Cumulative probability and probability density functions

The probability that an event will give values in the range x and $x + \Delta x$. Continuous distribution:

$$c.d.f.(x <= X) = \int_0^X p(x)dx \tag{3.1}$$

Discrete distribution:

$$c.d.f.(x <= X) = \sum_{i=1}^{N} p(x_i)dx \tag{3.2}$$

The expected value is

$$\langle x \rangle = \int_0^\infty xp(x)dx \tag{3.3}$$

Whereas the probability can be calculated from the ratio of the area of the *p.d.f* that corresponds to the condition that the event taking place to the area of the total curve, the probability can be read off directly from the *cumulative distribution function* (*c.d.f.*). The *c.d.f* starts with the value 0 and ends with 1.

An example of a *c.d.f* is shown in Fig. 4.2b, in Chapter 4. The equivalent to a *c.d.f* for actual observations is the *empirical cumulative function* (*e.d.f*). Table 2 gives an overview of the four main types of distribution functions.

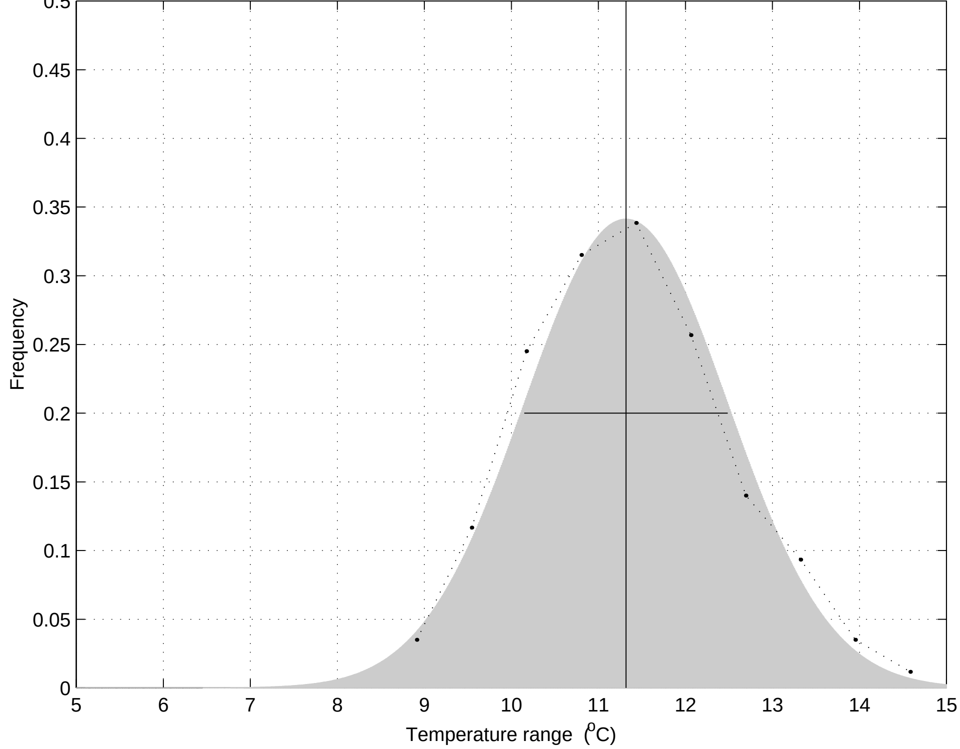

Figure 3.2: *An example of a Gaussian distribution curve. The vertical line marks the mean value and the horizontal line shows the* $\pm\sigma$ *range. The empirical probability distribution for the Bergen September temperature is also shown as black dots. [stats_uib_3_2.m]*

3.2 Normal/Gaussian distribution

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]. \qquad (3.4)$$

There are a number of commonly used theoretical distribution functions, which have been derived for ideal conditions. One such case is where the process $(y_i, i = [1...N])$ is random (stochastic), and whose distribution follows a Gaussian shape described by $f(x)$ in equation 3.4. This distribution function is widely used in statistical sciences. σ is in this case estimated by taking the standard deviation: $\sigma = std(y_i)$, and μ is taken as the mean value of y_i.

Fig. 3.2 shows a typical example of a Gaussian distribution. The values for σ and μ have been taken from the Bergen September 2-m temperature 1861-1997 record, and the empirical histogram for the temperature record is also shown as black dots.

The Gaussian distribution function in Fig 3.2 gives a concise and approximate description of the Bergen September temperature range and likelihood of occurrence. The mean and standard deviation, the two parameters used for fitting the Gaussian function to the observations, give a good description of the Bergen temperature statistics.