$Y = g(X)$

The expected value of Y is then

$$\mathbb{E}Y = \sum_y y\mathbb{P}(Y=y) = \sum_y y\left[\sum_{x:g(x)=y} \mathbb{P}(X=x)\right] = \sum_x g(x)\mathbb{P}(X=x).$$

This leads to the following natural defintion for the expectation of $g(X)$ for any real valued function g.

Definition. If X is a *discrete* random variable, then for any real-valued function g

$$\mathbb{E}\left[g(X)\right] = \sum_x g(x)\,\mathbb{P}(X=x)\,.$$

LOTUS: Law of the unconscius statistician.

Example. Find $\mathbb{E}[X/n]$ where $X \sim \mathsf{Binomial}(n,p)$. We have

$$\boxed{\mathbb{E}[X/n] = \sum_{x=0}^{n} \frac{x}{n}\cdot\binom{n}{x}p^x(1-p)^{n-x} = \frac{1}{n}\sum_{x=0}^{n} x\binom{n}{x}p^x(1-p)^{n-x}} = \frac{1}{n}\times np = p$$

In general, for any random variable X and real constants a and b

$$\boxed{\mathbb{E}[aX+b] = a\times\mathbb{E}(X) + b}$$

Example. Find $\mathbb{E}[X]$ for $X \sim \mathsf{Poisson}(\lambda)$. $\mathbb{E}(X) = \lambda$

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{x=0}^{\infty} x\cdot\mathbb{P}(X=x) \\
&= \sum_{x=0}^{\infty} x\cdot\frac{e^{-\lambda}\lambda^x}{x!} \\
&= \sum_{x=1}^{\infty} x\cdot\frac{e^{-\lambda}\lambda^x}{x\times(x-1)!} \\
&= \sum_{x=1}^{\infty} \frac{e^{-\lambda}\lambda^{x-1}\lambda}{(x-1)!} \\
&= \lambda\boxed{\sum_{k=0}^{\infty}\frac{e^{-\lambda}\lambda^k}{k!}} \\
&= \lambda\,.
\end{aligned}$$

$k = x-1$

pmf of poisson

$$\begin{aligned}
&\mathbb{E}[X(X-1)] \\
&= \sum_{x=0}^{\infty} x(x-1)\frac{e^{-\lambda}\lambda^x}{x!} \\
&= \sum_{x=2}^{\infty} \frac{\cancel{x}\cancel{(x-1)}e^{-\lambda}\lambda^x}{\cancel{x}\cdot\cancel{(x-1)}\cdot(x-2)!} \\
&= \sum_{x=2}^{\infty}\frac{e^{-\lambda}\lambda^{x-2}\times\lambda^2}{(x-2)!} \\
&= \lambda^2\sum_{k=0}^{\infty}\frac{e^{-\lambda}\lambda^k}{k!} \\
&= \lambda^2
\end{aligned}$$

$$Var(X) = \mathbb{E}(X-\mathbb{E}X)^2 = \sum_x (x-\mathbb{E}(X))^2\,\mathbb{P}(X=x) = \sum_x \left(x^2 - 2x\,\mathbb{E}X + (\mathbb{E}X)^2\right)\mathbb{P}(X=x)$$

$$= \sum_x x^2\,\mathbb{P}(X=x) - 2(\mathbb{E}X)\underbrace{\sum_x x\,\mathbb{P}(X=x)} + (\mathbb{E}X)^2\underbrace{\sum_x \mathbb{P}(X=x)}$$

$$= \mathbb{E}(X^2) - 2(\mathbb{E}X)(\mathbb{E}X) + (\mathbb{E}X)^2$$

Variance

Definition. The variance of a random variable X, denoted by $\mathrm{Var}(X)$ is defined by

$$\mathrm{Var}(X) = \mathbb{E}\,(X - \mathbb{E}\,X)^2\,.$$

sigma just a number

This *number*, sometimes written as σ_X^2, measures the *spread* or dispersion of the distribution of X. The square root of the variance is called the **standard deviation** and is denoted by σ_X.

It may be regarded as a measure of the *consistency* of outcome, as a smaller value of $\mathrm{Var}(X)$ implies that X is more often near $\mathbb{E}\,X$ than for a larger value of $\mathrm{Var}(X)$.

We finally list two properties of the variance:

- $\mathrm{Var}(X) = \mathbb{E}\,[X^2] - (\mathbb{E}\,X)^2$; and
- $\mathrm{Var}(a\,X + b) = a^2\,\mathrm{Var}(X)$.

$$Var(aX+b) = \sum_x \left(ax+\cancel{b} - (a\mathbb{E}X + \cancel{b})\right)^2\mathbb{P}(X=x)$$

$$= \sum_x a^2(x-\mathbb{E}X)^2\,\mathbb{P}(X=x)$$

Example. Find $\mathrm{Var}(X)$ where X is the outcome of a roll of a fair die.

$$\mathbb{E}\,[X^2] = \sum_{x=1}^{6} x^2\,\mathbb{P}(X=x) = 1^2\times\frac{1}{6} + 2^2\times\frac{1}{6} + \cdots + 6^2\times\frac{1}{6}$$

$$= 91/6$$

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}X)^2 = \frac{91}{6} - (3.5)^2 \simeq 2.92$$

The expectation and variance of the common discrete distributions can be found in the formula sheet. Familiarise yourself with these expressions. Are you able to derive them?

Example. From the formula sheet, if $X \sim \mathsf{Binomial}(n,p)$, then $\mathrm{Var}(X) = np(1-p)$. Find $\mathrm{Var}(X/n)$.

$$\mathrm{Var}(X/n) = \frac{1}{n^2}\,Var(X) = \frac{np(1-p)}{n^2} = \frac{p(1-p)}{n}$$

Question: How can we interpret this result?

Multiple Random Variables

When looking into detailed examples, we came across situations involving **multiple random variables** where *dependence* is an inherent model characteristic.

Examples.

1. We randomly select a person from a large population of twitter users and record the number of people they follow X and the number of people who follow them Y.

2. The number of swaps X and number of comparisons Y performed by a sorting algorithm such as quicksort.

3. We randomly select 20 people from a large population and ask their age. Number the people from 1 to 20, and let $X_1, \ldots, X_{20}$ be the measurements.

How can we specify a model these experiments?

We cannot just specify the pmf of the individual random variables. We also need to specify the "*interaction*" between the random variables. E.g., in Example 2, if the sorting algorithm needs to perform a large number of swaps X, then it also probably needs to do a large number of comparisons Y.

We need to specify the **joint distribution** of all the random variables $X_1, \ldots, X_n$ involved in the experiment. In other words, we need to specify the distribution of the random **vector** $\boldsymbol{X} := (X_1, \ldots, X_n)$. The **joint cumulative distribution function** F is defined by

$$F(x_1, \ldots, x_n) = \mathbb{P}(X_1 \leqslant x_1, \ldots, X_n \leqslant x_n) .$$

This completely specifies the probability distribution of the vector $\boldsymbol{X}$. However, if the X_i's are discrete, it suffices to only know the **joint probability mass function**.

Definition. Let $\boldsymbol{X} = (X_1, \ldots, X_n)$ be a *discrete* random vector. The function

$$(x_1, \ldots, x_n) \mapsto \mathbb{P}(X_1 = x_1, \ldots, X_n = x_n)$$

is called the **joint probability mass function** of $\boldsymbol{X}$.

We will often just work with pairs of random variables (X, Y) having a joint pmf $f_{X,Y}$. The generalisation to multiple random variables is usually straightforward.

Example: Suppose that a fair, six-sided die is rolled with an independently-tossed fair coin. Let X be the face value of the die, in $\{1, 2, 3, 4, 5, 6\}$, and let Y be the outcome of the coin, in $\{0, 1\}$.

0 - Heads

1 - Tails

y \ x	1	2	3	4	5	6	$\sum$
0	$1/12$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$1/2$
1	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$1/12$	$\frac{1}{2}$
$\sum$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	1

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

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

(by independence)

$= \frac{1}{6} \times \frac{1}{2} = \frac{1}{12}$

Example. In a box are three dice. Die 1 is a normal die; die 2 has no 6 face, but instead two 5 faces; die 3 has no 5 face, but instead two 6 faces.

The experiment consists of selecting a die at random, followed by a roll of that die. Let X be the die number that is selected, and let Y be the face value of that die. The joint pmf of X and Y is specified below.

x \ y	1	2	3	4	5	6	$\sum$
1	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{3}$
2	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{9}$	0	$\frac{1}{3}$
3	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	0	$\frac{1}{9}$	$\frac{1}{3}$
$\sum$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	1

$f_X(x)$

$f_Y(y)$

$$\begin{aligned}\mathbb{P}(X=3, Y=6) &= \mathbb{P}(\text{pick die 3 and roll 6}) \\ &= \mathbb{P}(\text{roll 6} \mid \text{pick die 3}) \times \mathbb{P}(\text{pick die 3}) \\ &= \frac{2}{6} \times \frac{1}{3} = \frac{1}{9}\end{aligned}$$

Marginal Distributions

Consider a set B that consists of pairs of points, where each pair is of the type (x, y), for real numbers x and y. For example,

> $B = \{(1,4), (2,4), (3,4)\}$ set of outcomes where we rolled 4
>
> $B = \{(x,y) : x \in \{1,2,3\}, y \in \{1,2,\ldots,6\}\}$
>
> This latter set describes the possible (x, y) values from the previous example.

We have for a set B containing elements of the type (x, y),

$$f_{X,Y}(B) = \mathbb{P}((X,Y) \in B) = \sum_{(x,y)\in B} \mathbb{P}(X = x, Y = y) .$$

The pmf's f_X of X, the so-called **marginal** pmf, can be found by *summing up* over respectively the y's:

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

Similarly, the marginal pmf of Y can be found by summing up over the x's.

Question: Is this another version of the Law of Total Probability?

> The Law of Total Probability gives
>
> $$f_X(x) = \mathbb{P}(X = x) = \sum_k \mathbb{P}(\{X=x\} \cap B_k) = \sum_k \mathbb{P}(X=x, Y=k)$$
>
> collection of sets $B_k = \{(x,y) : y = k\}$ $k = 1,2,\ldots,6$ forms a partion of Ω.

Independence of Discrete Random Variables

An important way of **creating** joint pmf's is by starting with the marginal pmf's of X and Y and then to define the events $\{X = x\}$ and $\{Y = y\}$ to be *independent*, for all x and y.

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

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

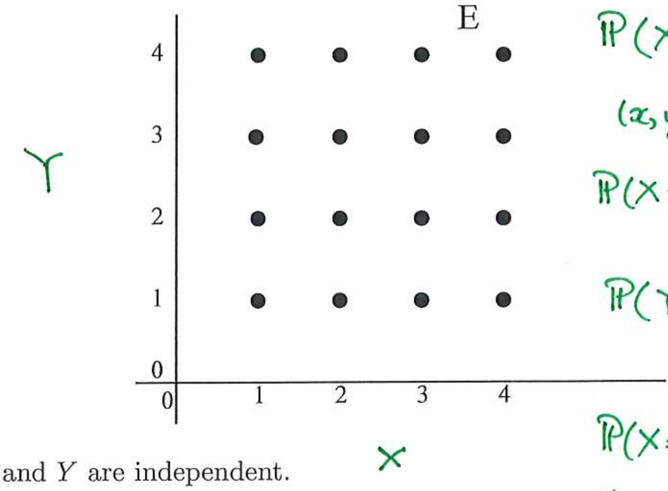
not for all x and y.

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

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