Problem

Let

- $\pi(\mathrm{d}y) = \pi(y)\mathrm{d}y$ be probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. As stressed by the expression $\pi(\mathrm{d}y) = \pi(y)\mathrm{d}y$, we assume that the measure π has a density on $\mathbb{R}$ with respect to the Lebesgue measure and by abuse of notation, we will also call π this density.
- $\phi(\mathrm{d}y) = \phi(y)\mathrm{d}y$ be another probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Again, the expression $\phi(\mathrm{d}y) = \phi(y)\mathrm{d}y$ means that we assume that the measure ϕ has a density on $\mathbb{R}$ with respect to the Lebesgue measure and by abuse of notation, we will also call ϕ this density.

In all the exercise, we assume that we can draw according to ϕ and that there exists a constant $\varepsilon > 0$ such that

$$\boxed{(A1) \qquad \forall x \in \mathbb{R}, \quad \pi(x) > \varepsilon\phi(x) > 0}$$

We now construct a family of random variables $(Z_t)_{t \geqslant 0}$ in the following way.

input : n
output: $Z_0, \ldots, Z_n$

At $t = 0$, draw $Z_0 \sim \mu$ where μ is arbitrary
for $t \leftarrow 1$ **to** n **do**

- Draw independently, $U_t \sim \mathrm{Unif}(0,1)$ and $Y_t \sim \phi$
- Letting $\beta : \mathbb{R} \to]0,1[$ be the function $\beta = \varepsilon\phi/\pi$, we set $Z_t = \begin{cases} Z_{t-1} & \text{if } U_t > \beta(Z_{t-1}) \\ Y_t & \text{if } U_t \leqslant \beta(Z_{t-1}) \end{cases}$

end

QUESTIONS

1. For any bounded measurable function $h : \mathbb{R} \to \mathbb{R}$, write $\mathbb{E}[h(Z_t)|Z_{t-1}]$. Deduce the expression of the Markov kernel P_1 associated to the Markov chain $(Z_k)_{k \in \mathbb{N}}$.
Solution.

For any bounded measurable function $h : \mathbb{R} \to \mathbb{R}$, we have

$$\begin{aligned}
\mathbb{E}[h(Z_t)|Z_{t-1}] &= \mathbb{E}[h(Z_{t-1})\mathbf{1}_{U_k > \beta(Z_{k-1})}|Z_{t-1}] + \mathbb{E}[h(Y_t)\mathbf{1}_{U_k \leqslant \beta(Z_{k-1})}|Z_{t-1}] \\
&= h(Z_{t-1})(1 - \beta(Z_{k-1})) + \beta(Z_{k-1}) \int \phi(y)h(y)\mathrm{d}y \\
&= \int \left[(1 - \beta(Z_{k-1}))\delta_{Z_{k-1}}(\mathrm{d}y) + \beta(Z_{k-1})\phi(y)\mathrm{d}y\right] h(y) \\
&= \int P_1(x, \mathrm{d}y)h(y)
\end{aligned}$$

where

$$P_1(x, \mathrm{d}y) = (1 - \beta(x))\delta_x(\mathrm{d}y) + \beta(x)\phi(y)\mathrm{d}y$$

□

2. Show that the Markov kernel P_1 is π-reversible.
Solution.

For any measurable function $h : \mathbb{R}^2 \to \mathbb{R}$,

$$\begin{aligned}
\int \pi(\mathrm{d}x)P_1(x, \mathrm{d}y)h(x,y) &= \int \pi(\mathrm{d}x)\left[(1 - \beta(x))\delta_x(\mathrm{d}y) + \beta(x)\phi(y)\mathrm{d}y\right] h(x,y) \\
&= \int \pi(\mathrm{d}x)(1 - \beta(x))h(x,x) + \int \pi(x)\frac{\varepsilon\phi(x)}{\pi(x)}\phi(y)h(x,y)\mathrm{d}x\mathrm{d}y \\
&= \int \pi(\mathrm{d}x)(1 - \beta(x))h(x,x) + \varepsilon \int \phi(x)\phi(y)h(x,y)\mathrm{d}x\mathrm{d}y
\end{aligned}$$

Similarly,

$$\int \pi(\mathrm{d}x)P_1(x,\mathrm{d}x)h(x,y) = \int \pi(\mathrm{d}y)(1-\beta(y))h(y,y) + \varepsilon \int \phi(y)\phi(x)h(x,y)\mathrm{d}x\mathrm{d}y$$

so that $\pi(\mathrm{d}x)P_1(x,\mathrm{d}y) = \pi(\mathrm{d}y)P_1(y,\mathrm{d}x)$ and P_1 is π-reversible.

□

3. Show that π is the unique invariant probability measure for P_1.
Solution.

For all $(x,A) \in \mathbb{R} \times \mathcal{B}(\mathbb{R})$ such that $\phi(A) = \int_A \phi(x)\mathrm{d}x > 0$, we have

$$P_1(x,A) \geqslant \beta(x) \int_A \phi(y)\mathrm{d}y = \beta(x)\phi(A) > 0$$

since $\beta(x) > 0$ for any $x \in \mathbb{R}$. Applying Proposition 2.10, we deduce that P_1 admits at most one invariant probability measure. Since it is π-reversible, it is also π-invariant. Therefore, π is the unique invariant probability measure for P_1.

□

4. Let $h : \mathbb{R} \to \mathbb{R}$ be a bounded measurable function such that $P_1 h = h$. Then, show that h is constant.
Solution.

For all $x \in \mathbb{R}$,

$$h(x) = P_1 h(x) = (1-\beta(x))h(x) + \beta(x) \int \phi(y)h(y)\mathrm{d}y$$

which is equivalent to $\beta(x)h(x) = \beta(x)\int \phi(y)h(y)\mathrm{d}y$. Dividing by $\beta(x)$ (since $\beta(x) > 0$), we get that h is constant. □

5. Let $(Z_k)_{k\in\mathbb{N}}$ be a Markov chain with Markov kernel P_1. Let $f : \mathbb{R} \to \mathbb{R}$ be a measurable function such that $\pi(|f|) < \infty$. Define

$$A = \left\{ \lim_{n\to\infty} n^{-1} \sum_{k=0}^{n-1} f(Z_k) = \pi(f) \right\}$$

Setting $h(x) = \mathbb{E}_x[\mathbf{1}_{A^c}] = \mathbb{P}_x(A^c)$, we admit that $P_1 h = h$ (it is actually proved in the Lecture Notes). Deduce from the previous question that the Law of Large Numbers holds for $(Z_k)_{k\in\mathbb{N}}$ starting from any initial distribution, i.e. for any probability measure ξ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$,

$$\lim_{n\to\infty} n^{-1} \sum_{k=0}^{n-1} f(Z_k) = \pi(f), \quad \mathbb{P}_\xi - a.s.$$

Solution.

Since P_1 admits a unique invariant probability measure π, the asssociated dynamical system is ergodic and the Birkhoff theorem then shows that $\mathbb{P}_\pi(A) = 1$ or equivalently $\mathbb{P}_\pi(A^c) = 0$. Then, $0 = \int \pi(\mathrm{d}x)\mathbb{P}_x(A^c) = \int \pi(\mathrm{d}x)h(x)$. Now, since $P_1 h = h$ and h is bounded, the previous question shows that h is constant. Combining with $\pi(h) = 0$, we get $h(x) = 0$ for all $x \in \mathbb{R}$. Hence, for probability measure ξ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$,

$$\mathbb{P}_\xi(A^c) = \int \xi(\mathrm{d}x)h(x) = 0$$

which is equivalent to $\mathbb{P}_\xi(A) = 1$ and the proof is completed. □

We now let

- $Q(x, \mathrm{d}y) = q(x,y)\mathrm{d}y$ be a Markov kernel on $\mathbb{R} \times \mathcal{B}(\mathbb{R})$. Thus, we assume that Q admits the Markov kernel density q with respect to the Lebesgue measure.
- P_0 be the Markov kernel associated to a "classical" Metropolis-Hastings algorithm, with proposal kernel Q and target distribution π, that is for any $x \in \mathbb{R}$,

$$P_0(x, \mathrm{d}y) = Q(x, \mathrm{d}y)\alpha(x,y) + \bar{\alpha}(x)\delta_x(\mathrm{d}y)$$

  where for any $x, y \in \mathbb{R}$,

$$\alpha(x,y) = \min\left(\frac{\pi(y)q(y,x)}{\pi(x)q(x,y)}, 1\right), \quad \bar{\alpha}(x) = 1 - \int Q(x, \mathrm{d}z)\alpha(x,z)$$

In addition to Assumption $(A1)$, we now also assume

$$\boxed{(A2) \qquad \forall x, y \in \mathbb{R}, \quad q(x,y) > 0}$$

We now construct a family of random variables $(X_t)_{t \geqslant 0}$ in the following way.

input : n
output: $X_0, \ldots, X_n$

At $t = 0$, draw $X_0 \sim \mu$ where μ is arbitrary
for $t \leftarrow 1$ **to** n **do**

- Draw independently, $X_t' \sim P_0(X_{t-1}, \cdot)$, $U_t \sim \mathrm{Unif}(0,1)$ and $Y_t \sim \phi$
- Letting $\beta : \mathbb{R} \to]0,1[$ be the function $\beta = \varepsilon\phi/\pi$, we set $X_t = \begin{cases} X_t' & \text{if } U_k > \beta(X_k') \\ Y_t & \text{if } U_k \leqslant \beta(X_k') \end{cases}$

end

QUESTIONS (CONTINUED)

6. For any bounded measurable function h, write $\mathbb{E}[h(X_t)|X_{t-1}]$ in terms of P_0,β and ϕ. Deduce that there exists functions γ_0 and γ_1 such that the Markov kernel P_2 associated to the Markov chain $(X_k)_{k \in \mathbb{N}}$ can be written as

$$P_2(x, \mathrm{d}y) = P_0(x, \mathrm{d}y)\gamma_0(y) + \gamma_1(x)\phi(y)\mathrm{d}y$$

   and give the expressions of the functions γ_0 and γ_1.
   Solution.

For any bounded measurable function h,

$$\begin{aligned}
\mathbb{E}[h(X_t)|X_{t-1}] &= \mathbb{E}[h(X_t')\mathbf{1}_{U_k > \beta(X_k')}|X_{t-1}] + \mathbb{E}[h(Y_t)\mathbf{1}_{U_k \leqslant \beta(X_k')}|X_{t-1}] \\
&= \int P_0(X_{t-1}, \mathrm{d}x')\left(\int_0^1 \mathbf{1}_{u > \beta(x')}\mathrm{d}u\right)h(x') + \int P_0(X_{t-1}, \mathrm{d}x')\left(\int_0^1 \mathbf{1}_{u \leqslant \beta(x')}\mathrm{d}u\right)\phi(y)h(y)\mathrm{d}y \\
&= \int P_0(X_{t-1}, \mathrm{d}x')(1 - \beta(x'))h(x') + \int P_0(X_{t-1}, \mathrm{d}x')\beta(x')\phi(y)h(y)\mathrm{d}y \\
&= \int \left[P_0(X_{t-1}, \mathrm{d}y)(1 - \beta(y)) + \int P_0(X_{t-1}, \mathrm{d}x')\beta(x')\phi(y)\mathrm{d}y\right]h(y) \\
&= \int P_2(X_{t-1}, \mathrm{d}y)h(y)
\end{aligned}$$

where

$$P_2(x, \mathrm{d}y) = P_0(x, \mathrm{d}y)\gamma_0(y) + \gamma_1(x)\phi(y)\mathrm{d}y, \quad \gamma_0(y) = 1 - \beta(y), \quad \gamma_1(x) = \int P_0(x, \mathrm{d}x')\beta(x')$$

□

7. Check that $P_2 = P_0 P_1$

   Solution.

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

$$
\begin{aligned}
P_0P_1(x,A) &= \int P_0(x,\mathrm{d}x') \int P_1(x',\mathrm{d}y)\mathbf{1}_A(y) \\
&= \int P_0(x,\mathrm{d}x') \int \left[(1-\beta(x'))\delta_{x'}(\mathrm{d}y) + \beta(x')\phi(y)\mathrm{d}y\right] \mathbf{1}_A(y) \\
&= \int P_0(x,\mathrm{d}x')(1-\beta(x'))\mathbf{1}_A(x') + \int P_0(x,\mathrm{d}x')\beta(x') \left(\int \phi(y)\mathrm{d}y\mathbf{1}_A(y)\right) \\
&= \int P_0(x,\mathrm{d}x')\gamma_0(x')\mathbf{1}_A(x') + \gamma_1(x) \left(\int \phi(y)\mathrm{d}y\mathbf{1}_A(y)\right) \\
&= \int \left[P_0(x,\mathrm{d}y)\gamma_0(y) + \gamma_1(x)\int \phi(y)\mathrm{d}y\right] \mathbf{1}_A(y) = P_2(x,A)
\end{aligned}
$$

□

8. Show that π is invariant for the Markov kernel P_2.
Solution.

Since P_0 and P_1 are π-invariant, we have $\pi P_2 = \pi P_0 P_1 = \pi P_1 = \pi$, which concludes the proof. □

9. Can we say that π is the unique invariant probability distribution for P_2?
Solution.

For any $(x,A) \in \mathbb{R} \times \mathcal{B}(\mathbb{R})$ such that $\phi(A) = \int_A \phi(x)\mathrm{d}x > 0$,

$$
P_2(x,A) \geqslant \gamma_1(x) \int_A \gamma_2(y)\mathrm{d}y = \gamma_1(x)\phi(A) > 0
$$

where we have used that

$$
\gamma_1(x) = \int P_0(x,\mathrm{d}x')\beta(x') > 0 \quad \text{since} \quad \beta(x') > 0, \forall x' \in \mathbb{R}
$$

Hence, by Proposition 2.10, P_2 admits at most one invariant probability measure and combining with the previous question, π is the unique invariant probability distribution for P_2. □

10. **(More Difficult)** Show that the Law of Large Numbers for $(X_t)_{t\geqslant 0}$ holds starting from any initial distribution ξ.
Solution.

Let $f : \mathbb{R} \to \mathbb{R}$ be a measurable function such that $\pi(|f|) < \infty$. Define

$$
A = \left\{ \lim_{n\to\infty} n^{-1} \sum_{k=0}^{n-1} f(X_k) = \pi(f) \right\}
$$

Since π is the unique invariant probability measure for P_2, we have by Birkhoff's ergodic theorem $0 = \mathbb{P}_\pi(A^c) = \int \pi(\mathrm{d}x)h(x) = \pi(h)$ where $h(x) = \mathbb{E}_x[\mathbf{1}_{A^c}] = \mathbb{P}_x(A^c)$. We now show that $h(x) = 0$ for all $x \in \mathbb{R}$. As seen in Question 5, we have $P_2 h = h$. Then, for all $x \in \mathbb{R}$,

$$
h(x) = P_2 h(x) = \int P_0(x,\mathrm{d}x')\left(\gamma_0(x')h(x')\right) + \gamma_1(x)\int \phi(y)h(y)\mathrm{d}y
$$

Since $\pi \geqslant \varepsilon\phi$, we get $\varepsilon\phi(h) \leqslant \pi(h) = 0$, this implies that $\phi(h) = 0$. Therefore the second term of the right-hand side cancels and we get

$$
h(x) = \int P_0(x,\mathrm{d}x')\left(\gamma_0(x')h(x')\right) = \int q(x,y)\alpha(x,y)\gamma_0(y)h(y)\mathrm{d}y + \bar{\alpha}(x)\gamma_0(x)h(x)
$$

Therefore, gathering all the terms including $h(x)$, we obtain

$$
(1-\bar{\alpha}(x)\gamma_0(x))\,h(x) = \int \underbrace{\frac{q(x,y)\alpha(x,y)\gamma_0(y)}{\pi(y)}}_{\psi_x(y)} \pi(y)h(y)\mathrm{d}y
$$

Since for all $y \in \mathbb{R}$, we have $\pi(y) > 0$, we deduce $0 \leqslant \psi_x(y) < \infty$ and since $\int \pi(y)h(y)\mathrm{d}y = 0$, we get by the monotone convergence

$$
0 \leqslant \int \psi_x(y)\pi(y)h(y)\mathrm{d}y = \lim_{M\to\infty} \int_{\psi_x(y)\leqslant M} \psi_x(y)\pi(y)h(y)\mathrm{d}y \leqslant \lim_{M\to\infty} M \underbrace{\int \pi(y)h(y)\mathrm{d}y}_{=0} = 0
$$

This implies $(1-\bar{\alpha}(x)\gamma_0(x))h(x)=0$ and since $0 \leqslant \gamma_0 = 1-\varepsilon\frac{\phi}{\pi} < 1$ and $0 \leqslant \bar{\alpha} \leqslant 1$ we get that $\bar{\alpha}\gamma < 1$. Therefore, for any $x \in \mathbb{R}$, $1-\bar{\alpha}(x)\gamma_0(x) \neq 0$ and hence $h(x)=0$ for any $x \in \mathbb{R}$. Finally,

$$\mathbb{P}_\xi(A^c) = \int \xi(\mathrm{d}x)\underbrace{h(x)}_{=0} = 0$$

which finally proves that for any initial distribution ξ,

$$\lim_{n\to\infty} n^{-1}\sum_{k=0}^{n-1} f(X_k) = \pi(f)\,, \quad \mathbb{P}_\xi - a.s.$$

□