Computational statistics - TP1 - BLANCHARD Théau

Exercice 1 :

1°) Soit h une fonction continue bornée. Alors

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

$$= \int h(r\cos\theta, r\sin\theta) f_R(r) f_\theta(\theta)\, dr d\theta, \text{ car } \theta \perp\!\!\!\perp R$$

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

On pose alors $\varphi : \begin{array}{|l} \mathbb{R}^+_* \times]0,2\pi[\mapsto \mathbb{R}^2 \\ (r,\theta) \mapsto (r\cos\theta, r\sin\theta) \end{array}$ qui est un difféomorphisme

de jacobien $J_\varphi(r,\theta) = \begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix}$ d'où $|\det(J_\varphi(r,\theta))| = r > 0$

Si on pose $f(x,y) = h(x,y)\exp(-\frac{x^2+y^2}{2})$ alors on a :

$$\mathbb{E}[h(X,Y)] = \int_{\mathbb{R}^+_* \times]0,2\pi[} (f \circ \varphi)(r,\theta) . |\det(J_\varphi(r,\theta))| \frac{1}{2\pi} dr d\theta$$

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

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

On en conclue donc que X et Y sont indépendants et suivent tous les deux une loi $\mathcal{N}(0,1)$.

2°) Soit R une variable aléatoire suivant une loi de Rayleigh de fonction de répartition F_R. Alors pour tout $t \in \mathbb{R}$:

$$F_R(t) = \int_0^t r\exp(-\frac{r^2}{2}) dr = \left[-\exp(-\frac{r^2}{2}) \right]_0^t = 1 - \exp(-\frac{t^2}{2}) .$$

Scanné avec CamScanner

F_R est continue et strictement croissante sur $]0,1[$ donc inversible : Soit $y\in]0,1[$.

$y = F_R(t) \Leftrightarrow 1-y = \exp(-\frac{t^2}{2}) \Leftrightarrow -\frac{t^2}{2} = \ln(1-y) \Leftrightarrow t = \sqrt{-2\ln(1-y)}$

Donc $F_R^{-1}(t) = \sqrt{-2\ln(1-t)}$.

Alors par le théorème d'inversion si $U \sim \mathcal{U}([0,1])$ alors $F_R^{-1}(U)$ suit une distribution de Rayleigh.

Algorithme :
- Sample $u_1 \sim \mathcal{U}([0,1])$
- Sample $u_2 \sim \mathcal{U}([0,1])$
- $r = \sqrt{-2\ln(1-u_1)}$
- return $r\cos(u_2), r\sin(u_1)$.

Par la question 1 on a donc bien généré deux variables aléatoires gaussiennes centrées réduites et indépendantes.

3) a) Dans la boucle V_1 et V_2 sont indépendants parce que U_1 et U_2 le sont et suivent une loi uniforme sur $[-1,1]$.

Si on note $E = \{(v_1,v_2) \in [-1,1]^2 / v_1^2+v_2^2 < 1\}$ alors $\int_E dv_1 dv_2 = \pi$.

On a donc que (V_1,V_2) suivent une loi de densité $f_{(V_1,V_2)}(v_1,v_2) = \frac{1}{\pi}\cdot \mathbb{1}_{v_1^2+v_2^2\leq 1}$

b) Soit h une fonction continue bornée. Alors

$$\mathbb{E}[h(T_1,V)] = \int_{-1}^{1}\int_{-1}^{1} h\left(\frac{v_1}{\sqrt{v_1^2+v_2^2}}, v_1^2+v_2^2\right)\cdot\frac{1}{\pi}\cdot\mathbb{1}_{v_1^2+v_2^2<1}\, dv_1 dv_2$$

$$= \frac{2}{\pi}\int_{\mathbb{R}^2} h\left(\frac{v_1}{\sqrt{v_1^2+v_2^2}}, v_1^2+v_2^2\right)\cdot\mathbb{1}_{v_1^2+v_2^2<1}\cdot\mathbb{1}_{v_2\geq 0}\, dv_1 dv_2, \text{ par parité en } v_2.$$

Soit $\varphi : \begin{array}{l}]-1,1[\times]0,1[\longrightarrow]-1,1[\times]0,1[\\ (v_1,v_2) \longmapsto \left(\frac{v_1}{\sqrt{v_1^2+v_2^2}}, v_1^2+v_2^2\right)\end{array}$ et $(t,v)\in]-1,1[\times]0,1[$.

$$\varphi(v_1,v_2) = (t,v) \Leftrightarrow \begin{cases} v = v_1^2+v_2^2 \\ t = \frac{v_1}{\sqrt{v_1^2+v_2^2}}\end{cases} \Leftrightarrow \begin{cases} v_1 = t\sqrt{v} \\ v_2 = \sqrt{v(1-t^2)}\end{cases}$$

Scanné avec CamScanner

Alors $\varphi^{-1} \left| \begin{array}{l}]-1,1[\times]0,1[\to]-1,1[\times]0,1[\\ (t,v) \mapsto t\sqrt{v}, \sqrt{v(1-t^2)} \end{array} \right.$

Donc φ est bijective, différentiable et φ^{-1} l'est aussi. Donc φ est un difféomorphisme et on a :

$$J_{\varphi^{-1}}(t,v) = \begin{pmatrix} \sqrt{v} & \frac{t}{2\sqrt{v}} \\ \frac{-t\sqrt{v}}{\sqrt{1-t^2}} & \frac{1-t^2}{2\sqrt{v(1-t^2)}} \end{pmatrix} \text{ et donc } |\det(J_{\varphi^{-1}}(t,v))| = \frac{1}{2\sqrt{1-t^2}} > 0.$$

Alors on a par changement de variable

$$\mathbb{E}[h(T_1,V)] = \int_{\mathbb{R}^2} h(t,v)\, \mathbb{1}_{v\in[0,1]} \frac{1}{\pi\sqrt{1-t^2}} \cdot \mathbb{1}_{t\in]-1,1[}\, dt\, dv.$$

On a alors que T_1 et V sont indépendants, que V suit une loi uniforme sur $[0,1]$ et que T_1 suit une loi de densité $f_{T_1}(t) = \frac{1}{\pi\sqrt{1-t^2}} \cdot \mathbb{1}_{[-1,1]}$.

Alors : $\mathbb{E}[h(T_1)] = \frac{1}{\pi} \int_{-1}^{1} h(t) \frac{dt}{\sqrt{1-t^2}}$. On fait le changement de variable $t = \cos\theta$

donc $\theta = \arccos(t)$ et $d\theta = \frac{-1}{\sqrt{1-t^2}} dt$ d'où

$$\mathbb{E}[h(T_1)] = -\frac{1}{\pi} \int_{\pi}^{0} h(\cos\theta)\, d\theta = \frac{1}{\pi} \int_0^{\pi} h(\cos(\theta)) d\theta$$

Or $\int_0^{\pi} h(\cos(\theta)) d\theta = \int_{-\pi}^{0} h(\cos(x))\, dx$ avec le changement de variable $x = -\theta$

$= \int_{\pi}^{2\pi} h(\cos(\theta - 2\pi))\, d\theta$, avec le cdv $\theta = x + 2\pi$

$= \int_{\pi}^{2\pi} h(\cos\theta)\, d\theta.$

Donc $\mathbb{E}[h(T_1)] = \frac{1}{2\pi} \cdot 2 \int_0^{\pi} h(\cos\theta) d\theta = \int_0^{2\pi} h(\cos\theta) \frac{1}{2\pi} d\theta.$

Donc T_1 suit la même loi que $\cos\theta$ avec $\theta \sim \mathcal{U}([0,2\pi])$

Scanné avec CamScanner

On montre exactement de la même manière que T_2 et V sont indépendants puis avec le changement de variable $\Theta = \sin(t)$ que T_2 suit la même loi que Θ avec $\Theta \sim \mathcal{U}([0, 2\pi])$.

c) On a $S = \sqrt{-2\ln(V)}$ avec $V \sim \mathcal{U}([0,1])$ donc S suit une loi de Rayleigh et $X = ST_1 = S\cos(\Theta)$ et $Y = ST_2 = S\sin(\Theta)$ avec $\Theta \sim \mathcal{U}([0,2\pi])$
Donc par la question 1, $(X,Y) \sim \mathcal{N}\left(\begin{pmatrix}0\\0\end{pmatrix}, I_2\right)$.

d) La probabilité de sortir de la boucle à chaque itération est donnée par $P(V_1^2+V_2^2 > 1) = \frac{\pi}{4}$ (Aire d'un quart de cercle unité). Ainsi le nombre d'itération suit une loi géométrique de paramètre $\frac{\pi}{4}$.
On a donc en moyenne $\frac{4}{\pi} \simeq 1,27$ itérations.

Exercice 2:

1) Soit $Q = \{\frac{1}{m}, m \in \mathbb{N}^*\}$. Alors si $X_m \in Q$, tel que $X_m = \frac{1}{m}$, on note B_m la variable de Bernoulli de paramètre $\frac{1}{m^2}$. Donc pour h une fonction continue bornée,

$$\mathbb{E}[h(X_{m+1}|X_m)] = \mathbb{E}[h(X_{m+1}|X_m, B_m = 0)]P(B_m = 0) + \mathbb{E}[h(X_{m+1}|X_m, B_m = 1)]P(B_m = 1)$$
$$= h\left(\frac{1}{m+1}\right)\cdot\left(1 - \frac{1}{m^2}\right) + \frac{1}{m^2}\int_0^1 h(y)\,dy$$

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

Si $X_m \notin Q$ on a immédiatement $X_{m+1} \sim \mathcal{U}([0,1])$ donc $P(x,A) = \int_{A\cap[0,1]} dt$.

On trouve donc bien que P est le noyau de transition.

Scanné avec CamScanner

2·) Soit μ une mesure uniforme sur $[0,1]$. Alors:

- par définition $\mu(A) = \int_{A \cap [0,1]} dy$

- $\int_0^1 \mu(dx) P(x,A) = \int_0^1 P(x,A)\,dx = \int_0^1 \int_{A\cap[0,1]} dt\,dx = \int_{A\cap[0,1]} dt$

car $\mathbb{Q}$ est de mesure nulle par rapport à la mesure de Lebesgue sur $[0,1]$.

- Finalement on a bien $\mu P = \mu$, ie μ est invariant par P.

3·) Soit $x \notin \mathbb{Q}$, alors:

$Pf(x) = \mathbb{E}[f(X_1) | X_0 = x] = \int f(y) P(x,dy) = \int f(y)\,\pi(y)\,dy = \int_0^1 f(y)\,dy$.

On a $P^2 f(x) = \int_0^1 Pf(t)\,dt$

$$= \int_{[0,1]\setminus\mathbb{Q}} Pf(t)\,dt + \underbrace{\int_{\mathbb{Q}} Pf(t)\,dt}_{=0}$$

$$= \int_0^1 \int_0^1 f(y)\,dy\,dt$$

$$= \int_0^1 f(y)\,dy.$$

Par récurrence on en déduit que pour tout $n \geq 1$, $P^n f(x) = \int_0^1 f(y)\,dy$

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

4·)