EXAMINATION SOLUTIONS

MATH 30015
BAYESIAN MODELLING
(Paper Code MATH-30015)

Summer 2018

2 hour and 30 minutes

---

1. (a) By definition, for $\theta \in \Theta$,

$$\pi(\theta|x^{(n)}) \propto f(x^{(n)}|\theta)\pi(\theta) = \theta^{\sum_{k=1}^n x_k}(1-\theta)^{n-\sum_{k=1}^n x_k}\theta^{\alpha_0-1}(1-\theta)^{\beta_0-1}$$
$$= \theta^{(\alpha_0+\sum_{k=1}^n x_k)-1}(1-\theta)^{(\beta_0+n-\sum_{k=1}^n x_k)-1}$$

where the right-hand side is proportional to the density of the Beta(α_n, β_n) distribution, with $\alpha_n = \alpha_0 + \sum_{k=1}^n x_k$ and $\beta_n = \beta_0 + n - \sum_{k=1}^n x_k$.

(b) Recall that an estimator δ is admissible if there exists no estimator δ' such that

$$R(\theta, \delta') \leq R(\theta, \delta), \quad \forall \theta \in \Theta$$

with the above inequality being strict for at least one $\theta \in \Theta$ and where the frequentist risk $R(\theta, \delta)$ is defined by

$$R(\theta, \delta) = \int_{\mathcal{X}} L(\theta, \delta(x)) f(x|\theta) \mathrm{d}x.$$

Using a quadratic loss function, $n = 1$ and the model we are considering in this question, we have

$$R(\theta, \delta) = \sum_{x=0}^{1} (\theta - \delta(x))^2 \theta^x (1-\theta)^{1-x}$$

and thus $R(\theta, \theta_n^*) = 0$ when $\theta = 0.2$. Because $(\theta - \delta(x))^2 = 0$ if and only if $\theta = \delta(x)$ while $0.2^x(1-0.2)^{1-x} > 0$ for all $x \in \{0, 1\}$, it follows that $R(0.2, \delta) = 0$ if and only if $\delta(x) = 0.2$ for all $x \in \{0, 1\}$. Hence, there exists no estimator $\delta \neq \theta_n^*$ such that $R(0.2, \delta) = 0$ and thus θ_n^* is admissible.

(c) By definition, $\delta^\pi(x^{(n)})$ minimizes the posterior expected loss; that is, $\delta^\pi(x^{(n)})$ minimizes the function $\rho : \Theta \to [0, +\infty)$ defined by

$$\rho(d) = \int_\Theta L(\theta, d)\pi(\theta|x^{(n)})\mathrm{d}\theta, \quad d \in \Theta.$$

Using the definition of L and the result in part (a), we have

$$\begin{aligned}
\rho(d) &= \frac{\Gamma(\alpha_n + \beta_n)}{\Gamma(\alpha_n)\Gamma(\beta_n)} \int_\Theta \frac{(\theta - d)^2}{\theta(1-\theta)}\theta^{\alpha_n - 1}(1-\theta)^{\beta_n - 1}\mathrm{d}\theta \\
&= \frac{\Gamma(\alpha_n + \beta_n)}{\Gamma(\alpha_n)\Gamma(\beta_n)} \int_\Theta (\theta - d)^2\theta^{\alpha_n - 2}(1-\theta)^{\beta_n - 2}\mathrm{d}\theta \\
&\propto \int_\Theta (\theta - d)^2\theta^{\alpha'_n - 1}(1-\theta)^{\beta'_n - 1}\mathrm{d}\theta
\end{aligned}$$

where $\alpha'_n = \alpha_n - 1$ and $\beta'_n = \beta_n - 1$. Note that because we have $\alpha_0 > 2$ and $\beta_0 > 2$, $\alpha'_n > 0$ and $\beta'_n > 0$ so that $\rho(d)$ is finite for any $d \in \Theta$.

Remark also that $\rho(d)$ is equal to the posterior expected loss associated to the quadric loss function $(\theta - d)^2$ when the prior distribution $\pi(\theta)$ is replaced by the $\mathrm{Beta}(\alpha_0 - 1, \beta_0 - 1)$ distribution. Hence, $\delta^\pi(x^{(n)})$ is the posterior mean under this new prior distribution; that is

$$\begin{aligned}
\delta^\pi(x^{(n)}) &= \frac{\Gamma(\alpha_n + \beta_n - 2)}{\Gamma(\alpha_n - 1)\Gamma(\beta_n - 1)} \int_\Theta \theta^{\alpha_n - 1}(1-\theta)^{\beta_n - 2}\mathrm{d}\theta \\
&= \Big(\frac{\Gamma(\alpha_n + \beta_n - 2)\Gamma(\alpha_n)}{\Gamma(\alpha_n + \beta_n - 1)\Gamma(\alpha_n - 1)}\Big) \int_\Theta \frac{\Gamma(\alpha_n + \beta_n - 1)}{\Gamma(\alpha_n)\Gamma(\beta_n - 1)}\theta^{\alpha_n - 1}(1-\theta)^{\beta_n - 2}\mathrm{d}\theta \\
&= \frac{\alpha_n - 1}{\alpha_n + \beta_n - 2}
\end{aligned}$$

where the last equality holds thanks to the *hint* given at the beginning of the question.

(d) To show that δ^π is admissible it is enough to show that it is the unique Bayes estimator associated to the loss L (Theorem 2.2). Unicity is obvious so that it remains to show that δ^π is a Bayes estimator; that is, that the Bayes risk $r(\pi)$ is finite.

By definition,

$$r(\pi)=\int_\Theta R(\theta,\delta^\pi)\pi(\theta)\mathrm{d}\theta=\int_\Theta\sum_{x=0}^{1}\frac{(\theta-\delta^\pi(x))^2}{\theta(1-\theta)}\theta^x(1-\theta)^{1-x}\pi(\theta)\mathrm{d}\theta$$
$$=\sum_{x=0}^{1}\frac{\Gamma(\alpha_0+\beta_0)}{\Gamma(\alpha_0)\Gamma(\beta_0)}\int_\Theta(\theta-\delta^\pi(x))^2\theta^{\alpha_0+x-2}(1-\theta)^{\beta_0-x-1}\mathrm{d}\theta$$
$$\le\sum_{x=0}^{1}\frac{\Gamma(\alpha_0+\beta_0)}{\Gamma(\alpha_0)\Gamma(\beta_0)}\int_\Theta\theta^{\alpha_0+x-2}(1-\theta)^{\beta_0-x-1}\mathrm{d}\theta.$$

Next, because $\alpha_0>2$, for $x\in\{0,1\}$ we have $\alpha_0+x-2>0$ and thus $\lim_{\theta\to 0}\theta^{\alpha_0+x-2}=0$. Similarly, as $\beta_0>2$, for $x\in\{0,1\}$ we have $\beta_0-x-1>0$ and thus $\lim_{\theta\to 1}\theta^{\beta_0-x-1}=0$. This shows that

$$\int_\Theta\theta^{\alpha_0+x-2}(1-\theta)^{\beta_0-x-1}\mathrm{d}\theta<+\infty,\quad\forall x\in\{0,1\}$$

and thus $r(\pi)<+\infty$ as required.

2. (a) We have $\mathbb{E}_{\theta_0}[X_1]=\theta_0$, $\mathrm{Var}_{\theta_0}(X_1)=\theta_0(1-\theta_0)$ and thus, since the observations are i.i.d., the central limit theorem yields

$$\sqrt{n}(\hat\theta_n-\theta_0)=\sqrt{n}\Big(\frac{1}{n}\sum_{k=1}^{n}X_k-\theta_0\Big)\overset{\text{dist.}}{\Longrightarrow}\mathcal{N}_1(0,\theta_0(1-\theta_0)).$$

For the posterior mean we have

$$\sqrt{n}(\mathbb{E}_\pi[\theta|X^{(n)}]-\theta_0)=\frac{\sqrt{n}\,\alpha_0}{\alpha_0+\beta_0+n}+\frac{n}{\alpha_0+\beta_0+n}\sqrt{n}\Big(\frac{1}{n}\sum_{k=1}^{n}X_k-\theta_0\Big)$$

where

$$\lim_{n\to+\infty}\frac{\sqrt{n}\,\alpha_0}{\alpha_0+\beta_0+n}=0,\quad\lim_{n\to+\infty}\frac{n}{\alpha_0+\beta_0+n}=1.$$

Hence, using the central limit theorem (and Slutsky's lemma),

$$\sqrt{n}(\mathbb{E}_\pi[\theta|X^{(n)}]-\theta_0)\overset{\text{dist.}}{\Longrightarrow}\mathcal{N}_1(0,\theta_0(1-\theta_0)).$$

(b) We have

$$
\begin{aligned}
\left|\mathbb{E}_\pi[\theta|X^{(n)}]-\hat\theta_n\right| &= \left|\frac{\frac{\alpha_0}{n}+\frac{1}{n}\sum_{k=1}^n X_k}{\frac{\alpha_0+\beta_0}{n}+1}-\frac{1}{n}\sum_{k=1}^n X_k\right|\\
&= \left|\frac{\frac{\alpha_0}{n}-\frac{\alpha_0+\beta_0}{n}\Big(\frac{1}{n}\sum_{k=1}^n X_k\Big)}{\frac{\alpha_0+\beta_0}{n}+1}\right|\\
&= \frac{1}{n}\left|\frac{\alpha_0}{\frac{\alpha_0+\beta_0}{n}+1}-\frac{\alpha_0+\beta_0}{\frac{\alpha_0+\beta_0}{n}+1}\Big(\frac{1}{n}\sum_{k=1}^n X_k\Big)\right|\\
&\leq \frac{1}{n}\frac{\alpha_0}{\frac{\alpha_0+\beta_0}{n}+1}+\frac{1}{n}\frac{\alpha_0+\beta_0}{\frac{\alpha_0+\beta_0}{n}+1}\Big(\frac{1}{n}\sum_{k=1}^n X_k\Big)\\
&\leq \frac{\alpha_0}{n}+\frac{\alpha_0+\beta_0}{n}\Big(\frac{1}{n}\sum_{k=1}^n X_k\Big).
\end{aligned}
$$

The law of large numbers ensures that $\lim_{n\rightarrow+\infty}\frac{1}{n}\sum_{k=1}^n X_k=\theta_0\leq 1$ ($\mathbb{P}_{\theta_0}$-almost surely) so that

$$
\limsup_{n\rightarrow+\infty} n\left|\mathbb{E}_\pi[\theta|X^{(n)}]-\hat\theta_n\right|\leq \alpha_0+(\alpha_0+\beta_0)\theta_0\quad (\mathbb{P}_{\theta_0}\text{-almost surely}).
$$

(c) A frequentist statistician may be willing to use the estimator $\mathbb{E}_\pi[\theta|X^{(n)}]$ because (i) this estimator has the same asymptotic distribution as the MLE and (ii) the difference $|\mathbb{E}_\pi[\theta|X^{(n)}]-\hat\theta_n|$ converges to zero much faster than $|\hat\theta_n-\theta_0|$. This last points means that very quickly has n increases the difference between the MLE and the posterior mean becomes negligible compared to the distance of these estimators to θ_0. (Another point that could be mentioned is that $\mathbb{E}_\pi[\theta|X^{(n)}]$ is admissible under the quadratic loss function, see part (d) of Question 1.)

(d) Let $\epsilon>0$, $\delta>0$, $V_\epsilon=\{\theta\in\Theta:\ |\theta-\theta_0|\geq\epsilon\}$ and

$$
\Sigma_{n,\delta}=\left\{x^{(n)}\in\mathcal{X}_1^n:\ \int_\Theta\prod_{k=1}^n\frac{\tilde f(x_k|\theta)}{\tilde f(x_k|\theta_0)}\pi(\theta)\mathrm{d}\theta\leq c\,\delta e^{-2n\delta^2}\right\}.
$$

Then,

$$\begin{aligned}
\mathbb{E}_{\theta_0}[\pi(V_\epsilon|X^{(n)})] &= \mathbb{E}_{\theta_0}[\phi_n(X^{(n)})\,\pi(V_\epsilon|X^{(n)})] + \mathbb{E}_{\theta_0}[\mathbb{1}_{\Sigma_{n,\delta}}(X^{(n)})(1-\phi_n(X^{(n)}))\pi(V_\epsilon|X^{(n)})] \\
&\quad + \mathbb{E}_{\theta_0}[\mathbb{1}_{\Sigma^c_{n,\delta}}(X^{(n)})(1-\phi_n(X^{(n)}))\pi(V_\epsilon|X^{(n)})] \\
&\leq \mathbb{E}_{\theta_0}[\phi_n(X^{(n)})] + \mathbb{P}_{\theta_0}(\Sigma_{n,\delta}) + \mathbb{E}_{\theta_0}[\mathbb{1}_{\Sigma^c_{n,\delta}}(X^{(n)})(1-\phi_n(X^{(n)}))\pi(V_\epsilon|X^{(n)})] \\
&\leq 2e^{-n\epsilon^2/2} + \frac{1}{n\delta^2} + \mathbb{E}_{\theta_0}[\mathbb{1}_{\Sigma^c_{n,\delta}}(X^{(n)})(1-\phi_n(X^{(n)}))\pi(V_\epsilon|X^{(n)})].
\end{aligned}$$

In addition (using Tonelli's theorem for the third inequality)

$$\begin{aligned}
&\mathbb{E}_{\theta_0}[\mathbb{1}_{\Sigma^c_{n,\epsilon}}(X^{(n)})(1-\phi_n(X^{(n)}))\pi(V_\epsilon|X^{(n)})] \\
&\qquad = \mathbb{E}_{\theta_0}\left[\mathbb{1}_{\Sigma^c_{n,\delta}}(X^{(n)})(1-\phi_n(X^{(n)}))\frac{\int_{V_\epsilon}\prod_{i=1}^n \frac{\tilde f(X_i|\theta)}{\tilde f(X_i|\theta_0)}\pi(\theta)\mathrm{d}\theta}{\int_{\Theta}\prod_{i=1}^n \frac{\tilde f(X_i|\theta)}{\tilde f(X_i|\theta_0)}\pi(\theta)\mathrm{d}\theta}\right] \\
&\qquad \leq \frac{e^{2n\delta^2}}{c\,\delta}\mathbb{E}_{\theta_0}\Big[\int_{V_\epsilon}(1-\phi_n(X^{(n)}))\prod_{i=1}^n\frac{\tilde f(X_i|\theta)}{\tilde f(X_i|\theta_0)}\pi(\theta)\mathrm{d}\theta\Big] \\
&\qquad = \frac{e^{2n\delta^2}}{c\,\delta}\int_{\{0,1\}^n}\left(\int_{V_\epsilon}(1-\phi_n(x^{(n)}))\prod_{i=1}^n\tilde f(x_i|\theta)\pi(\theta)\mathrm{d}\theta\right)\prod_{i=1}^n\mathrm{d}x_i \\
&\qquad = \frac{e^{2n\delta^2}}{c\,\delta}\int_{V_\epsilon}\left(\int_{\{0,1\}^n}\left((1-\phi_n(x^{(n)}))\prod_{i=1}^n\tilde f(x_i|\theta)\mathrm{d}x_i\right)\pi(\theta)\mathrm{d}\theta\right. \\
&\qquad \leq \frac{e^{2n\delta^2}}{c\,\delta}\sup_{\theta\in V_\epsilon}\mathbb{E}_\theta[1-\phi_n(X^{(n)})]\int_{V_\epsilon}\pi(\theta)\mathrm{d}\theta \\
&\qquad \leq \frac{2\,e^{-\frac{n}{2}(\epsilon^2-4\delta^2)}}{c\,\delta}.
\end{aligned}$$

Then, taking e.g. $\delta=\epsilon/\sqrt{8}$ yields

$$\mathbb{E}_{\theta_0}[\pi(V_\epsilon|X^{(n)})] \leq 2e^{-n\epsilon^2/2} + \frac{8}{n\epsilon^2} + \frac{2e^{-n\frac{\epsilon^2}{4}}}{c\,\epsilon}\sqrt{8}$$

and the result follows.

3. (a) Let $q_{1-\alpha/2}$ be the $(1-\alpha/2)$-quantile of the $\mathcal{N}_1(0,1)$ distribution. Then, as the posterior distribution is Gaussian, the HPD region at level $1-\alpha$ is the interval

$$\big[\mu_n - q_{1-\alpha/2}\sqrt{\sigma_n^2}\,,\,\mu_n + q_{1-\alpha/2}\sqrt{\sigma_n^2}\big].$$

(b) i. Let $\lambda_n = (1+n\sigma_0^2)^{-1}$, $\bar{X}_n = n^{-1}\sum_{k=1}^n X_k$ and $\pi_n^*(s|X^{(n)})$ be the p.d.f. of $\sqrt{n}(\theta - \hat{\theta}_n)|X^{(n)}$. Then, using the change of variable formula (see HW2, Problem 1)

$$\begin{aligned}
\pi_n^*(s|X^{(n)}) &= \frac{1}{\sqrt{n}}\pi\Big(\hat{\theta}_n + \frac{s}{\sqrt{n}}\,|X^{(n)}\Big)\\
&= \sqrt{\frac{1+n\sigma_0^2}{2\pi n\sigma_0^2}}\exp\Bigg(-(1+n\sigma_0^2)\frac{\Big(\lambda_n(\mu_0-\bar{X}_n)+\frac{s}{\sqrt{n}}\Big)^2}{2\sigma_0^2}\Bigg)\\
&= \sqrt{\frac{1+n\sigma_0^2}{2\pi n\sigma_0^2}}\exp\Bigg(-\frac{1+n\sigma_0^2}{2\,\sigma_0^2}\Big(\lambda_n^2(\mu_0-\bar{X}_n)^2+\frac{s^2}{n}+\frac{2s\lambda_n(\mu_0-\bar{X}_n)}{\sqrt{n}}\Big)^2\Bigg)\\
&= \sqrt{\frac{1+n\sigma_0^2}{2\pi n\sigma_0^2}}\exp\Bigg(-\frac{s^2}{2}-\frac{s^2}{2n\sigma_0^2}-\frac{1+n\sigma_0^2}{2\,\sigma_0^2}\Big(\lambda_n^2(\mu_0-\bar{X}_n)^2+\frac{2s\lambda_n(\mu_0-\bar{X}_n)}{\sqrt{n}}\Big)^2\Bigg)
\end{aligned}$$

where (almost surely)

$$\sqrt{\frac{1+n\sigma_0^2}{2\pi n\sigma_0^2}} \to \frac{1}{\sqrt{2\pi}},\qquad \lambda_n^2\to 0,\qquad \bar{X}_n \to \mathbb{E}[X_1] = \theta_0 \in (-\infty,+\infty).$$

Hence (almost surely),

$$\pi^*(s|X^{(n)}) \to \frac{1}{\sqrt{2\pi}}e^{-\frac{s^2}{2}}\quad \forall s\in\mathbb{R}$$

as required.

ii. Since for all $n\geq 1$ we have $\hat{\theta}_n \sim \mathcal{N}_1(\theta_0, n^{-1})$ it follows that

$$\big[\hat{\theta}_n - q_{1-\alpha/2}\,n^{-1/2}, \hat{\theta}_n + q_{1-\alpha/2}\,n^{-1/2}\big] = \big[\bar{X}_n - q_{1-\alpha/2}\,n^{-1/2}, \bar{X}_n + q_{1-\alpha/2}\,n^{-1/2}\big]$$

is a confidence interval at level $1-\alpha$.

iii. As n increases the posterior distribution $\pi^*(s|X^{(n)})$ becomes more and more similar to the distribution of $\sqrt{n}(\hat{\theta}_n - \theta_0)$ which is used to compute the confidence interval. Hence, as $n\to+\infty$, the difference between the credible interval and the confidence vanishes.

(c) i. Given the expression of the posterior distribution given in the statement of the question and the result in part (b)i. it should be clear that the new assumption on the data generating process does not modify the limiting distribution of $\sqrt{n}(\theta - \hat{\theta}_n)|X^{(n)}$. Then the result follows from part (c)i.

ii. Using the central limit theorem,

$$\sqrt{n}\big(\hat{\theta}_n - \mathbb{E}[X_1]\big) \Rightarrow \mathcal{N}_1(0, \mathrm{Var}(X_1)) \tag{1}$$

so that

$$\big[\bar{X}_n - q_{1-\alpha/2}\sqrt{1.5}\,n^{-1/2}, \bar{X}_n + q_{1-\alpha/2}\sqrt{1.5}\,n^{-1/2}\big]$$

is a confidence interval at level $1-\alpha$.

iii. As n increases the posterior distribution $\pi^*(s|X^{(n)})$ does not converge to the asymptotic distribution of $\sqrt{n}(\hat{\theta}_n - \theta_0)$ which is used to compute the confidence interval. Hence, as $n \to +\infty$, the difference between the credible interval and the confidence do not vanish.

4. (a) Let $\lambda_2 = 1 - \lambda_1$ and $p_{ij}^{\alpha,\beta}$ be the entry (i,j) of $P_{\alpha,\beta}$. Then, the likelihood $f(x^{(t+1)}|\theta)$ of the model is given by

$$f(x^{(t+1)}|\theta) = \lambda_{x_0} \prod_{s=1}^{t} p^{\alpha,\beta}_{x_{s-1}x_s}.$$

(b) i. If $\alpha = \beta = 1$ then $p_{11}^{(t)} = 1$ when t is even and 0 when t is odd. Hence, $1 = \limsup_{t\to+\infty} p_{11}^{(t)} > \liminf_{t\to+\infty} p_{11}^{(t)} = 0$. It is easily checked that this is the only set of parameters for which $\lim_{t\to+\infty} P^t_{\alpha,\beta}$ does not exist.

ii. If $\alpha = 0$ (resp. $\beta = 0$) then $p_{12}^{(t)} = 0$ (resp. $p_{21}^{(t)} = 0$) for all $t \geq 1$ and thus $P_{\alpha,\beta}$ is not irreducible. If $\alpha > 0$ and $\beta > 0$ then we have $p_{12}^{(1)} > 0$ and $p_{21}^{(1)} > 0$ so that $P_{\alpha,\beta}$ is irreducible. Hence, $P_{\alpha,\beta}$ is irreducible if and only if $\alpha, \beta > 0$.

iii. If $\alpha = \beta = 1$ then $p_{11}^{(t)} = 0$ when t is odd and thus $P_{\alpha,\beta}$ is periodic.
Let $\alpha = 0$ and $\beta = 1$. Then $p_{22}^{(t)} = 0$ for all $t \geq 1$ and thus $P_{\alpha,\beta}$ is periodic.
Similarly, when $\alpha = 1$ and $\beta = 0$, we have $p_{11}^{(t)} = 0$ for all $t \geq 1$ and thus $P_{\alpha,\beta}$ is periodic.
If $\alpha = \beta = 0$ then $p_{11}^{(t)} = p_{22}^{(t)} = 1$ for all $t \geq 1$ so that $P_{\alpha,\beta}$ is aperiodic.
Lastly, for any $(\alpha,\beta) \in (0,1)^2$, $p_{11}^{(t)} \to \alpha/(\alpha+\beta) > 0$ while $p_{22}^{(t)} \to \beta/(\alpha+\beta) > 0$ so that, for all t large enough $p_{11}^{(t)} > 0$ and $p_{22}^{(t)} > 0$.
Conclusion, $P_{\alpha,\beta}$ is aperiodic if and only if $(\alpha,\beta) \in (0,1)^2 \cup \{(0,0)\}$.

(c) From part (b) $P_{\alpha,\beta}$ is irreducible and aperiodic if and only if $(\alpha,\beta) \in (0,1)^2$. Then, we can take for instance the prior distribution defined for $\theta \in \Theta$ by

$$\begin{aligned}\pi(\theta) &= p(\lambda_1) \\ &\quad \times \Big(0.7p(\alpha)p(\beta) + 0.05\mathbb{1}_{\{(1,1)\}}(\alpha,\beta) + 0.125\mathbb{1}_{\{0\}\times[0,1]}(\alpha,\beta) + 0.125\mathbb{1}_{(0,1)\times\{0\}}(\alpha,\beta)\Big)\end{aligned}$$

where where e.g. p is the density of the Beta(a_0, b_0) distribution for some $a_0, b_0 > 0$.

(d) i. 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 $\min\left\{1, \frac{f(x^{(t+1)}|\tilde{\theta}_k)}{f(x^{(t+1)}|\theta_{k-1})}\right\}$ and $\theta_k = \theta_{k-1}$ otherwise.

end for

This algorithm is implementable because it is easy to sample from the distribution $\pi(\theta)$ specified in part (c).

ii. The proposal distribution $Q(\theta, \mathrm{d}\tilde{\theta})$ used in the above Metropolis-Hastings algorithm is such that $Q(\theta, A) > 0$ for all $\theta \in \Theta$ and (measurable) $A \subset \Theta$ such that $\pi(A|x^{(t+1)}) > 0$. Moreover, we clearly have

$$\mathbb{P}(\theta_k = \theta_{k-1}) > 0, \quad \forall k \geq 1$$

so that, by Theorem 7.6, the resulting Markov chain $(\theta_k)_{k\geq 0}$ is such that

$$\lim_{k\to+\infty} \mathbb{P}(\theta_k \in B) \to \pi(B|x^{(t+1)}), \qquad \text{for any measurable } B \subset \Theta.$$

Let $K > 1$ be the number of iterations we run the above Metropolis-Hastings algorithm. Then, we can approximate $\mathbb{E}_\pi[\theta|x^{(t+1)}]$ by $K^{-1}\sum_{k=1}^K \theta_k$.

End of solutions.