Chapter 3

Continuous random variables

3.1 Definitions, Mean and Variance

In Chapter 2 we defined a continuous random variable as one which can take. any value in an interval. Some continuous random variables are defined on a *finite* interval, others on the whole real line or the positive half of the real line Whichever is the case, we cannot give a non-zero probability to any particular value. Basically, if we could, then when we added up the uncountably many non-zero probabilities, they would sum to more than one. Instead we can find the probability that a continuous random variable lies in an interval. Note that because the probability of the end-points is zero it does not matter whether we include or exclude them.

Definition : If X is a continuous random variable then there exists a non-negative function, *f(x)*, called the *probability density function* of *X* such that

$$\int_{-\infty}^{\infty} f(x)dx = 1$$

And

$$\text{and} \quad P(a < X < b) = \int_{a}^{b} f(x)dx$$

Note that any function which is non-negative and integrates to one is a possible probability density function for a random variable X. As with discrete random variables some density functions are commonly used to model continuous random variables.

It is also convenient to define the following function.

Definition The *cumulative distribution function, F(x)* of a continuous ran variable x is defined by

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

Note that for a discrete random variable the cumulative distribution function $P(X \le x)$ will be a step function with steps of height $P(X = x)$ at the points at which X is defined. The continuous version can be thought of as a limiting case when all values of x in an interval are possible. Note that the cumulative distribution function is always non-decreasing and

$$\lim_{x \to \infty} F(x) = 0 \qquad \lim_{x \to \infty} F(x) = 1$$

Knowledge of either the density function or the cumulative distribution function is enough to define *X*.

We define the mean of a continuous random variable as follows.

Definition If *X* is a continuous random variable with probability density function *f(X)* then the mean or expected value of *X*, *E[X]* or p is defined by

$$E[X] = \mu = \int_{-\infty}^{\infty} x f(x) dx$$

We define the expectation of a function of *X* in a similar way

Definition if X is a continuous random variable with probability density function f(X) then the expected value of g(X) is defined by

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

Similarly the variance is defined by

Definition If *X* is a continuous random variable with probability density function f(x) then the variance of *X*, *Var [X]* is defined by

$$Var[x] = \int_{-\infty}^{\infty} (x-\mu)^2 f(x) dx$$

$$= \int_{-\infty}^{\infty} x^2 f(x) dx - \mu^2$$

We can also define the median of a continuous random variable.

Definition if *X* is a continuous random variable with probability density function f(s) then the median of *x* is the value *m* satisfying the equation

$$\int_{-\infty}^{m} f(x) dx = \int_{m}^{\infty} f(x) dx = \frac{1}{2}$$

It is the value such that *X* is equally likely to be more than the median as less than it.

Example: A random variable *X* has probability density function

$$f(x) = \begin{cases} cx^2(1-x) \quad if \quad 0 \le x \le 1 \\ 0 \qquad \text{otherwise} \end{cases}$$

1. Determine *c*.

2. find *E[X]*.

3. Find *Var[X]*.

4. Show that the median *m* satisfies the equation

$$6m^4 - 8m^3 + 1 = 0$$

1. We know that $\int_{-\infty}^{\infty} f(x)dx = 1$

so

$$\int_0^1 c(x^2 - x^3)dx = 1$$

$$\left[c(\tfrac{x^3}{3} - \tfrac{x^4}{4})\right]_0^1 = 1$$

$$c\tfrac{1}{12} =$$

and hence *c* = 12.

2.

$$E(X) = \int_0^1 12(x.^3 - x^4)dx$$

$$= \left[12(\frac{x^4}{4} - \frac{x^5}{5})\right]_0^1$$

$$= \frac{3}{5}$$

3.

$$E[X^2] = \int_0^1 12(x^4 - x^5)dx$$

$$= \left[12(\frac{x^5}{5} - \frac{x^6}{6})\right]_0^1$$

$$= \frac{12}{30}$$

$$= \frac{2}{3}$$

Thus $$Var[X] = \frac{2}{5} - \left(\frac{3}{5}\right)^2 = \frac{1}{25}$$

4.

$$\int_0^\infty 12(x^2 - x^3)dx = 0.5$$

$$\left[12(\frac{x^3}{3} - \frac{x^4}{4})\right]_0^m = 0.5$$

$$\left[12(\frac{m^3}{3} - \frac{m^4}{4})\right] = 0.5$$

$$4m^3 - 3m^4 = 0.5$$

$$6m^4 - 8m^3 + 1 = 0$$

3.2 The exponential distribution

The exponential distribution can be used to model the lifetimes of components. It is also linked to the Poisson distribution. If *X* has a Poisson distribution then the time between occurrences of *X* follows an exponential distribution.

The probability density function for an exponential distribution is

$$f(x) = \begin{cases} \lambda \exp(-\lambda x) , & if \;\; x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

We shall check first that this is a valid p d f. Clearly $f(x) \geq 0$. Also

$$\int_0^\infty \lambda \exp[-\lambda x]dx = \left[\frac{-\lambda}{\lambda}\exp[-\lambda x]\right]_0^\infty = 1$$

To find the mean we use integration by parts

$$E[X] = \int_0^\infty x\lambda \exp(-\lambda x)dx$$

$$= [-\exp[-\lambda x]x]_0^\infty + \int_0^\infty \exp(-\lambda x)dx$$

$$= \left[\frac{-1}{\lambda}\exp[-\lambda x\right]_0^\infty$$

$$= 1/\lambda$$

To find the variance we first find E[X^2]. This is also done by integration by parts

$$E[X^2] = \int_0^\infty x^2 \lambda \exp(-\lambda x) dx$$

$$= [-\exp[-\lambda x] x^2]_0^\infty + \int_0^\infty 2x \exp(-\lambda x) dx$$

$$= \left[\frac{-1}{\lambda} E(X) \right]$$

$$= \frac{2}{\lambda} \left(\frac{1}{\lambda} \right)$$

$$= \frac{2}{\lambda^2}$$

Therefore

$$Var[X] = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$$

The cumulative distribution function is given by

$$F(x) = P(X \le x) = \begin{cases} 0 & if \;\; x > 0 \\ \int_0^x f(t) dt & if \;\; x \ge 0 \end{cases}$$

Now

$$\int_0^\infty f(t) dt = \int_0^\infty \lambda \exp(-\lambda t) dt$$
$$= [-\exp[-\lambda t]_0^\infty$$
$$= 1 - \exp[-\lambda x]$$

The median m is given by F(m) = ½. Therefore substituting into the cdf

$$1 - \exp[-.\lambda m] = 1/2$$
$$\exp[-\lambda m] = 1/2$$
$$-\lambda m = \ln 1/2$$
$$\lambda m = \ln 2$$
$$m = \lambda^{-1} \ln 2.$$

3.3 The normal distribution

The normal, or Gaussian, distribution is the most commonly used distribution in statistics. There are goad theoretical reasons for using it in many cases but it is sometimes used because it has nice properties rather than because the analyst really believes the data are normally distributed.

The normal distribution is defined in terms of its mean and variance (or equivalently its standard deviation. If *X* is normally distributed with mean 0 and variance 1 then we write $X \sim N(0,1)$. Its probability density function is usually written as $\phi(x)$ and is given by

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \exp(-x^2/2) \quad \text{for} \quad -\infty < x < \infty$$

The cumulative distribution function is written as $\Phi(x)$

Table 3 gives the probability that a standard normal random variable, i.e. with mean zero and variance 1, is larger than specified value. That is it gives 1-$\Phi(x)$**.** In using the tables we utilise the symmetry of the normal distribution, and the fact that $P(Z < 0) = P(Z > 0) = 0.5$

For example suppose we are asked for the probabilities of the following events,
(i) Z < -2.45, (ii) (Z < 21)$\cup$(Z > 2.1) (iii) 0 < Z <1.2

(i) By symmetry P (Z < -2.45) = P (Z >2.45) = 0.00714
(ii) By symmetry P [(Z < -2.1) $\cup$ P (Z > 2.1)] = 2 P (Z > 2.1) = 2 x 0.01786 = 0.03572
(iii)

P [Z > 1.2] = 0.11507
P [Z < 1.2] = 1 – 0.11507
= 0.88493
P[0 < Z < 1.2] = 0.88493 – 0.5
= 0.38493

Note it often helps to draw a quick sketch of the area you want and to check if the answer seems reasonable in the light of this sketch.
Sometimes we want more than the two figures given in the table 3, for example Suppose we wanted P[Z > 1.453]. From the tables P(Z > 1.45) = 0.07353 and P(Z > 1.46) = 0.07215, to get the third figure we may linearly interpolate between these two probabilities. Thus we obtain

0.07353 – 0.3 x (0.07353 – 0.07215) = 0.07353 – 0.00041 = 0.07312

A general normally distributed random variable with mean μ and variance σ^2 has density

$$\phi(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp[-(x-\mu)^2 / 2\sigma^2 \quad \text{for} \quad -\infty < x < \infty$$

It is written $X \sim N(\mu, \sigma^2)$ fortunately we can calculate probabilities for a general normal distribution from the standard normal. If $X \sim N(\mu, \sigma^2)$ then

$$\frac{X - \mu}{\sigma} \sim N(0,1)$$

Example: It is known that in a certain district the heights of adult males are normally distributed with mean 175cm and standard deviation 7cm. Find the probability that a man selected at random from this district will be

(a) over 182cm tall

(b) between 170cm and 181cm tall

(c) under 179cm tall.

Let X be the height of the selected man. Then $X \sim N(175, 7^2)$ and Z = (X-175)/7 ~ N (0,1) Thus for

(a) P(X > 182) = P (Z > (182 – 175)/7) = P (Z > 1) = 0.159

(b) P(170 < X < 181) = P (-5/7 < Z < 6/7)
= P (Z >-5/7) – P (Z > 6/7)
= 0.7625 – 0.1968 = 0.566

(c) P (X< 179) = P (Z < 4/7) = 1 – P(Z > 4/7) $\approx$ 1 – 0.284 = 0.716

3.4 The normal as an approximating distribution

When *n* is large and p moderate we may use the normal distribution to approximate binomial probabilities. Note that as we are approximating a discrete random variable by a continuous one we have to employ the so-called continuity correction. For discrete random variable P(X < x) = P (X $\leq$ x-1) We approximate these quantities by P (Y < x - $\frac{1}{2}$) . We illustrate the technique in the following example.

Example A fair coin is tossed 150 times. Find a suitable approximation to the probabilities of the following events

(a) more than 70 heads

(b) fewer than 82 heads

(c) more than 72 but fewer than 79 heads.

Let X be the number of heads thrown, then X has a binomial distribution with n = 150 and p = ½ . As n is larger and p moderate we may approximate X by Y a normal random variable with mean np = 75 and variance np(1-p) = 37.5.

(a) We require (P X > 71) but this is the same as P(X $\geq$ 70) so we approximate by P (Y > 70.5). This is equal to
$P(Z > (70.5 - 75)/\sqrt{37.5}) \approx P(Z \geq -0.735) = 0.769$

(b) We require P(X < 82) but this is the same as P(X $\leq$81) so we approximate by P(Y < 81.5). This is equal to
P(Z < (81.5 – 75)/$\sqrt{37.5}$)$\approx$ P(Z < 1.06) = 1- 0.145 = 0.855

(c) We require P (72 < X < 79) which is the same as P (73 $\leq X \leq$78) and thus we approximate by (72.5 < y < 78.5). This approximately equal to
P(-0.408 < Z < 0.571) = 0.658 – 0.284 = 0.374

We may similarly approximate a Poisson random variable by a normal one of the same mean and variance so long as this mean is moderately large. We again have to use the continuity correction.

Example: A radioactive source emits particles at random at an average rate of 36 per hour. Find an approximation to the probability that more than 40 particles are emitted in one hour.

Let X be the number of particles emitted in one hour. Then X has a Poisson distribution with mean 36 and variance 36. We can approximate X by Y which has a N(36, 36) distribution. We require *P(X* > 40). This is approximately *P(Y* ≥ 40.5) or transforming to a standard normal random variable by subtracting the mean and dividing by the standard deviation, we have

$$P(Y \geq 40.5) = P(Z \geq \frac{40.5-36}{6})$$
$$= P(Z \geq 0.75)$$
$$= 0.2266$$

3.5 Exercises

1. The random variable X has probability density function $f(x) = c(a-x)^3$ **—**for 0 < z < a and zero otherwise. Determine ***c*.** Find the cumulative distribution function ***F(x)*.**

2. Assume that the continuous random variable ***Z*** has the probability density function

$$f(x) = \begin{cases} k(9-4x^2) & \text{for } 0 \leq x \leq 3/2 \\ 0 & \text{otherwise} \end{cases}$$

   (a) Calculate the value of ***k*.**
   (b) Find the mean and variance of ***Z*.**
   (c) Find the cumulative distribution function of ***Z*.**
   (d) Find the median of ***Z*.**
   (e) Find P(1/2 ≤ ***Z*** < 1).

3. The random variable X has probability density function given by

$$f(x) = \begin{cases} k\exp[-\lambda(x-\mu)] & \text{for } x \geq \mu \\ 0 & \text{otherwise} \end{cases}$$

   (a) Calculate the value of ***k*.**

   (b) Find the mean and variance of *X*.

   (c) Find the cumulative distribution function of *X*.

   (d) Find the median of *X*.

4. The lengths of a batch of bolts are assumed normally distributed with mean 4cm and standard deviation 0.1cm. What is the probability that a

bolt selected at random will be more than 4.1655cm in length? (Give answer to 5 dp)

5. Use a suitable approximation to determine the probability that there are between 75 and 82 heads inclusive when a fair coin is tossed 160 times.

6. A coin is to be tossed 10 times.

   (a) Assuming the coin is biased with P(head) =0.6, use a normal approximation to estimate the probability that between 56 and 63 heads occur.

   (b) Assume *P(head)=0.99.* Use a suitable approximation to estimate the probability that exactly 99 heads occur. (Do not calculate the exact binomial probability).