MAS3301 Bayesian Statistics

M. Farrow
School of Mathematics and Statistics
Newcastle University

Semester 2, 2008-9

15 Inference for Normal Distributions II

15.1 Student's t-distribution

When we look at the case where the precision is unknown we will find that we need to refer to Student's t-distribution. Suppose that $Y \sim N(0,1)$ and $X \sim \chi^2_d$, that is $X \sim \text{gamma}(d/2,\ 1/2)$, and that Y and X are independent. Let

$$T = \frac{Y}{\sqrt{X/d}}.$$

Then T has a Student's t-distribution on d degrees of freedom. We write $T \sim t_d$. "Student" was a pseudonym of W.S. Gosset, after whom the distribution is named.

The pdf is

$$f_T(t) = \frac{\Gamma(\{d+1\}/2)}{\sqrt{\pi d}\Gamma(d/2)}\left(1+\frac{t^2}{d}\right)^{-(d+1)/2} \qquad (-\infty < t < \infty).$$

Clearly this is a symmetric distribution with mean $\text{E}(T) = 0$. It can be shown that, for $d > 2$, the variance is $\text{var}(T) = d/(d-2)$. As $d \to \infty$, the t_d distribution tends to a standard normal $N(0,1)$ distribution. Figure 19 shows the pdf for different values of d.

A property of the Student's t-distribution which will be important for us is as follows. Suppose that τ and μ are two unknown quantities and that their joint distribution is such that the marginal distribution of τ is a gamma$(a,\ b)$ distribution and the conditional distribution of μ given τ is a normal $N(m,\ [c\tau]^{-1})$ distribution. Let us write $a = d/2$ and $b = dv/2$ so $\tau \sim \text{gamma}(d/2,\ dv/2)$. Then the marginal distribution of μ is such that

$$T = \frac{\mu - m}{\sqrt{v/c}} \sim t_d \tag{13}$$

where t_d is the Student's t-distribution on d degrees of freedom.

It is convenient to write the parameters of the gamma distribution as $d/2$ and $dv/2$ because then $dv\tau$ has a χ^2_d distribution.

<u>Proof</u> : The joint density of τ and μ is proportional to

$$\begin{aligned}
\tau^{d/2-1}e^{-dv\tau/2}\tau^{1/2}\exp\left\{-\frac{c\tau}{2}(\mu-m)^2\right\} &= \tau^{(d+1)/2-1}\exp\left\{-\frac{\tau}{2}\left[dv + c(\mu-m)^2\right]\right\} \\
&= \frac{\Gamma([d+1]/2)}{[\{dv+c(\mu-m)^2\}/2]^{(d+1)/2}}g(\tau;\ \tilde{a},\tilde{b})
\end{aligned}$$

where

$$g(\tau;\ \tilde{a},\tilde{b}) = \frac{\tilde{b}^{\tilde{a}}}{\Gamma(\tilde{a})}\tau^{\tilde{a}-1}\exp\{-\tilde{b}\tau\}$$

is the density of a gamma$(\tilde{a},\tilde{b})$ distribution for τ, where $\tilde{a} = (d+1)/2$ and $\tilde{b} = (\{dv + c(\mu - m)^2\}/2$. So the marginal density of μ is proportional to

$$\begin{aligned}
\int_0^\infty \frac{\Gamma([d+1]/2)}{[\{dv+c(\mu-m)^2\}/2]^{[d+1]/2}}g(\tau;\ \tilde{a},\tilde{b})\ d\tau &= \frac{\Gamma([d+1]/2)}{[\{dv+c(\mu-m)^2\}/2]^{[d+1]/2}} \\
&\propto [dv + c(\mu-m)^2]^{-[d+1]/2} \\
&\propto \left[1 + \frac{(\mu-m)^2}{dv/c}\right]^{-[d+1]/2}
\end{aligned}$$

If we write $t = (\mu - m)/\sqrt{v/c}$, then we get a density for t proportional to

$$\left[1+\frac{t^2}{d}\right]^{-[d+1]/2}$$

and this is proportional to the density of a Student's t-distribution on d degrees of freedom.

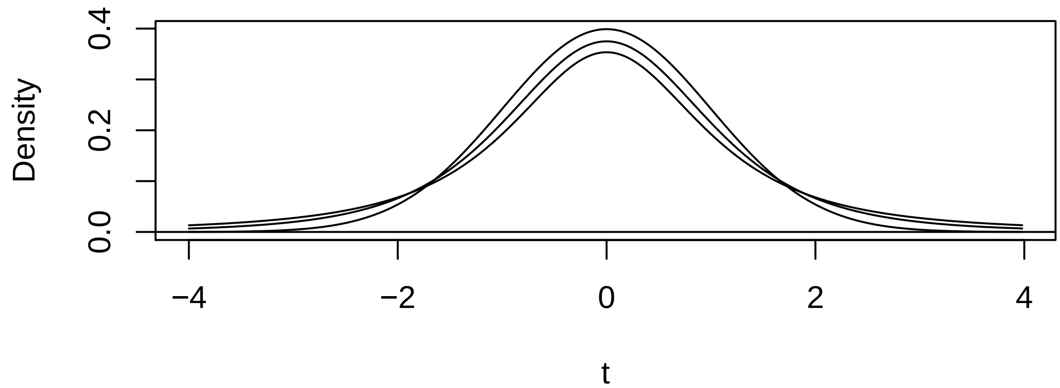

Figure 19: Probability density functions of Student's t-distributions with d degrees of freedom. From lowest to highest at $t = 0$, $d = 2$, $d = 4$, $d \to \infty$.

15.2 Unknown precision: conjugate prior

Now consider the case of inference for a normal distribution where the precision τ is unknown. Suppose we are going to observe data $Y_1, \dots, Y_n$ where $Y_i \sim N(\mu, \tau^{-1})$ and, given μ, τ, the observations are independent. Suppose now that both μ and τ are unknown.

There is a conjugate prior distribution. It is as follows.

We give τ a gamma$(d_0/2,\ d_0 v_0/2)$ prior distribution. We say that $\sigma^2 = \tau^{-1}$ has an "inverse gamma" prior because the reciprocal of σ^2 has a gamma prior. We can also say that $\sigma^2/(d_0 v_0)$ has an "inverse χ^2" distribution because

$$\frac{d_0 v_0}{\sigma^2} \sim \chi^2_{d_0}.$$

We then define the *conditional* prior distribution of μ given τ as a normal distribution with mean m_0 and precision $c_0\tau$, where the value of c_0 is specified. Thus the conditional prior precision of μ given τ is proportional to the error precision τ.

The joint prior density is therefore proportional to

$$\tau^{d_0/2-1} e^{-d_0 v_0 \tau/2} \tau^{1/2} \exp\left\{-\frac{c_0\tau}{2}(\mu - m_0)^2\right\}. \tag{14}$$

Using (13) we can see that the marginal prior distribution of μ is such that

$$T = \frac{\mu - m_0}{\sqrt{v_0/c_0}} \sim t_{d_0}$$

where t_{d_0} is the Student's t-distribution on d_0 degrees of freedom.

From (12) we see that our likelihood is proportional to

$$\begin{aligned}
\tau^{n/2} e^{-S_d\tau/2} \exp\left\{-\frac{n\tau}{2}(\bar{y} - \mu)^2\right\} &= \tau^{n/2} \exp\left\{-\frac{\tau}{2}\left[n(\bar{y} - \mu)^2 + S_d\right]\right\} \\
&= \tau^{n/2} \exp\left\{-\frac{\tau}{2}\left[n(\bar{y} - \mu)^2 + n s_n^2\right]\right\}
\end{aligned}$$

where

$$S_d = \sum_{i=1}^{n} (y_i - \bar{y})^2 = \sum_{i=1}^{n} y_i^2 - n\bar{y}^2$$

and

$$s_n^2 = \frac{S_d}{n}.$$

Multiplying this by the joint prior density (14) we obtain a posterior density proportional to

$$\tau^{d_1/2} e^{-d_0 v_0 \tau/2} \tau^{1/2} \exp\left\{-\frac{\tau}{2}\left[c_0(\mu - m_0)^2 + n(\bar{y} - \mu)^2 + n s_n^2\right]\right\}$$

where $d_1 = d_0 + n$. Now

$$\begin{aligned}
c_0(\mu - m_0)^2 + n(\bar{y} - \mu)^2 + n s_n^2 &= (c_0 + n)\mu^2 - 2(c_0 m_0 + n\bar{y})\mu + c_0 m_0^2 + n\bar{y}^2 + n s_n^2 \\
&= (c_0 + n)\left\{\mu^2 - 2\left(\frac{c_0 m_0 + n\bar{y}}{c_0 + n}\right)\mu + \left(\frac{c_0 m_0 + n\bar{y}}{c_0 + n}\right)^2\right\} \\
&\quad + c_0 m_0^2 + n\bar{y}^2 + n s_n^2 - \frac{(c_0 m_0 + n\bar{y})^2}{c_0 + n} \\
&= c_1\{\mu - m_1\}^2 + n v_d
\end{aligned}$$

where

$$\begin{aligned}
c_1 &= c_0 + n, \\
m_1 &= \frac{c_0 m_0 + n\bar{y}}{c_0 + n} \\
\text{and} \qquad n v_d &= c_0 m_0^2 + n\bar{y}^2 + n s_n^2 - \frac{(c_0 m_0 + n\bar{y})^2}{c_0 + n} \\
&= \frac{c_0 n(\bar{y} - m_0)^2 + n(c_0 + n)s_n^2}{c_0 + n} \\
&= n\left\{\frac{c_0 r^2 + n s_n^2}{c_0 + n}\right\}
\end{aligned}$$

where

$$\begin{aligned}
r^2 &= (\bar{y} - m_0)^2 + s_n^2 \\
&= \frac{1}{n}\left\{n m_0^2 + n\bar{y}^2 - 2n m_0 \bar{y} + \sum_{i=1}^{n} y_i^2 - n\bar{y}^2\right\} \\
&= \frac{1}{n}\sum_{i=1}^{n}(y_i - m_0)^2.
\end{aligned}$$

So the joint posterior density is proportional to

$$\tau^{d_1/2} e^{-d_1 v_1 \tau/2} \tau^{1/2} \exp\left\{-\frac{c_1 \tau}{2}(\mu - m_1)^2\right\}$$

where $d_1 v_1 = d_0 v_0 + n v_d$ and, since $d_1 = d_0 + n$,

$$v_1 = \frac{d_0 v_0 + n v_d}{d_0 + n}.$$

We see that this has the same form as (14) so the prior is indeed conjugate.

15.2.1 Summary

We can summarise the updating from prior to posterior as follows.

Prior :

$$\begin{aligned} d_0 v_0 \tau &\sim \chi^2_{d_0}, \\ \mu \mid \tau &\sim N(m_0,\ [c_0\tau]^{-1}) \\ \text{so} \qquad \frac{\mu - m_0}{\sqrt{v_0/c_0}} &\sim t_{d_0} \end{aligned}$$

Posterior : Given data $y_1, \ldots, y_n$ where these are independent observations (given $\mu,\ \tau$) from $n(\mu, \tau^{-1})$,

$$\begin{aligned} d_1 v_1 \tau &\sim \chi^2_{d_1}, \\ \mu \mid \tau &\sim N(m_1,\ [c_1\tau]^{-1}) \\ \text{so} \qquad \frac{\mu - m_1}{\sqrt{v_1/c_1}} &\sim t_{d_1} \end{aligned}$$

where

$$\begin{aligned} c_1 &= c_0 + n, \\ m_1 &= \frac{c_0 m_0 + n\bar{y}}{c_0 + n}, \\ d_1 &= d_0 + n, \\ v_1 &= \frac{d_0 v_0 + n v_d}{d_0 + n}, \\ v_d &= \frac{c_0 r^2 + n s_n^2}{c_0 + n}, \\ s_n^2 &= \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2 = \frac{1}{n}\left\{\sum_{i=1}^{n} y_i^2 - n\bar{y}^2\right\}, \end{aligned}$$

$$r^2 = (\bar{y} - m_0)^2 + s_n^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - m_0)^2.$$

15.2.2 Example

Suppose that our prior mean for τ is 1.5 and our prior standard deviation for τ is 0.5. This gives a prior variance for τ of 0.25. Thus, if the prior distribution of τ is gamma(a, b) then $a/b = 1.5$ and $a/b^2 = 0.25$. This gives $a = 9$ and $b = 6$. So $d_0 = 18$, $d_0 v_0 = 12$ and $v_0 = 2/3$.

Alternatively we could assess $v_0 = 2/3$ directly as a prior judgement about σ^2 and choose $d_0 = 18$ to reflect our degree of prior certainty. The prior mean of τ is $1/v_0 = 1.5$. The coefficient of variation of τ is $1/\sqrt{d_0/2}$. Since $1/\sqrt{9} = 1/3$, the prior standard deviation of τ is one third of its prior mean.

Using R we can find a 95% prior interval for τ.

```
> qgamma(0.025,9,6)
[1] 0.6858955
> qgamma(0.975,9,6)
[1] 2.627198
```

So the 95% prior interval is $0.686 < \tau < 2.627$. This corresponds to $0.3806 < \sigma^2 < 1.458$ or $0.6170 < \sigma < 1.207$. (We could approximate these intervals very roughly using the mean plus or minus two standard deviations for τ which gives $0.5 < \tau < 2.5$ which corresponds to $0.4 < \sigma^2 < 2.0$ or $0.632 < \sigma < 1.414$).

Suppose that our prior mean for μ is $m_0 = 20$ and that $c_0 = 1/10$ so that our conditional prior precision for μ, given that $\tau = 1.5$ is $1.5/10 = 0.15$. This corresponds to a conditional prior variance of 6.6667 and a conditional prior standard deviation of 2.5820. The marginal prior distribution for μ is therefore such that

$$t = \frac{\mu - 20}{\sqrt{(2/3)/0.1}} = \frac{\mu - 20}{2.582} \sim t_{18}.$$

Thus a 95% prior interval for μ is given by $20 \pm 2.101 \times 2.582$. That is $14.57 < \mu < 25.24$. (The value 2.101 is the 97.5% point of the t_{18} distribution).

Suppose that $n = 16$, $\bar{y} = 22.3$ and $S_d = 27.3$ so that $s_n^2 = S_d/16 = 1.70625$.

So

$$\begin{array}{lclcl}
c_1 & = & c_0 + 16 & = & 16.1, \\
m_1 & = & \frac{(1/10)\times 20 + 16\times 22.3}{(1/10)+16} & = & 22.2857, \\
d_1 & = & d_0 + 16 & = & 34, \\
(\bar{y} - m_0)^2 & = & (22.3 - 20)^2 & = & 5.29 \\
r^2 & = & (\bar{y} - m_0)^2 + s_n^2 & = & 6.99625, \\
v_d & = & \frac{c_0 r^2 + n s_n^2}{c_0 + n} & = & 1.7391, \\
v_1 & = & \frac{d_0 v_0 + n v_d}{d_0 + n} & = & 1.17134.
\end{array}$$

Thus the posterior distribution of τ is such that

$$d_1 v_1 \tau \sim \chi^2_{d_1}.$$

That is

$$39.826\tau \sim \chi^2_{34}.$$

The conditional posterior distribution of μ given τ is

$$\mu \mid \tau \sim N(m_1,\ [c_1\tau]^{-1}).$$

That is

$$\mu \mid \tau \sim N(22.2856,\ [16.1\tau]^{-1}).$$

The marginal posterior distribution of μ is such that

$$\frac{\mu - m_1}{\sqrt{v_1/c_1}} \sim t_{d_1}.$$

That is

$$\frac{\mu - 22.2857}{\sqrt{1.1734/16.1}} \sim t_{34}.$$

That is

$$\frac{\mu - 22.2857}{0.2697} \sim t_{34}.$$

Hence the marginal posterior 95% hpd interval for $t = (\mu - 22.2857)/0.2697$ is $-2.0322 < t < 2.0322$ where 2.0322 is the 97.5% point of the t_{34} distribution. (We can obtain this from tables or using the R command `qt(0.975,34)`). So, since $\mu = 22.2857 + 0.2697t$, the marginal 95% hpd interval for μ is $22.2857 \pm 2.0322 \times 0.2697$. That is

$$21.74 < \mu < 22.83.$$

16 More on Likelihood: Large Samples

16.1 Domination by the likelihood

We have an unknown parameter θ with prior pdf $f_\theta^{(0)}(\theta)$ and we will make observations $Y_1, \dots, Y_n$. Let us suppose that, given θ, $Y_1, \dots, Y_n$ are independent. (Similar results to the following can be proved in many cases where they are not). Let the pdf of Y_i given θ be $f_{Y_i|\theta}(y_i \mid \theta)$.

The likelihood is then

$$L(\theta;\ y) = \prod_{i=1}^{n} f_{Y_i|\theta}(y_i \mid \theta) = \prod_{i=1}^{n} L_i(\theta;\ y_i).$$

The posterior density is

$$f_{\theta|y}^{(1)}(\theta \mid y) = C f_\theta^{(0)}(\theta) \prod_{i=1}^{n} L_i(\theta;\ y_i).$$

Assuming that neither the prior density or the likelihood is zero, the log of the posterior density is

$$\log\{f_{\theta|y}^{(1)}(\theta \mid y)\} = \log(C) + \log\{f_\theta^{(0)}(\theta)\} + \sum_{i=1}^{n} l_i(\theta;\ y_i) \qquad (15)$$

where $l_i(\theta;\ y_i) = \log\{L_i(\theta;\ y_i)\}$.

Clearly, as n increases, the second (prior) term on the right hand side of (15) stays constant but the third (likelihood) term becomes more important. In the limit as $n \to \infty$ the posterior is dominated by the likelihood, provided that the prior is nonzero everywhere. Thus in large samples the posterior density is approximately proportional to the likelihood and the *posterior mode* is approximately equal to the maximum likelihood estimate.

16.2 Approximate normality

16.2.1 Scalar case

Subject to certain conditions, when we have a large sample the shape of the likelihood is approximately the shape of a normal distribution.

Suppose we have a scalar parameter θ. Let the log likelihood be $l(\theta)$ and the maximum likelihood estimate be $\hat{\theta}$. We can expand the log likelihood in a Taylor series about $\hat{\theta}$.

$$l(\theta) = l(\hat{\theta}) + (\theta - \hat{\theta}) \left(\frac{dl(\theta)}{d\theta}\right)_{\hat{\theta}} + \frac{1}{2!}(\theta - \hat{\theta})^2 \left(\frac{d^2 l(\theta)}{d\theta^2}\right)_{\hat{\theta}} + R(\theta, \hat{\theta})$$

where $R(\theta, \hat{\theta})$ represents the higher-order terms which are expected to be small when θ is close to $\hat{\theta}$.

Now

$$\left(\frac{dl(\theta)}{d\theta}\right)_{\hat{\theta}} = 0$$

so

$$l(\theta) \approx C - \frac{1}{2}(\theta - \hat{\theta})^2/V$$

where C is some constant and $V = -J^{-1}$ where

$$J = \left(\frac{d^2 l(\theta)}{d\theta^2}\right)_{\hat{\theta}}$$

is the (observed) *Fisher information.*
Exponentiating we see that the posterior density of θ is approximately proportional to

$$\exp\left\{-\frac{1}{2}(\theta - \hat{\theta})^2/V\right\}$$

so the posterior distribution of θ is approximately a normal distribution with mean $\hat{\theta}$ and variance $V = -J^{-1}$.

Note that we can often approximate the shape of a posterior distribution with a normal distribution. The mode of a normal $N(M, V)$ distribution is M so, if we can find the posterior mode, for example by maximising the sum of the log prior and log likelihood, we can approximate the posterior mean by M.

The normal probability density function is

$$(2\pi V)^{-1/2} \exp\left\{-\frac{1}{2V}(\theta - M)^2\right\}.$$

The log of this is

$$g(\theta) = -\frac{1}{2}\log(2\pi V) - \frac{1}{2V}(\theta - M)^2.$$

Differentiating this wrt θ we obtain

$$\frac{dg(\theta)}{d\theta} = -\frac{1}{V}(\theta - M).$$

Differentiating again we obtain

$$\frac{d^2 g(\theta)}{d\theta^2} = -\frac{1}{V}.$$

So, if we differentiate the sum of the log prior and the log likelihood twice and evaluate this second derivative at the posterior mode we get an approximation to $-1/V$.

These results allow us, for example, to find approximate credible intervals.

Note that this is an *asymptotic* result for large samples. We could derive similar results for some transformation of θ, e.g. $\log(\theta)$. The normal approximation might work better if we transform θ in some way. For example, a parameter θ might be restricted to an interval, such as $(0, 1)$ and the approximation might be better if we transform so that θ can take any value on the real line, e.g. $\eta = \log[\theta/(1-\theta)]$.

16.2.2 The multivariate normal distribution

Before we consider the vector case it is worth reminding ourselves about the multivariate normal distribution.

Suppose that $\underline{X}$ has a multivariate normal $N(\underline{M},\ V)$ distribution. This distribution has a *mean vector* $\underline{M} = (m_1, \ldots, m_n)^T$ where m_i is the mean of X_i, and a *variance matrix* V. The diagonal elements of V are the variances of $X_1, \ldots, X_n$ with the element in row and column i, v_{ii} being the variance of X_i. The covariance of X_i and X_j is v_{ij}, the element in row i and column j. Clearly $v_{ji} = v_{ij}$ and V is symmetric. It is also positive semi-definite.

The pdf is

$$f_X(\underline{x}) = (2\pi)^{-n/2}|V|^{-1/2}\exp\left\{-\frac{1}{2}(\underline{x}-\underline{M})^T V^{-1}(\underline{x}-\underline{M})\right\}.$$

(Here $\underline{x}^T$ denotes the transpose of $\underline{x}$). We often work in terms of the *precision matrix* $P = V^{-1}$. In this case, of course, we replace $(\underline{x}-\underline{M})^T V^{-1}(\underline{x}-\underline{M})$ with $(\underline{x}-\underline{M})^T P(\underline{x}-\underline{M})$.

If $\underline{X}$ have a multivariate normal $N(\underline{M},\ V)$ distribution and V is a diagonal matrix, that is if $\text{covar}(X_i, X_j) = 0$ when $i \neq j$, then $X_1, \ldots, X_n$ are independent.

16.2.3 Approximate normality: The vector case

Now θ is a vector parameter and its maximum likelihood estimate $\hat{\theta}$ is also a vector. The Taylor series expansion becomes

$$l(\theta) = l(\hat{\theta}) + (\theta-\hat{\theta})^T\left(\frac{\partial l(\theta)}{\partial\theta}\right)_{\hat{\theta}} + \frac{1}{2!}(\theta-\hat{\theta})^T\left(\frac{\partial^2 l(\theta)}{\partial\theta\partial\theta^T}\right)_{\hat{\theta}}(\theta-\hat{\theta}) + R(\theta,\hat{\theta})$$

This leads to an approximately multivariate normal posterior distrbution with mean vector M and variance matrix V where M is the posterior mode (or the maximum likelihood estimate) and

$$V = -J^{-1}$$

where J is the (observed) Fisher information matrix, the matrix of partial second derivatives of the log likelihood (or the log of the prior density plus the log likelihood), evaluated at the mode.

Practical 2

1: Project

You may wish to use this session as an opportunity to work on the project or ask any questions about it. Remember, for example, that Task 4 of the project is rather like Part 4 of Practical 1.

If you have already finished the project or you do not want to work on it for some other reason then there is another exercise below.

2: Two Normal Samples

We have looked at inference in the case of a single normal sample. We can, of course, extend this to more than one sample, just as, in non-Bayesian statistics, we would look at the two-sample t-test, the one-way analysis of variance etc. We do not have time in this module to look at the theory of this but we can do an example using the computer.

The following data originate with the Western Collaborative Group Study, which studied middle-aged men in California in 1960-61. The data are given by Selvin (1991) and by Hand *et al.* (1991). They give cholesterol measurements (mg per 100 ml) for two groups of heavy men. The groups are distinguised by behaviour type, A or B.

Type A

233	291	312	250	246	197	268	224	239	239
254	276	234	181	248	252	202	218	212	325

Type B

344	185	263	246	224	212	188	250	148	169
226	175	242	252	153	183	137	202	194	213

We regard these two sets of measurements as random samples from the "Type A" population and "Type B" population respectively and suppose that, given the values of $\mu_A,\ \mu_B, \tau$, an observation on a Type A subject will have a $N(\mu_A, \tau^{-1})$ distribution and an observation on a Type B subject will have a $N(\mu_B, \tau^{-1})$ distribution. It seems reasonable to give the same marginal prior distribution to both μ_A and μ_B. Prior knowledge of cholesterol levels might suggest a prior mean of, say, 200 but, for these groups, the actual value of μ might be somewhat different so a large prior standard deviation, say 50, might be appropriate. This gives us a prior variance for μ_A and μ_B of 2500. We also need a prior covariance. To get a value for this we can think about $\mu_A - \mu_B$. We have already implied that this difference has a prior mean of zero but we are unsure what it will actually be so we give it quite a large prior standard deviation, say 50 again. This implies a prior variance of 2500 for $\mu_A - \mu_B$. Now

$$\text{var}(\mu_A - \mu_B) = \text{var}(\mu_A) + \text{var}(\mu_B) - 2\ \text{covar}(\mu_A, \mu_B).$$

Hence $\text{covar}(\mu_A, \mu_B) = 1250$. Now these variances and covariances need to depend on τ in the conjugate prior so let us suppose that the individual standard deviation, that is the standard deviation of $Y - \mu$ where Y is an individual observation, is about 50, implying $\tau = 1/2500$. The conjugate prior distribution has a conditional bivariate normal distribution for $(\mu_A,\ \mu_B)^T$ given τ with precision matrix $C_0\tau$. Our reasoning about the prior variances and covariance gives us

$$C_0 = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}^{-1} = \frac{2}{3}\begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}.$$

Now, of course, we do not know the value of τ. If we take our value of 1/2500 as the prior median and also say that there is a 10% chance that the standard deviation is less than about 20 then this allows us to find a value for $d_0/2$ and $d_0v_0/2$. These turn out to be approximately $d_0/2 = 0.48$ and $d_0v_0/2 = 520$ so $d_0 = 0.96,\ d_0v_0 = 1040$ and $v_0 = 1083$.

1. The R function `findab`, which calls another function `findb`, can be used to find the values of a and b for a gamma distribution when two quantiles are specified. Both of these functions can be obtained from the module Web page. Install both of them.

```
findab<-function(starta,tau,p,niter)
{aa<-starta
 a<-numeric(niter)
 b<-numeric(niter)
 diff<-numeric(niter)
 c1<-findb(aa[1],tau,p)
 c2<-findb(aa[2],tau,p)
 while((c1[2]*c2[2])>0)
      {aa[1]<-aa[1]/2
       aa[2]<-aa[2]*2
       c1<-findb(aa[1],tau,p)
       c2<-findb(aa[2],tau,p)
       }
 for (i in 1:niter)
     {a[i]<-(aa[1]+aa[2])/2
      newc<-findb(a[i],tau,p)
      diff[i]<-newc[2]
      b[i]<-newc[1]
      if ((diff[i]*c1[2])>0)
         {aa[1]<-a[i]
          c1<-newc
          }
      else
         {aa[2]<-a[i]
          c2<-newc
          }
      }
a<-signif(a,4)
b<-signif(b,4)
diff<-signif(diff,4)
table<-data.frame(a,b,diff)
write.table(table,file="")
c(a[niter],b[niter])
 }


findb<-function(a,tau,p)
{q<-qgamma(p[2],a,1)
 b<-q/tau[2]
 diff<-pgamma(tau[1],a,b)-p[1]
 c(b,diff)
 }
```

Figure 20: Functions to find the parameters of a gamma distribution with two specified quantiles.

```
oneway<-function(d0,v0,C0,M0,y,group)
{J<-max(group)
 N<-numeric(J)
 ybar<-numeric(J)
 Sd<-0
 C1<-C0
 M1<-C0%*%M0
 for (j in 1:J)
     {N[j]<-sum(group==j)
      yj<-y[group==j]
      ybar[j]<-mean(yj)
      Sd<-Sd+sum((yj-ybar[j])^2)
      C1[j,j]<-C1[j,j]+N[j]
      }
 M1<-M1+N*ybar
 M1<-solve(C1,M1)
 R<-t(M0)%*%C0%*%M0+sum(N*ybar*ybar)-t(M1)%*%C1%*%M1
 N<-sum(N)
 Nvd<-Sd+R
 d1<-d0+N
 v1<-(d0*v0+Nvd)/d1
 list(d1=d1,v1=v1,C1=C1,M1=M1)
 }
```

Figure 21: R function for several normal samples, conjugate prior

2. To use `findab` we need to supply a lower and an upper starting value for a, the two specified quantiles and the two corresponding probabilities. So enter the following.

   ```
   t1<-1/2500
   t2<-1/(20^2)
   starta<-c(0.1,5)
   tau<-c(t1,t2)
   p<-c(0.5,0.9)
   ```

3. Now use the function.

   ```
   ab<-findab(starta,tau,p,20)
   ab
   ```

   Note that, if there is a 10% chance that the standard deviation is less than 20 then this corresponds to a 10% chance that $\tau > 1/(20^2)$. That is a probability of 0.9 that $\tau < 1/(20^2)$. The final argument of the function, 20, is the number of iterations to be done.

   The values of d_0 and $d_0 v_0$ are, of course, twice the values in `ab`.

4. It is convenient to use a R function to do the calculations to find the posterior distribution. Figure 21 shows a suitable function, called `oneway`. This function can be obtained from the module Web page.

5. Install the R function `oneway`.

6. Read the cholesterol data. For example:

   ```
   cholest<-scan("http://www.mas.ncl.ac.uk/~nmf16/teaching/mas3301/cholesterol.txt")
   ```

7. Set up a group variable and define the prior parameters.

```
group<-rep(1:2,c(20,20))
d0<-2*ab[1]
v0<-2*ab[2]/d0
C0<-matrix(c(2,-1,-1,2),ncol=2)*2/3
M0<-c(200,200)
```

8. Use the function `oneway` to find the posterior distribution.

```
> post<-oneway(d0,v0,C0,M0,cholest,group)
> post
```

9. The marginal posterior distribution of μ_A is such that

$$t_A = \frac{\mu_A - M_{1,A}}{\sqrt{v_1 w_{1,AA}}}$$

has a Student's t-distribution on d_1 degrees of freedom. Here

$$M_1 = \begin{pmatrix} M_{1,A} \\ M_{1,B} \end{pmatrix} \quad \text{and} \quad C_1^{-1} = \begin{pmatrix} w_{1,AA} & w_{1,AB} \\ w_{1,AB} & w_{1,BB} \end{pmatrix}.$$

Plot this marginal posterior distribution as follows.

```
w1<-solve(post$C1)
stepA<-1
muA<-seq(180,280,stepA)
seA<-sqrt(post$v1*w1[1,1])
tA<-(muA-post$M1[1])/seA
densA<-dt(tA,post$d1)/seA
plot(muA,densA,type="l",xlab=expression(mu[A]),ylab="Density")
```

Note that we have to scale the density by the standard error `seA` because of the transformation between t_A and μ_A.

10. Similarly the marginal posterior distribution of μ_B is such that

$$t_B = \frac{\mu_B - M_{1,B}}{\sqrt{v_1 w_{1,BB}}}$$

has a Student's t-distribution on d_1 degrees of freedom.

Plot this marginal posterior distribution as follows. (Note that some things are the same between the μ_A and μ_B cases so we could take some shortcuts).

```
stepB<-1
muB<-seq(180,280,stepB)
seB<-sqrt(post$v1*w1[2,2])
tB<-(muB-post$M1[2])/seB
densB<-dt(tB,post$d1)/seB
plot(muB,densB,type="l",xlab=expression(mu[B]),ylab="Density")
```

11. We can, in fact, also plot the joint posterior density of μ_A and μ_B. We can use a bit of theory which we do not cover in this module to help with this. It can be shown that the joint posterior density of μ_A and μ_B is proportional to

$$[d_1 v_1 + Q]^{-(d_1+2)/2}$$

where

$$\begin{aligned} Q &= (\underline{\mu} - \underline{M}_1)^T C_1 (\underline{\mu} - \underline{M}_1) \\ &= (\mu_A - M_{1,A})^2 C_{1,AA} + (\mu_B - M_{1,B})^2 C_{1,BB} + 2(\mu_A - M_{1,A})(\mu_B - M_{1,B}) C_{1,AB} \end{aligned}$$

We can do the calculations as follows.

```
mA<-matrix(muA,nrow=length(muA),ncol=length(muB))
mB<-matrix(muB,nrow=length(muA),ncol=length(muB),byrow=T)
dA<-mA-post$M1[1]
dB<-mB-post$M1[2]
B1<-post$d1*post$v1
q<-(dA^2)*post$C1[1,1] + (dB^2)*post$C1[2,2] + 2*dA*dB*post$C1[1,2]
dens<-(B1+q)^(-(post$d1+2)/2)
dens<-dens/(sum(dens)*stepA*stepB)
contour(muA,muB,dens)
contour(muA,muB,dens,xlab=expression(mu[A]),ylab=expression(mu[B]))
```

We can add a line showing where $\mu_A = \mu_B$ as follows.

```
abline(0,1,lty=2)
```