Exercise Session: MCMC methods

Exercise 1 (Simulation of a Pair)

Let (X, Y) be a random pair with joint density

$$f(x, y) = e^{-y}\mathbf{1}_{0\leq x\leq y}.$$

1. Determine the marginal density of X, denoted $f(x)$. What distribution do you recognize?

   We have

   $$f(x) = \int_{-\infty}^{\infty} f(x, y)dy = \int_{-\infty}^{\infty} e^{-y}\mathbf{1}_{0\leq x\leq y}dy = \int_{x}^{\infty} e^{-y}dy\mathbf{1}_{x\geq 0} = \Big[e^{-y}\Big]_{x}^{\infty} = e^{-x}\mathbf{1}_{x\geq 0}.$$

   We recognize the density of Exp(1).

2. Given $X = x \geq 0$, determine the conditional density $f(y|x)$. What distribution do you recognize?

   For any $x \geq 0$, we have

   $$f(y|x) = \frac{f(x, y)}{f(x)} = \frac{e^{-y}\mathbf{1}_{0\leq x\leq y}}{e^{-x}} = \exp(x - y)\mathbf{1}_{y-x\geq 0}.$$

   We recognize the density of $x + \mathrm{Exp}(1)$

3. Deduce a method to simulate a realization of the random pair (X, Y).

   Simulate $X, Z \sim \mathrm{Exp}(1)$ iid and ouput $(X, X + Z)$.

4. Given $Y = y \geq 0$, determine the conditional density $f(x|y)$. What distribution do you recognize?

   Given $y \geq 0$, we have

   $$f(y) = \int_{-\infty}^{\infty} e^{-y}\mathbf{1}_{0\leq x\leq y}dx = ye^{-y}.$$

   $$f(x|y) = \frac{f(x, y)}{f(y)} = \frac{e^{-y}\mathbf{1}_{0\leq x\leq y}}{ye^{-y}} = \frac{1}{y}\mathbf{1}_{0\leq x\leq y}.$$

   Density of $\mathrm{Unif}([0, y])$.

5. Starting, for example, from the point $(x_0, y_0) = (0, 1)$, write the pseudo-code of a Gibbs sampler with deterministic scanning to obtain a sample $(X_1, Y_1) \ldots, (X_n, Y_n)$ to sample from the target law f.

   We obtain the following pseudo-code

   Algorithm 1: Deterministic scan Gibbs sampler

   Initialization: $(x_0, y_0) = (0, 1)$

   For $t = 1, \ldots, n$:

   $x_t \sim \mathrm{Unif}([0, y_{t-1}])$

   $z_t \sim \mathrm{Exp}(1)$

   $y_t = x_t + z_t$

   Output $(x_0, y_0), \ldots (x_n, y_n)$.

6. Between the two proposed methods, which one would you choose to simulate according to the density $f(x,y)$?

   The first method directly generates iid samples, whereas the second generates a Markov chain, i.e. correlated samples. The first one should be preferred.

Exercise 2 (Metropolis-Hastings Algorithm)

Let E be a finite state space and π a probability measure on E such that $\pi(x) > 0$ for any $x \in E$. We recall the Metropolis-Hastings algorithm for an auxiliary transition kernel Q (such that $Q(x,y) = 0$ if and only if $Q(y,x) = 0$)

Step 0:
`Initialize` X_0 ;
Step n+1:
`Choose` y `according to the law` $Q(X_n,\cdot)$;
`Set` $X_{n+1} = y$ `with probability` $\min\left(1, \frac{\pi(y)Q(y,X_n)}{\pi(X_n)Q(X_n,y)}\right)$, `otherwise set` $X_{n+1} = X_n$

1) Show that if the acceptance-rejection probability $\min\left(1, \frac{\pi(y)Q(y,x)}{\pi(x)Q(x,y)}\right)$ is replaced by

$$\frac{\pi(y)Q(y,x)}{\pi(y)Q(y,x) + \pi(x)Q(x,y)},$$

   the measure π will still be invariant for the Markov chain defined by the Metropolis-Hastings algorithm.

2) More generally, replacing the acceptance-rejection probability $\min\left(1, \frac{\pi(y)Q(y,x)}{\pi(x)Q(x,y)}\right)$ by

$$\alpha\left(\frac{\pi(y)Q(y,x)}{\pi(x)Q(x,y)}\right),$$

   with $\alpha : \mathbb{R}_+ \to (0,1]$, give a sufficient condition on the function α for the measure π to be the invariant measure of the Markov chain defined by the Metropolis-Hastings algorithm.

   We address the two questions simultaneously. Let $x, y \in E$ such that $x \neq y$ and let us compute the coefficient $P(x,y)$ of the transition matrix P of the Markov chain.

$$\begin{aligned}
\mathbb{P}(X_{n+1} = y | X_n = x) &= \mathbb{P}\left(Y_{n+1} = y \text{ and } U_{n+1} \leq \alpha\left(\frac{\pi(Y_{n+1})Q(Y_{n+1},X_n)}{\pi(X_n)Q(X_n,Y_{n+1})}\right) \Big| X_n = x\right) \\
&= \mathbb{P}\left(Y_{n+1} = y \text{ and } U_{n+1} \leq \alpha\left(\frac{\pi(y)Q(y,x)}{\pi(x)Q(x,y)}\right) \Big| X_n = x\right) \\
&= \alpha\left(\frac{\pi(y)Q(y,x)}{\pi(x)Q(x,y)}\right) Q(x,y).
\end{aligned}$$

   We can now study the reversibility of π for the modified Metropolis-Hasting algorithm with acceptance probability $\alpha\left(\frac{\pi(y)Q(y,x)}{\pi(x)Q(x,y)}\right)$.

   The measure π is reversible for P if $\pi(x)P(x,y) = \pi(y)P(y,x)$ for any $x, y \in E$. This property is clear if $x = y$. Now, suppose $x \neq y$. Then

$$\begin{cases}
\pi(x)P(x,y) = \pi(x)Q(x,y)\,\alpha\left(\dfrac{\pi(y)Q(y,x)}{\pi(x)Q(x,y)}\right) \\
\pi(y)P(y,x) = \pi(y)Q(y,x)\,\alpha\left(\dfrac{\pi(x)Q(x,y)}{\pi(y)Q(y,x)}\right).
\end{cases}$$

Hence, $\pi(x)P(x,y) = \pi(y)P(y,x)$ iff

$$\pi(x)Q(x,y)\,\alpha\left(\frac{\pi(y)Q(y,x)}{\pi(x)Q(x,y)}\right) = \pi(y)Q(y,x)\,\alpha\left(\frac{\pi(x)Q(x,y)}{\pi(y)Q(y,x)}\right)$$

$$\iff \frac{\pi(x)Q(x,y)}{\pi(y)Q(y,x)}\,\alpha\left(\frac{\pi(y)Q(y,x)}{\pi(x)Q(x,y)}\right) = \alpha\left(\frac{\pi(x)Q(x,y)}{\pi(y)Q(y,x)}\right)$$

$$\iff z\,\alpha\left(\frac{1}{z}\right) = \alpha(z), \qquad \text{where} \qquad z = \frac{\pi(x)Q(x,y)}{\pi(y)Q(y,x)}.$$

For π to be reversible, it therefore suffices that $\forall z \in [0,1] : z\,\alpha\left(\frac{1}{z}\right) = \alpha(z)$.

This condition is met when $\alpha(z) = \min(1,z)$ since $z\min(1,1/z) = \min(z,1)$ if $z > 0$ (classical Metropolis-Hastings). Similarly, this condition holds when $\alpha(z) = \frac{1}{1+1/z}$, which is the setting of Question 1 (why?), as we have

$$z\alpha(1/z) = z \cdot \frac{1}{1+z} = \frac{1}{1+1/z} = \alpha(z).$$

Exercise 3 (Joint law and Gibbs sampler)

We consider the density

$$f(x,y) = C\exp\left(-\frac{y^2}{2} - \frac{x^2(y^2+4y+4)}{2}\right).$$

The renormalizing constant $C > 0$ is not specified here, but it does not matter for this exercise.

1. What is the law of X conditional on $Y = y$? What is the law of Y conditional on $X = x$?
   Conditional on $Y = y$, we have

$$f(x|y) = \frac{f(x,y)}{f(y)} = \frac{C\exp\left(-\frac{y^2}{2} - \frac{x^2(y^2+4y+4)}{2}\right)}{f(y)} = C'\exp\left(-\frac{x^2(y^2+4y+4)}{2}\right),$$

   where C' is independent of x. This is the density of $N(0, \frac{1}{(y+2)^2})$.

   Conditional on $X = x$, we have

$$\begin{aligned}
f(y|x) = \frac{f(x,y)}{f(x)} &= \frac{C\exp\left(-\frac{y^2}{2} - \frac{x^2(y^2+4y+4)}{2}\right)}{f(x)} \\
&= C''\exp\left(-\frac{1}{2}\left(y^2(1+x^2) + 4yx^2 + 4x^2\right)\right), \quad \text{where } C' \text{ is independent of } y \\
&= C''\exp\left(-\frac{(1+x^2)}{2}\left(y^2 + \frac{4yx^2}{(1+x^2)}\right)\right)\exp\left(-\frac{4x^2}{2}\right) \\
&= C''\exp\left(-\frac{(1+x^2)}{2}\left[\left(y + \frac{2yx^2}{(1+x^2)}\right)^2 - \frac{16x^4}{(1+x^2)^2}\right]\right)\exp\left(-\frac{4x^2}{2}\right) \\
&= C^*\exp\left(-\frac{(1+x^2)}{2}\left(y + \frac{2yx^2}{(1+x^2)}\right)^2\right).
\end{aligned}$$

   where C^* is independent of y. This is the density of $N\left(-\frac{2x^2}{1+x^2}, \frac{1}{1+x^2}\right)$.

2. Propose a deterministic scan Gibbs sampler pseudo-code to sample from the target f.

Algorithm 2: Deterministic scan Gibbs sampler

Initialization: Pick $(x_0, y_0) \in \mathbb{R}^2$

Until the termination condition:

$x_{t+1} \sim N(0, \frac{1}{(y_t+2)^2})$

$y_{t+1} \sim N\left(-\frac{2x_{t+1}^2}{(1+x_{t+1}^2)}, \frac{1}{(1+x_{t+1}^2)}\right)$.

Output $(x_0, y_0), \ldots, (x_T, y_T)$.

Exercise 4 (Markov chain)

For $\theta \in [0, \frac{1}{2}]$, we consider the Markov chain on $E = \{1, 2, 3\}$ with initial state $X_0 = 1$ and transition matrix

$$\Pi = \begin{pmatrix} \frac{1}{2} & \frac{\theta}{2} & \frac{1}{2}(1-\theta) \\ \theta & 1-2\theta & \theta \\ 0 & 1 & 0 \end{pmatrix}.$$

1. For what values of θ is the chain irreducible?

   The chain is not irreductible for $\theta = 0$, since $\mathbb{P}(X_1 = 2|X_0 = 2) = 1$ (one cannot go from 2 to another state different from 2 in finitely many steps).

   Conversely, the chain is irreducible for $\theta \in (0, 1/2]$. Indeed, 1 communicates with every other state, 3 communicates with 2 and 2 communicates with 1 and 3.

2. Suppose $\theta = 0$ and $X_1 = 1$. Compute $\mathbb{P}(X_3 = 1|X_1 = 1)$.

   If $\theta = 0$, then

$$\Pi = \begin{pmatrix} \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 1 & 0 \\ 0 & 1 & 0 \end{pmatrix}.$$

   We can check that

$$\Pi^2 = \begin{pmatrix} \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ 0 & 1 & 0 \\ 0 & 1 & 0 \end{pmatrix}.$$

   Therefore, $\mathbb{P}(X_3 = 1|X_1 = 1) = 1/4$.

3. Let $k \in \{1, 2, 3\}$. Suppose $\theta = \frac{1}{3}$. What is the almost sure limit of $\frac{1}{n}\sum_{j=1}^{n} \mathbf{1}_{X_j=k}$?

   The chain is irreducible and aperiodic, so the almost sure limit of $\frac{1}{n}\sum_{j=1}^{n} \mathbf{1}_{X_j=k}$ is the invariant probability measure $\pi = (x, y, z)$ such that $\pi\Pi = 1$, where

$$\Pi = \begin{pmatrix} \frac{1}{2} & \frac{1}{6} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 0 & 1 & 0 \end{pmatrix}$$

   Solving under the constraint that $x+y+z = 1$ since π is a probability vector, we get that $\pi = (\frac{3}{10}, \frac{9}{20}, \frac{1}{4})$.

Exercise 5 (Simultaneous Gibbs = Flawed Gibbs)

Consider the distribution $\pi = \pi_{X,Y}$ over $E = \{(0,0),(0,1),(1,0),(1,1)\}$ defined as follows

$$\pi(0,0) = \frac{2}{5}, \quad \text{and} \quad \pi(0,1) = \pi(1,0) = \pi(1,1) = \frac{1}{5}.$$

On this state space, we consider the Markov chain (X_n) defined by the following transitions: Letting $(x,y) = (X_t, Y_t)$, draw independently

$$X_{t+1} \sim \pi_{X|Y}\left(\cdot \mid y\right) \quad \text{and} \quad Y_{t+1} \sim \pi_{Y|X}\left(\cdot \mid x\right)$$

This way of moving in the state space is thus akin to a Gibbs sampler where all coordinates are updated simultaneously (that is to say, not using the most recent updates available so far).

1. Compute the probability distributions $\pi_{X|Y}\left(\cdot \mid y\right)$ and $\pi_{Y|X}\left(\cdot \mid x\right)$ for any $x, y \in \{0,1\}$ and show that $\pi_{X|Y} = \pi_{Y|X}$. Therefore, we will write

$$\pi_{a,b} := \pi_{X|Y}(a \mid b) = \pi_{Y|X}(a \mid b), \quad \forall a,b \in \{0,1\}.$$

Let π_X and π_Y denote the marginal distributions of X and Y, respectively. We have

$$\begin{aligned}\pi_X(0) &= \pi_{X,Y}(0,0) + \pi_{X,Y}(0,1) = \frac{2}{5} + \frac{1}{5} = \frac{3}{5} \\ \pi_X(1) &= \pi_{X,Y}(1,0) + \pi_{X,Y}(1,1) = \frac{1}{5} + \frac{1}{5} = \frac{2}{5}.\end{aligned}$$

Similarly, we can check that $\pi_Y(0) = \frac{3}{5}$ and $\pi_Y(1) = \frac{2}{5}$, which imples $\pi_X = \pi_Y$. Now, for any $a, b \in \{0,1\}$, we have

$$\pi_{X|Y}(a|b) = \frac{\pi(a,b)}{\pi_Y(b)} = \frac{\pi(a,b)}{\pi_X(b)} = \pi_{Y|X}(a|b).$$

Using the notation π_{ab} defined above, we obtain

$$\pi_{00} = \frac{\pi(0,0)}{\pi_X(0)} = \frac{2/5}{3/5} = \frac{2}{3} \qquad \pi_{01} = \frac{\pi(0,1)}{\pi_X(1)} = \frac{1/5}{2/5} = \frac{1}{2}$$

$$\pi_{10} = \frac{\pi(1,0)}{\pi_X(0)} = \frac{1/5}{3/5} = \frac{1}{3} \qquad \pi_{11} = \frac{\pi(1,1)}{\pi_X(1)} = \frac{1/5}{2/5} = \frac{1}{2}.$$

2. Justify that, for any $x, x', y, y' \in \{0,1\}$, we have

$$\mathbb{P}\left((X_{t+1}, Y_{t+1}) = (x', y') \mid (X_t, Y_t) = (x,y)\right) = \pi_{x'y}\,\pi_{y'x}.$$

By definition of the updates, we have

$$\begin{aligned}\mathbb{P}\left((X_{t+1}, Y_{t+1}) = (x', y') \mid (X_t, Y_t) = (x,y)\right) &= \mathbb{P}\big(X_{t+1} = x' \mid (X_t, Y_t) = (x,y)\big) \cdot \mathbb{P}\big(Y_{t+1} = y' \mid (X_t, Y_t) = (x,y)\big) \\ &= \pi_{X|Y}(x'|y) \cdot \pi_{Y|X}(y'|x) \\ &= \pi_{x'y}\,\pi_{y'x}.\end{aligned}$$

3. Show that the transition matrix P of the Markov chain (X_n) is

	$(0,0)$	$(0,1)$	$(1,0)$	$(1,1)$
$(0,0)$	$\frac{4}{9}$	$\frac{2}{9}$	$\frac{2}{9}$	$\frac{1}{9}$
$(0,1)$	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{6}$
$(1,0)$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{6}$	$\frac{1}{6}$
$(1,1)$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$

Is this chain irreducible? Is it aperiodic?

We apply the result of the previous question. We compute the first row of the transition matrix P.

Starting from the state $(0,0)$

$$P_{(0,0),(0,0)} = P\Big((X_{t+1}, Y_{t+1}) = (0,0) \,\big|\, (X_t, Y_t) = (0,0)\Big) = \pi_{00}\,\pi_{00} = \frac{2}{3} \times \frac{2}{3} = \frac{4}{9}.$$
$$P_{(0,0),(1,0)} = P\Big((X_{t+1}, Y_{t+1}) = (1,0) \,\big|\, (X_t, Y_t) = (0,0)\Big) = \pi_{10}\,\pi_{00} = \frac{1}{3} \times \frac{2}{3} = \frac{2}{9}.$$
$$P_{(0,0),(0,1)} = P\Big((X_{t+1}, Y_{t+1}) = (0,1) \,\big|\, (X_t, Y_t) = (0,0)\Big) = \pi_{00}\,\pi_{10} = \frac{2}{3} \times \frac{1}{3} = \frac{2}{9}.$$
$$P_{(0,0),(1,1)} = P\Big((X_{t+1}, Y_{t+1}) = (1,1) \,\big|\, (X_t, Y_t) = (0,0)\Big) = \pi_{10}\,\pi_{10} = \frac{1}{3} \times \frac{1}{3} = \frac{1}{9}.$$

The rest of the calculation follows the same lines and is omitted for brevity, but you are encouraged to do it for yourself.

Dince every entry of this matrix is positive, the Markov chain is aperiodic and irreducible.

4. Show that the proposed algorithm does not converge to the desired probability density.

Since the Markov chain is aperiodic and irreducible on a finite state space, the ergodic theorem ensures that $\frac{1}{n}\sum_{t=1}^{T} \mathbf{1}_{(X_t,Y_t)=x}$ converges to $\tilde{\pi}(x)$ as $T \to \infty$ for any $x \in E$, where $\tilde{\pi}$ is the stationary distribution of P. To show that the algorithm does not converge to π, it suffices to prove that π is not stationary for P, that is $\pi P \neq \pi$. Let

$$C_1 = \begin{pmatrix} 4/9 \\ 1/3 \\ 1/3 \\ 1/4 \end{pmatrix}$$

denote the first column of P. We can simply check that πC_1 is not equal to the first coordinate of π, which will yield the result. We have

$$\pi C_1 = \left(\frac{2}{5}, \frac{1}{5}, \frac{1}{5}, \frac{1}{5}\right) \begin{pmatrix} 4/9 \\ 1/3 \\ 1/3 \\ 1/4 \end{pmatrix} = \frac{47}{180} \neq \frac{2}{5}.$$

Therefore, π cannot be the stationary distribution of P, and the algorithm is flawed.

5. Show that the transition matrix P_a for the Gibbs sampler with random scanning is

	$(0,0)$	$(0,1)$	$(1,0)$	$(1,1)$
$(0,0)$	$\frac{2}{3}$	$\frac{1}{6}$	$\frac{1}{6}$	0
$(0,1)$	$\frac{1}{3}$	$\frac{5}{12}$	0	$\frac{1}{4}$
$(1,0)$	$\frac{1}{3}$	0	$\frac{5}{12}$	$\frac{1}{4}$
$(1,1)$	0	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{2}$

Verify that π is stationary for P_a. Is the chain reversible?