Basile TERVER

Copie 1

TP 1 Computational Statistics

Exercice 1

On note f_θ la densité de la variable θ : $f_\theta = \frac{1}{2\pi} \mathbb{1}_{[0,2\pi]}$

(1) Soit h une fonction continue bornée. On a

$$E_{X,Y}[h(X,Y)] = E_{R,\theta}[h(R\cos\theta, R\sin\theta)]$$

$$= \frac{1}{2\pi}\int_0^{2\pi}\int_0^{+\infty} h(r\cos\theta, r\sin\theta)\, r e^{-r^2/2}\, dr\, d\theta \qquad (*)$$

car, par indépendance de R et θ, la densité jointe du couple (R,θ), notée $f_{(R,\theta)}$ est égale au produit des densités marginales, soit $f_{(R,\theta)} = f_R \cdot f_\theta$.

- Soit $\phi : U := \mathbb{R}_*^+ \times [0, 2\pi[\to \mathbb{R}^2 \setminus \{(0,0)\} =: V$

$$(r,\theta) \mapsto (r\cos\theta, r\sin\theta)$$

U et V sont respectivement égaux à $\mathbb{R}^+ \times [0,2\pi]$ et $\mathbb{R}^2$ à ensembles de mesure nulle près

ϕ est bijective car pour tout $(x,y) \in \mathbb{R}^2 \setminus \{(0,0)\}$, $(r,\theta) \in \mathbb{R}_*^+ \times [0, 2\pi[$

$$\begin{cases} x = r\cos\theta \\ y = r\sin\theta \end{cases} \Leftrightarrow \begin{cases} r^2 = x^2+y^2 \\ \theta = \arctan\left(\frac{y}{x}\right) \end{cases}$$ et arctan est bijective

ϕ est un C^1-difféomorphisme donc par théorème de changement de variables, comme $J\phi(r,\theta) = \begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix}$ pour tout $(r,\theta) \in U$, et car $|J\phi^{-1}(x,y)| = |J\phi(x,y)|^{-1} = [r(\cos^2\theta + \sin^2\theta)]^{-1} = r^{-1}$ pour tout $(x,y) \in V$

et en utilisant $(*)$:

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

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

$$= \iint_{\mathbb{R}^2} h(x,y)\, \varphi(x)\varphi(y)\, dx\, dy$$

où φ est la densité d'une loi $\mathcal{N}(0,1)$.

Donc (X,Y) a une densité $f_{(X,Y)}$ qui se factorise en $f_{(X,Y)}(x,y) = \varphi(x)\varphi(y)$ pour tout $(x,y) \in \mathbb{R}^2$.

Scanned with CamScanner

Ainsi, X et Y sont indépendantes de loi $\mathcal{N}(0,1)$.

② Pour simuler une loi, la méthode la plus fiable numériquement est de simuler une loi uniforme sur $[0,1]$ puis d'utiliser la fonction de répartition inverse de la loi à simuler si celle-ci à une forme explicite, ce qui est le cas ici.

On simulera deux gaussiennes indépendantes en simulant une loi de Rayleigh(1) et une loi $\mathcal{U}[0,2\pi]$ puis en utilisant la transformation de la question ①.

- Pour simuler $\mathcal{U}[0,2\pi]$, on simule $\mathcal{U}[0,1]$ et on multiplie le résultat par 2π car $\mathcal{U}[0,2\pi] = 2\pi\,\mathcal{U}[0,1]$.

- Soit $R \sim \text{Rayleigh}(1)$, on note sa fonction de répartition F_R. On a pour tout x réel :

$$F_R(x) = P(R \leq x) = \int_{-\infty}^{x} f_R(r)\,dr = \int_{-\infty}^{0} r e^{-r^2/2}\,dr = 1 - e^{-x^2/2}$$

On a donc pour tout $u \in [0,1]$:

$$\forall x \in \mathbb{R},\ F_R(x) = u \iff 1 - e^{-x^2/2} = u \iff 1 - u = e^{-x^2/2}$$

$$\iff -\log(1-u) = \frac{x^2}{2} \iff x = \sqrt{-2\log(1-u)}$$

$\boxed{\text{On remarque que : } U \sim \mathcal{U}[0,1] \iff 1-U \sim \mathcal{U}[0,1]}$

Donc : $\forall u \in [0,1], F_R^{-1}(u) = \sqrt{-2\log(1-u)}$. Par le théorème de la transformée inverse on a donc $R \overset{\text{loi}}{=} \sqrt{-2\log(1-U)}$ avec $U \sim \mathcal{U}[0,1]$.

Donc si $U \sim \mathcal{U}[0,1]$, on a $\sqrt{-2\log(U)} \sim \text{Rayleigh}(1)$.

On propose donc l'algorithme suivant pour simuler 2 gaussiennes indépendantes :

Scanned with CamScanner

Algorithme :

Echantillonner indépendamment U_1, U_2 de loi $\mathcal{U}[0,1]$

Poser $\theta = 2\pi U_1$, $R = \sqrt{\log\left(\frac{1}{U_2^2}\right)}$

Poser $X = R\cos(\theta)$, $Y = R\sin(\theta)$

Retourner (X, Y)

③ ⓐ La loi de (V_1, V_2) à la fin de la boucle est la loi de $(V_1, V_2 \mid V_1^2 + V_2^2 \leq 1)$ car les tirages entre chaque itération de boucle sont indépendants.

Au sein d'une même boucle par définition de V_1, V_2 en fonction de U_1, U_2, on a $V_1, V_2 \sim \mathcal{U}[-1,1]$.

- Soit h une fonction continue bornée. On a :

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

- $\mathbb{1}_{\{V_1^2+V_2^2\leq 1\}}$ est une variable de Bernoulli de paramètre

$$\mathbb{P}(V_1^2+V_2^2 \leq 1) = \int_{-1}^{1}\int_{-1}^{1} \frac{dv_1}{2}\frac{dv_2}{2}\mathbb{1}_{\{v_1^2+v_2^2\leq 1\}} = \frac{1}{4}A(D(0,1)) = \frac{\pi}{4}$$

où $A(D(0,1))$ est l'aire du disque $D(0,1)$ de centre 0 et de rayon 1. En effet, par indépendance de U_1 et U_2, on a indépendance de V_1, V_2 donc la densité jointe $f(v_1,v_2)$ se factorise en $f(v_1,v_2) = f_{V_1} \cdot f_{V_2}$.

- On a donc :

$$E_{V_1,V_2}\left[h(V_1,V_2) \mid V_1^2+V_2^2 \leq 1\right] = \frac{4}{\pi}\iint_{D(0,1)} \frac{h(v_1,v_2)}{4}\,dv_1\,dv_2 = \frac{1}{\pi}\iint_{D(0,1)} h(v_1,v_2)\,dv_1\,dv_2$$

Donc $\underline{(V_1, V_2) \sim \mathcal{U}(D(0,1))}$.

ⓑ • Montrons que $T_1 \perp\!\!\!\perp V$. Soit h continue bornée. On considère désormais le couple $(V_1, V_2 \mid V_1^2+V_2^2 \leq 1)$ uniquement et on note (V_1, V_2) par souci de concision.

Scanned with CamScanner

Notons $D(0,1)^{\pm} = D(0,1) \cap (\mathbb{R} \times \mathbb{R}^{\pm})$. On a :

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

$$= \frac{1}{\pi} \iint_{D(0,1)^-} h\left(\frac{v_1}{\sqrt{v_1^2+v_2^2}}, v_1^2+v_2^2\right) dv_1\, dv_2 + \iint_{D(0,1)^+} h\left(\frac{v_1}{\sqrt{v_1^2+v_2^2}}, v_1^2+v_2^2\right) dv_1\, dv_2$$

• Posons $\varphi_+ : D(0,1)^+ \to [-1,1] \times [0,1]$

$$(v_1, v_2) \mapsto \left(\frac{v_1}{\sqrt{v_1^2+v_2^2}}, v_1^2+v_2^2\right)$$

$\varphi_- : D(0,1)^- \to [-1,1] \times [0,1]$

$$(v_1, v_2) \mapsto \left(\frac{v_1}{\sqrt{v_1^2+v_2^2}}, v_1^2+v_2^2\right)$$

• Si $(v_1, v_2) \in D(0,1)^{\pm}$ et $(t,v) \in [-1,1] \times [0,1]$, on a :

$$\varphi_{\pm}(v_1,v_2) = (t_1, v) \iff \begin{cases} t_1 = \frac{v_1}{\sqrt{v_1^2+v_2^2}} \\ v = v_1^2+v_2^2 \end{cases} \iff \begin{cases} v_1 = t_1\sqrt{v} \\ v_2 = \pm\sqrt{v - v_1^2} = \pm\sqrt{v(1-t_1^2)} \end{cases}$$

Donc $\varphi_{\pm}^{-1}(t_1, v) = (t_1\sqrt{v}, \pm\sqrt{v(1-t_1^2)})$ et φ^+ et φ^- sont des bijections.

De plus $J\varphi_{\pm}^{-1}(t_1, v) = \begin{pmatrix} \sqrt{v} & \frac{t_1}{2\sqrt{v}} \\ \pm\frac{(-t_1)\sqrt{v}}{\sqrt{1-t^2}} & \pm\frac{\sqrt{1-t_1^2}}{2\sqrt{v}} \end{pmatrix}$

Donc $|J\varphi_{\pm}^{-1}(t_1, v)| = \pm\frac{\sqrt{1-t_1^2}}{2} \pm \frac{t_1^2}{2\sqrt{1-t^2}} = \pm\frac{1}{2\sqrt{1-t^2}}$

Donc par changement de variables :

$$E[h(T_1,V)] = \frac{2}{\pi} \iint_{[-1,1]\times[0,1]} h(t_1, v) \frac{1}{2\sqrt{1-t_1^2}}\, dt_1\, dv$$

Donc (T_1, V) a pour densité $f_{(T_1,V)}(t_1,v) = \frac{\mathbb{1}_{[-1,1]}(t_1)}{\pi\sqrt{1-t_1^2}} \mathbb{1}_{[0,1]}(v)$ qui se factorise en une fonction de T_1 et une fonction de V donc $\underline{T_1 \perp\!\!\!\perp V}$. Le même raisonnement en permutant les rôles des indices 1 et 2 donne $\underline{T_2 \perp\!\!\!\perp V}$. Cette factorisation de densité donne également que $\underline{V \sim \mathcal{U}[0,1]}$ et $\forall t_1 \in [-1,1]$, $f_{T_1}(t_1) = \frac{\mathbb{1}_{[-1,1]}(t_1)}{\pi\sqrt{1-t_1^2}}$. Donc si h continue bornée :

$$E_{T_1}[h \circ \arccos(T_1)] = \int_{-1}^{1} \frac{h(\arccos(t_1))}{\pi\sqrt{1-t_1^2}}\, dt_1 \qquad \arccos(t_1) = \theta, \quad t_1 = \cos\theta, \quad dt_1 = -\sin\theta\, d\theta$$

$$= \int_{\pi}^{0} \frac{h(\theta)}{\pi\sqrt{1-\cos^2\theta}} (-\sin\theta)\, d\theta$$

$$= \int_0^{\pi} \frac{h(\theta)\, d\theta}{\pi}$$

Scanned with CamScanner

Basile TERVER

Copie 2

Donc $\arccos(T_1) \sim \mathcal{U}[0,\pi]$, donc par bijectivité de l'arccos, $T_1 = \cos\theta$ avec $\theta \sim \mathcal{U}[0,\pi]$. Or, On remarque que si $\theta_1 \sim \mathcal{U}[0,\pi]$, $\theta_2 \sim \mathcal{U}[0,2\pi]$, alors $\cos\theta_1 \overset{loi}{=} \cos\theta_2$. Donc $\underline{T_1 = \cos\theta \text{ avec } \theta \sim \mathcal{U}[0,2\pi]}$.

Le même raisonnement avec T_2 à la place de T_1 donne $\underline{T_2 = \sin\theta' \text{ avec } \theta' \sim \mathcal{U}[0,2\pi]}$.

Si l'on avait montré l'indépendance $(T_1, T_2) \perp\!\!\!\perp V$, on aurait obtenu que $(T_1, T_2) \overset{loi}{=} (\cos\theta, \sin\theta)$, $\theta \sim \mathcal{U}[0,2\pi]$.

c) Comme $V \sim \mathcal{U}[0,1]$, on a par la question 2, que $S = \sqrt{-2\log(V)} \sim \text{Rayleigh}(1)$.

De même, comme $(T_1, T_2) \overset{loi}{=} (\cos\theta, \sin\theta)$ avec $\theta \sim \mathcal{U}[0,2\pi]$ et $(T_1, T_2) \perp\!\!\!\perp V$, on a $\theta \perp\!\!\!\perp V$. Par la question 2 on a donc $(X, Y) \overset{loi}{=} (ST_1, ST_2) \overset{loi}{=} (S\cos\theta, S\sin\theta)$

$\underline{(X,Y) \sim \mathcal{N}_2(0,1)}$ un produit de 2 gaussiennes indépendantes.

Scanned with CamScanner

① Soit Z la variable aléatoire représentant le nombre de toutes boucles "while" nécessaires pour en sortir.

L'ensemble des tours sont des épreuves de Bernoulli indépendantes de probabilité de succès $\mathbb{P}(V_1^2+V_2^2 \leq 1) = \frac{\pi}{4}$ (V_1, V_2 sont ici entendus au sens de la question ③a).

Donc Z suit une loi géométrique de paramètre $\frac{\pi}{4} = p$.

Donc $\underline{E[Z] = \frac{1}{p} = \frac{4}{\pi}}$.

Exercice 2 : On note $1/\mathbb{N}^* = \{x = \frac{1}{m}, m \in \mathbb{N}^*\}$

① Cas $x \notin 1/\mathbb{N}^*$: Dans ce cas $\mathcal{L}(X_{n+1} \mid X_n = x) = \mathcal{U}[0,1]$.

D'où $P(x, A) = \int_{A \cap [0,1]} dt$ pour $A \in \mathcal{B}([0,1])$

Cas $x \in 1/\mathbb{N}^*$: Soit $x = \frac{1}{m}$, $m \in \mathbb{N}^*$. Soit h continue bornée, en posant $B_n = \mathbb{1}_{\{X_{n+1} = \frac{1}{m+1}\}}$, on a :

$E_{X_{n+1}}[h(X_{n+1}) \mid X_n = x] = E_{X_{n+1}, B_n}[h(X_{n+1})(\mathbb{1}_{B_n=0} + \mathbb{1}_{B_n=1}) \mid X_n = x]$

$= E_{X_{n+1}}[h(X_{n+1}) \mid B_n = 0, X_n = x]\mathbb{P}(B_n = 0 \mid X_n = x) + E_{X_{n+1}}[h(X_{n+1}) \mid B_n = 1, X_n = x]\mathbb{P}(B_n = 1 \mid X_n = x)$

$= E_{X_{n+1}}[h(X_{n+1}) \mid B_n = 0, X_n = x)\, x^2 + E_{X_{n+1}}[h(X_{n+1}) \mid B_n = 1, X_n = x](1 - x^2)$

$= E[h(U)]\, x^2 + h(\frac{1}{m+1})(1 - x^2)$ avec $U \sim \mathcal{U}[0,1]$

D'où : $P(x, A) = x^2 \int_{A \cap [0,1]} dt + (1 - x^2)\, \delta_{\frac{1}{m+1}}(A)$ pour tout $A \in \mathcal{B}([0,1])$

② Soit A un borélien de $\mathcal{B}([0,1])$. On a (en notant λ la mesure de Lebesgue) :

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

$= \sum_{m \in \mathbb{N}^*} \int_{\{x = \frac{1}{m}\}} \left[x^2 \lambda(A)\, dx + (1 - x^2)\, \delta_{\frac{1}{m+1}}(A)\, dx\right] + \int_{\{x \notin 1/\mathbb{N}^*\}} \lambda(A)\, dx$

$= \lambda(A) = \int_{[0,1] \cap A} dx$ car $1/\mathbb{N}^*$ est dénombrable donc de mesure de Lebesgue nulle.

Scanned with CamScanner

Donc $\underline{\pi P = \pi}$ et π est bien invariante pour P.

(3)(i) Si f est continue bornée, on a

$P f(x) = \int_{[0,1]} f(t)\,dt = E[f(U)]$, $U \sim \mathcal{U}[0,1]$.

Donc, avec probabilité 1, $P f(X_1) \notin 1/\mathbb{N}^*$.

(ii) Supposons que $X_n \notin 1/\mathbb{N}^*$ presque sûrement et que $P^n f(x) = E[f(U)]$, $U \sim \mathcal{U}[0,1]$. On a par définition de Pf et homogénéité de la chaîne (X_n) que :

$P^{n+1} f(x) = P(P^n f)(x) = E_\pi[E[f(U)]] = E[f(U')]$, $U' \sim \mathcal{U}[0,1]$.

Donc par récurrence : $\forall n \in \mathbb{N}$, $P^n f(x) = E[f(U)]$ pour une variable $U \sim \mathcal{U}[0,1]$

Donc $\boxed{\lim_n P^n f(x) = \int_{[0,1]} f(x)\,\pi(x)\,dx}$

(4)(a)(i) On a $P\left(x, \frac{1}{m+1}\right) = 1 - x^2 = 1 - \left(\frac{1}{m}\right)^2$

(ii) Soit $n \in \mathbb{N}^*$ tel que $P^n\left(x, \frac{1}{n+m}\right) = \prod_{k=0}^{n-1}\left[1 - \left(\frac{1}{m+k}\right)^2\right]$

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

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

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

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

Donc, par récurrence : $\forall m, n \in \mathbb{N}^*$, $P^n\left(\frac{1}{m}, \frac{1}{n+m}\right) = \prod_{k=0}^{n-1}\left(1 - \frac{1}{(m+k)^2}\right)$

Scanned with CamScanner

(b) Si $A = \bigcup_{q \in \mathbb{N}^*} \left\{ \frac{1}{m+1+q} \right\}$

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

d'après les questions précédentes.

Donc par la question (1 a) : $P^n(x, A) = \prod_{k=0}^{n-1} \left(1 - \frac{1}{(m+k)^2}\right)$

On remarque que $\pi(A) = 0$ car A est dénombrable.

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

$$= \sum_{k=0}^{n-1} \left[-\frac{1}{(m+k)^2} + o\left(\frac{1}{(m+k)^2}\right) \right]$$

$$= -\sum_{k=m}^{m+n-1} \left[\frac{1}{k^2} + o\left(\frac{1}{k^2}\right) \right]$$

Cette série à termes de signe constant converge par comparaison à la série $\sum_{n \geq 1} \frac{1}{n^2}$.

Donc $\lim_n \log(P^n(x, A)) < +\infty$ donc $\underline{\lim_n P^n(x, A) > 0 = \pi(A)}$

Scanned with CamScanner