TMA4300 Mod. stat. metoder

Norges teknisk-naturvitenskapelige universitet
Institutt for matematiske fag

Løsningsforslag - Eksamen mai 2009

Problem 1

a) We have

$$f(x;\mu) = \frac{\tau}{\sqrt{2\pi}} \exp\left\{-\frac{\tau^2}{2}\left(x^2 - 2x\mu + \mu^2\right)\right\}$$
$$= \frac{\tau \exp\left\{-\frac{\tau^2 x^2}{2}\right\}}{\sqrt{2\pi}} \exp\left\{\mu\tau^2 x - \frac{\tau^2\mu^2}{2}\right\}$$

Thus, we may choose for example

$$\underline{\underline{a(x) = \frac{\tau \exp\left\{-\frac{\tau^2 x^2}{2}\right\}}{\sqrt{2\pi}} \quad , \quad \phi(\mu) = \mu \quad , \quad t(x) = \tau^2 x \quad \text{and} \quad b(\mu) = -\frac{\tau^2\mu^2}{2}.}}$$

The conjugate prior distribution becomes

$$\pi(\mu) \propto \exp\left\{\phi(\mu)\alpha + b(\mu)\beta\right\} = \exp\left\{\mu\alpha - \frac{\tau^2\mu^2}{2}\beta\right\}$$

We see that the exponent is a second order function of μ. Thus, the conjugate prior is a normal distribution.

b) With

$$\pi(\mu) = \frac{r}{\sqrt{2\pi}} \exp\left\{-\frac{r^2}{2}\left(\mu - \nu\right)^2\right\}$$

the posterior distribution becomes

$$\pi(\mu|x_1,\ldots,x_n) \propto \pi(\mu)\prod_{i=1}^n f(x_i;\mu) \propto \exp\left\{-\frac{r^2}{2}(\mu-\nu)^2 - \sum_{i=1}^n \frac{\tau^2}{2}(x_i-\mu)^2\right\}$$
$$= \exp\left\{-\frac{r^2}{2}(\mu^2 - 2\mu\nu + \nu^2) - \frac{\tau^2}{2}\sum_{i=1}^n (x_i^2 - 2x_i\mu + \mu^2)\right\}$$

$$\propto \exp\left\{-\frac{1}{2}\left[\left(r^2+n\tau^2\right)\mu^2 - 2\left(r^2\nu+\tau^2\sum_{i=1}^n x_i\right)\mu\right]\right\}.$$

As expected from the results in **a**) we see that this is a normal distribution. To find $\mathrm{E}[\mu|x_1,\ldots,x_n]$ and $\mathrm{Var}[\mu|x_1,\ldots,x_n]$ define $\widetilde{\nu}=\mathrm{E}[\mu|x_1,\ldots,x_n]$ and $\widetilde{r}^2 = 1/\mathrm{Var}[\mu|x_1,\ldots,x_n]$. We then must have

$$\pi(\mu|x_1,\ldots,x_n) \propto \exp\left\{-\frac{\widetilde{r}^2}{2}\left(\mu-\widetilde{\nu}\right)^2\right\} \propto \exp\left\{-\frac{1}{2}\left[\widetilde{r}^2\mu^2 - 2\widetilde{r}^2\widetilde{\nu}\mu\right]\right\}.$$

Thus, we must have

$$\widetilde{r}^2 = r^2+n\tau^2 \quad \text{and} \quad \widetilde{r}^2\widetilde{\nu} = r^2\nu + \tau^2\sum_{i=1}^n x_i.$$

Solving with respect to $\widetilde{\nu}$ and $\widetilde{r}^2$ we get

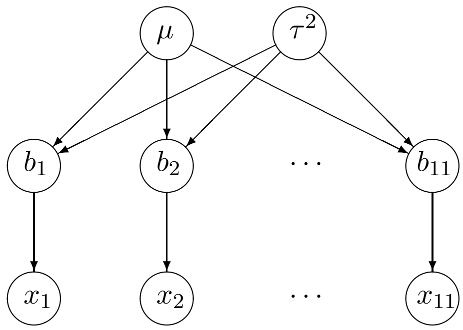
$$\widetilde{\nu} = \frac{r^2\nu+\tau^2\sum_{i=1}^n x_i}{r^2+n\tau^2} \quad \text{and} \quad \widetilde{r}^2 = r^2+n\tau^2$$

and we get

$$\underline{\underline{\mathrm{E}[\mu|x_1,\ldots,x_n] = \frac{r^2\nu+\tau^2\sum_{i=1}^n x_i}{r^2+n\tau^2} \quad \text{and} \quad \mathrm{Var}[\mu|x_1,\ldots,x_n] = \frac{1}{r^2+n\tau^2}.}}$$

Problem 2

a) The graphical model becomes



The full posterior distribution for μ is

$$\pi(\mu|\tau^2, b_1,\ldots,b_{11},x_1,\ldots,x_{11}) \propto \pi(\mu)\prod_{i=1}^{11}\pi(b_i;\mu,\tau^2),$$

where $\pi(\mu)$ is a $\mathrm{N}(\nu, 1/r^2)$ distribution and $\pi(b_i; \mu, \tau^2)$ is a $\mathrm{N}(\mu, 1/\tau^2)$ distribution. This is the same situation as discussed in Problem 1, except the $x_1, \ldots, x_n$ in Problem 1 is now replaced by $b_1, \ldots, b_{11}$. Thus, we have

$$\mu | \tau^2, b_1, \ldots, b_{11}, x_1, \ldots, x_{11} \sim \mathrm{N}\left(\frac{r^2\nu + \tau^2 \sum_{i=1}^{11} b_i}{r^2 + 11\tau^2}, \frac{1}{r^2 + 11\tau^2}\right).$$

Using this as a proposal distribution is a Gibbs step, in which case the acceptance probability becomes equal to one.

The full conditional distribution for τ^2 becomes

$$\pi(\tau^2 | \mu, b_1, \ldots, b_{11}, x_1, \ldots, x_{11}) \propto \pi(\tau^2) \prod_{i=1}^{11} \pi(b_i | \mu, \tau^2)$$

$$\propto \left(\tau^2\right)^{\alpha-1} \exp\left\{-\frac{\tau^2}{\beta}\right\} \prod_{i=1}^{11} \tau \exp\left\{-\frac{\tau^2}{2}(b_i - \mu)^2\right\}$$

$$= \left(\tau^2\right)^{\alpha+11/2-1} \exp\left\{-\frac{\tau^2}{\left(1/\beta + \frac{1}{2}\sum_{i=1}^{11}(b_i - \mu)^2\right)^{-1}}\right\}$$

Comparing this expression with the density of a gamma distribution we see that this is a gamma distribution with parameters

$$\underline{\underline{\widetilde{\alpha} = \alpha + \frac{11}{2} \quad \text{and} \quad \widetilde{\beta} = \frac{1}{\frac{1}{\beta} + \frac{1}{2}\sum_{i=1}^{11}(b_i - \mu)^2}.}}$$

b) The density of the proposal distribution is

$$q(\widetilde{b}_i | b_i) = \begin{cases} \frac{1}{2a} & \text{for } \widetilde{b}_i \in [b_i - a, b_i + a], \\ 0 & \text{otherwise.} \end{cases}$$

The acceptance probability is only of interest when $\widetilde{b}_i \in [b_i - a, b_i + a]$, in which case $q(\widetilde{b}_i | b_i) = 1/(2a) = q(b_i | \widetilde{b}_i)$ and the acceptance probability becomes

$$\alpha(\widetilde{b}_i | b_i) = \min\left\{1, \frac{\pi(b_1, \ldots, b_{i-1}, \widetilde{b}_i, b_{i+1}, \ldots, b_{11}, \mu, \tau^2 | x_1, \ldots, x_{11})}{\pi(b_1, \ldots, b_{i-1}, b_i, b_{i+1}, \ldots, b_{11}, \mu, \tau^2 | x_1, \ldots, x_{11})} \cdot \frac{q(b_i | \widetilde{b}_i)}{q(\widetilde{b}_i | b_i)}\right\}$$

$$= \min\left\{1, \frac{\pi(\widetilde{b}_i | \mu, \tau^2)\pi(x_i | \widetilde{b}_i)}{\pi(b_i | \mu, \tau^2)\pi(x_i | b_i)}\right\}$$

$$= \underline{\underline{\min\left\{1, \exp\left\{-\frac{\tau^2}{2}\left(\left(\widetilde{b}_i - \mu\right)^2 - (b_i - \mu)^2\right)\right\}\frac{\widetilde{p}_i^{x_i}(1-\widetilde{p}_i)^{n_i - x_i}}{p_i^{x_i}(1-p_i)^{n_i-x_i}}\right\}}},$$

where $p_i = 1/(1+e^{b_i})$ and $\widetilde{p}_i = 1/(1+e^{\widetilde{b}_i})$.

With proposal distribution $\widetilde{b}_i \sim \text{Unif}[b_i - 2a, b_i + a]$ the density of the proposal distribution becomes

$$q(\widetilde{b}_i|b_i) = \begin{cases} \frac{1}{3a} & \text{for } \widetilde{b}_i \in [b_i - 2a, b_i + a], \\ 0 & \text{otherwise.} \end{cases}$$

When $\widetilde{b}_i \in [b_i - a, b_i + a]$ we then have $q(\widetilde{b}_i|b_i) = 1/(3a) = q(b_i|\widetilde{b}_i)$ and the acceptance probability is given by the same formula as above. When $\widetilde{b}_i \in [b_i - 2a, b_i - a)$ we have $q(b_i|\widetilde{b}_i) = 0$ and thereby $\alpha(\widetilde{b}_i|b_i) = 0$. Thus, using computation time to propose values $\widetilde{b}_i \in [b_i - 2a, b_i - a)$ is a waste of computing time.

c) Of the three values for a tried it is preferable to use $a = 1.0$ as the estimated autocorrelation function is falling the fastest in this case. For $a = 4$ we can also observe that the simulated Markov chain has a very low acceptance probability for b_1, which is not a favourable property.

To decide the value of a one should also have studied corresponding plots for the other simulated values, $b_2, \ldots, b_{11}, \mu$ and τ^2. It is also natural to monitor the acceptance rates (for the b_i's) with the goal to get the acceptance rates close to 0.234.

d) First one needs to find the length of the burn-in phase of the chain. This can be done by output analysis. Assume the Markov chain has (essentially) converged after $T < M$ iterations.

1. $\text{E}[p_i|x_1, \ldots, x_{11}]$ can then be estimated by

$$\widehat{\text{E}}[p_i|x_1, \ldots, x_{11}] = \frac{1}{M - T + 1}\sum_{m=T}^{M}\frac{e^{b_i^m}}{1 + e^{b_i^m}}$$

2. $\text{Prob}(p_i < p_j|x_1, \ldots, x_n)$ can be estimated by

$$\widehat{\text{Prob}}(p_i < p_j|x_1, \ldots, x_n) = \frac{1}{M-T+1}\sum_{m=T}^{M} I\left(\frac{e^{b_i^m}}{1+e^{b_i^m}} < \frac{e^{b_j^m}}{1+e^{b_j^m}}\right)$$

$$= \frac{1}{M-T+1}\sum_{m=T}^{M} I\left(b_i^m < b_j^m\right)$$

3. $\text{Prob}(p_i < \min_{j\neq i} p_j | x_1, \ldots, x_{11})$ can be estimated by

$$\widehat{\text{Prob}}(p_i < \min_{j\neq i} p_j | x_1, \ldots, x_n) = \frac{1}{M-T-1} \sum_{m=T}^{M} I\left(b_i^m < \min_{j\neq i} b_j^m\right)$$

To find the best hospital given our data let us consider the outcome of a new surgery in each of the hospitals. Let y_i denote the result of the next surgery in hospital H_i, $y_i = 1$ if this surgery results in a death and $y_i = 0$ otherwise. To find the best hospital we then need to minimise

$$\text{Prob}(y_i = 1 | x_1, \ldots, x_{11}) = \int_{-\infty}^{\infty} \text{Prob}(y_i = 1 | p_i, x_1, \ldots, x_{11}) \pi(p_i | x_1, \ldots, x_{11}) \text{d}p_i$$

$$= \int_{-\infty}^{\infty} \text{Prob}(y_i = 1 | p_i) \pi(p_i | x_1, \ldots, x_{11}) \text{d}p_i = \int_{-\infty}^{\infty} p_i \pi(p_i | x_1, \ldots, x_{11}) \text{d}p_i = \text{E}[p_i | x_1, \ldots, x_{11}].$$

Thus, our estimated best hospital is the hospital with the lowest value for $\widehat{\text{E}}[p_i | x_1, \ldots, x_{11}]$.

Problem 3

a) Our estimated model is

$$x_i^\star \sim \text{bin}(n_i, \widehat{p}_i).$$

The bootstrap replication of $\widehat{p}_i$ becomes

$$\widehat{p}_i^\star = \frac{x_i^\star}{n_i}.$$

The ideal bootstrap estimator for the standard deviation of $\widehat{p}_i$ is thereby

$$\underline{\underline{\text{SD}_{\widehat{F}_{\text{par}}}[\widehat{p}_i^\star].}}$$

Compute first the corresponding variance,

$$\text{Var}_{\widehat{F}_{\text{par}}}[\widehat{p}_i^\star] = \text{Var}_{\widehat{F}_{\text{par}}}\left[\frac{x_i^\star}{n_i}\right] = \frac{\text{Var}_{\widehat{F}_{\text{par}}}[x_i^\star]}{n_i^2} = \frac{n_i \widehat{p}_i (1-\widehat{p}_i)}{n_i^2} = \frac{\widehat{p}_i (1-\widehat{p}_i)}{n_i}.$$

The ideal bootstrap estimator for the standard deviation of $\widehat{p}_i$ is thereby

$$\underline{\underline{\text{SD}_{\widehat{F}_{\text{par}}}[\widehat{p}_i^\star] = \sqrt{\frac{\widehat{p}_i (1-\widehat{p}_i)}{n_i}}.}}$$

b) The pseudo code becomes:

1. For $b = 1, \ldots, B$ and $i = 1, \ldots, 11$ generate $x_i^{\star b} \sim \text{bin}(n_i, \widehat{p}_i)$ independently.
2. For $b = 1, \ldots, B$ and $i = 1, \ldots, 11$ compute
$$\widehat{p}_i^{\star b} = \frac{x_i^{\star b}}{n_i}.$$
3. For each of $b = 1, \ldots, B$ order $\widehat{p}_1^{\star b}, \ldots, \widehat{p}_{11}^{\star b}$ from smallest to largest and let $\widehat{r}_i^{\star b}$ denote the rank of $\widehat{p}_i^{\star b}$ for $i = 1, \ldots, 11$.
4. For each of $i = 1, \ldots, 11$ order $\widehat{r}_i^{\star 1}, \ldots, \widehat{r}_i^{\star B}$ from smallest to largest and let $\widehat{r}_i^{\star (k)}$ denote the k'th smallest value.
5. For each of $i = 1, \ldots, 11$, the $(1 - \alpha) \cdot 100\%$ percentile interval for r_i is
$$\underline{\underline{[\widehat{r}_i^{\star (B\alpha/2)}, \widehat{r}_i^{\star (B(1-\alpha/2))}]}}.$$

That r_i takes only integer values is not important for the procedure except that it then requires that the interval is defined as a closed interval.