# Algorithme de Métropolis-Hastings

Master MIGS/PMG 2025-2026

Yoann Offret

## 1 Problématique

Soit $E$ un espace d'état fini. On cherche à simuler une distribution de Gibbs $\pi$ sur $E$ de potentiel $V : E \longmapsto \mathbb{R}$. Cela en simulant une chaîne de Markov (CM). Une telle distribution $\pi$ est donnée, pour tout $x \in E$, par

$$\pi(x) := \frac{1}{Z} e^{-V(x)}.$$

La constante $Z$ vérifie donc

$$Z = \sum_{x \in E} e^{-V(x)}.$$

En général, $E$ est de très grande taille et $V$ n'est pas directement calculable, mais seulement certaines de ses variations $\Delta V$.

**Exercice 1.** *Montrer que deux potentiels $V_1$ et $V_2$ sont égaux à une constante près si et seulement si les mesures de Gibbs associées $\pi_1$ et $\pi_2$ sont égales.*

Il est donc très difficile en pratique de simuler cette loi. L'idée est de construire une chaîne de Markov apériodique $(X_n)_{n \geq 0}$ sur $E$ de mesure de probabilité invariante $\pi$. En particulier, on aura par le théorème de Perron-Frobenius et le théorème ergodique :

$$X_n \xRightarrow[n \to +\infty]{} \pi,$$

et

$$\frac{1}{n} \sum_{k=0}^{n-1} f(X_k) \xrightarrow[n \to +\infty]{p.s.} \int f d\pi = \sum_{x \in E} f(x) \pi(x).$$

## 2 Algorithme

**Données :** Supposons connue une chaîne de Markov irréductible sur $E$ facilement simulable. On note $Q$ sa matrice de transition.

**Principe :** L'idée est de construire à partir de $Q$ une autre chaîne de Markov, irréductible et apériodique, de mesure réversible $\pi$ qui sera facilement simulable. On notera $P$ sa matrice de transition. Lorsque $x \neq y$, on cherche $P$ sous la forme

$$P(x, y) = Q(x, y) \alpha(x, y),$$

avec $0 \leq \alpha(x, y) \leq 1$. Notons qu'il suffit de définir la matrice de transition pour $x \neq y$ car pour $x = y$, on a forcément

$$P(x, x) = 1 - \sum_{y \neq x} P(x, y).$$

**Interprétation :** Pour simuler une transition de $x$ vers $y$ avec le noyau $P$, il suffit donc de suivre les deux étapes suivantes :

1. On simule une transition de $x$ vers $y_{\mathrm{p}}$ avec le noyau $Q$.
2. On choisit $y = y_{\mathrm{p}}$ avec probabilité $\alpha(x, y_{\mathrm{p}})$ et $y = x$ sinon.

L'état $y$ résultant est alors choisi avec probabilité $P(x, y)$. On appelle souvent $Q$ le noyau de proposition, la proposition étant l'état intermédiaire $y_{\mathrm{p}}$ dans l'étape 1. L'étape 2 est la phase d'acceptation-rejet de cette proposition.

**Exercice 2.** *Montrer que $\pi$ est une mesure de probabilité invariante pour $P$ si l'on choisit*

$$\alpha(x,y) := \min\left(1, \frac{Q(y,x)\pi(y)}{Q(x,y)\pi(x)}\right).$$

*Montrer que la chaîne de matrice de transition $P$ est irréductible et apériodique si $\pi \neq \mu$.*

Le plus souvent, la matrice $Q$ est symétrique. De plus, on n'a pas besoin de connaître $\pi$ complètement. En effet, il suffit de connaître les rapports $\pi(y)/\pi(x)$ lorsque $x$ et $y$ sont voisins au sens où $Q(x,y) > 0$. Cela revient à connaître les variations $\Delta V$ du potentiel entre deux voisins puisque

$$\frac{\pi(y)}{\pi(x)} = e^{-(V(y)-V(x))}.$$

Surtout, la constante de normalisation $Z$ n'a pas besoin d'être connue.

# 3 Modèle d'Ising

Soit $\Gamma = [|1, N|]^2$ le réseau carré où l'ensemble des voisins d'un point $(i,j) \in \Gamma$ est donné par

$$\mathcal{V}_{(i,j)} = \{(i+1,j); (i-1,j); (i,j+1); (i,j-1)\} \subset \Gamma.$$

On identifie $N+1$ avec 1 et 0 avec $N$ (périodicité). L'espace d'état est ici $E = \{-1,1\}^\Gamma$. On pose ensuite pour tout $x \in E$ et $\beta > 0$,

$$V(x) = -\sum_{(i,j)\in\Gamma} \sum_{(k,l)\in\mathcal{V}_{(i,j)}} x_{(i,j)}x_{(k,l)},$$

puis

$$\pi_\beta = \frac{1}{Z_\beta} e^{-\beta V(x)} \quad \text{où} \quad Z_\beta = \sum_{x\in E} e^{-\beta V(x)}.$$

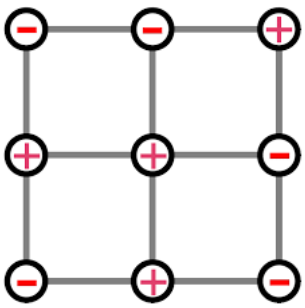


FIGURE 1 – Exemple de configuration $x$ avec $N = 3$. Que vaut $V(x)$ ici ?

**Exercice 3.** *Quel est le cardinal de $E$ ? Est-il raisonnable de calculer $Z_\beta$ ? Écrire un algorithme permettant de simuler $\pi_\beta$. On pose pour toute configuration $x \in E$,*

$$M(x) = \frac{1}{N^2} \sum_{(i,j)\in\Gamma} x_{i,j}.$$

*Tracer une approximation de la fonction*

$$\beta \longmapsto \sum_{x\in E} M(x)\pi_\beta(x).$$

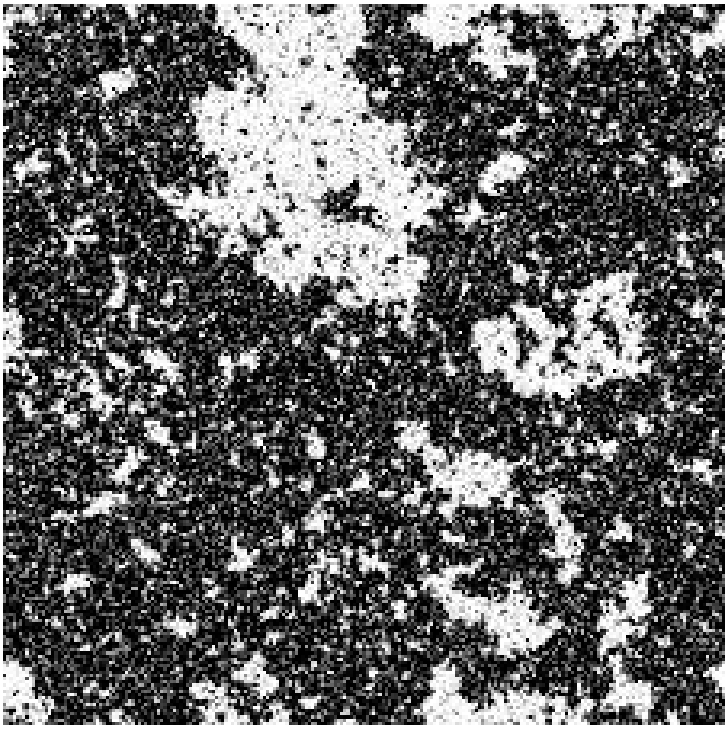


FIGURE 2 – Exemple de simulation de $\pi_\beta$ pour $N$ grand (noir=+, blanc=-)

# 4 Application en optimisation : Méthode du recuit simulé

On cherche à minimiser une fonction non constante $V$ sur $E$. Grâce à l'algorithme de Metropolis-Hastings, on sait, pour tout $T > 0$, simuler à partir d'une chaîne de Markov de matrice de transition $Q$, $\mu$-réversible, la mesure de Gibbs

$$\pi_T(x) := \frac{1}{Z_T} e^{-\frac{V(x)}{T}}, \quad \text{où} \quad Z_T := \sum_{x \in E} e^{-\frac{V(x)}{T}}.$$

Le paramètre $T$, souvent noté $1/\beta$, est pour des raisons physiques appelé la température. On rappelle que l'on accepte une transition de $x$ vers $y$ donnée par $Q(x, y)$ avec probabilité 1 quand $V(y) < V(x)$, et sinon avec une probabilité

$$P(x, y) := e^{-\frac{V(y)-V(x)}{T}}.$$

Ainsi, à haute température, on accepte plus volontiers un état de plus haute énergie $V$, tandis qu'à basse température, ces transitions sont très pénalisées.

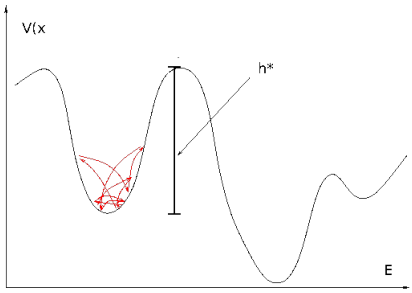
**Exercice 4.** *Montrer que $\pi_T$ converge en loi quand $T$ tend vers 0 vers la mesure uniforme sur l'ensemble des minima de $V$ sur $E$.*

Plus la température est basse, plus $\pi_T$ se concentre sur les minima de $V$. La méthode du recuit simulé tire son nom de la métallurgie. Le recuit est une opération consistant à laisser refroidir lentement un métal pour améliorer ses qualités. L'idée est qu'un refroidissement trop brutal peut figer le métal dans un état peu favorable (minimum local de $V$), tandis qu'un refroidissement assez lent permettra aux molécules de s'agencer au mieux dans une configuration stable (minimum global de $V$). Le principe est donc de faire varier la température $T$ en fonction du temps $n \geq 1$ dans l'algorithme de Metropolis-Hastings. On dispose du théorème suivant.

**Theorem 4.1** (Méthode de recuit simulé)**.** *Soit $h > 0$ et, pour tout $n \geq 1$,*

$$T_n^h := \frac{h}{\log(n+1)}.$$

*Il existe $h^* > 0$ tel que, pour tout $h > h^*$, la chaîne de Markov non homogène construite à l'aide de l'algorithme de Metropolis-Hastings en prenant à chaque étape $n$ la température $T_n^h$ converge en probabilité vers l'ensemble des minima de $V$.*




On peut montrer que $h^*$ représente la hauteur maximale des puits des minima locaux de $V$. En pratique, on doit régler le paramètre $h$ expérimentalement. Nous allons donner deux exemples d'applications parmi tant d'autres.

## 4.1 Déchiffrer un code secret

© Cet exemple est tiré du blog/chaîne YouTube de David Louapre, qui s'est librement inspiré de l'article

*Diaconis, P. (2009). The markov chain monte carlo revolution. Bulletin of the American Mathematical Society, 46(2), 179-205.*

Supposons que vous trouviez le message chiffré (codé) suivant :

FRNBRJMFBRNQXHFYNWSRNKRDNZLXDRDNMOOHCRFYNZXJJRNYSNKRDNB
RDHORDNJMHDNBRDHORNWSNRKKRDNMOOHCRFYNZXJJRNRKKRDNMOOH
CRFYNRYNYSNQOXDQROROMDNYXSVXSODNRQHZYRYR

Vous faites l'hypothèse que ce chiffrement a été fait par substitution de telle manière que chaque lettre et l'espace aient été substitués par une autre. Par exemple, l'espace " " a été remplacé par la lettre "E", la lettre "A" par la lettre "R", etc. Il y a donc $27! \simeq 10^{28}$ codes possibles. En supposant que vous puissiez tester un code par seconde (le temps de décoder puis surtout de le lire pour vérifier que le message est bien décodé, ce qui est généreux comme délai), il faudrait environ 345 milliards de milliards d'années pour tous les tester. Ainsi, la force brute n'est pas la meilleure technique...

Une autre méthode plus naïve est d'utiliser la fréquence des lettres dans la langue française. Par exemple, on sait que la fréquence de l'espace " " est d'environ 17.4% et que, parmi les lettres, la lettre "E" est la plus probable et apparaît environ 12.1% du temps. On peut estimer que la lettre la plus probable dans le message chiffré précédent est l'espace et la seconde est le "E" (ou inversement si l'on n'a pas de chance). Cependant, étant donné la faible longueur du message, rien n'est moins sûr. Et même si c'était le cas, cela ne nous donnerait pas les autres lettres (qui, pour certaines, ont des probabilités d'apparition très similaires).

Une méthode plus robuste est de considérer les fréquences des paires de lettres/espace (appelées bigrammes) dans la langue française. Par exemple, en analysant "Du côté de chez Swann" de Marcel Proust, on obtient les fréquences de la figure 4.

**Exercice 5.** *Proposer un algorithme basé sur les fréquences des bigrammes et le recuit simulé afin de décoder ce type de message chiffré.*

## 4.2 Problème du voyageur de commerce

On dispose des latitudes et longitudes des villes suivantes :

| | |
|---|---|
| "Bordeaux" | (44.833333,-0.566667)) |
| "Paris" | (48.8566969,2.3514616)) |
| "Nice" | (43.7009358,7.2683912)) |
| "Lyon" | (45.7578137,4.8320114)) |
| "Nantes" | (47.2186371,-1.5541362)) |
| "Brest" | (48.4,-4.483333)) |
| "Lille" | (50.633333,3.066667)) |
| "Clermont-Ferrand" | (45.783333,3.083333)) |
| "Strasbourg" | (48.583333,7.75)) |
| "Poitiers" | (46.583333,0.333333)) |
| "Angers" | (47.466667,-0.55)) |
| "Montpellier" | (43.6,3.883333)) |
| "Caen" | (49.183333,-0.35)) |
| "Rennes" | (48.083333,-1.683333)) |
| "Pau" | (43.3,-0.366667)) |

Vous démarrez de Rennes en hélicoptère et vous décidez de visiter chacune de ces villes puis de revenir à Rennes. Pour économiser du gazole vous aimeriez trouver le plus court chemin. Voir Figure 3.

**Exercice 6.** *Proposer un algorithme de type recuit simulé permettant de déterminer le plus court chemin.*

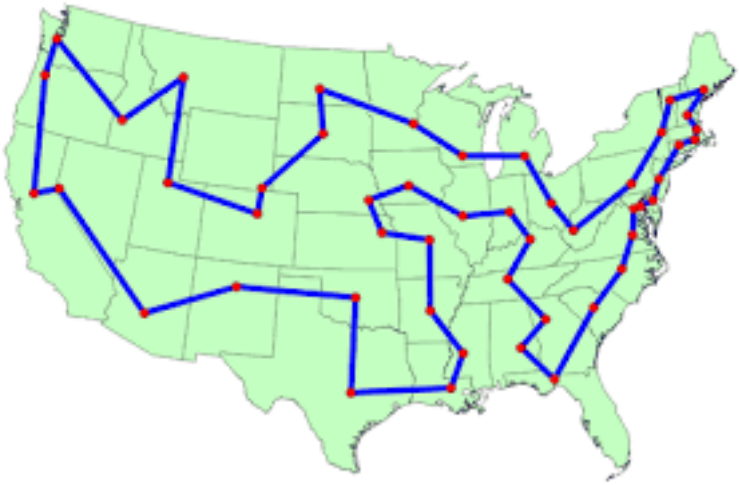


FIGURE 3 – Illustration du problème et du chemin optimal dans un autre contexte

# 5 Échantillonnage de Gibbs (Gibbs sampling)

On suppose ici que l'on veuille simuler une variable aléatoire $X = (X_1, \cdots, X_n)$ à valeurs dans un espace produit $E^n$. On fait l'hypothèse que la loi de $X$ est absolument continue par rapport à une mesure produit $\mu \otimes \cdots \otimes \mu$ sur $E$. Le plus souvent, $E$ est fini ou $E = \mathbb{R}$, auquel cas $\mu$ est la mesure de comptage ou la mesure de Lebesgue. On suppose également que les lois conditionnelles des marginales connaissant toutes les autres variables sont facilement simulables.

Plus précisément, on pose pour tout $1 \leq i \leq n$,

$$\widehat{X}_i = (X_1, \cdots, X_{i-1}, X_{i+1}, \cdots, X_n).$$

On note $\pi(dx_1 \cdots dx_n) = \pi(x_1, \cdots, x_n)\mu(dx_1) \cdots \mu(dx_n)$ la distribution de $X$ et $\pi_i(\hat{x}_i, dx_i)$ la loi conditionnelle de $X_i$ sachant que $\widehat{X}_i = \hat{x}_i$. On rappelle que cette dernière admet une densité conditionnelle donnée par

$$\pi_i((x_1, \cdots, x_{i-1}, x_{i+1}, \cdots, x_n); x_i) = \frac{\pi(x_1, \cdots, x_n)}{\int_E \pi(x_1, \cdots, x_n)\mu(dx_i)}.$$

On remarque également que $\pi$ a simplement besoin d'être connue à une constante près.

| | | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.0 | 9.1 | 1.5 | 7.0 | 11.6 | 9.7 | 2.6 | 1.0 | 0.7 | 3.4 | 2.1 | 0.0 | 9.0 | 5.3 | 2.9 | 2.2 | 8.1 | 5.3 | 2.7 | 7.0 | 2.8 | 2.7 | 2.7 | 0.0 | 0.0 | 0.4 | 0.0 |
| 15.6 | 0.0 | 1.6 | 2.3 | 0.8 | 0.0 | 0.2 | 1.8 | 0.2 | 26.1 | 0.2 | 0.0 | 3.4 | 2.5 | 14.7 | 0.0 | 2.0 | 0.4 | 7.2 | 4.5 | 3.3 | 6.5 | 6.2 | 0.0 | 0.0 | 0.6 | 0.0 |
| 0.2 | 8.3 | 0.1 | 0.0 | 0.1 | 15.2 | 0.0 | 0.0 | 0.0 | 17.0 | 0.9 | 0.0 | 25.6 | 0.0 | 0.0 | 12.9 | 0.0 | 0.0 | 14.3 | 3.2 | 0.4 | 1.7 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 8.2 | 6.8 | 0.0 | 1.3 | 0.0 | 25.3 | 0.0 | 0.0 | 15.6 | 4.9 | 0.0 | 0.0 | 2.1 | 0.0 | 0.0 | 24.8 | 0.0 | 0.1 | 3.9 | 0.1 | 3.2 | 3.7 | 0.0 | 0.0 | 0.0 | 0.1 | 0.0 |
| 18.1 | 8.5 | 0.0 | 0.0 | 0.0 | 48.5 | 0.0 | 0.0 | 0.0 | 9.2 | 0.0 | 0.0 | 0.0 | 0.2 | 0.0 | 5.2 | 0.0 | 0.0 | 3.1 | 0.7 | 0.0 | 6.4 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 39.7 | 0.9 | 0.1 | 2.1 | 0.3 | 1.3 | 0.5 | 0.6 | 0.1 | 0.7 | 0.2 | 0.0 | 4.6 | 3.5 | 10.8 | 0.1 | 1.0 | 0.1 | 7.7 | 11.4 | 7.8 | 4.5 | 1.1 | 0.0 | 0.5 | 0.0 | 0.6 |
| 1.8 | 24.9 | 0.0 | 0.0 | 0.0 | 13.8 | 9.6 | 0.0 | 0.0 | 13.8 | 0.0 | 0.0 | 5.1 | 0.0 | 0.0 | 15.7 | 0.0 | 0.0 | 9.5 | 0.5 | 0.1 | 5.2 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 1.9 | 10.8 | 0.0 | 0.0 | 0.0 | 33.7 | 0.0 | 0.2 | 0.1 | 7.7 | 0.0 | 0.0 | 3.4 | 0.2 | 9.7 | 4.6 | 0.0 | 0.0 | 17.2 | 0.1 | 1.9 | 8.5 | 0.0 | 0.0 | 0.0 | 0.1 | 0.0 |
| 3.0 | 28.9 | 0.0 | 0.0 | 0.0 | 42.5 | 0.0 | 0.0 | 0.0 | 6.5 | 0.0 | 0.0 | 0.0 | 0.1 | 0.0 | 13.5 | 0.0 | 0.0 | 1.8 | 0.0 | 0.0 | 2.7 | 0.0 | 0.0 | 0.8 | 0.0 | 0.0 |
| 13.5 | 0.9 | 0.6 | 1.5 | 1.2 | 10.5 | 0.8 | 1.5 | 0.0 | 0.0 | 0.0 | 0.0 | 9.8 | 2.3 | 9.9 | 3.1 | 0.2 | 1.0 | 7.3 | 14.2 | 19.8 | 0.0 | 1.5 | 0.0 | 0.3 | 0.0 | 0.1 |
| 17.1 | 12.9 | 0.0 | 0.0 | 0.0 | 39.3 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 23.3 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 7.3 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 8.7 | 8.7 | 0.0 | 0.0 | 0.0 | 34.8 | 0.0 | 0.0 | 0.0 | 21.7 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 4.3 | 0.0 | 0.0 | 4.3 | 4.3 | 0.0 | 4.3 | 0.0 | 0.0 | 0.0 | 8.7 | 0.0 |
| 18.3 | 16.1 | 0.3 | 0.1 | 0.0 | 33.0 | 0.0 | 0.2 | 0.1 | 4.9 | 0.0 | 0.0 | 13.1 | 0.1 | 0.0 | 3.1 | 0.1 | 1.0 | 0.0 | 1.3 | 0.3 | 7.8 | 0.0 | 0.0 | 0.0 | 0.1 | 0.0 |
| 4.8 | 19.2 | 4.0 | 0.0 | 0.0 | 35.2 | 0.0 | 0.0 | 0.0 | 6.1 | 0.0 | 0.0 | 0.2 | 10.0 | 0.1 | 11.8 | 6.7 | 0.0 | 0.0 | 0.1 | 0.1 | 1.4 | 0.0 | 0.0 | 0.0 | 0.2 | 0.0 |
| 25.1 | 3.7 | 0.0 | 5.1 | 6.7 | 12.5 | 0.7 | 1.4 | 0.1 | 2.5 | 0.0 | 0.0 | 0.0 | 0.0 | 4.5 | 3.9 | 0.0 | 0.3 | 0.1 | 8.9 | 22.0 | 1.4 | 0.9 | 0.0 | 0.0 | 0.0 | 0.0 |
| 1.5 | 0.0 | 1.0 | 1.6 | 1.7 | 0.8 | 0.5 | 0.4 | 0.1 | 11.2 | 0.1 | 0.0 | 2.2 | 8.4 | 25.1 | 0.0 | 1.1 | 0.3 | 7.4 | 3.0 | 2.9 | 29.3 | 0.2 | 0.0 | 0.0 | 1.3 | 0.0 |
| 1.6 | 25.0 | 0.0 | 0.2 | 0.0 | 16.6 | 0.0 | 0.0 | 1.4 | 3.0 | 0.0 | 0.0 | 10.5 | 0.0 | 0.0 | 16.1 | 3.9 | 0.0 | 14.9 | 2.2 | 1.1 | 3.5 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 0.2 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 99.8 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 23.2 | 11.2 | 0.3 | 2.2 | 2.6 | 30.8 | 0.3 | 0.6 | 0.0 | 7.3 | 0.0 | 0.0 | 0.6 | 1.5 | 0.8 | 5.8 | 0.3 | 0.3 | 1.9 | 3.9 | 4.7 | 1.2 | 0.6 | 0.0 | 0.0 | 0.1 | 0.0 |
| 49.3 | 7.7 | 0.0 | 0.6 | 0.0 | 12.6 | 0.1 | 0.0 | 0.0 | 6.3 | 0.0 | 0.0 | 0.0 | 0.1 | 0.0 | 6.3 | 1.0 | 1.0 | 0.0 | 6.4 | 4.4 | 2.9 | 0.0 | 1.2 | 0.0 | 0.1 | 0.0 |
| 44.7 | 9.4 | 0.0 | 0.0 | 0.0 | 18.0 | 0.0 | 0.0 | 0.4 | 7.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 4.5 | 0.0 | 0.0 | 8.2 | 1.8 | 3.9 | 1.9 | 0.0 | 0.0 | 0.0 | 0.1 | 0.0 |
| 19.0 | 1.8 | 0.5 | 1.3 | 0.8 | 13.2 | 0.4 | 0.3 | 0.1 | 10.0 | 0.5 | 0.0 | 2.6 | 0.8 | 9.5 | 0.2 | 1.2 | 0.1 | 14.6 | 9.4 | 6.8 | 0.0 | 3.4 | 0.0 | 3.2 | 0.1 | 0.0 |
| 0.3 | 28.6 | 0.0 | 0.0 | 0.0 | 31.4 | 0.0 | 0.0 | 0.0 | 14.6 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 19.3 | 0.0 | 0.0 | 4.2 | 0.0 | 0.0 | 1.6 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 0.0 | 99.7 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.3 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 70.1 | 3.1 | 0.0 | 3.8 | 0.0 | 4.4 | 0.0 | 0.0 | 0.3 | 5.7 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.1 | 6.2 | 2.0 | 0.0 | 0.0 | 3.7 | 0.0 | 0.5 | 0.0 | 0.0 | 0.1 | 0.0 |
| 35.3 | 25.1 | 0.0 | 0.4 | 0.1 | 17.6 | 0.0 | 0.3 | 0.0 | 0.2 | 0.0 | 0.0 | 1.2 | 2.9 | 0.2 | 3.8 | 0.9 | 0.0 | 1.0 | 10.5 | 0.5 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| 86.5 | 1.3 | 0.0 | 0.0 | 0.0 | 5.7 | 0.0 | 0.0 | 0.0 | 0.2 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 3.7 | 0.0 | 0.0 | 0.0 | 0.0 | 0.2 | 1.5 | 0.1 | 0.0 | 0.0 | 0.4 | 0.4 |

FIGURE 4 – Fréquences des bigrammes en % pour l'alphabet " ", "A", "B", e.t.c.

## 5.1 Balayage systématique

Pour tout $x, y \in E^n$, on introduit les probabilités de transition

$$P(x, dy_1 \cdots dy_n) = \prod_{i=1}^{n} \pi_i((y_1, \cdots, y_{i-1}, x_{i+1}, \cdots, x_n); dy_i).$$

Autrement dit, la densité de ce noyau de transition par rapport à $\mu^{\otimes n}$ est donnée par

$$P(x, y) = \prod_{i=1}^{n} \pi_i((y_1, \cdots, y_{i-1}, x_{i+1}, \cdots, x_n); y_i).$$

Algorithmiquement, appliquer le noyau $P(x, dy)$ consiste à partir d'un $n$-uplet $(x_1, x_2, \cdots, x_n)$, puis à remplacer sa première coordonnée $x_1$ par $y_1$ distribué comme $\pi_1((x_2, \cdots, x_n); dy_1)$, puis à remplacer la seconde coordonnée $x_2$ de $(y_1, x_2, \cdots, x_n)$ par $y_2$ suivant la distribution $\pi_2((y_1, x_3, \cdots, x_n); dy_2)$, et ainsi de suite jusqu'à remplacer la dernière coordonnée $x_n$ de $(y_1, y_2, \cdots, y_{n-1}, x_n)$ par $y_n$ distribué comme $\pi_n((y_1, \cdots, y_{n-1}); dy_n)$. On obtient ainsi un vecteur $(y_1, \cdots, y_n)$ distribué selon $P(x, dy)$.

## 5.2 Balayage aléatoire

On reprend ici les mêmes idées, mais à la place de changer systématiquement toutes les coordonnées d'un vecteur $x$ par des éléments $y_i$ distribués suivant les lois conditionnelles, on choisit au hasard, suivant une distribution $\nu_i$, l'indice $i$ du vecteur que l'on modifie. On obtient alors le noyau

$$Q(x, dy) = \sum_{i=1}^{n} \nu_i \, \pi_i(\hat{x}_i, dy_i) \, \mathbb{1}_{\hat{x}_i = \hat{y}_i}.$$

**Proposition 5.1.** *Montrer que $\pi$ est une mesure de probabilité invariante pour $P$ et $Q$.*

## 5.3 Applications

**Exercice 7.** *Soit $X_1, \cdots, X_n$ i.i.d. de loi exponentielle et $S$ leur somme. Proposer un algorithme d'échantillonnage de Gibbs pour simuler la loi de $(X_1, \cdots, X_n)$ sachant $S > s$.*

**Exercice 8.** *Simuler $X_1, \cdots, X_n$ à valeurs dans $[0, 1]$ de loi uniforme sachant que $|X_i - X_j| > \varepsilon$, $i \neq j$.*