M2DS

Problem: the auxiliary Metropolis Hasting algorithm

Nota Bene: *Recall that if φ is a concave function on $\mathbb{R}$, and if X is a random variable such that $\mathbb{E}[|X|] < \infty$, then, Jensen's inequality writes $\mathbb{E}[\varphi(X)] \leqslant \varphi(\mathbb{E}[X])$.*

In all the problem, π is a probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ (where $\mathcal{B}(\mathbb{R})$ is the Borel sigma-field on $\mathbb{R}$) and we assume that π has a **strictly positive density** with respect to the Lebesgue measure λ. For simplicity, we also denote by π its density with respect to λ, that is, $\pi(\mathrm{d}x) = \pi(x)\lambda(\mathrm{d}x) = \pi(x)\mathrm{d}x$ where for any $x \in \mathbb{R}$,

$$\boxed{\pi(x) > 0}$$

and where we recall the abuse of notation $\lambda(\mathrm{d}x) = \mathrm{d}x$.

We consider $\{q_v \,:\, v \in \mathbb{R}\}$ a family of **strictly positive** densities on $\mathbb{R}$ (with respect to the Lebesgue measure) $y \mapsto q_v(y)$ indexed by some $v \in \mathbb{R}$.

We also consider a **strictly positive** density $f : \mathbb{R} \to \mathbb{R}_*^+ = \mathbb{R}^+ \setminus \{0\}$ with respect to the Lebesgue measure on $\mathbb{R}$.

The Auxiliary Metropolis-Hasting algorithm (also called AuxMH) consists in constructing a Markov chain $\{X_n \,:\, n \in \mathbb{N}\}$ starting with initial distribution μ in the following way:

- Draw $X_0 \sim \mu$ where μ is arbitrary

for $t \leftarrow 1$ **to** n **do**

- Draw $V_t \sim f$ and then $Y_t|_{V_t} \sim q_{V_t}$.
- Draw independently $V_t^* \sim f$ and $U_t \sim \mathrm{Unif}([0,1])$
- Set $X_t = \begin{cases} Y_t & \text{if } U_t \leqslant \alpha_{V_t,V_t^*}(X_{t-1},Y_t) \quad \text{where } \alpha_{v,v^*}(x,y) = \min\left(\frac{\pi(y)q_{v^*}(x)}{\pi(x)q_v(y)}, 1\right) \\ X_{t-1} & \text{otherwise} \end{cases}$

end

Algorithm 1: the auxiliary MH algorithm

This algorithm can be seen as a variant of the classical Metropolis–Hastings scheme. Although not a standard MH algorithm *per se*, the introduction of an auxiliary variable V_t provides additional flexibility in designing the proposal mechanism through the family $\{q_v \,:\, v \in \mathbb{R}\}$.

The second auxiliary variable V_t^*, which only appears in the acceptance probability, plays a crucial role in ensuring that the resulting Markov chain is π-reversible (the last property will be established later in this problem).

1. For any bounded measurable function h on $\mathbb{R}$, show that

$$\mathbb{E}_\mu[h(X_t)\mathbf{1}_{\{U_t \leqslant \alpha_{V_t,V_t^*}(X_{t-1},Y_t)\}}|X_{t-1}] = \int \beta(X_{t-1},y)h(y)\mathrm{d}y$$

where $\beta(x,y)$ can be written as an integral involving the functions f, q_v and α_{v,v^*} . Check that for all $x, y \in \mathbb{R}$, we have $\beta(x,y) > 0$ and $\pi(x)\beta(x,y) = \pi(y)\beta(y,x)$.

Solution.

We have

$$\begin{aligned}\mathbb{E}_\mu[h(X_t)\mathbf{1}_{\{U_t \leqslant \alpha_{V_t,V_t^*}(X_{t-1},Y_t)\}}|X_{t-1}] &= \int\cdots\int h(y)f(v)f(v^*)q_v(y)\mathbf{1}_{\{u \leqslant \alpha_{v,v^*}(X_{t-1},y)\}}\mathbf{1}_{\{[0,1]\}}(u)\mathrm{d}v\mathrm{d}v^*\mathrm{d}y\mathrm{d}u \\ &= \int h(y)\underbrace{\iint f(v)f(v^*)q_v(y)\alpha_{v,v^*}(X_{t-1},y)\mathrm{d}v\mathrm{d}v^*}_{\beta(X_{t-1},y)}\mathrm{d}y\end{aligned}$$

where we have set

$$\beta(x,y) = \iint f(v)f(v^*)q_v(y)\alpha_{v,v^*}(x,y)\mathrm{d}v\mathrm{d}v^*.$$

Since for any $(x,y,v) \in \mathbb{R}^3$, $f(v) > 0$, $q_v(x,y) > 0$, $\pi(x) > 0$, we obtain that $\beta(x,y) > 0$. Moreover,

$$\begin{aligned}\pi(x)\beta(x,y) &= \iint \pi(x) f(v) f(v^*) q_v(y) \min\left(\frac{\pi(y) q_{v^*}(x)}{\pi(x) q_v(y)}, 1\right) \mathrm{d}v\mathrm{d}v^* \\ &= \iint f(v) f(v^*) \min\left(\pi(y) q_{v^*}(x), \pi(x) q_v(y)\right) \mathrm{d}v\mathrm{d}v^*\end{aligned}$$

which is symmetric wrt (x,y) and hence $\pi(x)\beta(x,y) = \pi(y)\beta(y,x)$. □

2. Show that

$$\mathbb{E}_\mu[h(X_t)\mathbf{1}_{\{U_t > \alpha_{V_t, V_t^*}(X_{t-1}, Y_t)\}} | X_{t-1}] = h(X_{t-1})\bar{\beta}(X_{t-1})$$

where $\bar{\beta}(x)$ should be expressed in terms of the function β only.
Solution.

We have

$$\begin{aligned}\mathbb{E}_\mu[h(X_t)\mathbf{1}_{\{U_t > \alpha_{V_t, V_t^*}(X_{t-1}, Y_t)\}} | X_{t-1}] &= \int \cdots \int h(X_{t-1}) \mathbf{1}_{\{u > \alpha_{v,v^*}(X_{t-1}, y)\}} f(v) f(v^*) q_v(y) \mathbf{1}_{\{[0,1]\}} \mathrm{d}v\mathrm{d}v^*\mathrm{d}y\mathrm{d}u \\ &= h(X_{t-1}) \int \cdots \int \left(1 - \alpha_{v,v^*}(X_{t-1}, y)\right) f(v) f(v^*) q_v(y) \mathrm{d}v\mathrm{d}v^*\mathrm{d}y \\ &= h(X_{t-1}) \Big(1 - \int \underbrace{\iint \alpha_{v,v^*}(X_{t-1}, y) f(v) f(v^*) q_v(y) \mathrm{d}v\mathrm{d}v^*}_{\beta(X_{t-1}, y)} \mathrm{d}y\Big)\end{aligned}$$

Hence $\bar{\beta}(x) = 1 - \int \beta(x,y)\mathrm{d}y$. □

3. Deduce the expression of the Markov kernel $P(x, \mathrm{d}y)$ of the Markov chain $\{X_k : k \in \mathbb{N}\}$ in terms of β and $\bar{\beta}$.
Solution.

From the two previous questions, $P(x, \mathrm{d}y) = \beta(x,y)\mathrm{d}y + \bar{\beta}(x)\delta_x(\mathrm{d}y)$ □

4. Show that P is π-invariant.
Solution.

We have

$$\pi(\mathrm{d}x)P(x, \mathrm{d}y) = \pi(\mathrm{d}x)[\beta(x,y)\mathrm{d}y + \bar{\beta}(x)\delta_x(\mathrm{d}y)] = \pi(\mathrm{d}x)\beta(x,y)\mathrm{d}y + \pi(\mathrm{d}x)\bar{\beta}(x)\delta_x(\mathrm{d}y)$$

Note that for any bounded measurable function h, $\iint h(x,y)\pi(\mathrm{d}x)\bar{\beta}(x)\delta_x(\mathrm{d}y) = \int h(x,x)\bar{\beta}(x)\pi(\mathrm{d}x) = \int h(y,y)\bar{\beta}(y)\pi(\mathrm{d}y) = \iint h(x,y)\pi(\mathrm{d}y)\bar{\beta}(y)\delta_y(\mathrm{d}x)$, showing that $\pi(\mathrm{d}x)\bar{\beta}(x)\delta_x(\mathrm{d}y) = \pi(\mathrm{d}y)\bar{\beta}(y)\delta_y(\mathrm{d}x)$. Moreover, for any $x, y \in \mathbb{R}$, we have $\pi(x)\beta(x,y) = \pi(y)\beta(y,x)$, showing that $\pi(\mathrm{d}x)\beta(x,y)\mathrm{d}y = \pi(\mathrm{d}y)\beta(y,x)\mathrm{d}x$. Finally, we get

$$\pi(\mathrm{d}x)P(x, \mathrm{d}y) = \pi(\mathrm{d}y)P(y, \mathrm{d}x)$$

The Markov kernel P is therefore π-reversible and hence π-invariant. □

5. Is π the unique invariant probability measure for P?
Solution.

By the expression of P, we have $P(x,A) \geqslant \int_A \beta(x,y)\mathrm{d}y$. Hence, since β is positive, we get that $\lambda(A) = \int_A \mathrm{d}y > 0$ implies that $\int_A \beta(x,y)\mathrm{d}y > 0$ (since $\beta(x,y) > 0$) and hence, $P(x,A) > 0$. The Markov kernel P is λ-irreducible and therefore admits at most one invariant probability measure. From the previous question, $\pi P = \pi$ and finally, π is the unique invariant probability measure for P. □

6. According to which theorem, we can obtain that for any measurable function g such that $\pi(|g|) < \infty$, we have $\mathbb{P}_\pi$-a.s.

$$\lim_{k \to \infty} \frac{1}{n} \sum_{k=0}^{n-1} g(X_k) = \pi(g),$$

Solution.

The Markov kernel P admits a unique invariant probability measure, therefore we can apply Birkhoff's ergodic theorem and we get the required result. □

7. Let h be a **bounded non-negative** measurable function such that $\pi(h) = 0$ and $Ph(x) = h(x)$ for any $x \in \mathbb{R}$. Show that for any $x \in \mathbb{R}$, $h(x) = 0$.
Solution.

From the expression of P, $h(x) = Ph(x) = h(x)\bar{\beta}(x) + \int h(y)\beta(x,y)\mathrm{d}y$. Hence,

$$h(x)(1-\bar{\beta}(x)) = \int h(y)\beta(x,y)\mathrm{d}y = \int \underbrace{\left(\int f(v)f(v^*)\min\left(\frac{q_{v^*}(x)}{\pi(x)}, \frac{q_v(y)}{\pi(y)}\right)\mathrm{d}v\mathrm{d}v^*\right)}_{\geqslant 0} \underbrace{h(y)}_{\geqslant 0}\pi(y)\mathrm{d}y = 0$$

where the last equality follows from $\pi(h) = 0$. Finally $h(x)(1-\bar{\beta}(x)) = 0$ for all $x \in \mathbb{R}$. But

$$1 - \bar{\beta}(x) = \int \beta(x,y)\mathrm{d}y > 0 \quad \text{since } \beta(x,y) > 0$$

Finally, we obtain $h(x) = 0$ for any $x \in \mathbb{R}$.

□

8. Let g be a measurable function such that $\pi(|g|) < \infty$. Define

$$A = \left\{ \lim_{n\to\infty} \frac{\sum_{i=0}^{n-1} g(X_i)}{n} = \pi(g) \right\}$$

For any $x \in \mathbb{R}$, $h(x) = \mathbb{E}_x[\mathbf{1}_{A^c}] = \mathbb{P}_x(A^c)$. We admit that $Ph(x) = h(x)$ for any $x \in \mathbb{R}$. Show that $\pi(h) = 0$. Deduce from the previous questions that the Law of Large Numbers actually holds for $\{X_n : n \in \mathbb{N}\}$ starting from any initial distribution, that is, for any probability measure ξ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, we have $\mathbb{P}_\xi - a.s.$,

$$\lim_{n\to\infty} \frac{\sum_{i=0}^{n-1} g(X_i)}{n} = \pi(g)$$

Solution.

We have $\pi(h) = \int \pi(\mathrm{d}x)\mathbb{P}_x(A^c) = \mathbb{P}_\pi(A^c) = 0$ where the last equality follows from Question 6. Hence, since $Ph = h$, the previous question shows that $h(x) = 0$ for all $x \in \mathbb{R}$. Finally, $\xi(h) = \int \xi(\mathrm{d}x)\mathbb{P}_x(A^c) = \mathbb{P}_\xi(A^c) = 0$. This concludes the proof. □

We will need to define the following assumption:

(**A1**) there exists a sequence of real numbers $(x_n)_{n\in\mathbb{N}}$ such that $\frac{\pi(x_n)}{\int f(v^*)q_{v^*}(x_n)\mathrm{d}v^*} = \infty$

In Questions 9-10-11, we assume (**A1**)

9. Show using Jensen's inequality with the concave function $u \mapsto \min(cu, 1)$ (where c is a positive constant), that for any $x, y \in \mathbb{R}$, $v \in \mathbb{R}$,

$$\int f(v^*)\min\left(\frac{\pi(y)q_{v^*}(x)}{\pi(x)q_v(y)}, 1\right)\mathrm{d}v^* \leqslant \min\left(\frac{\pi(y)\int f(v^*)q_{v^*}(x)\mathrm{d}v^*}{\pi(x)q_v(y)}, 1\right)$$

Deduce that $\lim_{n\to\infty}\bar{\beta}(x_n) = 1$.
Solution.

Taking $c = \frac{\pi(y)}{\pi(x)q_v(y)}$, we can apply Jensen's inequality with the concave function $u \mapsto \min(cu, 1)$, we get

$$\int f(v^*)\min\left(\frac{\pi(y)q_{v^*}(x)}{\pi(x)q_v(y)}, 1\right)\mathrm{d}v^* = \int f(v^*)\min\left(cq_{v^*}(x), 1\right)\mathrm{d}v^* \leqslant \min\left(c\int f(v^*)q_{v^*}(x)\mathrm{d}v^*, 1\right)$$
$$= \min\left(\frac{\pi(y)\int f(v^*)q_{v^*}(x)\mathrm{d}v^*}{\pi(x)q_v(y)}, 1\right)$$

Now, under (**A1**), there exists a sequence $\{x_n : n \in \mathbb{N}\}$ such that $\lim_{n\to\infty} \frac{\int f(v^*)q_{v^*}(x_n)\mathrm{d}v^*}{\pi(x_n)} = 0$. Then, using the previous inequality,

$$0 \leqslant 1 - \bar{\beta}(x_n) = \int\cdots\int f(v)f(v^*)q_v(y)\min\left(\frac{\pi(y)q_{v^*}(x_n)}{\pi(x_n)q_v(y)}, 1\right)\mathrm{d}v\mathrm{d}v^*\mathrm{d}y$$
$$\leqslant \iint f(v)q_v(y)\underbrace{\min\left(\frac{\pi(y)\int f(v^*)q_{v^*}(x_n)\mathrm{d}v^*}{\pi(x_n)q_v(y)}, 1\right)}_{\gamma_n(v,y)}\mathrm{d}y\mathrm{d}v$$

Note that $\gamma_n(v,y) \leqslant 1$ which is integrable wrt the probability measure $f(v)q_v(y)\mathrm{d}y\mathrm{d}v$. Moreover, for any $(v,y) \in \mathbb{R}\times\mathbb{R}$,

$$\gamma_n(v,y) = \min\left(\frac{\pi(y)\int f(v^*)q_{v^*}(x_n)\mathrm{d}v^*}{\pi(x_n)q_v(y)}, 1\right) \leqslant \frac{\pi(y)\int f(v^*)q_{v^*}(x_n)\mathrm{d}v^*}{\pi(x_n)q_v(y)} \to_{n\to\infty} 0.$$

The dominated convergence theorem then shows that

$$\lim_{n\to\infty} 1 - \bar{\beta}(x_n) = 0$$

which completes the proof of the question. □

Recall that a Markov kernel P is geometrically ergodic, if there exists a measurable non-negative function $V : \mathbb{R} \to \mathbb{R}^+$ and constants ρ such that $0 < \rho < 1$ satisfying: for any $n \in \mathbb{N}$ and any $x \in \mathbb{R}$,

$$\|P^n(x,\cdot) - \pi\|_{TV} \leqslant V(x)\rho^n. \tag{1}$$

We will show by contradiction that under condition (**A1**) (see above), the Markov kernel P cannot be geometrically ergodic.

10. Under (**A1**), using that $\pi(\{x\}) = 0$ for any singleton $\{x\}$, show that (1) implies that for any $x \in \mathbb{R}$ and any $n \in \mathbb{N}$, $2\bar{\beta}^n(x) \leqslant V(x)\rho^n$.
Solution.

We have for any $x \in \mathbb{R}$

$$\|P^n(x,\cdot) - \pi\|_{TV} = 2\sup\{|P^n f(x) - \pi(f)| \ : \ f \text{ measurable and } 0 \leqslant f \leqslant 1\}$$
$$\geqslant 2|P^n(x,\{x\}) - \pi(\{x\})|$$

where the last inequality follows by setting $f(u) = \mathbf{1}_{\{x\}}(u) \in [0,1]$. Noting that $\pi(\{x\}) = \int_{\{x\}}\pi(u)\mathrm{d}u = 0$ (since the Lebesgue measure of a singleton is null), we get $\|P^n(x,\cdot) - \pi\|_{TV} \geqslant 2P^n(x,\{x\}) \geqslant 2\mathbb{P}_x(X_1 = x,\ldots,X_n = x) = 2\bar{\beta}^n(x)$. Hence if (1) holds, then for any $x \in \mathbb{R}$ and any $n \in \mathbb{N}$, $2\bar{\beta}^n(x) \leqslant \|P^n(x,\cdot) - \pi\|_{TV} \leqslant V(x)\rho^n$. □

11. Show that if (**A1**) holds, then P is not geometrically ergodic.
Solution.

Since there exists a sequence of real numbers $(x_n)_{n\in\mathbb{N}}$ satisfying $\lim_{n\to\infty}\bar{\beta}(x_n) = 1$, there exists $x_\star$ such that $1 > \bar{\beta}(x_\star) > \rho$. By the previous question, we have for all $n \in \mathbb{N}$,

$$2\left(\frac{\bar{\beta}(x_\star)}{\rho}\right)^n \leqslant V(x_\star)$$

Letting $n \to \infty$, we finally get $\infty \leqslant V(x_\star)$ which contradicts $V(x_\star) \in \mathbb{R}^+$. By contradiction, we have proved that under (A1), the AuxMH is not geometrically ergodic. □

12. (*This question is independent from the previous questions.*) Let X, Y two independent random variables on the same probability space such that $\mathbb{E}[|X|] < \infty$ and $\mathbb{E}[|Y|] < \infty$. Using Jensen's inequality, show that $\mathbb{E}[X \wedge Y] \leqslant \mathbb{E}[X \wedge \mathbb{E}[Y]]$ where we have used the notation $u \wedge v = \min(u,v)$. Deduce that $\mathbb{E}[X \wedge Y] \leqslant \mathbb{E}[X] \wedge \mathbb{E}[Y]$.
Solution.

Since for any constant c, the function $u \mapsto c \wedge u$ is concave, Jensen's inequality shows that $\mathbb{E}[X \wedge Y|X] \leqslant X \wedge \mathbb{E}[Y|X] = X \wedge \mathbb{E}[Y]$. Taking the expectation on both sides of the inequality, we obtain $\mathbb{E}[X \wedge Y] \leqslant \mathbb{E}[X \wedge \mathbb{E}[Y]]$. Now, using again Jensen's inequality for the concave function $u \mapsto u \wedge \mathbb{E}[Y]$, we obtain $\mathbb{E}[X \wedge \mathbb{E}[Y]] \leqslant \mathbb{E}[X] \wedge \mathbb{E}[Y]$. Finally, we obtain $\mathbb{E}[X \wedge Y] \leqslant \mathbb{E}[X] \wedge \mathbb{E}[Y]$. □

13. Show that for any bounded measurable function h, we have $\mathbb{E}[h(Y_t)|X_{t-1}] = \int q(y)h(y)\mathrm{d}y$ where $q(y)$ can be expressed using the functions f and $q_v(y)$.
Solution.

From the description of the algorithm,

$$\mathbb{E}[h(Y_t)|X_{t-1}] = \iint f(v)q_v(y)h(y)\mathrm{d}v\mathrm{d}y = \int q(y)h(y)\mathrm{d}y$$

where we have set $q(y) = \int f(v)q_v(y)\mathrm{d}v$. □

14. Denote by P_1 the kernel of a Metropolis Hasting algorithm with proposal density $y \mapsto q(y)$ and acceptance probability α^{MH} as defined in the course. Show that for any $(x,A) \in \mathbb{R} \times \mathcal{B}(\mathbb{R})$, we have $P(x,A \setminus \{x\}) \leqslant P_1(x,A \setminus \{x\})$.
Solution.

For any $(x,A) \in \mathbb{R} \times \mathcal{B}(\mathbb{R})$,

$$\begin{aligned}
P_1(x,A \setminus \{x\}) &= \int_{A\setminus\{x\}} q(y)\alpha^{MH}(x,y)\mathrm{d}y \\
&= \int_{A\setminus\{x\}} q(y)\min\left(\frac{\pi(y)q(x)}{\pi(x)q(y)},1\right)\mathrm{d}y = \int_{A\setminus\{x\}} \pi(y)\left(\frac{q(x)}{\pi(x)} \wedge \frac{q(y)}{\pi(y)}\right)\mathrm{d}y \\
&= \int_{A\setminus\{x\}} \pi(y)\left(\frac{\int f(v^*)q_{v^*}(x)\mathrm{d}v}{\pi(x)} \wedge \frac{\int f(v)q_v(y)\mathrm{d}v}{\pi(y)}\right)\mathrm{d}y \\
&\geqslant \int_{A\setminus\{x\}} \pi(y)\left[\iint f(v)f(v^*)\left(\frac{q_{v^*}(x)}{\pi(x)} \wedge \frac{q_v(y)}{\pi(y)}\right)\mathrm{d}v\mathrm{d}v^*\right]\mathrm{d}y
\end{aligned}$$

where the last inequality follows from Question 12. Hence,

$$P_1(x,A \setminus \{x\}) \geqslant \int_{A\setminus\{x\}} \left[\iint f(v)f(v^*)q_v(y)\left(\frac{\pi(y)q_{v^*}(x)}{\pi(x)q_v(y)} \wedge 1\right)\mathrm{d}v\mathrm{d}v^*\right]\mathrm{d}y = P(x,A \setminus \{x\})$$

□

15. *(Optional)* Intuitively, which algorithm do you think performs better? Compare their advantages and limitations.
Solution.

Intuitively, both P_1 and P are π-reversible kernels. However, since $P(x,A \setminus \{x\}) \leqslant P_1(x,A \setminus \{x\})$, we can infer that a Markov chain with kernel P_1 achieves better exploration than one based on P. Nevertheless, using P_1 requires computing $q(y) = \int f(v)\,q_v(y)\,\mathrm{d}v$, which is not always tractable. □