Midterm

Bohan Niu

November 10, 2018

Problem 1
a)

$$p(\mu, \pi|x, y) = p(\mu|\pi, x, y)p(\pi|x, y) = p(\mu|x)p(\pi|y) \tag{1}$$

We can do this because x only depends on μ, y only depends on π and there is no relationship between x and y. So we can solve them separately:

$$\begin{aligned}
p(\mu|x) &= p(x|\mu)p(\mu) \\
&\propto e^{-\frac{\lambda}{2}(x-\mu)^2} e^{-\frac{\gamma}{2}\mu^2} \\
&\propto \exp\{-\frac{1}{2}(\lambda(x-\mu)^2 + \gamma\mu^2)\} \\
&\propto \exp\{-\frac{1}{2}(-2\lambda x\mu + \lambda\mu^2 + \gamma\mu^2)\} \\
&\propto \exp\{-\frac{\lambda+\gamma}{2}(\mu^2 - 2\frac{\lambda}{\lambda+\gamma}x\mu)\} \\
&\propto \exp\{-\frac{\lambda+\gamma}{2}(\mu - \frac{\lambda}{\lambda+\gamma}x)^2\} \\
&= \text{Normal}(\theta, \beta) \\
&\text{where } \theta = \frac{\lambda}{\lambda+\gamma}x, \beta = (\lambda+\gamma)^{-1}
\end{aligned} \tag{2}$$

$$\begin{aligned}
p(\pi|y) &= p(y|\pi)p(\pi) \\
&\propto \prod_{i=1}^{3} \pi_i^{\mathbb{1}[y=i]} \prod_{i=1}^{3} \pi_i^{\alpha-1} \\
&\propto \prod_{i=1}^{3} \pi_i^{\mathbb{1}[y=i]+\alpha_i-1} \\
&= \text{Dirichlet}(\mathbb{1}[y=i] + \alpha_i)
\end{aligned} \tag{3}$$

So in conclusion, we can get that:

$$p(\mu, \pi|x, y) = \text{Normal}(\theta, \beta)\text{Dirichlet}(\mathbb{1}[y=i] + \alpha_i) \tag{4}$$

b)

$$
\begin{aligned}
p(x_2, y_2|x_1, y_1) &= \int_\mu \int_\pi p(x_2, y_2|\mu, \pi) p(\mu, \pi|x_1, y_1) d\pi d\mu \\
&= \int_\mu \int_\pi p(x_2|\mu) p(y_2|\pi) p(\mu|x_1) p(\pi|y_1) d\pi d\mu \\
&= \int_\mu p(x_2|\mu) p(\mu|x_1) d\mu \int_\pi p(y_2|\pi) p(\pi|y_1) d\pi
\end{aligned}
\tag{5}
$$

Again, we take them separately for the same reason.

$$
\begin{aligned}
\int_\pi p(y_2|\pi) p(\pi|y_1) d\pi &= \int_\pi \prod_{i=1}^{3} \pi_i^{\mathbb{1}[y_2=i]} \prod_{j=1}^{3} \pi_j^{\mathbb{1}[y_1=j]+\alpha_j-1} \\
&= \frac{\mathbb{1}[y_2 = i] + \mathbb{1}[y_1 = i] + \alpha_i}{3 + \sum_j \mathbb{1}[y_1 = j] + \alpha_j}
\end{aligned}
\tag{6}
$$

$$
\begin{aligned}
\int_\mu p(x_2|\mu) p(\mu|x_1) d\mu &= \int_\mu \frac{\lambda}{\sqrt{2\pi}} \exp\{-\frac{\lambda}{2}(x_2 - \mu)^2\} \frac{\lambda+\gamma}{\sqrt{2\pi}} \exp\{-\frac{\lambda+\gamma}{2}(\mu - \frac{\lambda}{\lambda+\gamma} x_1)^2\} d\mu \\
&= \frac{\lambda(\lambda+\gamma)}{2\pi} \exp\{-\frac{\lambda}{2}(x_2 - \mu)^2 - \frac{\lambda+\gamma}{2}(\mu - \frac{\lambda}{\lambda+\gamma} x_1)^2\} \\
&= \frac{\lambda(\lambda+\gamma)}{2\pi} \exp\{-\frac{1}{2}(x_2^2 + \frac{\lambda^2}{\lambda+\gamma} x_1^2)\} \exp\{-\frac{1}{2}[(\lambda+\gamma+1)\mu^2 - 2(\lambda x_1 + x_2)\mu]\} \\
&= \frac{\lambda(\lambda+\gamma)}{2\pi} \exp\{-\frac{1}{2}(x_2^2 + \frac{\lambda^2}{\lambda+\gamma} x_1^2 + \frac{(\lambda x_1 + x_2)^2}{\lambda+\gamma+1})\} \\
&\int_\mu \exp\{-\frac{\lambda+\gamma+1}{2}(\mu - \frac{\lambda x_1 + x_2}{\lambda+\gamma+1})^2\} d\mu \\
&= \frac{\lambda(\lambda+\gamma)}{(\lambda+\gamma+1)\sqrt{2\pi}} \exp\{-\frac{1}{2}(x_2^2 + \frac{\lambda^2}{\lambda+\gamma} x_1^2 + \frac{(\lambda x_1 + x_2)^2}{\lambda+\gamma+1})\} \\
&\int_\mu \frac{\lambda+\gamma+1}{\sqrt{2\pi}} \exp\{-\frac{\lambda+\gamma+1}{2}(\mu - \frac{\lambda x_1 + x_2}{\lambda+\gamma+1})^2\} d\mu \\
&= \frac{\lambda(\lambda+\gamma)}{(\lambda+\gamma+1)\sqrt{2\pi}} \exp\{-\frac{1}{2}(x_2^2 + \frac{\lambda^2}{\lambda+\gamma} x_1^2 + \frac{(\lambda x_1 + x_2)^2}{\lambda+\gamma+1})\}
\end{aligned}
\tag{7}
$$

So after plugging in all the equation, the predictive distribution is:

$$
p(x_2, y_2|x_1, y_1) = \frac{\mathbb{1}[y_2 = i] + \mathbb{1}[y_1 = i] + \alpha_i}{3 + \sum_j \mathbb{1}[y_1 = j] + \alpha_j} \frac{\lambda(\lambda+\gamma)}{(\lambda+\gamma+1)\sqrt{2\pi}} \exp\{-\frac{1}{2}(x_2^2 + \frac{\lambda^2}{\lambda+\gamma} x_1^2 + \frac{(\lambda x_1 + x_2)^2}{\lambda+\gamma+1})\}
\tag{8}
$$

Problem 2
The EM equation in this case is:

$$\ln p(x|\lambda) = \int q(c) \ln \frac{p(x, c|\lambda)}{q(c)} dc + \int q(c) \ln \frac{q(c)}{p(c|x, \lambda)} dc \tag{9}$$

E-step:
To make $\mathcal{L}(\lambda)$ the largest, we first set $q(c)$ equals to $p(c|x, \lambda)$:

$$p(c|x, \lambda) \propto p(x|c, \lambda)p(c) \quad = \frac{\lambda_c^x e^{-\lambda_c}}{x!}\theta_c \tag{10}$$

After that, we can set $q_t(c) = p(c|x, \lambda_{t-1})$ at iteration t. Then we can calculate the expectation:

$$\mathcal{L} = \mathbb{E}_q[\ln p(x, c|\lambda)] - \mathbb{E}_q[\ln q(c)] \tag{11}$$

M-step: We can get the maximization of $\mathcal{L}(\lambda)$ by setting $\nabla_\lambda \mathcal{L}_t(\lambda)$ equals to zero.

In summary, the steps of the EM algorithm are:

1. Initialize λ, θ and c in some way.
2. For iteration $t = 1, \ldots, T$

   (a) E-Step: Calculate the matrix $\mathbb{E}[\phi]$, where:

   $$\mathbb{E}_q[c] =$$

   (b) M-Step: Update the vector λ using the expectations above in the following equation

   $$\lambda = \arg\max \mathcal{L}_t(\lambda)$$

   (c)Calculate $\ln p(x|\lambda)$ using the equation

Problem 3

From the problem, we can factorize the joint likelihood:

$$p(\lambda_{1:n}, \theta|x) = \prod_{i=1}^{n} p(\lambda_i|\theta)p(\theta) \prod_{i=1}^{n} p(x|\lambda_{1:n}, \theta) \tag{12}$$

We also have:

$$q(\lambda_{1:n}, \theta) = q(\theta) \prod_{i=1}^{n} q(\lambda_i) \tag{13}$$

By optimal method, we can find the distribution of $q(\lambda_k)$ and $q(\theta)$ respectively:

$$\begin{aligned}
q(\lambda_k) &\propto \exp\{\mathbb{E}_{q(\theta)}[\ln p(\lambda_{1:n}, \theta|x)]\} \\
&\propto \exp\{\mathbb{E}_{q(\theta)}[\ln p(\lambda_k|\theta) + \ln p(\theta) + \ln p(x_k|\lambda_k)]\} \\
&\propto \exp\{\mathbb{E}_{q(\theta)}[\ln p(\lambda_k|\theta)]\} p(x_k|\lambda_k) \\
&\propto \mathbb{E}_{q(\theta)}[\lambda_k^{a-1} e^{-\theta\lambda_k}] e^{-\lambda_k} \lambda^{x_k} \\
&\propto \lambda_k^{a-1} e^{-\mathbb{E}_{q(\theta)}[\theta]\lambda_k} e^{-\lambda_k} \lambda^{x_k} \\
&\propto \lambda_k^{x_k+a-1} e^{-(\mathbb{E}_{q(\theta)}[\theta]+1)\lambda_k} \\
&= \text{Gamma}(\lambda_k|a'_k, d'_k) \\
&\text{where } a'_k = x_k + a, d'_k = \mathbb{E}_{q(\theta)}[\theta] + 1
\end{aligned} \tag{14}$$

$$\begin{aligned}
q(\theta) &\propto \exp\{\mathbb{E}_{q(\lambda_{1:n})}[\ln p(\lambda_{1:n}, \theta|x)]\} \\
&\propto \exp\{\mathbb{E}_{q(\lambda_{1:n})}[\ln p(\lambda_{1:n}|\theta) + \ln p(\theta) + \ln p(x|\lambda_{1:n})]\} \\
&\propto \exp\{\mathbb{E}_{q(\lambda_{1:n})}[\ln p(\lambda_{1:n}|\theta) + \ln p(\theta))]\} \\
&\propto \exp\{\mathbb{E}_{q(\lambda_{1:n})}[\ln p(\lambda_{1:n}|\theta)]\} p(\theta) \\
&\propto \mathbb{E}_{q(\lambda_{1:n})}[\prod_{i=1}^{n} p(\lambda_i|\theta)] p(\theta) \\
&\propto \mathbb{E}_{q(\lambda_{1:n})}[\prod_{i=1}^{n} \theta^a e^{-\theta\lambda_k}] \theta^{b-1} e^{-c\theta} \\
&\propto \theta^{na} e^{-\sum_{i=1}^{n} \mathbb{E}_{q(\lambda_i)}[\lambda_i]\theta} \theta^{b-1} e^{-c\theta} \\
&\propto \theta^{na+b-1} e^{-(c+\sum_{i=1}^{n} \mathbb{E}_{q(\lambda_i)}[\lambda_i])\theta} \\
&= \text{Gamma}(\theta|b', c') \\
&\text{where } b' = na + b, c' = c + \sum_{i=1}^{n} \mathbb{E}_{q(\lambda_i)}[\lambda_i]
\end{aligned} \tag{15}$$

After getting $q(\lambda_k)$ and $q(\theta)$, we can calculate the expectation of them:

$$\begin{aligned}
\mathbb{E}_{q(\lambda_k)}[\lambda_k] &= \frac{a'_i}{d'_i} \\
\mathbb{E}_{q(\theta)}[\theta] &= \frac{b'}{c'}
\end{aligned} \tag{16}$$

We can also calculate $\mathcal{L}$:

$$
\begin{aligned}
\mathcal{L} &= \int\int q(\lambda_{1:n},\theta)\ln\frac{p(\lambda_{1:n},\theta|x)}{q(\lambda_{1:n},\theta)}d\theta d\lambda_{1:n} \\
&= \int\int q(\theta)\prod_{i=1}^{n} q(\lambda_i)\ln\frac{p(\lambda_{1:n},\theta|x)}{q(\theta)\prod_{i=1}^{n} q(\lambda_i)}d\theta d\lambda_{1:n} \\
&= \int\int q(\theta)\prod_{i=1}^{n} q(\lambda_i)\ln\frac{\prod_{i=1}^{n} p(\lambda_i|\theta)p(\theta)\prod_{i=1}^{n} p(x|\lambda_{1:n},\theta)}{q(\theta)\prod_{i=1}^{n} q(\lambda_i)}d\theta d\lambda_{1:n} \\
&= \int q(\theta)\ln p(\theta)d\theta + \int\int q(\theta)\prod_{i=1}^{n} p(\lambda_i|\theta)\ln\sum_{i=1}^{n} p(\lambda_i|\theta)d\theta d\lambda \\
&+ \int\prod_{i=1}^{n} p(\lambda_i|\theta)\ln\sum_{i=1}^{n} p(x_i|\lambda_i)d\lambda_{1:n} - \int\prod_{i=1}^{n} q(\lambda_i)\ln\sum_{i=1}^{n} q(\lambda_i)d\lambda_{1:n} \\
&- \int q(\theta)\ln q(\theta)d\theta
\end{aligned}
\tag{17}
$$

Since we know all of them, we can solve them separately.
VI algorithm for Bayesian regression model:

1. Initialize a, b, c in some way
2. For iteration $t = 1, \ldots, T$
   - Update $q(\lambda_k)$ by setting

     $$
     \begin{aligned}
     a_t^{k'} &= x_k + a \\
     d_t^{k'} &= \frac{b'_t}{c'_t} + 1
     \end{aligned}
     $$

   - Update $q(\theta)$ by setting

     $$
     \begin{aligned}
     b'_t &= na + b \\
     c'_t &= c + \sum_{i=1}^{n}\frac{a'_i}{d'_i}
     \end{aligned}
     $$

   - Evaluate $\mathcal{L}$ to access convergence