# 7 Continuous random variables

In the case of a discrete random variables we started with outcomes $P(e)$ which is the probability that $e$ is selected; this lead to the probablity of events $P(E)$ and then, for random variables, to the probability $p_X(x)$, which is the probability that the outcome corresponds to the value $x$. In the case where the outcomes are continuous, this story needs some adjustment because the probability of a particular outcome is zero. That is because you can come arbitrarily close to a continuous variable without equaling it. Imagine your random variable is the height of a tree, the chance of the tree being exactly six metres tall is zero, when you measure it, it might be 6.1m or 6.01m or 5.999m or six metres and three yocto metres, but the chance is it exactly six metres is zero. In fact, if we were to do this experiment, measuring trees, we wouldn't mean exactly exactly six metres, we would mean six metres within some percision, say within a centimetre $P(H \in [5.99, 6.01])$ and this is does make sense.

There are two approaches to dealing with this; the first is to define the **distribution function** or **cumulative**

$$F(x) = P(X < x) \tag{1}$$

and we then define the **density function**

$$f(x) = \frac{dF}{dx} \tag{2}$$

or we start with the density function and define the cumulative as

$$F(x) = \int_{-\infty}^{x} f(y)dy \tag{3}$$

Obviously

$$\lim_{x\to\infty} F(x) = 1 \tag{4}$$

or conversely

$$\int_{-\infty}^{\infty} f(y)dy = 1 \tag{5}$$

Now

$$P(x \in [x_1, x_2]) = P(x \leq x_2) - P(x < x_1) \tag{6}$$

For a continuous variable we don't have to be careful about the distinction between $x < x_2$ and $x \leq x_2$. Hence

$$P(x \in [x_1, x_2]) = F(x_2) - F(x_1) \tag{7}$$

or

$$P(x \in [x_1, x_2]) = \int_{x_1}^{x_2} f(y)dy \tag{8}$$

Obviously if $x > y$ then $F(x) \geq F(y)$, adding more outcomes can't reduce the probability. This means that $F(x)$ is a non-descreasing function and hence has a non-negative derivative, so

$$f(x) \geq 0 \tag{9}$$

However, while all probabilities have to be less than one:

$$P(x \in [a, b]) = \int_{a}^{b} f(y)dy \leq 1 \tag{10}$$

this doesn't mean $f(x)$ itself is bounded by one; it can exceed one provided its integral over any limit is less than one. We will see examples of this when we look at the Gaußian distribution.

We can see when the density is called the density; it is very similar to working out mass, say you had a rod with variable density $\rho(x)$ where $x$ is where you are along the rod. It wouldn't make sense to ask the total mass of the rod at a point $x$, that would be zero, but we could work out the mass of the rod between two points $x_1$ and $x_2$; you'd calculate that by integrating the density.

The simplest example is a constant distribution where the outcome is equally likely within some interval. This means that the probability over a subinterval is proportional to the length of the subinterval. Thus, say the interval is $[-1,1]$:

$$f(x) = \begin{cases} \frac{1}{2} & x \in [-1,1] \\ 0 & \text{otherwise} \end{cases} \tag{11}$$

This function and the corresponding cumulative is shown in Fig. 1.

The expected value is defined using the density:

$$\langle g(X)\rangle = \int_{-\infty}^{\infty} g(x)f(x)dx \tag{12}$$

Hence, for the constant example:

$$\mu = \langle X\rangle = \int_{-\infty}^{\infty} xf(x)dx = \frac{1}{2}\int_{-1}^{1} xdx = 0 \tag{13}$$

and

$$\langle X^2\rangle = \int_{-\infty}^{\infty} x^2 f(x)dx = \frac{1}{2}\int_{-1}^{1} x^2dx = \frac{1}{3} \tag{14}$$

This continuous version of the expected value has the same nice properties that the discrete version did: with the obvious notation

$$\langle c\rangle = c \tag{15}$$

and

$$\langle cg(X)\rangle = c\langle g(X)\rangle \tag{16}$$

and

$$\langle g_1(X) + g_2(X)\rangle = \langle g_1(X)\rangle + \langle g_2(X)\rangle \tag{17}$$

It is useful to note that these properties can be used to see how the mean and variance change under simple linear transformations. With the obvious notation let $Y = X + c$, then

$$\mu_Y = \langle Y\rangle = \langle X + c\rangle = \langle X\rangle + c = \mu_X + c \tag{18}$$

whereas

$$\sigma_Y^2 = \langle Y^2\rangle - \mu_Y^2 = \langle (X^2 + 2cX + c^2)\rangle - \mu_X^2 - 2c\mu_X - c^2 = \langle X^2\rangle - \mu_X^2 = \sigma_X^2 \tag{19}$$

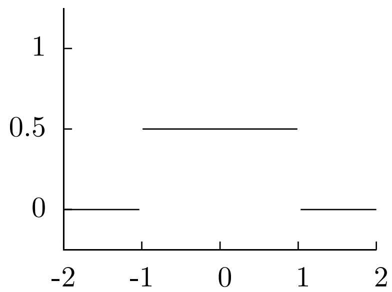
Similarly, if $Y = cX$ then

$$\mu_Y = \langle Y\rangle = \langle cX\rangle = c\mu_X \tag{20}$$

and

$$\sigma_Y^2 = \langle Y^2\rangle - \mu_Y^2 = c^2(\langle X^2\rangle - \mu_X^2) = c^2\sigma_Y^2 \tag{21}$$

**A**




**B**

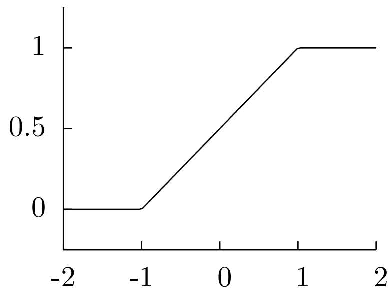



Figure 1: The density **A** and the cumulative **B** for the distribution constant over $[-1, 1]$.

## Summary

- The **distribution function** or **cumulative** is

$$F(x) = P(X < x) \tag{22}$$

so $\lim_{x\to-\infty} F(x) = 0$ and $\lim_{x\to\infty} F(x) = 1$.

- The **density function** is

$$f(x) = \frac{dF}{dx} \tag{23}$$

- By integrating we get

$$F(x) = \int_{-\infty}^{x} f(y)dy \tag{24}$$

and so

$$\int_{-\infty}^{\infty} f(y)dy = 1 \tag{25}$$

- Hence

$$P(x \in [x_1, x_2]) = F(x_2) - F(x_1) \tag{26}$$

or

$$P(x \in [x_1, x_2]) = \int_{x_1}^{x_2} f(y)dy \tag{27}$$

- $F(x)$ is a non-descreasing function so $f(x) \geq 0$. However while $\int_{x_0}^{x_1} f(x)dx \leq 1$ for any $x_1 > x_0$ there is no upperbound on $f(x)$.
- Expected values work much the same way they did for discrete random variables.
- If $Y = X + c$ then $\mu_Y = \mu_x + c$ and $\sigma_Y^2 = \sigma_X^2$.
- If $Y = cX$ then $\mu_Y = c\mu_X$ and $\sigma_Y^2 = c^2\sigma_X^2$.