We then have (from the definition of independent events)

$$\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\,\mathbb{P}(Y = y)\,. \qquad \text{for all } x \text{ and } y$$

or put another way

$$\boxed{f_{X,Y}(x, y) = f_X(x)\, f_Y(y) \qquad \text{for all } x \text{ and } y.}$$

We call the *random variables* X and Y independent when this holds.

There is an important difference between independent random variables and independent events. Recall that if x and y are given (fixed) then the *events* $\{X = x\}$ and $\{Y = y\}$ being independent only means

$$\boxed{f_{X,Y}(x, y) = f_X(x)\, f_Y(y) \qquad \text{for a given pair } (x,y)}$$

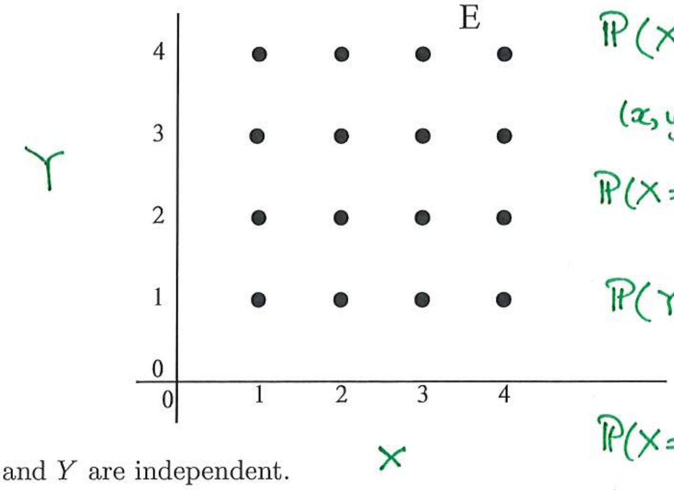
not for all x and y.

Note the similarity of these operations to those we performed earlier for events.

Example. Repeat the previous experiment with three ordinary dice. Since the events $\{X = x\}$ and $\{Y = y\}$ should be independent, each entry in the pmf table is $\frac{1}{3} \times \frac{1}{6}$.

Clearly in the first experiment not *all* events $\{X = x\}$ and $\{Y = y\}$ are independent (which are not?). Hence the random variables X and Y are not considered to be independent.

Example. We draw at random a point (X, Y) from the 16 points on the square E below.



$\mathbb{P}(X = x, Y = y) = \frac{1}{16}$

$(x, y) \in \{1, .. 4\} \times \{1, .., 4\}$

$\mathbb{P}(X = x) = \frac{4}{16} = \frac{1}{4}\,; \quad x = 1, .., 4$

$\mathbb{P}(Y = y) = \frac{4}{16} = \frac{1}{4}\,; \quad y = 1, .., 4$

Clearly X and Y are independent.

$\mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x)\,\mathbb{P}(Y = y)$ for all x, y so the random variables X and Y are independent.

Expectation Revisited

Similar to the one-dimensional case, the expected value of $Z = g(X, Y)$ can be evaluated as

$$\mathbb{E}[g(X, Y)] \quad \cancel{\mathbb{E}\,g(X, Y)} = \sum_x \sum_y g(x, y)\mathbb{P}(X = x, Y = y),$$

find pmf of Z and apply definition from pg 46.

and in general, the expected value of $Z = g(X_1, \dots, X_n)$ can be evaluated as

$$\mathbb{E}[g(X_1, \dots, X_n)] = \sum_{x_1} \cdots \sum_{x_n} g(x_1, \dots, x_n) \mathbb{P}(X_1 = x_1, \dots, X_n = x_n).$$

Example. We will often be interested in sums of random variables. Find the expected value of $X + Y$.

$$\begin{aligned}\mathbb{E}[X+Y] &= \sum_x \sum_y (x+y) \cdot \mathbb{P}(X=x, Y=y) \\ &= \sum_x \sum_y x\, \mathbb{P}(X=x, Y=y) + \sum_x \sum_y y\, \mathbb{P}(X=x, Y=y) \\ &= \sum_x x \underbrace{\boxed{\sum_y \mathbb{P}(X=x, Y=y)}}_{\text{marginal pmf of } X} + \sum_y y \underbrace{\boxed{\sum_x \mathbb{P}(X=x, Y=y)}}_{\text{marginal pmf } Y} \\ &= \sum_x x\, \mathbb{P}(X=x) + \sum_y y\, \mathbb{P}(Y=y) = \mathbb{E}[X] + \mathbb{E}[Y]\end{aligned}$$

In general, suppose that $X_1, \dots, X_n$ are random variables measured on the same random experiment. For arbitrary constants $b_0, b_1, \dots, b_n$, we have

$$\mathbb{E}[b_0 + b_1 X_1 + \cdots b_n X_n] = b_0 + b_1 \mathbb{E}[X_1] + \cdots + b_n \mathbb{E}[X_n]$$

Note: This (linearity of expectations) holds for any collection of random variables [measured on the same random experiment].

Example. Suppose that $X_1, \dots, X_n$ are independent Bernoulli(p) random variables. Compute $\mathbb{E}[X_1 + \cdots + X_n]$.

Recall that if $X_i \sim$ Bernoulli(p), then $\mathbb{P}(X_i = x) = p^x(1-p)^{1-x}$ for $x = 0, 1$.

$$\begin{aligned}\mathbb{E} X_i &= \sum_{x=0}^{1} x \cdot \mathbb{P}(X=x) \\ &= 0 \times (1-p) + 1 \times p = p\end{aligned}$$

Therefore,

$$\begin{aligned}\mathbb{E}[X_1 + \cdots + X_n] &= \mathbb{E}[X_1] + \cdots + \mathbb{E}[X_n] \\ &= p + p + \cdots + p = np\end{aligned}$$

We previously computed the expectation of a random variable having a Binomial(n, p) distribution. This required considerable effort. We know that a Binomial(n, p) random variable describes the total number of sucesses in a sequence of n Bernoulli trials with success probability p. In other words, if $X_1, \ldots, X_n$ are independent Bernoulli(p) random variables, then $\sum_{i=1}^{n} X_i \sim \text{Binomial}(n, p)$. Using linearity of expectations has greatly simplified this calculation.

Example. Suppose that X and Y are two *independent* random variables measured on the same random experiment. Find the expected value of XY.

$$\begin{aligned}
\mathbb{E}[XY] &= \sum_x \sum_y xy\, \mathbb{P}(X=x, Y=y) \\
&= \sum_x \sum_y xy\, \mathbb{P}(X=x)\, \mathbb{P}(Y=y) \quad \text{(by independence)} \\
&= \left\{\sum_x x\, \mathbb{P}(X=x)\right\}\left\{\sum_y y\, \mathbb{P}(Y=y)\right\} \\
&= \mathbb{E}[X]\, \mathbb{E}[Y]
\end{aligned}$$

Important: Suppose that $X_1, \ldots, X_n$ are *independent* random variables measured on the same random experiment. We have

$$\mathbb{E}[X_1 X_2 \cdots X_n] = \mathbb{E}[X_1]\, \mathbb{E}[X_2] \cdots \mathbb{E}[X_n]$$

Variance and Covariance

In the previous subsection, we saw that the expected value of the sum of random variables is equal to the sum of the expectations of the individual random variables. This gives a measure of location for the distribution of a sum of random variables. We have also seen that the spread of a distribution can be described by the variance. It is natural, therefore, to consider the variance of a sum of random variables. In general, the variance of a sum of random variables does not equal the sum of the variances, but includes an extra term called the covariance.

Definition. The covariance of a two random variables X and Y, denoted by $\text{Cov}(X, Y)$, is defined by

$\text{Cov}(X, X) = \text{Var}(X)$

$$\text{Cov}(X, Y) = \mathbb{E}\left[(X - \mathbb{E}X)(Y - \mathbb{E}Y)\right].$$

It is a measure of the amount of linear dependence between the two random variables.

Example. Suppose that X and Y are two random variables measured on the same random experiment. Find the variance of $X + Y$.

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

$$\begin{aligned}
\mathrm{Var}[X+Y] &= \sum_x \sum_y (x + y - \mathbb{E}[X+Y])^2 \, \mathbb{P}(X=x, Y=y) \\
&= \sum_x \sum_y (\boxed{x} + \boxed{y} - \boxed{\mathbb{E}[X]} - \mathbb{E}[Y])^2 \, \mathbb{P}(X=x, Y=y) \\
&= \\
&= \sum_x \sum_y \{(x - \mathbb{E}X)^2 + 2(x-\mathbb{E}X)(y - \mathbb{E}Y) + (y - \mathbb{E}Y)^2\} \, \mathbb{P}(X=x, Y=y) \\
&= \sum_x \sum_y (x - \mathbb{E}X)^2 \, \mathbb{P}(X=x, Y=y) \quad \swarrow \mathrm{Var}(X) \\
&\quad + 2 \sum_x \sum_y (x - \mathbb{E}X)(y - \mathbb{E}Y) \, \mathbb{P}(X=x, Y=y) \quad \swarrow \mathrm{cov}(X,Y) \\
&\quad + \sum_x \sum_y (y - \mathbb{E}Y)^2 \, \mathbb{P}(X=x, Y=y) \quad \swarrow \mathrm{Var}(Y) \\
&= \mathrm{Var}(X) + 2\,\mathrm{cov}(X,Y) + \mathrm{Var}(Y)
\end{aligned}$$

Recall that if X and Y are two independent random variables, then $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$. It follows that if X and Y are two independent random variables, then

$$\mathrm{Cov}(X,Y) = \mathbb{E}[(X - \mathbb{E}X)(Y - \mathbb{E}Y)] = \mathbb{E}[X - \mathbb{E}X]\,\mathbb{E}[Y - \mathbb{E}Y] = (\mathbb{E}X - \mathbb{E}X)\cdot(\mathbb{E}Y - \mathbb{E}Y) = 0$$

and

$$\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$$

Important: Suppose that $X_1, \ldots, X_n$ are *independent* random variables measured on the same random experiment. We have

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

Example. Find the variance of $X \sim \mathsf{Binomial}(n,p)$.

Let $X_1, \ldots, X_n$ be independent $\mathsf{Bernoulli}(p)$ random variables. Then $\sum_{i=1}^n X_i$ has a $\mathsf{Binomial}(n,p)$ distribution. We can therefore compute the variance of X by computing the variance of $\sum_{i=1}^n X_i$.

Recall that if $X_i \sim \mathsf{Bernoulli}(p)$, then $\mathbb{P}(X_i = x) = p^x(1-p)^{1-x}$ for $x = 0, 1$.

$$\mathbb{E}X_i = p$$

$$\begin{aligned}
\mathrm{Var}(X_i) &= \sum_{x=0}^{1} (x - \mathbb{E}X_i)^2 \, \mathbb{P}(X_i = x) \\
&= p^2 \times (1-p) + (1-p)^2 p = p(1-p)
\end{aligned}$$

Therefore,

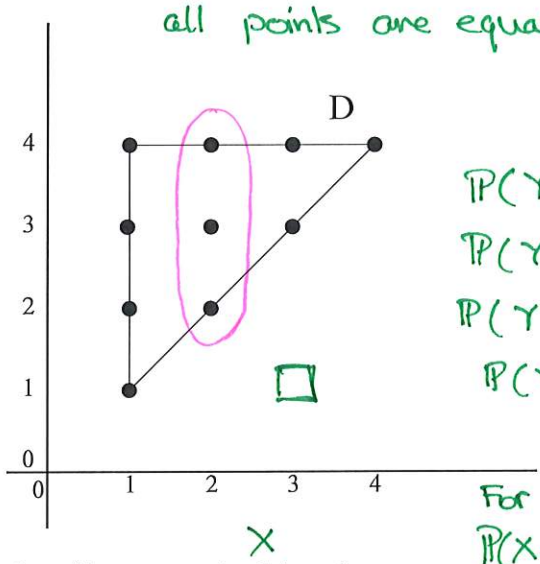
$$\mathrm{Var}(X_1 + X_2 + \cdots + X_n) = \cancel{p(1-p)}\ Var(X_1) + \cdots + Var(X_n)$$
$$= p(1-p) + \cdots + p(1-p) = np(1-p)$$

The Variance of a Binomial(n,p) r.v. is $np(1-p)$.

Conditional probability mass function

Example. We draw at random a point (X, Y) from the 10 points on the triangle D below.

all points are equally likely to be selected.



$\mathbb{P}(X=3, Y=1) = 0$

$\neq \mathbb{P}(X=3) \times \mathbb{P}(Y=1)$

$\mathbb{P}(Y=4) = 4/10$

$\mathbb{P}(Y=3) = 3/10$

$\mathbb{P}(Y=2) = 2/10$

$\mathbb{P}(Y=1) = 1/10$

For independence we need $\mathbb{P}(X=i, X=j) = \mathbb{P}(X=i)\,\mathbb{P}(Y=j)$ $\forall i,j$

The joint and marginal pmf's are easy to determine:

$$\mathbb{P}(X = i, Y = j) = 1/10 \qquad (i, j) \in D,$$
$$\mathbb{P}(X = i) = \frac{5-i}{10}, \qquad i \in \{1, 2, 3, 4\},$$
$$\mathbb{P}(Y = j) = j/10 \qquad j \in \{1, 2, 3, 4\}$$

Clearly X and Y are *not independent.* In fact, if we know that $X = 2$, then Y can only take the values $j = 2, 3$ or 4.

The corresponding probabilities are

$$f_{Y\,|\,X}(j, 2) = \begin{cases} \mathbb{P}(Y = j \,|\, X = 2) = \dfrac{\mathbb{P}(Y = j, X = 2)}{\mathbb{P}(X = 2)} = \dfrac{1/10}{3/10} = \dfrac{1}{3} & \text{if } j \in \{2, 3, 4\}, \\ 0 & \text{otherwise.} \end{cases}$$

We thus have determined the **conditional pmf** of Y given $X = 2$.

Definition. If X and Y are *discrete* and $\mathbb{P}(X = x) > 0$, then the probabilities

$$f_{Y\,|\,X}(y \,|\, x) = \mathbb{P}(Y = y \,|\, X = x) = \frac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(X = x)} = \frac{f_{X,Y}(x, y)}{f_X(x)},$$