Random Variables

Generally the object of an investigators interest is not necessarily the action in the sample space but rather some function of it. Technically a real valued function or mapping whose domain is the sample space is called a “Random Variable,” it is these that are usually the object of an investigators attention. If the mapping is onto a finite (or countably infinite) set of points on the real line, the random variable is said to be discrete. Otherwise, if the mapping is onto an uncountably infinite set of points, the random variable is continuous. This distinction is a nuisance because the nature of thing which describes the probabilistic behaviour of the random variable, called a **Probability Density Function** (denoted here as f(x) and referred to as a p.d.f.), will differ according to whether the variable is discrete or continuous.

In the case of a discrete random variable X, with typical outcome x_i, (it shall be assumed for convenience that the x_i's are ordered with I from smallest to largest), the probability density function $f(x_i)$ is simply the sum of the probabilities of outcomes in the sample space which result in the random variable taking on the value x_i. Basically the p.d.f. for a Discrete Random Variable obeys 2 rules:

i. $0 \leq f(x_i) \leq 1$

ii. $\sum_{all\ possible\ x_i} f(x_i) = 1$

In the discrete case $f(x_i) = P(X=x_i)$, in the continuous case it is not possible to interpret the p.d.f. in the same way indeed, since X can take on any one of an uncountably infinite set of values, we cannot give it a subscript “i”. when the random variable X is continuous with typical value x, the probability density function f(x) is a function that, when integrated over the range (a, b), will yield the probability that a sample is realized such that the resultant x would fall in that range (much as the probability function for continuous sample spaces was defined above). In this case the p.d.f. will obey three basic rules:

1. $f(x) \geq 0.$
2. $\int_a^b f(x)dx = P(a \leq x \leq b)$
3. $\int_{-\infty}^{\infty} f(x)dx = 1$

Associated with these densities are Cumulative Distribution Functions F(x) which in each case yield the probability that the random variable X is less than some value x. Algebraically these may be expressed as:

$$P(X \le x) = F(x) = \sum_{t=1}^{k} f(x_i); \quad where\ x_k \le x \le x_{k+1}$$

$$P(X \le x) = F(x) = \int_{-\infty}^{x} f(z)dz$$

Note that in the case of discrete random variables F(x) is defined over the whole range of the random variable and in the case of continuous distributions d(Fx)/dx = f(x) i.e. the derivative of the c.d.f. of x gives us the p.d.f. of x.

Expected Values and Variances

The Expected Value of a function g(x) of a random variable is a measure of its location and is defined for discrete and continuous random variables respectively as follows:

$$E(g(X)) = \sum_{all\ possible\ x_i} g(x_i) f(x_i)$$

$$E(g(X)) = \int_{-\infty}^{+\infty} g(x) f(x)dx$$

The expectations operator E( ) is simply another mathematical operator. Just as when the operator d/dx in front of a function g(x) tells us to "take the derivative of the function g(x) with respect to x according to a well specified set of rules" so E(g(X)) tells us to perform one of the above calculations dependent upon whether X is continuous or discrete. Like the derivative and integral operators, the expectations operator is a linear operator so that the expected value of a linear function of random variables is the same linear function of the expected values of those random variables. This property will be of useful shortly.

Aside from the general applicability of the above formulae there are many special types of g( ) function of interest to statisticians, each generating particular details of the nature of the random variable in question, things like moment generating functions and characteristic functions that are the material of a more advanced text and things like the Expected Value function and Variance function that are of interest to us here.

Two g( ) functions of special interest

1. g(X) = X

Obviously this yields E(X), the expected value of the random variable itself (frequently referred to as the mean and represented by the character μ), which is a constant providing a measure of where the centre of the distribution is located. The metric here is the same as that of the random variable so that, if f(x) is an income distribution measured in \$US, then its location will be in terms of a \$US value. The usefulness of the linearity property of the expectations operator can be seen by letting g(X) = a + bX where a and b are fixed constants. Taking expectations of this g(X) yields the expected value of a linear function of X which, following the respective rules of summation and integration, can be shown to be the same linear function of the expected value of X as follows:

$$E(a+bX) = \sum_{all\ possible\ x_i}(a+bx_i)f(x) = a\sum_{all\ possible\ x_i} f(x_i) + b\sum_{all\ possible\ x_i} xf(x_i) = a + bE(X)$$

$$E(a+bCX) = \int_{-\infty}^{+\infty}(a+bx)f(x)dx = a\int_{-\infty}^{+\infty} f(x)dx + b\int_{-\infty}^{+\infty} xf(x)dx = a + bE(X)$$

What the above demonstrates is two basic rules for expectations operators regardless of whether random variables are discrete or continuous:

i. **E(a) = a** "The expected value of a constant is a constant"

ii. **E(bX) = bE(X)** "The expected value of a constant times a random variable is equal to the constant times the expected value of the random variable"

2. $g(X) = (X-E(X))^i$ i = 2,…

Functions of this form yield "i"th moments about the mean." The metric here is the i'th power of that of the original distribution so that if f(x) relates to incomes measured in \$US then the i'th moment about the mean is measured in $(\$US)^i$. Sometimes the i'th root of g(X) is employed since it yields a measure of the appropriate characteristic in terms of the original units of the distribution. Furthermore to make distributions measured under different metrics comparable the function g(X) deflated by the appropriate power of E(X) (provided it is not 0) is considered, providing a metric free comparator. The second moment (i=2) is of particular interest[1] since as the variance (frequently represented as σ^2 or V(X), its square root is referred to as the standard deviation) it provides a measure of how spread out a distribution is. Of particular interest here is the Coefficient of Variation (CV) given by:

The Coefficient of Variation:

$$CV = \sqrt{\frac{E(X-E(X))^2}{E(X)^2}} = \frac{\sigma}{\mu} \qquad (1)$$

[1] The third moment provides a measure of skewness or how asymmetric a distribution is and the fourth, a measure of kurtosis or how peaked a distribution is but these will not be considered here.

Which may be interpreted as a metric free measure of dispersion.

Returning to the variance, regardless of whether the variable is discrete or continuous, notice that:

$$\begin{aligned} V(x) &= E((X-E(X))^2) \\ &= E(X^2 - 2E(X)X + (E(X))^2) \\ &= E(X^2) - 2E(E(X)X) + (E(X))^2 \\ &= E(X^2) - (E(X))^2 \end{aligned}$$

So that the variance is equal to the expected value of X^2 less the square of the expected value of X (these are not the same thing). Furthermore, again regardless of whether the random variable is discrete or continuous notice that for Y = a + bX:

$$\begin{aligned} V(Y) &= E((Y-E(Y))^2) \\ &= E((a + bX - (+bE(X)))^2) \\ &= E((bX - bE(X))^2) \\ &= b^2E((X-E(X))^2) \\ &= b^2V(X) \end{aligned}$$

So that the variance of a constant is zero and the variance of a constant times a random variable is the square of the constant times the random variable. The variance of a linear function of several random variables is a little more complicated, depending as it does on the relationships between the random variables it will be dealt with when multivariate analysis is considered later on. There are numerous discrete and continuous probability density functions to suit all kinds of purposes, one of the arts in practicing statistics is that of choosing the one most appropriate for a particular problem. Here we shall consider two examples of each to get an idea of what they are like.

Examples of Discrete Probability Density Functions

1) The Binomial Distribution.

The Binomial Distribution is founded upon a process in which the same experiment is **independently** repeated n times under **identical conditions**. The Sample Space for the experiment contains two possible events A and A^c and the issue at hand is how many times in n repetitions A occurs. Suppose that P(A)=p (and consequently $P(A^c)$=1-p) and when A happens in the i'th repetition of the experiment) so that Xi is a random variable such that $P(X_i = 1) = p$ and $P(X_i=0) = 1-p$ for i = 1, 2, …, n. Letting x_i be the outcome of the i'th experiment f(xi), the p.d.f. for the i'th experiment is given by:

$$f(x_i) = p^{x_i}(1-p)^{1-x_i}$$

So that when X_i returns a 1 the p.d.f. is p and when it returns a zero the p.d.f. is 1-p. using the notions of Expected Values and Variances it can be shown that E(Xi)=p and $V(x_i)=p(1-p)$ as follows:

$$E(X_i) = \sum_{all\ possible x_i} xf(x_i) = 0.p^0(1-p)^{1-0} + .1p^1(1-p)^{1-1} = p$$

$$V(X_i) = \sum_{all\ possible\ x_i}(x_i - E(x_i))^2 f(x_i) = (0-p)^2.p^0(1-p)^{1-0} + (1-p)^2.p^1(1-p)^{1-1}$$
$$= p^2(1-p) + (1-p)^2 p = p^2 - p^3 + p - 2p^2 + p^3 = p - p^2$$

Since the repetitions of the experiment are independent the joint probability of the n repetitions is:

$$f(x_1)f(x_2)...f(x_n) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} = p^{\sum_{i=1}^{n} x_i}(1-p)^{(n-\sum_{i=1}^{n} x_i)}$$

If A occurs k times the sum of the x_i's will equal k and this formula may be written as $p^k(1-p)^{n-k}$ and corresponds to the probability of getting a particular sequence of experiments where A occurred k times. The number of ways that A could happen k times in n experiments is n!/(k!(n-k)!) so that the probability of k occurrences in n experiments is given by:

$$\tfrac{n!}{k!(n-k)!} p^k(1-p)^{(n-k)} \quad k = 0,1,2,...,n.$$

Which is the Binomial Distribution.

The Poisson Distribution

The Poisson Distribution is employed in situations where the object of interest is the number of times given event occurs in a given amount of space or period of time. Thus for example, it could be used to study the number of crashes that take place at a particular spot over a period of a week or it could be used to investigate the number of faults in a fixed length of steel. The presumption in this model is that successive weeks, or successive lengths of steel are independent of one another and that the same probability model is applicable in each successive observation. This is much the same i.i.d. assumption we made in the case of the Binomial Distribution.

Letting x be the number of occurrences of the event the p.d.f. is given by:

$$f(x) = \tfrac{\lambda^x e^{-\lambda}}{x!} \quad \lambda > 0,\ x = 0,1,2,...$$

The unknown parameter is such that $E(x) = \lambda$ and $V(x) = \lambda$.

Examples of Continuous Probability Density Functions

The Uniform or Rectangular Distribution

This is perhaps the simplest distribution available. It describes the behaviour of a continuous random variable X that exists in the interval [a, b] whose probability of being in an interval [c, d] laying within [a, b] is given by (d-c)/(b-a). That is to say the probability of it laying in any interval in its range is equal to the proportionate size of the interval within the range. Furthermore X will have the same probability of landing in any one of a collection of equal sized intervals in the range (hence the name "uniform"). This distribution with a=0 and b=1 is frequently used as the basis for random number generators in software packages and computer games, largely because it is relatively easy to generate other more complex random variables from it.

Formally:

$$f(x) = \tfrac{1}{b-a} \text{ for } a \le X \le b$$
$$= 0 \text{ otherwise}$$

$$F(x) = 0 \text{ for } x < a$$
$$\tfrac{x-a}{b-a} \text{ for } a \le X \le b$$
$$= 1 \text{ for } x > b$$

The Normal Distribution

The normal distribution is probably the most frequently employed distribution in statistics with good reason, there are sound theoretical reasons why it can be employed in a wide range of circumstances where averages are used. Its p.d.f. is of the form:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

The parameters μ and σ^2 respectively correspond to the mean and variance of the random variable x. The fact that X is normally distributed with a mean μ and a variance σ^2 is often denoted by $X \sim N(\mu, \sigma^2)$. The normal distribution does not have a closed form representation for the cumulative density F(X) (i.e. we cannot write down an algebraic expression for it) but this will not present any difficulties since it is tabulated and most statistical software packages are capable of performing the appropriate calculations. The distribution is symmetric about the mean, bell shaped (hence the terminology "are you going to bell the mark sir?") with extremely thin tails to the extent that more than 99% of the distribution lays within $\mu \pm 2\sigma$.

Normal random variables possess the very useful property that linear functions of them are also normal. Hence if X is normal then Z = a + bX is also normal, and using our rules for expectations, E(Z) = a + bE(X) and V(Z) = bV(X). Letting a = $-\mu/\sigma$ and b = $1/\sigma$, Z ~ N(0, 1) which is referred to as a Standard Normal Random Variable (indeed the standard normal variable is frequently referred to with the letter z, hence the term "z score". This is most useful since N(0.1) is the distribution that is tabulated in textbooks and programmed in software packages. Suppose we need to calculate $P(X<x_0)$ is some known value and X ~ N(μ, σ^2) where the mean and variance are known values. Then since:

$$P(X < x_0) = P(X - \mu < x_0 - \mu) = P(\tfrac{X-\mu}{\sigma} < \tfrac{x_0-\mu)}{\sigma})$$

And since Z = $(X-\mu)/\sigma$:

$$P(X < x_0) = P(Z < \tfrac{x_0-\mu}{\sigma})$$

So all that is needed is to calculate the value $(x_0-\mu)/\sigma$ and employ the standard normal tables or software package to evaluate the probability.