STA 214 Mid-Term Exam Spring 1997

Full credit can be gained by correctly answering all questions (the questions are NOT equally weighted). Please make sure to pay careful attention to components of questions asking for comments and discussions, as well as more technical components.

Please do NOT write out the questions on your answer sheets.

The exam lasts exactly two hours.

Question 1

Data $x_1, \ldots, x_n$ follow a normal mixture model $(1-\pi)N(x|\mu_0, v_0) + \pi N(x|\mu_1, v_1)$ with known component variances v_0 and v_1. Assume a specified prior $p(\mu_0, \mu_1)$ on the component means, and, independently, a uniform prior on π, namely $U(\pi|0, 1)$.

(a) A modeller has a problem in which the variances are equal, $v_0 = v_1 = 1$. She wants to use a prior $p(\mu_0, \mu_1) = U(\mu_0|-a, a)U(\mu_1|-a, a)$ for some fixed a. Discuss the resulting identification problem for the component means, and ways of modifying the prior to fix the problem.

(b) In another application, the modeller knows that $v_0 = 1$ and $v_1 = 10$. How does this affect the identification issue for μ_0 and μ_1?

Question 2

An observation x follows the model $p(x|\mu)$ and μ has the mixture prior $(1-\pi)p_0(\mu) + \pi p_1(\mu)$ for some prior densities $p_0(\cdot)$ and $p_1(\cdot)$. As described on the course handout *General framework of mixture priors*, it follows that the resulting posterior is $p(\mu|x) = (1-\pi^*)p_0(\mu|x) + \pi^* p_1(\mu|x)$ where $p_i(\mu|x)$ is the posterior under prior $p_i(\mu)$ for each $i = 0, 1$, and $\pi^* = \pi p_1(x)/\{(1-\pi)p_0(x) + \pi p_1(x)\}$ with $p_i(x) = \int p(x|\mu)p_i(\mu)d\mu$ for $i = 0, 1$.

Suppose $p_0(\mu)$ is continuous, but that $p_1(\mu)$ is a uniform discrete density on a set $\{m_1, \ldots, m_k\}$,

$$p_1(\mu) = \sum_{j=1}^{k} p_j I_{m_j}(\mu)$$

or $Pr(\mu = m_j) = p_j$ for $j = 1, \ldots, k$.

(a) What is $p_1(x)$ in this case?

(b) What is the posterior probability that $\mu = m_1$?

(c) Describe in detail how you would simulate a value of μ from the posterior.

Question 3

In an MDP context, data are independently drawn from sampling densities $p(x_i|\mu_i)$ $(i = 1, \ldots, n)$. For each i, write $\mu^{(i)} = \{\mu_1, \ldots, \mu_{i-1}, \mu_{i+1}, \ldots, \mu_n\}$ for all the μ_j except μ_i. The conditional priors for each μ_i given the rest $\mu^{(i)}$ are

$$p(\mu_i|\mu^{(i)}) = (1-\pi)p_0(\mu_i) + \pi p_1(\mu_i|\mu^{(i)})$$

where $p_0(\cdot)$ is the continuous density of the DP base/prior $G_0(\cdot)$, with $\pi = (n-1)/(\alpha + n - 1)$ and where $p_1(\mu_i|\mu^{(i)})$ is the discrete density with probabilities $1/(n-1)$ at each of the elements of $\mu^{(i)}$.

(a) Using results in Question 2, or otherwise, write down the form of the conditional posterior $p(\mu_i|\mu^{(i)}, x_i)$.

(b) Briefly describe how to implement Gibbs sampling to simulate the posterior for $\{\mu_1, \ldots, \mu_n\}$ given data $\{x_1, \ldots, x_n\}$.

Question 4

An industrial process is being tested to assess whether or not an abrupt increase in process level has occurred. In standard operating mode the level is zero, and measurements made on the process are standard normally distributed about this level, $N(x|0,1)$. An abrupt level increase would change this to $N(x|\mu,1)$ for some $\mu > 0$; note that no level change corresponds to $\mu = 0$. Assume the level increase has prior probability 0.5, and assume a uniform prior $U(\mu|0,a)$ on μ if the level has changed. One observation x is made.

(a) Find the posterior density $p(\mu|x)$ given the observed value of x.

(b) What is the posterior probability (as a function of x) that the level has increased? Show that the corresponding Bayes' factor in favour of a level increase, i.e., in favour of $\mu > 0$ versus $\mu = 0$, is given by

$$\frac{\Phi(a-x)-\Phi(-x)}{a\phi(x)}$$

where $\phi(\cdot)$ is the standard normal density function and $\Phi(\cdot)$ the standard normal c.d.f.

(c) We observe $x = 2.5$, so that a standard one-sided significance test would reject the null hypothesis of no level change at better than the 1% level. Using the fact that

$$\Phi(a-2.5)-\Phi(-2.5) < 13.62\ a\ \phi(2.5)$$

for all $a > 0$, describe how the conclusions of this Bayesian test of $\mu > 0$ versus $\mu = 0$ differ from the significance test, and interpret the evidence that $x = 2.5$ provides in favour or against the level change.

Question 5

Vector observations $y_1, \dots, y_n$ are i.i.d. from the multivariate normal mixture model with density $\sum_{j=1}^{k} \pi_j N(\cdot|\mu_j, \Sigma_j)$. Write $\phi = \{\pi_j, \mu_j, \Sigma_j \ : j = 1, \dots, k\}$ for all model parameters, and introduce classification indicators $z = \{z_1, \dots, z_n\}$ where $z_i = j$ if observation y_i comes from component j.

(a) Given data $y = \{y_1, \dots, y_n\}$ and all model parameters ϕ, show how to calculate the conditional posterior probabilities

$$\pi^*_{ij} = Pr(z_i = j|y, \phi).$$

(b) An investigator wants to classify her data, i.e., to assess which of the components of the mixture each observation is most likely to have arisen from. Assuming the parameters ϕ are known, describe how this may be done.

(c) Suppose now (realistically) that the parameters ϕ are unknown, but that you have a sample from the posterior $p(\phi|y)$. Describe how the classification problem in (b) above is now addressed.

Question 1

(a) The prior and likelihood function are symmetric in μ_0 and μ_1, hence so is the posterior. The margins $p(\mu_i|x)$ are therefore identical, for example, and so inference is obscured. The posterior for π is symmetric about 0.5. The data cannot be classified, all posterior classification probabilities are 0.5.
Modify the prior to impose an identifying constraint; $\mu_0 < \mu_1$ is one example.

(b) The model is identified as $v_1 > v_0$, so no issue arises.

Question 2

(a) $p_1(x) = \sum_{j=1}^{k} p_j p(x|m_j)$, a mixture of the data densities $p(x|\cdot)$ evaluated at the prior support points m_j.

(b) The full posterior is

$$p(\mu|x) = (1-\pi^*)p_0(\mu|x) + \pi^* p_1(\mu|x)$$

with π^* as defined in the statement of the question. Now

$$\begin{aligned} p_1(\mu|x) &= p_1(\mu)p(x|\mu)/p_1(x) \\ &= \sum_{j=1}^{k} p_j p(x|m_j) I_{m_j}(\mu)/p_1(x) \\ &= \sum_{j=1}^{k} q_j I_{m_j}(\mu) \end{aligned}$$

where

$$q_j = p_j p(x|m_j) / \sum_{i=1}^{k} p_i p(x|m_i).$$

Interpret q_j as the posterior probability that $\mu = m_j$ given that it comes from the discrete component of the prior. Then, overall,

$$Pr(\mu = m_j|x) = \pi^* q_j.$$

for each $j = 1, \ldots, k$. For example, $Pr(\mu = m_1|x) = \pi^* q_1$.

(c) First, make a multinomial draw from $\{0, 1, \ldots, k\}$ with probabilities $(1-\pi^*)$, $\pi^* q_1, \ldots, \pi^* q_k$. If 0 is selected, draw μ from $p_0(\mu|x)$; otherwise, if $j > 0$ is selected, set $\mu = m_j$.
An alterative: draw a uniform deviate to choose between 0 or 1, with probability $1-\pi^*$ on 0. If 0 is selected, draw μ from $p_0(\mu|x)$; otherwise, make a multinomial draw from cells $\{1, \ldots, k\}$ with probabilities $q_1, \ldots, q_k$ and set $\mu = m_j$ where j is the cell selected.

Question 3

(a) Use the results stated in Question 2: for fixed i, match $k = n-1$ and, for $j = 1, \ldots, n$ but $j \neq i$, set $m_j = \mu_j$; also $p_j = 1/(n-1)$. Then we have

$$p(\mu|\mu^{(i)}, x_i) = (1-\pi_i^*)p_0(\mu_i|x_i) + \pi_i^* p_1(\mu_i|x_i)$$

where

$$p_0(\mu_i|x_i) = \int p(x_i|\mu_i)p_0(\mu_i)d\mu_i,$$

the prior marginal density of the datum under prior $p_0(\mu_i)$, and

$$p_1(\mu_i|x_i) = \sum_{j\neq i}^{n} q_{ij} I_{\mu_j}(\mu_i)$$

where

$$q_{ij} = p(x_i|\mu_j) / \sum_{r\neq i}^{n} p(x_i|\mu_r).$$

Also, again from the statement of Question 2,

$$\pi^* = \pi p_1(x)/\{(1-\pi)p_0(x) + \pi p_1(x)\}$$

with $\pi = (n-1)/(\alpha + n - 1)$ as given in this case.

(b) Question 2, part (c), describes exactly how to sample the posterior $p(\mu_i|\mu^{(i)}, x_i)$ for any i. We set up Gibbs sampling to iteratively sequence through $i = 1, \ldots, n$, generating a new μ_i from $p(\mu_i|\mu^{(i)}, x)$ where x is all the data. At each step, the most recently sampled values of elements of the conditioning parameters $\mu^{(i)}$ are used. Note that, under the conditional independence of the $(x_i|\mu_i)$, we have $p(\mu_i|\mu^{(i)}, x) \equiv p(\mu_i|\mu^{(i)}, x_i)$.

Question 4

(a) The prior is $p(\mu) = 0.5\{I_0(\mu) + I_{(0,a)}(\mu)/a\}$ so that, by Bayes' theorem,

$$p(\mu|x) = c\ p(x|\mu)p(\mu) = c\ \phi(x-\mu)p(\mu) = c\ \{\phi(x)I_0(\mu) + \phi(\mu - x)I_{(0,a)}(\mu)/a\}.$$

Normalising, $1 = \int\ c\ p(x|\mu)p(\mu)\ d\mu$ implies that $c = 1/\{a\phi(x) + \Phi(a-x) - \Phi(-x)\}$.

(b) Since the prior odds are 1, the Bayes' factor equals the posterior odds ratio, which is

$$\begin{aligned} O(x) &= \frac{Pr(\mu > 0|x)}{Pr(\mu = 0|x)} = \frac{p(x|\mu > 0)}{p(x|\mu = 0)} \\ &= \frac{\int p(x|\mu)U(\mu|0,a)d\mu}{\phi(x)} \\ &= \frac{\Phi(a-x) - \Phi(-x)}{a\phi(x)}. \end{aligned}$$

Then $Pr(\mu > 0|x) = O(x)/(1 + O(x))$.

(c) Using the stated fact we see that, whatever value of a is chosen for the prior, $O(2.5) < 13.62$. This implies that $Pr(\mu > 0|x) < 13.62/14.62 = 0.93$, or $Pr(\mu = 0|x) > 0.07$. Hence, for *any* uniform prior on values of $\mu > 0$, the posterior probability on the null hypothesis of no change in level is greater than 0.07. This is well above the extreme significance level of less that 0.01, indicating the typical "conservative" behaviour of Bayesian testers approaching the problem this way. The significance testing approach is, by Bayesian standards, guarranteed to reject null hypotheses too often. Finally, this is a gross bound: any specific prior, will more strongly favour $\mu = 0$ than 0.07, so the evidence seems relatively mild.

Question 4

(a) Immediately,

$$\pi^*_{ij} \propto \pi_j p(y|z_i = j, \phi) \propto \pi_j p(y_i|z_i = j, \phi) \propto \pi_j N(y_i|\mu_j, \Sigma_j).$$

So, for each $i = 1, \ldots, n$, compute the $j = 1, \ldots, k$ normal densities $N(y_i|\mu_j, \Sigma_j)$ at the data value y_i, find the normalising constant

$$c_i = 1/\sum_{r=1}^{k} \pi_r N(y_i|\mu_r, \Sigma_r)$$

and then we have

$$\pi^*_{ij} = c_i \pi_j N(y_i|\mu_j, \Sigma_j)$$

for each j.

(b) Look at the posterior classification probabilities π^*_{ij}, comparing them across components j.

(c) Now the π^*_{ij} are uncertain; formally $\pi^*_{ij} = \pi^*_{ij}(\phi)$ where ϕ is uncertain. The MC sample of parameters ϕ gives a corresponding MC sample of each of the $\pi^*_{ij}(\phi)$, hence an MC approximation to the full posterior for these probabilities. For example, we might estimate each of them by posterior means $E(\pi^*_{ij}|y)$, approximated by the sample mean of the MC sample values.
