Problem: the i-SIR Markov chain

In all the problem, π, resp. $\tilde{\pi}$, are probability measures on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ (where $\mathcal{B}(\mathbb{R})$ is the Borel sigma-field on $\mathbb{R}$) and we assume that these distributions have **strictly positive densities** with respect to the Lebesgue measure λ. For simplicity, we also denote by π, resp. $\tilde{\pi}$, their densities with respect to λ, that is, $\pi(\mathrm{d}x) = \pi(x)\lambda(\mathrm{d}x) = \pi(x)\mathrm{d}x$ and $\tilde{\pi}(\mathrm{d}x) = \tilde{\pi}(x)\lambda(\mathrm{d}x) = \tilde{\pi}(x)\mathrm{d}x$ where $\boxed{\pi(x) > 0 \text{ and } \tilde{\pi}(x) > 0 \text{ for any } x \in \mathbb{R}}$ and where we recall the abuse of notation $\lambda(\mathrm{d}x) = \mathrm{d}x$.

For integers $i \leqslant j$, the notation $[i : j]$ stands for $\{i, i+1, \ldots, j\}$. Define $\boxed{w(x) = \frac{\pi(x)}{\tilde{\pi}(x)}}$.

The i-SIR algorithm with d proposals (also called i-SIR(d)) 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 $k \leftarrow 1$ **to** n **do**
- Set $Y_0 = X_{k-1}$ and draw independently d random variables $Y_i \sim \tilde{\pi}$ for $i = 1, \ldots, d$.
- Draw a random variable J taking values on $[0 : d]$ with probabilities: $\mathbb{P}(J = k) = \frac{w(Y_k)}{\sum_{\ell=0}^d w(Y_\ell)}$ for $k \in [0 : d]$.
- Set $X_k = Y_J$.

end

Algorithm 1: the i-SIR(d) algorithm

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

$$\mathbb{E}_\mu[f(X_k)\mathbf{1}_{\{J=0\}} | X_{k-1}] = f(X_{k-1})\beta(X_{k-1})$$

where $\beta(x) = \int \cdots \int \frac{w(x)}{w(x) + \sum_{i=1}^d w(y_i)} \tilde{\pi}(\mathrm{d}y_1) \ldots \tilde{\pi}(\mathrm{d}y_d)$

Solution.

$f(X_k)\mathbf{1}_{\{J=0\}} = f(X_{k-1})\mathbf{1}_{\{J=0\}}$, hence

$$\mathbb{E}_\mu[f(X_k)\mathbf{1}_{\{J=0\}} | X_{k-1}] = f(X_{k-1})\mathbb{P}_\mu(J = 0 | X_{k-1}) = f(X_{k-1})\beta(X_{k-1})$$

where $\beta(x) = \int \cdots \int \frac{w(x)}{w(x) + \sum_{i=1}^d w(y_i)} \tilde{\pi}(\mathrm{d}y_1) \ldots \tilde{\pi}(\mathrm{d}y_d)$ □

2. Show that for any $\ell \in [1 : d]$,

$$\mathbb{E}_\mu[f(X_k)\mathbf{1}_{\{J=\ell\}} | X_{k-1}] = \int f(y_1) \left(\int \cdots \int \frac{1}{w(X_{k-1}) + \sum_{i=1}^d w(y_i)} \tilde{\pi}(\mathrm{d}y_2) \ldots \tilde{\pi}(\mathrm{d}y_d) \right) \pi(\mathrm{d}y_1)$$

Solution.

$$\mathbb{E}_\mu[f(X_k)\mathbf{1}_{\{J=\ell\}} | X_{k-1}] = \mathbb{E}_\mu[f(Y_\ell)\mathbf{1}_{\{J=\ell\}} | X_{k-1}] = \int \cdots \int f(y_\ell) \frac{w(y_\ell)}{w(X_{k-1}) + w(y_\ell) + \sum_{i \in [1:d] \setminus \{\ell\}} w(y_i)} \tilde{\pi}(\mathrm{d}y_\ell) \prod_{i \in [1:d] \setminus \{\ell\}} \tilde{\pi}(\mathrm{d}y_d)$$

$$= \int \cdots \int f(y_\ell) \frac{1}{w(X_{k-1}) + w(y_\ell) + \sum_{i \in [1:d] \setminus \{\ell\}} w(y_i)} \pi(\mathrm{d}y_\ell) \prod_{i \in [1:d] \setminus \{\ell\}} \tilde{\pi}(\mathrm{d}y_d)$$

where we have used $w(y_\ell)\tilde{\pi}(\mathrm{d}y_\ell) = \pi(\mathrm{d}y_\ell)$. The proof follows by renaming differently the variables and rearranging the terms. □

3. Deduce that the Markov kernel P of the Markov chain $\{X_k \,:\, k \in \mathbb{N}\}$ writes

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

where $\gamma(x, y)$ should be expressed explicitely. Check that for any $x, y \in \mathbb{R}$, we have $\gamma(x, y) = \gamma(y, x)$ and $\gamma(x, y) > 0$.

Solution.

We have

$$\mathbb{E}_\mu(f(X_k)|X_{k-1}) = \sum_{\ell=0}^{d} \mathbb{E}_\mu(f(X_k)\mathbf{1}_{\{J=\ell\}}|X_{k-1})$$
$$= f(X_{k-1})\beta(X_{k-1}) + d\int\cdots\int f(y_1)\frac{1}{w(X_{k-1})+\sum_{i=1}^d w(y_i)}\pi(\mathrm{d}y_1)\tilde\pi(\mathrm{d}y_2)\ldots\tilde\pi(\mathrm{d}y_d) = \int P(X_{k-1},\mathrm{d}y)f(y)$$

where $P(x,\mathrm{d}y) = \delta_x(\mathrm{d}y)\beta(x) + \pi(\mathrm{d}y)\gamma(x,y)$ and

$$\gamma(x,y) = d\int\cdots\int \frac{1}{w(x)+w(y)+\sum_{j=2}^d w(y_j)}\tilde\pi(\mathrm{d}y_2)\ldots\tilde\pi(\mathrm{d}y_d)$$

Obviously, for any $x,y\in\mathbb{R}$, we have $\gamma(x,y)=\gamma(y,x)$ and $\gamma(x,y)>0$.

□

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

We have
$$\pi(\mathrm{d}x)P(x,\mathrm{d}y) = \pi(\mathrm{d}x)[\delta_x(\mathrm{d}y)\beta(x)+\pi(\mathrm{d}y)\gamma(x,y)] = \pi(\mathrm{d}x)\delta_x(\mathrm{d}y)\beta(x)+\pi(\mathrm{d}x)\pi(\mathrm{d}y)\gamma(x,y)$$
Note that for any bounded measurable function h, $\iint h(x,y)\pi(\mathrm{d}x)\delta_x(\mathrm{d}y) = \int h(x,x)\pi(\mathrm{d}x) = \int h(y,y)\pi(\mathrm{d}y) = \iint h(x,y)\pi(\mathrm{d}y)\delta_y(\mathrm{d}x)$, showing that $\pi(\mathrm{d}x)\delta_x(\mathrm{d}y) = \pi(\mathrm{d}y)\delta_y(\mathrm{d}x)$. Moreover, for any $x,y\in\mathbb{R}$, we have $\gamma(x,y)=\gamma(y,x)$, showing that $\pi(\mathrm{d}x)\pi(\mathrm{d}y)\gamma(x,y) = \pi(\mathrm{d}y)\pi(\mathrm{d}x)\gamma(y,x)$. 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 $P(x,A)\geqslant\int_A\pi(\mathrm{d}y)\gamma(x,y)$. Hence, since γ is positive, we get that $\pi(A)>0$ implies that $P(x,A)>0$. The Markov kernel P is π-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 f such that $\pi(|f|)<\infty$, we have $\mathbb{P}_\pi$-a.s.
$$\lim_{k\to\infty}\frac{1}{n}\sum_{k=0}^{n-1} f(X_k) = \pi(f),$$
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, $h(x) = Ph(x) = h(x)\beta(x)+\int h(y)\gamma(x,y)\pi(\mathrm{d}y)$. Hence,

$$h(x)(1-\beta(x)) = \int h(y)\gamma(x,y)\pi(\mathrm{d}y) = 0$$

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

$$\beta(x) = \int\cdots\int\frac{w(x)}{w(x)+\sum_{i=1}^d w(y_i)}\tilde\pi(\mathrm{d}y_1)\ldots\tilde\pi(\mathrm{d}y_d) < 1 \quad \text{since } w>0$$

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

□

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

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

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 $\mathbb{P}_\xi - a.s.$,

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

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. □

9. We now assume that

   (A1) $\sup_{x\in\mathbb{R}} w(x) = \infty$

   Show that there exists a sequence of real numbers $(x_n)_{n\in\mathbb{N}}$ such that $\lim_{n\to\infty} \beta(x_n) = 1$.
   Solution.

   Under (A1), there exists a sequence $\{x_n : n \in \mathbb{N}\}$ such that $\lim_{n\to\infty} w(x_n) = \infty$. Then,

   $$\beta(x_n) = \int\cdots\int \underbrace{\frac{w(x_n)}{w(x_n) + \sum_{i=1}^d w(y_i)}}_{g_n(y_{1:d})} \tilde{\pi}(\mathrm{d}y_1)\ldots\tilde{\pi}(\mathrm{d}y_d)$$

   We have $g_n(y_{1:d}) \leqslant 1$ which is integrable wrt the probability measure $\tilde{\pi}(\mathrm{d}y_1)\ldots\tilde{\pi}(\mathrm{d}y_d)$. Moreover, $\lim_{n\to\infty} g_n(y_{1:d}) = 1$. The dominated convergence theorem then shows that

   $$\lim_{n\to\infty}\beta(x_n) = \lim_{n\to\infty}\int\cdots\int g_n(y_{1:n})\tilde{\pi}(\mathrm{d}y_1)\ldots\tilde{\pi}(\mathrm{d}y_d) = \int\cdots\int \lim_{n\to\infty} g_n(y_{1:n})\tilde{\pi}(\mathrm{d}y_1)\ldots\tilde{\pi}(\mathrm{d}y_d) = 1$$

   □

10. 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) (of the previous question), the Markov kernel P cannot be geometrically ergodic. Indeed, using that $\pi(\{x\}) = 0$ for any singleton $\{x\}$, show that (1) implies that for any $x \in \mathbb{R}$ and any $n \in \mathbb{N}$, $2\beta^n(x) \leqslant V(x)\rho^n$.
    Solution.

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

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

    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 $\|P^n(x,\cdot) - \pi\|_{TV} \geqslant 2P^n(x,\{x\}) \geqslant 2\mathbb{P}_x(X_1 = x, \ldots, X_n = x) = 2\beta^n(x)$. Hence if (1) holds, then for any $x \in \mathbb{R}$ and any $n \in \mathbb{N}$, $2\beta^n(x) \leqslant \|P^n(x,\cdot) - \pi\|_{TV} \leqslant V(x)\rho^n$. □

11. Conclude.
    Solution.

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

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

    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 i-SIR(d) is not geometrically ergodic. □