MATH 559: BAYESIAN THEORY AND METHODS

BAYESIAN MODELLING WITH THE NORMAL MODEL

Suppose that a model is to be constructed under an assumption of exchangeability with the following components:

- Finite realization $y_1, \ldots, y_n$ recorded;
- $\mathcal{Y} \equiv \mathbb{R}$;
- $\theta = (\mu, \sigma)$;
- $f_Y(y;\theta) \equiv Normal(\mu, \sigma^2)$
- $\pi_0(\mu, \sigma)$ a prior density on $\mathbb{R} \times \mathbb{R}^+$.

In the book *Bayesian Theory*, Chapter 4, pp 183 – 186, the authors point out that this model can be justified by a requirement for *centered spherical symmetry* amongst the Y variables, that is, for all $n \geq 1$ the quantities

$$Y_1 - \overline{Y}_n, \ldots, Y_n - \overline{Y}_n$$

exhibit *spherical symmetry*, that is, if $Z_i = Y_i - \overline{Y}_n$ for $i = 1, \ldots, n$, then the vectors

$$\mathbf{Z}_n = (Z_1, \ldots, Z_n)^\top \qquad \text{and} \qquad \mathbf{A}\mathbf{Z}_n$$

have the same distribution, for any $n \times n$ matrix $\mathbf{A}$ such that $\mathbf{A}^\top\mathbf{A} = \mathbf{I}_n$. The de Finetti representation then takes the form

$$f_{Y_1,\ldots,Y_n}(y_1, \ldots, y_n) = \int_{-\infty}^{\infty}\int_0^{\infty} \left\{\prod_{i=1}^n f_Y(y_i;\mu,\sigma)\right\} \pi_0(d\mu, d\sigma)$$

For illustration, we consider a data-generating scenario where the true values of the parameters are $\mu_0 = 2, \sigma_0 = 1$ and consider a sample of size $n = 20$.

```
set.seed(234)
n<-20
theta0<-2;sigma0<-1
y<-rnorm(n,theta0,sigma0)
round(y,4)

+  [1]  2.6608 -0.0530  0.5008  3.4712  3.4591  2.1401  2.2092 -1.0361  1.5131
+ [10]  0.9121  2.0579  3.1040  1.9744  2.5148  2.9901  2.3035  1.0699  2.0840
+ [19]  2.5268  2.0159

(ybar<-mean(y))

+ [1] 1.920928
```

Bayesian inference for θ with σ known.

First consider the case where σ is treated as known and correctly specified as σ_0; we can consider this specification by imagining that

$$\pi_0(\mu, \sigma) = \pi_0(\mu|\sigma)\delta_{\sigma_0}(\sigma)$$

that is, the prior for σ is a degenerate function with mass 1 at σ_0. Then the Bayesian calculation becomes

$$f_{Y_1,\ldots,Y_n}(y_1, \ldots, y_n) = \int_{-\infty}^{\infty} \left\{\prod_{i=1}^n f_Y(y_i;\mu,\sigma_0)\right\} \pi_0(d\mu|\sigma_0) \tag{1}$$

We have

$$\prod_{i=1}^{n} f_Y(y_i;\mu,\sigma_0) = \prod_{i=1}^{n} \left(\frac{1}{2\pi\sigma_0^2}\right)^{1/2} \exp\left\{-\frac{1}{2\sigma_0^2}(y_i-\mu)^2\right\}$$
$$= \left(\frac{1}{2\pi\sigma_0^2}\right)^{n/2} \exp\left\{-\frac{1}{2\sigma_0^2}\sum_{i=1}^{n}(y_i-\mu)^2\right\}$$
$$= \left(\frac{1}{2\pi\sigma_0^2}\right)^{n/2} \exp\left\{-\frac{1}{2\sigma_0^2}\left[n(\overline{y}_n-\mu)^2+\sum_{i=1}^{n}(y_i-\overline{y}_n)^2\right]\right\}$$

where

$$\overline{y}_n = \frac{1}{n}\sum_{i=1}^{n} y_i \qquad \sum_{i=1}^{n}(y_i-\mu)^2 = n(\overline{y}_n-\mu)^2+\sum_{i=1}^{n}(y_i-\overline{y}_n)^2$$

by the usual sums-of-squares decomposition. Ignoring constants that do not involve μ, we see that

$$\prod_{i=1}^{n} f_Y(y_i;\mu,\sigma_0) \propto \exp\left\{-\frac{n}{2\sigma_0^2}(\overline{y}_n-\mu)^2\right\}.$$

Suppose that we specify that

$$\pi_0(\mu|\sigma_0) \equiv Normal(\eta,\sigma_0^2/\lambda)$$

for some fixed $\eta \in \mathbb{R}$ and $\lambda > 0$; that is

$$\pi_0(\mu|\sigma_0) = \left(\frac{\lambda}{2\pi\sigma_0^2}\right)^{1/2} \exp\left\{-\frac{\lambda}{2\sigma_0^2}(\mu-\eta)^2\right\}. \tag{2}$$

Then, up to proportionality, we have that

$$\pi_n(\mu|\sigma_0) \propto \exp\left\{-\frac{n}{2\sigma_0^2}(\overline{y}_n-\mu)^2\right\}\exp\left\{-\frac{\lambda}{2\sigma_0^2}(\mu-\eta)^2\right\} = \exp\left\{-\frac{1}{2\sigma_0^2}\left[n(\overline{y}_n-\mu)^2+\lambda(\mu-\eta)^2\right]\right\}.$$

To simplify this expression, we may use the complete-the-square formula

$$A(x-a)^2+B(x-b)^2 = (A+B)\left(x-\frac{Aa+Bb}{A+B}\right)^2+\frac{AB}{A+B}(a-b)^2$$

to deduce that

$$n(\overline{y}_n-\mu)^2+\lambda(\mu-\eta)^2 = (n+\lambda)\left(\mu-\frac{n\overline{y}_n+\lambda\eta}{n+\lambda}\right)^2+\frac{n\lambda}{n+\lambda}(\overline{y}_n-\eta)^2.$$

Thus

$$\pi_n(\mu|\sigma_0) \propto \exp\left\{-\frac{1}{2\sigma_0^2}\left[(n+\lambda)\left(\mu-\frac{n\overline{y}_n+\lambda\eta}{n+\lambda}\right)^2+\frac{n\lambda}{n+\lambda}(\overline{y}_n-\eta)^2\right]\right\}$$
$$\propto \exp\left\{-\frac{(n+\lambda)}{2\sigma_0^2}\left(\mu-\frac{n\overline{y}_n+\lambda\eta}{n+\lambda}\right)^2\right\}$$

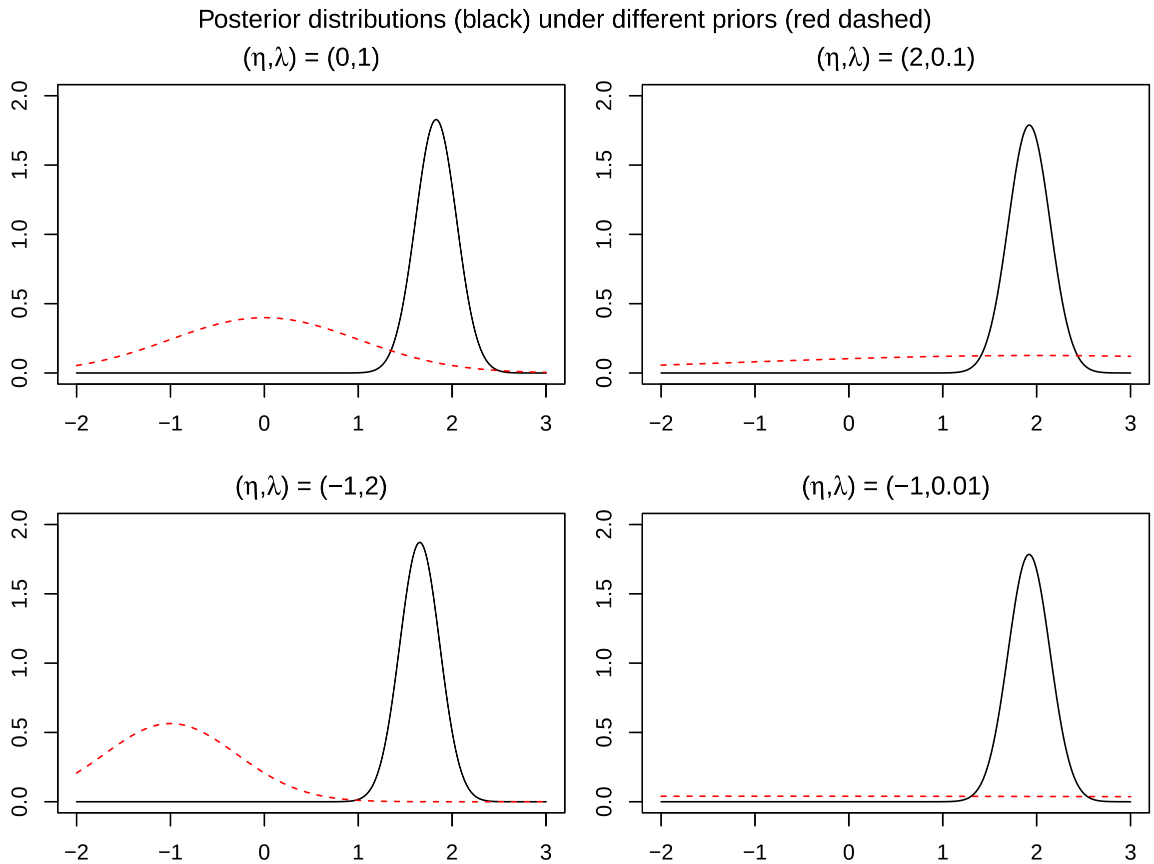
and from this we may deduce that $\pi_n(\mu|\sigma_0) \equiv Normal(\eta_n,\sigma_0^2/\lambda_n)$, where

$$\eta_n = \frac{n\overline{y}_n+\lambda\eta}{n+\lambda} \qquad \lambda_n = n+\lambda.$$

We can investigate the behaviour of the posterior for different settings of the prior parameters (or *hyper-parameters*):

- $(\eta, \lambda) = (0, 1)$;
- $(\eta, \lambda) = (2, 0.1)$;
- $(\eta, \lambda) = (-1, 2)$;
- $(\eta, \lambda) = (-1, 0.01)$.

```
par(oma=c(2,2,1,4),mar=c(3,3,2,0),mfrow=c(2,2))
xv<-seq(-2,3,by=0.01)

##Prior 1
eta<-0;lambda<-1
eta.n<-(n*ybar+lambda*eta)/(n+lambda)
lambda.n<-n+lambda
yv1<-dnorm(xv,eta.n,sqrt(sigma0^2/lambda.n))
mtxt<-substitute(paste('(',eta,',',lambda,')',' = (',ev,',',lam,')'),
    list(ev=eta,lam = lambda))
plot(xv,yv1,type='l',main=mtxt,ylim=range(0,2))
lines(xv,dnorm(xv,eta,sqrt(sigma0^2/lambda)),col='red',lty=2)

##Prior 2
eta<-2;lambda<-0.1
eta.n<-(n*ybar+lambda*eta)/(n+lambda)
lambda.n<-n+lambda
yv2<-dnorm(xv,eta.n,sqrt(sigma0^2/lambda.n))
mtxt<-substitute(paste('(',eta,',',lambda,')',' = (',ev,',',lam,')'),
    list(ev=eta,lam = lambda))
plot(xv,yv2,type='l',main=mtxt,ylim=range(0,2))
lines(xv,dnorm(xv,eta,sqrt(sigma0^2/lambda)),col='red',lty=2)

##Prior 3
eta<--1;lambda<-2
eta.n<-(n*ybar+lambda*eta)/(n+lambda)
lambda.n<-n+lambda
yv3<-dnorm(xv,eta.n,sqrt(sigma0^2/lambda.n))
mtxt<-substitute(paste('(',eta,',',lambda,')',' = (',ev,',',lam,')'),
    list(ev=eta,lam = lambda))
plot(xv,yv3,type='l',main=mtxt,ylim=range(0,2))
lines(xv,dnorm(xv,eta,sqrt(sigma0^2/lambda)),col='red',lty=2)

##Prior 4
eta<--1;lambda<-0.01
eta.n<-(n*ybar+lambda*eta)/(n+lambda)
lambda.n<-n+lambda
yv4<-dnorm(xv,eta.n,sqrt(sigma0^2/lambda.n))
mtxt<-substitute(paste('(',eta,',',lambda,')',' = (',ev,',',lam,')'),
    list(ev=eta,lam = lambda))
plot(xv,yv4,type='l',main=mtxt,ylim=range(0,2))
lines(xv,dnorm(xv,eta,sqrt(sigma0^2/lambda)),col='red',lty=2)
mtext(text="Posterior distributions (black) under different priors (red dashed)",
    side=3,line=0,outer=TRUE)
```



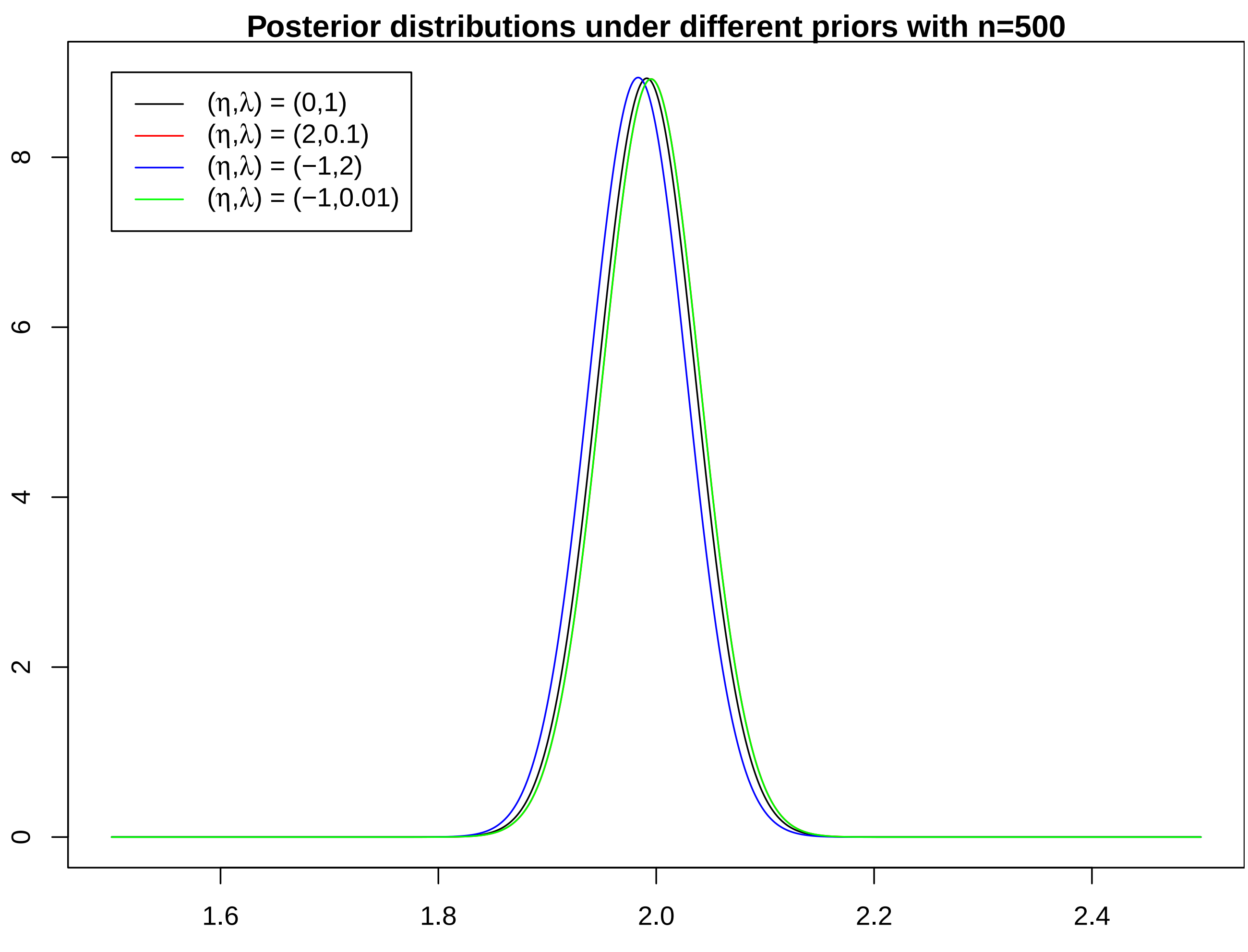
Note that as n increases,

$$\eta_n = \frac{n\overline{y}_n + \lambda\eta}{n+\lambda} \doteq \overline{y}_n \qquad \lambda_n = n + \lambda \doteq n$$

and the influence of the prior diminishes. We can see the effect of changing λ if the sample size is increased to 500;

```
set.seed(234);n<-500
y<-rnorm(n,theta0,sigma0);ybar<-mean(y)
par(mar=c(3,3,1,0));xv<-seq(1.5,2.5,by=0.001)

eta<-0;lambda<-1        ##Prior 1
eta.n<-(n*ybar+lambda*eta)/(n+lambda)
lambda.n<-n+lambda
yv1<-dnorm(xv,eta.n,sqrt(sigma0^2/lambda.n))
mtxt1<-expression(paste('(',eta,',',lambda,')',' = (0,1)'))
plot(xv,yv1,type='l',ylim=range(0,9),xlab=expression(mu))
title("Posterior distributions under different priors with n=500")

eta<-2;lambda<-0.1      ##Prior 2
eta.n<-(n*ybar+lambda*eta)/(n+lambda)
lambda.n<-n+lambda
yv2<-dnorm(xv,eta.n,sqrt(sigma0^2/lambda.n))
mtxt2<-expression(paste('(',eta,',',lambda,')',' = (2,0.1)'))
lines(xv,yv2,col='red')
```

```
eta<--1;lambda<-2      ##Prior 3
eta.n<-(n*ybar+lambda*eta)/(n+lambda)
lambda.n<-n+lambda
yv3<-dnorm(xv,eta.n,sqrt(sigma0^2/lambda.n))
mtxt3<-expression(paste('(',eta,',',lambda,')',' = (-1,2)'))
lines(xv,yv3,col='blue')

eta<--1;lambda<-0.01   ##Prior 4

eta.n<-(n*ybar+lambda*eta)/(n+lambda)
lambda.n<-n+lambda
yv4<-dnorm(xv,eta.n,sqrt(sigma0^2/lambda.n))
mtxt4<-expression(paste('(',eta,',',lambda,')',' = (-1,0.01)'))
lines(xv,yv4,col='green')
legend(1.5,9.0,c(mtxt1,mtxt2,mtxt3,mtxt4),lty=1,col=c('black','red','blue','green'))
```



In this plot, the green and red lines are exactly overlapping.

From (1), we have for the exchangeable (marginal) specification

$$f_{Y_1,\ldots,Y_n}(y_1,\ldots,y_n) = \int_{-\infty}^{\infty} \left\{ \prod_{i=1}^{n} f_Y(y_i;\mu,\sigma_0) \right\} \pi_0(d\mu|\sigma_0) \tag{3}$$

$$= \int_{-\infty}^{\infty} \left(\frac{1}{2\pi\sigma_0^2}\right)^{n/2} \exp\left\{-\frac{1}{2\sigma_0^2}\left[n(\overline{y}_n-\mu)^2+\sum_{i=1}^{n}(y_i-\overline{y}_n)^2\right]\right\} \left(\frac{\lambda}{2\pi\sigma_0^2}\right)^{1/2} \exp\left\{-\frac{\lambda}{2\sigma_0^2}(\mu-\eta)^2\right\}\,d\mu$$

$$= \left(\frac{1}{2\pi\sigma_0^2}\right)^{(n+1)/2} \lambda^{1/2} \exp\left\{-\frac{1}{2\sigma_0^2}\left[\sum_{i=1}^{n}(y_i-\overline{y}_n)^2+\frac{n\lambda}{n+\lambda}(\overline{y}_n-\eta_n)^2\right]\right\} \int_{-\infty}^{\infty} \exp\left\{-\frac{\lambda_n}{2\sigma_0^2}(\mu-\eta_n)^2\right\}\,d\mu$$

$$= \left(\frac{1}{2\pi\sigma_0^2}\right)^{n/2} \left(\frac{\lambda}{\lambda_n}\right)^{1/2} \exp\left\{-\frac{1}{2\sigma_0^2}\left[\sum_{i=1}^{n}(y_i-\overline{y}_n)^2+\frac{n\lambda}{n+\lambda}(\overline{y}_n-\eta)^2\right]\right\} \tag{4}$$

Prediction: The predictive distribution for Y_{n+1} given $Y_1 = y_1,\ldots,Y_n = y_n$ is given by

$$f_{Y_{n+1}|Y_1,\ldots,Y_n}(y_{n+1}|y_1,\ldots,y_n) = \int_{-\infty}^{\infty} f_Y(y_{n+1};\mu,\sigma_0)\pi_n(\mu|\sigma_0)\,d\mu$$

$$= \int_{-\infty}^{\infty} \left(\frac{1}{2\pi\sigma_0^2}\right)^{1/2} \exp\left\{-\frac{1}{2\sigma_0^2}(y_{n+1}-\mu)^2\right\} \left(\frac{\lambda_n}{2\pi\sigma_0^2}\right)^{1/2} \exp\left\{-\frac{\lambda_n}{2\sigma_0^2}(\mu-\eta_n)^2\right\}\,d\mu$$

$$= \left(\frac{1}{2\pi\sigma_0^2}\right)(\lambda_n)^{1/2} \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2\sigma_0^2}\left[(y_{n+1}-\mu)^2+\lambda_n(\mu-\eta_n)^2\right]\right\}\,d\mu$$

We may write the term in the square brackets as

$$(1+\lambda_n)\left(\mu-\frac{y_{n+1}+\lambda_n\eta_n}{1+\lambda_n}\right)^2+\frac{\lambda_n}{1+\lambda_n}(y_{n+1}-\eta_n)^2$$

so therefore the predictive distribution $f_{Y_{n+1}|Y_1,\ldots,Y_n}(y_{n+1}|y_1,\ldots,y_n)$ takes the form

$$\left(\frac{1}{2\pi\sigma_0^2}\right)(\lambda_n)^{1/2} \exp\left\{-\frac{\lambda_n}{2(1+\lambda_n)\sigma_0^2}(y_{n+1}-\eta_n)^2\right\} \int_{-\infty}^{\infty} \exp\left\{-\frac{(1+\lambda_n)}{2\sigma_0^2}\left(\mu-\frac{y_{n+1}+\lambda_n\eta_n}{1+\lambda_n}\right)^2\right\}\,d\mu$$

$$= \left(\frac{\lambda_n}{2\pi(1+\lambda_n)\sigma_0^2}\right)^{1/2} \exp\left\{-\frac{\lambda_n}{2(1+\lambda_n)\sigma_0^2}(y_{n+1}-\eta_n)^2\right\}.$$

We conclude that

$$f_{Y_{n+1}|Y_1,\ldots,Y_n}(y_{n+1}|y_1,\ldots,y_n) \equiv Normal\left(\mu_{n,1},\frac{\sigma_0^2}{\lambda_{n,1}}\right) \tag{5}$$

where

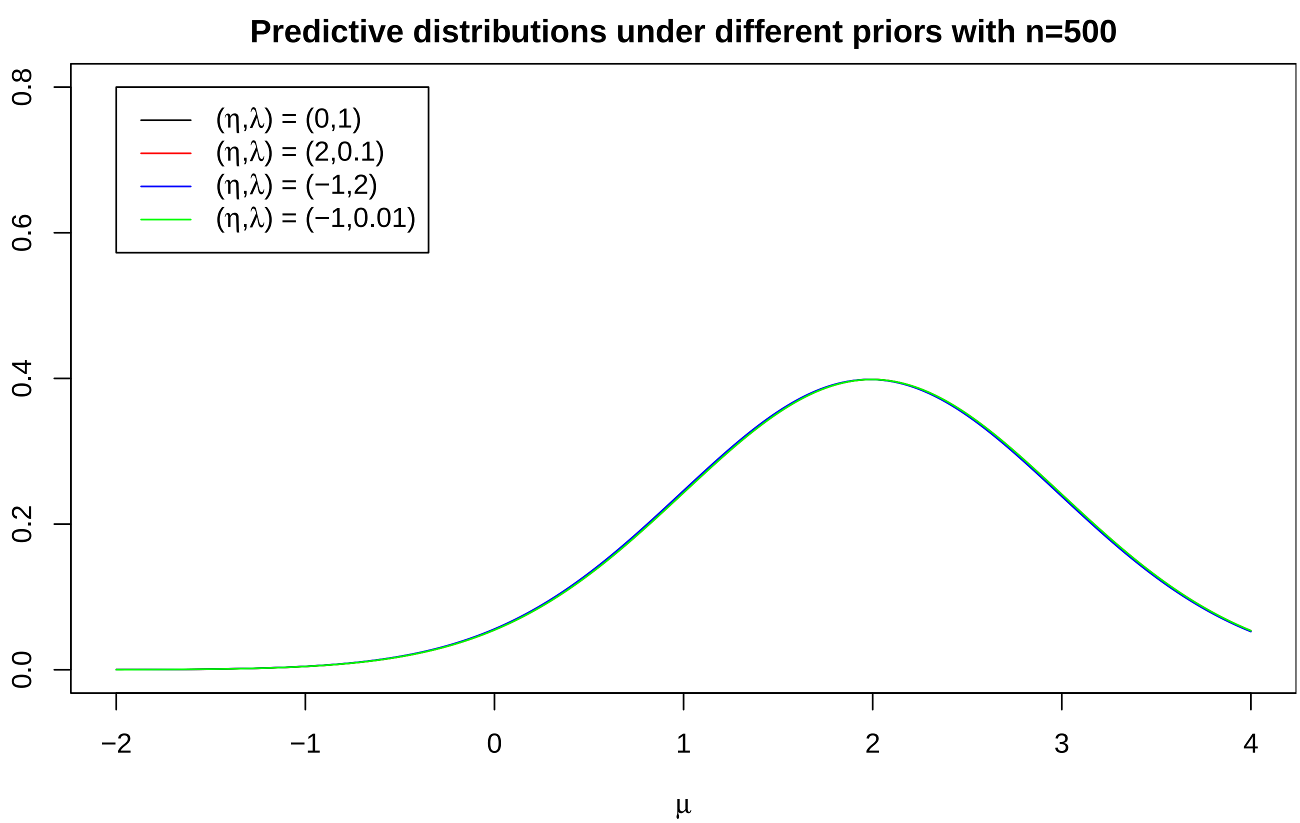
$$\mu_{n,1} = \eta_n \qquad \lambda_{n,1} = \frac{\lambda_n}{1+\lambda_n}.$$

Choice of λ: Note that as $\lambda_n = n + \lambda$, setting $\lambda \longrightarrow 0$ may be regarded as an attempt to express prior ignorance about μ; recall that the prior on μ is $Normal(\eta, \sigma_0^2/\lambda)$, so as λ gets smaller the prior variance increases. If $\lambda = 0$, the posterior distribution is well-defined as $\pi_n(\mu|\sigma_0) \equiv Normal\left(\overline{y}_n, \sigma_0^2/n\right)$. For the predictive distribution

$$\lambda_{n,1} = \frac{\lambda_n}{1+\lambda_n} \qquad \therefore \qquad \lambda_{n,1}^{-1} = 1 + \lambda_n^{-1}$$

As $n \longrightarrow \infty$, $\lambda_{n,1} \longrightarrow 1$. However, this also reveals that as $\lambda \longrightarrow 0$, whatever the value of n, the prior predictive distribution (4) is not a proper distribution. The effect of changing λ on the posterior predictive is illustrated in the following plots with $n = 500$.

```
set.seed(234);n<-500
y<-rnorm(n,theta0,sigma0);ybar<-mean(y)
par(mar=c(4,3,2,0)); xv<-seq(-2,4,by=0.001)

eta<-0;lambda<-1   ##Prior 1
eta.n<-(n*ybar+lambda*eta)/(n+lambda)
lambda.n<-n+lambda
lambda.npred<-(lambda.n)/(lambda.n+1)
yv1<-dnorm(xv,eta.n,sqrt(sigma0^2/lambda.npred))
mtxt1<-expression(paste('(',eta,',',lambda,')',' = (0,1)'))
plot(xv,yv1,type='l',ylim=range(0,0.8),xlab=expression(mu))
title("Predictive distributions under different priors with n=500")

eta<-2;lambda<-0.1  ##Prior 2
eta.n<-(n*ybar+lambda*eta)/(n+lambda)
lambda.n<-n+lambda
lambda.npred<-(lambda.n)/(lambda.n+1)
yv2<-dnorm(xv,eta.n,sqrt(sigma0^2/lambda.npred))
mtxt2<-expression(paste('(',eta,',',lambda,')',' = (2,0.1)'))
lines(xv,yv2,col='red')

eta<--1;lambda<-2   ##Prior 3
eta.n<-(n*ybar+lambda*eta)/(n+lambda)
lambda.n<-n+lambda
lambda.npred<-(lambda.n)/(lambda.n+1)
yv3<-dnorm(xv,eta.n,sqrt(sigma0^2/lambda.npred))
mtxt3<-expression(paste('(',eta,',',lambda,')',' = (-1,2)'))
lines(xv,yv3,col='blue')

eta<--1;lambda<-0.01  ##Prior 4
eta.n<-(n*ybar+lambda*eta)/(n+lambda)
lambda.n<-n+lambda
lambda.npred<-(lambda.n)/(lambda.n+1)
yv4<-dnorm(xv,eta.n,sqrt(sigma0^2/lambda.npred))
mtxt4<-expression(paste('(',eta,',',lambda,')',' = (-1,0.01)'))
lines(xv,yv4,col='green')
legend(-2,0.8,c(mtxt1,mtxt2,mtxt3,mtxt4),lty=1,col=c('black','red','blue','green'))
```



Bayesian inference with both μ and σ unknown.

If both μ and σ are unknown we can still consider a factorization of the prior

$$\pi_0(\mu, \sigma) = \pi_0(\mu|\sigma)\pi_0(\sigma)$$

and then note that if we retain the same conditional prior for μ as above

$$\pi_0(\mu|\sigma) \equiv Normal(\eta, \sigma^2/\lambda)$$

then as

$$\pi_n(\mu, \sigma) \propto \prod_{i=1}^{n} f_Y(y_i; \mu, \sigma)\pi_0(\mu|\sigma)\pi_0(\sigma)$$

it must be that

$$\pi_n(\mu|\sigma) \propto \prod_{i=1}^{n} f_Y(y_i; \mu, \sigma)\pi_0(\mu|\sigma)$$

as all other terms are constant in μ. Noting this, we have precisely the same calculation as if σ were known, and it follows that

$$\pi_n(\mu|\sigma) \equiv Normal(\eta_n, \sigma^2/\lambda_n)$$

Therefore it is required only to compute the posterior for σ, $\pi_n(\sigma)$. To do this, we first write an equivalent model in terms of σ^2, and consider the joint posterior

$$\pi_n(\mu, \sigma^2) \propto \prod_{i=1}^{n} f_Y(y_i; \mu, \sigma)\pi_0(\mu|\sigma^2)\pi_0(\sigma^2)$$

which, now retaining all terms in σ, can be written

$$\pi_n(\mu,\sigma) \propto \left(\frac{1}{\sigma^2}\right)^{(n+1)/2} \exp\left\{-\frac{1}{2\sigma^2}\left[(n+\lambda)\left(\mu-\frac{n\overline{y}_n+\lambda\eta}{n+\lambda}\right)^2+\frac{n\lambda}{n+\lambda}(\overline{y}_n-\eta)^2\right]\right\}\pi_0(\sigma^2)$$

where the leading term is comprised of a product of n factors of $(1/\sigma)$ from the likelihood term, and then one factor $(1/\sigma)$ from $\pi_0(\mu|\sigma^2)$ – see equation (2). For convenience we try to choose a prior $\pi_0(\sigma)$ that combines with the other terms in a tractable fashion. If

$$\pi_0(\sigma^2) \propto \left(\frac{1}{\sigma^2}\right)^{\alpha/2+1} \exp\left\{-\frac{\beta}{2\sigma^2}\right\}$$

then a simple calculation will follow; this specification is possible as it corresponds to assuming that

$$\frac{1}{\sigma^2} \sim Gamma(\alpha/2, \beta/2).$$

This distribution for σ^2 is referred to as the *Inverse Gamma* distribution: if $X \sim InvGamma(\alpha,\beta)$ then the pdf for X takes the form

$$f_X(x) = \frac{\beta^\alpha}{\Gamma(\alpha)}\left(\frac{1}{x}\right)^{\alpha+1}\exp\left\{-\frac{\beta}{x}\right\} \qquad x>0$$

for $\alpha, \beta > 0$ – the distribution arises by transformation as the distribution of the quantity $X = 1/Z$ where $Z \sim Gamma(\alpha,\beta)$.

Using this prior specification, we conclude that

$$\pi_n(\mu,\sigma^2) = c\left(\frac{1}{\sigma^2}\right)^{(n+\alpha+1)/2+1} \exp\left\{-\frac{1}{2\sigma^2}\left[(n+\lambda)\left(\mu-\frac{n\overline{y}_n+\lambda\eta}{n+\lambda}\right)^2+\frac{n\lambda}{n+\lambda}(\overline{y}_n-\eta)^2+\sum_{i=1}^n(y_i-\overline{y}_n)^2+\beta\right]\right\}$$

where c is a constant that does not depend on μ or σ. To compute the marginal posterior density for σ, we must integrate out μ from $\pi_n(\mu,\sigma)$; this can be achieved in a straightforward fashion as most of the terms are constant in μ. We have that

$$\begin{aligned}
\pi_n(\sigma^2) &= \int_{-\infty}^{\infty}\pi_n(\mu,\sigma)\,d\mu\\
&= c\left(\frac{1}{\sigma^2}\right)^{(n+\alpha+1)/2+1} \exp\left\{-\frac{1}{2\sigma^2}\left[\frac{n\lambda}{n+\lambda}(\overline{y}_n-\eta)^2+\sum_{i=1}^n(y_i-\overline{y}_n)^2+\beta\right]\right\}\\
&\qquad\times\int_{-\infty}^{\infty}\exp\left\{-\frac{(n+\lambda)}{2\sigma^2}\left(\mu-\frac{n\overline{y}_n+\lambda\eta}{n+\lambda}\right)^2\right\}d\mu\\
&= c\left(\frac{1}{\sigma^2}\right)^{(n+\alpha+1)/2+1} \exp\left\{-\frac{1}{2\sigma^2}\left[\frac{n\lambda}{n+\lambda}(\overline{y}_n-\eta)^2+\sum_{i=1}^n(y_i-\overline{y}_n)^2+\beta\right]\right\}\left(\frac{2\pi\sigma^2}{n+\lambda}\right)^{1/2}
\end{aligned}$$

as the integrand is a Normal density with the normalizing constant removed. Hence

$$\pi_n(\sigma^2) \propto \left(\frac{1}{\sigma^2}\right)^{(n+\alpha)/2+1} \exp\left\{-\frac{1}{2\sigma^2}\left[\frac{n\lambda}{n+\lambda}(\overline{y}_n-\eta)^2+\sum_{i=1}^n(y_i-\overline{y}_n)^2+\beta\right]\right\}$$

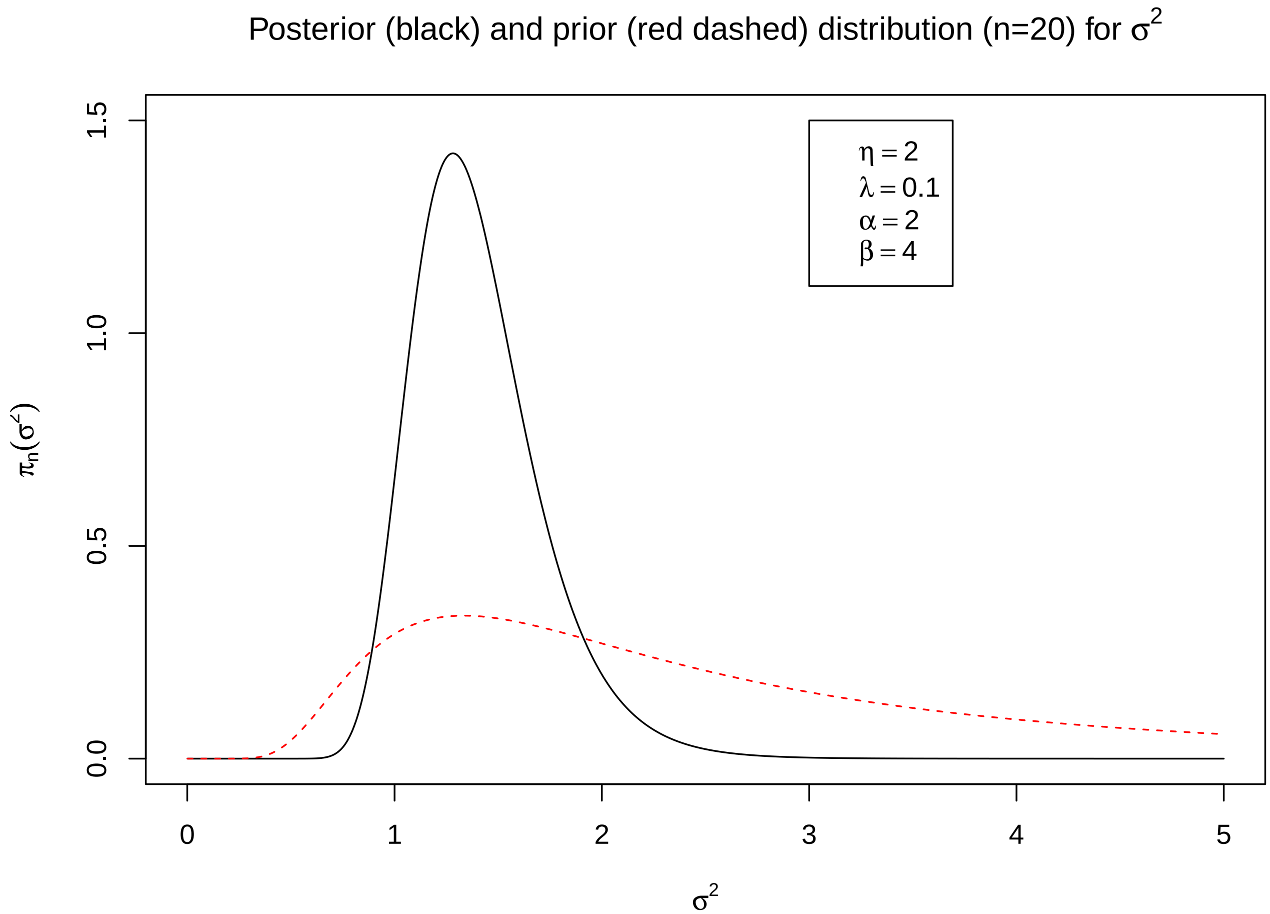
and we can deduce that the posterior distribution of σ^2 is defined by the fact that $1/\sigma^2$ has a $Gamma(a_n/2, b_n/2)$ distribution with parameters

$$a_n = n + \alpha \qquad\qquad b_n = \frac{n\lambda}{n+\lambda}(\overline{y}_n - \eta)^2 + \sum_{i=1}^{n}(y_i - \overline{y}_n)^2 + \beta$$

The following code computes the posterior distribution for σ^2 for a specific choice of prior, namely

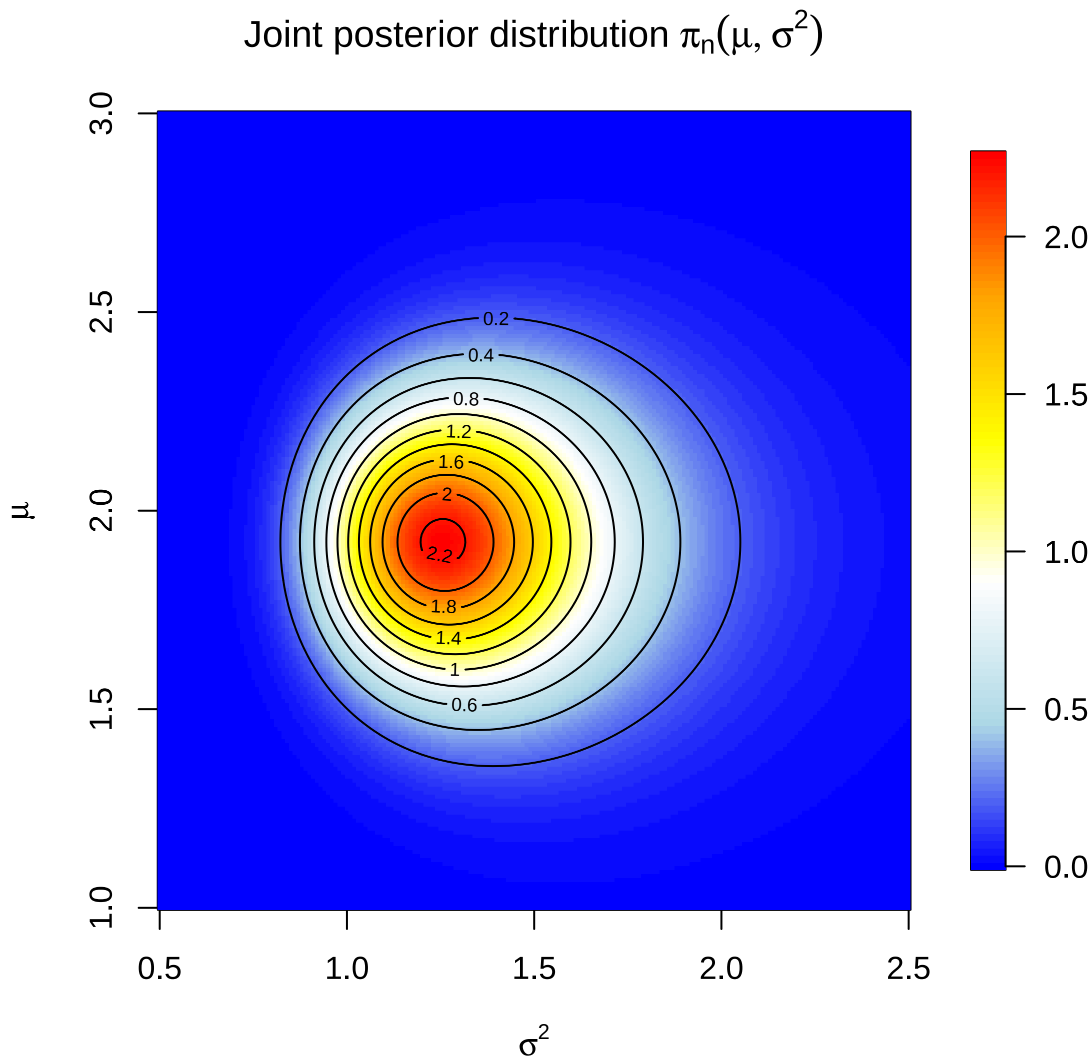
$$\eta = 2 \qquad \lambda = 0.1 \qquad \alpha = 2 \qquad \beta = 4.$$

```
set.seed(234)
dinvgamma<-function(x,a,b,log=FALSE){
    dx<-(b^a/gamma(a))*(1/x)^(a+1)*exp(-b/x)
    if(log){
        dx<-log(dx)
    }
    return(dx)
}
n<-20
theta0<-2; sigma0<-1
y<-rnorm(n,theta0,sigma0)
ybar<-mean(y)
yssq<-sum((y-ybar)^2)
eta<-2
lambda<-0.1
al<-2
be<-4
al.n<-n+al
be.n<-((n*lambda)/(n+lambda))*(ybar-eta)^2+yssq+be
xv<-seq(0,5,by=0.001)
yv<-dinvgamma(xv,al.n,be.n)
plot(xv,yv,type='l',xlab=expression(sigma^2),ylab=expression(pi[n](sigma^2)),
    ylim=range(0,1.5))
yv0<-dinvgamma(xv,al,be)
lines(xv,yv0,col='red',lty=2)
ttxt<-expression(paste("Posterior (black) and prior (red dashed) distribution (n=20) for ",sigma^2
title(ttxt)
legend(3,1.5,c(expression(eta==2),expression(lambda==0.1),
        expression(alpha==2),expression(beta==4)))
```

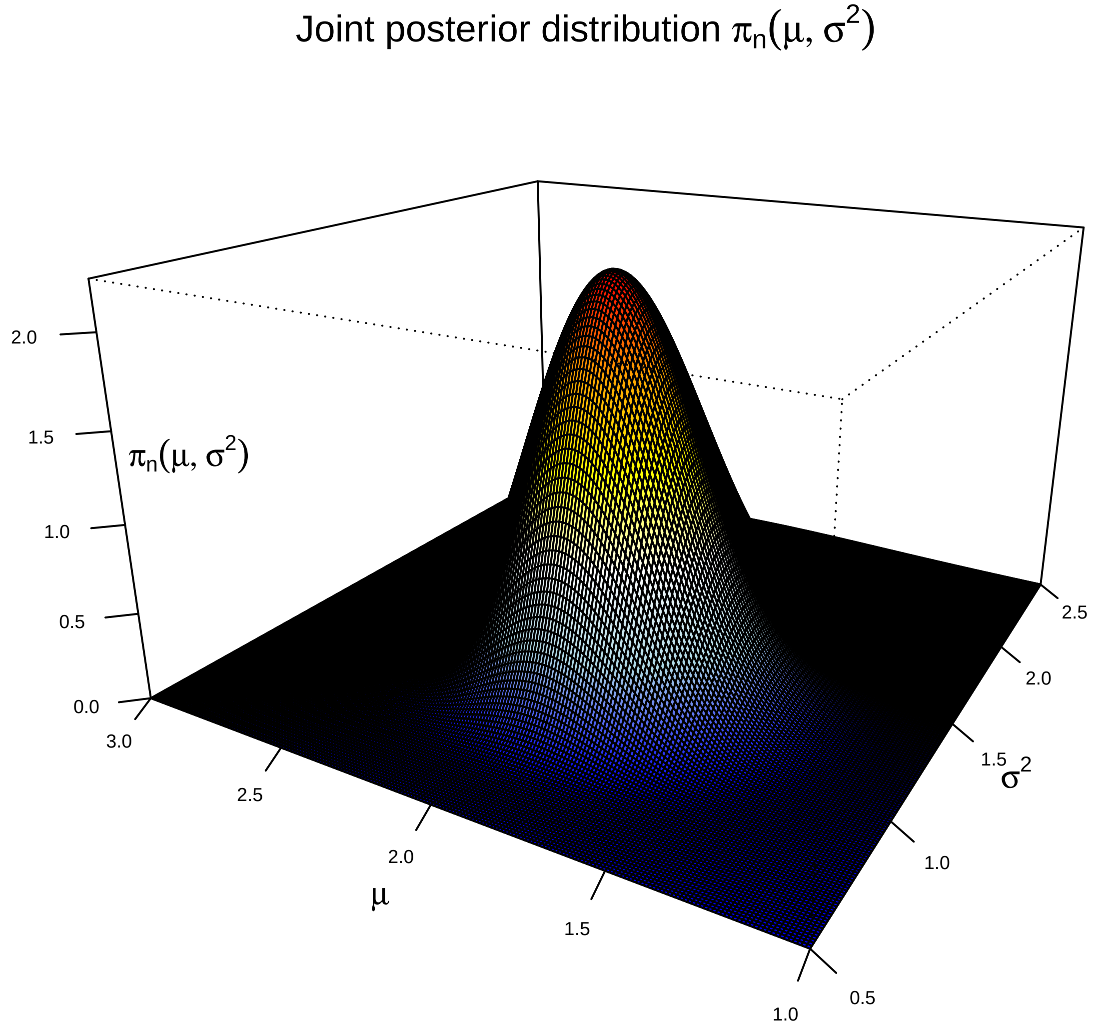


For this prior, we can also compute the joint posterior distribution $\pi_n(\mu, \sigma^2)$ on a suitable grid of points

```
library(fields,quietly=TRUE)
colfunc <- colorRampPalette(c("blue","lightblue","white","yellow","orange","red"))
sigsq.v<-seq(0.5,2.5,by=0.01)
mu.v<-seq(1,3,by=0.01)
pi.n.jt<-matrix(0,nrow=length(sigsq.v),ncol=length(mu.v))
eta.n<-(n*ybar+lambda*eta)/(n+lambda)
lambda.n<-n+lambda
for(i in 1:length(sigsq.v)){
    for(j in 1:length(mu.v)){
         pi.n.jt[i,j]<-dinvgamma(sigsq.v[i],al.n,be.n)*
                        dnorm(mu.v[j],eta.n,sqrt(sigsq.v[i]/lambda.n))
    }
}
par(pty='s')
image.plot(sigsq.v,mu.v,pi.n.jt,col=colfunc(100),
    xlab=expression(sigma^2),ylab=expression(mu))
contour(sigsq.v,mu.v,pi.n.jt,add=T)
title(expression(paste('Joint posterior distribution ',pi[n](mu,sigma^2))))
```




```
z<-pi.n.jt
z.facet.center <- (z[-1,-1] + z[-1,-ncol(z)] + z[-nrow(z),-1] + z[-nrow(z),-ncol(z)])/4
z.facet.range<-cut(z.facet.center, 100)
colvec<-colfunc(100)[z.facet.range]
par(mar=c(2,2,4,0))
persp(sigsq.v,mu.v,pi.n.jt,col=colvec,phi=20,theta=-60,
   ticktype="detailed",expand=0.6,xlab=' ',ylab=' ',zlab=' ', cex.axis=0.6,
   main=expression(paste('Joint posterior distribution ',pi[n](mu,sigma^2))))
text(-0.35,0,expression(pi[n](mu,sigma^2)))
text(-0.2,-0.35,expression(mu))
text(0.3,-0.25,expression(sigma^2))
```



Finally, the marginal posterior distribution for μ can be computed analytically as

$$
\begin{aligned}
\pi_n(\mu) &= \int_0^\infty \pi_n(\mu|\sigma^2)\pi_n(\sigma^2)\, d\sigma^2 \\
&= \int_0^\infty \left(\frac{\lambda_n}{2\pi\sigma^2}\right)^{1/2} \exp\left\{-\frac{\lambda_n}{2\sigma^2}\left(\mu-\eta_n\right)^2\right\} \frac{(b_n/2)^{a_n/2}}{\Gamma(a_n/2)} \left(\frac{1}{\sigma^2}\right)^{a_n/2+1} \exp\left\{-\frac{b_n}{2\sigma^2}\right\}\, d\sigma^2 \\
&= \frac{(b_n/2)^{a_n/2}}{\Gamma(a_n/2)} \left(\frac{\lambda_n}{2\pi}\right)^{1/2} \int_0^\infty \left(\frac{1}{\sigma^2}\right)^{(a_n+1)/2+1} \exp\left\{-\frac{1}{2\sigma^2}\left[\lambda_n\left(\mu-\eta_n\right)^2+b_n\right]\right\}\, d\sigma^2 \\
&= \frac{(b_n/2)^{a_n/2}}{\Gamma(a_n/2)} \left(\frac{\lambda_n}{2\pi}\right)^{1/2} \frac{\Gamma((a_n+1)/2)}{\left\{\frac{1}{2}\left[\lambda_n\left(\mu-\eta_n\right)^2+b_n\right]\right\}^{(a_n+1)/2}} \\
&= \frac{\Gamma((a_n+1)/2)}{\Gamma(a_n/2)\sqrt{\pi}} \left(\frac{\lambda_n}{b_n}\right)^{1/2} \left\{1+\frac{\lambda_n}{b_n}\left(\mu-\eta_n\right)^2\right\}^{-(a_n+1)/2}
\end{aligned}
$$

Often, this density is reparameterized by writing

$$\frac{\lambda_n}{b_n} = \frac{1}{a_n \phi_n}$$

so that

$$\pi_n(\mu) = \frac{\Gamma((a_n+1)/2)}{\Gamma(a_n/2)\sqrt{\pi}} \frac{1}{a_n^{1/2}} \left(\frac{1}{\phi_n}\right)^{1/2} \left\{1 + \frac{1}{a_n} \frac{(\mu - \eta_n)^2}{\phi_n}\right\}^{-(a_n+1)/2}$$

This is the general *Student*-t distribution with a_n degrees of freedom, centered at η_n, with scale parameter $\phi_n^{1/2}$.

Note: As before, we may consider limiting cases of the prior specification where $\lambda \longrightarrow 0$, which induces a prior for μ that becomes more and more diffuse as λ increases, and $\alpha, \beta \longrightarrow 0$ which induces a diffuse prior for σ^2. Under these limiting specifications $\eta_n = \overline{y}_n$ and $\lambda_n = n$, and

$$a_n = n \qquad\qquad b_n = \sum_{i=1}^{n} (y_i - \overline{y}_n)^2$$

so that the resulting posterior is still a well-defined distribution. However, also as before the prior predictive distribution becomes an improper distribution in these limiting cases; for the posterior predictive distribution we have as in (5), for $y_{n+1} \in \mathbb{R}$,

$$\begin{aligned}
f_{Y_{n+1}|Y_1,\ldots,Y_n}(y_{n+1}|y_1,\ldots,y_n) &= \int_0^\infty \int_{-\infty}^\infty f_Y(y_{n+1};\mu,\sigma)\pi_n(\mu|\sigma^2)\pi_n(\sigma^2)\, d\mu\, d\sigma \\
&= \int_0^\infty \left(\frac{\lambda_{n,1}}{2\pi\sigma^2}\right)^{1/2} \exp\left\{-\frac{\lambda_{n,1}}{2\sigma^2}(y_{n+1}-\mu_{n,1})^2\right\} \pi_n(\sigma^2)\, d\sigma \\
&= \left(\frac{\lambda_{n,1}}{2\pi}\right)^{1/2} \frac{(b_n/2)^{a_n/2}}{\Gamma(a_n/2)} \int_0^\infty \left(\frac{1}{\sigma^2}\right)^{(a_n+1)/2+1} \exp\left\{-\frac{1}{2\sigma^2}\left[\lambda_{n,1}(y_{n+1}-\mu_{n,1})^2 + b_n\right]\right\} d\sigma \\
&= \left(\frac{\lambda_{n,1}}{\pi}\right)^{1/2} \frac{(b_n/2)^{a_n/2}}{\Gamma(a_n/2)} 2^{a_n/2} \frac{\Gamma\left(\dfrac{a_n+1}{2}\right)}{\left\{\lambda_{n,1}(y_{n+1}-\mu_{n,1})^2 + b_n\right\}^{(a_n+1)/2}} \\
&= \frac{\Gamma((a_n+1)/2)}{\Gamma(a_n/2)\sqrt{\pi}} \left(\frac{\lambda_{n,1}}{b_n}\right)^{1/2} \left\{1 + \frac{\lambda_{n,1}}{b_n}(y_{n+1}-\mu_{n,1})^2\right\}^{-(a_n+1)/2} \\
&= \frac{\Gamma((a_n+1)/2)}{\Gamma(a_n/2)\sqrt{\pi}} \frac{1}{a_n^{1/2}} \left(\frac{1}{\phi_{n,1}}\right)^{1/2} \left\{1 + \frac{1}{a_n}\frac{(y_{n+1}-\mu_{n,1})^2}{\phi_{n,1}}\right\}^{-(a_n+1)/2}
\end{aligned}$$

which is again a *Student*-t distribution with a_n degrees of freedom, centered at $\mu_{n,1} = \eta_n$, with scale

$$\phi_{n,1}^{1/2} = \sqrt{\frac{b_n}{a_n \lambda_{n,1}}}.$$