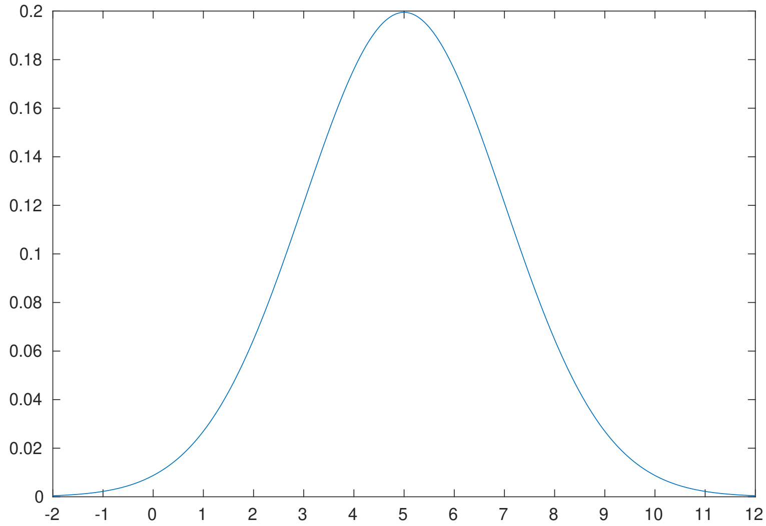
1 The normal distribution

One of the most common continuous distributions is the normal distribution. A random variable X is said to be **normally distributed with parameters μ and σ**, in short $X \in N(\mu, \sigma)$, if the probability density function of X is given by

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



Figur 1: The probability density function of a normal distribution (N(5,2)).

The distribution is symmetrical around μ, therefore

$$E[X] = \mu$$

and it can be shown that

$$D(X) = \sigma.$$

Remark 1 The parameters of the normal distribution are thus the distribution's expectation and standard deviation. It is quite common to use the variance as a parameter instead of the standard deviation. Again make sure that you know which parameterization that is used!

As has already been mention the normal distribution is a frequently appearing distribution. This is partly due to its nice mathematical properties, and partly due to the Central limit theorem.
We begin with a nice mathematical property (which is also pivotal to the central limit theorem).

Proposition 1 *Let $X_1, X_2; \ldots, X_n$ be independent random variables such that $X_i \in N(\mu_i, \sigma_i)$, $i = 1, \ldots, n$. Furthermore, let*

$$Y = a_1X_1 + a_2X_2 + \ldots + a_nX_n + b = \sum_{i=1}^{n} a_iX_i + b.$$

Then $Y \in N(\mu, \sigma)$ *where*

$$\begin{aligned}
\mu &= E[Y] = E\left[\sum_{i=1}^{n} a_i X_i + b\right] \\
&= \sum_{i=1}^{n} a_i E[X_i] + b \\
&= \sum_{i=1}^{n} a_i \mu_i + b \\
\sigma^2 &= V(Y) = V\left(\sum_{i=1}^{n} a_i X_i + b\right) = \{\mathit{independence}\} \\
&= \sum_{i=1}^{n} V(a_i X_i) = \sum_{i=1}^{n} a_i^2 V(X_i) \\
&= \sum_{i=1}^{n} a_i^2 \sigma_i^2
\end{aligned}$$

Remark 2 In words the proposition states that "Linear combinations of independent normally distributed random variables are normally distributed".

Let X be $N(\mu, \sigma)$ and let

$$Z = \frac{X - \mu}{\sigma} = \frac{1}{\sigma} X - \frac{\mu}{\sigma}$$

According to Proposition 1 Z is a normally distributed random variable with

$$\begin{aligned}
E[Z] &= E\left[\frac{1}{\sigma} X - \frac{\mu}{\sigma}\right] = \frac{1}{\sigma} E[X] - \frac{\mu}{\sigma} = 0 \\
V(Z) &= V\left(\frac{1}{\sigma} X - \frac{\mu}{\sigma}\right) = \frac{1}{\sigma^2} V[X] = 1
\end{aligned}$$

i.e. $Z \in N(0, 1)$ which is known as the *standard normal distribution*. We have that

$$\begin{aligned}
\varphi(x) &= f_Z(x) = \frac{2}{\sqrt{2\pi}} e^{-x^2/2} \\
\Phi(x) &= F_Z(t) = P(Z \le t) = \int_{-\infty}^{t} \varphi(x) dx
\end{aligned}$$

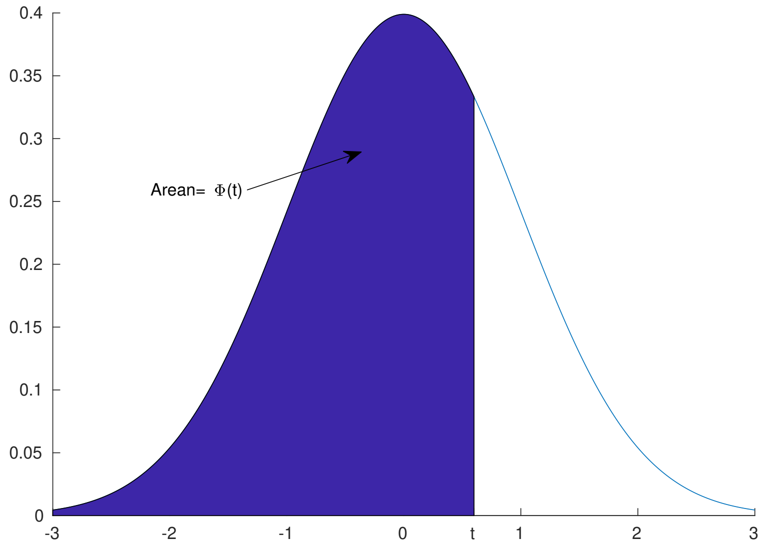
There is no closed form expression for the distribution function Φ. In the compiled formulae there is a table of certain values and more advanced calculators have the function programmed.

Note that if $X \in N(\mu, \sigma)$ then

$$P(a < X \le b) = P\left(\frac{a - \mu}{\sigma} < \underbrace{\frac{X - \mu}{\sigma}}_{N(0,1)} \le \frac{b - \mu}{\sigma}\right)$$

$$P\left(\frac{a - \mu}{\sigma} < Z \le \frac{b - \mu}{\sigma}\right) = P\left(Z \le \frac{b - \mu}{\sigma}\right) - P\left(Z \le \frac{a - \mu}{\sigma}\right)$$

$$= \Phi\left(\frac{b - \mu}{\sigma}\right) - \Phi\left(\frac{a - \mu}{\sigma}\right).$$



Figur 2: The distribution function of the standard normal distribution.

Remark 3 Since the normal distribution is a continuous distribution it does not matter if you change a non-strict inequality to a strict inequality (or the other way around). This is because the probability of the random variable assuming any single value is 0.

1.1 The Central Limit Theorem

Proposition 2 (The Central limit theorem) *Let $X_1, X_2, \ldots$ be independent identically distributed random variables with expectation μ and standard deviation $\sigma > 0$ and let*

$$\bar{Y}_n = \sum_{i=1}^{n} X_i.$$

Then the following holds

$$P\left(a < \frac{\bar{Y}_n - n\mu}{\sigma\sqrt{n}} \leq b\right) \to \Phi(b) - \Phi(a) \qquad \text{as } n \to \infty$$

Remark 4 In words the proposition says that the sum of many independent identically distributed random variables is approximately normally distributed".

In particular it holds that

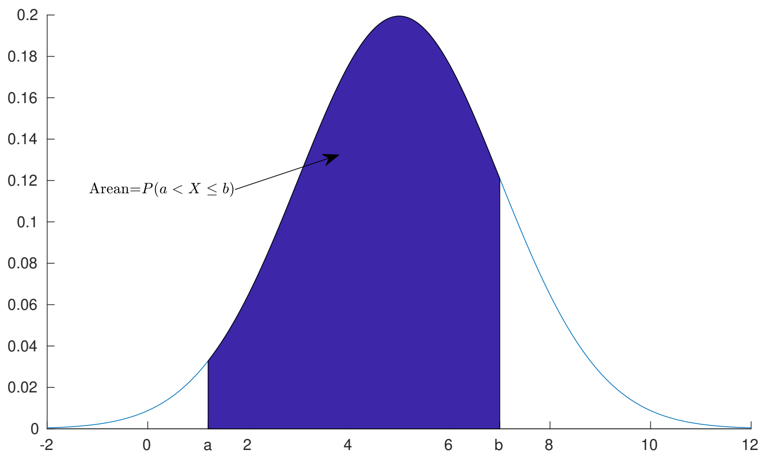
$$\sum_{i=1}^{n} X_i \quad \text{approximately} \quad N(n\mu, \sqrt{n}\sigma)$$

and that

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \quad \text{approximately} \quad N(\mu, \sigma)$$

given that n is large enough.

Example 1 Do problem 818 in [2]. □



Figur 3: $P(a < X \leq b)$ då $X \in N(5, 2)$.

Referenser

[1] Blom, G., Enger, J., Englund, G., Grandell, J., och Holst, L., (2005). Sannolikhetsteori och statistikteori med tillämpningar.

[2] Blom, Gunnar, (1989). Probability and Statistics. Theory and Applications.