Continuous Random Variables

Pablo Arés Gastesi

Departamento de Matemáticas y Ciencia de Datos
Universidad San Pablo CEU
pablo.aresgastesi@ceu.es

Septembre 2024

Density Function

Continuos random variables take infinitely many values, so we cannot compute the probability of the variable being equal to a point; that is, $P(X = x)$ does not make sense. Instead, we compute probabilities of intervals.

The **density function** $f(x)$ *of a continuos random variable* X *is a function satifying the following:*

- $f(x) \geq 0$ *for all* x*.*
- $\int_{-\infty}^{\infty} f(x)\, dx = 1$
- $P(a \leq X \leq b) = \int_a^b f(x)\, dx$

Cumulative Distribution Function

As in the discrete case, the **cumulative distribution function** $F(x)$ gives the probability of the random variable X not exceeding the value x:

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(s)\, ds$$

Important

Given the density function $f(x)$ and the cumulative distribution function $F(x)$ of X we have $F'(x) = f(x)$.

We also have a relation between the cumulative distribution function and the probability of X being in an interval:

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

Expected Value and Variance

For a continuous random variable X we must integrate to compute the expected value $E[X]$ and the variance $V[X]$.

$$\mu_X = E[X] = \int_{-\infty}^{\infty} x\, f(x)\, dx$$

$$\sigma_X^2 = V[X] = \int_{-\infty}^{\infty} (x-\mu)^2\, f(x)\, dx$$

Important

$$\sigma_x^2 = E[X^2] - \mu^2$$

Uniform Distribution

The **uniform distribution** between a and b, with $a < b$, $U(a, b)$ gives equal probabilities to all parts of the interval $[a, b]$. Its density function is $f(x) = \frac{1}{b-a}$ for $a \leq x \leq b$, and $f(x) = 0$ otherwise.

Important

If $X \sim U(a, b)$ we have

- *The density function is given by*

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

- $E[X] = \frac{a+b}{2}$
- $V[X] = \frac{(b-a)^2}{12}$

Normal Distribution

A random variable X is **normally distributed**, with mean μ and standard deviation σ, written as $X \sim N(\mu, \sigma)$ if its density function is given by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

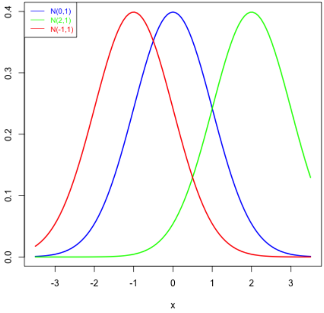
Do not panic!! You don't need to learn this formula.

Important

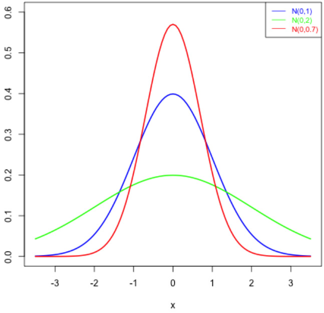
If $X \sim N(\mu, \sigma)$ *then*

$$E[X] = \mu, \qquad V[X] = \sigma^2$$

Normal Distribution



Change in mean



Change in standard deviation

Standardization

Important (Standardization)

If $Y \sim N(\mu, \sigma)$ *then*

$$X = \frac{Y - \mu}{\sigma} \sim N(0, 1)$$

Central Limit Theorem

Important (Central Limit Theorem)

Let $X_1, X_2, \ldots, X_n$ *be independent, identically distributed random variables with* $E[X_i] = \mu$ *and* $Var[X_i] = \sigma^2 < \infty$ *for all variables. Set* $Y = \frac{1}{n}(X_1 + X_2 + \cdots + X_n)$. *Then*

$$\sqrt{n}\,(Y - \mu) \sim N(0, \sigma)$$

when n is big enough. Or, equivalently,

$$Y \sim N(\mu, \sigma/\sqrt{n})$$

The binomial and the normal distributions

Since the binomial distribution is a sum of independent Bernoulli distributions with the same probability, we can use the normal distribution to approximate the binomial distribution when the parameter n is big (probabilities for the binomial are hard to compute; for the normal, we have tables).

Important

If $X \sim B(n, p)$, with $np(1-p) > 5$, then the variable

$$Z = \frac{X - np}{\sqrt{np(1-p)}}$$

is normally distributed: $Z \sim N(0, 1)$. Or, equivalently, for n big enough, $B(n, p)$ can be approximated by

$$B(n, p) \approx N(np, \sqrt{np(1-p)})$$

The binomial and the normal distributions

For example, if $X \sim B(100, 0.4)$, we have $np(1-p) = 100(0.4)(0.6) = 24 > 5$. To compute the probability of X being between 45 and 50 (that is, $X = 45, 46, 47, 48, 49, 50$) we can do with a normal distribution, after standardization. We have $np = 40$ and $np(1-p) = 24$, $\sqrt{24} = 4.9$, so $Z = \frac{X-40}{4.9} \sim N(0,1)$. We compute the normalized values of the end points, 45 and 50:

$$\frac{45-40}{4.9} = 1.02, \qquad \frac{50-40}{4.9} = 2.04$$

So

$$P(45 \leq X \leq 50) \approx P(1.02 \leq Z \leq 2.024) = F(2.04) - F(1.02) = 0.133$$

Computing by the binomial formula we have the probability is approximately 0.16

χ^2 Distribution

The chi-squared distribution χ^2_n with n degrees of freedom is obtained as the sum of of the squared of independent, identically distributed standard normal distributions. That is, if $Z_1, Z_2, \ldots, Z_n$ are independent, all of the $Z_i \sim N(0,1)$, then

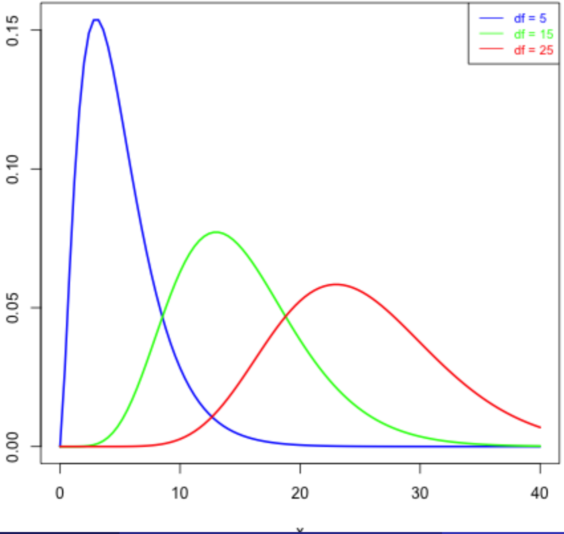
$$\chi^2_n = Z_1^2 + Z_2^2 + \cdots + Z_n^2$$

If $X \sim \chi^2_n$ then we have the following important facts:

Important

- *X only takes only positive values.*
- *The density function $f(x) = \ldots$ something very complicated.*
- $E[X] = n$
- $V[X] = 2n$

χ^2 Distribution



Student's t Distribution

If $Z \sim N(0,1)$ and $X \sim \chi^2_n$ (with $n \geq 1$) are independent, then the **Student's t distribution** with n degrees of freedom t_n is a continuous random varible that generalizes the normal distribution. It is given by

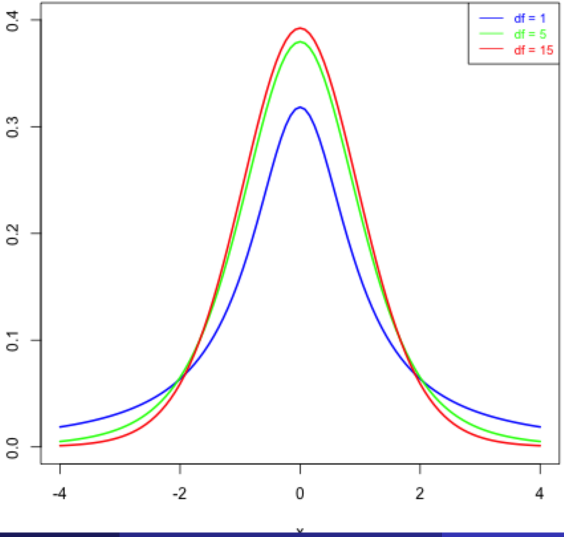
$$t_n = \frac{Z}{\sqrt{X/n}}$$

Important

If $T \sim t_n$ then we have

- *T takes all real values*
- *The density function is, again, complicated.*
- *$E[T] = 0$ if $n > 1$; undefined if $n = 1$*
- *$V[T] = \frac{n}{n-2}$ if $n > 2$; ∞ for $n = 2$; undefined for $n = 1$*

Students' t Distribution



Snedecor's F Distribution

The Fisher-Snedecor distribution, or **Snedecor's F** distribution, with m and n degrees of freedom, $F_{m,n}$ is given by

$$F_{m,n} = \frac{X/m}{Y/n}$$

where $X \sim \chi^2_m$ and $Y \sim \chi^2_n$ are independent χ^2 distributions.

Important

If $X \sim F_{m,n}$ we have

- *The order of m and n is important!*
- *X takes only positive values.*
- $E[X] = \frac{n}{n-2}$ *for* $n > 2$
- $V[X] = \frac{2n^2\,(m+n-2)}{m\,(n-2)^2\,(n-4)}$ *for* $n > 4$

Students' t Distribution