Bayesian Modelling – Problem Sheet 7 (Solutions)

Problem 1

1. We have

$$\pi(\theta|x) \propto f(x|\theta)\pi(\theta) \propto \Big(\theta^{-n} e^{-\frac{\sum_{k=1}^n |x_k|}{\theta}}\Big)\theta^{-\alpha_0-1}e^{-\frac{\alpha_0}{\theta}}$$
$$= \theta^{-\alpha_0-n-1}\exp\Big(-\frac{\alpha_0+\sum_{k=1}^n |x_k|}{\theta}\Big)$$

where we recognize the unnormalized density of the Inv-Gamma(α_n, β_n) distribution where

$$\alpha_n = \alpha_0 + n, \quad \beta_n = \alpha_0 + \sum_{k=1}^n |x_k|.$$

The prior is conjugate because $\pi(\theta)$ and $\pi(\theta|x)$ belong to the same family of distributions (i.e. the family of inverse Gamma distributions).

2. Note that, when $n \geq 2$, we have $\alpha_n > 2$ so that all the integrals below are finite.

For the posterior mean:

$$\begin{aligned}
\mathbb{E}_\pi[\theta|x] = \int_\Theta \theta\pi(\theta|x)\mathrm{d}\theta &= \frac{\beta_n^{\alpha_n}}{\Gamma(\alpha_n)}\int_\Theta \theta\Big(\theta^{-\alpha_n-1}e^{-\frac{\beta_n}{\theta}}\Big)\mathrm{d}\theta \\
&= \frac{\beta_n\Gamma(\alpha_n-1)}{\Gamma(\alpha_n)}\frac{\beta_n^{\alpha_n-1}}{\Gamma(\alpha_n-1)}\int_\Theta \theta^{-(\alpha_n-1)-1}e^{-\frac{\beta_n}{\theta}}\mathrm{d}\theta \\
&= \frac{\beta_n\Gamma(\alpha_n-1)}{\Gamma(\alpha_n)} \\
&= \frac{\beta_n}{(\alpha_n-1)}
\end{aligned}$$

where the last equality uses the hint.

For the posterior variance we have, using similar computations,

$$\begin{aligned}
\mathbb{E}_\pi[\theta^2|x] = \int_\Theta \theta^2\pi(\theta|x)\mathrm{d}\theta &= \frac{\beta_n^{\alpha_n}}{\Gamma(\alpha_n)}\int_\Theta \theta^2\Big(\theta^{-\alpha_n-1}e^{-\frac{\beta_n}{\theta}}\Big)\mathrm{d}\theta \\
&= \frac{\beta_n^2\Gamma(\alpha_n-2)}{\Gamma(\alpha_n)}\frac{\beta_n^{\alpha_n-2}}{\Gamma(\alpha_n-2)}\int_\Theta \theta^{-(\alpha_n-2)-1}e^{-\frac{\beta_n}{\theta}} \\
&= \frac{\beta_n^2}{(\alpha_n-1)(\alpha_n-2)}
\end{aligned}$$

and thus

$$\mathrm{Var}_\pi(\theta|x) = \mathbb{E}_\pi[\theta^2|x] - \mathbb{E}_\pi[\theta|x]^2 = \frac{\beta_n^2}{(\alpha_n-1)^2(\alpha_n-2)}.$$

3. When $\alpha_0 < 1$ and $n = 1$ we have $\alpha_n < 2$ and thus

$$\mathbb{E}_\pi[\theta^2|x] = \frac{\beta_n^{\alpha_n}}{\Gamma(\alpha_n)} \int_\Theta \theta^{-(\alpha_n-2)-1} e^{-\frac{\beta_n}{\theta}} = +\infty$$

implying that $\mathrm{Var}_\pi(\theta|x) = +\infty$. In this case, and under a quadratic loss function, the posterior expected loss of $\mathbb{E}_\pi[\theta|x]$ is not finite because

$$\rho(\pi, \mathbb{E}_\pi[\theta|x]|x) = \int_\Theta (\theta - \mathbb{E}_\pi[\theta|x])^2 \pi(\theta|x) \mathrm{d}\theta = \mathrm{Var}_\pi(\theta|x) = +\infty.$$

Since this holds for any observation x, it follows that the Bayes risk of the estimator $\delta^\pi(\cdot) := \mathbb{E}_\pi[\theta|\cdot]$ is infinite since

$$r(\pi) = \int_{\mathbb{R}^n} \rho(\pi, \delta^\pi(x)|x) m(x) \mathrm{d}x = +\infty, \quad m(x) = \int_\Theta f(x|\theta)\pi(\theta)\mathrm{d}\theta$$

and thus $\mathbb{E}_\pi[\theta|\cdot]$ is not a Bayes estimator.

4. First, note that if $Y \sim \mathrm{Laplace}(\theta)$ for a $\theta > 0$, we have (using integration by parts)

$$\begin{aligned} \mathbb{E}\big[|Y|\big] = \frac{1}{2\theta} \int_{-\infty}^{+\infty} |y| e^{-\frac{y}{\theta}} \mathrm{d}y &= \frac{1}{\theta} \int_0^{+\infty} y e^{-\frac{y}{\theta}} \mathrm{d}y \\ &= \frac{1}{\theta}\Big(-\theta y e^{-\frac{y}{\theta}}\Big|_0^{+\infty} + \theta \int_0^{+\infty} e^{-\frac{y}{\theta}} \mathrm{d}y \Big) \\ &= \int_0^{+\infty} e^{-\frac{y}{\theta}} \mathrm{d}y \\ &= \theta \Big(\frac{1}{2\theta} \int_{-\infty}^{+\infty} e^{-\frac{|y|}{\theta}} \mathrm{d}y \Big) \\ &= \theta. \end{aligned} \tag{1}$$

Next, remark that the posterior expectation can be rewritten as

$$\mathbb{E}_\pi[\theta|X^{(n)}] = \frac{\beta_n}{\alpha_n - 1} = \frac{\frac{\alpha_0}{n} + \frac{1}{n}\sum_{k=1}^n |X_k|}{\frac{\alpha_0}{n} + 1 - \frac{1}{n}}.$$

Using (1), $\lim_{n\to+\infty} \frac{1}{n}\sum_{k=1}^n |X_k| = \theta_0$ (almost surely) by the law of large numbers and thus $\lim_{n\to+\infty} \mathbb{E}_\pi[\theta|X^{(n)}] = \theta_0$ (almost surely).

The posterior variance can be rewritten as

$$\mathrm{Var}_\pi(\theta|X^{(n)}) = \frac{\beta_n}{(\alpha_n - 1)^2(\alpha_n - 2)} = \frac{\mathbb{E}_\pi[\theta|X^{(n)}]}{(\alpha_n - 1)(\alpha_n - 2)}$$

and therefore, since $\lim_{n\to+\infty} \alpha_n = +\infty$ and $\lim_{n\to+\infty} \mathbb{E}_\pi[\theta|X^{(n)}] = \theta_0 < +\infty$ (almost surely), it follows that $\lim_{n\to+\infty} \mathrm{Var}_\pi(\theta|X^{(n)}) = 0$ (almost surely).

5. By definition, $\delta^\pi(x) \in \operatorname{argmin}_{d\in\Theta} \rho(\pi, d|x)$ where

$$\begin{aligned}
\rho(\pi, d|x) &= \int_\Theta L(\theta, d)\pi(\theta|x)\mathrm{d}\theta \\
&= \int_\Theta \left(\frac{d}{\theta} - 1\right)^2 \pi(\theta|x)\mathrm{d}\theta \\
&= d^2 \int_\Theta \theta^{-2}\pi(\theta|x)\mathrm{d}\theta + \int_\Theta \pi(\theta|x)\mathrm{d}\theta) - 2d \int_\Theta \theta^{-1}\pi(\theta|x)\mathrm{d}\theta \\
&= d^2\,\mathbb{E}_\pi[\theta^{-2}|x] + 1 - 2d\,\mathbb{E}_\pi[\theta^{-1}|x].
\end{aligned}$$

Since

$$\begin{aligned}
\frac{\partial\rho(\pi, d|x)}{\partial d} &= 2d\,\mathbb{E}_\pi[\theta^{-2}|x] - 2\mathbb{E}_\pi[\theta^{-1}|x] \\
\frac{\partial^2\rho(\pi, d|x)}{\partial d^2} &= 2\mathbb{E}_\pi[\theta^{-2}|x] > 0, \quad \forall d \in \Theta
\end{aligned}$$

it follows that $\delta^\pi(x)$ is the unique minimizer of the function $f(d) = \rho(\pi, d|x)$ and is such that

$$2\delta^\pi(x)\mathbb{E}_\pi[\theta^{-2}|x] - 2\mathbb{E}_\pi[\theta^{-1}|x] = 0$$

i.e.

$$\delta^\pi(x) = \frac{\mathbb{E}_\pi[\theta^{-1}|x]}{\mathbb{E}_\pi[\theta^{-2}|x]}.$$

Next, we have

$$\begin{aligned}
\mathbb{E}_\pi[\theta^{-1}|x] &= \frac{\beta_n^{\alpha_n}}{\Gamma(\alpha_n)} \int_\Theta \theta^{-1}\left(\theta^{-\alpha_n+1}e^{-\frac{\beta_n}{\theta}}\right)\mathrm{d}\theta \\
&= \frac{\Gamma(\alpha_n+1)}{\beta_n\Gamma(\alpha_n)} \frac{\beta_n^{\alpha_n+1}}{\Gamma(\alpha_n+1)} \int_\Theta \theta^{-(\alpha_n+1)-1}e^{-\frac{\beta_n}{\theta}}\,\mathrm{d}\theta \\
&= \frac{\Gamma(\alpha_n+1)}{\beta_n\Gamma(\alpha_n)} \\
&= \frac{\alpha_n}{\beta_n}
\end{aligned}$$

and, similarly,

$$\begin{aligned}
\mathbb{E}_\pi[\theta^{-2}|x] &= \frac{\beta_n^{\alpha_n}}{\Gamma(\alpha_n)} \int_\Theta \theta^{-2}\left(\theta^{-\alpha_n+1}e^{-\frac{\beta_n}{\theta}}\right)\mathrm{d}\theta \\
&= \frac{\Gamma(\alpha_n+2)}{\beta_n^2\Gamma(\alpha_n)} \frac{\beta_n^{\alpha_n+2}}{\Gamma(\alpha_n+2)} \int_\Theta \theta^{-(\alpha_n+2)-1}e^{-\frac{\beta_n}{\theta}}\,\mathrm{d}\theta \\
&= \frac{\Gamma(\alpha_n+2)}{\beta_n^2\Gamma(\alpha_n)} \\
&= \frac{\alpha_n(\alpha_n+1)}{\beta_n^2}.
\end{aligned}$$

Consequently,

$$\delta^\pi(x) = \frac{\mathbb{E}_\pi[\theta^{-1}|x]}{\mathbb{E}_\pi[\theta^{-2}|x]} = \frac{\beta_n}{\alpha_n + 1}.$$

The fact that $\lim_{n\to+\infty} \delta^\pi(X^{(n)}) = \theta_0$ (almost surely) is a direct consequence of the computations done in part 4.

6. a) By definition,

$$B_{01}^\pi(x) = \frac{f(x|1)}{m_1(x)} = \frac{2^{-n}e^{-\sum_{k=1}^n |x_k|}}{m_1(x)}$$

where

$$\begin{aligned} m_1(x) = \int_\Theta f(x|\theta)\pi(\theta)\mathrm{d}\theta &= \frac{\alpha_0^{\alpha_0}}{2^n\Gamma(\alpha_0)} \int_\Theta \theta^{-\alpha_0-n-1} \exp\Big(- \frac{\alpha_0 + \sum_{k=1}^n |x_k|}{\theta}\Big)\mathrm{d}\theta \\ &= \frac{\alpha_0^{\alpha_0}}{2^n\Gamma(\alpha_0)} \int_\Theta \theta^{-\alpha_n-1}e^{-\frac{\beta_n}{\theta}}\mathrm{d}\theta \\ &= \frac{\alpha_0^{\alpha_0}\Gamma(\alpha_n)}{2^n\beta_n^{\alpha_n}\Gamma(\alpha_0)}. \end{aligned}$$

Therefore,

$$B_{01}^\pi(x) = \frac{\beta_n^{\alpha_n}\Gamma(\alpha_0)}{\alpha_0^{\alpha_0}\Gamma(\alpha_n)e^{\sum_{k=1}^n |x_k|}} = \frac{\beta_n^{\alpha_n}\Gamma(\alpha_0)}{\alpha_0^{\alpha_0}\Gamma(\alpha_n)e^{\beta_n-\beta_0}}.$$

b) We show below that

$$\lim_{\alpha_0\to 0_+} \alpha_0^{\alpha_0} = 1, \quad \lim_{\alpha_0\to 0_+} \Gamma(\alpha_0) = +\infty.$$

Note first that

$$\alpha_0^{\alpha_0} = \exp\big(\alpha_0\log\alpha_0\big) = \exp\left(\frac{\log\alpha_0}{\frac{1}{\alpha_0}}\right) = \exp\left(- \frac{-\log\alpha_0}{\frac{1}{\alpha_0}}\right)$$

where, using Hospital's rule,

$$\lim_{\alpha_0\to 0_+} \frac{-\log\alpha_0}{\frac{1}{\alpha_0}} = \lim_{\alpha_0\to 0_+} \frac{-\alpha_0^{-1}}{-\alpha_0^{-2}} = \lim_{\alpha_0\to 0_+} \alpha_0 = 0.$$

Therefore, since the mapping $z \mapsto e^{-z}$ is continuous,

$$\lim_{\alpha_0\to 0_+} \alpha_0^{\alpha_0} = \exp\left(- \lim_{\alpha_0\to 0_+} \frac{-\log\alpha_0}{\frac{1}{\alpha_0}}\right) = e^0 = 1.$$

Next, because

$$\Gamma(t+1) = t\Gamma(t), \quad \forall t > 0$$

and because the Gamma function is continuous on $(0, +\infty)$, it follows that

$$\lim_{\alpha_0 \to 0_+} \Gamma(\alpha_0) = \lim_{\alpha_0 \to 0_+} \frac{\Gamma(\alpha_0 + 1)}{\alpha_0} = \frac{1}{0} = +\infty$$

and thus, as $\alpha_0 \to 0$, $B_{01}^{\pi}(x) \to +\infty$.

As $\alpha_0 \to 0$, it is easily checked that the prior variance goes to $+\infty$. Hence, using a vague prior in the context of hypothesis testing can lead to meaningless conclusions as illustrated in the above computations where the Bayes factor can be made arbitrary large by increasing the prior variance.

Problem 2

1. a) Let $\epsilon > 0$, $V_\epsilon = \{\theta : \|\theta - \theta_0\| \geq \epsilon\}$ and $C > 0$ be such that $\|\theta - \theta_0\| \leq C$ for all $\theta \in \Theta$. Notice that such a constant C exists since Θ is bounded. Then,

$$\begin{aligned}
\|\mathbb{E}_\pi[\theta|X^{(n)}] - \theta_0\| &= \left\| \int_\Theta (\theta - \theta_0)\pi(\theta|X^{(n)})\mathrm{d}\theta \right\| \\
&\leq \int_\Theta \|\theta - \theta_0\|\pi(\theta|X^{(n)})\mathrm{d}\theta \\
&= \int_{V_\epsilon} \|\theta - \theta_0\|\pi(\theta|X^{(n)})\mathrm{d}\theta + \int_{V_\epsilon^c} \|\theta - \theta_0\|\pi(\theta|X^{(n)})\mathrm{d}\theta \\
&\leq C\, \pi(V_\epsilon|X^{(n)}) + \epsilon
\end{aligned}$$

where the first inequality uses Jensen's inequality. By assumption, $\pi(\theta|X^{(n)})$ is consistent which, together with the above computations, implies that

$$0 \leq \limsup_{n \to +\infty} \|\mathbb{E}_\pi[\theta|X^{(n)}] - \theta_0\| \leq \epsilon, \quad \mathbb{P}_{\theta_0} - \text{almost surely.}$$

Since $\epsilon > 0$ is arbitrary the result follows. (Remark: If a deterministic sequence $(x_n)_{n\geq 1}$ is such that $\limsup_{n\to+\infty} |x_n| \leq \epsilon$ for all $\epsilon > 0$ then we trivially have $\lim_{n\to+\infty} x_n = 0$. The same result holds for the almost sure convergence of random variables but in this case some effort is needed to prove it.)

b) Recall that

$$\rho^\pi(X^{(n)}) = \int_\Theta \|\mathbb{E}_\pi[\theta|X^{(n)}] - \theta\|^2 \pi(\theta|X^{(n)})\mathrm{d}\theta$$

and let $\epsilon > 0$. Then, with V_ϵ as above and using the inequality $\|a + b\|^2 \leq 2(\|a\|^2 + \|b\|^2)$, $a, b \in \mathbb{R}^d$, we have

$$\begin{aligned}
\rho^\pi(X^{(n)}) &= \int_\Theta \|\mathbb{E}_\pi[\theta|X^{(n)}] - \theta_0 + \theta_0 - \theta\|^2 \pi(\theta|X^{(n)})\mathrm{d}\theta \\
&\leq 2\int_\Theta \|\mathbb{E}_\pi[\theta|X^{(n)}] - \theta_0\|^2 \pi(\theta|X^{(n)})\mathrm{d}\theta + 2\int_\Theta \|\theta - \theta_0\|^2 \pi(\theta|X^{(n)})\mathrm{d}\theta \\
&\leq 2\|\mathbb{E}_\pi[\theta|X^{(n)}] - \theta_0\|^2 + 2C^2\pi(V_\epsilon|X^{(n)}) + 2\epsilon^2
\end{aligned}$$

where the inequality uses similar computations as in part 1.a). Therefore, using the consistency of $\pi(\theta|X^{(n)})$ and the result of part 1.a), we have

$$0 \leq \limsup_{n\to+\infty} \rho^\pi(X^{(n)}) \leq \epsilon, \quad \mathbb{P} - \text{almost surely.}$$

Since $\epsilon > 0$ is arbitrary this implies that $\lim_{n\to+\infty} \rho^\pi(X^{(n)}) = 0$, $\mathbb{P}_{\theta_0}$-almost surely.

2. By definition

$$B^\pi_{01}(X^{(n)}) = c_0 \frac{\pi(\Theta_0|X^{(n)})}{1 - \pi(\Theta_0|X^{(n)})}, \quad c_0 = \frac{\pi(\Theta_1)}{\pi(\Theta_0)}$$

where $c_0 \in (0, +\infty)$ since Θ_0 is a non-empty open set and $\pi(\theta)$ is a strictly positive proper prior density on Θ (and thus $\pi(\Theta_0) > 0$ and $\pi(\Theta_1) < 1$).

Let $C > 0$ and remark that

$$B^\pi_{01}(X^{(n)}) \geq C \Leftrightarrow \pi(\Theta_0|X^{(n)}) - \frac{C}{C + c_0} \geq 0. \tag{2}$$

Assume first that $\theta_0 \in \Theta_0$ and let $\epsilon > 0$ be such that $V^c_\epsilon \subset \Theta_0$. Notice that such an $\epsilon > 0$ exists because Θ_0 is open and non-empty. Then,

$$\pi(\Theta_0|X^{(n)}) \geq \pi(V^c_\epsilon|X^{(n)}) = 1 - \pi(V_\epsilon|X^{(n)})$$

where, by the consistency of $\pi(\theta|X^{(n)})$, $\lim_{n\to+\infty} \pi(V_\epsilon|X^{(n)}) = 0$, $\mathbb{P}_{\theta_0}$-almost surely. This shows that $\lim_{n\to+\infty} \pi(\Theta_0|X^{(n)}) = 1$, $\mathbb{P}_{\theta_0}$-almost surely, implying that

$$\liminf_{n\to+\infty} \Big(\pi(\Theta_0|X^{(n)}) - \frac{C}{C + c_0}\Big) \geq \frac{c_0}{C + c_0} > 0, \quad \mathbb{P}_{\theta_0} - \text{almost surely.} \tag{3}$$

Let $Z_n = \mathbb{1}_{[C,+\infty]}\big(B^\pi_{01}(X^{(n)})\big)$ so that, by (2) and (3), $\liminf_{n\to+\infty} Z_n = 1$, $\mathbb{P}_{\theta_0}$-almost surely. Therefore,

$$\liminf_{n\to+\infty} \mathbb{P}\big(B^\pi_{01}(X^{(n)}) \geq C\big) = \liminf_{n\to+\infty} \mathbb{E}[Z_n] \geq \mathbb{E}\big[\liminf_{n\to+\infty} Z_n\big] = 1$$

where the second inequality uses Fatou's lemma.

If $\theta_0 \in \Theta_1$, since

$$B^\pi_{01}(X^{(n)}) \leq C \Leftrightarrow B^\pi_{10}(X^{(n)}) \geq \frac{1}{C}, \quad B^\pi_{10}(X^{(n)}) := \frac{1}{B^\pi_{01}(X^{(n)})}$$

the result follows from the above computations.

Problem 3

1. Let $\epsilon > 0$. Then, using Markov's inequality,

$$\mathbb{P}\big(|Y_t - Y| \geq \epsilon\big) \leq \frac{\mathbb{E}[|Y_t - Y|^2]}{\epsilon^2}, \quad \forall t \geq 0$$

so that,

$$0 \leq \limsup_{t\to+\infty} \mathbb{P}\big(|Y_t - Y| \geq \epsilon\big) \leq \limsup_{t\to+\infty} \frac{\mathbb{E}[|Y_t - Y|^2]}{\epsilon^2} = 0.$$

Hence, for all $\epsilon > 0$ we have $\lim_{t\to+\infty} \mathbb{P}\big(|Y_t - Y| \geq \epsilon\big) = 0$; that is, $Y_t \to Y$ in probability.

2. a) If the Y_t's are i.i.d. then for any $t \geq 0$ and $(i,j) \in \mathcal{Y}^2$ we have

$$\mathbb{P}(Y_{t+1} = j | Y_t = i) = \mathbb{P}(Y_{t+1} = j) = \mathbb{P}(Y_0 = j).$$

Hence,

$$P = \begin{pmatrix} \mathbb{P}(Y_0 = 1) & \mathbb{P}(Y_0 = 2) \\ \mathbb{P}(Y_0 = 1) & \mathbb{P}(Y_0 = 2) \end{pmatrix} = \begin{pmatrix} \lambda_1 & 1-\lambda_1 \\ \lambda_1 & 1-\lambda_1 \end{pmatrix}.$$

To find the value(s) of λ_1 such that P is aperiodic and irreducible we consider three different cases:

- If $\lambda_1 = 0$ then $P = \left(\begin{smallmatrix} 0 & 1 \\ 0 & 1 \end{smallmatrix}\right)$ so that $P^t = P$ for all $t \geq 1$. In this case, P is neither aperiodic nor irreducible.
- If $\lambda_1 = 1$ then $P = \left(\begin{smallmatrix} 1 & 0 \\ 1 & 0 \end{smallmatrix}\right)$ so that $P^t = P$ for all $t \geq 1$. In this case, P is neither aperiodic nor irreducible.
- if $\lambda_1 \in (0,1)$ then it easily checked that P is aperiodic and irreducible.

Consequently, P is aperiodic and irreducible if an only if $\lambda_1 \in (0,1)$.

b) It is easily checked that the system of equation $P^T\mu = \mu$ has a unique solution at $\mu = (\lambda_0, 1-\lambda_0)$.

3. a) Let $\mathbf{1} = (1,1)$. Then,

$$R\mathbf{1} = \alpha(Q_1\mathbf{1}) + (1-\alpha)(Q_2\mathbf{1}) = \alpha(\mathbf{1}) + (1-\alpha)(\mathbf{1}) = \mathbf{1}$$

so that R is a transition matrix (it is obvious that all its components are non-negative).

To show that μ is an invariant distribution of R it suffices to notice that

$$\mu^T R = \alpha(\mu^T Q_1) + (1-\alpha)(\mu^T Q_2) = \alpha(\mu^T) + (1-\alpha)(\mu^T) = \mu^T.$$

b) We have

$$R\mathbf{1} = Q_1(Q_2\mathbf{1}) = Q_1\mathbf{1} = \mathbf{1}$$

so that R is a transition matrix (it is obvious that all its components are non-negative).

To show that μ is an invariant distribution of R it suffices to notice that

$$\mu^T R = \mu^T(Q_1 Q_2) = (\mu^T Q_1)Q_2 = (\mu^T)Q_2 = \mu^T.$$

4. a) Write P as $P^{(\alpha,\beta)} = \left(\begin{smallmatrix} \alpha & 1-\alpha \\ 1-\beta & \beta \end{smallmatrix}\right)$ and λ as $\lambda = (\lambda_1, 1-\lambda_1)$. Then, the parameter of the model is $\theta = (\lambda_1, \alpha, \beta) \in [0,1]^3$. Hence, the parameter space is $\Theta = [0,1]^3$.

   b) Let $\lambda_2 = 1 - \lambda_1$ and $p_{ij}^{(\alpha,\beta)}$ denote component (i,j) of the matrix $P^{(\alpha,\beta)}$ defined in part 4.a). Then, the likelihood function of the model is defined by

$$f(y|\theta) = \lambda_{y_0} \prod_{s=1}^{t} p^{(\alpha,\beta)}_{y_{t-1} y_t}, \quad \theta \in \Theta.$$

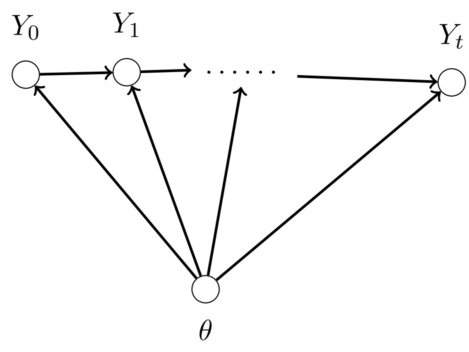
   c) Using the result in part 2.a), assuming that the Y_t's are i.i.d. with prior probability $1/2$ amounts to assuming that with prior probability $1/2$ have $(\alpha, \beta) = (\lambda_1, 1-\lambda_1)$.

   A possible prior distribution is then

$$\pi(\theta) = f(\lambda_1)\Big(\frac{1}{2}\mathbb{1}_{\{(\lambda_1, 1-\lambda_1)\}}((\alpha,\beta)) + \frac{1}{2} f(\alpha) f(\beta)\Big)$$

   where $f(\cdot)$ is e.g. the p.d.f. of the Beta(a_0, b_0) distribution for some hyper-parameters $a_0, b_0 > 0$.

   d) Here is the DAG for (Y, θ):



   e) We can for instance use the following algorithm:

   Input: $x_0 \in [0,1]^3$.
   Set $\theta_0 = x_0$
   for $k \geq 1$ **do**
   $\tilde{\theta}_k \sim \pi(\theta)$

Set $\theta_k = \tilde{\theta}_k$ with probability $\min\left\{1, \frac{f(y|\tilde{\theta}_k)}{f(y|\theta_{k-1})}\right\}$ and $\theta_k = \theta_{k-1}$ otherwise.

end for

Using the prior distribution as proposal distribution is an easy (but inefficient!) way to ensure that the proposal distribution has the same support as the posterior distribution (i.e. that the first sufficient condition given in Theorem 7.6 is verified). It is also obvious that for this proposal distribution $\mathbb{P}(\theta_k = \theta_{k-1}) > 0$ so that Theorem 7.6 ensures that the above Metropolis-Hastings algorithm defines a Markov chain $(\theta_k)_{k\geq 1}$ that converges (in the sense of Theorem 7.6) to $\pi(\theta|y) \propto f(y|\theta)\pi(\theta)$.