9 — CONTINUOUS DISTRIBUTIONS

A random variable whose value may fall *anywhere* in a range of values is a continuous random variable and will be associated with some continuous distribution. Continuous distributions are to discrete distributions as type `real` is to type `int` in ML.

Many formulae for discrete distributions can be adapted for continuous distributions. Very often, little more is required than the translation of sigma signs into integral signs. The main bad news is that there is no equivalent of probability generating functions.

Adapting the $\mathrm{P}(X = r)$ Notation

In discussions which involve a single discrete random variable, the notation $\mathrm{P}(X = r)$ has been used. When required, mapping is employed to ensure that $r \in \mathbb{N}$.

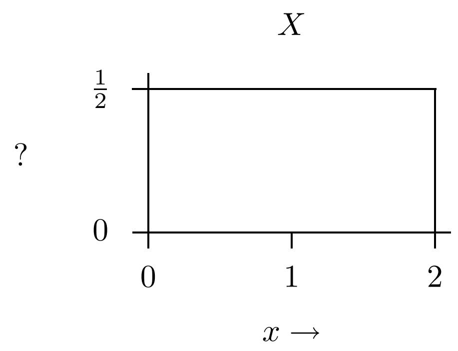
In discussing a single continuous random variable, X will again be used as the name but x will be used instead of r for the value. In probability theory r strongly implies a non-negative integer whereas $x \in \mathbb{R}$ and may range from $-\infty$ to $+\infty$.

There is at once a problem with the notation $\mathrm{P}(X = x)$ for the probability is zero for any particular x. Even if x is constrained to be in some finite range, such as -1 to $+1$, there are an infinite number of possible values for x.

Fortunately, many variants of the $\mathrm{P}(X = x)$ notation are still useful. For example:

$$\mathrm{P}(X < 0.5) \qquad \mathrm{P}(-1 \leqslant X < +1) \qquad \mathrm{P}(a \leqslant X < b)$$

There is an obvious difficulty with a graphical representation of a continuous random variable. A plot of $\mathrm{P}(X = x)$ against x serves no useful purpose! Nevertheless, graphical representations are both possible and useful and here is a first attempt at representing a continuous random variable X which is distributed Uniform(0,2):



It is not immediately clear what label should be attached to the vertical axis but this representation has the right feel about it. The height of the plot is constant over the range 0 to 2 and is zero outside this range.

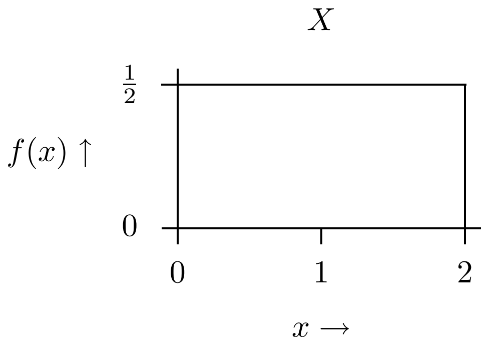
The constant height is $\frac{1}{2}$ to ensure that the total area under the curve is 1 and this is the clue to much of what follows. The idea of area corresponding to probability was introduced on page 1.6 and with continuous random variables area is often the most convenient way of representing probability.

Probability Density Functions

In the present case, the area under the curve between $x = 1$ and $x = 1\frac{1}{4}$ is $(1\frac{1}{4} - 1) \times \frac{1}{2} = \frac{1}{8}$ so the probability $\mathrm{P}(1 \leqslant X < 1\frac{1}{4}) = \frac{1}{8}$.

In general, this calculation will be an integration and some consideration needs to be given to the function to be integrated.

The function is called a *probability density function* or pdf. In the case of a single random variable it is often named $f(x)$ and this is the appropriate label for the vertical axis:



In the case of the random variable X which is distributed Uniform(0,2):

$$f(x) = \begin{cases} \frac{1}{2}, & \text{if } 0 \leqslant x < 2 \\ 0, & \text{otherwise} \end{cases}$$

The calculation just undertaken is more formally written:

$$\mathrm{P}(1 \leqslant X < 1\tfrac{1}{4}) = \int_1^{1\frac{1}{4}} f(x)\,dx = \int_1^{1\frac{1}{4}} \frac{1}{2}\,dx = \left[\frac{x}{2}\right]_1^{\frac{5}{4}} = \frac{5}{8} - \frac{1}{2} = \frac{1}{8}$$

In general:

$$\mathrm{P}(a \leqslant X < b) = \int_a^b f(x)\,dx$$

Both common sense and the axioms of probability impose certain constraints that have to be met by any probability density function:

I $f(x)$ must be single valued for all x

II $f(x) \geqslant 0$ for all x

III $\displaystyle\int_{-\infty}^{+\infty} f(x)\,dx = 1$

This last is sometimes expressed as:

$$\int_R f(x)\,dx = 1$$

Here R refers to the range of interest, where the probability density function is non-zero. For the uniform distribution above, the range R is 0 to 2.

Continuous Distributions

Informally, a discrete distribution has been taken as almost any indexed set of probabilities whose sum is 1. The index has always been $r = 0, 1, 2, \ldots$

Equally informally, almost any function $f(x)$ which satisfies the three constraints can be used as a probability density function and will represent a continuous distribution.

Expectation

With discrete distributions, the general formula for the mean or expectation of a single random variable X is:

$$\mu = \mathrm{E}(X) = \sum_r r.\mathrm{P}(X = r)$$

This is the first example of a formula used with discrete distributions which can be readily adapted for continuous distributions. The mean μ or expectation $\mathrm{E}(X)$ of a random variable X whose probability distribution function is $f(x)$ is:

$$\mu = \mathrm{E}(X) = \int_R x.f(x)\,dx$$

The general form for the expectation of a function of a random variable adapts too but since f is used as the name of the general probability density function some other name has to be used for the function of the random variable. In the following, h is taken as some function of the random variable X and the expectation:

$$\mathrm{E}\big(h(X)\big) = \int_R h(x).f(x)\,dx$$

In the particular case of the square of X, when $h(X) = X^2$:

$$\mathrm{E}(X^2) = \int_R x^2.f(x)\,dx$$

Variance

The definition of variance is exactly the same for continuous random variables as for discrete random variables:

$$\text{Variance} = \sigma^2 = \mathrm{V}(X) = \mathrm{E}\big((X-\mu)^2\big) = \mathrm{E}(X^2) - \big(\mathrm{E}(X)\big)^2$$

Thus the variance can be determined by first evaluating $\mathrm{E}(X)$ and $\mathrm{E}(X^2)$. Alternatively, $(X - \mu)^2$ can be regarded as a special case of the function h and the variance can be directly computed thus:

$$\text{Variance} = \sigma^2 = \mathrm{V}(X) = \mathrm{E}\big((X-\mu)^2\big) = \int_R (x-\mu)^2.f(x)\,dx$$

Illustration — Uniform(0,2)

Consider the random variable X which is distributed Uniform(0,2) and whose probability density function is:

$$f(x) = \begin{cases} \frac{1}{2}, & \text{if } 0 \leqslant x < 2 \\ 0, & \text{otherwise} \end{cases}$$

The expectation $\mathrm{E}(X)$ is:

$$\mathrm{E}(X) = \int_0^2 x.\frac{1}{2}\,dx = \left[\frac{x^2}{4}\right]_0^2 = 1$$

The expectation $\mathrm{E}(X^2)$ is:

$$\mathrm{E}(X^2) = \int_0^2 x^2.\frac{1}{2}\,dx = \left[\frac{x^3}{6}\right]_0^2 = \frac{4}{3}$$

The variance $\mathrm{V}(X)$ is:

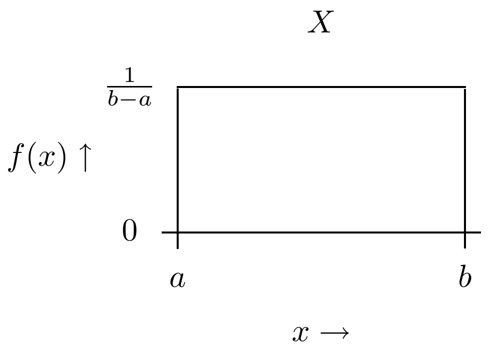
$$\mathrm{V}(X) = \mathrm{E}(X^2) - \big(\mathrm{E}(X)\big)^2 = \frac{4}{3} - 1^2 = \frac{1}{3}$$

The General Uniform Distribution

In the general case, a random variable which is distributed Uniform(a,b) is uniformly distributed over the range a to b. To ensure that the integral of the associated probability density function $f(x)$ over this range is 1 the function is defined as:

$$f(x) = \begin{cases} \frac{1}{b-a}, & \text{if } a \leqslant x < b \\ 0, & \text{otherwise} \end{cases}$$

This function can be represented graphically:



By analogy with discrete distributions, the first check is that the integration over the appropriate range is 1:

$$\int_a^b \frac{1}{b-a}\,dx = \left[\frac{x}{b-a}\right]_a^b = 1$$

The expectation $\mathrm{E}(X)$ is:

$$\mathrm{E}(X) = \int_a^b x.\frac{1}{b-a}\,dx = \frac{1}{b-a}\left[\frac{x^2}{2}\right]_a^b = \frac{1}{b-a}\cdot\frac{b^2-a^2}{2} = \frac{b+a}{2}$$

This is simply confirming that the mean is halfway between a and b and this was seen earlier with the distribution Uniform(0,2) where the mean was 1.

The expectation $\mathrm{E}(X^2)$ is:

$$\mathrm{E}(X^2) = \int_a^b x^2.\frac{1}{b-a}\,dx = \frac{1}{b-a}\left[\frac{x^3}{3}\right]_a^b = \frac{1}{b-a}\cdot\frac{b^3-a^3}{3} = \frac{b^2+ab+a^2}{3}$$

The variance $\mathrm{V}(X)$ is:

$$\mathrm{V}(X) = \mathrm{E}(X^2) - \big(\mathrm{E}(X)\big)^2 = \frac{b^2+ab+a^2}{3} - \frac{b^2+2ab+a^2}{4}$$

$$= \frac{4b^2+4ab+4a^2 \; - \; 3b^2-6ab-3a^2}{12} = \frac{b^2-2ab+a^2}{12} = \frac{(b-a)^2}{12}$$

With the distribution Uniform(0,2) $a=0$ and $b=2$ giving the result $\frac{1}{3}$ noted earlier.

Illustration — Roulette Wheel

Let X be the angle between some reference radius on a roulette wheel and some fixed direction on the casino table. The angle X is a random variable which is distributed Uniform$(0, 2\pi)$.

Taking the values $a=0$ and $b=2\pi$, the expectation and variance are:

$$\mathrm{E}(X) = \frac{2\pi}{2} = \pi \qquad \text{and} \qquad \mathrm{V}(X) = \frac{(2\pi)^2}{12} = \frac{\pi^2}{3}$$

Mode and Median

Informally, the *mode* of any distribution is the most probable value. This is the value for which $f(x)$ is a maximum. Clearly the Uniform distribution does not have a mode in any useful sense.

Informally, the *median* of any distribution is the middle value. This is the value of x which is such that the area under $f(x)$ to the left of x is equal to the area under $f(x)$ to the right of x. If the value of the median is M then M must be such that:

$$\int_{-\infty}^{\mathrm{M}} f(x)\,dx = \int_{\mathrm{M}}^{+\infty} f(x)\,dx$$

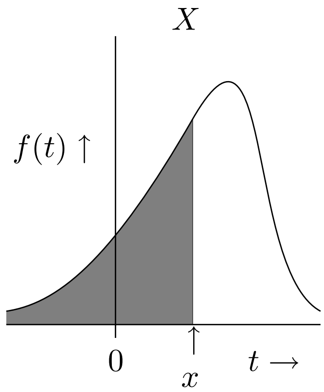
In the case of the Uniform distribution, the median is the same as the mean since the halfway point divides the area into two equal parts.

Probability Distribution Functions

Related to any probability density function $f(x)$ there is an associated function $F(x)$ which is known as the *probability distribution function.* The relationship is:

$$F(x) = \mathrm{P}(X < x) = \int_{-\infty}^{x} f(t)\,dt$$

The following figure shows the relationship diagrammatically. The function $F(x)$ is the area under the curve from the leftmost end of the region of the distribution (which may be $-\infty$) up to x:



Two points stem directly from the definition of a probability distribution function. First:

$$\mathrm{P}(a \leqslant X < b) = F(b) - F(a)$$

Secondly, given that $F(x)$ is the integral of $f(x)$, the derivative of $F(x)$ must be $f(x)$:

$$\frac{d}{dx}\,F(x) = f(x)$$

It is unfortunate that two important functions have the same initial letters. Some writers distinguish the two thus:

pdf	stands for probability density function
PDF	stands for probability distribution function

Given such obvious scope for confusion, the abbreviations will not be used. Moreover, only limited use will be made of probability distribution functions.

The Exponential Distribution

The Exponential distribution, sometimes known as the Negative Exponential distribution, is related to the Geometric and Poisson discrete distributions.

The main design criterion for this distribution is to find, for some random variable X, a probability density function which is such that:

$$\mathrm{P}(X \geqslant x) = e^{-\lambda x}$$

This is a tail probability whose value decreases exponentially as x increases.

In the context of the Poisson distribution, imagine that a town averages one murder a year. The probability of the town having a run of 'at least 10 years' without a murder is substantially less than the probability of lasting 'at least one year' without a murder.

The first step in determining the appropriate probability density function is to find the probability distribution function. Given that $\mathrm{P}(X < x) + \mathrm{P}(X \geqslant x) = 1$:

$$\mathrm{P}(X \geqslant x) = 1 - \mathrm{P}(X < x) = 1 - F(x)$$

Hence:

$$F(x) = 1 - e^{-\lambda x}$$

Differentiate with respect to x:

$$f(x) = \lambda.e^{-\lambda x}$$

This is not quite suitable as a probability density function because the range has not been specified. Clearly the range cannot start from $-\infty$ for this would lead to an infinite area under the curve. The appropriate formal specification of the probability density function for the exponential distribution is:

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

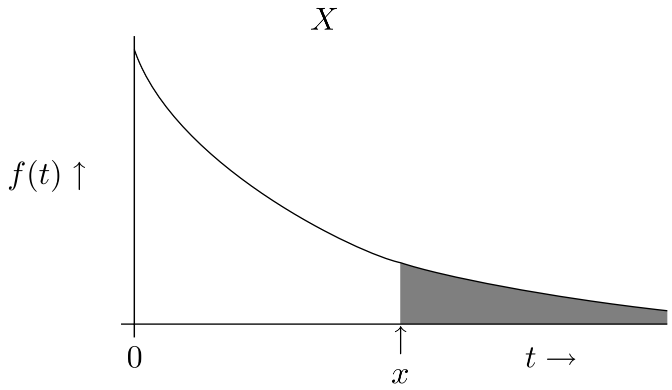
It is simple to check that, without any need for scaling, the integration over the range 0 to ∞ is 1:

$$\int_0^\infty \lambda.e^{-\lambda x}\,dx = \Big[-e^{-\lambda x}\Big]_0^\infty = 1$$

A second check is to confirm that the probability density function satisfies the design criterion that $\mathrm{P}(X \geqslant x) = e^{-\lambda x}$:

$$\mathrm{P}(X \geqslant x) = \int_x^\infty \lambda.e^{-\lambda t}\,dt = \Big[-e^{-\lambda t}\Big]_x^\infty = e^{-\lambda x}$$

A graphical representation of the Exponential distribution is:



The probability $\mathrm{P}(X \geqslant x)$ corresponds to the shaded area. Quite clearly, the larger the value of x the smaller this probability.
The expectation $\mathrm{E}(X)$ is:

$$\mathrm{E}(X) = \int_0^\infty x\lambda.e^{-\lambda x}\,dx$$

Let $t = \lambda x \quad$ so $dx = \frac{1}{\lambda}dt$. Then:

$$\mathrm{E}(X) = \int_0^\infty t.e^{-t}\frac{1}{\lambda}\,dt = \frac{1}{\lambda}\Big[-(t+1)e^{-t}\Big]_0^\infty = \frac{1}{\lambda}$$

The expectation $\mathrm{E}(X^2)$ is:

$$\mathrm{E}(X^2) = \int_0^\infty x^2\lambda.e^{-\lambda x}\,dx$$

Let $t = \lambda x \quad$ so $dx = \frac{1}{\lambda}dt$. Then:

$$\mathrm{E}(X^2) = \int_0^\infty \frac{t}{\lambda}t.e^{-t}\frac{1}{\lambda}\,dt = \frac{1}{\lambda^2}\int_0^\infty t^2.e^{-t}\,dt = \frac{1}{\lambda^2}\Big[-(t^2+2t+2)e^{-t}\Big]_0^\infty = \frac{2}{\lambda^2}$$

The variance $\mathrm{V}(X)$ is:

$$\mathrm{V}(X) = \mathrm{E}(X^2) - \big(\mathrm{E}(X)\big)^2 = \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 = \frac{1}{\lambda^2}$$

In the case of the Poisson distribution both the expectation and the variance are λ.
In the case of the Exponential distribution the expectation is $\frac{1}{\lambda}$ but the variance is $\frac{1}{\lambda^2}$.
An important consideration in the context of the Exponential distribution is that the time you may expect to wait for a No. 9 bus does not depend on when you start waiting for it.

Glossary

The following technical terms have been introduced:

probability density function — median — mode
probability distribution function

Exercises — IX

1. The distribution of the angle α (to the vertical) at which meteorites strike the Earth has probability density function:

$$f(\alpha) = \sin(2\alpha) \quad \text{where } 0 \leqslant \alpha \leqslant \tfrac{\pi}{2}$$

   Find the expectation and variance of the distribution.

2. Find the expectation and variance of the double exponential distribution:

$$f(x) = \tfrac{1}{2}ce^{-c|x|}$$

3. If X has the exponential distribution show that:

$$\mathrm{P}(X > u + v \mid X > u) = \mathrm{P}(X > v) \quad \text{for all } u, v > 0$$

   This is the 'lack of memory property' (c.f. Exercises IV, question 2).