Bastien LE CHENADEC
bastien.le-chenadec@eleves.enpc.fr

TP 1
Computational statistics

3 novembre 2023

Exercices

Exercice 1 Box-Muller and Marsaglia-Bray algorithm

1. On se donne une fonction h qu'on suppose continue et bornée. Si on peut montrer (1) avec f_X et f_Y des densités gaussiennes centrées réduites, alors on aura montré que X et Y sont indépendantes et suivent une loi normale centrée réduite par le théorème d'identification des lois.

$$\forall h \quad \mathbb{E}[h(X, Y)] = \int_{\mathbb{R}^2} h(x, y) f_X(x) f_Y(y) dxdy \tag{1}$$

Soit h continue et bornée.

$$\begin{aligned}
\mathbb{E}[h(R\cos\Theta, R\sin\Theta)] &= \int_{\mathbb{R}^2} h(r\cos\theta, r\sin\theta) f_{R,\Theta}(r,\theta) drd\theta \\
&= \int_{\mathbb{R}^+}\int_0^{2\pi} \frac{h(r\cos\theta, r\sin\theta)}{2\pi} r\exp\left(\frac{-r^2}{2}\right) drd\theta \qquad = (*)
\end{aligned}$$

Soit $\Psi : (r,\theta) \mapsto (x,y)$. Ψ est un $\mathcal{C}^1$-difféomorphisme (bijection différentiable) de $\mathbb{R}^+ \times [0, 2\pi]$ dans $\mathbb{R}^2$. On a :

$$\begin{aligned}
\nabla\Psi &= \begin{pmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta} \end{pmatrix} \\
&= \begin{pmatrix} \cos\theta & -r\sin\theta \\ sin\theta & r\cos\theta \end{pmatrix}
\end{aligned}$$

De plus $\det(\nabla\Psi) = r\cos^2\theta + r\sin^2\theta = r$. Par le théorème de changement de variable :

$$\begin{aligned}
(*) &= \int_{\mathbb{R}^2} h(x,y)\exp\left(\frac{-(x^2+y^2)}{2}\right)\frac{1}{2\pi} dxdy \\
&= \int_{\mathbb{R}^2} h(x,y)\frac{\exp(-x^2/2)}{\sqrt{2\pi}}\frac{\exp(-y^2/2)}{\sqrt{2\pi}} dxdy
\end{aligned}$$

On a montré (1), donc $\boxed{X, Y \sim \mathcal{N}(0,1) \text{ et sont indépendantes}}$.

2. Écrire un algorithme pour échantillonner (sampler) deux gaussiennes indépendantes de loi $\mathcal{N}(0,1)$.

Rappel (Théorème de la transformée inverse)

> Soit $U \sim \mathcal{U}([0,1])$ et soit X une variable aléatoire qui a pour fonction de répartition F telle que $F(x) = \int_{-\infty}^{x} f_X(x)dx$. Alors $F^{-1}(U) \sim X$.

La fonction de répartition de la loi de Rayleigh de paramètre 1 est :

$$\begin{aligned} F_R(r) &= \int_0^r t\exp(-t^2/2)dt \\ &= \left[-\exp(-t^2/2)\right]_0^r \\ &= 1-\exp(-r^2/2) \end{aligned}$$

Par le théorème de la transformée inverse, $F_R^{-1}(U) \sim R(1)$. $F_R : \mathbb{R} \to [0,1]$ est monotone et continue sur un intervalle fermé. Soit $y \in [0,1]$ tel que $y = F_R(r)$.

$$\begin{aligned} y = F_R(r) &\Leftrightarrow y = 1-\exp(-r^2/2) \\ &\Leftrightarrow \exp(-r^2/2) = 1-y \\ &\Leftrightarrow -r^2/2 = \ln(1-y) \\ &\Leftrightarrow r^2 = -2\ln(1-y) \\ &\Leftrightarrow \boxed{r = \sqrt{-2\ln(1-y)}} \qquad\qquad \text{car } y \in [0,1]. \end{aligned}$$

On en déduit l'algorithme suivant :

Algorithme :

> 1. Tirer U_1, U_2 de loi uniforme sur $[0,1]$.
> 2. Prendre $R = -\sqrt{2\log(1-U_1)}$ et $\Theta = 2\pi U_2$. R suit une loi de Rayleigh de paramètre 1 et Θ suit une loi uniforme sur $[0,2\pi]$.
> 3. Prendre $(X,Y) = (R\cos(\Theta), R\sin(\Theta))$. D'après la question 1., X et Y sont indépendantes et suivent une loi $\mathcal{N}(0,1)$.

3. a) À la fin de la boucle (V_1, V_2) suivent la distribution uniforme de :

$$E = \left\{(V_1,V_2) | V_1^2 + V_2^2 \leq 1\right\}$$

(V_1, V_2) a pour densité :

$$\begin{aligned} f_{V_1,V_2}(v_1,v_2) &= \frac{1}{|E|}\mathbb{1}_{\{v_1^2+v_2^2\leq 1\}}(v_1,v_2)\mathbb{1}_{[-1,1]^2}(v_1,v_2) \\ &= \boxed{\tfrac{1}{\pi}\mathbb{1}_{\{v_1^2+v_2^2\leq 1\}}(v_1,v_2)} \end{aligned}$$

b) On utilise à nouveau le théorème d'identification des lois. Soit h continue et bornée. On veut montrer que :

$$\mathbb{E}[h(T_1,V)] = \int h(t_1,v)f_{T_1}(t_1)f_V(v)dvdt_1$$

avec f_{T_1} et f_V les densités qui nous intéressent.

Distribution de $\cos(\Theta)$:

Soit $\Theta \sim \mathcal{U}([0,2\pi])$. On note que $\cos([0,\pi]) = \cos([\pi,2\pi])$. Par conséquent travaillons avec $Y = \cos(\Theta/2)$ afin d'avoir une fonction monotone et inversible. On cherche la densité de Y.

$$\begin{aligned}
F_Y(y) &= P(Y \leq y) \\
&= P(\cos(\Theta/2) \leq y) \\
&= P(\Theta/2 \geq \arccos(y)) \\
&= P(\Theta/2 \leq \arccos(-y)) \\
&= \arccos(-y)/\pi \\
f_Y(y) &= \frac{d}{dy} F_Y(y) \\
&= \frac{1}{\pi}\frac{d}{dy}\arccos(-y) \\
&= \boxed{\tfrac{1}{\pi\sqrt{1-y^2}}}
\end{aligned}$$

Distribution de T_1 :

On note $\Psi : v_1, v_2 \mapsto (t_1, v)$. On note que ψ n'est pas injective. Calculons le jacobien de Ψ :

$$\begin{aligned}
\nabla\Psi &= \begin{pmatrix} \frac{\partial t_1}{\partial v_1} & \frac{\partial t_1}{\partial v_2} \\ \frac{\partial t_2}{\partial v_1} & \frac{\partial t_2}{\partial v_2} \end{pmatrix} \\
&= \begin{pmatrix} \frac{v_2^2}{(v_1^2+v_2^2)^{3/2}} & \frac{-v_1v_2}{(v_1^2+v_2^2)^{3/2}} \\ 2v_1 & 2v_2 \end{pmatrix}
\end{aligned}$$

On a donc le jacobien $\det \nabla\Psi = \frac{2v_2}{\sqrt{v_1^2+v_2^2}} = 2\sqrt{1-t_1^2}$. Par le théorème de changement de variable :

$$\begin{aligned}
&\int_{-1}^{1}\int_{-1}^{1} h(t_1(v_1,v_2), v(v_1,v_2))\frac{1}{\pi}\mathbb{1}_{\{v_1^2+v_2^2\leq 1\}}(v_1,v_2)dv_1dv_2 \\
&= 2\int_{-1}^{1}\int_{0}^{1} h(t_1,v)\mathbb{1}_{\{v\leq 1\}}\frac{1}{2\pi\sqrt{1-t_1^2}}dvdt_1 && \text{(le facteur 2 provient de} \\
& && \text{la non-injectivité de } \Psi).
\end{aligned}$$

On reconnaît bien la densité de $\cos(\Theta)$ et celle de $\mathcal{U}([0,1])$. Donc T_1 et V sont indépendantes, T_1 suit la même distribution que $\cos(\Theta)$ et V suit une loi uniforme sur $[0,1]$.

Distribution de T_2 :

De façon similaire, on peut cacluler la densité de $\sin(\Theta)$, puis on peut prouver que T_2 et V sont indépendantes et que T_2 suit la même distribution que $\sin(\Theta)$.

c) Distribution de (X, Y) ?

$S = F^{-1}(V)$ avec $V \sim \mathcal{U}([0,1])$ et F^{-1} l'inverse de la fonction de répartition d'une loi de Rayleigh de paramètre 1 (d'après la question 2.). Donc S suit une loi de Rayleigh de paramètre 1 d'après le théorème de la transformée inverse.

Comme les variables aléatoires T_1 et T_2 sont indépendantes de V, elles sont indépendantes de S. D'après la question 1., $\boxed{X \text{ et } Y \text{ sont indépendantes}}$ et $\boxed{X \text{ et } Y \sim \mathcal{N}(0,1)}$.

d) Le nombre moyen d'étapes dans la boucle revient à regarder la probabilité de se trouver dans un disque de rayon 1 sachant qu'on est uniformément dans $[-1,1]^2$. En faisant le ratio des surfaces, on trouve que cette probabilité est $\frac{\pi}{4}$.

On note C la variable aléatoire du nombre de coups au bout duquel on sort de la boucle. Elle suit une loi géométrique, pour $n \geq 1$:

$$P(C = n) = \frac{\pi}{4}\left(1 - \frac{\pi}{4}\right)^{n-1}$$

On calcule donc l'espérance d'une loi géométrique :

$$\begin{aligned}
\mathbb{E}[C] &= \sum_{n=1}^{+\infty} nP(C = n) \\
&= \sum_{n=1}^{+\infty} \frac{n\pi}{4}\left(1 - \frac{\pi}{4}\right)^{n-1} \\
&= \frac{\pi}{4}\frac{1}{(\frac{\pi}{4})^2} && \text{par (2).} \\
&= \boxed{\frac{4}{\pi}}
\end{aligned}$$

<u>Justification de (2) :</u>

Notons $f_n(q) = \sum_{k=0}^{n} q^k$. Pour $0 < q < 1$, on reconnaît une série géométrique de raison q :

$$f_n(q) = \frac{1 - q^{n+1}}{1 - q}$$

Dérivons par raport à la variable q :

$$\sum_{k=1}^{n} kq^{k-1} = \frac{-(n+1)q^n(1-q) + (1 - q^{n+1})}{(1-q)^2}$$

Par croissance comparée, il est clair qu'on a :

$$\lim_{n \to +\infty} \sum_{k=1}^{n} kq^{k-1} = \frac{1}{(1-q)^2} \tag{2}$$

Exercice 2 Invariant distribution

Rappel (Noyau de transition)

Soit $\mathcal{X}$ l'ensemble des états d'une chaîne de Markov, $P : (\mathcal{X} \times \mathcal{X}) \to \mathbb{R}$ est une matrice de transition si :

$$\forall (x,y) \in (\mathcal{X} \times \mathcal{X}), P(x,y) \geq 0 \quad \text{et} \quad \sum_y P(x,y) = 1 \quad \forall x \in \mathcal{X}$$

En particulier, $\forall x \in \mathcal{X}, Q(y) = P(x,y)$ est une loi de probabilités sur $\mathcal{X}$.

Par extension on définit le noyau de transition pour $A \subset \mathcal{X}$:

- $\forall x \in \mathcal{X}, A \mapsto P(x,A)$ est une loi de probabilité.
- $\forall A \subset \mathcal{X}, x \mapsto P(x,A)$ est mesurable.

Et on définit aussi l'application de P sur les fonctions ou mesures :

$$\begin{cases} Pf : & \mathcal{X} \to \mathbb{R}^+ \\ & x \mapsto \sum_{y \in \mathcal{X}} P(x,y) f(y) \end{cases}$$

$$\begin{cases} \mu P : & \mathcal{X} \to \mathbb{R}^+ \\ & y \mapsto \sum_{x \in \mathcal{X}} P(x,y) \mu(x) \end{cases}$$

1. On note $Q = \left\{\frac{1}{m}, m \in \mathbb{N}^*\right\}$. Le noyau de transition de (X_n) est la distribution de $X_{n+1}|X_n$. Si on peut montrer 3 pour toute fonction h continue et bornée, on aura identifié le noyau de transition de (X_n) par le théorème d'identification des lois.

$$\mathbb{E}[h(X_{n+1})|X_n] \sim \int_{\mathbb{R}} h(y) P(x, dy) \tag{3}$$

<u>Si $x \notin Q$:</u>

$$\begin{aligned} \mathbb{E}[h(X_{n+1})|X_n] &= \int_{\mathbb{R}} h(y) f_{X_{n+1}|X_n}(y) dy \\ &= \int_{\mathbb{R}} h(y) \mathbb{1}_{\{[0,1]\}}(y) dy \qquad \text{par l'énoncé.} \end{aligned}$$

Donc $P(x,A) = \int_A P(x,dy) = \boxed{\int_{A \cap [0,1]} dy}$.

<u>Si $x \in Q$:</u>

On note $B_m \sim \mathcal{B}(\frac{1}{m^2})$.

$$\begin{aligned} \mathbb{E}[h(X_{n+1})|X_n] &= \mathbb{E}[h(X_{n+1})|X_n, B_m = 0] P(B_m = 0) + \mathbb{E}[h(X_{n+1})|X_n, B_m = 1] P(B_m = 1) \\ &= h\left(\frac{1}{m+1}\right)\left(1 - \frac{1}{m^2}\right) + \frac{1}{m^2} \int_{A \cap [0,1]} h(y) dy \end{aligned}$$

D'où :

$$
\begin{aligned}
P(x,A) &= \int_A P(x,dy) \\
&= \boxed{(1-x^2)\delta_{\frac{1}{m+1}}(A) + x^2 \int_{A\cap[0,1]} dy} \qquad \text{avec } x^2 = \frac{1}{m^2}
\end{aligned}
$$

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

2. On note $\pi \sim \mathcal{U}([0,1])$, montrons que $\pi P = \pi$.

Soit A une partie de $\mathbb{R}$.

$$
\begin{aligned}
\pi P(A) &= \int_{\mathbb{R}} P(x,A)\pi(x) \\
&= \int_{[0,1]} P(x,A)dx && \text{car } \pi \text{ uniforme.} \\
&= \int_{[0,1]\cap Q} P(x,A)dx + \int_{[0,1]\cap \bar{Q}} P(x,A)dx && \text{par la relation de Chasles.} \\
&= \int_{[0,1]\cap \bar{Q}} P(x,A)dx && \text{car } Q \text{ est de mesure nulle.} \\
&= \int_{[0,1]\cap A} dt && \text{par la question 1.} \\
&= \pi(A) && \text{car } \pi \text{ est uniforme sur } [0,1].
\end{aligned}
$$

On a bien $\boxed{\pi P = \pi}$.

3. Soit $x \notin Q$. Calculons $P^n f(x)$ pour tout $n \geq 1$ et $\lim_{n\to+\infty} P^n f(x)$ en fonction de $\int f(t)\pi(t)dt$. Nous procédons par récurrence.

Soit la propriété $\mathcal{P}(n)$: $P^n f(x) = \int_0^1 f(y)dy$.

<u>Initialisation :</u>

$$
Pf(x) = \mathbb{E}[f(X_1)|X_0 = x] = \int_0^1 f(y)dy \qquad \text{car } x \notin Q.
$$

Donc $\mathcal{P}(1)$ est vraie.

<u>Récurrence :</u>

Soit $n \in \mathbb{N}^*$, supposons $\mathcal{P}(n)$. Montrons $\mathcal{P}(n+1)$:

$$\begin{aligned}
P^{n+1}f(x) &= PP^n f(x) \\
&= \int_0^1 \left(\int_0^1 f(y)dy \right) dx && \text{par } \mathcal{P}(n) \text{ et car } x \notin Q. \\
&= \int_0^1 f(y)dy && \text{(intégration d'une constante).}
\end{aligned}$$

On a bien $\mathcal{P}(n+1)$.

<u>Conclusion :</u>

On a bien $\boxed{\mathcal{P}(n) \text{ pour tout } n \in \mathbb{N}^*}$.

On s'intéresse maintenant à $\lim_{n\to+\infty} P^n f(x)$:

$$\begin{aligned}
\lim_{n\to+\infty} P^n f(x) &= \lim_{n\to+\infty} \int_0^1 f(y)dy && \text{par ce qui précède.} \\
&= \int_0^1 f(y)dy && \text{car le terme est indépendant de } n. \\
&= Pf(x) \\
&= \pi Pf(x) && \text{par la question 2.} \\
&= \boxed{\int_0^1 f(y)\pi(y)dy}
\end{aligned}$$

4. a) Soit $x = \frac{1}{m}$ avec $m \geq 2$. Nous procédons par récurrence.

Soit la propriété $\mathcal{P}(n)$: $P^n(\frac{1}{m}, \frac{1}{n+m}) = \prod_{k=1}^n \left(1 - \frac{1}{(m+k-1)^2}\right)$.

<u>Initialisation :</u>

$$\begin{aligned}
P\left(\frac{1}{m}, \frac{1}{m+1}\right) &= \frac{1}{m^2} \int_{\{\frac{1}{m+1}\}\cap[0,1]} dt + \left(1 - \frac{1}{m^2}\right) \delta_{\frac{1}{m+1}} \left(\left\{\frac{1}{m+1}\right\}\right) && \text{par la question 1.} \\
&= 1 - \frac{1}{m^2}
\end{aligned}$$

Donc $\mathcal{P}(1)$ est vraie.

<u>Récurrence :</u>

Soit $n \in \mathbb{N}^*$, supposons $\mathcal{P}(n)$. Montrons $\mathcal{P}(n+1)$:

$$
\begin{aligned}
P^{n+1}\left(\frac{1}{m}, \frac{1}{m+n+1}\right) &= P\left(\frac{1}{m+n}, \frac{1}{m+n+1}\right) P^n\left(\frac{1}{m}, \frac{1}{m+n}\right) \\
&\quad + \left(1 - P^n\left(\frac{1}{m}, \frac{1}{m+n}\right)\right) \int_{\left\{\frac{1}{m+n+1}\right\}\cap[0,1]} dt \\
&= \left(1 - \frac{1}{(m+n)^2}\right) \prod_{k=1}^{n} \left(1 - \frac{1}{(m+k-1)^2}\right) \\
&\quad + 0 \\
&= \prod_{k=1}^{n+1} \left(1 - \frac{1}{(m+k-1)^2}\right)
\end{aligned}
$$

On a bien $\mathcal{P}(n+1)$.

Conclusion :

Pour tout $n \in \mathbb{N}^*$, $\boxed{P^n(\frac{1}{m}, \frac{1}{n+m}) = \prod_{k=1}^{n} \left(1 - \frac{1}{(m+k-1)^2}\right)}$.

On peut développer l'expression précédente :

$$
\begin{aligned}
P^n\left(\frac{1}{m}, \frac{1}{n+m}\right) &= \prod_{k=1}^{n} \left(1 - \frac{1}{(m+k-1)^2}\right) \\
&= \prod_{k=1}^{n} \left(\frac{(m+k-1)^2 - 1}{(m+k-1)^2}\right) \\
&= \prod_{k=1}^{n} \left(\frac{(m+k-2)(m+k)}{(m+k-1)^2}\right) \\
&= \prod_{k=1}^{n} \left(\frac{m+k-2}{m+k-1}\right) \prod_{k=1}^{n} \left(\frac{m+k}{m+k-1}\right) \\
&= \frac{m-1}{m+n-1} \times \frac{m+n}{m} \qquad \text{en simplifiant les termes qui s'annulent.} \\
&= \boxed{\frac{m-1}{m} \times \frac{m+n}{m+n-1}}
\end{aligned}
$$

b) Par la question précédente, il est clair que :

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

On note $A = \bigcup_{q\in\mathbb{N}} \left\{\frac{1}{m+q+1}\right\}$. Il est clair que pour des valeurs de m élevées, on a $\frac{m-1}{m} \approx 1$ et on est presque sûr de se trouver dans A asymptotiquement. En revanche, pour des valeurs de m faibles, on a des chances de sortir de A, et par les mêmes arguments que précédemment (la mesure de A est nulle), on n'a aucune chance de revenir dans A. On distingue donc deux cas :

- On sort de A avec probabilité $\frac{1}{m}$. Dans ce cas :

$$
\lim_{n\to+\infty} P^n(x, A) = \pi([0, 1])
$$

- On reste dans A avec probabilité $\frac{m-1}{m}$. Dans ce cas on n'a pas $\lim_{n \to +\infty} P^n(x, A) = \pi(A)$ puisqu'on a montré que les probabilités de transitions entre les états $\frac{1}{m+n}$ et $\frac{1}{m+n+1}$ tendent vers 1.

Donc d'une façon générale, $\lim_{n \to +\infty} P^n(x, A) \neq \pi(A)$.

Exercice 3 Stochastic Gradient Learning in Neural Networks

Les réponses aux questions sont dans le Jupyter Notebook.