Discrete probability functions and the mean and variance

Discrete probability functions

We have talked about outcomes to random experiments and the individual probabilities of each outcome, or a set of outcomes. It is convention, if possible, to list all possible outcomes and their associated probabilities in tabular form to summarise an experiment. This is called **the probability distribution of X**.

Notation:- The random variable associated with an experiment has a **capital letter**, whereas a particular outcome is indicated by the equivalent **lower case letter**.

Example:- A fair coin is tossed twice and the total number of heads, X, is noted. A convenient sample space for the experiment is a list of all paired outcomes, i.e.,

$$S = \{TT, TH, HT, HH\}.$$

Each outcome is equally likely and so we associate the probability of $\frac{1}{4}$ to each. The possible values of X are 0,1 and 2, with probabilities $\frac{1}{4}, \frac{2}{4}$ and $\frac{1}{4}$ respectively. These are displayed in tabular form as:-

x	0	1	2
p_x	$\frac{1}{4}$	$\frac{2}{4}$	$\frac{1}{4}$

Note that $\sum_{all\ x} p_x = \frac{1}{4} + \frac{2}{4} + \frac{1}{4} = 1$, as it should be, to be a valid probability distribution.

Note:- If you are asked to find the probability distribution of a random variable X you either display it in tabular form, showing all (x, p_x) pairs, or define this pair mathematically.

Cumulative distribution function

The notation for a cumulative distribution function of a random variable X is usually $F(x)$ and it is defined as $F(x) = P(X \leq x)$ for all values of $x \in (-\infty, \infty)$

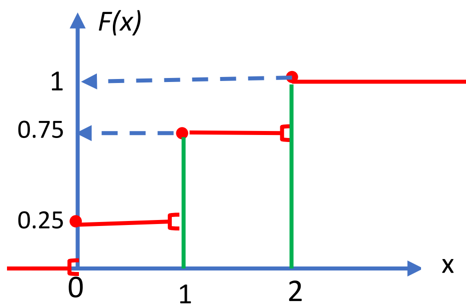
For the above example

$$F(x) = 0 \quad if \quad x < 0$$

$$F(x) = \tfrac{1}{4} \quad if \quad 0 \leq x < 1$$

$$F(x) = \tfrac{3}{4} \quad if \quad 1 \leq x < 2$$

$$F(x) = 1 \quad if \quad 2 \leq x$$



The dot means the point is included and the open bracket means the point is excluded. The size of the jump in the curve is the probability at that point. So, we can see that probability only occurs at the points 0,1 and 2 with sizes of 0.25, 0.5 and 0.25 respectively. Note the value of $F(x)$ always starts from zero to the far left and finishes up at 1 to the far right. It can never go down because that would mean a negative probability which isn't allowed. It is also clear that probabilities are found by differencing the values of $F(x)$ at two points i.e., $P(X = 1) = F(1) - F(0) = 0.75 - 0.25 = 0.5$.

Note:- Most statistical tables plot $F(x)$ not the probability function.

Mean and Variance of a Random Variable, or equivalently it's distribution

Suppose the above experiment was repeated N times and the values 0,1 and 2 occurred f_0, f_1 and f_2 respectively. The straightforward average of these numbers would be

$$\bar{x} = \frac{\sum_{i=1}^{N} x_i}{N} = \frac{(0\times f_0 + 1\times f_1 + 2\times f_2)}{N} = 0 \times \left(\frac{f_0}{N}\right) + 1 \times \left(\frac{f_1}{N}\right) + 2 \times \left(\frac{f_2}{N}\right)$$

The terms in brackets are the relative frequencies of each event and we define these to be the equivalent probabilities as $N \to \infty$. This prompts the following definition for the mean, written E(X), of the random variable X.

If the probability distribution of X is the paired values (x, p_x) listed for all x, then

$$\boldsymbol{E(X) = \sum_{all\ x} x p_x = \mu},$$ where μ is the usual notation for the mean.

Generally, if you have any function of X, its mean value is

$$\boldsymbol{E(U(X)) = \sum_{all\ x} u(x) p_x}$$

The mean is the most popular measure of centre but others include the median and mode. To describe a distribution, we need to indicate where its centre is, and how spread out(dispersed) it is. For each observation x_i, it's deviation from centre, using the mean, is $x_i - \mu$. If $\mu = 2$ and $x_i = 3$ the deviation is 3 – 2 = 1. However, if $x_i = 1$ the deviation is -1. From the point of view of spread, both are the same distance from the mean μ and we wouldn't really want to distinguish between them. We can remove the sign problem by squaring the deviation and so we evaluate the average value of $(X - \mu)^2$. We call this average the variance of the distribution of X and denote it by σ^2. It follows that the

Variance of X is $\boldsymbol{\sigma^2 = E((X-\mu)^2) = \sum_{all\ x}(x-\mu)^2 p_x}$. Now, $\boldsymbol{\sigma = \sqrt{\sigma^2}}$, has the same units as the random variable X, and is called the **standard deviation of X**, which is also used to indicate the dispersion of the distribution of X.

Note:- A little algebra will confirm that $\boldsymbol{\sigma^2 = E((X-\mu)^2) = E(X^2) - E^2(X)}$

A useful result:- If a and b are any two constants then

$$\boldsymbol{E(aX+b) = aE(X) + b} \quad \textbf{and} \quad \boldsymbol{Var(aX+b) = a^2 Var(X)}$$