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.

$y \backslash x$	-1	0	1
1	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$
0	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$
-1	$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$

$y \backslash x$	-1	0	1
1	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{1}{16}$
0	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{8}$
-1	$\frac{1}{16}$	$\frac{1}{8}$	$\frac{1}{16}$

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

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

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

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

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

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

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

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

Some consequences of independence

If X and Y are independent, then

1. $\mathrm{E}[XY] = \mathrm{E}[X]\,\mathrm{E}[Y]$
   This is easy to show:
$$\begin{aligned}\mathrm{E}[XY] &= \sum_x \sum_y xy p_{X,Y}(x,y) \\ &= \sum_x \sum_y xy p_X(x) p_Y(y) \\ &= \left(\sum_x x p_X(x)\right)\left(\sum_y y p_Y(y)\right) \\ &= \mathrm{E}[X]\,\mathrm{E}[Y]\end{aligned}$$

2. Using the same line of reasoning:
$$E[g(X)h(Y)] = \mathrm{E}[g(X)]\,\mathrm{E}[h(Y)]$$

3. $\mathrm{var}(X+Y) = \mathrm{var}(X) + \mathrm{var}(Y)$
   This follows directly from 1. above:
$$\begin{aligned}\mathrm{var}(X+Y) &= \mathrm{E}[(X+Y)^2] - (\mathrm{E}[X+Y])^2 \\ &= \mathrm{E}[X^2 + 2XY + Y^2] - (\mathrm{E}[X])^2 - 2\,\mathrm{E}[X]\,\mathrm{E}[Y] - (\mathrm{E}[Y])^2 \\ &= \mathrm{E}[X^2] + 2\,\mathrm{E}[XY] + \mathrm{E}[Y^2] \\ &\qquad - (\mathrm{E}[X])^2 - 2\,\mathrm{E}[X]\,\mathrm{E}[Y] - (\mathrm{E}[Y])^2 \\ &= \mathrm{E}[X^2] + 2\,\mathrm{E}[X]\,\mathrm{E}[Y] + \mathrm{E}[Y^2] \qquad \text{(by 1.)} \\ &\qquad - (\mathrm{E}[X])^2 - 2\,\mathrm{E}[X]\,\mathrm{E}[Y] - (\mathrm{E}[Y])^2 \\ &= \mathrm{E}[X^2] - (\mathrm{E}[X])^2 + \mathrm{E}[Y^2] - (\mathrm{E}[Y])^2 \\ &= \mathrm{var}(X) + \mathrm{var}(Y)\end{aligned}$$

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

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

Again, a consequence of this independence is that

$$\mathrm{var}(X_1 + X_2 + \cdots + X_n) = \mathrm{var}(X_1) + \mathrm{var}(X_2) + \cdots + \mathrm{var}(X_n)$$

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 $\mathrm{E}[X_i] = p$ we have that

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

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

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

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

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

   $\mathrm{var}(X_i) =$

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

   $\mathrm{var}(\widehat{P}) =$

3. How does $\mathrm{var}(\widehat{P})$ behave as n gets large? What does this say (qualitatively) about how $\widehat{P}$ 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?