Chapter 2

Inference for a normal population

This chapter shows how to make inferences for the mean and variance of a normal population using a conjugate prior distribution. First we need the multi-parameter version of Bayes Theorem.

2.1 Bayes Theorem for many parameters

Suppose that now the probability (density) function we used to describe the data depends on many parameters, that is, $f(\boldsymbol{x}|\boldsymbol{\theta})$ where $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_p)^T$. After observing the data, the likelihood function for $\boldsymbol{\theta}$ is $f(\boldsymbol{x}|\boldsymbol{\theta})$. Prior beliefs about $\boldsymbol{\theta}$ are represented through a probability (density) function $\pi(\boldsymbol{\theta})$. Therefore, using Bayes Theorem, the posterior probability (density) function for $\boldsymbol{\theta}$ is

$$\pi(\boldsymbol{\theta}|\boldsymbol{x}) = \frac{\pi(\boldsymbol{\theta})\, f(\boldsymbol{x}|\boldsymbol{\theta})}{f(\boldsymbol{x})}$$

where

$$f(\boldsymbol{x}) = \begin{cases} \int_\Theta \pi(\boldsymbol{\theta})\, f(\boldsymbol{x}|\boldsymbol{\theta})\, d\boldsymbol{\theta} & \text{if } \boldsymbol{\theta} \text{ is continuous,} \\ \\ \sum_\Theta \pi(\boldsymbol{\theta})\, f(\boldsymbol{x}|\boldsymbol{\theta}) & \text{if } \boldsymbol{\theta} \text{ is discrete.} \end{cases}$$

As in Chapter 1, this can be rewritten as

$$\begin{aligned} \pi(\boldsymbol{\theta}|\boldsymbol{x}) &\propto \pi(\boldsymbol{\theta}) \times f(\boldsymbol{x}|\boldsymbol{\theta}) \\ i.e. \quad \text{posterior} &\propto \text{prior} \times \text{likelihood.} \end{aligned}$$

Next we introduce a new distribution which will be useful later on.

Example 2.1

If X has a generalised $t_a(b, c)$ distribution (see page 101) then show that $Y = (X - b)/\sqrt{c} \sim t_a \equiv t_a(0, 1)$.

Recall the general result: if X is a random variable with probability density function $f_X(x)$ and g is a bijective (1–1) function then the random variable $Y = g(X)$ has probability density function

$$f_Y(y) = f_X\left\{g^{-1}(y)\right\}\left|\frac{d}{dy}\,g^{-1}(y)\right|. \qquad (2.1)$$

Solution

Here we take $Y = g(X) = (X - b)/\sqrt{c}$ from which we obtain $X = g^{-1}(Y) = b + \sqrt{c}\,Y$. Therefore using (2.1) we have

$$\begin{aligned}
f_Y(y) &= f_X\left\{g^{-1}(y)\right\}\left|\frac{d}{dy}\,g^{-1}(y)\right| \\
&= f_Y\left(b + \sqrt{c}\,y\right) \times \sqrt{c} \\
&= \frac{\Gamma\left(\frac{a+1}{2}\right)}{\sqrt{ac\pi}\,\Gamma\left(\frac{a}{2}\right)}\left(1 + \frac{y^2}{a}\right)^{-\frac{a+1}{2}} \times \sqrt{c}, \quad y \in \mathbb{R} \\
&= \frac{\Gamma\left(\frac{a+1}{2}\right)}{\sqrt{a\pi}\,\Gamma\left(\frac{a}{2}\right)}\left(1 + \frac{y^2}{a}\right)^{-\frac{a+1}{2}}, \quad y \in \mathbb{R}.
\end{aligned}$$

This is the t_a density and so $Y = (X - b)/\sqrt{c} \sim t_a$.

Comment

Values for the density function $f_Y(y)$ and the distribution function $F_Y(y)$ can be obtained by using the `R` functions `dgt` and `pgt` in the package `nclbayes`.

It is clear that $t_a(0, 1) \equiv t_a$ by examining their densities. Therefore, it makes sense to think of the t_a distribution as the standard t_a–distribution and make all calculations for the generalised $t_a(b, c)$ distribution from this standard distribution. The relationship between this standard and generalised version of the t-distribution is directly analogous to that between the standard normal $N(0, 1)$ distribution and its more general version: the $N(b, c)$ distribution. In both cases the relationship is one of location and scale:

$$Y \sim N(b, c) \quad \Longrightarrow \quad \frac{Y - b}{\sqrt{c}} \sim N(0, 1)$$

$$Y \sim t_a(b, c) \quad \Longrightarrow \quad \frac{Y - b}{\sqrt{c}} \sim t_a.$$

2.2 Prior to posterior analysis

Suppose we have a random sample from a normal distribution in which both the mean μ and the precision τ are unknown, that is, $X_i|\mu,\tau \sim N(\mu, 1/\tau)$, $i = 1, 2, \ldots, n$ (independent). We shall adopt a (joint) prior distribution for μ and τ for which

$$\mu|\tau \sim N\left(b, \frac{1}{c\tau}\right) \qquad \text{and} \qquad \tau \sim Ga(g, h)$$

for known values b, c, g and h. This distribution has density function

$$\begin{aligned}\pi(\mu,\tau) &= \pi(\mu|\tau)\pi(\tau) \\ &= \left(\frac{c\tau}{2\pi}\right)^{1/2} \exp\left\{-\frac{c\tau}{2}(\mu-b)^2\right\} \times \frac{h^g \tau^{g-1} e^{-h\tau}}{\Gamma(g)}, \quad \mu \in \mathbb{R},\ \tau > 0 \\ &\propto \tau^{g-\frac{1}{2}} \exp\left\{-\frac{\tau}{2}\left[c(\mu-b)^2 + 2h\right]\right\}, \quad \mu \in \mathbb{R},\ \tau > 0. \qquad (2.2)\end{aligned}$$

We will use the notation $NGa(b, c, g, h)$ for this distribution. Thus we take the prior distribution

$$\begin{pmatrix}\mu \\ \tau\end{pmatrix} \sim NGa(b, c, g, h).$$

Determine the posterior distribution for $\begin{pmatrix}\mu \\ \tau\end{pmatrix}$.

Hint:

$$c(\mu-b)^2 + n(\bar{x}-\mu)^2 = (c+n)\left\{\mu - \left(\frac{cb+n\bar{x}}{c+n}\right)\right\}^2 + \frac{nc(\bar{x}-b)^2}{c+n}.$$

Solution

From (1.8), the likelihood function is

$$f(\boldsymbol{x}|\mu,\tau) = \left(\frac{\tau}{2\pi}\right)^{n/2} \exp\left[-\frac{n\tau}{2}\left\{s^2 + (\bar{x}-\mu)^2\right\}\right].$$

Using Bayes Theorem, the posterior density is

$$\pi(\mu,\tau|\boldsymbol{x}) \propto \pi(\mu,\tau)\, f(\boldsymbol{x}|\mu,\tau)$$

and so, for $\mu \in \mathbb{R}$, $\tau > 0$

$$\begin{aligned}\pi(\mu,\tau|\boldsymbol{x}) &\propto \tau^{g-\frac{1}{2}} \exp\left\{-\frac{\tau}{2}\left[c(\mu-b)^2 + 2h\right]\right\} \\ &\qquad\qquad \times \tau^{\frac{n}{2}} \exp\left[-\frac{n\tau}{2}\left\{s^2 + (\bar{x}-\mu)^2\right\}\right] \\ &\propto \tau^{g+\frac{n}{2}-\frac{1}{2}} \exp\left\{-\frac{\tau}{2}\left[c(\mu-b)^2 + n(\bar{x}-\mu)^2 + 2h + ns^2\right]\right\} \\ &\propto \tau^{g+\frac{n}{2}-\frac{1}{2}} \exp\left\{-\frac{\tau}{2}\left[(c+n)\left\{\mu - \left(\frac{cb+n\bar{x}}{c+n}\right)\right\}^2 + \frac{nc(\bar{x}-b)^2}{c+n} + 2h + ns^2\right]\right\}\end{aligned}$$

using the hint. Let

$$B = \frac{bc + n\bar{x}}{c+n}, \qquad C = c + n,$$
$$G = g + \frac{n}{2}, \qquad H = h + \frac{cn(\bar{x}-b)^2}{2(c+n)} + \frac{ns^2}{2}. \tag{2.3}$$

Then the posterior density is

$$\pi(\mu,\tau|\boldsymbol{x}) \propto \tau^{G-\frac{1}{2}} \exp\left\{-\frac{\tau}{2}\left[C(\mu-B)^2 + 2H\right]\right\},$$
$$\mu \in \mathbb{R}, \tau > 0$$

Notice that this posterior density is of the same form as the prior density (2.2). Therefore, we can conclude that the posterior distribution is

$$\begin{pmatrix}\mu\\ \tau\end{pmatrix}\Bigg|\, \boldsymbol{x} \sim NGa(B,C,G,H).$$

Thus, the NGa distribution is conjugate to this data model.

2.2.1 Marginal distributions

Suppose $(\mu,\tau)^T \sim NGa(b,c,g,h)$. From the definition of the NGa distribution we know that $\tau \sim Ga(g,h)$. This also means that $\sigma = 1/\sqrt{\tau} \sim$ *Inv-Chi(g,h)*; see page 101.

The (marginal) density for μ is, for $\mu \in \mathbb{R}$

$$\begin{aligned}\pi(\mu) &= \int_0^\infty \pi(\mu,\tau)\,d\tau \\ &\propto \int_0^\infty \tau^{g-\frac{1}{2}} \exp\left\{-\frac{\tau}{2}\left[c(\mu-b)^2+2h\right]\right\}\,d\tau.\end{aligned}$$

Now, as the integral of a gamma density over its entire range is one, we have

$$\int_0^\infty \frac{b^a\theta^{a-1}e^{-b\theta}}{\Gamma(a)}\,d\theta = 1 \quad\Longrightarrow\quad \int_0^\infty \theta^{a-1}e^{-b\theta}\,d\theta = \frac{\Gamma(a)}{b^a}.$$

Therefore, for $\mu \in \mathbb{R}$

$$\begin{aligned}\pi(\mu) &\propto \int_0^\infty \tau^{g+\frac{1}{2}-1} \exp\left\{-\frac{\tau}{2}\left[c(\mu-b)^2+2h\right]\right\}\,d\tau \\ &\propto \frac{\Gamma\left(g+\frac{1}{2}\right)}{[\{c(\mu-b)^2+2h\}/2]^{g+\frac{1}{2}}} \\ &\propto h^{-g-1/2}\left\{1+\frac{c(\mu-b)^2}{2h}\right\}^{-g-1/2} \\ &\propto \left\{1+\frac{c(\mu-b)^2}{2h}\right\}^{-\frac{2g+1}{2}}.\end{aligned}$$

Comparing this density with that of the generalised t–distribution (on page 101) gives

$$\mu \sim t_{2g}\left(b, \frac{h}{gc}\right). \qquad (2.4)$$

Thus, marginally, the prior distribution for μ is a t–distribution.

Similar calculations can be used to determine the (marginal) posterior distributions.

Summary of marginal distributions

The prior $\begin{pmatrix}\mu\\ \tau\end{pmatrix} \sim NGa(b, c, g, h)$ has marginal distributions

- $\mu \sim t_{2g}\left(b, \frac{h}{gc}\right)$
- $\tau \sim Ga(g, h)$

Also $\sigma = 1/\sqrt{\tau} \sim \textit{Inv-Chi}(g, h)$.

The posterior $\left.\begin{pmatrix}\mu\\ \tau\end{pmatrix}\right| \boldsymbol{x} \sim NGa(B, C, G, H)$ has marginal distributions

- $\mu|\boldsymbol{x} \sim t_{2G}\left(B, \frac{H}{GC}\right)$
- $\tau|\boldsymbol{x} \sim Ga(G, H)$

Also $\sigma|\boldsymbol{x} \sim \textit{Inv-Chi}(G, H)$.

It can be shown that the posterior mean of μ is greater than its prior mean if and only if the sample mean (likelihood mode) is greater than its prior mean, that is,

$$E(\mu|\boldsymbol{x}) > E(\mu) \quad \Longleftrightarrow \quad \bar{x} > b.$$

The relationships between the prior and posterior variance of μ and mean and variance of τ and of σ are rather more complex.

Example 2.2

Recall Example 1.4 on the earth's density. Previously we assumed that the measurements followed a $N(\mu, 0.2^2)$ distribution, that is, the standard deviation of the measurements was known to be $0.2\,g/cm^3$. Now we consider the case where this standard deviation is unknown and determine posterior distributions using the theory in section 2.2.

Before we can proceed, we must specify the parameters in the $NGa(b, c, g, h)$ prior distribution for (μ, τ). In the previous analysis, we assumed that the population measurement precision was $\tau = 1/0.2^2 = 25$ and assumed a $N(5.41, 0.4^2)$ prior distribution for the population mean, that is, $\mu|\tau = 25 \sim N(5.41, 0.4^2)$.

Choice of b and c: the conditional prior distribution for μ is $\mu|\tau \sim N\{b, 1/(c\tau)\}$ and so matching the prior distributions for μ (when $\tau = 25$) gives $b = 5.41$ and $c = 0.25$.

Choice of g and h: the marginal prior distribution for τ is $\tau \sim Ga(g, h)$. Previously, we assumed $\tau = 25$ (with $Var(\tau) = 0$) and so take this value as the prior mean: $E(\tau) = 25$. Suppose we also decide that $Var(\tau) = 250$. These two requirements give $g = 2.5$ and $h = 0.1$. Therefore, we will assume the prior distribution

$$\begin{pmatrix} \mu \\ \tau \end{pmatrix} \sim NGa(5.41, 0.25, 2.5, 0.1).$$

We have seen that if $(\mu, \tau)^T \sim NGa(b, c, g, h)$ then the marginal distribution of μ is $\mu \sim t_{2g}\{b, h/(gc)\}$. Therefore, with this choice of prior distribution, the marginal prior distribution for μ is

$$\mu \sim t_5(5.41, 0.16).$$

Figure 2.1 shows the close match between the new (marginal) prior distribution for μ and that used previously.

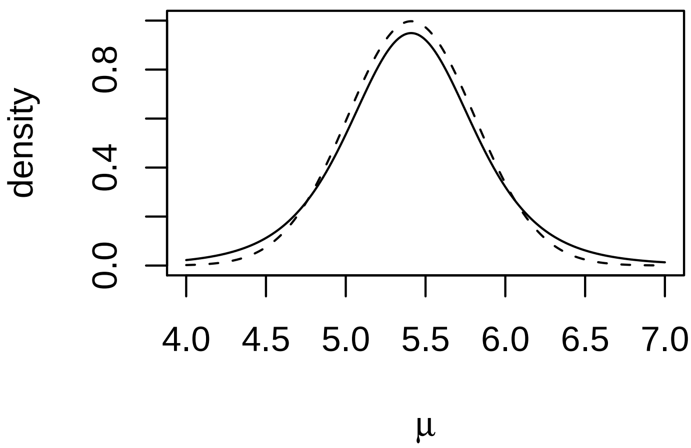

Figure 2.1: Marginal prior density for μ: new version (solid) and previous version (dashed)

Determine the posterior distribution for $(\mu, \tau)^T$. Also determine the marginal prior distribution for τ and for σ, and the marginal posterior distribution for each of μ, τ and σ.

Solution

We can combine the information in the $NGa(5.41, 0.25, 2.5, 0.1)$ prior distribution for $(\mu, \tau)^T$ with that in the data ($n = 23$, $\bar{x} = 5.4848$, $s = 0.1882$) using the results in section 2.2 to obtain a $NGa(B,C,G,H)$ posterior distribution, where

$$B = \frac{bc+n\bar{x}}{c+n} = \frac{(5.41\times 0.25)+(23\times 5.4848)}{23.25} = 5.4840,$$
$$C = c+n = 23.25,$$
$$G = g+\frac{n}{2} = 14,$$
$$H = h+\frac{cn(\bar{x}-b)^2}{2(c+n)}+\frac{ns^2}{2} = 0.1+\frac{5.75}{46.5}(5.4848-5.41)^2+11.5\times 0.1882^2 = 0.5080.$$

The marginal prior distributions for τ and σ are

$$\tau \sim Ga(g,h) \equiv Ga(2.5,0.1)$$
$$\sigma \sim \textit{Inv-Chi}(g,h) \equiv \textit{Inv-Chi}(2.5,0.1)$$

Also the marginal posterior distributions for μ, τ and σ are

$$\mu|\boldsymbol{x} \sim t_{2G}\left(B,\frac{H}{GC}\right) \equiv t_{28}(5.4840, 0.001561)$$
$$\tau|\boldsymbol{x} \sim Ga(G,H) \equiv Ga(14,0.5080)$$
$$\sigma|\boldsymbol{x} \sim \textit{Inv-Chi}(G,H) \equiv \textit{Inv-Chi}(14,0.5080)$$

Plots of the (marginal) prior and posterior distributions of μ, τ and σ are given in Figure 2.2. Note that the (marginal) prior and posterior distributions for σ can be determined from that of τ. We can also examine the joint prior and posterior distributions for $(\mu,\tau)^T$ via the contour plots of their densities to see if there is any change in the dependence structure; see Figure 2.3. This figure is produced by using the `R` command `NGacontour` in the `nclbayes` package as follows:

```
mu=seq(4.5,6.5,len=1000)
tau=seq(0,71,len=1000)
NGacontour(mu,tau,b,c,g,h,lty=3)
NGacontour(mu,tau,B,C,G,H,add=TRUE)
```

in which the variables `b,c,g,h,B,C,G,H` have already been set to their prior/posterior values. A careful look at the values of the contour levels plotted shows that the highest

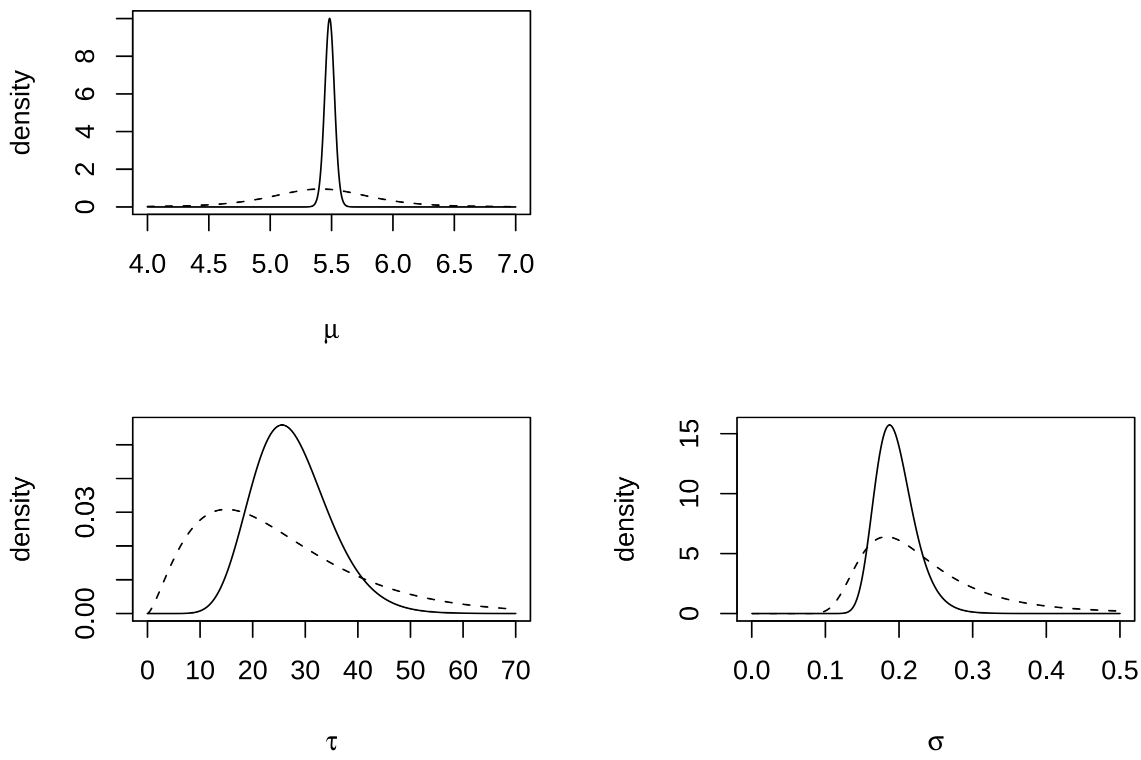

Figure 2.2: Prior (dashed) and posterior (solid) densities for μ, τ and σ

contour level plotted for the prior density is 0.024 and the lowest level for the posterior density is 0.05. From this we can conclude that the posterior distribution is far more concentrated than the prior distribution. Also the contours for the posterior distribution are much more elliptical than those for the prior distribution. This indicates a change in the dependence structure. However, the main changes shown by the figure are in the mean and variability of μ and τ.

Wikipedia tells us that the actual mean density of the earth is $5.515\,g/cm^3$. We can determine the (posterior) probability that the mean density is within 0.1 of this value as follows. We already know that $\mu|\boldsymbol{x} \sim t_{28}(5.484, 0.001561)$ and so we can calculate

$$Pr(5.415 < \mu < 5.615|\boldsymbol{x}) = 0.9529$$

using `pgt(5.615,28,5.484,0.001561)-pgt(5.415,28,5.484,0.001561)`.

Without the data, the only basis for determining the earth's density is via the prior distribution. Here the prior distribution is $\mu \sim t_5(5.41, 0.16)$ and so the (prior) probability that the mean density is within 0.1 of the (now known) true value is

$$Pr(5.415 < \mu < 5.615) = 0.1802,$$

calculated using `pgt(5.615,5,5.41,0.16)-pgt(5.415,5,5.41,0.16)`.

These probability calculations demonstrate that the data have been very informative and changed our beliefs about the earth's density.

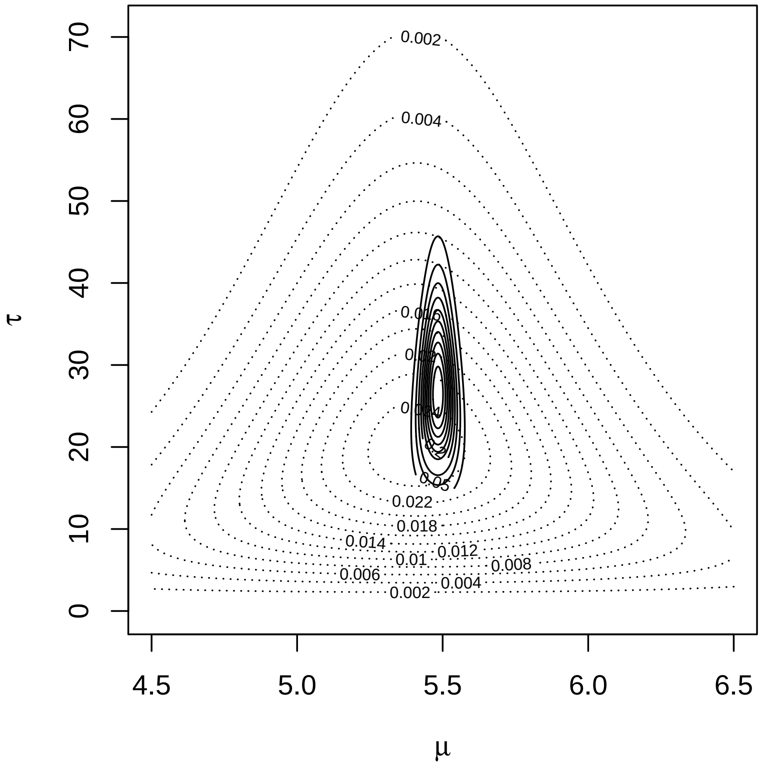

Figure 2.3: Contour plot of the prior (dashed) and posterior (solid) densities for $(\mu, \tau)^T$.

2.3 Confidence intervals and regions

Example 2.3

Determine the $100(1-\alpha)\%$ highest density interval (HDI) for the population mean μ in terms of quantiles of the standard t-distribution.

Solution

The marginal posterior distribution is $\mu|\boldsymbol{x} \sim t_{2G}\left(B, \frac{H}{GC}\right)$. This is a symmetric distribution and so the HDI is an equi-tailed interval. Therefore the HDI (ℓ, u) for μ must satisfy

$$Pr(\mu < \ell|\boldsymbol{x}) = \alpha/2 \qquad \text{and} \qquad Pr(\mu > u|\boldsymbol{x}) = \alpha/2.$$

Now, given the data $\boldsymbol{x}$

$$\frac{\mu - B}{\sqrt{H/(GC)}} \sim t_{2G}$$

and so

$$Pr(\mu > u|\boldsymbol{x}) = \alpha/2 \quad \Rightarrow \quad Pr\left(\frac{\mu - B}{\sqrt{H/(GC)}} > \frac{u - B}{\sqrt{H/(GC)}}\middle|\boldsymbol{x}\right) = \alpha/2$$
$$\Rightarrow \quad \frac{u - B}{\sqrt{H/(GC)}} = t_{2G;\alpha/2}$$

where $t_{2G;p}$ is the upper p point of the t_{2G} distribution. Therefore

$$u = B + t_{2G;\alpha/2}\sqrt{\frac{H}{GC}}.$$

Similar calculations give

$$\ell = B + t_{2G;1-\alpha/2}\sqrt{\frac{H}{GC}} = B - t_{2G;\alpha/2}\sqrt{\frac{H}{GC}}$$

since the t distribution is symmetric about zero. Thus the $100(1-\alpha)\%$ HDI for μ is

$$\left(B - t_{2G;\alpha/2}\sqrt{\frac{H}{GC}},\ B + t_{2G;\alpha/2}\sqrt{\frac{H}{GC}}\right).$$

These intervals can be calculated easily using the `R` function `qgt` in the package `nclbayes`. For example, the prior and posterior 95% HDIs for μ can be calculated using

```
c(qgt(0.025,2*g,b,h/(g*c)),qgt(0.975,2*g,b,h/(g*c)))
c(qgt(0.025,2*G,B,H/(G*C)),qgt(0.975,2*G,B,H/(G*C)))
```

Determining a highest density interval (HDI) for the population precision τ or standard deviation σ is more complicated as their posterior distributions are not symmetric. The (marginal) posterior for τ is $\tau|\boldsymbol{x} \sim Ga(G, H)$ and the (marginal) posterior for σ is $\sigma|\boldsymbol{x} \sim Inv\text{-}Chi(G, H)$. HDIs can be found by using the `R` functions `hdiGamma` and `hdiInvchi` in the package `nclbayes`. More standard equi-tailed confidence intervals can be found using the functions `qgamma` and `qinvchi`.

For example, the prior and posterior 95% HDIs for τ can be calculated using `R` commands `hdiGamma(0.95,g,h)` and `hdiGamma(0.95,G,H)`, and those for σ using commands `hdiInvchi(0.95,g,h)` and `hdiInvchi(0.95,G,H)`. The 95% equi-tailed confidence intervals are calculated in a similar way to the HDIs for μ above. So for τ, the prior and posterior intervals are calculated using

```
c(qgamma(0.025,g,h),qgamma(0.975,g,h))
c(qgamma(0.025,G,H),qgamma(0.975,G,H))
```

and those for σ using

```
c(qinvchi(0.025,g,h),qinvchi(0.975,g,h))
c(qinvchi(0.025,G,H),qinvchi(0.975,G,H))
```

	Prior	Posterior	
μ:	$(4.3818, 6.4382)$	$(5.4031, 5.5649)$	
τ:	$(1.4812, 55.9573)$	$(14.0193, 42.2530)$	$\leftarrow$ HDI
	$(4.1561, 64.1625)$	$(15.0674, 43.7625)$	
σ:	$(0.1062, 0.4246)$	$(0.1466, 0.2505)$	$\leftarrow$ HDI
	$(0.1248, 0.4905)$	$(0.1512, 0.2576)$	

Table 2.1: Prior and posterior 95% intervals for the analysis in Example 2.2

The numerical values for the prior and posterior 95% intervals for the analysis in Example 2.2 are given in Table 2.1. Notice that there is little difference between the posterior HDI and equi-tailed intervals for τ and for σ, whereas the prior intervals are fairly different. This is because the prior distributions are quite skewed but the posterior distributions are fairly symmetric; see Figure 2.2.

In Bayesian inference it can also be useful to determine (joint) confidence regions for several parameters, in this case, for $(\mu, \tau)^T$. In general this is a difficult problem to solve mathematically, and it is in this case.

Example 2.4

Determine a joint confidence region for $(\mu, \tau)^T$.

Solution

We know that the (joint) prior distribution for these parameters is

$$\begin{pmatrix}\mu\\ \tau\end{pmatrix} \sim NGa(b, c, g, h).$$

Therefore an HDI–type confidence region takes the form

$$\begin{aligned}
&\left\{\begin{pmatrix}\mu\\ \tau\end{pmatrix} : \ \pi(\mu,\tau) > k\right\}\\
&= \left\{\begin{pmatrix}\mu\\ \tau\end{pmatrix} : \ \tau^{g-\frac{1}{2}} \exp\left\{-\frac{\tau}{2}\left[c(\mu-b)^2 + 2h\right]\right\} > k'\right\}\\
&= \left\{\begin{pmatrix}\mu\\ \tau\end{pmatrix} : \ \left(g-\frac{1}{2}\right)\log\tau - \frac{\tau}{2}\left[c(\mu-b)^2 + 2h\right] > k''\right\}\\
&= \left\{\begin{pmatrix}\mu\\ \tau\end{pmatrix} : \ \frac{\tau c(\mu-b)^2}{2} + h\tau - \left(g-\frac{1}{2}\right)\log\tau < k_\alpha\right\}
\end{aligned}$$

where k_α will depend on the confidence level of the region. These regions are not difficult to draw. The difficult part is determining the appropriate value for k_α to get say a 95% confidence region. If we could determine the distribution of

$$Y = \frac{\tau c(\mu-b)^2}{2} + h\tau - \left(g-\frac{1}{2}\right)\log\tau$$

when

$$\begin{pmatrix}\mu\\ \tau\end{pmatrix} \sim NGa(b, c, g, h)$$

then we could get the value for k_α. Unfortunately it is quite difficult to do this mathematically. However, we can use simulation methods to get a pretty accurate value for k_α (for a given confidence level).

Using an additional argument in the `R` function `NGacontour` produces plots of confidence regions. For example

```
mu=seq(3.5,7.5,len=1000)
tau=seq(0,80,len=1000)
NGacontour(mu,tau,b,c,g,h,p=c(0.95,0.9,0.8),lty=3)
NGacontour(mu,tau,B,C,G,H,p=c(0.95,0.9,0.8),add=TRUE)
```

produces a plot containing the 95%, 90% and 80% prior and posterior confidence regions for $(\mu,\tau)^T$ for the prior and posterior distributions in Example 2.2; see Figure 2.4. The upper plot shows contours of both prior and posterior densities. The numbers within the plot are the contour levels. The largest prior confidence region is the 95% region. The next largest is the 90% prior confidence region and the smallest is the 80% prior confidence region. The same ordering holds for the posterior confidence regions. The posterior contours are so concentrated in the middle of the plot that there is no room to put in the contour levels. However, these can be see on the lower plot which also shows the contours but focuses the parameter range to highlight the contours of the posterior density. The values of the contours in this lower plot show that the posterior density is much more peaked, that is, the posterior has a much reduced variability. The location of the centre of the central contour for both the prior and posterior densities shows that there has been little change in the mean/mode.

2.4 Predictive distribution

Suppose we sample another value y randomly from the population. What values is it likely to take? This is described by its predictive distribution. We can determine this distribution by using the definition of the predictive density

$$f(y|\boldsymbol{x}) = \int f(y|\mu,\tau)\,\pi(\mu,\tau|\boldsymbol{x})\,d\mu\,d\tau$$

or by using Candidate's formula (as this is a conjugate analysis). However, for this model/prior, there is a more straightforward method to determine the predictive distribution in this model.

As Y is a random value from the population, we have that $Y|\mu,\tau \sim N(\mu,1/\tau)$. We also know that the posterior distribution is $(\mu,\tau)^T|\boldsymbol{x} \sim NGa(B,C,G,H)$. Therefore, we can write

$$Y = \mu + \varepsilon,$$

where

$$\varepsilon|\tau \sim N(0,1/\tau) \qquad \text{and} \qquad \mu|\boldsymbol{x},\tau \sim N\left(B,\frac{1}{C\tau}\right).$$

Hence Y is the sum of two independent normal random quantities, and so

$$Y|\boldsymbol{x},\tau \sim N\left(B,\frac{1}{\tau}+\frac{1}{C\tau}\right) \equiv N\left(B,\frac{C+1}{C\tau}\right).$$

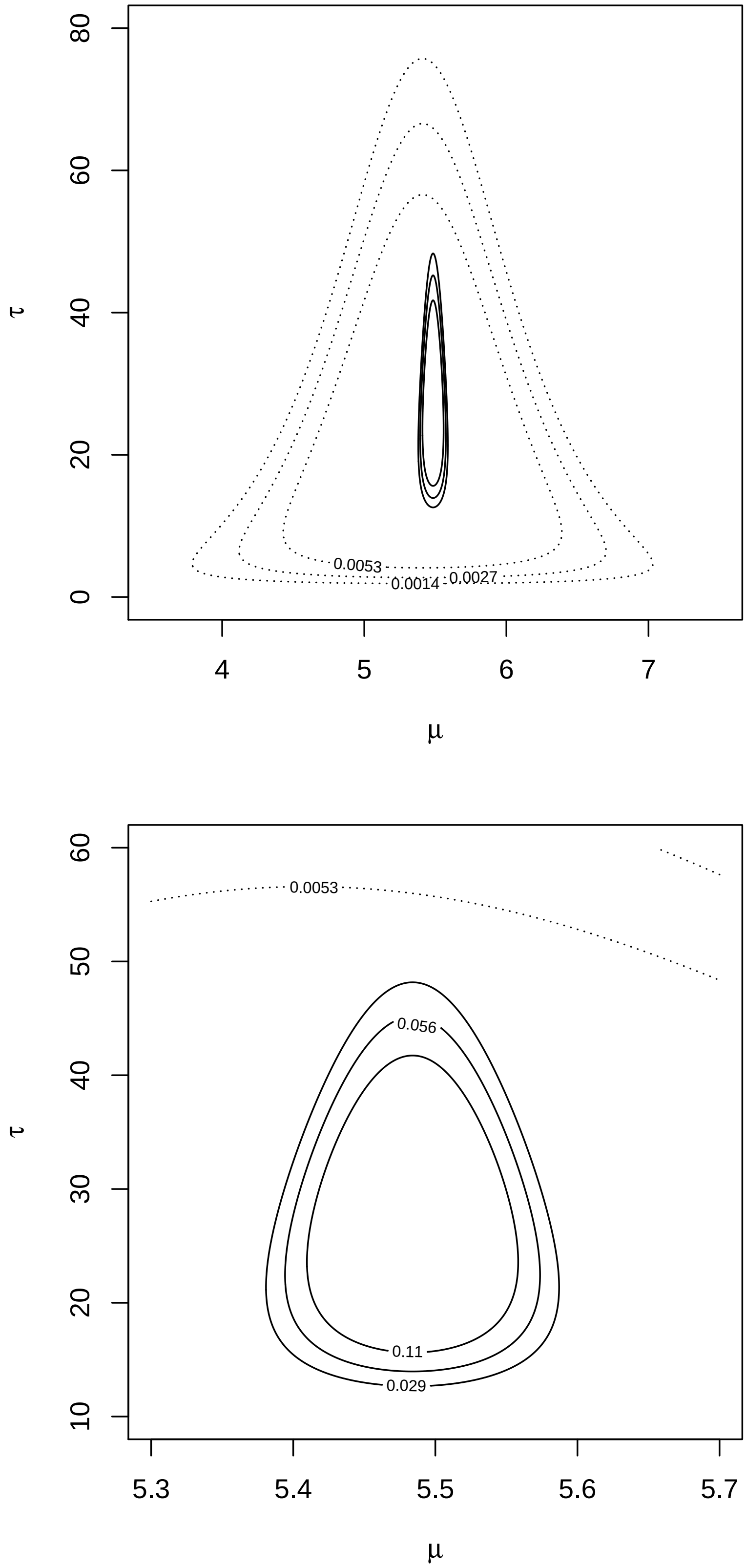

Figure 2.4: 95%, 90% and 80% prior (dashed) and posterior (solid) confidence regions for $(\mu, \tau)^T$

Thus, as $\tau|\boldsymbol{x} \sim Ga(G, H)$

$$\left.\begin{pmatrix} Y \\ \tau \end{pmatrix}\right| \boldsymbol{x} \sim NGa\left(B, \frac{C}{C+1}, G, H\right)$$

and so, using (2.4)

$$Y|\boldsymbol{x} \sim t_{2G}\left\{B, \frac{H(C+1)}{GC}\right\}.$$

We can determine $100(1-\alpha)\%$ predictive intervals by noting that the predictive distribution is symmetric about its mean and therefore the HDI is

$$\left(B - t_{2G;\alpha/2}\sqrt{\frac{H(C+1)}{GC}},\ B + t_{2G;\alpha/2}\sqrt{\frac{H(C+1)}{GC}}\right).$$

These predictive intervals can be calculated easily using the R function qgt. For example, in Example 2.2, the prior and posterior predictive HDIs for a new value Y from the population are $(4.2604, 6.5596)$ and $(5.0855, 5.8825)$ respectively, calculated using

```
c(qgt(0.025,2*g,b,h*(c+1)/(g*c)),qgt(0.975,2*g,b,h*(c+1)/(g*c)))
c(qgt(0.025,2*G,B,H*(C+1)/(G*C)),qgt(0.975,2*G,B,H*(C+1)/(G*C)))
```

2.5 Summary

Suppose we have a normal random sample with $X_i|\mu,\tau \sim N(\mu, 1/\tau)$, $i = 1, 2, \dots, n$ (independent).

(i) $(\mu,\tau)^T \sim NGa(b, c, g, h)$ is a conjugate prior distribution.

(ii) The posterior distribution is $(\mu,\tau)^T|\boldsymbol{x} \sim NGa(B, C, G, H)$ where the posterior parameters are given by (2.3).

(iii) The marginal prior distributions are $\mu \sim t_{2g}\{b, h/(gc)\}$, $\tau \sim Ga(g, h)$, $\sigma = 1/\sqrt{\tau} \sim$ *Inv-Chi*(g, h).

(iv) The marginal posterior distributions are $\mu|\boldsymbol{x} \sim t_{2G}\{B, H/(GC)\}$, $\tau|\boldsymbol{x} \sim Ga(G, H)$, $\sigma|\boldsymbol{x} \sim$ *Inv-Chi*(G, H).

(v) Prior and posterior means and standard deviations for μ, τ and σ can be calculated from the properties of the t, *Gamma* and *Inv-Chi* distributions.

(vi) Prior and posterior probabilities and densities for μ, τ and σ can be calculated using the R functions pgt, dgt, pgamma, dgamma, pinvchi, dinvchi.

(vii) HDIs or equi-tailed CIs for μ, τ and σ can be calculated using qgt, hdiGamma, hdiInvchi, qgamma, qinvchi.

(viii) Contour plots of the prior and posterior densities for $(\mu,\tau)^T$ can be plotted using the NGacontour function.

(ix) Prior and posterior confidence regions for $(\mu,\tau)^T$ can be plotted using the NGacontour function.

(x) The predictive distribution for a new observation Y from the population is $Y|\boldsymbol{x} \sim t_{2G}\{B, H(C+1)/(GC)\}$ and its HDI can be calculated using the qgt function.

2.6 Why do we have so many different distributions?

So far we have used many distributions, some you will have met before and some will be new. After a while the variety and sheer number of different distributions can become overwhelming. Why do we need so many distributions and why do we name so many of them?

Statistics studies the random variation in experiments, samples and processes. The variety of applications leads to their randomness being described by many different distributions. In many applications, bespoke distributions will need to be formulated. However, some distributions come up time and time again for modelling random variation in data and for describing prior beliefs. It is helpful for us to be able to refer to these distributions – and so we give each one a name – and also to be able to quote known results for these distributions such as their mean and variance. In this chapter you have been introduced to a generalisation of the t-distribution and the inverse chi distribution, and we have been able to use results for their mean and variance to study prior and posterior distributions and have been able to plot these distributions using functions in the `R` package.

You will meet several other new distributions in the remainder of the module. You won't be surprised to hear that it is useful to have a working knowledge of each of these distributions but perhaps not vital to remember all their properties listed in these notes. To help in this regard, the exam paper will contain a list of all the distributions used in the exam, together with their density (or probability function) and any useful results such as their mean and variance (as needed for the exam); see the specimen exam paper at the back of this booklet.

2.7 Learning objectives

By the end of this chapter, you should be able to:

- determine the posterior distribution for $(\mu, \tau)^T$
- determine and use the univariate prior and posterior distributions
- determine confidence intervals, HDIs and confidence regions
- determine the predictive distribution of another value from the population, and its predictive interval
- determine the predictive distribution of the mean of another random sample from the population

both in general and for a particular prior and data set. Also you should be able to:

- appreciate the benefit of naming distributions and for having lists of properties for these distributions