This is not finished, but may be useful for definitions of expectation, variance, covariance, density function etc..

1 Basic definitions

In the following, we imagine an experiment whose outcome cannot be predicted with certainty. Although we do not know the outcome of the experiment, we will suppose we know the different possible outcomes.

Definition: The set of all possible outcomes of an experiment is called the *sample space* and will be denoted Ω.

> *Example:* We flip two coins and denote the results HH, HT etc. For this experiment, $\Omega = \{HH, HT, TH, TT\}$

Definition: Any subset A included in Ω ($A \subset \Omega$) is called an event.

> *Example:* If we call A the event 'to obtain at least one head' in the above experiment, then $A = \{HH, HT, TH\}$.

Definition: A probability is a function $\mathrm{P} : \Omega \rightarrow R$ which has the following properties:

- $\forall E \subset \Omega, 0 \leq \mathrm{P}(E) \leq 1$
- $\mathrm{P}(\Omega) = 1$
- for all sequence of events $E_1, E_2, \ldots$ verifying

$$\forall (i,j) \in I^2, i \neq j \Longrightarrow E_i \cap E_j = 0$$

(such events are said to be mutually exclusive), we have:

$$\mathrm{P}(\bigcup_i^{\infty} E_i) = \sum_i^{\infty} \mathrm{P}(E_i) \tag{1}$$

Note: there is a great deal which is skipped here, and the definition of a probability measure and its scope of application involve some mathematical subtlety that we will spare ourselves. The interested reader could refer to:

- *A First Course in Probability Theory*, S. M. Ross, Prentice Hall, 6th Edition
- *Probability and Random Processes*, G. Grimmett and Stirzaker D., 3rd Edition, Oxford University Press

When dealing with probabilities, there are a number of 'useful equalities and inequalities', such as

$$\mathrm{P}(A \cap B) = \mathrm{P}(A) + \mathrm{P}(B) - \mathrm{P}(A \cup B)$$

2 Conditional probabilities

We might be interested in the probability of A occurring given that B has occurred. This is called the conditional probability of A given B and is denoted $\mathrm{P}(A|B)$. Clearly, for A to occur given that B occurred, the event which caused B to happen must also belong to A. It follows that we are concerned with the probability of $A \cap B$ in the reduced sample space B:

$$P(A|B) = \frac{\mathrm{P}(A \cap B)}{\mathrm{P}(B)} \quad \text{if} \quad \mathrm{P}(B) > 0$$

Example: consider again $\Omega = \{HH, HT, TH, TT\}$ and denote B 'the first coin results in H', and A 'the second coin results in T'. Under the condition that A is realised the sample space is reduced to $B = \{HH, HT\}$. Since $A = \{HT, TT\}$, it is clear that $A \cap B = \{HT\}$ and its probability is $\frac{1}{2}$. We should have expected this result: the result of the second coin is not dependent on that of the first.

2.1 Independent events

As is the case in the example above, the probability of an event A may well not depend on that of B, in this case:

$$\mathrm{P}(A|B) = \mathrm{P}(A)$$

which is to say:

$$\mathrm{P}(A \cap B) = \mathrm{P}(A)\mathrm{P}(B)$$

3 Random variables

We are most often concerned with quantities associated with events rather than events themselves. If for example, you are playing a game where your score depends on the outcome, you will be more interested in the score than the events themselves.

Definition: A random variable X is a function $X : \Omega \rightarrow \Re$ which associates a value $X(A)$ to an event A.

> *Example:* if we flip a coin three times, we have: $\Omega = \{HHH, HHT, HTH, THH, TTH, THT, HTT, TTT\}$ we can define the random variable X as being the number of heads in the outcome of an experiment.

Obviously, our attention will now shift from the probability of an event to the probability of the random variable taking a given value.

Definition: The (probability) mass function of a random variable X is the function:

$$\begin{aligned} p_X : \Re &\rightarrow [0,1] \\ a &\rightarrow p(a) = \mathrm{P}(X = a) \end{aligned}$$

Note that the notation $P(X = a)$ is not rigorous, we really should write: $\mathrm{P}(\{\omega \in \Omega, X(\omega) = a\})$.

> *Example:* what is the mass function of X as defined above ?
>
> $$\begin{aligned} \mathrm{P}(X=0) &= \mathrm{P}(\{TTT\}) = 1/8 \\ \mathrm{P}(X=1) &= 3/8 \\ \mathrm{P}(X=2) &= 3/8 \\ \mathrm{P}(X=3) &= 1/8 \end{aligned}$$

3.1 Expectation

Imagine that you are taking part in a game where a coin is tossed three times, you earn X =3$ if the result is three heads, 1$ for all combination of heads and tails, but lose 9$ for three tails. At this point, you have more interested in the value you expect X to take on average, its *expectation* or *expected value*, than the probability of a single event.

Definition: The expectation of a random variable X, denoted $E(X)$, is given by:

$$E(X) = \sum x p_X(x)$$

where $p_X(x)$ is the mass function of X and the sum is over all the possible values of X.

> *Example:* What is the expected value of your gain/loss if you play the game described above ?
> The mass function of X is:
>
> $$\begin{aligned} \mathrm{P}(X=3) &= \mathrm{P}(\{HHH\}) = 1/8 \\ \mathrm{P}(X=1) &= 6/8 \\ \mathrm{P}(X=-9) &= 1/8 \end{aligned}$$
>
> so that:
>
> $$E(X) = 3 * \frac{1}{8} + 1 * \frac{6}{8} - 9 * \frac{1}{8} = 0$$

Property of $E(X)$ It is easy to show, using the definition, that:

$$\forall (a,b) \in \Re^2,\ E(aX+b) = aE(X) + b$$

3.2 Variance

The variance quantifies the average 'distance' between X and its expectation $E(X)$.

Definition: The variance of a random variable X, denoted $\mathrm{var}(X)$, is given by:

$$\mathrm{var}(X) = E\left((X - E(X))^2\right) = E(X^2) - [E(X)]^2$$

Property of the variance:

$$\forall (a,b) \in \Re^2,\ \mathrm{var}(aX+b) = a^2 \mathrm{var}(X)$$

4 Continuous random variables

Definition: A continuous random variable is a random variable that takes any value in $\Re$, rather than a set of discrete values.

> *Example:* The life of a light bulb can be measured in hours. The result t is a continuous random variable that takes value in $\Re^+$ (real positive).

The probability for the random variable X to take a value between x and $x + \mathrm{d}x$ is given by:

$$P(x \leq X \leq x + \mathrm{d}x) = f_X(x)\mathrm{d}x$$

f_x is called the density function of X. We now define the expectation and variance:

$$E(X) = \int_{\Re} x f_X(x)\mathrm{d}x \quad (2)$$

$$\mathrm{var}(X) = E[(x - E(X))^2] = \int_{\Re} (x - E(X))^2 f_X(x)\mathrm{d}x \quad (3)$$

5 Joint distributions

In many cases, we will be interested in the distribution of more than one random variable. If, for example, we measure the height and weight of a sample of population, we might be interested in the probability to have the height in an interval and, at the same time, the weight in another interval.

In this case, the joint distribution is:

$$f_{XY}(x, y)\mathrm{d}x\mathrm{d}y = P(x \leq X \leq x + \mathrm{d}x, y \leq Y \leq y + \mathrm{d}y) \quad (4)$$

In addition to expectation and variance, another important characteristic of joint distribution is the *covariance* which measure how strongly the value of one variable influences the other. The covariance is defined by:

$$\mathrm{cov}(X, Y) = E(XY) - E(X)E(Y)$$

random variables verifying $\mathrm{cov}(X, Y) = 0$ are said to be uncorrelated.