Lecture 3 — Basic Probability and Statistics

The aim of this lecture is to provide an extremely speedy introduction to the probability and statistics which will be needed for the rest of this lecture course. The majority of mathematics students should already be familiar with this material.

Definition 1. A *sample space* is the set of all possible outcomes from an experiment. For example, if we consider tossing two coins, the possible outcomes are HH, HT, TH and TT. The sample space may be discrete (as in the previous example) or continuous (for example a measurement of a person's height in metres). Formally, a *discrete sample space* is one with a finite or countably infinite number of possible values. A *continuous sample space* is one which takes values in one or more intervals.

Example One: The number of bytes counted past a point in a network in a second is a disrete sample space with the possible outcomes between 0 and the bandwidth of the link in bytes per second.

Example Two: The number of bytes counted at the queues of n nodes in a network is a discrete sample space with the possible outcomes in $\mathbb{Z}_+^n$.

Definition 2. An *event* is a subset of a sample space amd a *simple event* is one member of the sample space. Often an event is described in words rather than by explicit enumeration of the subspace. For example, if the event is "getting exactly one head in two coin tosses" then it would be the subset HT and TT. An example of an event on a continuous sample space is measuring a height which is between 1.5 and 2.0 metres.

Definition 3. A *probability measure* $\mathbb{P}$ is a real-valued set function defined on a sample space S which satisfies

1. $0 \leq \mathbb{P}[A] \leq 1$ for every event A
2. $\mathbb{P}[S] = 1$
3. $\mathbb{P}[A_1 \cup A_2 \cup \ldots] = \mathbb{P}[A_1] + \mathbb{P}[A_2] + \ldots$ for every finite or infinite sequence of disjoint events $A_1, A_2, \ldots$.

Example: The probability of throwing two or more heads when throwing three unbiased coins is the probability of the event {HHH HHT HTH THH}. This is a union of four disjoint simple events and is ($4/8 = 1/2$) since there are eight possibilities in total each with equal probability $1/8$.

Example: The Poisson distribution is given by

$$\mathbb{P}[X = x] = \frac{\lambda^x e^{-\lambda}}{x!} \qquad x \in \mathbb{Z}_+.$$

It is left as an exercise for the student to show that the first and second condition are both met. That is,

$$0 \leq \frac{\lambda^x e^{-\lambda}}{x!} \leq 1 \qquad x \in \mathbb{Z}_+,$$

and

$$\sum_{x=0}^{\infty} \frac{\lambda^x e^{-\lambda}}{x!} = 1.$$

The notation $\mathbb{P}[A, B]$ refers to the probability that events A and B both occur also known as the *joint probability*. The notation $\mathbb{P}[A|B]$ is the probability that event A occurs given that event B occurs or the *conditional probability*.

Example: If A is the event "the first coin shows H" and B is the event "more than two heads are thrown" then of the above examples, A and B both occur for HHH, HHT and HTH — there are four possibilities which have A true HTT is the remaining one. Therefore $\mathbb{P}[B|A] = 3/4$ — by coincidence $\mathbb{P}[A|B] = 3/4$ as well in this case (this is not generally true). In this case $\mathbb{P}[A, B] = 3/8$.

Theorem 1. Bayes theorem states that

$$\mathbb{P}[A, B] = \mathbb{P}[B]\,\mathbb{P}[A|B]$$

The reason for this can be trivially seen. The probability of A and B is the probability of B multiplied by the probability of A given that B has occurred.

Definition 4. A *random variable* is a real-valued function defined on a sample space. For example, X might be the number of heads in two coin tosses or the height of a given measurement in metres. The domain of X is the sample space and its range is within the real numbers $\mathbb{R}$. A *discrete random variable* is a random variable defined on a discrete sample space and a *continuous random variable* is a random variable defined on a continuous sample space for which the probability is zero that it will assume any given value in an interval.

It should also be noted that it follows from these definitions that a real-valued function of a random variable (or a set of random variables) is itself a random variable.

Definition 5. The *discrete density function* $f(x)$ for a discrete random variable X is given by the equation

$$f(x) = \mathbb{P}[X = x].$$

The *distribution function* (sometimes called the cumulative distribution function) $F(x)$ for a discrete random variable X is given by

$$F(x) = \mathbb{P}[X \leq x] = \sum_{y \leq x} \mathbb{P}[X = y].$$

Definition 6. The *continuous density function* $f(x)$ for a continuous random variable X is uniquely determined by the following properties:

1. $f(x) \geq 0$ for all $x \in \mathbb{R}$
2. $\int_{-\infty}^{\infty} f(x)dx = 1$
3. $\int_a^b f(x)dx = \mathbb{P}[a < x < b]$ for all $a, b \in \mathbb{R}$ where $a \leq b$.

Example: Consider the so called "flat" or "constant" distribution where the sample space is some interval (a, b) and $\mathbb{P}[c < X < d] \propto (d - c)$ where $a \leq c \leq d \leq b$. The third part of the definition will give us that

$$f(x) = \begin{cases} k & a < x < a \\ 0 & \text{otherwise,} \end{cases}$$

where k is some constant. From the second part we get that

$$\int_a^b k dx = 1,$$

and hence $k = 1/(b-a)$.

The distribution function (sometimes called the cumulative distribution function) is the sum of the density function,

$$F(x) = \int_{-\infty}^{x} f(y)dy,$$

where $f(y)$ is the density function.

Example: For the flat distribution defined above it is left as an exercise to the student to show that

$$F(x) = \begin{cases} 0 & x \leq a \\ (x-a)/(b-a) & a < x < b \\ 1 & x \geq b. \end{cases}$$

Often it is useful to deal with more than one random variable at once. If two variables X and Y are considered then it is useful to know probabilities about what happens with both variables.

Definition 7. The *joint density function* of two random variables X and Y is defined by $f(x,y)$. In the discrete case this is defined by the equation

$$f(x,y) = \mathbb{P}\left[\, X = x, Y = y \,\right].$$

In the continuous case it must possess the following properties:

1. $f(x,y) \geq 0$
2. $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)dxdy = 1$
3. $\int_{a}^{b}\int_{c}^{d} f(x,y)dxdy = \mathbb{P}\left[\, a < X < b \text{ and } c < Y < d \,\right]$, for all $a \leq b \in \mathbb{R}$ and $c \leq d \in \mathbb{R}$.

Definition 8. The random variables X and Y with density functions $g(x)$ and $h(x)$ and the joint density function $f(x,y)$ are said to be *independent* if and only if

$$f(x,y) = g(x)h(y),$$

for all x and y.

Definitions 7 and 8 can be extended in the obvious way to more than two variables. It should be noted, however, that merely because each pair of events is independent does NOT mean that the entire set of events is independent. It is instructive to come up with examples where this is not the case.

Definition 9. The *expected value* or *expectation* of the function $g(X)$ on a discrete random variable X is given by

$$\mathrm{E}\left[\, g(X) \,\right] = \sum_{i=1}^{\infty} g(x_i)f(x_i),$$

where x_i are all the possible values of X (that is all the members of its sample space) and $f(x)$ is the density fucntion for X.

For a continuous variable the sum in the above changes to an integral.

Definition 10. The *expected value* or *expectation* of the function $g(X)$ on a continuous random variable X is given by

$$\mathrm{E}\left[\, g(X) \,\right] = \int_{-\infty}^{\infty} g(x)f(x)dx,$$

where $f(x)$ is the density function for X.

It should be noted that in Defintions 9 and 10 there is no guarantee that either the sum or the integral converge. If they diverge then the expectation is undefined.

We can extend the definition of expectation to a set of random variables $X_1, \ldots, X_n$.

Definition 11. For random variables $X_1, \ldots, X_n$ with density function $f(x_1, \ldots, x_n)$ then the expectation value of a function $h(X_1, \ldots, X_n)$ is given by:

$$\mathrm{E}\left[\, h \,\right] = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} h(x_1, \ldots, x_n) f(x_1, \ldots, x_n) dx_1 \ldots dx_n.$$

Expectation E is a linear operator. If g, g_1 and g_2 are three functions of a set of random variables then the following properties follow from the previous definitions:

- $\mathrm{E}\left[\, cg \,\right] = c\mathrm{E}\left[\, g \,\right]$ for any constant c.
- $\mathrm{E}\left[\, g_1 + g_2 \,\right] = \mathrm{E}\left[\, g_1 \,\right] + \mathrm{E}\left[\, g_2 \,\right]$.
- $\mathrm{E}\left[\, g_1 g_2 \,\right] = \mathrm{E}\left[\, g_1 \,\right] \mathrm{E}\left[\, g_2 \,\right]$, if g_1 and g_2 are independent.

The first two properties follow trivially from substituting $h = cg$ and $h = g_1 + g_2$ into Definition 11. The third property is derived as follows:

$$\mathrm{E}\left[\, g_1 g_2 \,\right] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g_1 g_2 f(g_1, g_2) dg_1 dg_2,$$

where $f(g_1, g_2)$ is the joint density function of g_1 and g_2. Since g_1 and g_2 are independent then from Definition 8

$$\begin{aligned}
\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g_1 g_2 f(g_1, g_2) dg_1 dg_2 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g_1 g_2 f_1(g_1) f_2(g_2) dg_1 dg_2 \\
&= \int_{-\infty}^{\infty} g_1 f_1(g_1) dg_1 \int_{-\infty}^{\infty} g_2 f_2(g_2) dg_2 \\
&= \mathrm{E}\left[\, g_1 \,\right] \mathrm{E}\left[\, g_2 \,\right],
\end{aligned}$$

where $f_1(g_1)$ and $f_2(g_2)$ are the density functions of g_1 and g_2 respectively.

Using the Definitions 9 and 10 for expectation then *mean* μ and *variance* σ^2 of a random variable X can be defined.

Definition 12. The *mean* μ of a random variable X (either discrete or continuous) is given by

$$\mu = \mathrm{E}\left[\, X \,\right].$$

Definition 13. The *variance* σ^2 of a random variable X (either discrete or continuous) is denoted by $\mathrm{var}\left(\, X \,\right)$ and is given by

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

Examples: Show that the flat distribution as previously defined, has mean $(a+b)/2$ and variance $(b-a)^2/12$. Show that the Poisson distribution has mean λ and variance λ.

As previously noted, the expectation is not guaranteed to converge and, for some random variables, μ and σ^2 do not exist.

Instructive Example: What is the expected payout of a coin-tossing game defined as follows. The player tosses a coin until they get a head. If the first head occurs on throw n then they are paid £2^n. Calculate the expected payout.