TD 1 - Computational statistics

Exercice 1

méthode de la fonction muette

$X \underset{\text{sur } \mathbb{R}}{\sim} \mu$ $\iff$ $\forall h : \mathbb{R} \to \mathbb{R}$ continue et bornée

de densité f $\qquad \mathbb{E}[h(X)] = \int h(x) f(x) dx$

$$\begin{cases} \mu_\Theta(d\theta) = \frac{1}{2\pi} \mathbb{1}_{[0,2\pi]}(\theta) d\theta = f_\theta(\theta) d\theta \\ \mu_R(dr) = f_R(r) dr \end{cases}$$

1 _ Soit $h : \mathbb{R}^2 \to \mathbb{R}$, continue bornée

$$\mathbb{E}_{X,Y}[h(X,Y)] = \mathbb{E}_{R,\Theta}[h(R\cos\Theta, R\sin\Theta)] = (1)$$

Objectif : $\mathbb{E}[h(X,Y)] = \iint h(x,y)\gamma(x)\gamma(y) dx dy$,

γ densité de $\mathcal{N}(0,1)$.

$\gamma : x \mapsto \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-x^2}{2}\right)$

$$(1) = \int_{\mathbb{R}^2} h(r\cos\theta, r\sin\theta) f_r(r) f_\theta(\theta) dr d\theta$$

$$= \int_{\mathbb{R}_+} \int_0^{2\pi} h(r\cos\theta, r\sin\theta) \hat{f_r}(r) \hat{f_\theta}(\theta) dr d\theta$$

$(\hat{f}(x) = f(x) \mathbb{1}_{x \in A})$

Changement de variable :

$$\Phi : \begin{vmatrix} \mathbb{R} \times [0, 2\pi] \longrightarrow \mathbb{R} \times \mathbb{R} \\ (r, \theta) \longmapsto (r\cos\theta, r\sin\theta) \end{vmatrix}$$

Cette fonction est bijective :

$\triangle$

, $\text{Jac}[f][x] = \left(\frac{\partial f_i(x)}{\partial x_j}\right)_{i,j}$

Chgt de variable : Φ bijectif

$$\int_{\mathcal{U}} h(\phi(u))\,du = \int_{\phi(\mathcal{U})} h(v)\,|\text{Jac}[\Phi](v)|\,dv$$

$\hookrightarrow |\det J(\phi)(v)|$

Pour calculer le déterminant :

- soit on trouve Φ^{-1} : long
- $|\text{Jac}[\Phi^{-1}](v)| = |\text{Jac}[\Phi](v)|^{-1}$

Soit $r \in \mathbb{R}$, $\theta \in [0, 2\pi]$, $\text{Jac}[\Phi](r,\theta) = \begin{bmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{bmatrix}$

$$|\text{Jac}[\Phi](r,\theta)| = r$$

Donc : $$|\text{Jac}[\phi^{-1}](x,y)| = |\text{Jac}[\phi][\phi^{-1}(x,y)]|^{-1}$$

$$= \frac{1}{\sqrt{x^2+y^2}}$$

On applique le changement de variable :

$$= \int_{\mathbb{R}^2} h(x,y) \frac{1}{\sqrt{x^2+y^2}} \times \frac{1}{2\pi} \sqrt{x^2+y^2} \exp\left(-\frac{(x^2+y^2)}{2}\right) dx\,dy$$

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

Ainsi, $\forall h$ C^0 bornée, $\mathbb{E}[h(R\cos\theta, R\sin\theta)] = \mathbb{E}[h(X,Y)]$
$X, Y \sim \mathcal{N}(0,1)$, $X \perp Y$

donc : $R\cos\theta \sim \mathcal{N}(0,1)$, $R\sin\theta \sim \mathcal{N}(0,1)$ et sont indépendantes

2_ Idée : simuler R et θ et on construit $X = R\cos\theta$, $Y = R\sin\theta$

On sait simuler $\mathcal{U}[0,1]$ donc $2\pi\,\mathcal{U}[0,1]$.

Méthode : simulation avec la répartition inverse.

$$f : \begin{cases} \mathbb{R} \to [0,1] \\ x \mapsto \mathbb{P}(X \leq x) \end{cases}$$ par X une variable aléatoire réelle.

Soit $N \sim \mathcal{U}[0,1]$, supposons que f est bijective.
$f^{-1}(N)$ a la même loi que X.

En effet, $x \in \mathbb{R}$, $\quad$ F croissante

$$\mathbb{P}(F^{-1}(N) \leq x) \overset{\nearrow}{=} \mathbb{P}(N \leq F(x))$$
$$= F(x) = \mathbb{P}(X \leq x)$$

Soit F la df de Rayleigh :

Soit $x > 0$, $\quad F(x) = \int_0^x f_R(r)\,dr$

$$= \int_0^x r\exp\left(-\frac{r^2}{2}\right) dr$$

$$F(x) = 1 - \exp\left(-\frac{x^2}{2}\right)$$

F est strictement croissante sur $\mathbb{R}_+$ donc bijective dans $[0,1[$.

F^{-1} ? $\quad F(r) = u$

$$1 - \exp\left(-\frac{r^2}{2}\right) = u$$
$$-\frac{r^2}{2} = \ln(1-u)$$
$$r = \sqrt{2\ln\left(\frac{1}{1-u}\right)}$$

<u>Résumé de la technique de sampling</u>

1 _ Sample $U_1, U_2 \sim \mathcal{U}[0,1]$

2 _ ~~return~~ set $\theta = 2\pi U_1$

3 _ set $R = F_R^{-1}(U_2)$

4 _ Return $R\cos\theta, R\sin\theta$

F^{-1} est la fonction quantile.

On peut donc simuler R en tirant une $\mathcal{U}[0,1]$, ~~en multipliant par 2π le tirage~~ et on applique F^{-1} pour obtenir un échantillon de R.

3 _ 1 _ Loi de (V_1, V_2) conditionnée à $\{V_1^2 + V_2^2 \leq 1\}$

Méthode de la fonction muette :

$V_1, V_2 \sim \mathcal{U}([-1,1])$. $\rightarrow \mathbb{1}_{V_1^2+V_2^2 \leq 1}$, bernoulli

$\mathbb{E}[h(V_1,V_2) \mid V_1^2+V_2^2 \leq 1]$ $\qquad$ $\mathbb{E}[X|A] = \frac{\mathbb{E}[X\mathbb{1}_A]}{\mathbb{P}(A)}$, $\mathbb{P}(A) > 0$

$$= \frac{\mathbb{E}[h(V_1,V_2)\mathbb{1}_{V_1^2+V_2^2 \leq 1}]}{\mathbb{P}(V_1^2+V_2^2 \leq 1)}$$

$$= \int_{[-1,1]\times[-1,1]} h(v_1,v_2)\, \mathbb{1}_{\{v_1^2+v_2^2 \leq 1\}} \times cste \times dv_1 dv_2$$

$$cste = \frac{1}{2} \times \frac{1}{2} \times \frac{1}{\mathbb{P}(V_1^2+V_2^2 \leq 1)} = \frac{1}{4}\, \frac{1}{\mathbb{P}(V_1^2+V_2^2 \leq 1)}$$

$\rightarrow$ (2)

b) $T_1 = \frac{V_1}{\sqrt{V_1^2+V_2^2}}$ $\quad T_2 = \frac{V_2}{\sqrt{V_1^2+V_2^2}}$ $\quad U = V_1^2+V_2^2$

$h : \mathbb{R} \to \mathbb{R}$, continue bornée :

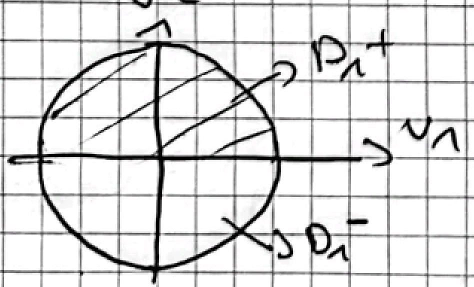
$$\mathbb{E}[h(T_1,V)] = \int_{D(0,1)} h(\Phi(v_1,v_2))\, f_C(v_1,v_2)\, dv_1 dv_2$$

$$\Phi : \left|\begin{array}{l} D(0,1) \longrightarrow [-1,1]\times[0,1] \\ \\ (v_1,v_2) \longmapsto \left(\frac{v_1}{\sqrt{v_1^2+v_2^2}},\ v_1^2+v_2^2\right) \end{array}\right.$$

Si on explicite Φ^{-1}

$V_1 = T_1\sqrt{U}$

$V_2^2 = U - T_1^2 U = U(1-T_1^2) \Rightarrow V_2 = \pm\sqrt{U(1-T_1^2)}$



$\phi_{|D_1^+}$ bijective

$\phi_{|D_1^-}$ bijective

(2) : ~~On a~~

We have : $\mathbb{E}[h(V_1,V_2) \mid V_1^2+V_2^2 \leq 1]$

$$= \int_{[-1,1]\times[-1,1]} h(v_1,v_2) \frac{\mathbb{1}_{\{v_1^2+v_2^2\leq 1\}}}{2\times 2\times \mathbb{P}(V_1^2+V_2^2\leq 1)} dv_1 dv_2$$

$$\mathbb{P}(V_1^2+V_2^2 \leq 1) = \frac{|B(0,1)|}{4}, \quad B(0,1) = \{x\in\mathbb{R}^2, \|x\|_2\leq 1\}$$

Donc ~~Hence~~:

$$\mathbb{E}[h(V_1,V_2) \mid V_1^2+V_2^2\leq 1] = \int_{B(0,1)} \frac{h(v_1,v_2)}{|B(0,1)|} dv_1 dv_2$$

$$= \int_{\mathbb{R}^2} h(v_1,v_2) \frac{\mathbb{1}_{(v_1,v_2)\in B(0,1)}}{|B(0,1)|} dv_1 dv_2$$

$= \mathbb{E}[h(X_1,X_2)]$, $(X_1,X_2) \sim$ uniform law dimension 2 on $B(0,1)$.

Donc ~~Hence~~: $\boxed{V_1,V_2 \mid V_1^2+V_2^2\leq 1 \sim \mathcal{U}(B(0,1))}$

b) Soit h $\mathcal{C}^0$ et bornée.

On considère : $\varphi : \begin{vmatrix} \mathbb{R}_+\times[0,2\pi[\longrightarrow \mathbb{R}^2 \\ (r,\theta) \longmapsto (\sqrt{r}\cos\theta, \sqrt{r}\sin\theta) \end{vmatrix}$

φ est $\mathcal{C}^1$-difféomorphisme sur $]0,1[\times[0,2\pi[$ →

$\varphi^{-1}(v_1,v_2) = (r,\theta)$

S

$$|J_\varphi(v,\theta)| = \begin{vmatrix} \frac{1}{2\sqrt{v}}\cos\theta & -\sqrt{v}\sin\theta \\ \frac{1}{2\sqrt{v}}\sin\theta & \sqrt{v}\cos\theta \end{vmatrix}$$

$$= \frac{1}{2}\cos^2\theta + \frac{1}{2}\sin^2\theta = \frac{1}{2}$$

$$\mathbb{E}[h(T_1,R)] = \iint_{\mathbb{R}^2} \frac{1}{\pi} \mathbb{1}_{(v_1,v_2)\in B(0,1)}\, h\left(\frac{v_1}{\sqrt{v_1^2+v_2^2}}, v_1^2+v_2^2\right) dv_1 dv_2$$

$$= \int_0^1\int_0^{2\pi} \frac{1}{\pi} |J_\varphi(v,\theta)|\, h\left(\frac{\sqrt{v}\cos\theta}{(\sqrt{v}\cos\theta)^2+(\sqrt{v}\sin\theta)^2}, v\right) dv\, d\theta$$

$$= \iint_{\mathbb{R}^2} h(\cos\theta, v)\, \mathbb{1}_{v\in[0,1]}(v) \frac{1}{2\pi} \mathbb{1}_{\theta\in[0,\pi]}(\theta)\, d\theta\, dv$$

Ainsi, $\boxed{\begin{cases} T_1 \sim \cos\theta \quad \text{avec} \quad \theta\sim\mathcal{U}[0,2\pi] \\ V\sim\mathcal{U}[0,1] \\ T_1 \perp\!\!\!\perp V \end{cases}}$

Montrons que $T_1, T_2 \sim \cos\theta, \sin\theta, \quad \theta\sim\mathcal{U}[0,2\pi]$

En utilisant le même changement de variable :

h c° bornée,

$$\mathbb{E}[h(T_1,T_2)] = \iint_{\mathbb{R}^2} h(\cos\theta,\sin\theta)\, \mathbb{1}_{x\in[0,1]}(x) \frac{1}{2\pi}\mathbb{1}_{\theta\in[0,2\pi]}(\theta)\, d\theta\, dx$$

$$= \int_{\mathbb{R}} h(\cos\theta,\sin\theta) \frac{1}{2\pi} \mathbb{1}_{\theta\in[0,2\pi]}(\theta)\, d\theta \times 1$$

Donc on a bien prouvé le résultat.

Enfin, en utilisant les mêmes méthodes de calcul,

6) on peut montrer que $T_2 \perp\!\!\!\perp V$.

c) $S = \sqrt{-2\log(V_1^2+V_2^2)}$

$S = \sqrt{-2\log V}$

$(X, Y) = (ST_1, ST_2)$

$V \sim \mathcal{U}([0,1])$, et $\underset{\text{Rayleigh}}{F_R^{-1}}(x) = \sqrt{-2\log(\underbrace{1-V_1}_{\sim V_1})}$

donc d'après les résultats précédents,

$S \sim \text{Rayleigh}(1)$

Ainsi, $(X,Y) = (ST_1, ST_2) \sim (R\cos\Theta, R\sin\Theta)$ en conservant les notations précédentes.

Ainsi : $X \perp\!\!\!\perp Y$, $X \sim \mathcal{N}(0,1)$, $Y \sim \mathcal{N}(0,1)$

d) On pose $Z \sim \begin{cases} 1 & \text{avec proba } \mathbb{P}(V_1^2+V_2^2 \leq 1) \\ 0 & \text{avec proba } 1-\mathbb{P}(V_1^2+V_2^2\leq 1) \end{cases}$

Soit N le nombre d'itérations.

On reconnait alors une loi géométrique :

$n \geq 0$ $\mathbb{P}(N=n) = \mathbb{P}\left(\prod_{i=1}^{n} Z_i = 0, \prod_{i=1}^{n-1} Z_i = 1\right)$

$= \mathbb{P}(Z_1 = 1, \ldots, Z_{n-1} = 1, Z_n = 0)$

Donc : $N \sim \mathcal{G}(p = \mathbb{P}(V_1^2+V_2^2 \leq 1))$

On a calculé que $\mathbb{P}(V_1^2+V_2^2\leq 1) = \frac{\pi}{4}$,

et $X \sim \mathcal{G}(p) \Rightarrow \mathbb{E}[X] = \frac{1}{p}$,

Donc :

$\boxed{\mathbb{E}[N] = \frac{1}{\frac{\pi}{4}} = \frac{4}{\pi} < 2}$

Exercice 2

1 - Soit ~~$x \neq \frac{1}{m}$~~ $x \in [0,1]$ tel que

$x \neq \frac{1}{m}$, $\forall m \in \mathbb{N}^*$

Avec h C^0 et bornée :

$$\mathbb{E}[h(X_{n+1}) \mid X_n = x] = \int_{\mathbb{R}} h(y) \underbrace{P(x,dy)}_{\text{ici uniforme sur } [0,1]}$$

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

Ainsi : $\boxed{P(x,A) = \int_A P(x,dy) = \int_A \mathbb{1}_{[0,1]}(y)dy = \int_{A \cap [0,1]} dt}$

- $\exists m > 0, \in \mathbb{N}$, $x = \frac{1}{m}$

Dans ce cas :

$$\mathbb{E}[h(X_{n+1}) \mid X_n = x] = h\left(\frac{1}{m+1}\right)(1 - x^2) + \frac{1}{x^2}\int_{[0,1]} h(t)dt$$

Par conséquent :

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

δ

On note $B = \{\frac{1}{n}, n \in \mathbb{N}^*\}$

On veut montrer que :

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

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

$$= \int_{[0,1]} P(x,A) \left[1\!\!1_{x\in B} + 1\!\!1_{x \notin B} \right] dx$$

$$\underset{?}{=} \int_{[0,1]} \int_{A\cap[0,1]} dt \; dx = \int_{[0,1]\cap A} dt \times 1$$

$\int 1\!\!1_{x\in B}\,dx = 0$ car $\lambda(B) = 0$
Lebesgue

$$\boxed{= \pi(A)}$$

3 _ Soit $x \notin B$.

$X_1 | X_0 = x \sim \mathcal{U}[0,1]$

donc pour f, $\mathbb{E}[f(X_1) | X_0 = x] = \int_0^1 f(x)\,dx$

Montrons par récurrence que : $\forall n \geq 1$, $X_n | X_0 = x \sim \mathcal{U}[0,1]$

C'est vrai pour $n = 1$.

Soit $n \geq 1$ telle que ce soit vrai.

9

On a par ~~construction~~ : définition :
$\forall h$ $\mathcal{E}_0$ bornée

$\mathbb{E}[h(X_{n+1}) \mid X_0 = x]$

$= \mathbb{E}[h(X_{n+1}) \mid X_n \in [0,1] \setminus B, X_0 = x]$
$\times \mathbb{P}(X_n \in [0,1] \setminus B, X_0 = x)$
$+ \mathbb{E}(h(X_{n+1}) \mid X_n \in B, X_0 = x) \mathbb{P}(X_n \in B, X_0 = x)$

Par hypothèse, $X_n \underset{|X_0 = x}{\sim} \mathcal{U}[0,1]$ donc $\begin{cases} \mathbb{P}(X_n \in B, X_0 = x) = 0 \\ \mathbb{P}(X_n \in [0,1] \setminus B, X_0 = x) = 1 \end{cases}$

d'où $\mathbb{E}[h(X_{n+1}) \mid X_0 = x] =$
$\mathbb{E}[h(X_{n+1}) \mid X_n \in [0,1] \setminus B, X_0 = x]$

Or, $X_{n+1} \mid X_n \in [0,1] \setminus B, X_0 = x \sim \mathcal{U}[0,1]$

Donc : l'hypothèse est vérifiée pour tout $n \geq 1$.

Par conséquence : $\forall n \geq 1$, $\mathbb{E}[f(X_n) \mid X_0 = x] = \int_0^1 f(x)\,dx$

ce qui implique aussi :

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

4_ $x = \frac{1}{m}$, $m \geq 2$

Montrons que, $n \in \mathbb{N}^*$, $\boxed{P^n\left(x, \frac{1}{n+m}\right) = \prod_{j=0}^{n-1} \left(1 - \frac{1}{(j+m)^2}\right)}$

Par récurrence :

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

▸ vrai pour $n \in \mathbb{N}^*$,

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

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

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

Par récurrence, on a prouvé le résultat.

b) $n \geq 1$, $P^n(x, A) = P^n\left(x, \bigcup_{q \geq 0} \frac{1}{m+1+q}\right)$

$$\overset{\text{sigma-additivité}}{=} \sum_{q \geq 0} P^n\left(x, \frac{1}{m+1+q}\right)$$

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

car en n étapes, on ne peut pas arriver à $\frac{1}{m+1+q}$, $1+q \neq n$ avec probabilité non nulle !!

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

$$\ln\left(\prod_{j=0}^{n-1}\left(1-\left(\frac{1}{n+j}\right)^2\right)\right) = \sum_{j=0}^{n-1} \ln\left(1-\left(\frac{1}{n+j}\right)^2\right)$$

$$\ln\left(1-\left(\frac{1}{n+j}\right)^2\right) \sim -\frac{1}{(n+j)^2}$$, comme

ces deux suites sont de même signe,

$\ln\left(\prod_{j=0}^{n-1}\left(1-\left(\frac{1}{n+j}\right)^2\right)\right)$ converge car le

terme sommé est équivalent à un terme sommable

Donc : $\ln \prod_{j=0}^{n-1}\left(1-\left(\frac{1}{n+j}\right)^2\right) \longrightarrow a > 0$

($a \neq 0$ impossible car terme de signe constant)

Donc : $\boxed{p^n\left(\frac{x}{x}, \theta\right) \xrightarrow[n \to +\infty]{} e^a > 0 \neq \underset{\overset{\shortparallel}{0}}{\pi(\theta)}}$

Montrons plus précisément que $\mathbb{P}(U_1^2+U_2^2 \leq 1) = \frac{\pi}{4}$

- $\boxed{\mathbb{P}(V_1^2+V_2^2 \leq 1) = \mathbb{E}[\mathbb{1}_{V_1^2+V_2^2 \leq 1}]}$

$$= \int_{\mathbb{R}^2} \mathbb{1}_{\{(2u_1-1)^2+(2u_2-1)^2 \leq 1\}} \mathbb{1}_{[0,1]}(u_1)\, \mathbb{1}_{[0,1]}(u_2)\, du_1 du_2$$

$$= \int_0^1 \int_0^1 \mathbb{1}_{\{(2u_1-1)^2+(2u_2-1)^2 \leq 1\}}(u_1,u_2)\, du_1 du_2$$

On pose les changements de variable affines:

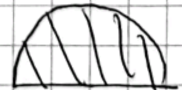
$$x = 2u_1 - 1,\ y = 2u_2 - 1 \Rightarrow \begin{cases} dx = 2du_1 \\ dy = 2du_2 \end{cases}$$

$$= \frac{1}{4} \int_{-1}^1 \int_{-1}^1 \mathbb{1}_{x^2+y^2 \leq 1}\, dy\, dx$$

$$= \frac{1}{4} \int_{-1}^1 \left(\int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} dy \right) dx$$

$$= \frac{1}{4} \int_{-1}^1 2\sqrt{1-x^2}\, dx = \frac{1}{2} \int_{-1}^1 \sqrt{1-x^2}\, dx$$

$x \mapsto \sqrt{1-x^2}$ est l'équation d'un arc de cercle de rayon 1 centré en 0 :



Ainsi :

$$\boxed{= \frac{1}{2} \times \underbrace{\frac{\pi}{2}}_{\text{aire demi-cercle}} = \frac{\pi}{4}}$$