Dorian Gaillard

TP 1 :

Exercice 1)

1) Let H be a measurable function.

$$\mathbb{E}(H(X)) = \int_0^{+\infty}\int_0^{2\pi} H(r\cos\theta)\,\frac{re^{-\frac{r^2}{2}}}{2\pi}\,dr\,d\theta$$

By changing variables to polar coordinates, (ie $\begin{cases} x = r\cos\theta \\ y = r\sin\theta \end{cases}$ and $J = \begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix}$

so $\det J = r$)

$$\mathbb{E}(H(X)) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} H(x)\,\frac{e^{-\frac{x^2+y^2}{2}}}{2\pi}\,dx\,dy$$

$$= \frac{1}{2\pi}\int_{-\infty}^{+\infty} H(x)\,e^{-\frac{x^2}{2}}\,dx \underbrace{\int_{-\infty}^{+\infty} e^{-\frac{y^2}{2}}\,dy}_{=\sqrt{2\pi}}$$

$$= \int_{-\infty}^{+\infty} H(x)\,\frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}\,dx$$

so $X \sim \mathcal{N}(0;1)$.

By doing the same for Y, $Y \sim \mathcal{N}(0;1)$.

Then $$\mathbb{E}(H(X,Y)) = \int_0^{+\infty}\int_0^{2\pi} H(r\cos\theta, r\sin\theta)\,\frac{re^{-\frac{r^2}{2}}}{2\pi}\,dr\,d\theta$$

$$= \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} H(x,y)\,\frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}\,\frac{e^{-\frac{y^2}{2}}}{\sqrt{2\pi}}\,dx\,dy$$

so the joint density is indeed the product of the densities of X and Y ie $X \perp\!\!\!\perp Y$.

2) We need to have R such that R follows a Rayleigh distribution of parameter 1, which is done by taking $R=\sqrt{-2\ln(V)}$ where $V \sim \mathcal{U}([0;1])$. Indeed, $R\leq t \Leftrightarrow V \geq e^{-\frac{t^2}{2}}$ and $\frac{dP(V\geq e^{-\frac{t^2}{2}})}{dt} = t e^{-\frac{t^2}{2}}$. Then $X = R\cos(\Theta)$ and $Y = R\sin(\Theta)$ are two independant gaussian distributions.

3) a) At the end of the loop, the distribution of (V_1, V_2) is the distribution of (V_1, V_2) initially knowing $V_1^2 + V_2^2 \leq 1$.

So by denoting $f(v_1, v_2) = \frac{1}{4}\mathbb{1}_{(v_1,v_2)\in[0;1]^2}$ the initial density, at the end of the loop, the density becomes:

$$f_{endloop}(v_1, v_2) = \frac{f(v_1,v_2)\,\mathbb{1}_{v_1^2+v_2^2\leq 1}}{P(V_1^2+V_2^2\leq 1)}$$

$$= \frac{\mathbb{1}_{v_1^2+v_2^2\leq 1}}{\pi}$$

ie $(V_1, V_2) \sim \mathcal{U}(\mathcal{B}(0;1))$.

b) ~~$\ast$~~ $V \sim \mathcal{U}([0;1])$:

Let $t \in [0;1]$, [0;1],

$$P(V \leq t) = P(R \leq \sqrt{t}) \quad \text{where } R = \sqrt{V_1^2+V_2^2}$$

$$= \iint_{\mathcal{B}(0;\sqrt{t})} \frac{dv_1\,dv_2}{\pi}$$

(by changing variables to polar coordinates)
$$= \int_0^{\sqrt{t}}\int_0^{2\pi} \frac{r}{\pi}\,dr\,d\theta$$

$$= t$$

If $t \in \mathbb{R}^-$, $P(V\leq t) = 0$, if $t \geq 1$, $P(V \leq t) = 1$.

So the density is $\frac{dP(V\leq t)}{dt} = \mathbb{1}_{[0;1]}(t)$ ie $V \sim \mathcal{U}([0;1])$

$*$ (T_1, T_2) has the same distribution as $(\cos\Theta, \sin\Theta)$ with $\Theta \sim \mathcal{U}([0,2\pi])$:

$$\mathbb{E}(H(T_1, T_2)) = \iint_{B(0,1)} H(t_1, t_2) \frac{d\sigma_1 d\sigma_2}{\pi}$$

in polar coordinates $= \int_0^1 \int_0^{2\pi} H(\cos(\theta), \sin(\theta)) \frac{r}{\pi} dr\, d\theta$

$$= \int_0^{2\pi} H(\cos(\theta), \sin(\theta)) \frac{d\theta}{2\pi}$$

so it is the case.

$*$ $(T_1, T_2) \perp\!\!\!\perp V$:

$$\mathbb{E}(H(T_1, T_2, V)) = \iint_{B(0,1)} H(t_1, t_2, \sigma) \frac{d\sigma_1 d\sigma_2}{\pi}$$

$$= \int_0^1 \int_0^{2\pi} H(\cos\theta, \sin\theta, r^2) \frac{r\, dr\, d\theta}{\pi}$$

$$= \int_0^{2\pi}\int_0^1 H(\cos\theta, \sin\theta, \sigma) \frac{d\theta}{2\pi} d\sigma$$

by changing variable $\sigma = r^2$, ie $d\sigma = 2\, dr$

So we do have that the joint density is the product of the densities.

c) We have $X = S T_1$ and $Y = S T_2$.
Let's compute the distribution of S:
Let $t \in \mathbb{R}^+$,

$$S \leq t \iff V \geq e^{-\frac{t^2}{2}}$$

So $P(S \leq t) = \int e^{-\frac{t^2}{2}} \, ds$

$$= 1 - e^{-\frac{t^2}{2}}$$

and $\frac{dP(S \leq t)}{dt} = t e^{-\frac{t^2}{2}}$

ie S follow a Rayleigh distribution with parameter 1.

Then we have $S \perp\!\!\!\perp (T_1, T_2)$ because $V \perp\!\!\!\perp (T_1, T_2)$ and by question 1 we have that X and Y are two independant Gaussian distribution $\mathcal{N}(0,1)$.

d) The loop can be modelled by a geometric law with parameter $p = P(V_1^2 + V_2^2 \leq 1) = \pi$.
The expected number of steps is then $\frac{1}{p} = \frac{1}{\pi}$ (expectation of a geometric law).

Exercice 2)

1) If $x_n = \frac{1}{n}$, let's write $Y_n \sim \mathcal{B}(1 - X_n^2)$, such that $X_{n+1} = \frac{1}{n+1}$ when $Y = 1$ and $X_{n+1} \sim \mathcal{U}([0;1])$ when $Y = 0$.

Then let $A \subset [0;1]$,
$P(x, A) = P(X_{n+1} \in A \mid X_n = x)$

If $x = \frac{1}{n}$, then

$$P(X_{n+1} \in A) = P(X_{n+1} \in A \mid Y=0) P(Y=0)$$

$$+ P(X_{n+1} \in A \mid Y=1) P(Y=1)$$

$$= P(X_{n+1} \in A \mid X_{n+1} \sim \mathcal{U}([0;1])) P(Y=0)$$

$$+ P\left(\frac{1}{n+1} \in A\right) P(Y=1)$$

$$= x^2 \int_{A \cap [0;1]} dt + (1-x^2) \delta_{\frac{1}{n+1}}(A)$$

If $x \neq \frac{1}{n}$, $X_{n+1} \sim \mathcal{U}([0;1])$ so

$$P(X_{n+1} \in A) = \int_{A \cap [0;1]} dt$$

So indeed,

$$P(x,A) = \begin{cases} x^2 \int_{A \cap [0;1]} dt + (1-x^2) \delta_{\frac{1}{n+1}}(A) & \text{if } x = \frac{1}{n} \\ \int_{A \cap [0;1]} dt & \text{else} \end{cases}$$

2) We want to prove that $\int_{[0;1]} \pi(dx) P(x,A) = \pi(A)$.
Let $S = \left\{ \frac{1}{n}, n \in \mathbb{N}^* \right\}$.

Then $\int_{[0;1]} \pi(dx) P(x,A) = \int_{[0;1] \setminus S} \pi(dx) P(x,A)$

$$+ \underbrace{\int_S \pi(dx) P(x,A)}_{=0 \text{ because } S \text{ is countable}}$$

$$= \int_{[0;1] \setminus S} P(x,A)\, dx$$

$$= \int_{[0;1]} \int_{[0;1] \cap A} dt$$

$$= \pi(A)$$

3) * $Pf(x) = \mathbb{E}(f(x_1) | X_0 = x)$

$$= \int_{[0;1]} f(x) \, P(x, dx)$$

$$= \int_{[0;1]} f(x) \, dx$$

* Let's show by recurrence that for $x \neq \frac{1}{n}$, $P^n(x, A) = \pi(A)$.

- For $n = 2$, $P^2(x, A) = \int_{[0;1]} P(x, dz) P(z, A)$

$$= \int_{[0;1]} P(z, A) \, dz \text{ because } x \neq \frac{1}{n}$$

but S is countable so its integral is 0 and

$$P^2(x, A) = \int_{[0;1]\setminus S} P(z, A) \, dz$$

$$= \int_{[0;1]} \pi(A) \, dz$$

$$= \pi(A)$$

- For any n: let's suppose that this is true for a certain $n \in \mathbb{N}$, then

$$P^{n+1}(x, A) = \int_{[0;1]} P(x, dz) \, P^n(z, A)$$

$$= \int_{[0;1]\setminus S} P^n(z, A) \, dz$$

$$= \int_{[0;1]} \pi(A) \, dz$$

$$= \pi(A)$$

Finally, $P^n f(x) = \int_{[0;1]} P(x, dz)\, f(z)$

$= \int_{[0;1]\setminus S} \pi(z)\, f(z)\, dz$

So $\lim_{n \to +\infty} P^n f(x) = \int_{[0;1]} f(x) \pi(x)\, dx$

4) a) Let's show by recurrence that

$\forall n \in \mathbb{N}^*,\quad P^n\left(x, \frac{1}{n+m}\right) = \prod_{i=0}^{n-1} \left(1 - \frac{1}{(m+i)^2}\right)$

For $n=1$, $P\left(x, \frac{1}{m+1}\right) = x^2 \int_{\{\frac{1}{m+1}\}} dt + (1-x^2)\delta_{\frac{1}{m+1}}\left(\frac{1}{m+1}\right)$

$= (1-x^2)$

$= \left(1 - \frac{1}{m^2}\right)$

For any n ; let's suppose this is true for a certain $n \in \mathbb{N}^*$,

$P^{n+1}\left(x, \frac{1}{m+n+1}\right) = \int_{[0;1]} P(x, dz)\; P^n\left(z, \frac{1}{m+n+1}\right)$

$= \int_{[0;1]} \left(x^2 dz + (1-x^2)\delta_{z, \frac{1}{m+1}}\right) P^n\left(z, \frac{1}{m+n+1}\right)$

$= \int_{[0;1]} (1-x^2)\, \delta_{\frac{1}{m+1}}(z)\, P^n\left(z, \frac{1}{m+n+1}\right)$

$0 = \{ \quad + \int_{[0;1]} x^2 P^n\left(z, \frac{1}{m+n+1}\right) dz$

because $P^n\left(z, \frac{1}{m+n+1}\right) = 0$ on $[0;1]\setminus S$ and S is countable

$= (1-x^2)\, P^n\left(\frac{1}{m+1}, \frac{1}{m+n+1}\right)$

$= (1-x^2) \prod_{i=0}^{n-1} \left(1 - \frac{1}{(m+1+i)^2}\right)$

$= \prod_{i=0}^{n} \left(1 - \frac{1}{(m+i)^2}\right)$

b) $\forall n \in \mathbb{N}^*$,

$$P^n(x, A) \geq P^n\left(x, \frac{1}{m+n}\right)$$
$$= \prod_{i=0}^{n-1} \left(1 - \frac{1}{(m+i)^2}\right)$$

which doesn't go to 0 when $n \to +\infty$, indeed

$$\ln\left(P^n(x, A)\right) \geq \sum_{i=0}^{n-1} \ln\left(1 - \frac{1}{(m+i)^2}\right)$$

but $\ln\left(1 - \frac{1}{(m+i)^2}\right) \underset{i \to +\infty}{\sim} - \frac{1}{(m+i)^2}$

and $-\sum_{i \geq 0} \frac{1}{(m+i)^2}$ is convergent,

so the sum is convergent but if the product went to 0 then the logarithm would diverge to $-\infty$, so the product does not go to 0.

Thus we do not have

$\lim_{n \to +\infty} P^n(x, A) = \pi(A) = 0$ because A is countable.

Exercise 3)

1) We want to find the parameter w minimizing the empirical risk. The stochastic gradient descent step is:

Take one sample x_{k+1} randomly across the n samples.

$$w_{k+1} = w_k + 2\,{}^t x_{k+1}\left(y_{k+1} - {}^t w_k x_i\right) \times t_k$$

where y_{k+1} is label of x_{k+1}

and t_k is the step length

i.e. $w_{k+1} = w_k - t_k \nabla_w j(w_k, x_{k+1})$

with $j'(w_k, x_{k+1}) = (y_{k+1} - {}^t w_k x_{k+1})$

I took $t_k = \frac{1}{k}$ such that $\sum t_k = +\infty$ and $\sum t_k^2 < +\infty$.

In the course a number of hypotheses are given to ensure convergence:

* the RV $j(w, X)$ is measurable and have an expectation
* $w \mapsto j(w, x)$ is convex and continuous
* the gradient need to be bounded so in the implementation I chose to normalize it
* $t_k = \frac{1}{k}$ checks $\sum t_k = +\infty$ and $\sum t_k^2 < +\infty$