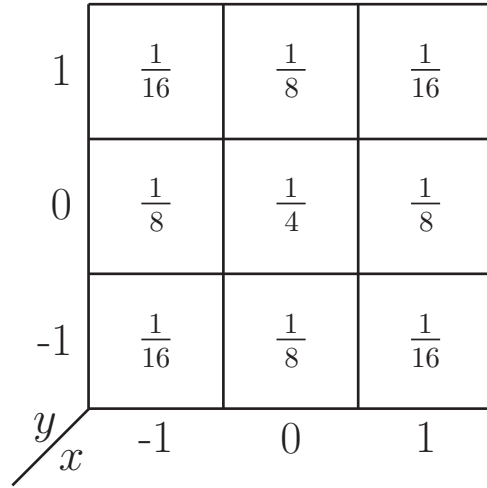
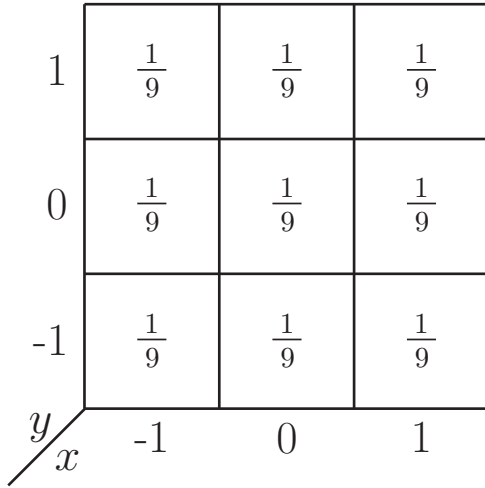
Since we can always factor  $p_{X,Y}(x, y) = p_{X|Y}(x|y)p_Y(y)$ , independence means that

$$p_{X|Y}(x|y) = p_X(x) \quad \text{for all } y \text{ with } p_Y(y) > 0.$$

That is, learning the value of  $Y$  tells us nothing about  $X$ . (More precisely, learning the value of  $Y$  does not change the pmf for  $X$ .)

### Exercise:

Suppose that the random variables  $X$  and  $Y$  can take three different values:  $-1, 0, 1$ . For each joint pmf below, decide whether  $X$  and  $Y$  are independent.



1	$\frac{1}{3}$	0	$\frac{1}{6}$	
0	0	0	0	
-1	$\frac{1}{6}$	0	$\frac{1}{3}$	
$y$	$x$	-1	0	1

1	$\frac{1}{4}$	$\frac{1}{4}$	0
0	$\frac{1}{4}$	$\frac{1}{4}$	0
-1	0	0	0
$y$ $x$	-1	0	1

1	$\frac{1}{4}$	$\frac{1}{4}$	0	
0	0	0	0	
-1	$\frac{1}{4}$	$\frac{1}{4}$	0	
$y$	$x$	-1	0	1

1	$\frac{1}{4}$	$\frac{1}{4}$	0
0	0	$\frac{1}{4}$	$\frac{1}{4}$
-1	0	0	0
$y$ $x$	-1	0	1

## Some consequences of independence

If  $X$  and  $Y$  are independent, then

1.  $E[XY] = E[X] E[Y]$

This is easy to show:

$$\begin{aligned} E[XY] &= \sum_x \sum_y xyp_{X,Y}(x, y) \\ &= \sum_x \sum_y xyp_X(x)p_Y(y) \\ &= \left( \sum_x xp_X(x) \right) \left( \sum_y yp_Y(y) \right) \\ &= E[X] E[Y] \end{aligned}$$

2. Using the same line of reasoning:

$$E[g(X)h(Y)] = E[g(X)] E[h(Y)]$$

3.  $\text{var}(X + Y) = \text{var}(X) + \text{var}(Y)$

This follows directly from 1. above:

$$\begin{aligned} \text{var}(X + Y) &= E[(X + Y)^2] - (E[X + Y])^2 \\ &= E[X^2 + 2XY + Y^2] - (E[X])^2 - 2 E[X] E[Y] - (E[Y])^2 \\ &= E[X^2] + 2 E[XY] + E[Y^2] \\ &\quad - (E[X])^2 - 2 E[X] E[Y] - (E[Y])^2 \\ &= E[X^2] + 2 E[X] E[Y] + E[Y^2] \quad (\text{by 1.}) \\ &\quad - (E[X])^2 - 2 E[X] E[Y] - (E[Y])^2 \\ &= E[X^2] - (E[X])^2 + E[Y^2] - (E[Y])^2 \\ &= \text{var}(X) + \text{var}(Y) \end{aligned}$$

## Independence of many random variables

The definition of independence extends naturally to more than two random variables.

We say  $X_1, X_2, \dots, X_n$  are independent if

$$p_{X_1, \dots, X_n}(x_1, \dots, x_n) = p_{X_1}(x_1) \cdot p_{X_2}(x_2) \cdot \dots \cdot p_{X_n}(x_n)$$

Again, a consequence of this independence is that

$$\text{var}(X_1 + X_2 + \dots + X_n) = \text{var}(X_1) + \text{var}(X_2) + \dots + \text{var}(X_n)$$

**Example. (Polling)** Suppose we are trying to determine whether Candidate R or Candidate D is going to win the state of Florida in an upcoming election. We will let  $p$  be the proportion of voters that will vote for  $R$ ; if  $p > 0.5$ , then R will win, if  $p < 0.5$ , then  $D$  will win.

We will try to determine  $p$  by randomly selecting a subset of size  $n$  of the population and polling them. A decent model for this process is treat the responses of the  $n$  people as independent Bernoulli random variables  $X_1, X_2, \dots, X_n$  with pmfs

$$p_{X_i}(k) = \begin{cases} p & k = 1 \\ 1 - p & k = 0 \end{cases},$$

so  $X_i = 1$  if you plan to vote for R, and  $X_i = 0$  if you plan to vote for D.

It is easy to see that since  $E[X_i] = p$  we have that

$$E[X_1 + X_2 + \cdots + X_n] = E[X_1] + E[X_2] + \cdots + E[X_n] = np$$

and so a reasonable way to estimate  $p$  is by calculating the **sample mean**:

$$\hat{P} = \frac{1}{n}(X_1 + X_2 + \cdots + X_n).$$

Notice that  $\hat{P}$  is itself a random variable; its expectation is  $E[\hat{P}] = p$  no matter what  $n$  is. But as we will see below, the variance of  $\hat{P}$  **decreases** as  $n$  gets bigger.

1. What is the variance of a single polling result  $X_i$ ?

$$\text{var}(X_i) =$$

2. What is the variance of  $\hat{P}$ ?

$$\text{var}(\hat{P}) =$$

3. How does  $\text{var}(\hat{P})$  behave as  $n$  gets large? What does this say (qualitatively) about how  $\hat{P}$  concentrates around its mean  $p$ ?

**Exercise:**

When I drive to work I pass through 27 traffic lights. Assume that each light is equally likely to be green or red when I arrive, independent of all others.

1. What is the pmf of the number of red lights that I hit?
2. What is the mean and variance of the number of red lights that I hit?
3. Suppose that each red light causes a 1.5 minute delay. What is the variance of my commuting time?