Computational Statistique TP1
Samuel Bazaz Jazayeri

I 1) Soit $h : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ bornée,

$$\mathbb{E}_{X,Y}[h(X,Y)] = \mathbb{E}_{R,\Theta}[(R\cos\Theta, R\sin\Theta)] = \int_{\mathbb{R}^2} h(r\cos\theta, r\sin\theta)\, f_R(r)\, 1_{r\geq 0}\, f_\Theta(\theta)\, 1_{\theta\in[0,2\pi]}\, d\theta\, dr$$

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

Posant $\Phi : \begin{cases} \mathbb{R}_+ \times [0,2\pi] \to \mathbb{R}\times\mathbb{R} \\ (r,\theta) \mapsto (r\cos\theta, r\sin\theta) \end{cases}$, comme $|\mathrm{Jac}(\Phi^{-1}(u))| = \left|\begin{pmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{pmatrix}\right|^{-1} = \frac{1}{r\cos\theta^2 + r\sin\theta^2} = \frac{1}{r}$,

par changement de variables

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

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

On identifie des densités de $\mathcal{N}(0,1)$ séparées en x et y, donc par proposition des fonctions muettes :

$$\boxed{\begin{array}{l} X \sim \mathcal{N}(0,1) \\ Y \sim \mathcal{N}(0,1) \\ X \perp\!\!\!\perp Y \end{array}}$$

2) Soit F f.d.r de R et $r > 0$:

$$F(r) = \int_0^r u e^{-\frac{u^2}{2}} du = \left[-e^{-\frac{u^2}{2}}\right]_0^r = 1 - e^{-\frac{r^2}{2}}$$

Soit $u \in [0,1[$:

$$\begin{aligned} F(r) = u &\iff e^{-\frac{r^2}{2}} = 1-u \\ &\iff r^2 = -2\log(1-u) \\ &\iff r = \sqrt{2\log\left(\frac{1}{1-u}\right)} \end{aligned}$$

Donc $F :]0,+\infty[\to [0,1[$ est une bijection et $F^{-1}(u) = \sqrt{2\log\left(\frac{1}{1-u}\right)}$

Ainsi, pour U_1, U_2 iid de loi $\mathcal{U}_{[0,1]}$ et posant :

$$\boxed{\begin{cases} g_R(u) = F^{-1}(u) \\ g_\Theta(u) = 2\pi . u \end{cases}}$$

On a : $\begin{cases} g_R(U_1) \perp\!\!\!\perp g_\theta(U_2) \\ g_R(U_1) \overset{(d)}{=} R \\ g_\theta(U_2) \overset{(d)}{=} \theta \end{cases}$

On peut donc simuler (X,Y) par l'algorithme suivant :

```
Sample U1, U2 indépendant r.v with distribution U[0,1]
Set R = gR(U1)
Set θ = gθ(U2)
Set X = R cos θ
Set Y = R sin θ
return (X,Y)
```

3)

a) Par définition $(U_1, U_2) \sim \mathcal{U}_{[0,1]^2}$

Ensuite par lemme des coalitions $2U_1 - 1 \perp\!\!\!\perp 2U_2 - 1$ puis

$$(2U_1 - 1, 2U_2 - 1) \sim \mathcal{U}_{[-1,1]^2}$$

Soit $(C_i)_{1 \le i \le n}$ iid de loi $\mathcal{U}_{[-1,1]^2}$,

Notant $N = \inf\{n \geq 1, C_n \in B_{\|\|_2}(0,1)\}$, On a à la fin de la boucle while :

$$(V_1, V_2) = C_N$$

Il s'agit d'une méthode de rejet et comme C_n est uniforme, la densité de C_N est donnée par :

$$(v_1, v_2) \mapsto \frac{\mathbb{1}_{(v_1,v_2) \in B_{\|\|_2}(0,1)}}{\int_{B_{\|\|_2}(0,1)} d(v_1,v_2)} = \frac{\mathbb{1}_{(v_1,v_2) \in B_{\|\|_2}(0,1)}}{\pi}$$

On a donc $(V_1, V_2) \in \mathcal{U}(B_{\|\|_2}(0,1))$ de densité

$$\boxed{f_{V_1,V_2}(v_1, v_2) = \frac{1}{\pi} \mathbb{1}_{(v_1,v_2) \in B_{\|\|_2}(0,1)}}$$

b) Soit h mesurable bornée, par le théorème transfert

$$\mathbb{E}[h(T_1, T_2, V)] = \int_{B_{\|\|_2}(0,1)} h\left[\frac{v_1}{\sqrt{v_1^2+v_2^2}}, \frac{v_2}{\sqrt{v_1^2+v_2^2}}, v_1^2+v_2^2\right] \frac{1}{\pi} d(v_1, v_2)$$

D'après un résultat de trigonométrie :

$$\forall (t_1,t_2) \in \mathbb{R}^2,\ t_1^2+t_2^2 = 1 \Leftrightarrow \exists! \theta \in [0, 2\pi[,\ \begin{cases} t_1 = \cos\theta \\ t_2 = \sin\theta \end{cases}$$

Posant

$$\Phi : \begin{array}{l} B(0,1) \to B(0,1) \\ (v_1, v_2) \mapsto \left(\frac{v_1}{\sqrt{v_1^2+v_2^2}}, \frac{v_2}{\sqrt{v_1^2+v_2^2}} \right) \end{array}$$

$\forall (v_1, v_2) \in B(0,1)$, $\Phi(v_1, v_2) = (\cos\theta, \sin\theta)$ avec $\theta = \begin{cases} \arccos\left(\frac{v_1}{\sqrt{v_1^2+v_2^2}}\right) & \text{si } v_1 \geq 0,\ v_2 \geq 0 \\ \frac{\pi}{2} + \arccos\left(\frac{v_1}{\sqrt{v_1^2+v_2^2}}\right) & \text{si } v_1 \leq 0,\ v_2 \geq 0 \\ \pi + \arccos\left(\frac{v_1}{\sqrt{v_1^2+v_2^2}}\right) & \text{si } v_1 \leq 0,\ v_2 \leq 0 \\ \frac{3\pi}{2} + \arccos\left(\frac{v_1}{\sqrt{v_1^2+v_2^2}}\right) & \text{si } v_1 \geq 0,\ v_2 \leq 0 \end{cases}$

D'après l'expression de θ, ce dernier est unique et parcourt $[0, 2\pi[$ quand (v_1, v_2) parcourt $S(0,1)$

Soit $\Phi_2(v, \theta) = (\sqrt{v}\cos\theta, \sqrt{v}\sin\theta)$ avec $v = v_1^2 + v_2^2$

Φ_2 définit bien une bijection sur $B(0,1)$ et :

$$|\operatorname{Jac} \Phi_2| = \begin{vmatrix} \frac{1}{2\sqrt{v}}\cos\theta & -\sqrt{v}\sin\theta \\ \frac{1}{2\sqrt{v}}\sin\theta & \sqrt{v}\cos\theta \end{vmatrix} = \frac{1}{2}\cos\theta^2 + \frac{1}{2}\sin\theta^2 = \frac{1}{2}$$

d'où :

$$\mathbb{E}[h(T_1, T_2, V)] = \int_{[0,1]} \int_{[0,2\pi]} h(\cos\theta, \sin\theta, v) \frac{1}{\pi} \frac{1}{2} \, dv \, d\theta$$

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

On a montré l'existence de θ tel que

$$\boxed{\begin{array}{l} \theta \sim \mathcal{U}_{[0,2\pi]} \\ \theta \perp\!\!\!\perp V = V_1^2 + V_2^2 \\ T_1 = \cos\theta \\ T_2 = \sin\theta \end{array}}$$

Aussi $\underline{V \sim \mathcal{U}_{[0,1]}}$

c) On a $S = g_R(U)$ avec $U = 1 - V \sim \mathcal{U}_{[0,1]}$ et g_R définit en $\underline{2}$

Donc S suit une distribution de Rayley et par $\underline{1}$:

$$\boxed{(X, Y) \sim \mathcal{N}(0, I_2)}$$

d)

Soit T le nombre d'étapes avant une sortie de boucle.
Comme les V_1, V_2 sont définis à chaque itération de manière indépendante
$T \sim \mathcal{G}(p)$ où p est la probabilité de sortir.

Or :

$$p = \mathbb{P}(V_1^2 + V_2^2 \leq 1)$$
$$= \mathbb{E}_{\mathcal{U}_{[-1,1]^2}}[\mathbb{1}_{V_1^2+V_2^2 \leq 1}]$$
$$= \int_{B(0,1)} \frac{1}{4} \, d(v_1, v_2)$$
$$= \frac{\pi}{4}$$

Donc le nombre moyen d'itérations de la boucle while est donnée par :

$$\boxed{\mathbb{E}[T] = \frac{1}{p} = \frac{4}{\pi}}$$

II)

1)

Par définition :

$$P(x_n, x_{n+1}) = \begin{cases} (1-x_n^2)\, \mathbb{1}_{x_{n+1} = \frac{1}{m+1}} + x_n^2\, \mathbb{1}_{x_{n+1} \in [0,1]} & \text{si } x_n = \frac{1}{m} \\ \mathbb{1}_{x_{n+1} \in [0,1]} & \text{sinon} \end{cases}$$

$$\text{ie } P(x, t) = \begin{cases} x^2\, \mathbb{1}_{t \in [0,1]} + (1-x^2)\, \delta_{\frac{1}{m+1}}(t) & \text{si } x = \frac{1}{m} \\ \mathbb{1}_{t \in [0,1]} & \end{cases}$$

Puis par intégration de t sur A

$$\boxed{P(x, A) = \begin{cases} x \int_{A \cap [0,1]} dt + (1-x^2)\, \delta_{\frac{1}{m+1}}(A) & \text{si } x = \frac{1}{m} \\ \int_{A \cap [0,1]} dt & \text{sinon} \end{cases}}$$

2) Soit $A \subset \mathbb{R}$,

$$\tilde{\pi} P(A) = \int_{\mathbb{R}} \tilde{\pi}(dx)\, P(x, A)$$
$$= \int_{[0,1]} P(x, A)\, dx$$
$$= \int_{\bigcup_{m \in \mathbb{N}} \{\frac{1}{m}\}} P(\tfrac{1}{m}, A)\, dx + \int_{[0,1] \setminus \bigcup_{m \in \mathbb{N}} \{\frac{1}{m}\}} P(x, A)\, dx$$

Comme $\lambda(\bigcup_{m \in \mathbb{N}} \{\frac{1}{m}\}) = \sum_{m \in \mathbb{N}} \lambda(\{\frac{1}{m}\}) = 0$ par σ-additivité :

$$\int_{[0,1] \setminus \bigcup_{m \in \mathbb{N}} \{\frac{1}{m}\}} \lambda(A \cap [0,1])\, dx = \lambda(A \cap [0,1]) = \tilde{\pi}(A)$$

Aussi :

$$\forall x \in \mathbb{R}, \quad 0 \leq P(x,A) \leq 1$$

donc :

$$0 \leq \int_{\bigcup_{n\in\mathbb{N}^*}\{\frac{1}{n}\}} P(\tfrac{1}{n},A)\,dx \leq \lambda\Big(\bigcup_{n\in\mathbb{N}^*}\{\tfrac{1}{n}\}\Big) = 0$$

Puis :

$$\tilde{\pi} P(A) = 0 + \tilde{\pi}(A)$$

On a montré l'invariance de $\tilde{\pi}$ pour P :

$$\boxed{\tilde{\pi} P = \tilde{\pi}}$$

3) Soit f mesurable bornée, $x \notin \bigcup_{n\in\mathbb{N}^*}\{\frac{1}{n}\}$

$$Pf(x) = \int_{\mathbb{R}} f(y)\, P(x,dy)$$

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

$$= \int_{\mathbb{R}} f(y)\, \tilde{\pi}(dy)$$

Ensuite pour $A \subset \mathbb{R}$

- $$P^2(x,A) = \int_{\mathbb{R}} P(x,dy)\, P(y,A)$$

$$= \int_{[0,1]} P(y,A)\, dy$$

$$= \tilde{\pi} P(A)$$

- Si pour $n \geq 2$, $P^n(x,A) = \tilde{\pi} P(A)$

$$P^{n+1}(x,A) = \int_{\mathbb{R}} P^n(x,dy)\, P(y,A)$$

$$= \int_{[0,1]} P(y,A)\, dy$$

$$= \tilde{\pi} P(A)$$

Ainsi par récurrence :

$$\forall n \geq 2, \quad P^n f(x) = \int_{\mathbb{R}} f(y)\, P^n(x,dy)$$

$$= \int_{\mathbb{R}} f(y)\, \tilde{\pi} P(dy)$$

Or comme $\tilde{\pi}$ est invariante :

$$\tilde{\pi}P(dy) = \tilde{\pi}(dy)$$

D'où :

$$\boxed{\forall n \geq 1,\ P^n f(x) = \int_{\mathbb{R}} f(y)\, \tilde{\pi}(dy)}$$

Comme la suite est constante, on a une forme d'ergodicité pour $x \notin \{\frac{1}{m}, m \in \mathbb{N}\}$:

$$\lim_{n \to +\infty} P^n f(x) = \int_{\mathbb{R}} f(y)\, \tilde{\pi}(dy)$$

4)

a)

$$\forall m \in \mathbb{N}^*,\ P\left(\frac{1}{m}, \frac{1}{m+1}\right) = \frac{1}{m^2} \underbrace{\tilde{\pi}\left(\left\{\frac{1}{m+1}\right\}\right)}_{0} + \left(1 - \frac{1}{m^2}\right) \underbrace{\delta_{\frac{1}{m+1}}\left(\frac{1}{m+1}\right)}_{1}$$

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

$$\forall m \in \mathbb{N}^*,\ P^2\left(\frac{1}{m}, \frac{1}{m+2}\right) = \int_{\mathbb{R}} P\left(\frac{1}{m}, dy\right) P\left(y, \frac{1}{m+2}\right)$$

$$= \int_{\mathbb{R}} P\left(y, \frac{1}{m+2}\right) \left(\frac{1}{m^2} \tilde{\pi}(dy) + \left(1 - \frac{1}{m^2}\right) \delta_{\frac{1}{m+1}}(dy)\right)$$

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

$$= \frac{1}{m^2} \underbrace{\tilde{\pi}\left(\left\{\frac{1}{m+2}\right\}\right)}_{0} + \left(1 - \frac{1}{m^2}\right)\left(1 - \frac{1}{(m+1)^2}\right) \quad \text{car } m+1 \in \mathbb{N}^*$$

o si pour $n \in \mathbb{N}^*$ fixé ; $\forall m \in \mathbb{N}^*$, $P^n\left(\frac{1}{m}, \frac{1}{m+n}\right) = \prod_{k=0}^{n} \left(1 - \frac{1}{(m+k)^2}\right)$

$$P^{n+1}\left(\frac{1}{m}, \frac{1}{m+n+1}\right) = \int_{\mathbb{R}} P\left(\frac{1}{m}, dy\right) P^n\left(y, \frac{1}{m+n+1}\right)$$

$$= \int_{\mathbb{R}} P^n\left(y, \frac{1}{m+n+1}\right)\left(\frac{1}{m^2} \tilde{\pi}(dy) + \left(1 - \frac{1}{m^2}\right) \delta_{\frac{1}{m+1}}(dy)\right)$$

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

$$= \frac{1}{m^2} \underbrace{\tilde{\pi}\left(\left\{\frac{1}{m+n+1}\right\}\right)}_{0} + \left(1 - \frac{1}{m^2}\right) \prod_{k=0}^{n} \left(1 - \frac{1}{m+k+1}\right) \quad \text{car } \tilde{\pi} \text{ invariant et } m+1 \in \mathbb{N}^*$$

Par récurrence :

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

b) Soit $A = \bigcup_{q \in \mathbb{N}} \left\{ \frac{1}{m+1+q} \right\}$

- par σ-additivité :

$$\pi(A) = \sum_{q \in \mathbb{N}} \lambda\left(\left\{ \cancel{\frac{1}{m+1+q}}^{\,0} \right\}\right)$$

$$= 0$$

- pour $n \in \mathbb{N}^*$, $\frac{1}{n+m} \in A$ en prenant $q = n-1 \in \mathbb{N}$

donc par croissance de la mesure positive :

$$\hat{P}\left(\tfrac{1}{m}, A\right) \geq \hat{P}\left(\tfrac{1}{m} \cdot \tfrac{1}{n+m}\right)$$

Par $\underline{b}$:

$$\lim_{n \to +\infty} \hat{P}\left(\tfrac{1}{m}, \tfrac{1}{n+m}\right) = \lim_{n \to +\infty} \prod_{k=0}^{n} \left(1 - \tfrac{1}{(m+k)^2}\right)$$

On a :

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

Par développement limité

$$\log\left(1 - \tfrac{1}{(m+k)^2}\right) \underset{k \to +\infty}{\sim} -\tfrac{1}{(m+k)^2} = O\left(\tfrac{1}{k^2}\right)$$

Par comparaison à une série de Riemann convergente :

$$\sum_{k \geq 0} \log\left(1 - \tfrac{1}{(m+k)^2}\right) \quad \text{converge}$$

Donc :

$$\lim_{n \to +\infty} \sum_{k=0}^{n} \log\left(1 - \tfrac{1}{(m+k)^2}\right) \in \mathbb{R}$$

Puis par continuité :

$$\lim_{n \to +\infty} \prod_{k=0}^{n} \left(1 - \tfrac{1}{(m+k)^2}\right) = e^{\sum_{k=0}^{\infty} \log\left(1 - \frac{1}{(m+k)^2}\right)} > 0$$

Par minoration on en déduit la non ergodicité de notre markovien :

$$\lim_{n \to +\infty} P^n(x, A) > 0 \quad \text{d'où} \quad \boxed{\lim_{n \to +\infty} P^n(x, A) \neq \pi(A)}$$

III / 1)

A défault de calculer l'erreur de généralisation, le problème de regression consiste à minimiser le risque empirique.

Ie: nous étudions ce risque d'un point de vue stochastique, d'où le problème suivant :

$$\min_{\Theta \in \Theta_c} \mathbb{E}[R_n(\Theta)]$$

Dans notre cas, nous étudions une régression linéaire de paramètres : $w \in \mathbb{R}^d$, $\tau \in \mathbb{R}$ (biais)

$$\boxed{\min_{\substack{w \in \mathbb{R}^d \\ \tau \in \mathbb{R}}} \mathbb{E}_{x,y}\left[\frac{1}{n}\sum_{i=1}^{n}(y_i - w^T x_i - \tau)^2\right]}$$

Notre problème est ouvert, il ne sera donc pas nécessaire de réaliser des projections pour se maintenir dans le domaine de faisabilité.

Notant $j(w, \tau, (y_i)_i, (x_i)_i) = \frac{1}{n}\sum_{i=1}^{n}(y_i - w^T x_i - \tau)^2$

$$j(w, \tau, (y_i)_i, (x_i)_i) = \frac{1}{n}\|\vec{y} - Xw - \tau . 1_n\|_2^2 \quad \text{où } X = \begin{bmatrix} x_1^T \\ \vdots \\ x_n^T \end{bmatrix},\ 1_n = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$$

$$= \frac{1}{n}\left(n\tau^2 + \vec{y}^T\vec{y} + w^T X^T X w - 2\vec{y}^T X w + 2\tau (Xw - \vec{y})^T 1_n\right)$$

$$\nabla_{w,\tau} j(w, \tau, \vec{y}, X) = \frac{1}{n}\begin{pmatrix} 2(X^T X w - \vec{y} + \tau . X^T 1_n) \\ \hline 2(n\tau + (Xw - \vec{y})^T 1_n) \end{pmatrix}$$

$$\boxed{\nabla_{w,\tau} j(w, \tau, \vec{y}, X) = \frac{2}{n}\begin{pmatrix} X^T X w - \vec{y} + \tau . X^T 1_n \\ \hline n\tau + (Xw - \vec{y})^T 1_n \end{pmatrix}}$$

Notant $u = (w, \tau)$ et $\mathcal{E} = (X, \vec{y})$

La descente de gradient stochastique sera implémentée de la manière suivante :

- On tire aléatoirement $u_0 = (w_0, \tau_0)$, $\begin{cases} w_0 \in \mathbb{R}^d \\ \tau_0 \in \mathbb{R} \end{cases}$ et z_k
- Tant que $k < k_{max}$:
  - On tire z_{k+1} de loi $\mathbb{P}_{x,y}$ tel que $z_{k+1} \perp\!\!\!\perp z_i$, $i \in [0, k]$
  - On calcule $\nabla_u j(u_k, z_{k+1})$
  - $\varepsilon_k = \frac{1}{\sqrt{k}}$
  - $u_{k+1} = u_k - \varepsilon \nabla_u j(u_k, z_{k+1})$
- On retourne u_k

On pourra simuler les z_k à partir de tirages uniformes indépendants.

Aussi, pour simplifier le calcul on ajoute une colonne de 1 dans $\mathbb{X}$ afin d'obtenir le biais, à partir d'une régression sans biais :

$$\boxed{\begin{cases} \tilde{w} = \begin{bmatrix} \underline{w} \\ b \end{bmatrix} \\ \tilde{\mathbb{X}} = \begin{bmatrix} \mathbb{1}_n \,|\, \mathbb{X} \end{bmatrix} \end{cases} \quad \text{et} \quad \nabla_{\tilde{w}}\, j(\tilde{w}, \vec{y}, \tilde{\mathbb{X}}) = \frac{2}{n}\left(\tilde{\mathbb{X}}^T \tilde{\mathbb{X}} w - \vec{y}\right)}$$