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 $\pi(x)\beta(x,y) = \pi(y)\beta(y,x)$.

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.

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 $\bar{\beta}$.

4. Show that P is π-invariant.

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

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 $x \in \mathbb{R}$, $h(x) = 0$.

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 $\mathbb{P}_\xi - a.s.$,
$$\lim_{n\to\infty} \frac{\sum_{i=0}^{n-1} g(X_i)}{n} = \pi(g)$$

We will need to define the following assumption:

($\mathbf{A1}$) 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 ($\mathbf{A1}$)

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

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 ($\mathbf{A1}$) (see above), the Markov kernel P cannot be geometrically ergodic.

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

11. Show that if ($\mathbf{A1}$) holds, then P 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 $\mathbb{E}[X \wedge Y] \leqslant \mathbb{E}[X] \wedge \mathbb{E}[Y]$.

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 $q_v(y)$.

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 $P(x,A \setminus \{x\}) \leqslant P_1(x,A \setminus \{x\})$.

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