CHAPTER 3

POSTERIOR DISTRIBUTIONS

3.1. Notation

The letter ‘f’ will always be used for probability density functions. Thus f(x) and f(y) are different functions (for example, a normal function and a gamma function) although both are probability density functions.

The variables will be in red colour, independently on whether they are after a statement of ‘|’ (given) or not. For example,

$$f(y \mid \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(y-\mu)^2}{2\sigma^2}\right]$$

is a function of σ^2, but it is not a function of y or a function of μ, that are considered as constants. Thus, in this example we represent a family of density functions for different values of σ^2. An example of this type of function is the likelihood (see chapter 1).

The sign $\propto$ means “proportional to”. We will often work with proportional functions, since it is easier and as we will see later, the results from proportional functions can be reduced to (almost) exact results easily with MCMC. For example, if c and k are constants, the distribution N(0,1) is

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \propto \exp\left(x^2\right) \propto k \cdot \exp(c) \cdot \exp\left(x^2\right) = k \cdot \exp\left(c + x^2\right)$$

Thus we can add or subtract constants from an exponent, or multiply or divide if the constants are not in the exponent. What we cannot do is

$$f(x) \propto \exp\left(x^2\right) \neq k + \exp\left(x^2\right)$$

$$f(x) \propto \exp\left(x^2\right) \neq \exp\left(c \cdot x^2\right) = \left[\exp\left(x^2\right)\right]^c$$

3.2. Cumulative distribution

Take the example of a single gene with two alleles (A,a) producing three genotypes aa, Aa, AA. We define a variable x with three values 0,1,2 corresponding to the three genotypes. If we cross two individuals Aa, according to Mendel’s law, the probability of each offspring genotype is

	aa	Aa	AA
x_0 =	0	1	2
P(x = x_0) =	¼	½	¼

We define the cumulative distribution function as

$F(x_0) = P(x \le x_0)$

in this example, this cumulative distribution function is (figure 3.1)

$F(0) = P(x \le 0) = 1/4$
$F(1) = P(x \le 1) = 1/4 + 1/2 = 3/4$
$F(2) = P(x \le 2) = 1/4 + 1/2 + 1/4 = 1$
For all values $x<0$ we have $F(x) = 0$
For all values $0 \le x < 1$ we have $F(x) = 1/4$
For all values $1\le x < 2$ we have $F(x) = 3/4$
For all values $x \ge 2$ we have $F(x) = 1$

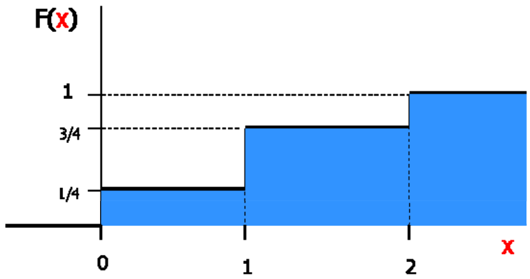

Figure 3.1. Cumulative distribution function

The probability of x to be between 1 and 2 (equal or lower than 2 but higher than 1)

$P(1 < x \le 2) = P(x\le2) – P(x\le1) = F(2) – F(1) = 1 – 3/4 = 1/4$

Usually the cumulative distribution functions are continuous, for example

$$F(x) = x = P(x_0 \le x)$$

The probability of x to be between a and b is

$P(a < x \le b) = P(x \le b) – P(x \le a) = F(a) – F(b)$

3.3. Density distribution

3.3.1. *Definition*

We define the probability density function (figure 3.3.) as

$$f(\mathbf{x}) = \frac{\Delta F(\mathbf{x})}{\Delta \mathbf{x}}$$

this function is always positive because $\Delta F(x)$ and Δx are both positive. In the example of figure 3.1., the values of $\Delta F(x)$ are ¼ until 0, ½ from 0 to 1 and ½ from 1 upwards. Then,

$\sum f(x) \cdot \Delta x = \sum \Delta F(x) = 1/4 + 1/2 + 1/4 = 1$

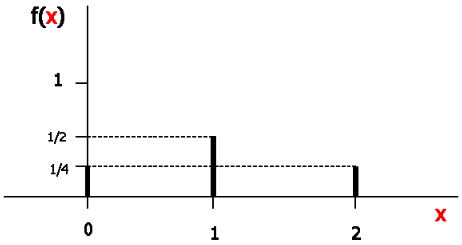

Figure 3.2. Probability density function

in the continuous case

$$f(x) = \frac{dF(x)}{dx}$$

this means (figure 3.3a) that

$$F(x_0) = \int_{-\infty}^{x_0} f(x)\,dx = P(x \le x_0)$$

and, analogously as in the discrete case

$$\int_{-\infty}^{+\infty} f(x)\,dx = 1$$

The probability of x to be between two values 'a' and 'b' is

$$P(x \le b) - P(x \le a) = F(b) - F(a) = \int_{-\infty}^{b} f(x)dx - \int_{-\infty}^{a} f(x)dx = \int_{a}^{b} f(x)dx$$

which is the area defined by f(x) between 'a' and 'b' (figure 2.3**b**)

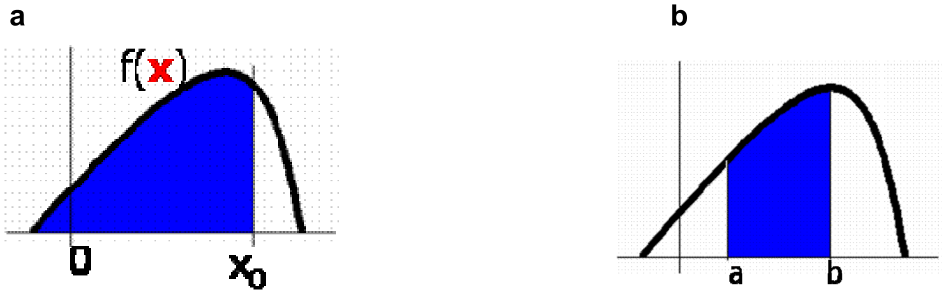

Figure 3.3. Probability density functions. **a.** In blue, $P(x \le x_0)$, **b.** In blue, $P(a \le x \le b)$

Notice that $f(x_0)$ is not a probability. $F(x_0)$ is a probability, by definition. Areas defined by f(x) are probabilities, and the area $f(x_0) \cdot \Delta x$ is approximately a probability when Δx is small (figure 3.4). These small probabilities are usually expressed as f(x)dx.

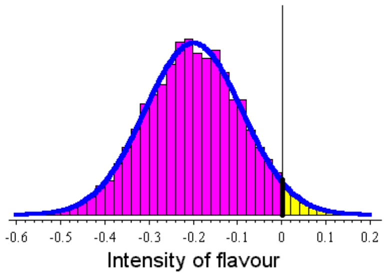

Figure 3.4. Probabilities are areas of f(x), for example the shadowed area of f(x) for all x>0. The small rectangles $f(x)\cdot\Delta x$ are approximate probabilities.

3.4. Features of a density distribution

3.4.1. *Mean*

The *expectation* or *mean* of a density function is

$$E(x) = \int_{-\infty}^{\infty} x \cdot f(x)dx$$

For example, if

$y = x^2$

$$E(y) = \int_{-\infty}^{\infty} y \cdot f(y)dy = \int_{-\infty}^{\infty} x^2 \cdot f(x)dx$$

3.4.2. *Median*

The median is the value dividing the density distribution in two parts ach one with a 50% of probability

3.4.3. *Mode*

The mode is the maximum of the density function. The mode it is the most probable value in the discrete case, and in the continuous case is the value around which the probability is maximum (i.e.: the value of x for which $f(x) \cdot \Delta x$ is maximum).

3.4.4. *Credibility intervals*

A credibility interval of a given probability, for example a 90%, is the interval [a, b] containing a probability of 90%. Any values ‘a’ and ‘b’ for which

$$\int_a^b f(x)dx = 0.90$$

constitute a credibility interval [a, b] at 90%. Observe that there are infinite credibility intervals at 90% of probability, but they have different length (see figure 2.5). One of them is the shortest one, and in Bayesian inference, when the density function used is the posterior density, it is called the Highest Posterior Density interval at 90% ($HPD_{90\%}$).

3.5. Conditional distribution

3.5.1. *Definition*

We say that the conditional distribution of x given $y=y_0$ is

$$f(x \mid y=y_0) = \frac{f(x, y=y_0)}{f(y=y_0)}$$

Following our notation, in which the variables are in red and the constants are in black, we can express the same formula as

$$f(x \mid y) = \frac{f(x, y)}{f(y)}$$

If we consider that y can take several values, the formula can be expressed as

$$f(x \mid y) = \frac{f(x, y)}{f(y)}$$

and in this case it represents a family of density functions, with as different density function for each value of y = y_0, y_1, y_2, …

For two given values 'x' and 'y', the formula is

$$f(x \mid y) = \frac{f(x, y)}{f(y)}$$

3.5.2. *Bayes theorem*

Although f(x) is not a probability, we found that f(x)·Δx is indeed a probability (see fig. 3.4), thus applying Bayes theorem, we have

$$P(A \mid B) = \frac{P(B \mid A) \cdot P(A)}{P(B)}$$

$$f(x \mid y)\Delta x = \frac{f(y \mid x)\Delta y \cdot f(x)\Delta x}{f(y)\Delta y} \longrightarrow f(x \mid y) = \frac{f(y \mid x) \cdot f(x)}{f(y)}$$

thus we have now a version of Bayes theorem for density functions. Considering x as the variable and 'y' as a given (constant) value,

$$f(x \mid y) = \frac{f(y \mid x) \cdot f(x)}{f(y)}$$

which can be expressed proportionally, since f(y) is a constant,

$$f(x \mid y) \propto f(y \mid x) \cdot f(x)$$

For example, if we have a normal distribution y ~ N(μ,σ^2) in which we do not know the parameters μ, σ^2, the uncertainty about both parameters can be expressed as

$$f(\mu \mid \sigma^2, y) \propto f(y \mid \mu, \sigma^2) f(\mu)$$

$$f(\sigma^2 \mid \mu, y) \propto f(y \mid \sigma^2, \mu,) f(\sigma^2)$$

3.5.3. *Conditional distribution of the sample of a Normal distribution*

Let us consider now a random sample **y** from a normal distribution

$$f(\mathbf{y} \mid \mu, \sigma^2) = f(y_1, y_2 \ldots y_n \mid \mu, \sigma^2) = f(y_1 \mid \mu, \sigma^2) \cdot f(y_2 \mid \mu, \sigma^2) \cdot \ldots \cdot f(y_n \mid \mu, \sigma^2) =$$

$$= \prod_1^n f(y_i \mid \mu, \sigma^2) = \prod_1^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(y_i - \mu)^2}{2\sigma^2}\right] = \frac{1}{\left(\sqrt{2\pi}\right)^n \left(\sigma^2\right)^{\frac{n}{2}}} \exp\left[-\frac{\sum_1^n (y_i - \mu)^2}{2\sigma^2}\right]$$

this is the conditional distribution of the data because it is conditioned to the given values of 'μ' and 'σ^2'; i.e., for each given value of the mean and the variance we have a different distribution. For example; for a given mean μ=5 and variance σ^2=2, and for a sample of three elements **y'** = [y_1, y_2, y_3] we have

$$f(\mathbf{y} \mid \mu, \sigma^2) = \frac{1}{\left(\sqrt{2\pi}\right)^3 (2)^{\frac{3}{2}}} \exp\left[-\frac{(y_1 - 5)^2 + (y_2 - 5)^2 + (y_3 - 5)^2}{2 \cdot 2}\right]$$

This is a trivariate multinormal distribution, with three variables y_1, y_2 and y_3.

3.5.4. *Conditional distribution of the variance of a Normal distribution*

We can also write the conditional distribution of the variance for a given sample '**y**' and a given mean 'μ'. We do not know f(σ^2 | μ, **y**), but we can apply Bayes theorem and we obtain

$$f(\sigma^2 \mid \mu, \mathbf{y}) \propto f(\mathbf{y} \mid \sigma^2, \mu) f(\sigma^2)$$

applying the principle of indifference (see 2.1.3), we will consider in this example that *a priori* all values of the variance have the same probability density; i.e., f(σ^2)=constant. This leads to

$$f\left(\sigma^2 \mid \mu, \mathbf{y}\right) \propto f\left(\mathbf{y} \mid \sigma^2, \mu\right)$$

but we know the distribution of the data, that we have assumed to be Normal, thus we can write the conditional distribution of the variance,

$$f\left(\sigma^2 \mid \mu, \mathbf{y}\right) \propto f\left(\mathbf{y} \mid \sigma^2, \mu\right) = \frac{1}{\left(\sqrt{2\pi}\right)^n \left(\sigma^2\right)^{\frac{n}{2}}} \exp\left[-\frac{\sum_1^n (y_i - \mu)^2}{2\sigma^2}\right]$$

Notice that the variable is coloured in red, thus here the variance is the variable and the sample and the mean are given constants. For example, if the mean and the sample are

μ = 1
y' = [2, 3, 4]

for this given mean and this given sample, the conditional distribution of the variance is

$$f\left(\sigma^2 \mid \mu, \mathbf{y}\right) \propto \frac{1}{\left(\sigma^2\right)^{\frac{3}{2}}} \exp\left[-\frac{(2-1)^2 + (3-1)^2 + (4-1)^2}{2\sigma^2}\right] = \frac{1}{\left(\sigma^2\right)^{\frac{3}{2}}} \exp\left[-\frac{7}{\sigma^2}\right]$$

Notice that this is not a Normal distribution. A Normal distribution does not have the variable (in red) there; a Normal distribution looks like

$$f(x) = \frac{1}{\left(\sqrt{2\pi}\right)(10)^{\frac{1}{2}}} \exp\left[-\frac{(x-5)^2}{2 \cdot 10}\right]$$

where 5 and 10 are the mean and the variance of the variable x.

Which is, then, the conditional distribution of the variance? There is a type of distribution called “inverted gamma” that looks like

$$f(x \mid \alpha, \beta) \propto \frac{1}{x^{\alpha+1}} \exp\left[-\frac{\beta}{x}\right]$$

where ‘α’ and ‘β’ are parameters that determine the shape of the function. In figure 2.2 we find three different shapes we can obtain by given different values to ‘α’ and ‘β’, a flat line and two curves, one of them sharper than the other one. We can see that the conditional distribution of the variance is of the type “inverted gamma”. If we take

$$\beta = 7$$

$$\alpha + 1 = \frac{3}{2}$$

we obtain an “inverted “gamma”, and in general the variance of a Normal distribution is an inverted gamma ([1]) with parameters

$$\beta = \frac{1}{2}\sum_{1}^{n}\left(y_i - \mu\right)^2$$

$$\alpha = \frac{n}{2} - 1$$

3.5.5. *Conditional distribution of the mean of a Normal distribution*

We will write now the conditional distribution of the mean for a given sample ‘**y**’ and a given variance ‘σ^2’. We do not know f(μ |σ^2, **y**), but we can apply Bayes theorem and we obtain

$$f\left(\mu \mid \sigma^2, \mathbf{y}\right) \propto f\left(\mathbf{y} \mid \mu, \sigma^2\right) f\left(\mu\right)$$

Applying the principle of indifference (see 2.1.3), we will consider in this example that *a priori* all values of the variance have the same probability density; i.e., f(μ)=constant. This leads to

$$f\left(\mu \mid \sigma^2, \mathbf{y}\right) \propto f\left(\mathbf{y} \mid \mu, \sigma^2\right)$$

but we know the distribution of the data, that we have assumed to be Normal, thus we can write the conditional distribution of the mean,

$$f\left(\mu \mid \sigma^2, \mathbf{y}\right) \propto \frac{1}{\left(\sigma^2\right)^{\frac{n}{2}}} \exp\left[-\frac{\sum_{1}^{n}\left(y_i - \mu\right)^2}{2\sigma^2}\right]$$

Notice that the variable is coloured in red, thus here the mean is the variable and the sample and the variance are given constants. For example, if the variance and the sample are

σ^2 = 9
y' = [2, 3, 4]

for this given variance and this given sample, the conditional distribution of the mean is

[1] This type of inverted gamma is usually named “inverted chi-square”

$$f\left(\mu \mid \sigma^2,\mathbf{y}\right) \propto \frac{1}{(9)^{\frac{3}{2}}} \exp\left[-\frac{(2-\mu)^2+(3-\mu)^2+(4-\mu)^2}{2\cdot 9}\right] = \frac{1}{27}\exp\left[-\frac{3\mu^2-18\mu+32}{18}\right]$$

This can be transformed in a Normal distribution easily

$$f\left(\mu \mid \sigma^2,\mathbf{y}\right) \propto \frac{1}{27}\exp\left[-\frac{\mu^2-6\mu+\frac{32}{3}}{\frac{18}{3}}\right] = \frac{1}{27}\exp\left[-\frac{\mu^2-2\cdot 3\mu+9-9+\frac{32}{3}}{\frac{18}{3}}\right] =$$

$$= \frac{1}{27}\exp\left[-\frac{-9+\frac{32}{3}}{\frac{18}{3}}\right]\exp\left[-\frac{1}{2}\frac{(\mu-3)^2}{\frac{18}{2\cdot 3}}\right] \propto \frac{1}{\sqrt{2\pi}\cdot\sqrt{2}}\exp\left[-\frac{1}{2}\frac{(\mu-3)^2}{2}\right]$$

which is a Normal distribution with mean 3 and variance 2. In a general form, we have (Appendix 3.2)

$$f\left(\mu \mid \sigma^2,\mathbf{y}\right) \propto \frac{1}{\left(\sigma^2\right)^{\frac{n}{2}}}\exp\left[-\frac{\sum_1^n (y_i-\mu)^2}{2\sigma^2}\right] \propto \frac{1}{\sqrt{2\pi}\left(\frac{\sigma^2}{n}\right)^{\frac{1}{2}}}\exp\left[-\frac{(\mu-\bar{y})^2}{2\frac{\sigma^2}{n}}\right]$$

which is a Normal distribution with mean the sample mean and variance the given variance divided by the sample size.

3.6. Marginal distribution

3.6.1. *Definition*

We saw in 2.2.3 the advantages of marginalisation. When we have a bivariate density f(x, y), a marginal density f(x) is

$$f(x) = \int_{-\infty}^{\infty} f(x,y)\,dy = \int_{-\infty}^{\infty} f(x \mid y) f(y)\,dy$$

The marginal density f(x) takes the *average* of all values of 'y' for each x; i.e., takes all values of 'y', multiplies them by their probability and sums.

A density can be marginal for a variable and conditional for another one. For example, a marginal density conditioned in 'z' is

$$f(x \mid z) = \int_{-\infty}^{\infty} f(x,y \mid z)\,dy = \int_{-\infty}^{\infty} f(x \mid y,z) f(y \mid z)\,dy$$

where 'y' has been marginalised and 'z' is conditioning the values of 'x'. Notice that the marginalised variable 'y' does not appear in the formula because for each value of x all its possible values have been considered, multiplied by their respective probability and summed up, thus we do not need to give a value to 'y' in order to obtain f(x|z). However, the conditional variable appears in the formula because for each given value of 'z' we obtain a different value of f(x|z). We will see an example in next paragraph.

3.6.2. *Marginal distribution of the variance of a normal distribution*

The marginal density of the variance *conditioned to the data* is

$$f\left(\sigma^2 \mid \mathbf{y}\right) = \int_{-\infty}^{\infty} f\left(\mu, \sigma^2 \mid \mathbf{y}\right) d\mu$$

Here the mean is *marginalised* and the data are *conditioning* the values of the variance, which means that we will obtain a different distribution for each sample. We will call this distribution "the marginal distribution of the variance" as a short name, because in Bayesian inference it is implicit that we are always conditioning in the data. Bayesian inference is always based in the sample, not in conceptual repetitions of the experiment; the sample is always 'given'.

We do not know $f\left(\mu, \sigma^2 \mid \mathbf{y}\right)$, but we can find it applying Bayes theorem because we know the distribution of the data $f\left(\mathbf{y} \mid \mu, \sigma^2\right)$. If the prior information $f\left(\mu, \sigma^2\right)$ is constant because we apply the principle of indifference as before, we have

$$f\left(\mu, \sigma^2 \mid \mathbf{y}\right) = \frac{f\left(\mathbf{y} \mid \mu, \sigma^2\right) f\left(\mu, \sigma^2\right)}{f(\mathbf{y})} \propto f\left(\mathbf{y} \mid \mu, \sigma^2\right) f\left(\mu, \sigma^2\right) \propto f\left(\mathbf{y} \mid \mu, \sigma^2\right)$$

and the marginal density is

$$f\left(\sigma^2 \mid \mathbf{y}\right) = \int_{-\infty}^{\infty} f\left(\mu, \sigma^2 \mid \mathbf{y}\right) d\mu \propto \int_{-\infty}^{\infty} f\left(\mathbf{y} \mid \mu, \sigma^2\right) d\mu = \int_{-\infty}^{\infty} \frac{1}{(2\pi)^{\frac{n}{2}} \left(\sigma^2\right)^{\frac{n}{2}}} \exp\left[-\frac{\sum_1^n (y_i - \mu)^2}{2\sigma^2} \right] d\mu$$

We solved this integral in Appendix 3.3, and we know that the solution is

$$f\left(\sigma^2 \mid \mathbf{y}\right) \propto \frac{1}{\left(\sigma^2\right)^{\frac{n-1}{2}}} \exp\left[-\frac{\sum_1^n (y_i - \bar{y})^2}{2\,\sigma^2} \right]$$

which is an Inverted Gamma distribution, as in 3.4.4, with parameters α, β

$$\beta = \frac{1}{2}\sum_{1}^{n}\left(y_i - \bar{y}\right)^2$$

$$\alpha = \frac{n-1}{2} - 1 = \frac{n-3}{2}$$

For example, if we take the same sample as in 3.4.4

y' = [2, 3, 4]

for this given sample, the marginal distribution of the variance is

$$f\left(\sigma^2 \mid \mathbf{y}\right) \propto \frac{1}{\left(\sigma^2\right)^{\frac{3-1}{2}}} \exp\left[-\frac{\left(2-\frac{2+3+4}{3}\right)^2 + \left(3-\frac{2+3+4}{3}\right)^2 + \left(4-\frac{2+3+4}{3}\right)^2}{2\sigma^2}\right] = \frac{1}{\sigma^2}\exp\left[-\frac{1}{\sigma^2}\right]$$

Notice that the mean does not appear in the formula. In 3.5.4, when we calculated the density of the variance conditioned to the mean and to the data we had to give a value for the mean and we had to give the data. Here we only should give the data because the mean has been *marginalised.*

3.6.3. *Marginal distribution of the mean of a Normal distribution*

The marginal density of the mean *conditioned to the data* is

$$f\left(\mu \mid \mathbf{y}\right) = \int_0^{\infty} f\left(\mu, \sigma^2 \mid \mathbf{y}\right) d\sigma^2$$

Here the variance is *marginalised* and the data are *conditioning* the values of the variance, which means that we will obtain a different distribution for each sample. We will call this distribution "the marginal distribution of the mean" as a short name, because, as we said before, in Bayesian inference it is implicit that we are always conditioning in the data.

We do not know the function $f\left(\mu, \sigma^2 \mid \mathbf{y}\right)$, but applying Bayes theorem we can find it, because we now the distribution of the data, $f\left(\mathbf{y} \mid \mu, \sigma^2\right)$.

$$f\left(\mu, \sigma^2 \mid \mathbf{y}\right) = \frac{f\left(\mathbf{y} \mid \mu, \sigma^2\right) f\left(\mu, \sigma^2\right)}{f\left(\mathbf{y}\right)} \propto f\left(\mathbf{y} \mid \mu, \sigma^2\right) f\left(\mu, \sigma^2\right)$$

if we admit the indifference principle to show vague prior information, then $f(\mu,\sigma^2)$ is a constant, thus

$$f\left(\mu, \sigma^2 \mid \mathbf{y}\right) \propto f\left(\mathbf{y} \mid \mu, \sigma^2\right)$$

and the marginal density of the mean is

$$f(\mu\,|\,\mathbf{y}) \propto \int f(\mathbf{y}\,|\,\mu,\sigma^2)\,d\sigma^2 = \int_0^\infty \frac{1}{(2\pi)^{\frac{n}{2}}(\sigma^2)^{\frac{n}{2}}} \exp\left[-\frac{\sum_1^n (y_i-\mu)^2}{2\sigma^2}\right] d\sigma^2$$

This integral is solved in Appendix 3.4, and the result is

$$f(\mu\,|\,\mathbf{y}) \propto \left[1+\frac{n}{n-1}\cdot\frac{(\mu-\bar{y})^2}{s^2}\right]^{-\frac{n-2}{2}}$$

where

$$\bar{y} = \frac{1}{n}\sum_1^n y_i \quad ; \quad s^2 = \frac{1}{n-1}\sum_1^n (y_i-\bar{y})^2$$

This is a Student t-distribution with n–1 degrees of freedom, having a mean which is the sample mean and a variance that is the sample quasi-variance, thus

$$f(\mu\,|\,\mathbf{y}) \propto t_{n-1}(\bar{y}, s^2)$$

For example, if **y’** = [2, 3, 4]

$$\bar{y} = \frac{1}{3}(2+3+4) = 3$$

$$s^2 = \frac{1}{3-1}\left[(2-3)^2+(3-3)^2+(4-3)^2\right] = 1$$

$$f(\mu\,|\,\mathbf{y}) \propto \left[1+\frac{3}{2-1}\cdot\frac{(\mu-3)^2}{1}\right]^{-\frac{3}{2}}$$

Notice that the variance does not appear in the formula. In 3.5.5, when we calculated the density of the mean conditioned to the variance and to the data, we had to give a value for the variance and we had to give the data. Here we only should give the data because the variance has been *marginalised.*