Stat 544 - Spring 2005

First Mid-Term - March 11, 2005
KEY

Read all the questions carefully and start answering the ones for which you know the answer. None of the questions require extensive calculations or derivations. Be precise. Show all your work; partial credit will be given. Each part of a question is worth 10 points, for a total of 100 points.

Good luck!

Problem 1 (Source: WinBUGS)

Hospitals are often ranked on the basis of patient survival after surgery. Mortality rates in 12 hospitals that specialize in cardiac surgery in babies were obtained. Data on each hospital included the number of operations performed during a pre-specified period and the number of deaths resulting from those operations. Data are presented in the table below. In the table, n_i and y_i denote the number of operations and the number of deaths in the ith hospital, respectively.

Hospital	A	B	C	D	E	F	G	H	I	J	K	L
n_i	47	148	119	810	211	196	148	215	207	97	256	360
y_i	0	18	8	46	8	13	9	31	14	8	29	24

(a) What might be an appropriate sampling model for the number of deaths in the ith hospital? Please define the parameter(s) in the model.

A binomial model for y is appropriate. We let

$$y_i \sim B(n_i, \theta_i),$$

for n_i known and θ_i the probability of death in ith hospital.

(b) Assume that all hospitals are *independent*. For any one hospital, derive the conjugate prior distribution for the parameter(s) in the model. [You do not need to choose numerical values for the parameters of the prior.]

Since

$$p(y_i|\theta_i) \propto \theta_i^{y_i}(1-\theta_i)^{n_i-y_i},$$

we know that the conjugate prior for θ_i will have to have the same form. The Beta distribution with parameters (α, β) has this form and thus is conjugate to the binomial likelihood. Thus,

$$p(\theta_i|\alpha, \beta) \propto \theta_i^{\alpha-1}(1-\theta_i)^{\beta-1}.$$

(c) Given your prior in (b), write down the posterior distribution of the unknown parameter(s) of interest for hospital B. What is the posterior mean of the parameter(s)? [The expression for the posterior mean will be a function of the parameters of the prior distribution rather than an actual number.]

$$p(\theta_2|y_2) \propto \text{ Beta}(\alpha + y_2 - 1, \beta + n_2 - y_2 - 1),$$

for $n_2 = 148$ and $y_2 = 18$. The posterior mean of θ_2 is given by

$$E(\theta_2|y_2) = \frac{\alpha + y_2 - 1}{\alpha + \beta + n_2 - 2}.$$

(d) It might be more reasonable to assume that given some hyperparameter, the hospitals are *exchangeable* rather than independent. Formulate a hierarchical model for these data. Carefully define the sampling distribution, the population distribution on the exchangeable parameters, and the joint posterior distribution of all model parameters and hyperparameters.

There are several ways in which we can formulate an exchangeable model on the probabilities of death. One option is to model the logit of the probabilities as a normal random variable. This approach is similar to the one we described in the toxicity in rats example, except that in that example, we had a covariate (dose) associated to the logit of the probability of death.

$$p(y_i|\theta_i) \propto \theta_i^{y_i}(1-\theta_i)^{n_i-y_i}$$

$$
\begin{aligned}
logit(\theta_i) &= b_i \\
p(b_i|\mu, \tau^2) &\propto \exp\{-\frac{1}{2\tau^2}(b_i - \mu)^2\} \\
p(\mu, \tau^2) &\propto Inv - \chi^2(\nu, \tau_0^2),
\end{aligned}
$$

with (ν, τ_0^2) fixed at some value.

Another option is to model the exchangeable probabilities of death as Beta,

$$\theta_i \sim \text{ Beta}(\alpha, \beta)$$

and then place priors on (α, β). There are sevearal ways to do so. Stroud (1994, *Canadian Journal of Statistics*) proposes the following re-parametrization:

$$\alpha = \nu\mu, \text{ and } \beta = \nu(1 - \mu),$$

with $\nu \sim \text{ Exp}(\delta)$, $\mu \sim \text{ Beta}(\phi, \rho)$ and (δ, ϕ, ρ) fixed at some value.

(e) Consider the model you formulated in (d) and suppose that you wish to make inferences about the overall probability of death of babies undergoing heart surgery. On which of the marginal posterior distributions in the model would you base your inferences?

We would be interested in the marginal posterior distribution of μ, which represents the overall mean (across hospitals) probability of death (or logit of the probability of death).

Problem 2

Let $y_i \sim N(\mu, 1)$, with μ unknown. Suppose that you choose a normal prior for μ so that

$$p(\mu|\mu_0, \tau^2) = N(\mu_0, \tau^2),$$

with $\mu_0 = 0$ and $\tau_0^2 = 1$.

(a) You obtain a random sample of size $n_1 = 10$ $\{y_1, y_2, ..., y_{10}\}$ from the population. The sample average $\bar{y}_1$ equals 10. Write down the mean and variance of the posterior distribution of μ.

The posterior distribution $p_1(\mu|\mu_1, \tau_1^2)$ of μ given that $\sigma^2 = 1$ is normal, with mean μ_1 and variance τ_1^2, where

$$
\begin{aligned}
\mu_1 &= \frac{\frac{1}{\tau_0^2}\mu_0 + \frac{n_1}{\sigma^2}\bar{y}}{\frac{1}{\tau_0^2} + \frac{n_1}{\sigma^2}} \\
&= \frac{\mu_0 + 100\tau_0^2}{1 + 10\tau_0^2} \\
&= 9.09 \\
\tau_1^2 &= \frac{1}{\frac{1}{\tau_0^2} + \frac{n_1}{\sigma^2}} \\
&= \frac{\tau_0^2}{1 + 10\tau_0^2} \\
&= 0.091.
\end{aligned}
$$

Thus, μ_1 is a weighted average of the prior mean (0) and the sample mean (10), with weights given by the inverse of the variances (1 and 0.1, respectively).

(b) You now draw an *additional* random sample of size $n_2 = 10$ from the same population. The sample average in this second sample is $\bar{y}_2 = 20$. What is the mean and the variance of the posterior distribution of μ after observing the second set of observations?

Since we are updating the posterior sequentially as we obtain more data, we will use $p_1(\mu|\mu_1, \tau_1^2)$ as the prior. That is, as additional data become available, the posterior obtained in the previous step is used as the prior in the current step. Again, μ will be distributed a posteriori as a normal random variable with mean μ_2 and variance τ_2^2, where

$$
\begin{aligned}
\mu_2 &= \frac{\frac{1}{\tau_1^2}\mu_1 + \frac{n_2}{\sigma^2}\bar{y}_2}{\frac{1}{\tau_1^2} + \frac{n_2}{\sigma^2}} \\
&= 14.29 \\
\tau_2^2 &= \frac{1}{\frac{1}{\tau_1^2} + \frac{n_2}{\sigma^2}} \\
&= 0.048.
\end{aligned}
$$

(c) Suppose now that instead of collecting two samples of 10 observations each, you had obtained all 20 observations at the same time. Obtain the posterior mean of μ and compare the result to the one you obtained in part (b). Any comments?

If instead of proceeding sequentially we had obtained all 20 observations at the same time, then $n = 20$ and $\bar{y} = 15$. The posterior of μ would still be normal with mean m and variance v^2, where now

$$
\begin{aligned}
m &= \frac{\frac{1}{\tau_0^2}\mu_0 + \frac{n}{\sigma^2}\bar{y}}{\frac{1}{\tau_0^2} + \frac{n}{\sigma^2}} \\
&= 14.29 \\
v^2 &= \frac{1}{\frac{1}{\tau_0^2} + \frac{n}{\sigma^2}} \\
&= 0.049.
\end{aligned}
$$

The mean is a weighted average of the prior mean (0) and the sample mean (15) that is based on 20 observations. We are not surprised that the result is the same as in part (b). We discussed sequential updating early in the semester and argued that we would arrive to the same posterior by processing data in sequential 'batches' and using each posterior as the prior in the next step, or by processing all of the data at the same time.

Problem 3

It is well known that a person who has had too much to drink finds it difficult to touch his own nose tip with a finger. In fact, the probability that a person who is drunk will hit the tip of his own nose with a finger is about 0.2. Persons who have not had too much to drink, on the other hand, have a much higher chance (about 0.90) of hitting the tip of their nose with a finger.

You are a policeman in Ames and have been stationed at the corner of Welch Ave. and Lincoln Way on a balmy Friday night. From past experience, you know that about 60% of the drivers in that area have had too much to drink.

Let θ denote the probability that an individual accurately touches the tip of his nose. Write down an informative prior distribution for θ that makes use of the information about θ presented above. Justify your choice of distributional form and prior parameter values.

The population appears to be composed of two groups: the drunk and the sober drivers. The probability of 'success' (touching the tip of your nose) is different in each of the two groups. The appropriate prior for θ is a mixture of distributions, and since θ is a probability, we might consider a mixture of two conjugate distributions such as the Beta.

We have additional information. We know, for example, that among drunks, θ is known to be approximately equal to 0.2 and that among sober people, that probability climbs to perhaps 0.9. Thus, under one of the Beta components θ should have prior expectation approximately equal to 0.2 and under the other, the prior expected value for θ ought to be about 0.9.

If $\theta|\alpha, \beta \sim \text{Beta}(\alpha, \beta)$, the $E(\theta|\alpha, \beta) = \alpha/(\alpha + \beta)$. Assuming that these prior guesses for θ among drunk and sober people are more or less believable, we can fix the 'prior sample size' $\alpha + \beta$ to about 10 in both mixture components. If we do so, the resulting mixture components are

$$\text{Beta}(2, 8) \text{ and } \text{Beta}(9, 1).$$

Since we know that about 60% of the drivers on Welch and Lincoln Way on a balmy Friday evening will have had too much to drink, we can complete the specification of the prior. The proposed prior is

$$p(\theta) \propto 0.6 \text{ Beta}(2, 8) + 0.4 \text{ Beta}(9, 1).$$

A different prior would result from changing $(\alpha + \beta)$ in each of the two components to reflect more or less certainty about the prior guesses for θ in sober and drunken drivers.

Problem 4

Consider the following setup

1. $Y_1 \sim \mathrm{N}(\mu_1, 1)$
2. $Y_2 \sim \mathrm{N}(\mu_2, 1)$
3. $\lambda \in \{0, 1\}$ with $P(\lambda = 1) = \pi$
4. $Y = (1 - \lambda)Y_1 + \lambda Y_2$
5. Y_1 and Y_2 are independent

To complete the model specification, we choose the following priors for the parameters in the model:

$$\begin{aligned} p(\mu_1, \mu_2) &\propto 1 \\ p(\pi) &\propto 1. \end{aligned}$$

Note that λ_i is an indicator variable that indicates whether the ith observation comes from the first or from the second normal. You observe 20 values of Y: $(y_1, \ldots, y_{20})$. The histogram below was obtained from the 20 observations.

You wish to approximate the posterior distributions of μ_1 and μ_2 using the Gibbs sampler. Carefully describe all of the steps for implementing a Gibbs sampler in this problem. [Hint: if at iteration t you knew the values of $\lambda_1, ...\lambda_{20}$, obtaining the conditional distributions of μ_1 and μ_2 would be trivial. Remember the Poisson model with a change point example that we did in class several weeks ago.]

This problem is nicely suited for the Gibbs sampler. Notice that the full conditional distributions are trivial, reflecting the fact that if we knew which observation comes from which component (if we knew the λ_i's), then estimating μ_1, μ_2, π would be very easy.

We begin by writing the joint posterior distribution and then derive the full conditional distributions.

$$p(\mu_1, \mu_2, \lambda_1, ...\lambda_n, \pi | y) \propto \Pi_i[(1 - \lambda_i)\ \mathrm{N}_1(\mu_1, 1) + \lambda_i\ \mathrm{N}_2(\mu_2, 1)]\pi^{\lambda_i}(1 - \pi)^{1-\lambda_i}.$$

$$
\begin{aligned}
p(\mu_1 | \text{ all }) &\propto \Pi_{(i,\lambda_i=0)} \text{ N}_1(\mu_1, 1) \\
&\propto \exp\{-\frac{1}{2} \sum_{(i,\lambda_i=0)} (y_i - \mu_1)^2\} \\
&\propto \text{ N}(\bar{y}_1, 1/n_1),
\end{aligned}
$$

where $n_1 = \sum(1 - \lambda_i)$ and

$$
\bar{y}_1 = \frac{\sum(1 - \lambda_i) y_i}{\sum(1 - \lambda_i)}.
$$

Similarly,

$$
p(\mu_2 | \text{ all }) \propto \text{ N}(\bar{y}_2, 1/n_2),
$$

where $n_2 = 20 - n_1$ and

$$
\bar{y}_2 = \frac{\sum \lambda_i y_i}{\sum \lambda_i}.
$$

The conditional distribution of π is

$$
\begin{aligned}
p(\pi | \text{ all }) &\propto \Pi_i \pi^{\lambda_i} (1 - \pi)^{1-\lambda_i} \\
&\propto \pi^{n_1} (1 - \pi)^{n_2} \\
&\propto \text{ Beta}(n_1 + 1, n_2 + 1).
\end{aligned}
$$

Finally,

$$
\begin{aligned}
\Pr(\lambda_i = 1 | \text{ all }) &\propto \frac{\pi \phi_2(y_i)}{(1 - \pi)\phi_1(y_i) + \pi \phi_2(y_i)} \\
&= \gamma(\mu_1, \mu_2, \pi),
\end{aligned}
$$

where ϕ_i is the density function of a $\text{N}(\mu_i, 1)$, $i = 1, 2$ evaluated at the previous draw of μ_i. Steps then are the following:

1. Start with a guess for $\lambda_1, ..., \lambda_n$.
2. Draw μ_1 and μ_2 from the normal conditionals.
3. Draw π from its conditional.
4. Update the λ_i by setting them to 1 with probability $\gamma(\mu_1, \mu_2, \pi)$.
5. Repeat.