which is read as saying the probability of getting a data value $y$, given $(\mu, \sigma^2)$ a mean of $\mu$ and a variance of $\sigma^2$, is calculated from this rather complicated-looking two-parameter exponential function. For any given combination of $\mu$ and $\sigma^2$, it gives a value between 0 and 1. Recall that likelihood is the product of the probability densities, for each of the values of the response variable, $y$. So if we have $n$ values of $y$ in our experiment, the likelihood function is

$$L(\mu, \sigma) = \prod_{i=1}^{n} \left( \frac{1}{\sigma \sqrt{2\pi}} \exp \left[ -\frac{(y_i - \mu)^2}{2\sigma^2} \right] \right),$$

where the only change is that $y$ has been replaced by $y_i$ and we multiply together the probabilities for each of the $n$ data points. There is a little bit of algebra we can do to simplify this: we can get rid of the product operator, $\Pi$, in two steps. First, for the constant term: that, multiplied by itself $n$ times, can just be written as $1/(\sigma \sqrt{2\pi})^n$. Second, remember that the product of a set of antilogs (exp) can be written as the antilog of a sum of the values of $x_i$ like this: $\prod \exp(x_i) = \exp(\sum x_i)$. This means that the product of the right-hand part of the expression can be written as

$$\exp \left[ -\frac{\sum_{i=1}^{n} (y_i - \mu)^2}{2\sigma^2} \right]$$

so we can rewrite the likelihood of the normal distribution as

$$L(\mu, \sigma) = \frac{1}{(\sigma \sqrt{2\pi})^n} \exp \left[ -\frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \mu)^2 \right].$$

The two parameters $\mu$ and $\sigma$ are unknown, and the purpose of the exercise is to use statistical modelling to determine their maximum likelihood values from the data (the $n$ different values of $y$). So how do we find the values of $\mu$ and $\sigma$ that maximize this likelihood? The answer involves calculus: first we find the derivative of the function with respect to the parameters, then set it to zero, and solve.

It turns out that because of the exp function in the equation, it is easier to work out the log of the likelihood,

$$l(\mu, \sigma) = -\frac{n}{2} \log(2\pi) - n \log(\sigma) - \sum (y_i - \mu)^2/2\sigma^2,$$

and maximize this instead. Obviously, the values of the parameters that maximize the log-likelihood $l(\mu, \sigma) = \log(L(\mu, \sigma))$ will be the same as those that maximize the likelihood. From now on, we shall assume that summation is over the index $i$ from 1 to $n$.

Now for the calculus. We start with the mean, $\mu$. The derivative of the log-likelihood with respect to $\mu$ is

$$\frac{dl}{d\mu} = \sum (y_i - \mu)/\sigma^2.$$  

Set the derivative to zero and solve for $\mu$:

$$\sum (y_i - \mu)/\sigma^2 = 0 \quad \text{so} \quad \sum (y_i - \mu) = 0.$$
Taking the summation through the bracket, and noting that $\sum \mu = n\mu$,

$$\sum y_i - n\mu = 0 \quad \text{so} \quad \sum y_i = n\mu \quad \text{and} \quad \mu = \frac{\sum y_i}{n}.$$  

The maximum likelihood estimate of $\mu$ is the arithmetic mean.

Next we find the derivative of the log-likelihood with respect to $\sigma$:

$$\frac{d}{d\sigma} l = -\frac{n}{\sigma} + \frac{\sum (y_i - \mu)^2}{\sigma^3},$$

recalling that the derivative of $\log(x)$ is $1/x$ and the derivative of $-1/x^2$ is $2/x^3$. Solving, we get

$$-\frac{n}{\sigma} + \frac{\sum (y_i - \mu)^2}{\sigma^3} = 0 \quad \text{so} \quad \sum (y_i - \mu)^2 = \sigma^3 \left(\frac{n}{\sigma}\right) = \sigma^2 n \quad \sigma^2 = \frac{\sum (y_i - \mu)^2}{n}.$$  

The maximum likelihood estimate of the variance $\sigma^2$ is the mean squared deviation of the $y$ values from the mean. This is a biased estimate of the variance, however, because it does not take account of the fact that we estimated the value of $\mu$ from the data. To unbias the estimate, we need to lose 1 degree of freedom to reflect this fact, and divide the sum of squares by $n - 1$ rather than by $n$ (see p. 52 and restricted maximum likelihood estimators in Chapter 19).

Here, we illustrate R’s built-in probability functions in the context of the normal distribution. The density function $\text{dnorm}$ has a value of $z$ (a quantile) as its argument. Optional arguments specify the mean and standard deviation (default is the standard normal with mean 0 and standard deviation 1). Values of $z$ outside the range $-3.5$ to $+3.5$ are very unlikely.

```r
par(mfrow=c(2,2))
curve(dnorm,-3,3,xlab="z",ylab="Probability density",main="Density")
```

The probability function $\text{pnorm}$ also has a value of $z$ (a quantile) as its argument. Optional arguments specify the mean and standard deviation (default is the standard normal with mean 0 and standard deviation 1). It shows the cumulative probability of a value of $z$ less than or equal to the value specified, and is an S-shaped curve:

```r
curve(pnorm,-3,3,xlab="z",ylab="Probability",main="Probability")
```

Quantiles of the normal distribution $\text{qnorm}$ have a cumulative probability as their argument. They perform the opposite function of $\text{pnorm}$, returning a value of $z$ when provided with a probability.

```r
curve(qnorm,0,1,xlab="p",ylab="Quantile (z)",main="Quantiles")
```

The normal distribution random number generator $\text{rnorm}$ produces random real numbers from a distribution with specified mean and standard deviation. The first argument is the number of numbers that you want to be generated: here are 1000 random numbers with mean 0 and standard deviation 1: