Two events are said to be independent if

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

which implies that

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

If events are not independent, then they are said to be dependent.

So we can tell if two events are independent or dependent by looking at either their joint probability or their conditional probability.

Mutually exclusive events A and B such that $P(A) \neq 0$ and $P(B) \neq 0$, are dependent because

$$P(A \cap B) = 0 \neq P(A) \times P(B)$$

An exception is when if either $P(A) = 0$ or $P(B) = 0$ then

$$P(A \cap B) = P(\emptyset) = 0 = P(A) \times P(B)$$

This is the only case where on set of events can be both independent and mutually exclusive

Random Variables

We can define a function or rule that assigns a numerical value to each outcome of a categorical experiment. Such a rule is known as a random variable. For example, in the case of a coin we could define a random variable:

$$X = \textit{the total number of heads observed}$$

Discrete random variable can take countable number of distinct values, meaning that the set of values are a subset of the natural numbers $\{1,2,3,4,\ldots\}$ so random variables can also take an infinite number of distinct values. Example: number of applicants to a university.

A random variable is described entirely by its **probability distribution.** For simple DRV's this is just a list or table of values that the RV can possibly take and their associated probabilities. This is also known as a probability mass function.

Example: Let X be the # of heads observed after 3 coin flips. The PMF is given by:

X	$P(X = x_i)$
0	$1/8$
1	$3/8$
2	$3/8$
3	$1/8$

Another way we can describe a random variable is via its cumulative distribution function which gives the probability being **less than or equal to** some value c.

$$Fx(c) = P(X \leq c), c \in \mathbb{R}$$

CDF's satisfy the following properties:

1. $Fx(-\infty) = 0$
2. $Fx(\infty) = 1$ and
3. $0 \leq Fx(c) \leq 1 \; for \; all - \infty < c < \infty$

The expected value (mean) of a random variable is the weighted sum of all the possible outcomes in which the weights are of associated probabilities.

Given a DRV X with possible values $x_1, x_2, \dots, x_k$ that occur with probabilities $P(X = x_i)$, for $i = 1, \dots, k$, the expected value of X is

$$\mu = E(X) = \sum_{i=1}^{k} x_i P(X = x_i)$$

To determine the how spread out or dispersed our observations would be if we were to observe many realisations of a random variable we would compute the variance.

$$\sigma^2 = V(X) = E[(X - \mu)^2] = \sum_{i=1}^{k} (x_i - \mu)^2 P(X = x_i)$$

another way to calculate variance is

$$\sigma^2 = E(X^2) - \mu^2$$

Discrete Bivariate Distributions

Consider an experiment where the outcomes can be described in terms of two random variables (X, Y), with X $\in$ {x_1, x_2} and Y $\in$ {y_1, y_2, y_3}. We can represent all possible outcomes in a table:

$X \backslash Y$	y_1	y_2	y_3
x_1	(x_1, y_1)	(x_1, y_2)	(x_1, y_3)
x_2	(x_2, y_1)	(x_2, y_2)	(x_2, y_3)

Observe that each outcome is a **joint event** of the form

$$\{X = x \cap Y = y\}.$$

If we add row and column sums to the previous table, then we obtain a marginal probability distribution

$X \backslash Y$	y_1	y_2	y_3	$P(X)$
x_1	$P(x_1, y_1)$	$P(x_1, y_2)$	$P(x_1, y_3)$	$P(x_1)$
x_2	$P(x_2, y_1)$	$P(x_2, y_2)$	$P(x_2, y_3)$	$P(x_2)$
$P(Y)$	$P(y_1)$	$P(y_2)$	$P(y_3)$	1

If we condition on the event X=x_1 then we see that the various probabilities of Y the we know that the various probabilities of Y occurring are P(x1, y1), P(x1,y2), and P(x1,y3). These values are not a valid probability distribution because the do not add up to 1. Instead:

$$P(x_1, y_1) + P(x_1, y_2) + P(x_1, y_3) = P(x_1)$$

To obtain a valid probability distribution we simply scale it by its sum. Thus

$$P(Y = y_1 | X = x_1) = \frac{P(x_1, y_1)}{P(x_1)}, \quad P(Y = y_2 | X = x_1) = \frac{P(x_1, y_2)}{P(x_1)},$$
$$P(Y = y_3 | X = x_1) = \frac{P(x_1, y_3)}{P(x_1)}$$

Here P(x_1) plays the role of a normalizing constant, which scales probabilities so that they sum to unity. To scale for independence one needs to check that for every cell in the table,

$$P(x_i, y_j) = P(x_i) \times P(y_j)$$

- If this relationship fails to hold for at least one cell, then the RV's are dependent.

If the joint probabilities are specified as a function, then that function must factor according to

$$f_{X,Y}(x, y) = f_X(x) \times f_Y(y).$$

If $P(x_i, y_j) = 0$ for any pair (i, j) then the events X=x_i and Y=y_i are **mutually exclusive.** The only way the RV's X and Y can be mutually exclusive for all of their possible values is if one or both have zero probability of ever occurring.

We can obtain a measure of the association between two random variables by computing their covariance,

$$\sigma_{XY} = COV[X, Y] = \sum_{i=1}^{k} \sum_{j=1}^{l} (x_i - \mu_x)(y_j - \mu_y) P(X = x_i \cap Y = y_j)$$

As is the case with the sample covariance, we can rescale this quantity to obtain the coefficient of correlation,

$$\rho = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

If 2 RV's X and Y are independent, then their correlation is zero. However, the reverse is not true.