Suggested solutions to extra exercises MSA101/MVE187, autumn 2017

Petter Mostad

October 19, 2017

1. (a) We have

$$\begin{aligned} & \Pr(x_0,\dots,x_k,x_{k+1},y_0,\dots,y_k,y_{k+1}) \\ = \;& \Pr(y_{k+1}\mid x_{k+1})\Pr(x_{k+1}\mid x_k)\Pr(x_0,\dots,x_k,y_0,\dots,y_k) \\ = \;& \exp(-\lambda_{x_{k+1}})\frac{\lambda_{x_{k+1}}^{y_{k+1}}}{y_{k+1}!}0.8^{I[x_k=x_{k+1}]}0.1^{I[x_k\neq x_{k+1}]}\Pr(x_0,\dots,x_k,y_0,\dots,y_k)\end{aligned}$$

   To maximize this for fixed x_k and x_{k+1}, we may use the values of $x_0,\dots,x_{k-1}$ optiming the last factor for this x_k. Thus, the above is equal to

$$\exp(-\lambda_{x_{k+1}})\frac{\lambda_{x_{k+1}}^{y_{k+1}}}{y_{k+1}!}0.8^{I[x_k=x_{k+1}]}0.1^{I[x_k\neq x_{k+1}]}b_{k,x_k}$$

   Now, for each possible value of x_{k+1}:

   - Compute the above expression for the different values of x_k.
   - Store the maximum of these values as $b_{k+1,x_{k+1}}$
   - Store the value of x_i giving the maximum as $c_{k+1,x_{k+1}}$.

   (b) First, we find the j maximizing $b_{T,j}$, and set $x_T=j$. Then, for each $i=T-1,\dots,0$, we set

$$x_i = c_{i+1,x_{i+1}}$$

2. (a) Using that

$$X_{i+1}\mid X_i \sim \text{Normal}(X_i,\frac{1}{\tau_x})$$

   and

$$X_i\mid y_0,\dots,y_i \sim \text{Normal}(a_i,\frac{1}{t_i})$$

   we get that

$$X_{i+1}\mid y_0,\dots,y_i \sim \text{Normal}(a_i,\frac{1}{\tau_x}+\frac{1}{t_i}) = \text{Normal}(a_i,\frac{1}{\tau_x t_i/(\tau_x+t_i)})$$

Further, using also that

$$y_{i+1} \mid X_{i+1} \sim \text{Normal}(X_{i+1}, \frac{1}{\tau_y})$$

and the standard formula for updating a normal-normal conjugacy, we get

$$X_{i+1} \mid y_0, \ldots, y_{i+1} \sim \text{Normal}\left(\frac{y_{i+1}\tau_y + a_i\tau_x t_i/(\tau_x + t_i)}{\tau_y + \tau_x t_i/(\tau_x + t_i)}, \frac{1}{\tau_y + \tau_x t_i/(\tau_x + t_i)}\right)$$

and thus that

$$a_{i+1} = \frac{y_{i+1}\tau_y + a_i\tau_x t_i/(\tau_x + t_i)}{\tau_y + \tau_x t_i/(\tau_x + t_i)}$$

and

$$t_{i+1} = \tau_y + \tau_x t_i/(\tau_x + t_i)$$

(b) We get

$$\begin{aligned}
& \pi(y_i, y_{i+1}, \ldots, y_T \mid x_i) \\
= \quad & \pi(y_i \mid x_i)\pi(y_{i+1}, \ldots, y_T \mid x_i) \\
\propto \quad & \exp\left(-\frac{\tau_y}{2}(y_i - x_i)^2\right)\exp\left(-\frac{s_i}{2}(b_i - x_i)^2\right) \\
\propto \quad & \exp\left(-\frac{\tau_y + s_i}{2}\left(\frac{\tau_y y_i + s_i b_i}{\tau_y + s_i} - x_i\right)^2\right) \\
\propto \quad & \text{Normal}\left(x_i; \frac{\tau_y y_i + s_i b_i}{\tau_y + s_i}, \frac{1}{\tau_y + s_i}\right)
\end{aligned}$$

(c) As

$$x_i \mid x_{i-1} \sim \text{Normal}(x_{i-1}, \frac{1}{\tau_x})$$

we get, as in (a), and using the result of (b) above,

$$\pi(y_i, \ldots, y_T \mid x_{i-1}) \propto \text{Normal}\left(x_{i-1}; \frac{\tau_y y_i + s_i b_i}{\tau_y + s_i}, \frac{1}{\tau_y + s_i} + \frac{1}{\tau_x}\right)$$

and thus

$$b_{i-1} = \frac{\tau_y y_i + s_i b_i}{\tau_y + s_i}$$

and

$$s_{i-1} = 1/\left(\frac{1}{\tau_y + s_i} + \frac{1}{\tau_x}\right)$$

(d) Bayes formula gives

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

With the prior Normal$(x_i; a_i, \frac{1}{t_i})$ and the likelihood Normal$(x_i; b_i, \frac{1}{s_i})$ we get the posterior

$$\pi(x_i \mid y_0, \ldots, y_T) = \text{Normal}\left(x_i; \frac{a_i t_i + b_i s_i}{t_i + s_i}, \frac{1}{t_i + s_i}\right).$$

(e) From (d), we know the marginal distribution of X_T given the data, and we may start with simulating x_T. Then, using that

$$\pi(X_{T-1} \mid x_T, y_0, \ldots, y_T) \propto \pi(X_{T-1} \mid y_0, \ldots, y_{T-1})\pi(x_T \mid X_{T-1})$$

we may use the computed values of a_i and t_i, together with the likelihood of $\pi(x_T \mid X_{T-1})$, to compupte the normal posterior, and then simulate from it. Similar steps can now be repeated for $i = T-2, \ldots, 0$.

3. (a) We get

$$\pi(X_0, \ldots, X_T, y_0, \ldots, y_T \mid \lambda_1, \lambda_2, \lambda_3) \propto_{\lambda_1,\lambda_2,\lambda_3} \prod_{i=0}^{T} \pi(y_i \mid X_i)$$

$$\propto_{\lambda_1,\lambda_2,\lambda_3} \prod_{i=0}^{T}\prod_{j=1}^{3} \left[\exp(-\lambda_j)\lambda_j^{y_i}\right]^{I(X_i=j)}$$

(b) One may use the Forward-Backward algorithm. Define, for $i = 0, \ldots, T$ and $j = 1, 2, 3$,

$$a_{ij} = \Pr(X_i = j \mid y_0, \ldots, y_i, \theta')$$

and for $i = 0, \ldots, T-1$ and $j = 1, 2, 3$,

$$b_{ij} = \Pr(y_{i+1}, \ldots, y_T \mid X_i, \theta'),$$

writing also $b_{Tj} = 1$. We use the forward-backward algorithm to recursively compute values for a_{ij} and b_{ij}. With these computed, we can write

$$\pi(X_i \mid y_0, \ldots, y_T, \theta') \propto_{X_i} \pi(y_{i+1}, \ldots, y_T \mid X_i, \theta')\pi(X_i \mid y_0, \ldots, y_i, \theta')$$

and so we get, for $i = 0, \ldots, T$,

$$w_{ij} = \Pr(X_i = j \mid y_0, \ldots, y_T, \theta') = \frac{a_{ij}b_{ij}}{a_{i1}b_{i1} + a_{i2}b_{i2} + a_{i3}b_{i3}}$$

which are the numbers we needed to compute.
The forward-backward algorithm has the following steps:

- As
$$\pi(X_0 \mid y_0, \theta') \propto_{X_0} \pi(y_0 \mid X_0, \theta')\pi(X_0)$$
we get
$$a_{0j} = \frac{\exp(-\lambda_j)\lambda_j^{y_0}}{\exp(-\lambda_1)\lambda_1^{y_0} + \exp(-\lambda_2)\lambda_2^{y_0} + \exp(-\lambda_3)\lambda_3^{y_0}}.$$
- When $i > 0$ we get
$$\begin{aligned} & \pi(X_i = j \mid y_0, \ldots, y_{i-1}, \theta') \\ = & \sum_{s=1}^{3} \pi(X_i = j \mid X_{i-1} = s)\pi(X_{i-1} = s \mid y_0, \ldots, y_{i-1}, \theta') \\ = & 0.1 \cdot (a_{i-1,1} + a_{i-1,2} + a_{i-1,3}) + 0.7a_{i-1,j} \\ = & 0.1 + 0.7 \cdot a_{i-1,j}. \end{aligned}$$
Using
$$\pi(X_i \mid y_0, \ldots, y_i, \theta') \propto_{X_i} \pi(y_i \mid X_i, \theta')\pi(X_i \mid y_0, \ldots, y_{i-1}, \theta')$$
we get
$$a_{ij} = \frac{\exp(-\lambda_j)\lambda_j^{y_i}\left[0.1 + 0.7a_{i-1,j}\right]}{\sum_{s=1}^{3} \exp(-\lambda_s)\lambda_s^{y_i}\left[0.1 + 0.7a_{i-1,s}\right]}$$
- For $i < T$, we geet
$$\begin{aligned} b_{i-1,j} &= \pi(y_i, \ldots, y_T \mid X_{i-1} = j, \theta') \\ &= \sum_{s=1}^{3} \pi(X_i = s \mid X_{i-1} = j)\pi(y_i \mid X_i = s, \theta')\pi(y_{i+1}, \ldots, y_T \mid X_i = s, \theta') \\ &= \sum_{s=1}^{3} 0.1^{I(s \neq j)} 0.8^{I(s=j)} \frac{\exp(-\lambda_s)\lambda_s^{y_i}}{y_i!} b_{is} \end{aligned}$$

(c) We get
$$\begin{aligned} Q(\theta \mid \theta') &= E_{\theta'}\left[\log\left(\pi\left(X_0, \ldots, X_T, y_0, \ldots, y_T \mid \lambda_1, \lambda_2, \lambda_3\right)\right)\right] \\ &= E_{\theta'}\left[\sum_{i=0}^{T} \sum_{j=1}^{3} I(X_i = j)(-\lambda_j + y_i \log(\lambda_j))\right] \\ &= \sum_{i=0}^{T} \sum_{j=1}^{3} w_{ij}(-\lambda_j + y_i \log(\lambda_j)) \end{aligned}$$
where the w_{ij} were defined in the answer of (b).

(d) For $j = 1, 2, 3$ we may differentiate $Q(\theta \mid \theta')$ with respect to λ_j and set to zero. This results is
$$\lambda_j = \frac{\sum_{i=0}^{T} w_{ij} y_i}{\sum_{i=0}^{T} w_{ij}}.$$

(e) The full algorithm would go as follows: First, derive some starting values for $\lambda_1, \lambda_2, \lambda_3$. Note that the model is completely symmetric under permutation of the indices of the λ's, so for every local maximum we find, there will be others where the λ's are permuted. Which one we reach will be determined by the starting vector of the λ's; one may choose $\lambda_1 < \lambda_2 < \lambda_3$ with reasonable values compared to the mean values of the y_i. Then, the algorithm would iterate between running the forward-backward algorithm to compute the values of the w_{ij}, and updating the values of the λ_j according to (d), until connvergence is reached.

4. (a) we get

$$\begin{aligned}
& \pi(X_0, \dots, X_T, Y_0, \dots, Y_T \mid \tau_0, \tau_x, \tau_y) \\
= \quad & \left[\prod_{i=0}^{T} \pi(Y_i \mid X_i, tau_y)\right] \left[\prod_{i=1}^{T} \pi(X_i \mid X_{i-1}, \tau_x)\right] \pi(X_0) \\
\propto_\theta \quad & \left[\prod_{i=0}^{T} \tau_y^{1/2} \exp\left(-\frac{\tau_y}{2}(Y_i - X_i)^2\right)\right] \left[\prod_{i=1}^{T} \tau_x^{1/2} \exp\left(-\frac{\tau_x}{2}(X_i - X_{i-1})^2\right)\right] \tau_0^{1/2} \exp\left(-\frac{\tau_0}{2} X_0^2\right)
\end{aligned}$$

(b) This is exactly the answer to exercise 2(d) above. In the following, let us use the notation

$$X_i \mid y_0, \dots, y_T, \theta' \sim \text{Normal}\left(m_i, \frac{1}{u_i}\right)$$

where m_i and u_i are the values computed as in 2(d).

(c) Using (a) we get

$$\begin{aligned}
Q(\theta \mid \theta') &= E_{\theta'}\left[\log(\pi(X_0, \dots, X_T, Y_0, \dots, Y_T \mid \theta))\right] \\
&= E_{\theta'}\left[\sum_{i=0}^{T}\left(\frac{1}{2}\log(\tau_y) - \frac{\tau_y}{2}(Y_i - X_i)^2\right)\right. \\
&\quad \left. + \sum_{i=1}^{T}\left(\frac{1}{2}\log(\tau_x) - \frac{\tau_x}{2}(X_i - X_{i-1})^2\right) + \frac{1}{2}\log(\tau_0) - \frac{\tau_0}{2}X_0^2\right] \\
&= \sum_{i=0}^{T}\left(\frac{1}{2}\log(\tau_y) - \frac{\tau_y}{2}(Y_i^2 - 2Y_i E_{\theta'}[X_i] + E_{\theta'}[X_i^2])\right) \\
&\quad + \sum_{i=1}^{T}\left(\frac{1}{2}\log(\tau_x) - \frac{\tau_x}{2}(E_{\theta'}[X_i^2] - 2E_{\theta'}[X_i X_{i-1}] + E_{\theta'}[X_{i-1}])\right) \\
&\quad + \frac{1}{2}\log(\tau_0) - \frac{\tau_0}{2}E_{\theta'}[X_0^2]
\end{aligned}$$

We see from this that we actually need to go back to the Forward-Backward algorithm to be able to compute $E_{\theta'}[X_i X_{i-1}]$.

Note that

$$\begin{aligned}\pi(X_{i-1} \mid X_i, y_0, \ldots, y_T, \theta') &= \pi(X_{i-1} \mid X_i, y_0, \ldots, y_{i-1}, \theta') \\ &\propto_{X_{i-1}} \pi(X_i \mid X_{i-1}, \theta')\pi(X_{i-1} \mid y_0, \ldots, y_{i-1}, \theta')\end{aligned}$$

Using the notation from exercise 2, computing with parameters θ', we have

$$X_{i-1} \mid y_0, \ldots, y_{i-1}, \theta' \sim \text{Normal}(a_{i-1}, \frac{1}{t_{i-1}})$$

and

$$X_i \mid X_{i-1}, \theta' \sim \text{Normal}(X_{i-1}, \frac{1}{\tau'_x})$$

Using the conjugacy formula, we get

$$X_{i-1} \mid X_i, y_0, \ldots, y_T, \theta' \sim \text{Normal}\left(\frac{\tau'_x X_i + a_{i-1}t_{i-1}}{\tau'_x + t_{i-1}}, \frac{1}{\tau'_x + t_{i-1}}\right)$$

thus

$$E_{\theta'}\left[X_i X_{i-1}\right] = E_{\theta'}\left[X_i \frac{\tau'_x X_i + a_{i-1}t_{i-1}}{\tau'_x + t_{i-1}}\right] = \frac{\tau'_x}{\tau_x + t_{i-1}} E_{\theta'}[X_i^2] + \frac{a_{i-1}t_{i-1}}{\tau'_x + t_{i-1}} E_{\theta'}[X_i]$$

For the remaining expressions, we have, from (b),

$$E_{\theta'}[X_i] = m_i$$

and

$$E_{\theta'}[X_i^2] = Var_{\theta'}[X_i] + E_{\theta'}[X_i]^2 = \frac{1}{u_i} + m_i^2$$

(d) Differentiating $Q(\theta \mid \theta')$ for each parameter and setting the result to zero yields

$$\begin{aligned}\tau_0 &= \frac{1}{E_{\theta'}[X_0^2]} \\ \tau_x &= \frac{1}{\frac{1}{n}\sum_{i=1}^n E_{\theta'}[X_i^2] - 2E_{\theta'}[X_i X_{i-1}] + E_{\theta'}[X_{i-1}^2]} \\ \tau_y &= \frac{1}{\frac{1}{n+1}\sum_{i=0}^n Y_i^2 - 2Y_i E_{\theta'}[X_i] + E_{\theta'}[X_i^2]}\end{aligned}$$

(e) The full EM algorithm would consist of first estimating reasonable starting values for τ_0, τ_x, and τ_y. Then, one would iterate between computing the expectations as in (c) and the new variable values as in (d), until converrgence.