STEP Support Programme

STEP 2 Statistics Topic Notes

Probability A good introduction to basic probability can be found here.

- $\mathrm{P}(A \cap B)$ means the probability that both A and B happen.
  If A and B are *mutually exclusive* then $\mathrm{P}(A \cap B) = 0$. Mutually exclusive events cannot happen together.
  If A and B are *independent* then $\mathrm{P}(A \cap B) = \mathrm{P}(A) \times \mathrm{P}(B)$. If A and B are independent then whether A happens or not has no effect on the probability of B happening and vice versa. Mutually exclusive events cannot be independent.
- The *complement* to A is "A doesn't happen" or "not A". It can be written as A' and we have $\mathrm{P}(A') = 1 - \mathrm{P}(A)$. The probabilities of *complementary* events sum to 1.
- $\mathrm{P}(A \cup B)$ means the probability that A or B (or both) happen.
- $\mathrm{P}(A \cup B) = \mathrm{P}(A) + \mathrm{P}(B) - \mathrm{P}(A \cap B)$. To show this, draw a Venn diagram with two overlapping circles, the area inside one representing $\mathrm{P}(A)$ and the other representing $\mathrm{P}(B)$.
- $\mathrm{P}(A|B)$ means the probability that event A happens **given** that we know event B happens. This is a *conditional* probability. A lot of conditional probability questions can be done informally using tree diagrams, or by considering a population (e.g. $100,000$ people).
- **Bayes' Theorem**

  $$\mathrm{P}(A|B) = \frac{\mathrm{P}(A \cap B)}{P(B)}\,.$$

  It can also be useful to consider the equivalent statement $\mathrm{P}(A \cap B) = P(B) \times P(A|B)$.

  Writing $\mathrm{P}(A \cap B)$ in two different ways and equating gives us $\mathrm{P}(A|B) = \dfrac{\mathrm{P}(B|A) \times \mathrm{P}(A)}{P(B)}$.

Combinations, permutations and arrangements

- The number of ways of choosing[1] r objects from n objects is ${}^nC_r = \dbinom{n}{r} = \dfrac{n!}{r!(n-r)!}$.
- The number of permutations[2] of r objects taken from a selection of n different objects is ${}^nP_r = \dfrac{n!}{(n-r)!}$.
- The number of different arrangements[3] of r objects where a of them are all the same, another b are all the same (but different to the first lot) etc. is $\dfrac{r!}{a! \times b! \times \cdots}$.

[1] "Choosing" implies that the order doesn't matter.
[2] "Permutations" means that order does matter.
[3] Order matters.

Discrete Probability Distributions

"Discrete" means that only certain values can be taken (such as the numbers on a dice — we cannot get a value between 2 and 3)[4].
Let X be a discrete random variable.

- The *expectation* (or mean) is given by:

  $$\mathrm{E}(X) = \sum i \times \mathrm{P}(X = i).$$

  So if the possible values of X are $1, 2, \ldots, k$ then
  $\mathrm{E}(X) = 1 \times \mathrm{P}(X = 1) + 2 \times \mathrm{P}(X = 2) + \ldots + k \times \mathrm{P}(X = k)$.

- The *variance* is given by:

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

  where $\mathrm{E}(X^2) = \sum i^2 \times \mathrm{P}(X = i)$. $\mathrm{Var}(X)$ is never negative, and can be thought of as "the mean of the squares − the square of the mean"[5].

- The *Mode* or *Modal Value* is the value of x for which $\mathrm{P}(X = x)$ is the greatest (there may be more than one mode).

The *binomial* distribution, $X \sim B(n, p)$, is the distribution of the number of "successes" in a sequence of n independent yes/no trials each of which has a probability p of success. An example would be the number of sixes you get when you roll a dice 10 times.

- $\mathrm{P}(X = r) = \dbinom{n}{r} p^r (1 - p)^{n-r}$
- $\mathrm{E}(X) = np$
- $\mathrm{Var}(X) = np(1 - p)$

The *Poisson* distribution, $X \sim \mathrm{Po}(\lambda)$, is the distribution of the number of occurrences of an event in a given "interval" (which can be time, length, etc.). An example could be the number of meteors greater than 1 metre diameter that strike earth in a year.

- $\mathrm{P}(X = r) = \dfrac{\mathrm{e}^{-\lambda}\lambda^r}{r!}$
- $\mathrm{E}(x) = \mathrm{Var}(X) = \lambda$
- If the number of occurrences in an interval of length T follows a Poisson distribution with mean λ, then the number of occurrences in an interval of length kT follows a Poisson distribution with mean $k\lambda$.
- if X and Y are two *independent* Poisson random variables with means λ and μ respectively then $X + Y$ has a Poisson distribution with mean $\lambda + \mu$.
- If n is "large" and p is "very small" then a Poisson distribution with mean np can be used to approximate a Binomial distribution, $X \sim B(n, p)$.

[4] There can be infinitely many values, such as the number of coin tosses until you get a head, or non-integer values, such as the value of one dice roll divided by another dice roll — we can still only get certain values, such as $\frac{4}{3}$, but not others, such as $\frac{2}{7}$.

[5] The above definition of variance is usually the easiest to work with, but variance is really the mean squared distance of values from the mean. This gives $\mathrm{Var}(X) = \mathrm{E}\left((X - \mathrm{E}(X))^2\right) = \sum (i - \mathrm{E}(X))^2 \times \mathrm{P}(X = i)$ which can be expanded to give the above result.

Continuous Probability Distributions

"Continuous" means that all the values in a certain range are possible. Examples include height of a person, or the half life of a radioactive element. Continuous random variables are usually defined by a probability distribution function, $\mathrm{f}(x)$.

- $\mathrm{P}(a \leqslant X \leqslant b) = \displaystyle\int_a^b \mathrm{f}(x)\,\mathrm{d}x$

- $\displaystyle\int_{-\infty}^{\infty} \mathrm{f}(x)\,\mathrm{d}x = 1$ (as the total probability must be 1). This means that the total area under the curve $y = \mathrm{f}(x)$ must be 1.

  If $\mathrm{f}(x)$ is equal to 0 for some ranges of x then you will be able to change the limits accordingly. For example if $\mathrm{f}(x) = kx$ for $0 \leqslant x \leqslant 10$, and is equal to 0 elsewhere, then we could write $\int_0^{10} \mathrm{f}(x)\,\mathrm{d}x = 1$ (and hence find the value of k).

- The *expectation* (or mean) is given by:

$$\mu = \int_{-\infty}^{\infty} x\mathrm{f}(x)\,\mathrm{d}x\,.$$

- The *variance* is given by:

$$\int_{-\infty}^{\infty} x^2\mathrm{f}(x)\,\mathrm{d}x\, - \mu^2.$$

  Note that the formulae for expectation and variance of a continuous distribution are very similar to the ones for a discrete distribution, all that has happened is that the sum has been replaced by an integral. In fact Leibniz considered integration to be an infinite sum of infinitesimal "bits". and so he based the integral symbol $\int$ on the "long s" character (for "summation").

- The *cumulative distribution function* is defined by:

$$\mathrm{F}(x) = \mathrm{P}(X \leqslant x) = \int_{-\infty}^{x} \mathrm{f}(t)\,\mathrm{d}t$$

  Here we have taken the lower limit as $-\infty$. It may be that $\mathrm{f}(x) = 0$ for $-\infty < x < a$ say, in which case we could write the lower limit as a. Note the use of "dummy variable" t inside the integral — we cannot use x inside the integral as it is used as a limit.

- The *median*, m satisfies $P(X \leqslant m) = P(X \geqslant m) = \frac{1}{2}$, i.e. $\displaystyle\int_{-\infty}^{m} \mathrm{f}(x)\,\mathrm{d}x = \int_{m}^{\infty} \mathrm{f}(x)\,\mathrm{d}x = \frac{1}{2}$.

- The *mode* is where the probability distribution function has a maximum (there may be more than one!).

The Normal distribution $X \sim N(\mu, \sigma^2)$

- If $X \sim B(n, p)$ and n is "large" and/or p is "close to" $\frac{1}{2}$ then X can be approximated by a normal distribution, $X \sim N(np, np(1-p))$.

- If $X \sim \mathrm{Po}(\lambda)$ and λ is"large"then X can by approximated by a normal distribution, $X \sim N(\lambda, \lambda)$.

More on the Poisson Distribution

The *Poisson* distribution measures the number of occurrences of an event in a given time interval. It was first used by Ladislaus Josephovich Bortkiewicz to model the number of deaths of Prussian cavalry-men by horse kicks in a year.

A Poisson random variable satisfies the following conditions:

I Occurrences are independent.

II The mean (or expected) number of occurrences during a time interval is proportional to the length of the time interval.

As well as modelling the number of occurrences in a given time interval it can be used to model the number of occurrences in a given space interval. Some applications are the number of car accidents in a mile of road, the number of people joining a queue every 5 minutes and the number of hairs in a burger.

The number of occurrences in a given time interval is given by:

$$\mathrm{P}(X = n) = \frac{\mathrm{e}^{-\lambda}\lambda^n}{n!}$$

where n is an integer, with $n \geqslant 0$, λ is the mean number of occurrences in the given interval and (by convention) $0! = 1$.

Note that the sum of all the probabilities is given by:

$$\sum_{n=0}^{\infty} \frac{\mathrm{e}^{-\lambda}\lambda^n}{n!} = \mathrm{e}^{-\lambda} \times \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} = \mathrm{e}^{-\lambda} \times \mathrm{e}^{\lambda} = 1\,.$$

For the last equality, we used the exponential series. You may like to show that $\mathrm{E}(X) = \lambda$.