Petter Mostad
Applied Mathematics and Statistics
Chalmers

Suggested solutions for
MSA101 / MVE187 Computational methods for Bayesian statistics
Exam 27 October 2018

1. (a) The likelihood $\pi(X \mid \theta)$ is non-zero for $\theta \geq X$. As the prior is non-zero for $\theta \geq M$ and $\pi(\theta \mid X) \propto \pi(X \mid \theta)\pi(\theta)$, we get that the posterior $\pi(\theta \mid X)$ is nonzero for $\theta \geq \max(M, X)$.

   (b) If $\theta \sim \text{Pareto}(M, \alpha)$, we get, for $\theta \geq \max(M, X)$,

$$\pi(\theta \mid X) \propto_\theta \pi(X \mid \theta)\pi(\theta) \propto_\theta \frac{1}{\theta} \cdot \theta^{-(\alpha+1)} = \theta^{-(\alpha+1+1)}$$

   This means that $\theta \mid X \sim \text{Pareto}(\max(X, M), \alpha + 1)$, so the posterior is in the same family as the prior, and conjugacy is proved.

   (c) We get, for $X > 0$,

$$\pi(X) = \frac{\pi(X \mid \theta)\pi(\theta)}{\pi(\theta \mid X)} = \frac{\frac{1}{\theta} \cdot \alpha M^\alpha \theta^{-(\alpha+1)}}{(\alpha+1)(\max(M, X))^{\alpha+1}\theta^{-(\alpha+1+1)}} = \frac{\alpha}{\alpha+1} \cdot \frac{M^\alpha}{(\max(M, X))^{\alpha+1}}$$

2. Assume we want to simulate from a density $\pi(x) = Cf(x)$. Assume there is another density $g(x)$ and a constant M such that $Mg(x) \geq Cf(x)$ for all x for which $\pi(x) > 0$. Rejection sampling then means sampling from $\pi(x)$ using the the following steps:

   (a) Sample x from the density $g(x)$.

   (b) Sample u from the Uniform(0, 1) density.

   (c) If $u > \frac{Cf(x)}{Mg(x)}$ go back to step (a); otherwise, return x as the sampled value.

   Note that the algorithm depends on C/M and not on C and M separately. Thus, it can be performed also when the factor C is unknown, as long as one can find M/C so that $M/C \cdot g(x) \geq f(x)$.

3. (a)

(b) We may compute

$$\begin{aligned}
f(\theta) &= C' + \log(\pi(y \mid \theta)\pi(\theta)) \\
&= C' + \log\left[\prod_{i=1}^{3}\prod_{j=1}^{3}\text{Normal}(y_{ij};\mu_i,\tau_i^{-1})\cdot\prod_{i=1}^{3}\text{Gamma}(\tau_i;2,\beta)\right. \\
&\quad \left.\cdot\prod_{i=1}^{3}\text{Normal}(\mu_i;0,1)\cdot\text{Gamma}(\beta;2,2)\right] \\
&= C'' + \sum_{i=1}^{3}\sum_{j=1}^{3}\log\left[\tau_i^{1/2}\exp\left(-\frac{\tau_i}{2}(y_{ij}-\mu_i)^2\right)\right] + \sum_{i=1}^{3}\log\left[\beta^2\tau_i\exp(-\beta\tau_i)\right] \\
&\quad + \sum_{i=1}^{3}\log\left[\exp\left(-\frac{1}{2}\mu_i^2\right)\right] + \log(\beta\exp(-2\beta)) \\
&= C'' + \frac{3}{2}\sum_{i=1}^{3}\log\tau_i - \frac{1}{2}\sum_{i=1}^{3}\tau_i\sum_{j=1}^{3}(y_{ij}-\mu_i)^2 + 6\log\beta + \sum_{i=1}^{3}\log\tau_i - \beta\sum_{i=1}^{3}\tau_i \\
&\quad -\frac{1}{2}\sum_{i=1}^{3}\mu_i^2 + \log\beta - 2\beta \\
&= C'' + \frac{5}{2}\sum_{i=1}^{3}\log\tau_i - \frac{1}{2}\sum_{i=1}^{3}\tau_i\sum_{j=1}^{3}(y_{ij}-\mu_i)^2 - \frac{1}{2}\sum_{i=1}^{3}\mu_i^2 \\
&\quad +7\log\beta - \beta\left(2+\sum_{i=1}^{3}\tau_i\right)
\end{aligned}$$

(c) The acceptance probability is given by

$$p = \min\left(1, \frac{\pi(\theta^{new} \mid y)q(\theta \mid \theta^{new})}{\pi(\theta \mid y)q(\theta^{new} \mid \theta)}\right)$$

where q denotes the proposal function. The proposal function is symmetric in the μ_i variables, but not in the remaining variables. Thus we get

$$\begin{aligned}
\frac{q(\theta \mid \theta^{new})}{q(\theta^{new} \mid \theta)} &= \frac{\text{Exponential}(\beta;2)\prod_{i=1}^{3}\text{Exponential}(\tau_i;2)}{\text{Exponential}(\beta^{new};2)\prod_{i=1}^{3}\text{Exponential}(\tau_i^{new};2)} \\
= \quad & \frac{\exp\left(-2\beta-2\tau_1-2\tau_2-2\tau_3\right)}{\exp\left(-2\beta^{new}-2\tau_1^{new}-2\tau_2^{new}-2\tau_3^{new}\right)}.
\end{aligned}$$

We can use the function f computed in (a) to obtain $\frac{\pi(\theta^{new}|y)}{\pi(\theta|y)} = \exp\left(f(\theta^{new}) - f(\theta)\right)$. This gives us

$$p = \min\left(1, \exp\left(f(\theta^{new}) - f(\theta) + 2(\beta^{new}-\beta) + 2\sum_{i=1}^{3}(\tau_i^{new}-\tau_i)\right)\right).$$

(d) The conditional distributions for each variable given the other variables and the data can most easily be read off from the logposterior computed in (b), at least for β and the τ_i. We get

$$\begin{aligned}\beta \mid \ldots &\sim \text{Gamma}\left(8, 2 + \sum_{i=1}^{3} \tau_i\right) \\ \tau_i \mid \ldots &\sim \text{Gamma}\left(\frac{7}{2}, \beta + \frac{1}{2}\sum_{j=1}^{3}(y_{ij} - \mu_i)^2\right)\end{aligned}$$

We may do the same thing for the μ_i, but it may be easier to use the formula for conjugate updating of normal distributions with fixed precisions. We then get

$$\mu_i \mid \cdots \sim \text{Normal}\left(\frac{0 \cdot 1 + \tau_i y_{i1} + \tau_i y_{i2} + \tau_i y_{i3}}{1 + 3\tau_i}, \frac{1}{1 + 3\tau_i}\right) \sim \text{Normal}\left(\frac{3\tau_i \overline{y}_{i\cdot}}{1 + 3\tau_i}, \frac{1}{1 + 3\tau_i}\right)$$

where $\overline{y}_{i\cdot} = \sum_{j=1}^{3} y_{ij}$.

4. (a) In the graph, both X and Y are decendants of Z_1. This means that there may be a dependency between them via this variable, so X and Y are not necessarily independent. One may also say that X and Y are not d-separated, as there is an active trail X, A_1, Z_1, Y.

   (b) X and Y are conditionally independent given Z_1, i.e., it is true that $X \coprod Y \mid Z_1$. As soon as Z_1 is observed, there is no dependency between Y and Z as they are not directly linked; the variables A_2, A_3, A_4, Z_1 make no difference as they are not observed (or conditioned on). One may also say that X and Y are d-separated given Z_1 as there is then no active trail from X to Y.

   (c) X and Y are not necessarily conditionally independent given Z_1 and Z_2, i.e., it is not necessarily true that $X \coprod Y \mid \{Z_1, Z_2\}$. The reason is that the conditioning on the Z_2 creates a dependency between X and Y, as Z_2 is a decendant of A_1 and Y, and X is a decendant of A_1. One may also say that X and Y are not d-separated given $\{Z_1, Z_2\}$, as there is an active trail X, A_1, A_2, Z_2, A_3, Y.

5. (a) We have

$$\begin{aligned}\log(\pi(x_0, \ldots, x_T, y_0, \ldots, y_T \mid \tau)) &= \log\left[\prod_{i=0}^{T} \pi(y_i \mid x_i, \tau) \prod_{i=1}^{T} \pi(x_i \mid x_{i-1})\pi(x_0)\right] \\ &= C + \sum_{i=0}^{T} \log(\pi(y_i \mid x_i, \tau)) \\ &= C' + \sum_{i=0}^{T} \log\left(\tau^{1/2} \exp(-\frac{\tau}{2}(y_i - x_i)^2)\right) \\ &= C' + \frac{T+1}{2} \log\tau - \frac{\tau}{2}\sum_{i=0}^{T}(y_i - x_i)^2\end{aligned}$$

(b) The expectation becomes

$$\mathrm{E}\left[\log(\pi(x_0,\dots,x_T,y_0,\dots,y_T \mid \tau))\right] = C' + \frac{T+1}{2}\log\tau - \frac{\tau}{2}\sum_{i=0}^{T}\left(y_i^2 - 2y_i\,\mathrm{E}[x_i] + \mathrm{E}\left[x_i^2\right]\right),$$

so what we need is to compute the expectations $\mathrm{E}[x_i]$ and $\mathrm{E}\left[x_i^2\right]$ for all i. This can be done with the Forward-Backward algorithm, using τ fixed at some value τ^{old}. The Forward part will calculate recursively, for $i = 0,\dots,T$, the probability mass functions $\pi(x_i \mid y_0,\dots,y_i)$, while the Backward part will calculate recursively, for $i = T-1,\dots,0$, the probability mass functions $\pi(y_{i+1},\dots,y_T \mid x_i)$. Then, the marginal distribution for each x_i can be computed using

$$\pi(x_i \mid y_0,\dots,y_T) \propto \pi(y_{i+1},\dots,y_T \mid x_i)\pi(x_i \mid y_0,\dots,y_i)$$

and from this probability mass function the two expectations $\mathrm{E}[x_i]$ and $\mathrm{E}\left[x_i^2\right]$ can be computed.

(c) As $\log(\pi(\tau)) = \log(\exp(-\tau)) = -\tau$, we have

$$\begin{aligned}
& \mathrm{E}\left[\log(\pi(x_0,\dots,x_T,y_0,\dots,y_T \mid \tau))\right] + \log(\pi(\tau)) \\
= \; & C' + \frac{T+1}{2}\log\tau - \frac{\tau}{2}\sum_{i=0}^{T}\left(y_i^2 - 2y_i\,\mathrm{E}[x_i] + \mathrm{E}\left[x_i^2\right]\right) - \tau
\end{aligned}$$

One way to find the τ maximizing this expression is to recognize that exponentiating it gives a function proportional to the density for a

$$\mathrm{Gamma}\left(\frac{T+1}{2}, 1 + \frac{1}{2}\sum_{i=0}^{T}\left(y_i^2 - 2y_i\,\mathrm{E}[x_i] + \mathrm{E}\left[x_i^2\right]\right)\right)$$

density. Such a density has mode given by

$$\hat{\tau} = \frac{\frac{T+1}{2}}{1 + \frac{1}{2}\sum_{i=0}^{T}\left(y_i^2 - 2y_i\,\mathrm{E}[x_i] + \mathrm{E}\left[x_i^2\right]\right)} = \frac{T+1}{2 + \sum_{i=0}^{T}\left(y_i^2 - 2y_i\,\mathrm{E}[x_i] + \mathrm{E}\left[x_i^2\right]\right)}$$

(d) The algorithm presented is deterministic, so it should never give different results when started at the same point. However, when started at different points, it may give different results, as it is an optimization algorithm, converging to a local, and not necessarily a global, optimum.

6. Assume that $i < j$. That the joint distribution is a Gaussian Markov Random Field implies that the density can be written on the form

$$\pi(x_1,\dots,x_n) = \exp(-f(x_1,\dots,x_n))$$

where $f(x_1, \ldots, x_n)$ is a quadratic polynomial, i.e., we can write

$$f(x_1, \ldots, x_n) = \sum_{k=1}^{n} a_k x_k^2 + \sum_{1 \leq k \leq s \leq n} b_{ks} x_k x_s + \sum_{k=1}^{n} c_k x_k + d$$

where a_k, b_{ks}, c_k, and d are real numbers. Thus we can also write

$$\pi(x_1, \ldots, x_n) = \left[\prod_{k=1}^{n} \exp(-a_k x_k^2)\right]\left[\prod_{1 \leq k \leq s \leq n} \exp(-b_{ks} x_k x_s)\right]\left[\prod_{k=1}^{n} \exp(-c_k x_k)\right] \exp(-d).$$

In the Markov network for the model, there is no line between variables X_i and X_j if and only if the density can be written as a product where no factors depend on both x_i and x_j. We see that this happens if and only if the number b_{ij} is zero.

The density of a Gaussian Markov Random Field can also be written as

$$\pi(x) \propto \exp\left(-\frac{1}{2}(x-\mu)^t P(x-\mu)\right)$$

where $x = (x_1, \ldots, x_n)$, μ is a vector of real numbers, and P is the symmetric precision matrix. Comparing the two representations of the density, we see that $\frac{1}{2}(p_{ij} + p_{ji}) = b_{ij}$, so that $p_{ij} = b_{ij}$, as $p_{ij} = p_{ji}$ because of symmetry. Thus there is no line between variables X_i and X_j if and only if $p_{ij} = 0$.