7 Continuous Variables

7.1 Distribution function

With continuous variables we can again define a probability distribution but instead of specifying $\Pr(X = j)$ we specify $\Pr(X < u)$ since $\Pr(u < X < u + \delta) \to 0$ as $\delta \to 0$. For example, if X is the amount of oil (in barrels) which will be extracted from a particular well then we can find $\Pr(X < 1000000)$ or $\Pr(X < 500000)$ but it makes no sense to talk about $\Pr(X = 750000)$ because we could measure the amount more precisely and find that it was not *exactly* 750000 barrels.

We can plot a graph of $\Pr(X < u)$ against u. This graph represents a *function* of u. This function is called the *distribution function* of X. We might write it as $F(u)$. Hence $\Pr(X < u) = F(u)$. Because this is a probability, it is always between 0 and 1. That is $0 \leq F(u) \leq 1$. Also, because, for example, the probability that $X < 600000$ must be at least as big as the probability that $X < 500000$, we can say that if $w > u$ then $F(w) \geq F(u)$.

7.2 Probability density function

Another important way of representing a continuous probability distribution is the *probability density function* or *pdf.* This is actually the gradient of the distribution function. That is the pdf is

$$f(u) = \frac{dF(u)}{du}.$$

If we draw a graph of the probability density function then probabilities are represented by *areas under the curve* on the graph. For example, the probability that $500000 < X < 600000$ is the area above the axis and below the curve between the limits at 500000 and 600000.

Note that

$$\begin{aligned} \int_{-\infty}^{w} f(u).du &= F(w), \\ \int_{-\infty}^{\infty} f(u).du &= 1. \end{aligned}$$

7.3 Mode and median

The *mode* of a continuous probability distribution is the point at which the probability density function attains its maximum value. The *median* of a continuous probability distribution is the point at which the distribution function has the value 0.5.

7.4 Example

A continuous random variable X has the following distribution function.

$$F_X(u) = \begin{cases} 0 & (u < 0) \\ \frac{3}{4}(u^2 - u^3/3) & (0 \leq u \leq 2) \\ 1 & (u > 2) \end{cases}$$

The median is 1.0.
The probability density function is as follows.

$$f_X(u) = \begin{cases} 0 & (u < 0) \\ \frac{3}{4}(2u - u^2) & (0 \leq u \leq 2) \\ 0 & (u > 2) \end{cases}$$

The mode is 1.0.
The probability that $0.5 < X < 1.5$ can be found as follows.

$$\begin{aligned} \Pr(0.5 < X < 1.5) &= F_X(1.5) - F_X(0.5) \\ &= \frac{3}{4}\left(1.5^2 - \frac{1.5^3}{3} - 0.5^2 + \frac{0.5^3}{3}\right) \\ &= \frac{3}{4}(1.125 - 0.208) \\ &= 0.687. \end{aligned}$$

Alternatively

$$\Pr(0.5 < X < 1.5) = \int_{0.5}^{1.5} f_X(u).du.$$

7.5 Expectation

The results concerning expectation etc. for continuous random variables are similar to those for discrete random variables with the summations replaced with integrals.

Let X be a continuous random variable with pdf $f_X(u)$. Then

$$\mathrm{E}(X) = \int_{-\infty}^{\infty} u f_X(u).du.$$

For example, in the example above,

$$\begin{aligned} \mathrm{E}(X) &= \int_{-\infty}^{\infty} u f_X(u).du \\ &= \int_0^2 \frac{3}{4}(2u^2 - u^3).du \\ &= \frac{3}{4}\left[\frac{2u^3}{3} - \frac{u^4}{4}\right]_0^2 \\ &= 1. \end{aligned}$$

Let $g(X)$ be a function of X. Then its expectation is

$$\mathrm{E}[g(X)] = \int_{-\infty}^{\infty} g(u) f_X(u).du.$$

For example, in the example above,

$$\begin{aligned}
\mathrm{E}(X^2) &= \int_{-\infty}^{\infty} u^2 f_X(u).du \\
&= \int_0^2 \frac{3}{4}(2u^3 - u^4).du \\
&= \frac{3}{4}\left[\frac{2u^4}{4} - \frac{u^5}{5}\right]_0^2 \\
&= \frac{6}{5}.
\end{aligned}$$

As for discrete random variables, for two continuous random variables X and Y, we have

$$\mathrm{E}(aX + bY) = a\mathrm{E}(X) + b\mathrm{E}(Y).$$

7.6 Variance and covariance

The variance of any random variable is given (provided it exists) by

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

In the example

$$\mathrm{var}(X) = \frac{6}{5} - 1^2 = \frac{1}{5}.$$

As for discrete random variables, for two continuous random variables X and Y, we have

$$\begin{aligned}
\mathrm{covar}(X, Y) &= \mathrm{E}[\{X - \mathrm{E}(X)\}\{Y - \mathrm{E}(Y)\}] \\
&= \mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y) \\
\mathrm{var}(aX + bY) &= a^2\mathrm{var}(X) + b^2\mathrm{var}(Y) + 2ab\mathrm{covar}(X, Y)
\end{aligned}$$

7.7 Independence

If $\Pr(X < u|Y < v) = \Pr(X < u)$ for all u and v then X and Y are independent. If X and Y are independent then $E(XY) = E(X)E(Y)$.

Note that

$$\mathrm{E}(XY) = \mathrm{E}(X)\mathrm{E}(Y) \Leftrightarrow \mathrm{covar}(X, Y) = 0 \Leftrightarrow \mathrm{var}(aX + bY) = a^2\mathrm{var}(X) + b^2\mathrm{var}(Y).$$

7.8 The continuous uniform distribution

Let a and b be two real numbers such that $a < b$. If a random variable X has the probability density function

$$f_X(u) = \begin{cases} (b-a)^{-1} & a \le u \le b \\ 0 & \text{otherwise} \end{cases},$$

we say that X has a continuous uniform distribution on the interval $[a, b]$. That is $X \sim U(a, b)$. The distribution function is

$$F_X(u) = \begin{cases} 0 & u < a \\ (u-a)/(b-a) & a \le u \le b \\ 1 & u > b \end{cases}.$$

The mean of X is obviously $(b-a)/2$ but we can find this formally as follows.

$$\begin{aligned} \mathrm{E}(X) &= (b-a)^{-1}\int_a^b u.du \\ &= (b-a)^{-1}\left\{\frac{1}{2}(b^2-a^2)\right\} \\ &= \frac{1}{2}\frac{(b-a)(b+a)}{(b-a)} \\ &= \frac{1}{2}(b+a) \end{aligned}$$

Similarly we can find the variance. First recall that $\mathrm{var}(X) = \mathrm{E}\left(\{X - \mathrm{E}[X]\}^2\right)$.

$$\begin{aligned} \mathrm{var}(X) &= (b-a)^{-1}\int_a^b \left\{u - \frac{1}{2}(b+a)\right\}^2 .du \\ &= (b-a)^{-1}\int_{-(b-a)/2}^{(b-a)/2} v^2.dv \\ &= \frac{(b-a)^2}{12} \end{aligned}$$

where $v = u - (b+a)/2$.

Example: A car ferry over a short crossing leaves from one side every 20 minutes. A car arrives at a random point in time to use the ferry. The distribution of waiting time in minutes is uniform$(0, 20)$. The mean is 10 minutes, the variance is 33.333 and the standard deviation is 5.77 minutes. The probability that the waiting time is more than 5 minutes is 0.25 etc.

7.9 The negative exponential distribution

Let λ be a positive real number. If a random variable T has the probability density function

$$f_T(t) = \begin{cases} 0 & (t < 0) \\ \lambda e^{-\lambda t} & (0 \leq t < \infty) \end{cases}$$

we say that T has a **negative exponential distribution** (sometimes called simply an "exponential distribution.")

The distribution function is

$$F_T(t) = \begin{cases} 0 & (t < 0) \\ 1 - e^{-\lambda t} & (0 \leq t < \infty) \end{cases} .$$

In finding the mean and variance it helps first to evaluate $\mathrm{E}(T^n)$ for $n \geq 1$.

$$\begin{aligned} \mathrm{E}(T^n) &= \int_{-\infty}^{\infty} t^n f_T(t).dt \\ &= \int_0^{\infty} t^n \lambda e^{-\lambda t}.dt \\ &= \left[-t^n e^{-\lambda t}\right]_0^{\infty} + \int_0^{\infty} n t^{n-1} e^{-\lambda t}.dt \\ &= \frac{n}{\lambda}\mathrm{E}(T^{n-1}) \end{aligned}$$

Now $\mathrm{E}(T^0) = \mathrm{E}(1) = 1$ so

$$\mathrm{E}(T) = \frac{1}{\lambda}$$

and $\mathrm{E}(T^2) = 2/\lambda^2$ so

$$\begin{aligned} \mathrm{var}(T) &= \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 \\ &= \frac{1}{\lambda^2} \end{aligned}$$

7.10 Relationship of the negative exponential distribution to the Poisson distribution

Consider a series of "events" at points in time which are random and independent in such a way that they are equally as likely to occur at any time as at any other time and the occurrence of one event does not affect the probability of the occurrence of an event at any other time. Examples of this sort of thing might be the emissions of particles from a radioactive source (with a long half-life) the occurrence of accidents or insurance claims of certain types etc. The number of "events" in a time interval of a given length has a Poisson distribution. Suppose that the number X of "events" occurring in any time interval of duration t time-units (e.g. t minutes) has a Poisson distribution with mean λt. The average rate of occurrence of "events" is therefore λ per unit time and

$$\Pr(X = i) = \frac{e^{-\lambda t}(\lambda t)^i}{i!}$$

$(i = 0, 1, 2, \ldots)$.

Consider the time up to the first occurrence of one of these events. Let the time elapsed, from some arbitrary starting point in time, up to the first event be T units. (Thus T is a random variable). Now $T > t$ if and only if there are no events in the first t time units $(t \geq 0)$. Therefore $\Pr(T > t) = \Pr(X = 0)$ and $\Pr(T \leq t) = 1 - \Pr(X = 0) = 1 - e^{-\lambda t}$, the distribution function of a negative exponential random variable $(t \geq 0)$. The p.d.f. may be obtained by differentiation. Hence the waiting time T has a negative exponential distribution.

7.11 Extreme value distribution

Subject to certain conditions, the distribution of the maximum (i.e. largest value) in random samples of size m from a continuous distribution tends to a *type 1 extreme vale distribution* (EV1) as m tends to infinity. (The conditions are that the distribution has an unbounded upper tail which decays at least as fast as that of the exponential distribution). For example, the annual maximum sea level is the maximum of about 700 high water levels. Thes are *not* independent and this means that m is effectively less than 700. Also the upper tail of the sea-level distribution is not strictly unbounded. Nevertheless the EV1 distribution often provides a good fit to yearly maxima. The EV1 distribution has distribution function

$$F(x) = \exp\{-e^{-(x-\xi)/\theta}\}$$

for $-\infty < x < \infty$.

The probability density function is

$$f(x) = \theta^{-1}\exp\{-(x-\xi)/\theta - e^{-(x-\xi)/\theta}\}.$$

The parameter ξ is the mode of the distribution and θ, which must be positive, is a scale factor proportional to the standard deviation.

median	$\xi - \theta\ln\{\ln(2)\} = \xi - 0.36651\theta$
mean	$\xi + \gamma\theta = \xi + 0.57722\theta$
	where $\gamma = 0.57722$ is Euler's constant
variance	$\pi^2\theta^2/6 = 1.6449\theta^2$
standard deviation	$\pi\theta/\sqrt{6} = 1.28255\theta$

In the case where $\xi = 0$ and $\theta = 1$ the EV1 distribution is known as a Gumbel distribution.

We can fit an EV1 distribution to data by comparing the sample mean and sample standard deviation of the data to the formulae above. So, for example, if the sample mean is $\bar{x} = 16.3$ and the sample standard deviation is $s = 5.7$ then we would estimate θ using $\hat{\theta} = s/1.28255 = 5.7/1.28255 = 4.444$ and estimate ξ using $\hat{\xi} = \bar{x} - 0.57722\hat{\theta} = 16.3 - 0.57722 \times 4.444 = 13.73$.

7.12 Rayleigh distribution

The Rayleigh distribution is often used to model wave heights and wind speeds.

The distribution function is

$$F(x) = 1 - \exp\left\{-\frac{x^2}{2\theta^2}\right\}$$

for $0 < x < \infty$.

The probability density function is

$$f(x) = \frac{xe^{-x^2/2\theta^2}}{\theta^2}.$$

mean	$\theta\sqrt{\pi/2} = 1.2533\theta$
variance	$(4-\pi)\theta^2/2 = 0.429201\theta^2$
standard deviation	$\theta\sqrt{\{(4-\pi)/2\}} = 0.655136\theta$

Note that $\mathrm{E}(X^2) = \mathrm{var}(X) + [\mathrm{E}(X)]^2 = 2\theta^2$.

There is only one parameter, θ. We can estimate it using the sample mean so that $\hat{\theta} = \bar{x}/\sqrt{\pi/2} = \bar{x}/1.2533$.

7.13 Problems

1. The number of failures occurring in a machine of a certain type in a year has a Poisson distribution with mean 0.4. In a factory there are ten of these machines. What is

   (a) the expected total number of failures in the factory in a year?

   (b) the probability that there are fewer than two failures in the factory in a year?

2. A man goes fishing. The number of fish he catches in one hour has a Poisson distribution with mean 1.25. He continues fishing for four hours, then goes home, unless he catches five fish before the four hours are up, in which case he goes home as soon as he has caught the fifth fish.

   Find

   (a) the probability that he goes home with three fish.

   (b) the expected number of fish he takes home.

3. There are five machines in a factory. Of these machines, three are working properly and two are defective. Machines which are working properly produce articles each of which has independently a probability of 0.1 of being imperfect. For the defective machines this probability is 0.2.

   A machine is chosen at random and five articles produced by the machine are examined. What is the probability that the machine chosen is defective given that, of the five articles examined, two are imperfect and three are perfect?

4. A continuous random variable T has the following probability density function.

$$f_T(u) = \begin{cases} 0 & (u < 0) \\ 3(1 - u/k) & (0 \le u \le k) \\ 0 & (u > k) \end{cases} .$$

Find

(a) k.

(b) $\mathrm{E}(T)$.

(c) $\mathrm{E}(T^2)$.

(d) $\mathrm{var}(T)$.

5. A continuous random variable X has the following probability density function

$$f_X(u) = \begin{array}{cc} 0 & (u < 0) \\ ku & (0 \le u \le 1) \\ 0 & (u > 1) \end{array}$$

(a) Find k.

(b) Find the distribution function $F_X(u)$.

(c) Find $\mathrm{E}(X)$.

(d) Find $\mathrm{var}(X)$.

(e) Find $\mathrm{E}(e^X)$.

(f) Find $\mathrm{var}(e^X)$.

(g) Find the distribution function of e^X. (Hint: For what values of X is $e^X < u$?)

(h) Find the probability density function of e^X.

(i) Sketch $f_X(u)$.

(j) Sketch $F_X(u)$.