Simon Queric
simon.queric@telecom-paris.fr
Computational Statistics
Prof. Stéphanie Allassonnière

11/13/23
TP 1

TP1 : Reminder on Markov Chains - Stochastic gradient descent

Exercice 1 : Box-Muller and Marsaglia-Bray algorithm

1. Let $f : \mathbb{R}^2 \to \mathbb{R}$ integrable.

   We note $\phi : (r, \theta) \in \mathbb{R}_+^* \times]0, 2\pi[\mapsto (r\cos\theta, r\sin\theta) \in \mathbb{R}^2 - \{(0,x), x \geqslant 0\}$, which is a diffeomorphism, we can apply the change of variable theorem with $(x, y) = (r\cos\theta, r\sin\theta)$. On a $\det J_\phi = r$.

$$\begin{aligned}
\mathbb{E}[f(X,Y)] &= \mathbb{E}[f(\phi(R,\Theta))] \\
&= \frac{1}{2\pi}\int_{\mathbb{R}_+^* \times]0,2\pi[} f(r\cos\theta, r\sin\theta) r e^{-\frac{r^2}{2}} \mathrm{d}r\mathrm{d}\theta \\
&= \frac{1}{2\pi}\int_{\mathbb{R}^2 - \{(0,x), x\geqslant 0\}} f(x,y) e^{-(x^2+y^2)/2} \mathrm{d}x\mathrm{d}y \\
&= \frac{1}{2\pi}\int_{\mathbb{R}^2} f(x,y) e^{-(x^2+y^2)/2} \mathrm{d}x\mathrm{d}y
\end{aligned}$$

   We deduce the p.d.f of couple (X, Y) which is $f_{X,Y}(x,y) = \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}} = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} = f_X(x) f_Y(y)$. We showed that $X, Y \sim \mathcal{N}(0,1)$ and the independance of variables X and Y.

2. The previous question show that if we know how to generate a uniform variable Θ, on $[0, 2\pi]$, and a Rayleigh variable R, which are independant, then we know how to simulate two independant variables of distribution $\mathcal{N}(0,1)$.

   The Python language allows us to generate a uniform distribution on $[0,1]$, hence on $[0, 2\pi]$. Let's proof that we can simulate a Rayleigh variable by Inverse transform sampling.

   Let R be a Rayleigh variable. We have

$$F_R(x) = \mathbb{P}(R \leqslant x) = \int_0^x r e^{-\frac{-r^2}{2}} \mathrm{d}r = \left[-e^{-\frac{r^2}{2}}\right]_0^x = 1 - e^{\frac{-x^2}{2}}$$

   This c.d.f is invertible of inverse $F_R^{-1}(y) = \sqrt{-2\ln(1-y)}$.

Algorithm 1 Sampling independant gaussian variables

Sample independent random variables $U_1, U_2 \sim \mathcal{U}([0,1])$.
$\Theta \leftarrow 2\pi U_1$
$R \leftarrow \sqrt{-2\ln(1-U_2)}$
$X \leftarrow R\cos\Theta$
$Y \leftarrow R\sin\Theta$
Output : X, Y

3. (a) The while loop correspond to a rejection sampling. The target distribution is the uniform distribution on the unit disk (of p.d.f $\frac{1}{\pi}\mathbb{1}_{D(0,1)}$), and the distribution that we know how to generate is the uniforme distribution on square $[-1,1]^2$ of p.d.f $\frac{1}{4}\mathbb{1}_{[-1,1]^2}$). Hence, the random couple (V_1, V_2) follows a uniform distribution on the unit disk.

   Let's proof that rigorously. For sake of clarity, we'll denote $V_{i,1}, V_{i,2}$ the variables generated at the i-th step of the loop.
   Let $T = \inf\left\{i \in \mathbb{N}^*, V_{i,1}^2 + V_{i,2}^2 \leqslant 1\right\}$.
   We want to show that $(V_{T,1}, V_{T,2})$ follows a uniform distribution on the unit disk.
   Let $f : [-1,1]^2 \to \mathbb{R}$ a bounded and measurable function.
   We have :

$$
\begin{aligned}
\mathbb{E}[f(V_{T,1},V_{T,2})] &= \mathbb{E}\left[\sum_{k=1}^{\infty} f(V_{k,1},V_{k,2})\,\mathbb{1}_{(T=k)}\right] \\
&= \sum_{k=1}^{\infty} \mathbb{E}[f(V_{k,1},V_{k,2})\,\mathbb{1}_{(V_{k,1}^2+V_{k,2}^2\leqslant 1)} \prod_{i<k} \mathbb{1}_{(V_{i,1}^2+V_{i,2}^2>1)}] \\
&= \sum_{k=1}^{\infty} \mathbb{E}[f(V_{k,1},V_{k,2})\,\mathbb{1}_{(V_{k,1}^2+V_{k,2}^2\leqslant 1)}] \left(1-\frac{\pi}{4}\right)^{k-1} \text{ by independance} \\
&= \frac{4}{\pi}\frac{1}{4}\int_{[-1,1]^2} f(v_1,v_2)\,\mathbb{1}_{v_1^2+v_2^2\leqslant 1}\,\mathrm{d}v_1\mathrm{d}v_2 \\
&= \frac{1}{\pi}\int_{D(0,1)} f(v_1,v_2)\mathrm{d}v_1\mathrm{d}v_2
\end{aligned}
$$

(b) Let $f:[0,1]^2 \to \mathbb{R}, g:[0,1]\to\mathbb{R}$ bounded, measurable.
We show, by using the diffeomorphisme $\phi:(v,\theta)\in]0,1[\times]0,2\pi[\mapsto (\sqrt{v}\sin\theta,\sqrt{v}\cos\theta)\in D(0,1)$ of inverse $\phi^{-1}(v_1,v_2)=\left(v_1^2+v_2^2, 2\arctan\left(\frac{v_2}{\sqrt{v_1^2+v_2^2}+v_1}\right)\right)$ and such that $|\det J(\phi)(v,\theta)| = 1/2$, and by applying the change of variable theorem with $(v_1,v_2)=(\sqrt{v}\cos\theta,\sqrt{v}\sin\theta)$, that :

$$
\begin{aligned}
\mathbb{E}[f(T_1,T_2)] &= \frac{1}{\pi}\int_{D(0,1)} f\left(\frac{v_1}{\sqrt{v_1^2+v_2^2}},\frac{v_2}{\sqrt{v_1^2+v_2^2}}\right)\mathrm{d}v_1\mathrm{d}v_2 \\
&= \frac{1}{\pi}\int_0^1\int_0^{2\pi} f(\cos\theta,\sin\theta)|\det J(\phi)(v,\theta)|\mathrm{d}\theta\mathrm{d}v \\
&= \frac{1}{2\pi}\int_0^{2\pi} f(\cos\theta,\sin\theta)\mathrm{d}\theta
\end{aligned}
$$

et

$$
\begin{aligned}
\mathbb{E}[g(V)] &= \frac{1}{\pi}\int_{D(0,1)} g(v_1^2+v_2^2)\mathrm{d}v_1\mathrm{d}v_2 \\
&= \frac{1}{\pi}\int_0^1\int_0^{2\pi} g(v)|\det J(\phi)(v,\theta)|\mathrm{d}\theta\mathrm{d}v \\
&= \int_0^1 g(v)\mathrm{d}v
\end{aligned}
$$

Hence, $V\sim\mathcal{U}([0,1])$ and $(T_1,T_2)\sim(\cos\Theta,\sin\Theta)$ where $\Theta\sim\mathcal{U}([0,2\pi])$.

$$
\begin{aligned}
\mathbb{E}[f(T_1,T_2)g(V)] &= \mathbb{E}\left[f\left(\frac{V_1}{\sqrt{V_1^2+V_2^2}},\frac{V_2}{\sqrt{V_1^2+V_2^2}}\right)g\left(\sqrt{V_1^2+V_2^2}\right)\right] \\
&= \frac{1}{\pi}\int_{D(0,1)} f\left(\frac{v_1}{\sqrt{v_1^2+v_2^2}},\frac{v_2}{\sqrt{v_1^2+v_2^2}}\right) g\left(v_1^2+v_2^2\right)\mathrm{d}v_1\mathrm{d}v_2 \\
&= \frac{1}{\pi}\int_{]0,1[\times]0,2\pi[} f(\cos\theta,\sin\theta)g(v)|\det J(\phi)(v,\theta)|\mathrm{d}v\mathrm{d}\theta \\
&= \frac{1}{2\pi}\int_0^{2\pi} f(\cos\theta,\sin\theta)\mathrm{d}\theta\int_0^1 g(v)\mathrm{d}v \\
&= \mathbb{E}[f(T_1,T_2)]\mathbb{E}[g(V)]
\end{aligned}
$$

The couple (T_1,T_2) and the variable V are independant.

(c) Let's show that S follows a Rayleigh distribution.
Let $s\geqslant 0$.
$\mathbb{P}(S\leqslant s)=\mathbb{P}(\sqrt{-2\log V}\leqslant s)=\mathbb{P}(V\geqslant e^{-s^2/2})=1-e^{-s^2/2}$.
The c.d.f of S is the c.d.f of a Rayleigh distribution. According to question 1, the couple $(X,Y)=(ST_1,ST_2)$ follows a multivariate gaussian distribution $\mathcal{N}(0,I_2)$.

(d) We are looking for $\mathbb{E}[T]$ where T is the stopping time defined at question 3. (a)
The variable T follows a geometric distribution $\mathcal{G}\left(\frac{\pi}{4}\right)$, then we deduce that $\mathbb{E}[T]=\frac{4}{\pi}$.

Exercice 2 : Invariant distribution

Let $I = \left\{\frac{1}{m}, m \in \mathbb{N}^*\right\}$.

1. The definition of $\mathcal{L}(X_{n+1}|X_n)$ is equivalent to define implicitely a sequence of independant random variable $(Z_{n,m})_{n,m}$ of distribution $\mathcal{B}\left(\frac{1}{m}\right)$ such that $Z_{n,m}$ is independante of X_n and :

$$\begin{cases} X_{n+1} & = \frac{1}{m+1} \text{ si } X_n = \frac{1}{m} \text{ et } Z_{n,m} = 1 \\ X_{n+1} & \sim \mathcal{U}([0,1]) \text{ sinon} \end{cases}$$

We have

$$\mathbb{P}[X_{n+1} \in A | X_n \notin I] = \int_{A\cap[0,1]} \mathrm{d}t$$

and

$$\begin{aligned} \mathbb{P}\left[X_{n+1} \in A | X_n = \frac{1}{m}\right] &= \mathbb{P}[Z_{n,m} = 1]\mathbb{P}\left[X_{n+1} \in A | X_n = \frac{1}{m}, Z_{n,m} = 1\right] \\ &\quad + \mathbb{P}[Z_{n,m} = 0]\mathbb{P}\left[X_{n+1} \in A | X_n = \frac{1}{m}, Z_{n,m} = 0\right] \\ &= \frac{1}{m^2}\delta_{\frac{1}{m+1}}(A) + \left(1 - \frac{1}{m}\right)^2 \int_{A\cap[0,1]} \mathrm{d}t \end{aligned}$$

We deduce the transition kernel of the Markov chain.

2. Let $A \in \mathcal{B}([0,1])$.

We have :

$$P(x, A) = \mathbb{1}_{x\notin I}\, \pi(A) + \sum_{m=1}^{+\infty} \mathbb{1}_{x=1/m}\left(x^2\pi(A) + (1-x^2)\delta_{\frac{1}{m+1}(A)}\right)$$

.

Le right term is null almost everywhere with respect to the measure $\pi(\mathrm{d}x) = \mathrm{d}x$. We deduce that :

$$\int_0^1 P(x, A)\mathrm{d}x = \int_0^1 \mathbb{1}_{x\notin I}\, \pi(A)\mathrm{d}x = \pi(A)\int_0^1 \mathbb{1}_{x\notin I}\, \mathrm{d}x = \pi(A)$$

The π is invariant with respect to the transition kernel P.

3. We have $x \notin I$, hence $P(x, \cdot) = \pi(\cdot)$:

$Pf(x) = \int_0^1 f(y)P(x, \mathrm{d}y) = \int_0^1 f(y)\pi(dy)$

The function Pf si constant. We deduce that $x \notin I, P^n f(x) = \int_0^1 f(y)\pi(\mathrm{d}y)$ for all $n \geqslant 1$. Then $\lim P^n f(x) = \int_0^1 f(y)\pi(\mathrm{d}y)$.

4. (a) We have $P\left(x, \frac{1}{m+1}\right) = 1 - x^2$. We show by induction the property $P^n\left(x, \frac{1}{m+n}\right) = \prod_{k=0}^{n-1}\left(1 - \frac{1}{(m+k)^2}\right)$.

Let's show the induction step :

$$\begin{aligned} P^{n+1}\left(x, \frac{1}{m+n+1}\right) &= \int_0^1 P\left(z, \frac{1}{m+n+1}\right) P^n(x, \mathrm{d}z) \text{ according to Chapman-Kolmogorov equation} \\ &= P\left(\frac{1}{m+n}, \frac{1}{m+n+1}\right) P^n\left(\frac{1}{m}, \frac{1}{m+n}\right) \text{ car } P\left(z, \frac{1}{m+n+1}\right) = 0 \text{ if } z \neq \frac{1}{m+n} \\ &= \prod_{k=0}^{n}\left(1 - \frac{1}{(m+k)^2}\right) \text{by induction hypothesis} \end{aligned}$$

(b) It's easy to show that for $q \neq n-1$, $P^n\left(x, \frac{1}{m+1+q}\right) = 0$

Then,

$$P^n(x, A) = \sum_{q\in\mathbb{N}} P^n\left(x, \frac{1}{m+1+q}\right) = P^n\left(x, \frac{1}{m+n}\right) = \prod_{k=0}^{n-1}\left(1 - \frac{1}{(m+k)^2}\right)$$

We obtain :

$$P^n\left(x, \frac{1}{m+n}\right) = \prod_{k=0}^{n-1}\frac{m+k-1}{m+k}\prod_{k=0}^{n-1}\frac{m+k+1}{m+k} = \frac{m-1}{m+n-1}\frac{m+n}{m}$$

Then $\lim P^n(x, A) = \dfrac{m-1}{m} \neq 0 = \pi(A)$.

Exercice 3 : Stochastic Gradient Learning in Neural Networks

1. We can notice that

$$R_n(w) = \frac{1}{n}\sum_{i=1}^{n}(y_i - w^T x_i)^2 = \mathbb{E}_{\mu_n}\left[(Y - w^T X)^2\right]$$

where $\mu_n = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}\delta_{(x_i,y_i)}$ is the empirical measure of (X, Y).

Algorithm 2 Stochastic Gradient Descent for minimizing empirical risk

---

Input : a sample $(x_i, y_i)_{1\leqslant i\leqslant n}$ of size n
$w \leftarrow w_0$, with random w_0 e.g $w_0 \sim \mathcal{N}(0, \sigma^2 I_2)$
$k \leftarrow 1$
while not stopping criterion **do**
    Sample $(x^k, y^k) \sim \mu_n$
    $w \leftarrow w - \gamma_k \nabla_w F(x^k, y^k, w)$
    $k \leftarrow k+1$
end while
Output : w

---

où $F(x, y, w) = (y - w^T x)^2$ d'où $\nabla_w F(x, y, w) = -2(y - w^T x)x$, et $\gamma_k = \gamma/\sqrt{k}$.