1

Emilio PICARD

Computational Statistics
TP-1

Exercice 1:

1) R and Θ are independant, so the joint ~~probability~~ distribution $P_{(R,\Theta)}$ is

$$f_{(R,\Theta)}(r,\theta) = \frac{r}{2\pi}\exp\left(-\frac{r^2}{2}\right)\mathbb{1}_{\mathbb{R}_+}(r)\mathbb{1}_{[0,2\pi]}(\theta).$$

Let consider $h: \mathbb{R}^2 \to \mathbb{R}^2$ bounded and mesurable.

Let $X = R\cos(\Theta)$, $Y = R\sin(\Theta)$.

$$E[h(X,Y)] = E[h(R\cos(\Theta), R\sin(\Theta))]$$

$$= \int_{\mathbb{R}}\int_{\mathbb{R}} h(r\cos\theta, r\sin\theta) f_{r,\theta}(r,\theta)\, dr\, d\theta$$

$$= \int_{\mathbb{R}}\int_{\mathbb{R}} h(r\cos\theta, r\sin\theta) \frac{\mathbb{1}_{[0,2\pi]}(\theta)}{2\pi} r e^{-\frac{r^2}{2}} \mathbb{1}_{\mathbb{R}_+}(r)\, dr\, d\theta$$

we define $g(r,\theta) = (r\cos\theta, r\sin\theta)$, $\forall (r,\theta) \in \{\mathbb{R}_+ \times [0,2\pi]\}$.

we will ~~do~~ apply a change of variables.

g is a C^1-diffeomorphism, we can calculate the Jacobian matrix J_g:

$$J_g = \begin{pmatrix} \cos\theta & \sin\theta \\ -r\sin\theta & r\cos\theta \end{pmatrix} \Rightarrow \det(J_g) = r\cos^2\theta + r\sin^2\theta = r$$

$$g(r,\theta) = (x,y) \iff \begin{cases} r = \sqrt{x^2+y^2} \\ \theta = \arctan\left(\frac{y}{x}\right) \end{cases}$$

Thus,

$$E[h(X,Y)] = \int_{\mathbb{R}}\int_{\mathbb{R}} h(x,y) \frac{\sqrt{x^2+y^2}}{2\pi} e^{-\frac{x^2+y^2}{2}} \times \frac{1}{\sqrt{x^2+y^2}}\, dx\, dy$$

$$= \int_{\mathbb{R}}\int_{\mathbb{R}} h(x,y) \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}\, dx\, dy$$

Thus, $X \perp\!\!\!\perp Y$ because $f_{(X,Y)}(x,y) = f_X(x) f_Y(y)$,

where f_X ~~(x)~~ and f_Y are Gaussian density functions, with $X \sim \mathcal{N}(0,1)$, $Y \sim \mathcal{N}(0,1)$.

Both X and Y have $\mathcal{N}(0,1)$ distribution and are independant.

□

2) Let $U_1 \sim \mathcal{U}([0,1])$, $U_2 \sim \mathcal{U}([0,1])$

- to simulate Θ; we can simulate U_2, and multiply by 2π;

  if $U_2 \sim \mathcal{U}[0,1]$, $2\pi U_2 \sim \mathcal{U}[0,2\pi]$. Thus, we will simulate Θ with $2\pi U_2$.

- to simulate R:

let first compute the cdf: let $x \in \mathbb{R}_+$.

$$P(R<x) = \int_{-\infty}^{x} r e^{-\frac{r^2}{2}} \mathbb{1}_{\mathbb{R}_+}(r)\, dr$$

$$= \left[-e^{-\frac{r^2}{2}}\right]_0^x = 1 - e^{-\frac{x^2}{2}}.$$

$P(R<x) = u$

$$\iff 1 - e^{-\frac{x^2}{2}} = u \iff -\frac{x^2}{2} = \ln(1-u)$$

$$\iff x = \sqrt{-2\ln(1-u)}$$

To simulate R, we simulate U_1, and so we can have

$$R_{sim} = \sqrt{-2\ln(1-U_1)} = \sqrt{-2\ln(U_1)}$$

Thus, let's right an algorithm:

- simulate $U_1 \sim \mathcal{U}[0,1]$, $U_2 \sim \mathcal{U}[0,1]$.
- simulate $\begin{cases} R = \sqrt{-2\ln U_1} \\ \Theta = 2\pi U_2 \end{cases}$
- simulate $\begin{cases} X = R\cos\Theta = \cos(2\pi U_2)\sqrt{-2\ln U_1} \\ Y = R\sin\Theta = \sin(2\pi U_2)\sqrt{-2\ln U_1} \end{cases}$

Thus, X and Y are two ~~simulations~~ samplings of independant gaussians distribution $\mathcal{N}(0,1)$. $\square$

3) a) $V_1 \sim \mathcal{U}[-1,1]$, $V_2 \sim \mathcal{U}[-1,1]$ (before the while loop).

The while loop is used to simulate V_1 and V_2 in the unit disk, uniformly.

Hence, after the while loop, the couple (V_1, V_2) has uniform distribution on the unit disk. $\square$

b) We are looking at the parameters of a geometric probability distribution.

The ~~var~~ random variable $\mathbb{1}_{\{V_1^2+V_2^2<1\}} \sim Geom(p)$, where p is what we are looking for (the first success).

during the while loop, V_1 and V_2 are uniform random variable between -1 and 1.

$$\mathbb{E}\left[\mathbb{1}_{V_1^2+V_2^2\leq 1}\right]$$

$$= \int_{\mathbb{R}}\int_{\mathbb{R}} \mathbb{1}_{\{u^2+v^2\leq 1\}} \frac{1}{2}\mathbb{1}_{[-1,1]}(u)\frac{1}{2}\mathbb{1}_{[-1,1]}(v)\,du\,dv$$

$$= \frac{1}{4}\int_{-1}^{1}\int_{-1}^{1} \mathbb{1}_{\{|u|\leq\sqrt{1-v^2}\}}\,du\,dv$$

$$= \frac{1}{4}\int_{-1}^{1}\int_{-\sqrt{1-v^2}}^{\sqrt{1-v^2}} 1\,du\,dv$$

$$= \frac{1}{4}\int_{-1}^{1}\left(\sqrt{1-v^2}-(-\sqrt{1-v^2})\right)dv \qquad \begin{cases} v = \cos x \\ dv = -\sin x\,dx \\ -1\leq v\leq 1 \Leftrightarrow -\pi\leq x\leq 0 \end{cases}$$

$$= \frac{1}{4}\int_{-\pi}^{0} 2\sqrt{1-\cos^2(x)}\,(-\sin(x))\,dx$$

$$= \frac{1}{2}\int_{-\pi}^{0}\sin^2(x)\,dx$$

or, $\sin^2(x) = \frac{1-\cos(2x)}{2}$

Thus:

$$\mathbb{E}\left[\mathbb{1}_{V_1^2+V_2^2\leq 1}\right] = \frac{1}{2}\int_{-\pi}^{0}\frac{1-\cos(2x)}{2}\,dx$$

$$= \frac{1}{4}\left(\int_{-\pi}^{0}1\,dx - \int_{-\pi}^{0}\cos(2x)\,dx\right)$$

$$= \frac{1}{4}\left(\pi - \underbrace{\left[\frac{1}{2}\sin(2x)\right]_{-\pi}^{0}}_{=0}\right) = \frac{\pi}{4}$$

Thus $\mathbb{E}\left[\mathbb{1}_{V_1^2+V_2^2\leq 1}\right] = \frac{\pi}{4} = \frac{1}{p} \Leftrightarrow \boxed{p = \frac{4}{\pi}}$

The expected number of steps in the "while" loop is 2! $\square$

(c) $T_1 = \frac{V_1}{\sqrt{V_1^2+V_2^2}}$, $V = V_1^2 + V_2^2$.

Let's use the same idea of change of variable as in Q1.

Let h measurable and bounded function in $\mathbb{R}^2$.

$$\mathbb{E}[h(T_1, V)] = \mathbb{E}\left[h\left(\frac{V_1}{\sqrt{V_1^2+V_2^2}}, V_1^2+V_2^2\right)\right]$$

$$= \int_{\mathbb{R}^2} h\left(\frac{u}{\sqrt{u^2+v^2}}, u^2+v^2\right)\frac{1}{\pi}\mathbb{1}_{\{u^2+v^2\leq 1\}}\,du\,dv$$

(because $(V_1,V_2)\sim U(D(0,1))$ so the joint density is $\frac{1}{\pi}\mathbb{1}_{\{u^2+v^2\leq 1\}}$)

change of variable: $\begin{cases} u = r\cos\theta \\ v = r\sin\theta \end{cases} \Leftrightarrow \begin{cases} r = \sqrt{u^2+v^2} \\ \theta = \arctan\left(\frac{v}{u}\right)\end{cases}$

$du\,dv = r\,dr\,d\theta$ and $u^2+v^2\leq 1 \Leftrightarrow 0\leq r\leq 1$ and $0\leq\theta\leq 2\pi$.

finally, $\begin{cases} \frac{u}{\sqrt{u^2+v^2}} = \frac{r\cos\theta}{r} = \cos\theta \\ u^2+v^2 = r^2 \end{cases}$

we get:

$$\mathbb{E}[h(T_1,V)] = \int_{\mathbb{R}^2} h(\cos\theta, r^2)\frac{r}{\pi}\mathbb{1}_{[0,1]}(r)\mathbb{1}_{[0,2\pi]}(\theta)\,dr\,d\theta$$

Then, we apply $x = r^2$ ~~$\Leftrightarrow dx = 2r\,dr$~~ $\Leftrightarrow \sqrt{x} = r$

$dr = \frac{1}{2r}dx$

$$\Rightarrow \mathbb{E}[h(T_1,V)] = \int_{\mathbb{R}^2} h(\cos\theta, x)\frac{r}{\pi\, 2r}\mathbb{1}_{[0,1]}(x)\mathbb{1}_{[0,2\pi]}(\theta)\,dx\,d\theta$$

$$= \int_{\mathbb{R}^2} h(\cos\theta, x)\,\mathbb{1}_{[0,1]}(x)\frac{1}{2\pi}\mathbb{1}_{[0,2\pi]}(\theta)\,dx\,d\theta.$$

We recognize the joint density function of X, Y independant where $X\sim U([0,1])$ and $Y\sim U([0,2\pi])$

This ~~proves~~ shows that $T_1 \perp\!\!\!\perp V$, and

$\begin{cases} T_1 \equiv \cos(\Theta),\ \Theta\sim U([0,2\pi]) \\ V \sim U([0,1]) \end{cases}$ $\square$

(d) $X = S\frac{V_1}{\sqrt{V_1^2+V_2^2}} = ST_1 = S\cos(\Theta)$

with $S = \sqrt{-2\log(V_1^2+V_2^2)}$, according Q2, $S\sim \text{Rayleigh}(1)$

by symetry, $Y = S\frac{V_2}{\sqrt{V_1^2+V_2^2}} = S\sin(\Theta)$

Thus, with Q1, we can conclude that $(X,Y)\sim\mathcal{N}(0,1)^2$. $\square$

Exercice 2:

(1) By definition, $P(x,A) = \mathbb{P}(X_{n+1}\in A \mid X_n = x)$, $\forall n\in\mathbb{N}$.

let $x\in[0,1]$. According the exercise, (X_n) can take 2 values.

if $X_n = \frac{1}{k}$ $(k>0)$: set $B = \{X_{n+1}\in A \mid X_n = \frac{1}{k}\}$, $A_0 = \{X_{n+1} = \frac{1}{k+1}\}$ and $A_1 = \{X_{n+1}\sim U([0,1])\}$.

We can use the total probabilities formula:

$$P(x,A) = \mathbb{P}(B) = \mathbb{P}(B\cap A_0) + \mathbb{P}(B\cap A_1)$$

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

as we know that $\begin{cases}\mathbb{P}(A_0 \mid X_n = x (=\frac{1}{k})) = 1-x^2 \\ \mathbb{P}(A_1 \mid X_n = x) = x^2\end{cases}$

otherwise (if $X_n \neq \frac{1}{k}$) we set $x \neq \frac{1}{k}$ ($\nexists k \in \mathbb{N}^* / x = \frac{1}{k}$)

we get $P(x,A) = \mathbb{P}(X_{n+1} \in A \mid X_n = x)$

$$= \int_{\mathbb{R}} \mathbb{1}_{t \in A}(t)\, \mathbb{1}_{t \in [0,1]}(t)\, dt = \int_{[0,1]\cap A} dt .$$

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

$\square$

(2) let π be the Uniform distribution on $[0,1]$.

$\pi \overset{Law}{\equiv} \mathcal{U}([0,1])$. we have $\pi(x)dx = \mathbb{1}_{[0,1]}(x)\,dx$.

Let show that $\pi P = \pi$ ie show that

$$\pi(A) = \int_{\mathbb{R}} P(x,A)\, \pi(x)\, dx$$

$$\int_{\mathbb{R}} \pi(x) P(x,A)\, dx = \int_{\mathbb{R}} \mathbb{1}_{[0,1]}(x)\, P(x,A)\, dx .$$

$$= \int_{\mathbb{R}} \Big(\underbrace{\mathbb{1}_{[0,1]}(x)\, \mathbb{1}_{\{x = \frac{1}{k}\}}(x) \big[\ \dots\ \big]}_{=0 \text{ with respect to the Lebesgue measure}} + \mathbb{1}_{[0,1]}(x) \int_{A\cap[0,1]} dt \Big)\, dx$$

$$= \int_{\mathbb{R}} \Big(\int_{A\cap[0,1]} dt \Big)\, \mathbb{1}_{[0,1]}(x)\, dx$$

$$= \int_{\mathbb{R}} \Big(\underbrace{\int_A \mathbb{1}_{[0,1]}(t)\, dt}_{=\pi(A)} \Big)\, \mathbb{1}_{[0,1]}(x)\, dx$$

$$= \pi(A) \int_{\mathbb{R}} \mathbb{1}_{[0,1]}(x)\, dx$$

and $\int_{\mathbb{R}} \mathbb{1}_{[0,1]}(x)\, dx = 1$ (because $x \mapsto \mathbb{1}_{[0,1]}(x)$ is a density function of a $\mathcal{U}([0,1])$).

finally, we have

$$\int_{\mathbb{R}} \pi(x) P(x,A)\, dx = \pi(A) \times 1 = \pi(A).$$

Thus, $\pi P = \pi$.

$\square$

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

$Pf(x) = \mathbb{E}[f(X_1) \mid X_0 = x]$. $(X_1 \mid X_0 = x \sim \mathcal{U}([0,1]))$

$$= \int_{\mathbb{R}} f(y)\, \mathbb{1}_{[0,1]}(y)\, dy \ \cancel{= \int_{\mathbb{R}} f(y)\, \pi(y)\, dy}$$

~~Thus,~~ $\cancel{Pf(x) = \mathbb{E}_\pi[f(x)]}$.

The exercise told us that if $X_n \neq \frac{1}{k}$, then X_{n+1} is a $\mathcal{U}([0,1])$.

Let's compute $P^2 f(x)$:

$$P^2 f(x) = P(Pf(x)) = \mathbb{E}[Pf(X_1) \mid X_0 = x]$$

But, as $\mathcal{U}([0,1])$ is invariant for P, we get

$$P^2 f(x) = Pf(x) = \int_{\mathbb{R}} f(y)\, \mathbb{1}_{[0,1]}(y)\, dy .$$

Extending this to n, we get

$$P^n f(x) = P^{n-1}(Pf(x)) = \dots = Pf(x)$$

Thus, $P^n f(x) = \int f(y) \mathbb{1}_{[0,1]}(y)\,dy$
(for $x \notin \{\frac{1}{k}, k \in \mathbb{N}^*\}$).

Since $\forall n \geq 1$, $P^n f(x) = \int_0^1 f(y)\,dy$, we can have the limit easily:

$$\lim_{n \to +\infty} P^n f(x) = \int_0^1 f(y)\,dy$$

For this Markov chain, $\int_0^1 f(y)\,dy = \int_{\mathbb{R}} f(x)\pi(x)\,dx$.

Thus, we get

$$P^n f(x) \xrightarrow[n \to +\infty]{} \int_{\mathbb{R}} f(x)\pi(x)\,dx.$$

$\square$

4 a Let $x = \frac{1}{k}$, $k \geq 2$. Let $n \in \mathbb{N}^*$.
To move from $\frac{1}{k}$ to $\frac{1}{n+k}$ after n steps, there exists just one way for the chain, it's moving through $\frac{1}{k}, \frac{1}{k+1}, \dots, \frac{1}{k+n}$ after n steps.

Thus, $P^n\left(\frac{1}{k}, \frac{1}{n+k}\right) = P\left(\frac{1}{k}, \frac{1}{k+1}\right) P\left(\frac{1}{k+1}, \frac{1}{k+2}\right) \dots P\left(\frac{1}{k+n-1}, \frac{1}{n+k}\right)$

and $P\left(\frac{1}{k}, \frac{1}{k+1}\right) = 1 - \frac{1}{k^2}$, $P\left(\frac{1}{k+1}, \frac{1}{k+2}\right) = 1 - \frac{1}{(k+1)^2}$, etc.

Thus, we obtain

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

$\square$

3 I have choosen $\overline{w} = \begin{pmatrix} -1/2 \\ 1/2 \end{pmatrix}$.
The estimated w^* is the result of the SGD algorithm after 1000 iterations.
His value is $w^* = \begin{pmatrix} -0{,}1 \\ 0{,}12 \end{pmatrix}$. As it set randomly initially, w^* is close to $\overline{w}$, as it's not perfect we can improve the value of w^* by changing the learning rate or take small batches of data

$\square$

4 Adding Gaussian noise make w^*_{noise} less accurate.

(b) if $A = \bigcup_{q \in \mathbb{N}} \left\{ \frac{1}{k+1+q} \right\}$, A is a set of specific points, so we have $\pi(A) = \int dt = 0$.
(because $\pi \equiv \mathcal{U}[0,1]$)

Also, $P^n\left(x, \frac{1}{m+k}\right) \xrightarrow[n \to +\infty]{} 0$ as its a product of values between 0 and 1 $A \cap [0,1]$

Thus, here, we have

$$\lim_{m \to +\infty} P^m(x, A) = \pi(A) = 0.$$

$\square$

Exercice 3:

(1) one way to minimize $R_n(\omega)$ is to use the Stochastic gradient descent algorithm. Unlike the standart gradient descent (which uses the whole dataset), SGD uses one (or a small batch) of samples at a time to compute the gradient.
Here is an SGD algorithm: (that uses one sample):

1- Init w_0 (the parameter to estimate).

2- for each iteration:

- select random sample from the dataset (x_i, y_i)
- compute gradient: $\nabla_\omega R_n(\omega) = -2x_i(y_i - x_i^T \omega)$.
- update parameters: $\omega_{k+1} = \omega_k - \eta \nabla_\omega R_{(x_i,y_i)}(\omega_k)$

η is the learning rate

3 repeat 2 until convergence. (fixed number of iterations or when ω values stop changing significantly).

$\square$