MAS3301 Bayesian Statistics

M. Farrow
School of Mathematics and Statistics
Newcastle University

Semester 2, 2008-9

21 Mixture Priors

21.1 Introduction

In earlier lectures we have looked at the use of conjugate prior distributions. We have also looked at the use of numerical methods when the prior is not conjugate. Conjugate priors make the calculations simple but might not represent the genuine prior beliefs. In this lecture we will look at a way to introduce more flexibility into our choice of prior distribution while still retaining some of the convenience of using a conjugate prior.

21.2 Mixture distributions

Suppose that $f_1(x), \ldots, f_J(x)$ are all proper probability density functions. Then we introduce numbers $k_1, \ldots, k_J$ such that $k_j \geq 0$ for all j and $\sum_{j=1}^{J} k_j = 1$. Then

$$f(x) = \sum_{j=1}^{J} k_j f_j(x)$$

is also a proper probability density.

The resulting distribution is called a *mixture distribution.* Provided $J < \infty$ it is a *finite mixture distribution.* In this module we will only consider finite mixtures.

Suppose we assume only that $k_j \geq 0$ for all j. Then $f(x)$ is a proper density if and only if

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

Now

$$\begin{aligned} \int_{-\infty}^{\infty} f(x)\ dx &= \int_{-\infty}^{\infty} \sum_{j=1}^{J} k_j f_j(x)\ dx \\ &= \sum_{j=1}^{J} k_j \int_{-\infty}^{\infty} f_j(x)\ dx \\ &= \sum_{j=1}^{J} k_j \end{aligned}$$

so $f(x)$ is proper if and only if $\sum_{j=1}^{J} = 1$.

21.3 Mixture priors

The use of mixtures as prior distributions allows greater flexibility in the shape of the prior density. If the individual mixture components are conjugate prior distributions then updating is straightforward and the posterior distribution is also a mixture of conjugate distributions.

Consider what happens when we update our beliefs using Bayes' theorem. Suppose we have a prior density $f_j^{(0)}(\theta)$ for a parameter θ and suppose the likelihood is $L(\theta)$. Then our posterior density is

$$f_j^{(1)}(\theta) = \frac{f_j^{(0)}(\theta)L(\theta)}{C_j}$$

where

$$C_j = \int_{-\infty}^{\infty} f_j^{(0)}(\theta)L(\theta)\ d\theta.$$

Now let our prior density for a parameter θ be

$$f^{(0)}(\theta) = \sum_{j=1}^{J} k_j^{(0)} f_j^{(0)}(\theta).$$

Our posterior density is

$$\begin{aligned} f^{(1)}(\theta) &= \frac{\sum_{j=1}^{J} k_j^{(0)} f_j^{(0)}(\theta) L(\theta)}{C} \\ &= \frac{\sum_{j=1}^{J} k_j^{(0)} C_j f_j^{(0)}(\theta) L(\theta)/C_j}{C} \\ &= \frac{\sum_{j=1}^{J} k_j^{(0)} C_j f_j^{(1)}(\theta)}{C} \end{aligned}$$

Hence we require

$$\frac{\sum_{j=1}^{J} k_j^{(0)} C_j}{C} = 1$$

so

$$C = \sum_{j=1}^{J} k_j^{(0)} C_j$$

and the posterior density is

$$f^{(1)}(\theta) = \sum_{j=1}^{J} k_j^{(1)} f_j^{(1)}(\theta)$$

where

$$k_j^{(1)} = \frac{k_j^{(0)} C_j}{\sum_{i=1}^{J} k_i^{(0)} C_i}.$$

21.4 Example: beta-binomial

Suppose we will make an observation from a binomial($n,\ \theta$) distribution. Component j of our mixture prior distribution is a beta($a_j,\ b_j$) distribution.

In particular, suppose that $J = 2$ and our prior density is

$$f^{(0)}(\theta) = 0.5 f_1^{(0)}(\theta) + 0.5 f_2^{(0)}(\theta)$$

where $f_1^{(0)}(\theta)$ is a beta(2,4) density and $f_2^{(0)}(\theta)$ is a beta(4,2) density. Figure 22 shows the two component densities and the overall prior density.

Suppose that our data are $n = 10,\ x = 8,\ n - x = 2$.
Then:

$f_1^{(1)}$: beta$(2+8,\ 4+2) =$ beta$(10,\ 6)$, $f_2^{(1)}$: beta$(4+8,\ 2+2) =$ beta$(12, 4)$

$$C_j = \int_0^1 \frac{\Gamma(a_j + b_j)}{\Gamma(a_j)\Gamma(b_j)} \theta^{a_j - 1}(1-\theta)^{b_j - 1} \begin{pmatrix} n \\ x \end{pmatrix} \theta^x (1-\theta)^{n-x}\ d\theta$$

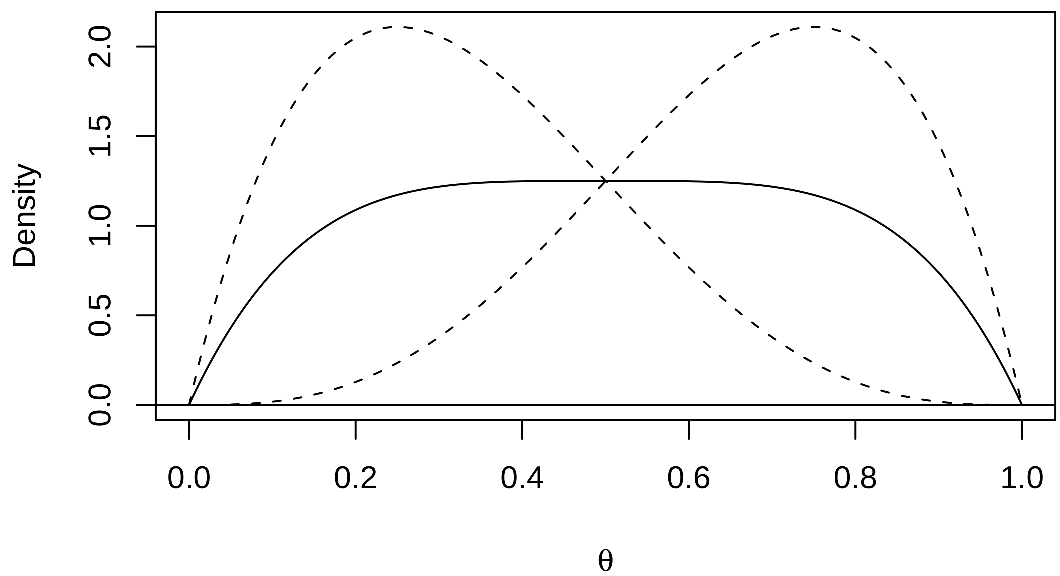

Figure 22: Beta mixture prior density. Dashed lines are components. Solid line is mixture density.

$$
\begin{aligned}
C_j &= \binom{n}{x} \frac{\Gamma(a_j+b_j)}{\Gamma(a_j)\Gamma(b_j)} \frac{\Gamma(a_j+x)\Gamma(b_j+n-x)}{\Gamma(a_j+b_j+n)} \\
&\quad \times \int_0^1 \frac{\Gamma(a_j+b_j+n)}{\Gamma(a_j+x)\Gamma(b_j+n-x)} \theta^{a_j+x-1}(1-\theta)^{b_j+n-x-1}\, d\theta \\
&= \binom{n}{x} \frac{\Gamma(a_j+b_j)}{\Gamma(a_j)\Gamma(b_j)} \frac{\Gamma(a_j+x)\Gamma(b_j+n-x)}{\Gamma(a_j+b_j+n)}
\end{aligned}
$$

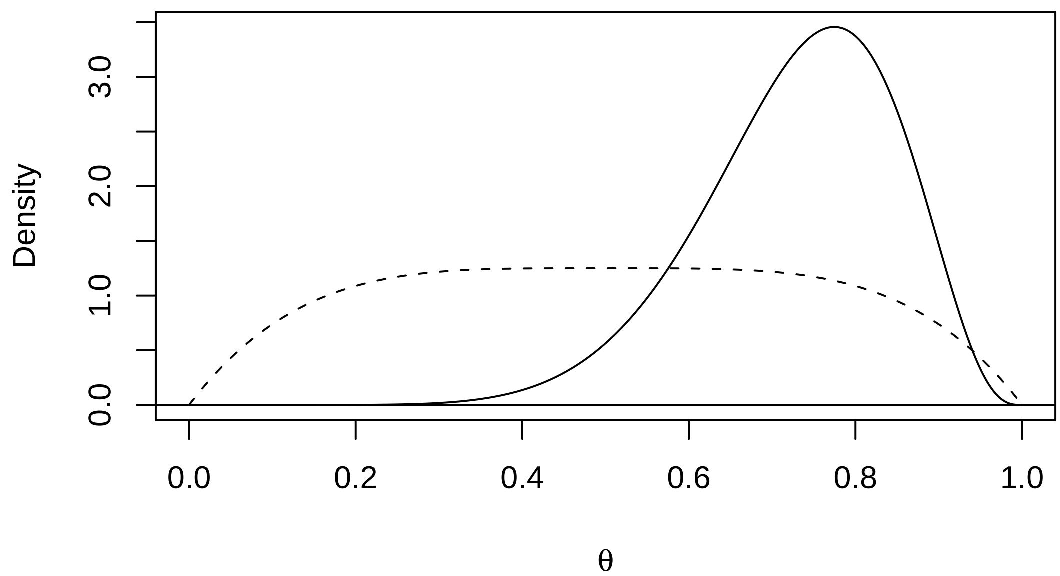

Figure 23: Beta mixture prior density (dashed line) and posterior (solid line).

Now

$$k_j^{(1)} = \frac{k_j^{(0)} C_j}{\sum k_i^{(0)} C_i}$$

but the binomial coefficient

$$\begin{pmatrix} n \\ x \end{pmatrix}$$

cancels and $k_1^{(0)} = k_2^{(0)} = 0.5$ cancel.

$$\begin{aligned} C_1 &= \begin{pmatrix} n \\ x \end{pmatrix} \frac{5!}{1!3!} \frac{9!5!}{15!} = \begin{pmatrix} n \\ x \end{pmatrix} \frac{5 \times 4 \times 5!}{15 \times 14 \times \cdots \times 10} = 6.66 \times 10^{-4} \begin{pmatrix} n \\ x \end{pmatrix} \\ C_2 &= \begin{pmatrix} n \\ x \end{pmatrix} \frac{5!}{3!1!} \frac{11!3!}{15!} = \begin{pmatrix} n \\ x \end{pmatrix} \frac{5!}{15 \times 14 \times 13 \times 12} = 36.63 \times 10^{-4} \begin{pmatrix} n \\ x \end{pmatrix} \end{aligned}$$

So

$$
\begin{aligned}
k_1^{(1)} &= \frac{6.66}{6.66+36.63} = 0.1538 \\
k_2^{(1)} &= \frac{36.63}{6.66+36.63} = 0.8462
\end{aligned}
$$

The posterior density is

$$f^{(1)}(\theta) = 0.1538 f_1^{(1)}(\theta) + 0.8462 f_2^{(1)}(\theta).$$

Figure 23 shows the prior and posterior densities.

MAS3301, Tutorial Example

This is a long example, longer than one examination question, which combines a number of topics, particularly mixture priors and inference for a normal mean.

Suppose that we will observe a sample $Y_1, \dots, Y_n$ from a normal distribution $N(\mu,\ \tau^{-1})$ where μ is unknown but τ is known. Given μ and τ, the observations $Y_1, \dots, Y_n$ are independent.

Our prior distribution for μ is a mixture distribution with density

$$f_0(\mu) = \sum_{j=1}^{J} \pi_{0,j} f_{0,j}(\mu)$$

where $f_{0,j}(\mu)$ is the density of a normal $N(M_{0,j},\ P_{0,j}^{-1})$ distribution and $\pi_{0,j} \geq 0$ for $j = 1, \dots, J$.

The observed data are $y_1, \dots, y_n$.

1. Show that $f_0(\mu)$ is a proper density if and only if $\sum_{j=1}^{J} \pi_{0.j} = 1$.

2. Show that
$$f_{0,j}(\mu) \propto P_{0,j}^{1/2} \phi(P_{0,j}^{1/2}[\mu - M_{0,j}])$$
where
$$\phi(z) = (2\pi)^{-1/2} \exp(z^2/2)$$
is the standard normal probability density function..

3. Show that the posterior distribution of μ is a mixture distribution with a density which can be written in the form
$$f_1(\mu) = \sum_{j=1}^{J} \pi_{1,j} f_{1,j}(\mu)$$
and find expressions for $f_{1,j}(\mu)$ and $\pi_{1,j}$.

4. Suppose that $J = 3$ and that
$$\begin{array}{ccc} \pi_{0,1} = 0.25, & \pi_{0,2} = 0.5, & \pi_{0,3} = 0.25, \\ M_{0,1} = 7.0, & M_{0,2} = 9.0, & M_{0,3} = 11.0, \\ P_{0,1} = 1.0, & P_{0,2} = 0.25, & P_{0,3} = 1.0 \\ n = 20, & \sum_{i=1}^{n} y_i = 206.3, & \tau = 0.4. \end{array}.$$

   (a) Find the prior mean and prior standard deviation of μ.

   (b) Find the posterior mean and posterior standard deviation of μ.

Solution

1.

$$\begin{aligned} \int_{-\infty}^{\infty} f_0(\mu)\ d\mu &= \int_{-\infty}^{\infty} \sum_{j=1}^{J} \pi_{0,j} f_{0,j}(\mu)\ d\mu \\ &= \sum_{j=1}^{J} \pi_{0,j} \int_{-\infty}^{\infty} f_{0,j}(\mu)\ d\mu \\ &= \sum_{j=1}^{J} \pi_{0,j} \end{aligned}$$

since
$$\int_{-\infty}^{\infty} f_{0,j}(\mu)\ d\mu = 1.$$

Therefore $f_0(\mu)$ is a proper density if and only if $\sum_{j=1}^{J} \pi_{0,j} = 1$ (given that $\pi_{0,j} \geq 0$).

2.

$$f_{0,j}(\mu) = (2\pi)^{-1/2} P_{0,j}^{1/2} \exp\left\{-\frac{P_{0,j}}{2}(\mu - M_{0,j})^2\right\} = (2\pi)^{-1/2} P_{0,j}^{1/2} \exp\left\{-\frac{z^2}{2}\right\}$$

where

$$z = P_{0,j}^{1/2}(\mu - M_{0,j}).$$

The result follows.

3. The posterior density is

$$\begin{aligned} f_1(\mu) &= \frac{\sum_{j=1}^J \pi_{0,j} f_{0,j}(\mu) L(\mu)}{C} \\ &= \frac{\sum_{j=1}^J \pi_{0,j} C_j f_{0,j}(\mu) L(\mu)/C_j}{C} \end{aligned}$$

where

$$C_j = \int_{-\infty}^{\infty} f_{0,j}(\mu) L(\mu)\ d\mu$$

so

$$f_1(\mu) = \frac{\sum_{j=1}^J \pi_{0,j} C_j f_{1,j}(\mu)}{C}$$

where

$$f_{1,j}(\mu) = \frac{f_{0,j}(\mu) L(\mu)}{C_j}$$

which is the posterior density given the prior $f_{0,j}(\mu)$.

Hence we require

$$\frac{\sum_{j=1}^J \pi_{0,j} C_j}{C} = 1$$

so

$$C = \sum_{i=1}^{J} \pi_{0,j} C_j$$

and

$$f_1(\mu) = \sum_{j=1}^{J} \pi_{1,j} f_{1,j}(\mu)$$

where

$$\pi_{1,j} = \frac{\pi_{0,j} C_j}{\sum_{i=1}^J \pi_{0,j} C_j}$$

and

$$C_j = \int_{-\infty}^{\infty} f_{0,j}(\mu) L(\mu)\ d\mu.$$

Now

$$f_{0,j}(\mu) = (2\pi)^{-1/2} P_{0,j}^{1/2} \exp\left\{-\frac{P_{0,j}}{2}(\mu - M_{0,j})^2\right\}$$

(we can drop the $(2\pi)^{-1/2}$ since it does not depend on μ or j and cancels) and the likelihood is

$$L(\mu) = \prod_{i=1}^{n} (2\pi)^{-1/2} \tau^{1/2} \exp\left\{-\frac{\tau}{2}(y_i - \mu)^2\right\}$$

(again, we can drop the $(2\pi)^{-1/2}$ and, in fact, the $\tau^{1/2}$) so

$$\begin{aligned} L(\mu) &\propto \prod_{i=1}^{n} \exp\left\{-\frac{\tau}{2}(y_i - \mu)^2\right\} \\ &\propto \exp\left\{-\frac{\tau}{2}\sum_{i=1}^{n}(y_i - \mu)^2\right\} \end{aligned}$$

Now

$$\begin{aligned}
\sum_{i=1}^{n}(y_i-\mu)^2 &= \sum_{i=1}^{n}(y_i-\bar{y}+\bar{y}-\mu)^2 \\
&= \sum_{i=1}^{n}(y_i-\bar{y})^2+n(\bar{y}-\mu)^2+2(\bar{y}-\mu)\sum_{i=1}^{n}(y_i-\bar{y}) \\
&= \sum_{i=1}^{n}(y_i-\bar{y})^2+n(\bar{y}-\mu)^2
\end{aligned}$$

So

$$\begin{aligned}
L(\mu) &\propto \exp\left\{-\frac{\tau}{2}\left(\sum_{i=1}^{n}(y_i-\bar{y})^2+n(\bar{y}-\mu)^2\right)\right\} \\
&\propto \exp\left\{-\frac{n\tau}{2}(\bar{y}-\mu)^2\right\}
\end{aligned}$$

(We can drop

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

since τ is known and this term does not depend on either j or μ).
So

$$\begin{aligned}
f_{0,j}(\mu)L(\mu) &\propto P_{0,j}^{1/2}\exp\left\{-\frac{P_{0,j}}{2}(\mu-M_{0,j})^2\right\}\exp\left\{-\frac{n\tau}{2}(\bar{y}-\mu)^2\right\} \\
&\propto P_{0,j}^{1/2}\exp\left\{-\frac{1}{2}[P_{0,j}(\mu-M_{0,j})^2+n\tau(\bar{y}-\mu)^2]\right\} \\
&\propto P_{0,j}^{1/2}\exp\left\{-\frac{1}{2}[(P_{0,j}+n\tau)\mu^2-2(P_{0,j}M_{0,j}+n\tau\bar{y})\mu]\right\} \\
&\qquad\times\exp\left\{-\frac{1}{2}[P_{0,j}M_{0,j}^2+n\tau\bar{y}^2]\right\} \\
&\propto P_{1,j}^{1/2}\exp\left\{-\frac{P_{i,j}}{2}(\mu^2-2M_{1,j}\mu+M_{1,j}^2)\right\} \\
&\qquad\times\left(\frac{P_{0,j}}{P_{1,j}}\right)^{1/2}\exp\left\{-\frac{P_{1,j}}{2}\left(\frac{P_{0,j}M_{0,j}^2+n\tau\bar{y}^2}{P_{1,j}}-M_{1,j}^2\right)\right\} \\
&\propto (2\pi)^{-1/2}P_{1,j}^{1/2}\exp\left\{-\frac{P_{1,j}}{2}(\mu-M_{1,j})^2\right\} \\
&\qquad\times\left(\frac{P_{0,j}}{P_{1,j}}\right)^{1/2}\exp\left\{-\frac{1}{2}\left(P_{0,j}M_{0,j}^2+P_d\bar{y}^2-P_{1,j}M_{1,j}^2\right)\right\}
\end{aligned}$$

where

$$\begin{aligned}
P_{1,j} &= P_{0,j}+P_d \\
P_d &= n\tau \\
\text{and } M_{1,j} &= \frac{P_{0,j}M_{0,j}+P_d\bar{y}}{P_{1,j}}.
\end{aligned}$$

So, since the second term does not depend on μ (and is therefore a constant in terms of μ),

$$f_{1,j}(\mu)=P_{1,j}^{1/2}\phi(P_{1,j}^{1/2}[\mu-M_{1,j}]).$$

Furthermore

$$C_j=\int_{-\infty}^{\infty}f_{0,j}(\mu)L(\mu)\,d\mu=\left(\frac{P_{0,j}}{P_{1,j}}\right)^{1/2}\exp\left\{-\frac{1}{2}\left(P_{0,j}M_{0,j}^2+P_d\bar{y}^2-P_{1,j}M_{1,j}^2\right)\right\}.$$

4.

$$\bar{y} = \frac{206.3}{20} = 10.315$$

$$P_d = n\tau = 8$$

j	$P_{0,j}$	$M_{0,j}$	$P_{1,j}$	$M_{1,j}$
1	1.00	7.0	$1+8=9$	$(1\times 7+8\times 10.315)/9 = 9.94667$
2	0.25	9.0	$0.25+8=8.25$	$(0.25\times 9+8\times 10.315)/8.25 = 10.27515$
3	1.00	11.0	$1+8=9$	$(1\times 11+8\times 10.315)/9 = 10.39111$

j	$(P_{0.j}/P_{1,j})$	$P_{0,j}M_{0,j}^2 + P_d\bar{y}^2 - P_{1,j}M_{1,j}^2$	C_j	$\pi_{0,j}C_j$	$\pi_{1,j}$
1	0.11111	9.76820	0.00252	0.00063	0.00454
2	0.03030	0.41920	0.14116	0.07058	0.50829
3	0.11111	0.41709	0.27059	0.06765	0.48717
				0.13886	

(a) Prior mean:

$$0.25\times 7 + 0.5\times 9 + 0.25\times 11 = \underline{9.0}$$

$$\mathrm{E}_j(\mu^2) = \mathrm{var}_j(\mu) + [\mathrm{E}_j(\mu)]^2$$

$$\mathrm{E}(\mu^2) = 0.25\times\left[\frac{1}{1}+7^2\right] + 0.5\times\left[\frac{1}{0.25}+9^2\right] + 0.25\times\left[\frac{1}{1}+11^2\right] = 85.5$$

Prior variance

$$\mathrm{var}(\mu) = 85.5 - 9^2 = 4.5$$

Prior standard deviation

$$\text{std.dev.}(\mu) = \sqrt{4.5} = \underline{2.121}$$

(b) Posterior mean:

$$0.00454\times 9.4667 + 0.50829\times 10.27515 + 0.48717\times 10.39111 = \underline{10.330}$$

$$\mathrm{E}_j(\mu^2) = \mathrm{var}_j(\mu) + [\mathrm{E}_j(\mu)]^2$$

$$\begin{aligned}\mathrm{E}(\mu^2) &= 0.00454\times\left[\frac{1}{9}+9.4667^2\right] + 0.50829\times\left[\frac{1}{8.25}+10.27515^2\right] \\ &\quad +0.48717\times\left[\frac{1}{9}+10.39111^2\right] = 106.8323\end{aligned}$$

Posterior variance

$$\mathrm{var}(\mu) = 106.8323 - 10.330^2 = 0.12026$$

Posterior standard deviation

$$\text{std.dev.}(\mu) = \sqrt{0.12026} = \underline{0.347}$$

Notes

1. Alternative form for C_j :

$$
\begin{aligned}
P_{0,j}M_{0,j}^2 + P_d\bar{y}^2 - P_{1,j}M_{1,j}^2 &= P_{1,j}^{-1}\{P_{1,j}P_{0,j}M_{0,j}^2 + P_{1,j}P_d\bar{y}^2 - P_{1,j}^2M_{1,j}^2\} \\
&= P_{1,j}^{-1}\{P_{0,j}^2M_{0,j}^2 + P_dP_{0,j}M_{0,j}^2 + P_{0,j}P_d\bar{y}^2 + P_d^2\bar{y}^2 \\
&\quad -(P_{0,j}M_{0,j} + P_d\bar{y})^2\} \\
&= P_{1,j}^{-1}\{P_dP_{0,j}M_{0,j}^2 + P_dP_{0,j}\bar{y}^2 - 2P_{0,j}P_dM_{0,j}\bar{y}\} \\
&= P_{1,j}^{-1}P_dP_{0,j}\{\bar{y} - M_{0,j}\}^2 \\
&= \left(\frac{P_{0,j} + P_d}{P_{0,j}P_d}\right)^{-1}(\bar{y} - M_{0,j})^2 \\
&= (P_{0,j}^{-1} + P_d^{-1})^{-1}(\bar{y} - M_{0,j})^2 \\
&= P_R(\bar{y} - M_{0,j})^2
\end{aligned}
$$

where

$$P_R = (P_{0,j}^{-1} + P_d^{-1})^{-1}.$$

Also

$$\left(\frac{P_{0,j}}{P_{1,j}}\right)^{1/2} \propto \left(\frac{P_{0,j}P_d}{P_{1,j}}\right)^{1/2} = \left(\frac{P_{0,j} + P_d}{P_{0,j}P_d}\right)^{1/2} = (P_{0,j}^{-1} + P_d^{-1})^{-1/2} = P_R^{1/2}.$$

So we can use, for C_j

$$C_j = P_R^{1/2}\exp\left\{-\frac{P_R}{2}(\bar{y} - M_{0,j})^2\right\}.$$

This form of C_j is just proportional to the original form but, of course, that is all that is required.

2. The graph shows the prior density (dashes) and posterior density (solid). The prior might have been more "sensible" if we had increased the weight on the middle component a little.