Exam: From MCMC to data-based generative models

23 October 2023

i) Justify your answers and draw a box around your results.

ii) Duration 2 hours.

iii) All documents are allowed except electronic devices (laptop, tablets, phones...)

Exercise

We recall that the density of $\mathcal{N}(0, \sigma^2)$ is $f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}}$.

1. Consider the Markov chain produced by the following recursion

$$\forall k \geqslant 0 \,, \quad X_{k+1} = X_k + |\eta_{k+1}| \,,$$

where $\eta_1, \ldots, \eta_n, \ldots$ are i.i.d. of law $\mathcal{N}(0, \sigma^2)$. Write down $Q : (\mathbb{R}, \mathcal{B}(\mathbb{R})) \to [0, 1]$ its Markov kernel as $Q(x, \mathrm{d}y) = 2l(x, y)\, \mathrm{d}y$ for some function $l : \mathbb{R}^2 \to \mathbb{R}$ that you will specify.

2. We consider two intervals $I_1 = [a_1, b_1]$ and $I_2 = [a_2, b_2]$. Give a condition on a_1 and b_2 such that for any probability measure ν on $\mathbb{R}$,

$$\mathbb{P}_\nu(X_0 \in I_1, X_1 \in I_2) = 0 \,.$$

3. Is Q reversible relatively to any probability measure ν?

4. Show that for any initialization $\lim_{k \to +\infty} \frac{X_k}{k} = \mathbb{E}[|\eta_1|]$ almost-surely. To what quantity converges X_k?

Let $\pi(\mathrm{d}x) = \pi(x)\, \mathrm{d}x$ be a probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ with a density $\pi(x)$ relatively to the Lebesgue measure. We consider the extended space $\mathsf{X} = \mathbb{R} \times \{-1, 1\}$ with its product sigma-algebra $\mathcal{X}$ and the extended distribution Π on X defined for any $A \in \mathcal{B}(\mathbb{R})$ as $\Pi(A \times \{1\}) = \Pi(A \times \{-1\}) = \frac{1}{2}\pi(A)$. From this moment we are analyzing Algorithm 1.

5. Write down $P : \mathsf{X} \times \mathcal{X} \to [0, 1]$ the kernel of Algorithm 1.

6. We notice that P can be written as a product of two kernels $P = SR$, where $R(x, p, \mathrm{d}y, \mathrm{d}p') = \delta_x(\mathrm{d}y)\delta_{-p}(\mathrm{d}p')$. Write down the kernel $S : \mathcal{X} \to [0, 1]$. Notice that it is the kernel of a Metropolis-Hastings algorithm for which you will specify the proposal kernel $\bar{Q} : \mathcal{X} \to [0, 1]$ and the acceptance probability $\alpha(x, p_1, y, p_2)$.

7. Show the reversibility of S relatively to Π by checking the detailed balance condition. Show the reversibility of R.

8. Is Π an invariant distribution of P? Is it reversible? Justify your answers.

9. Let $x \in \mathbb{R}$ and $A \in \mathcal{B}(\mathbb{R})$ be of positive (Lebesgue) measure. What are the conditions on x, A to obtain $P(x, 1, A, \{1\}) > 0$? On what conditions $P(x, 1, A, \{1\}) = 0$?

10. Show that P is irreducible relatively to some irreducibility measure ν on $(\mathsf{X}, \mathcal{X})$ that you will specify. Deduce a way to approximate $\mathbb{E}_{X \sim \pi}[h(X)]$ through Algorithm 1, for any bounded function h.

Algorithm 1 Guided Random Walk. Input: $x_0 \in \mathbb{R}$, $\sigma > 0$.

$X_0 = x_0 \in \mathbb{R}$, $p_0 \in \{-1, 1\}$.
for $k \geqslant 0$ **do**
  Sample $Z_{k+1} \sim \mathcal{N}(0, \sigma^2)$ and U_{k+1} independently from a uniform distribution on $[0, 1]$.
  $Y_{k+1} = X_k + p_k |Z_{k+1}|$
  if $U_{k+1} \leqslant \frac{\pi(Y_{k+1})}{\pi(X_k)}$ **then**
    $(X_{k+1}, p_{k+1}) = (Y_{k+1}, p_k)$
  else
    $(X_{k+1}, p_{k+1}) = (X_k, -p_k)$
  end if
end for

Problem: the Rejection Markov chain

Notation. We use the following notation:

- $\mathsf{F}^+(\mathbb{R})$ is the set of non-negative measurable functions on $\mathbb{R}$
- $\mathsf{F}_b(\mathbb{R})$ is the set of bounded measurable functions on $\mathbb{R}$
- $\mathsf{F}_b^+(\mathbb{R}) = \mathsf{F}^+(\mathbb{R}) \cap \mathsf{F}_b(\mathbb{R})$
- For a given Markov kernel Q on $\mathbb{R} \times \mathcal{B}(\mathbb{R})$, we denote by $\mathbb{P}_\xi^Q$ the probability measure induced on $(\mathbb{R}^\mathbb{N}, \mathcal{B}(\mathbb{R})^{\otimes \mathbb{N}})$ by the Markov kernel Q and the initial distribution ξ. We write $\mathbb{E}_\xi^Q$ the associated expectation operator.

If $\xi = \delta_x$ for some $x \in \mathbb{R}$, we simply use $\mathbb{E}_x^Q := \mathbb{E}_{\delta_x}^Q$. In what follows, π, resp. $\tilde{\pi}$, are probability measures on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ 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 $\pi(x) > 0$ and $\tilde{\pi}(x) > 0$ for any $x \in \mathbb{R}$ and where we recall the abuse of notation $\lambda(\mathrm{d}x) = \mathrm{d}x$.

The following assumption holds in all the exercise:

(A0) The Markov kernel Q is $\tilde{\pi}$-invariant i.e. $\tilde{\pi}Q = \tilde{\pi}$

For any function $\rho : \mathbb{R} \to (0,1)$, define the rejection kernel S_ρ as follows:

$$\forall h \in \mathsf{F}_b^+(\mathbb{R}), \quad S_\rho h(x) = \sum_{k=1}^{\infty} \mathbb{E}_x^Q \left[h(\tilde{X}_k)\rho(\tilde{X}_k) \prod_{i=1}^{k-1} (1 - \rho(\tilde{X}_i)) \right] \tag{1}$$

with the convention $\prod_{i=1}^{0} = 1$. In words, a transition according to S_ρ can be seen as follows: from the current state x, we run a Markov chain $(\tilde{X}_i)_{i \geqslant 0}$ with Markov kernel Q and at each time step k, we flip a coin with success probability $\rho(\tilde{X}_k)$. The next state of our transition according to S_ρ corresponds to the first $\tilde{X}_k$ with a successful coin.

1. For any $h \in \mathsf{F}_b^+(\mathbb{R})$, show that $S_\rho h = Q(\rho h) + Q(\bar{\rho} S_\rho h)$ where we set $\bar{\rho} = 1 - \rho$.
2. Deduce that $\nu S_\rho = \nu$ where the probability measure ν is defined by

$$\nu(A) = \frac{\int_A \rho(x)\tilde{\pi}(\mathrm{d}x)}{\int_\mathbb{R} \rho(x)\tilde{\pi}(\mathrm{d}x)}, \quad A \in \mathcal{B}(\mathbb{R}), \tag{2}$$

   [**Hint**: you can use the previous question and apply $\tilde{\pi}$ on both sides of the equation]

3. **In the rest of the problem,** , we assume that there exists a constant M such that

$$\forall x \in \mathbb{R}, \quad \pi(x) < M\tilde{\pi}(x). \tag{3}$$

   Give a function ρ such that $\pi S_\rho = \pi$. In what follows, ρ is chosen as in this question.

4. Show that π is the unique invariant distribution for S_ρ.
5. According to which theorem, we can obtain that for any measurable function f such that $\pi(|f|) < \infty$, we have $\mathbb{P}_\pi - a.s.$,

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

   where $\{X_n \, : \, n \in \mathbb{N}\}$ is a Markov chain with Markov kernel S_ρ.

6. **In the rest of the problem**, we assume that any measurable bounded function h on $\mathbb{R}$ satisfying $Qh = h$ is a constant function. Assume that for some measurable bounded function h on $\mathbb{R}$, we have $S_\rho h = h$. Show that the function h is constant (**Hint:** Use the first question of the exercise).

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

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

Denote $h(x) = \mathbb{E}_x[\mathbf{1}_A]$. We admit that $S_\rho(h) = h$ (it is actually proved in the Lecture Notes). 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 $\mathbb{P}_\xi - a.s.$,

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