Student Name: Essakine Amer
Student ID: amer.essakine@ens-paris-saclay.fr

Computational statistics
TP 1

Exercise 1

1. Let ψ be a mesurable function, and lets prove that X and Y are independant.

$$\mathbb{E}[\phi(X,Y)] = \frac{1}{2\pi}\int_0^{\infty}\int_0^{2\pi}\phi(rcos(\theta), rsin(\theta))r\exp^{-r^2}\,dr\,d\theta$$

   We perform the change of variables $(x,y) = (rcos(\theta), rsin(\theta))$, the function describing this change is C^1 Diffeomorphism, and the Jacobian is given by :

$$det(J)(r,\theta) = \det\begin{bmatrix} cos(\theta) & -rsin(\theta) \\ sin(\theta) & rcos(\theta)\end{bmatrix} = r(cos(\theta)^2 + sin(\theta)^2) = r$$

   Hence we have :

$$\mathbb{E}[\phi(X,Y)] = \frac{1}{2\pi}\int_{\mathbb{R}^2}\phi(x,y)\exp(-x^2-y^2)\,dx\,dy$$

   Hence, the density function of (X,Y) is given by :

$$f_{(X,Y)}(x,y) = \frac{1}{2\pi}\exp(-x^2-y^2) = f_X(x)f_Y(y)$$

   Hence X, Y are independant and their density functions are given by :

$$f_X(x) = \frac{1}{\sqrt{2\pi}}\exp(-x^2) \quad \text{and} \quad f_Y(y) = \frac{1}{\sqrt{2\pi}}\exp(-y^2)$$

   Which means that both X and Y have $\mathcal{N}(0,1)$ distribution and are independant.

2. From the previous exercise, we know that we can sample an independent Gaussian distribution given that we can sample a Rayleigh distribution and a uniform distribution. The CDF of the Rayleigh distribution is given by :

$$F_R(r) = \mathbb{P}(R \le r) = \int_0^r s\exp(-s^2/2)\,ds = 1-\exp(-\frac{x^2}{2})$$

   Let $u \in]0,1[$:

$$F_R(r) = u \Leftrightarrow r = \sqrt{-2\ln(1-u)}$$

   Hence sampling a Rayleigh distribution is the same as sampling $\sqrt{-2\ln(1-U)}$, where U follows a uniform distribution on $[0,1]$, which is the same as sampling $\sqrt{-2\ln(U)}$. Hence we have an algorithm to sample (X,Y)

Algorithm 1 Sample an independant Gaussian distibuition

Sample Θ with distribution $\mathcal{U}(0, 2\pi)$
Sample U with distribution $\mathcal{U}(0, 1)$
Set R $= \sqrt{-2\ln(U)}$
Set $X = Rcos(\theta)$ and $Y = Rsin(\theta)$
return (X,Y)

3. (a) V_1 and V_2 both have $\mathcal{U}([-1, 1])$ distribution, and the loop continues until (V_1, V_2) lies in the unit disk. Hence, at the end of the while loop (V_1, V_2) follows a uniform distribution over the unit disk.

(b) Let N denote the number of steps in the 'while loop', and let

$$p = \mathbb{P}(V_1^2 + V_2^2 > 1)$$

, then for j $\in \mathbb{N}$:

$$\mathbb{P}(N = j) = p(1 - p)^{j-1}$$

Hence, N follows a geometric law of parameter p and its expected value is :

$$\mathbb{N} = \frac{1}{p}$$

Moreover, $1 - p$ is the probability that (V_1, V_2) lies in the unit disk, which is the ratio of the surface of the square $[-1, 1]^2$ and unit circle. Hence :

$$p = \frac{4 - \pi}{4}$$

Finally, the expected number of steps is $\frac{4}{4-\pi} \approx 4.65$

(c) Let h be a measurable function,

$$\mathbb{E}(h(T_1, V)) = \int_{(v_1, v_2) \in \mathbb{S}^1} h(\frac{v_1}{\sqrt{v_1^2 + v_2^2}}, v_1^2 + v_2^2)\frac{1}{\pi}\, dv_1\, dv_2$$

We perform a variable change to go into polar coordinates :—

$$\mathbb{E}(h(T_1, V)) = \int_0^1 \int_0^{2\pi} h(\cos(\theta), r^2)\frac{1}{\pi} r\, dr\, d\theta = \int_0^1 \int_0^{2\pi} h(\cos(\theta), r)\frac{1}{2\pi}\, dr\, d\theta = \mathbb{E}(h(\cos(\Theta)), R))$$

Where $\Theta \sim \mathcal{U}([0, 2\pi])$ and $R \sim \mathcal{U}([0, 1])$, and R and Θ are independent.
This implies that T_1 has the same distribution as $\cos(\Theta)$ when we fix h as a constant in the second parameter. Furthermore, V has the same distribution as R, which follows a uniform distribution on the interval $[0, 1]$, again with h held constant in the first parameter. Since also the pairs (T_1, V) and $(\cos(\Theta), R)$ share the same joint distribution, we conclude that T_1 and V are independent random variables.

(d) Since $S = \sqrt{-2log(V)}$, we can show that S follows a Rayleigh distribution. In fact for $s \in \mathbb{R}$:

$$P(S \leq s) = P(V \geq \exp(-\frac{s^2}{2}) = rexp(-\frac{s^2}{2})\mathbb{I}_{\mathbb{R}^+}(r)$$

Since $T_1 = \frac{V_1}{\sqrt{V_1^2+V_2^2}}$ has the same distribution as $\cos(\Theta)$, by similar reasoning we find that $\frac{V_2}{\sqrt{V_1^2+V_2^2}}$ has the same distribution as $\sin(\Theta)$. This implies that X and Y have the same distributions as $S\cos(\Theta)$ and $S\sin(\Theta)$, respectively.

Furthermore, since S and Θ are independent—given that S can be expressed as a function of V and Θ as a function of T_1, and we established in the previous question that V and T_1 are independent—we conclude that X and Y indeed have the same distribution as $S\cos(\Theta)$ and $S\sin(\Theta)$. By the first question we conclude that (X, Y) follows $\mathcal{N}(0, I_2)$

Exercise 2

1. Let $n \in \mathbb{N}$ and $A \subset [0,1]$,

$$\mathbb{P}(X_{n+1} \in A | X_n \notin \{\frac{1}{k}, k \in \mathbb{N}^*\}) = \mathbb{P}_{X_{n+1}\sim\mathcal{U}([0,1])}(X_{n+1} \in A) = \int_{A\cap[0,1]} dt$$

   And we have :

$$\mathbb{P}(X_{n+1} \in A | X_n = \frac{1}{k}) = \mathbb{P}(X_{n+1} \in A | X_{n+1} = \frac{1}{k+1}, X_n = \frac{1}{k}) \cdot (1 - (\frac{1}{k})^2) + \mathbb{P}(X_{n+1} \in A | X_{n+1} \neq \frac{1}{k+1}, X_n = \frac{1}{k}) \cdot (\frac{1}{k})^2$$

   Hence :

$$\mathbb{P}(X_{n+1} \in A | X_n = \frac{1}{k}) = (1 - (\frac{1}{k})^2)\delta_{\frac{1}{k+1}}(A) + (\frac{1}{k})^2 \int_{A\cap[0,1]} dt$$

   From this, we conclude that the transition kernel is exactly given by the expression in the kernel.

2. Let $A \subset [0,1]$,

$$\int \pi(\,dx)P(x, A) = \int_0^1 P(x, A)\,dx = \int_{x\in[0,1]\setminus\{\frac{1}{k}, k\in\mathbb{N}^*\}} P(x, A)\,dx$$

   This is true since the set $\{\frac{1}{k}, k \in \mathbb{N}^*\}$ is a countable set, so its measure is 0 by π. Hence :

$$\int \pi(\,dx)P(x, A) = \int_{x\in[0,1]\setminus\{\frac{1}{k}, k\in\mathbb{N}^*\}} \int_{A\cap[0,1]} dt\,dx = \pi(A)$$

   This proves that the uniform distribution π is invariant for P

3. Let $x \notin \{\frac{1}{k}, k \in \mathbb{N}^*\}$, and f a bounded measurable function

$$Pf(x) = \mathbb{E}(f(X_1)|X_0 = x) = \int_0^1 f(t)\pi(t)\,dt$$

   Hence $Pf(x)$ does not depend on x. In general, if a is a constant, and $g(x) = a$ for all x :

$$Pg(x) = \int_0^1 a\pi(t)\,dt = a$$

We thus conclude that for all $n \geq 1$,

$$P^n f(x) = \int_0^1 f(t)\pi(t)\,dt$$

Since other than the first step, we keep applying P to a constant which gives the same constant. We conclude then that :

$$\lim_{n\to\infty} P^n f(x) = \int_0^1 f(t)\pi(t)\,dt$$

4. (a) Let $n \in \mathbb{N}^*$, and $k \in \mathbb{N}$ with $k \geq 2$

$$\begin{aligned} P^{n+1}(x, \frac{1}{n+k+1}) &= \int P^n(x,\, dy)P(y, \frac{1}{n+k+1}) \\ &= P^n(x, \frac{1}{n+k})P(\frac{1}{n+k}, \frac{1}{n+k+1}) \end{aligned}$$

Since $P(y, \frac{1}{n+k+1}) = 0$ unless $y = \frac{1}{n+k}$
Lets prove by recursion that

$$P^n(x, \frac{1}{n+k}) = \prod_{j=0}^{n-1}(1 - \frac{1}{(j+k)^2})$$

For $n = 1$: By the first question :

$$P(x, \frac{1}{k+1}) = 1 - \frac{1}{k^2}$$

Let suppose the formula is valid for $n \in \mathbb{N}$, then :

$$\begin{aligned} P^{n+1}(x, \frac{1}{n+k+1}) &= P^n(x, \frac{1}{n+k})P(\frac{1}{n+k}, \frac{1}{n+k+1}) \\ &= \prod_{j=0}^{n-1}(1 - \frac{1}{(j+k)^2})(1 - \frac{1}{(n+k)^2}) \\ &= \prod_{j=0}^{n}(1 - \frac{1}{(j+k)^2}) \end{aligned}$$

Hence the result

(b) Let $A = \cup_{q\in\mathbb{N}}\{\frac{1}{k+q+1}\}$ and $n \in \mathbb{N}$,

$$\begin{aligned}
P^n(x, A) &= \sum_{q\in\mathbb{N}} P^n(x, \frac{1}{k+q+1}) \\
&= P^n(x, \frac{1}{n+k}) \\
&= \prod_{j=0}^{n-1}(1 - \frac{1}{(j+k)^2}) \\
&= \prod_{j=0}^{n-1} \frac{j+k-1}{j+k}\frac{j+k+1}{j+k} \\
&= \frac{k-1}{n-1+k}\frac{n+k}{k} \\
&\xrightarrow{n\to\infty} \frac{k-1}{k}
\end{aligned}$$

Hence $\lim_{n\to\infty} P^n(x, A) \neq \pi(A)$ since $k \geq 2$