Elementary Statistics Course

Alfredo Sánchez Alberca (asalber@ceu.es)

Feb 2016

Department of Applied Math and Statistics
CEU San Pablo

CEU Universidad San Pablo

License terms

Contents

Continuous random variables

Continuous random variables

Continuous random variables, unlike discrete random variables, can take any value in a real interval. Thus the range of a continuous random variables is infinite and uncountable.

Such a density of values makes impossible to compute the probability for each one of them, and therefore, it's not possible to define a probabilistic model trough a probability function like with discrete random variables.

Besides, usually the measurement of continuous random variable is limited by the precision of the measuring instrument. For instance, when somebody says that is 1.68 meters tall, his or her true height is no exactly 1.68 meters, cause the precision of the measuring instrument is only cm (two decimal places). This means that the true height of that person is between 1.675 y 1.685 meters.

Hence, for continuous variables, *it make no sense to calculate the probability of an isolated value, and we will calculate probabilities for intervals.*

To model the probability distribution of a continuous random variable is used a probability density function.

Definition (Probability density function)

The *probability density function* of a continuous random variable X is a function $f(x)$ that meets the following conditions:

- Is non-negative: $f(x) \geq 0 \; \forall x \in \mathbb{R}$,
- The area bounded by the curve of the density function and the x-axis is equal to 1, that is,
$$\int_{-\infty}^{\infty} f(x)\, dx = 1.$$
- The probability that X assumes a value between a and b is equal to the area under the density function bounded by a and b, that is,
$$P(a \leq X \leq b) = \int_{a}^{b} f(x)\, dx$$

The probability density function measures the relative likelihood of every

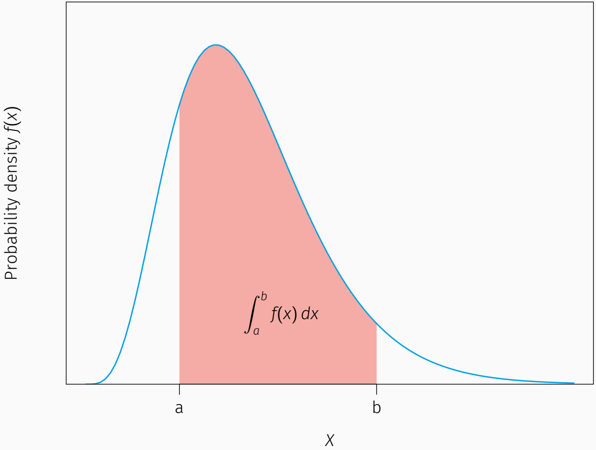
The same way that for discrete random variables, for continuous random variables it makes sense to calculate cumulative probabilities.

Definition (Distribution function)

The *distribution function* of a continuous random variable X is a function $F(x)$ that maps every value a to the probability that X takes on a value less than or equal to a, that is,

$$F(a) = P(X \leq a) = \int_{-\infty}^{a} f(x)\, dx.$$

Probabilities as areas



$$P(a \leq X \leq b) = \int_a^b f(x)\,dx = F(b) - F(a)$$

Given the following function

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

let's check that is a density function.

As this function is clearly non-negative, we have to check that total area bounded by the curve and the x-axis is 1.

$$\int_{-\infty}^{\infty} f(x)\, dx = \int_{-\infty}^{0} f(x)\, dx + \int_{0}^{\infty} f(x)\, dx = \int_{-\infty}^{0} 0\, dx + \int_{0}^{\infty} e^{-x}\, dx =$$
$$= \left[-e^{-x}\right]_0^{\infty} = -e^{-\infty} + e^0 = 1.$$

Now, let's calculate the probability of X having a value between 0 and 2.

$$P(0 \leq X \leq 2) = \int_0^2 f(x)\, dx = \int_0^2 e^{-x}\, dx = \left[-e^{-x}\right]_0^2 = -e^{-2} + e^0 = 0.8646.$$

Population statistics

The calculation of the population statistics is similar to the case of discrete variables, but using the density function instead of the probability function, and extending the discrete sum to the integral.

The most important are:

- Mean:
$$\mu = E(X) = \int_{-\infty}^{\infty} xf(x)\,dx$$
- Variance:
$$\sigma^2 = Var(X) = \int_{-\infty}^{\infty} x^2 f(x)\,dx - \mu^2$$
- Standard deviation:
$$\sigma = +\sqrt{\sigma^2}$$

Let X be a variable with the following probability density function

$$f(x) = \begin{cases} 0 & \text{si } x < 0 \\ e^{-x} & \text{si } x \geq 0 \end{cases}$$

The mean is

$$\mu = \int_{-\infty}^{\infty} xf(x)\,dx = \int_{-\infty}^{0} xf(x)\,dx + \int_{0}^{\infty} xf(x)\,dx = \int_{-\infty}^{0} 0\,dx + \int_{0}^{\infty} xe^{-x}\,dx =$$
$$= \left[-e^{-x}(1+x)\right]_0^{\infty} = 1.$$

and the variance is

$$\sigma^2 = \int_{-\infty}^{\infty} x^2 f(x)\,dx - \mu^2 = \int_{-\infty}^{0} x^2 f(x)\,dx + \int_{0}^{\infty} x^2 f(x)\,dx - \mu^2 =$$
$$= \int_{-\infty}^{0} 0\,dx + \int_{0}^{\infty} x^2 e^{-x}\,dx - \mu^2 = \left[-e^{-x}(x^2+2x+2)\right]_0^{\infty} - 1^2 = 2e^0 - 1 = 1.$$

According to the type of experiment where the random variable is measured, there are different probability distributions models. The most common are

- Continuous uniform.
- Normal.
- Student's T.
- Chi-square.
- Fisher-Snedecor's F.

Continuous uniform probability distribution model $U(a, b)$

When all the values of a random variable X have equal probability, the probability distribution of X is uniform.

Definition (Continuous uniform distribution $U(a, b)$)

A continuouo random variable X follows a probability distribution model *uniform* of parameters a and b, noted $X \sim U(a, b)$, if its range is $\mathrm{Ran}(X) = [a, b]$ and its density function is

$$f(x) = \frac{1}{b-a} \quad \forall x \in [a, b]$$

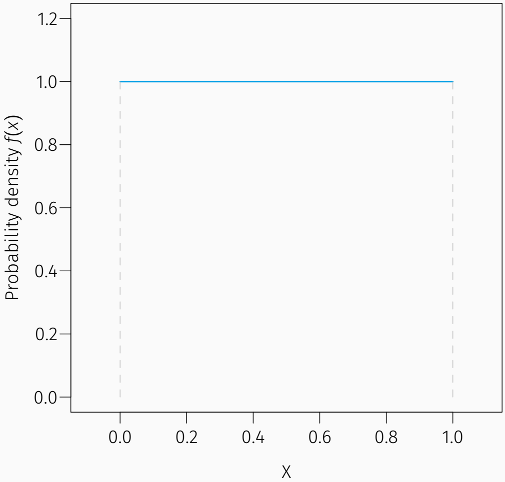
Observe that a and b are the minimum and the maximum of the range respectively, and that the density function is constant.

The mean and the variance are

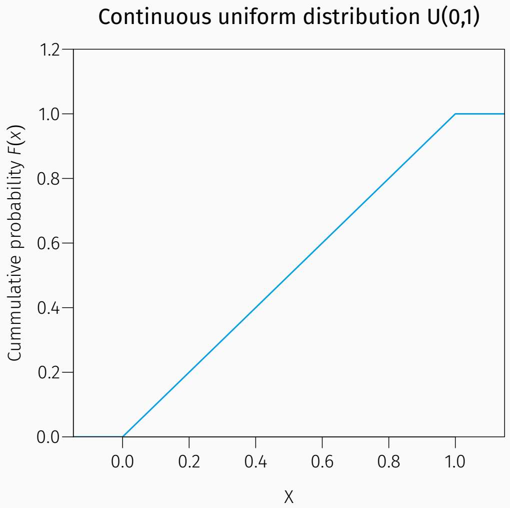
$$\mu = \frac{a+b}{2} \qquad \sigma^2 = \frac{(b-a)^2}{12}.$$

The generation of a random number between 0 and 1 is follows a continuous uniform distribution $U(0, 1)$.

Continuous unniform distribution U(0,1)



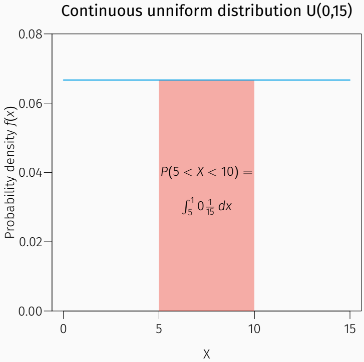
As the density function is constant, the distribution function has a linear growth.



Supóngase que un autobús pasa por una parada cada 15 minutos. Si una persona puede llegar a la parada en cualquier instante, *¿cuál es la probabilidad de que espere entre 5 y 10 minutos?*

En este caso, la variable X que mide el tiempo de espera sigue un modelo de distribución uniforme continua $U(0,15)$ ya que cualquier valor entre los 0 y los 15 minutos es equipobrable. Así pues, la probabilidad que nos piden es

$$P(5 \leq X \leq 10) = \int_5^{10} \frac{1}{15}\, dx = \left[\frac{X}{15}\right]_5^{10} =$$
$$= \frac{10}{15} - \frac{5}{15} = \frac{1}{3}.$$



Además, el tiempo medio de espera será $\mu = \frac{0+15}{2} = 7.5$ minutos.