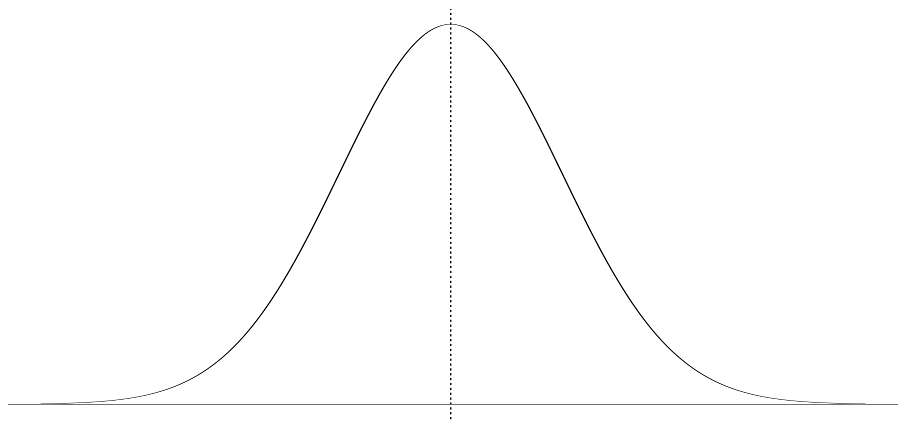
Chapter 10

More Continuous Probability Models

10.1 Introduction

Over the past few weeks we have discussed some "standard" probability distributions which can be used to model data. We have looked at two such distributions for *discrete* data – the binomial distribution and the Poisson distribution – and last week the Normal distribution was introduced as a probability model for *continuous* data.

Recall the *probability density function* of the Normal distribution, which is often referred to as a "bell–shaped curve":



We saw in the lecture last week that many naturally occurring continuous measurements (such as height, weight, time, rainfall etc.) often resemble this bell–shaped curve when plotted using a histogram, for example. But what if we cannot assume "Normality" for our data?

We now consider two other probability models which can be used to model continuous data when the Normal distribution isn't appropriate.

10.2 The Uniform Distribution

The *uniform distribution* is the most simple continuous distribution. As the name suggests, it describes a variable for which all possible outcomes are equally likely. For example, suppose you manage a group of Environmental Health Officers and need to decide at what time they should inspect a local hotel. You decide that any time during the working day (9.00 to 18.00) is okay but you want to decide the time "randomly". Here "randomly" is a short–hand for "a random time, where all times in the working day are equally likely to be chosen". Let X be the time to their arrival at the hotel measured in terms of minutes from the start of the day. Then X is a uniform random variable between 0 and 540:

As with the Normal distribution, the total area (base $\times$ height) under pdf must equal one. Therefore, as the base is 540, the height must be $1/540$. Hence the probability density function (pdf) for the continuous random variable X is

$$f(x) = \begin{cases} \dfrac{1}{540} & \text{for } 0 \leq x \leq 540 \\ 0 & \text{otherwise.} \end{cases}$$

In general, we say that a random variable X which is equally likely to take any value between a and b has a uniform distribution on the interval a to b, i.e.

$$X \sim U(a, b).$$

The random variable has probability density function (pdf)

$$f(x) = \begin{cases} \dfrac{1}{b-a} & \text{for } a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}$$

and probabilities can be calculated using the formula

$$P(X \leq x) = \begin{cases} 0 & \text{for } x < a \\ \dfrac{x-a}{b-a} & \text{for } a \leq x \leq b \\ 1 & \text{for } x > b. \end{cases}$$

We use this formula directly because it is fairly simple, unlike the Normal distribution for which we used tables of probabilities. Therefore, for example, the probability that the inspectors visit the hotel in the morning (within 180 minutes after 9am) is

$$P(X \leq 180) = \frac{180-0}{540-0} = \frac{1}{3}.$$

The probability of a visit during the lunch hour (12.30 to 13.30) is

$$\begin{aligned} P(210 \leq X \leq 270) &= P(X \leq 270) - P(X < 210) \\ &= \frac{270-0}{540-0} - \frac{210-0}{540-0} \\ &= \frac{270-210}{540} \\ &= \frac{60}{540} \\ &= \frac{1}{9}. \end{aligned}$$

10.2.1 Mean and Variance

The mean and variance of a continuous random variable can be calculated in a similar manner to that used for a discrete random variable. However, the specific techniques required to do this are outside the scope of this course and so we will simply state the results.

If X is a uniform random variable on the interval a to b then its mean and variance are

$$E(X) = \mu = \frac{a+b}{2}, \qquad Var(X) = \sigma^2 = \frac{(b-a)^2}{12}.$$

In the above example, we have

$$E(X) = \frac{a+b}{2} = \frac{0+540}{2} = 270,$$

so that the mean arrival of the inspectors is 9am $+$ 270 minutes $=$ 13.30. Also

$$Var(X) = \frac{(540-0)^2}{12} = 24300$$

and therefore $SD(X) = \sqrt{Var(X)} = \sqrt{24300} = 155.9$ minutes.

10.3 The Exponential Distribution

The *exponential distribution* is another common distribution that is used to describe continuous random variables. It is often used to model lifetimes of products and times between "random" events such as arrivals of customers in a queueing system or arrivals of orders. The distribution has one parameter, λ. If our random variable X follows an exponential distribution, then we say

$$X \sim \exp(\lambda).$$

Its probability density function is

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{for } x \geq 0, \\ 0 & \text{otherwise} \end{cases}$$

and probabilities can be calculated using

$$P(X \leq x) = \begin{cases} 0 & \text{for } x < 0 \\ 1 - e^{-\lambda x} & \text{for } x > 0. \end{cases}$$

The main features of this distribution are:

1. an exponentially distributed random variable can only take positive values
2. larger values are increasingly unlikely – "exponential decay"
3. the value of λ fixes the rate of decay – larger values correspond to more rapid decay.

Consider an example in which the time (in minutes) between successive users of a pay phone can be modelled by an exponential distribution with $\lambda = 0.3$. The probability of the gap between phone users being less than 5 minutes is

$$P(X < 5) = 1 - e^{-0.3\times 5} = 1 - 0.223 = 0.777.$$

Also the probability that the gap is more than 10 minutes is

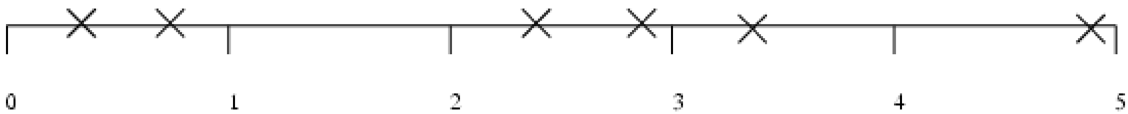
$$P(X > 10) = 1 - P(X \leq 10) = 1 - \left(1 - e^{-0.3\times 10}\right) = e^{-0.3\times 10} = 0.050$$

and the probability that the gap is between 5 and 10 minutes is

$$P(5 < X < 10) = P(X < 10) - P(X \leq 5) = 0.950 - 0.777 = 0.173.$$

One of the main uses of the exponential distribution is as a model for the times between events occurring randomly in time. We have previously considered events which occur at random points in time in connection with the Poisson distribution. The Poisson distribution describes probabilities for the number of events taking place in a given time period. The exponential distribution describes probabilities for the times between events. Both of these concern events occurring randomly in time (at a constant average rate, say λ). This is known as a *Poisson process*.

Consider a series of randomly occurring events such as calls at a credit card call centre. The times of calls might look like



There are two ways of viewing these data. One is as the number of calls in each minute (here 2, 0, 2, 1 and 1) and the other as the times between successive calls. For the Poisson process,

- the number of calls has a Poisson distribution with parameter λ, and
- the time between successive calls has an exponential distribution with parameter λ.

10.3.1 Mean and Variance

The mean and variance of the exponential distribution can be shown to be

$$E(X) = \frac{1}{\lambda}, \qquad Var(X) = \frac{1}{\lambda^2}.$$

10.4 Exercises

1. An express coach is due to arrive in Newcastle from London at 23.00. However, in practice, it is equally likely to arrive anywhere between 15 minutes early to 45 minutes late, depending on traffic conditions. Let the random variable X denote the amount of time (in minutes) that the coach is delayed.

   (a) Sketch the pdf.

   (b) Calculate the mean and standard deviation of the delay time.

   (c) What is the probability that the coach is less than 5 minutes late?

   (d) What is the probability that the coach is more than 20 minutes late?

   (e) What is the probability that the coach arrives between 22.55 and 23.20?

   (f) What is the probability that the coach arrives at 23.00?

   (g) What is the probability that the coach arrives at 0.00?

   (h) Do you think that this is a good model for the coach's arrival time?

2. A network server receives incoming requests according to a Poisson process with rate $\lambda = 2.5$ per minute.

   (a) What is the expectation of the time between arrivals of requests?

   (b) What is the probability that the time between requests is less than 2 minutes?

   (c) What is the probability that the time between requests is greater than 1 minute?

   (d) What is the probability that the time between requests is between 30 seconds and 50 seconds?

3. As Production Manager, you are responsible for buying a new piece of equipment for your company's production process. A salesman from one company has told you that he can supply you with equipment for which the time to first breakdown (in months) follows an exponential distribution with $\lambda = 0.11$. Another salesman (from another company) has told you that the time to first breakdown of their machines is also exponentially distributed but with $\lambda = 0.01$. It is very important that the equipment you purchase does not break down for at least six months. Calculate the probability of this outcome for both suppliers and make a recommendation to the company board about which machine should be bought.

   How might you take into account a difference between the prices for the machines?

4* Calls made to a Company's complaints hotline arrive according to a *Poisson process* with rate two per minute.

   (a) Write down the distribution of X, the number of calls to the complaints hotline in any five minute period. Find $E[X]$, and s.d.(X).

   (b) Write down the distribution of Y, the time between successive calls to the complaints hotline in any five minute period. Find $E[Y]$ and s.d.(Y).

   (c) Use the distribution in part (a) to find the probability that there will be fewer than four calls made to the complaints hotline in any five minute period.

   (d) Use the distribution in part (b) to find the probability that the time between any two successive calls to the complaints line is greater than ten seconds.

   (e) This company has twenty branches throughout Northeast England. Each week, all employees at a particular branch receive a bonus if, on average, there have been less than four calls per five minutes to the complaints hotline. Find the probability that employees at fewer than two of the twenty branches receive a bonus.

   (f) This company manufactures notebook PCs. Weekly sales of their products via their website tend to be Normally distributed with mean $\mu = £7550$ and standard deviation $\sigma = £255$. Find the probability that weekly sales exceed $\mu + 2\sigma$.

* Prize question – to be submitted by no later than the 11 o'clock tutorial this Wednesday!