Petter Mostad
Applied Mathematics and Statistics
Chalmers

Suggested solutions for
MSA101 / MVE187 Computational methods for Bayesian statistics
Exam 21 October 2017

1. (a) We get

$$\pi(p \mid x = 3) \propto \pi(x = 3 \mid p)\pi(p) = (1-p)^3 p = p^{2-1}(1-p)^{4-1}.$$

Thus the posterior is a Beta(2, 4) distribution.

(b) If $x \mid p \sim \text{Geometric}(p)$ and $p \sim \text{Beta}(\alpha, \beta)$, then

$$\pi(p \mid x) \propto \pi(x \mid p)\pi(p) \propto (1-p)^x p p^{\alpha-1}(1-p)^{\beta-1} = p^{\alpha+1-1}(1-p)^{\beta+x-1}$$

so $p \mid x \sim \text{Beta}(\alpha + 1, \beta + x)$, and the Beta family is conjugate to the Geometric distribution.

(c) If p has a uniform prior on $[0, 1]$, the posterior given x is $\text{Beta}(2, 1 + x)$ and the posterior given x and y is $\text{Beta}(3, 1 + x + y)$. Thus

$$\begin{aligned}
\pi(y \mid x) &= \frac{\pi(y \mid p)\pi(p \mid x)}{\pi(p \mid x, y)} \\
&= \frac{\text{Geometric}(y; p)\,\text{Beta}(p; 2, 1 + x)}{\text{Beta}(p; 3, 1 + x + y)} \\
&= \frac{(1-p)^y p \frac{\Gamma(2+1+x)}{\Gamma(2)\Gamma(1+x)} p^1 (1-p)^x}{\frac{\Gamma(3+1+x+y)}{\Gamma(3)\Gamma(1+x+y)} p^2 (1-p)^{x+y}} \\
&= \frac{\Gamma(3 + x)\Gamma(3)\Gamma(1 + x + y)}{\Gamma(4 + x + y)\Gamma(2)\Gamma(1 + x)}
\end{aligned}$$

When $x = 3$ this becomes

$$\pi(y \mid x) = \frac{\Gamma(6)\Gamma(3)\Gamma(4 + y)}{\Gamma(7 + y)\Gamma(2)\Gamma(4)} = \frac{40}{(4 + y)(5 + y)(6 + y)}$$

2. Note first that $\text{Gamma}(1, \beta) = \text{Exponential}(\beta)$. As the Exponential distribution has a cumulative distribution function that is easy to compute, we may simulate from $\text{Exponential}(\beta)$ by simulating $U \sim \text{Uniform}[0, 1]$ and computing $-\log(U)/\beta$. Thus, when α is an integer, we may simulate from $\text{Gamma}(\alpha, \beta)$, by simulating $U_1, \ldots, U_\alpha \sim \text{Uniform}[0, 1]$ and computing

$$-\frac{1}{\beta}\sum_{i=1}^{\alpha} \log(U_i)$$

3. (a) We get

$$\begin{aligned}
a_0 &= \Pr(x_0 = 1 \mid y_0) \\
&= \frac{\Pr(y_0 \mid x_0 = 1)\Pr(x_0 = 1)}{\Pr(y_0 \mid x_0 = 1)\Pr(x_0 = 1) + \Pr(y_0 \mid x_0 = 0)\Pr(x_0 = 0)} \\
&= \frac{0.8 \cdot 0.1}{0.8 \cdot 0.1 + 0.3 \cdot 0.9} = \frac{8}{35}
\end{aligned}$$

Also,

$$\begin{aligned}
& \Pr(x_1 = 1 \mid y_0) \\
= \; & \Pr(x_1 = 1 \mid x_0 = 1)\Pr(x_0 = 1 \mid y_0) + \Pr(x_1 = 1 \mid x_0 = 0)\Pr(x_0 = 0 \mid y_0) \\
= \; & 0.4 \cdot \frac{8}{35} + 0.2 \cdot \left(1 - \frac{8}{35}\right) = \frac{43}{175}
\end{aligned}$$

and thus (assuming $y_0 = 1$ and $y_1 = 0$)

$$\begin{aligned}
a_1 &= \Pr(x_1 = 1 \mid y_0, y_1) \\
&= \frac{\Pr(y_1 \mid x_1 = 1)\Pr(x_1 = 1 \mid y_0)}{\Pr(y_1 \mid x_1 = 1)\Pr(x_1 = 1 \mid y_0) + \Pr(y_1 \mid x_1 = 0)\Pr(x_1 = 0 \mid y_0)} \\
&= \frac{0.8 \cdot \frac{43}{175}}{0.8 \cdot \frac{43}{175} + 0.3 \cdot \left(1 - \frac{43}{175}\right)} = \frac{86}{185}
\end{aligned}$$

(Full points were given to those who used the right formulas without completing the numerical calculations.)

(b) We get

$$\begin{aligned}
b_{T-1} &= \Pr(y_T \mid x_{T-1} = 1) \\
&= \Pr(y_T \mid x_T = 1)\Pr(x_T = 1 \mid x_{T-1} = 1) + \Pr(y_T \mid x_T = 0)\Pr(x_T = 0 \mid x_{T-1} = 1) \\
&= 0.8 \cdot 0.4 + 0.3 \cdot 0.6 = 0.5.
\end{aligned}$$

(c) For $i = 0, \dots, T-1$, we can write

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

Thus we get

$$\begin{aligned}
\pi(x_i = 1 \mid y_0, \dots, y_t) &= \frac{\pi(x_i = 1 \mid y_0, \dots, y_i)\pi(y_{i+1}, \dots, y_T \mid x_i = 1)}{\sum_{j=0}^{1} \pi(x_i = j \mid y_0, \dots, y_0)\pi(y_{i+1}, \dots, y_T \mid x_i = j)} \\
&= \frac{a_i b_i}{a_i b_i + (1 - a_i)(1 - b_i)}
\end{aligned}$$

4. (a) For the posterior we have

$$
\begin{aligned}
& \pi((\lambda_1,\ldots,\lambda_n,\beta \mid x_{11},\ldots,x_{nm}) \\
\propto\ & \pi(x_{11},\ldots,x_{nm} \mid \lambda_1,\ldots,\lambda_n)\pi(\lambda_1,\ldots,\lambda_n \mid \beta)\pi(\beta) \\
\propto\ & \left[\prod_{i=1}^{n}\prod_{j=1}^{m}\lambda_i \exp(-\lambda_i x_{ij})\right]\left[\prod_{i=1}^{n}\frac{\beta^4}{\Gamma(4)}\lambda_i^{4-1}\exp(-\beta\lambda_i)\right]\beta^{3-1}\exp(-4\beta)
\end{aligned}
$$

Thus the logarithm of the posterior density becomes, up to an additive constant,

$$
\begin{aligned}
& 2\log(\beta)-4\beta+\sum_{i=1}^{n}\left[4\log(\beta)+3\log(\lambda_i)-\beta\lambda_i+\sum_{j=1}^{m}\left[\log(\lambda_i)-\lambda_i x_{ij}\right]\right] \\
=\ & (m+3)\sum_{i=1}^{n}\log(\lambda_i)-\sum_{i=1}^{n}\lambda_i\left(\sum_{j=1}^{m}x_{ij}\right)-\beta\sum_{i=1}^{n}\lambda_i+(4n+2)\log(\beta)-4\beta
\end{aligned}
$$

(b) Fixing all values except λ_i, the logarith of the posterior becomes, up to an additive constant,

$$
(m+3)\log(\lambda_i)-\lambda_i\sum_{j=1}^{m}x_{ij}-\beta\lambda_i
$$

From this we can read off that the conditional distribution to be used for λ_i in the Gibbs sampling is

$$
\text{Gamma}\left(m+4,\beta+\sum_{j=1}^{m}x_{ij}\right)
$$

Fixing all values except β we get

$$
(4n+2)\log(\beta)-\left(4+\sum_{i=1}^{n}\lambda_i\right)\beta
$$

from which we get that the conditional distribution for β is

$$
\text{Gamma}\left(4n+3,4+\sum_{i=1}^{n}\lambda_i\right)
$$

A Gibbs sampler for this model would initiate the simulation with reasonable values for $\lambda_1,\ldots,\lambda_n,\beta$: For example we could set

$$
\lambda_i=\frac{1}{m}\sum_{j=1}^{m}x_{ij}
$$

and then

$$
\beta=\frac{1}{n}\sum_{i=1}^{n}\lambda_i
$$

The algorithm would then iterate between simulating the λ_i and β according to the conditional distributions found above.

(c) We could extend the simulation by simulating values in the Gibbs sampler for all x_{ij} that are censored. Specifically, censored x_{ij} should be simulated from the truncated Exponential distribution with parameter λ_i, truncated so that $x_{ij} > 10$. With the x_{ij} simulated in this way, the remaining Gibbs sampling steps could be performed as above.

More formally, let c_{ij} be the censored data, so that $c_{ij} = x_{ij}$ when $x_{ij} < 10$ and $c_{ij} = 10$ when $x_{ij} \geq 10$. The full posterior then gets an extra factor

$$\prod_{i=1}^{n} \prod_{j=1}^{m} I[x_{ij} = c_{ij}]^{I[c_{ij}<10]} I[x_{ij} \geq 10]^{I[c_{ij}=10]}.$$

Removing the factors not containing x_{ij} from the posterior, we get that $x_{ij} = c_{ij}$ when $c_{ij} < 10$ and

$$\pi(x_{ij} \mid \ldots) \propto I[x_{ij} \geq 10]\lambda_i \exp(-\lambda_i x_{ij})$$

when $c_{ij} = 10$. Thus, in the Gibbs sampling, any censored x_{ij} should be simpulated from an Exponential distribution with parameter λ_i truncated to be greater than or equal to 10. In other words, one may simulate from an Exponential distribution with parameter λ_i and then add 10.

5. (a) Assume you want to simulate from a density proportional to $f(x)$ and that

$$f(x) = \prod_{i=1}^{n} g_i(x)$$

for some non-negative functions $g_1(x), \ldots, g_n(x)$. Define instrumental variables $y_1, \ldots, y_n$ with

$$y_i \mid x \sim \text{Uniform}[0, g_i(x)]$$

Then the joint density can be written

$$\pi(x, y_1, \ldots, y_n) \propto \prod_{i=1}^{n} g_i(x) \prod_{i=1}^{n} \frac{I(0 \leq y_i \leq g_i(x))}{g_i(x)} = \prod_{i=1}^{n} I(0 \leq y_i \leq g_i(x))$$

Thus a Gibbs sampler will iterate between sampling the y_i from the uniform densities given above, and sampling x from the uniform distribution on the set

$$\bigcap_{i=1}^{n} \{x : y_i \leq g_i(x)\}$$

(b) In this case, we can use

$$g_1(x) = \exp\left(-(x+1)^2\right)$$

and

$$g_2(x) = \frac{1}{3 + x^4}$$

Indeed, for positive x, we get that $g_1'(x) = \exp(-(x+1)^2)(-2(x+1)) < 0$ and $g_2'(x) = -(3+x^4)^{-2}4x^3 < 0$, so both functions are strictly decreasing. We see that $y_1 \leq \exp\left(-(x+1)^2\right)$ is equivalent to $x \leq \sqrt{-\log(y_1)} - 1$ and that $y_2 \leq 1/(3+x^4)$ is equivalent to $x \leq (1/y_2 - 3)^{1/4}$. Thus we simulate

$$x \mid y_1, y_2 \sim \text{Uniform}\left[0, \min\left(\sqrt{-\log(y_1)} - 1, (1/y_2 - 3)^{1/4}\right)\right]$$

6. (a) We get

$$\begin{aligned}
& \log\left(\pi\left(y_1, \ldots, y_n, X_1, \ldots, X_n \mid \theta\right)\right) \\
= \; & \log\left(\prod_{i=1}^{n}\left[\left((1-\theta)\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}y_i^2\right)\right)^{I(X_i=0)}\left(\theta\frac{1}{\pi(1+y_i^2)}\right)^{I(X_i=1)}\right]\right) \\
= \; & \sum_{i=1}^{n}\left[I(X_i=0)\left(\log(1-\theta) - \frac{1}{2}\log(2\pi) - \frac{1}{2}y_i^2\right) + \right. \\
& \left. I(X_i=1)\left(\log(\theta) - \log(\pi) - \log(1+y_i^2)\right)\right]
\end{aligned}$$

(b) We have

$$\frac{\Pr[X_i = 1 \mid y_1, \ldots, y_n, \theta']}{\Pr[X_i = 0 \mid y_1, \ldots, y_n, \theta']} = \frac{\Pr[y_i \mid X_i = 1]}{\Pr[y_i \mid X_i = 0]} \cdot \frac{\Pr[X_i = 1 \mid \theta']}{\Pr[X_i = 0 \mid \theta']} = \frac{\frac{1}{\pi(1+y_i^2)}}{\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}y_i^2\right)} \cdot \frac{\theta'}{1-\theta'}$$

Thus

$$w_i = \Pr[X_i = 1 \mid y_1, \ldots, y_n, \theta'] = \frac{\frac{1}{\pi(1+y_i^2)}\theta'}{\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}y_i^2\right)(1-\theta') + \frac{1}{\pi(1+y_i^2)}\theta'}$$

(c) We get

$$\begin{aligned}
Q(\theta \mid \theta') &= E_{\theta'}\left[\log(\pi(y_1, \ldots, y_n, X_1, \ldots, X_n \mid \theta))\right] \\
&= E_{\theta'}\left[\sum_{i=1}^{n}\left[I(X_i=0)\left(\log(1-\theta) - \frac{1}{2}\log(2\pi) - \frac{1}{2}y_i^2\right) + \right.\right. \\
&\quad \left.\left. I(X_i=1)\left(\log(\theta) - \log(\pi) - \log(1+y_i^2)\right)\right]\right] \\
&= \sum_{i=1}^{n}\left[(1-w_i)\left(\log(1-\theta) - \frac{1}{2}\log(2\pi) - \frac{1}{2}y_i^2\right) + \right. \\
&\quad \left. w_i\left(\log(\theta) - \log(\pi) - \log(1+y_i^2)\right)\right]
\end{aligned}$$

(d) From (c) we get that the value of $Q(\theta \mid \theta')$ is, except for an additive term not depending on θ,

$$\log(1-\theta)\sum_{i=1}^{n}(1-w_i) + \log(\theta)\sum_{i=1}^{n}w_i$$

Differentiating with respect to θ, setting to zero, and solving, gives

$$\theta = \frac{1}{n} \sum_{i=1}^{n} w_i$$

Thus this value maximizes $Q(\theta, | \theta')$.

(e) The EM algorithm would start with a reasonable estimate for θ and for the X_i. Then, one would iterate between computing the w_i as in (b) and computing the θ as in (d) until convergence.