TEMPS DE MÉLANGE DE LA MARCHE ALÉATOIRE SUR L'HYPERCUBE

mots-clés : chaînes de Markov, théorème ergodique, temps de mélange.

L'objectif de ce texte est de modéliser la perte de données dans un message chiffré, lorsque l'endroit où est stocké le message est soumis à divers aléas (chocs, changements de température ou de pression, *etc.*). On modélisera le message par un mot binaire

$$m(t) = (m(t))_1(m(t))_2 \dots (m(t))_N \in \{0,1\}^N,$$

de taille $N \geq 2$ fixée et dépendant du temps t, qu'on supposera discrétisé et appartenant à $\mathbb{N} = \{0, 1, 2, \dots\}$. Le mot $m(0)$ est le message originellement stocké (par exemple, 0101110100), et à chaque instant $t \in \mathbb{N}$:

— soit l'espace de stockage n'est pas modifié, et $m(t+1) = m(t)$; ceci se produit avec probabilité $1 - \rho$, où ρ est un certain paramètre dans $(0, 1)$ qui mesure la sensibilité de l'espace de stockage aux aléas extérieurs.

— sinon, avec probabilité ρ, l'une des lettres de $m(t)$ est modifiée. On supposera qu'une seule lettre peut être modifiée à chaque instant, et que toutes les lettres ont la même probabilité d'être modifiées. Ainsi, pour tout $i \in [\![1, N]\!]$, avec probabilité $\frac{\rho}{N}$,

$$(m(t+1))_j = \begin{cases} (m(t))_j & \text{si } j \neq i, \\ 1 - (m(t))_i & \text{si } j = i. \end{cases}$$

On suppose par ailleurs que les aléas subis par le message à des instants différents sont indépendants. On se demande alors combien de temps on peut laisser le message $m(t)$ se détériorer tout en récupérant au moins quelques informations concernant le message original. Plus précisément, on veut répondre aux deux questions suivantes :

(1) En fonction de t, quelle proportion les messages $m(0)$ et $m(t)$ ont-ils en commun ?

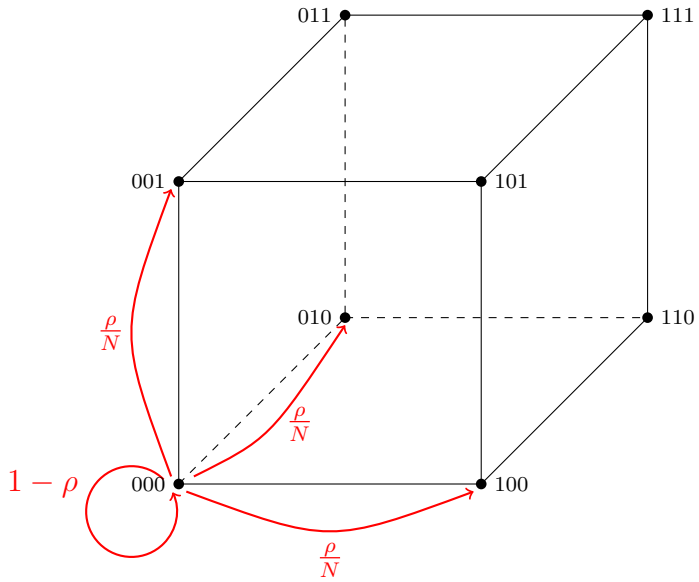
(2) On suppose qu'on dispose d'un très grand nombre de copies du message original $m(0) = m^1(0) = m^2(0) = \dots = m^k(0)$, chacun de ces messages étant stocké dans un endroit indépendant des autres, et soumis au processus aléatoire de dégradation précédemment décrit. On a donc $k \gg 1$ processus indépendants $(m^1(t))_{t\in\mathbb{N}}$, $(m^2(t))_{t\in\mathbb{N}}$, *etc.* Si t est petit, on peut s'attendre à ce qu'on puisse effectivement reconstituer $m(0)$ à partir des différentes copies dégradées $m^1(t), \dots, m^k(t)$. Mais est-ce encore vrai pour t grand (quitte à augmenter le nombre de copies k) ? ou existe-t-il un moment T à partir duquel on ne puisse plus du tout retrouver $m(0)$ à partir de ces copies ?

1. Marche aléatoire sur l'hypercube

Le processus de dégradation des messages constitue une chaîne de Markov $(m(t))_{t\in\mathbb{N}}$ d'espace d'états $M = \{0,1\}^N$, et de matrice de transition

$$P(m,m') = \begin{cases} 1-\rho & \text{si } m = m', \\ \frac{\rho}{N} & \text{si } m \text{ et } m' \text{ diffèrent en exactement une lettre } i \in [\![1,N]\!], \\ 0 & \text{si } m \text{ et } m' \text{ diffèrent en plus de deux lettres.} \end{cases}$$

Géométriquement, l'espace des états peut être représenté par un hypercube de dimension N, et la chaîne de Markov ci-dessus est une marche aléatoire sur cet hypercube :



Dans ce qui suit, on fixe un mot initial $m(0)$, et on note π_t la loi de $m(t)$:

$$\pi_t(w) = \mathbb{P}[m(t) = w],$$

où $(m(t))_{t\in\mathbb{N}}$ est la chaîne de Markov d'espace d'états M, d'état initial $m(0)$ et de matrice de transition P. Si l'on représente une probabilité sur M par un vecteur ligne de taille 2^N dont les coordonnées sont indexées par M, alors

$$\pi_t = \pi_0\, P^t, \quad \text{avec } \pi_0 = (0,\dots,0,1_{m(0)},0,\dots,0).$$

Théorème 1. *La chaîne de Markov $(m(t))_{t\in\mathbb{N}}$ est irréductible, récurrente positive et apériodique, de loi invariante la mesure uniforme*

$$\nu(w) = \frac{1}{2^N} \quad \textit{pour tout mot } w \in M.$$

Pour tout mot initial $m(0)$, on a la convergence en loi

$$\forall w \in M, \ \lim_{t\to\infty} \pi_t(w) = \nu(w) = \frac{1}{2^N}.$$

Idée de preuve. On peut changer les lettres une à une par des transitions de la chaîne de Markov, donc celle-ci est irréductible. D'autre part, la matrice de transition P est bistochastique :

$$\forall m \in M, \ \sum_{m'\in M} P(m,m') = \sum_{m'\in M} P(m',m) = 1.$$

Ceci implique que la mesure invariante est la loi uniforme. Finalement, comme $P(m,m) \neq 0$ pour tout m, la matrice est apériodique. La convergence en loi découle alors du théorème ergodique pour les chaînes irréductibles récurrentes positives et apériodiques. □

On munit M de la distance de Hamming

$$d(m,m') = \operatorname{card}\{i \in [\![1,N]\!] \mid m_i \neq m'_i\}$$

et on s'intéresse à la distribution de $D(t) = d(m(0),m(t))$. On a le résultat théorique suivant :

Proposition 2. *Le processus $(D(t))_{t\in\mathbb{N}}$ est une chaîne de Markov sur $[\![0,N]\!]$, d'état initial 0 et de matrice de transition*

$$Q(k,l) = \begin{cases} 1-\rho & si\ l = k, \\ \frac{\rho\, k}{N} & si\ l = k-1, \\ \frac{\rho\,(N-k)}{N} & si\ l = k+1. \end{cases}$$

À l'aide de ce résultat, on peut calculer en fonction de t le nombre moyen $\mathbb{E}[D(t)]$ de lettres de $m(t)$ qui diffèrent de $m(0)$. En effet,

$$\begin{aligned} \mathbb{E}[D(t+1)] &= \sum_{k=0}^{N} \mathbb{E}[D(t+1) \,|\, D(t)=k]\, \mathbb{P}[D(t)=k] \\ &= \sum_{k=0}^{N} \left((1-\rho)k + \frac{\rho\, k}{N}(k-1) + \frac{\rho\,(N-k)}{N}(k+1) \right) \mathbb{P}[D(t)=k] \\ &= \rho + \sum_{k=0}^{N} \left(1 - \frac{2\rho}{N}\right)\, k\, \mathbb{P}[D(t)=k] = \rho + \left(1-\frac{2\rho}{N}\right)\, \mathbb{E}[D(t)]. \end{aligned}$$

Les espérances $\mathbb{E}[D(t)]$ vérifient donc une équation de récurrence d'ordre 1, dont la solution est

$$\mathbb{E}[D(t)] = \frac{N}{2}\left(1 - \left(1-\frac{2\rho}{N}\right)^t\right).$$

On peut de même calculer

$$\operatorname{var}(D(t)) = \frac{N}{4}\left(1 + (N-1)\left(1-\frac{4\rho}{N}\right)^t - N\left(1-\frac{2\rho}{N}\right)^{2t}\right) \leq \frac{N}{4}.$$

L'inégalité de Bienaymé-Chebyshev donne alors :

Théorème 3. *Pour tout temps t et tout $\varepsilon > 0$,*

$$\mathbb{P}\left[\left|\frac{D(t)}{N} - \frac{1}{2}\right| \geq \varepsilon\right] \leq \frac{1}{4N\varepsilon^2}\, \left(1 + N\left(1-\frac{2\rho}{N}\right)^{2t}\right).$$

Ces calculs démontrent les faits suivants :

(1) Supposons N assez grand (disons $N \geq 10$), et ρ fixé dans $(0,1)$. Le nombre de lettres modifiées $D(t)$ dans le message $m(t)$ est avec grande probabilité de l'ordre de N (c'est-à-dire une proportion positive du message) si et seulement si t est également au moins de l'ordre de N.

(2) Si $t \geq \frac{N \log N}{4\rho}$, alors $\mathbb{P}\left[\left|\frac{D(t)}{N} - \frac{1}{2}\right| \geq \varepsilon\right] \leq \frac{1}{2N\varepsilon^2}$, et on a une concentration des valeurs de $D(t)$ autour de la valeur $\frac{N}{2}$.

2. Temps de mélange de la chaîne de Markov

Les résultats précédents donnent une première information sur le temps t requis pour que le message $m(t)$ se dégrade significativement. Pour aller plus loin, on va calculer exactement la distribution π_t de $m(t)$. On équipe l'espace vectoriel V des fonctions de M dans $\mathbb{R}$ du produit scalaire

$$\langle f \mid g \rangle = \frac{1}{2^N} \sum_{m \in M} f(m) g(m),$$

et si $w, m \in M$, on note $\varepsilon_w(m) = (-1)^{\sum_{i=1}^N w_i m_i}$. D'autre part, on identifie $\{0,1\}$ et le groupe additif $\mathbb{Z}/2\mathbb{Z}$ (c'est-à-dire que $1+1=0$ dans ce groupe), puis M et le groupe additif $(\mathbb{Z}/2\mathbb{Z})^N$: ainsi, étant donnés deux mots l et m, $l+m$ est le mot

$$(l_1 + m_1)(l_2 + m_2) \cdots (l_N + m_N) \in (\mathbb{Z}/2\mathbb{Z})^N.$$

Proposition 4. *Les fonctions ε_w avec $w \in M$ forment une base orthonormée de V. De plus, si l'on identifie M et le groupe additif $(\mathbb{Z}/2\mathbb{Z})^N$, alors :*

(1) Les fonctions ε_w sont des morphismes de groupes de $((\mathbb{Z}/2\mathbb{Z})^N, +)$ vers $(\{\pm 1\}, \times)$.

(2) Pour tous $m, w_1, w_2 \in M$,

$$\frac{1}{2^N} \sum_{m_1, m_2 \in M \mid m = m_1 + m_2} \varepsilon_{w_1}(m_1)\, \varepsilon_{w_2}(m_2) = \begin{cases} \varepsilon_{w_1}(m) & \text{si } w_1 = w_2, \\ 0 & \text{sinon.} \end{cases}$$

La matrice de transition P et ses puissances peuvent s'exprimer simplement grâce aux fonctions ε_w. En effet, notons δ_0 le Dirac en le mot nul $00\cdots0$, et pour $i \in [\![1, N]\!]$, δ_i le Dirac en le mot $0\cdots01_i0\cdots0$ avec un 1 seulement en i-ième position. On peut alors écrire :

$$P(m, m') = (1-\rho)\, \delta_0(m - m') + \frac{\rho}{N} \sum_{i=1}^N \delta_i(m - m') = p(m - m')$$

avec $p = (1-\rho)\, \delta_0 + \frac{\rho}{N} \sum_{i=1}^n \delta_i$. Pour tout mot w, on a alors $\langle p \mid \varepsilon_w \rangle = \frac{1}{2^N}\left(1 - \frac{2\rho|w|}{N}\right)$ avec $|w| = \operatorname{card}\{i \in [\![1, N]\!] \mid w_i = 1\}$. Comme ε_w est une base orthonormée, ceci implique :

$$P(m, m') = \frac{1}{2^N} \sum_{w \in M} \left(1 - \frac{2\rho|w|}{N}\right) \varepsilon_w(m - m').$$

Puis, on calcule $P^2(m, m')$ comme suit :

$$\begin{aligned} P^2(m, m') &= \sum_{l \in M} p(m-l)\, p(l-m') = \sum_{l, w_1, w_2} \langle p \mid \varepsilon_{w_1} \rangle\, \langle p \mid \varepsilon_{w_2} \rangle\, \varepsilon_{w_1}(m-l)\, \varepsilon_{w_2}(l-m') \\ &= 2^N \sum_{w_1, w_2} \langle p \mid \varepsilon_{w_1} \rangle\, \langle p \mid \varepsilon_{w_2} \rangle\, 1_{w_1 = w_2}\, \varepsilon_{w_1}(m - m') \\ &= \frac{1}{2^N} \sum_{w \in M} \left(1 - \frac{2\rho|w|}{N}\right)^2 \varepsilon_w(m - m') \end{aligned}$$

en utilisant à la seconde ligne la seconde partie de la proposition 4. Par récurrence sur t, on obtient la formule explicite suivante pour P^t, ou de façon équivalente pour la loi π_t :

Théorème 5. *Pour tout temps $t \in \mathbb{N}$, tout mot initial $m(0)$ et tout mot $m \in M$,*

$$\pi_t(m) = \frac{1}{2^N} \sum_{w \in M} \left(1 - \frac{2\rho|w|}{N}\right)^t \varepsilon_w(m - m(0)).$$

Si $w \neq 00\cdots0$, alors $(1 - \frac{2\rho|w|}{N}) \in (-1, 1)$ et dans la somme ci-dessus, le terme correspondant à w tend vers 0. On retrouve donc la convergence en loi $\pi_t \rightharpoonup \nu$, mais en plus on peut en calculer la vitesse. Supposons pour simplifier $\rho = \frac{1}{2}$.

Théorème 6. *On définit la distance en variation totale entre π_t et ν par les formules équivalentes*

$$d_{\mathrm{VT}}(\pi_t, \nu) = \max_{A \subset M} |\pi_t(A) - \nu(A)| = \frac{1}{2} \sum_{m \in M} |\pi_t(m) - \nu(m)|.$$

Il existe une constante $K > 0$ telle que, si $\rho = \frac{1}{2}$ et si $t = \frac{N}{2}(\log N + c)$ avec $c > 0$, alors

$$d_{\mathrm{VT}}(\pi_t, \nu) \leq K\, \mathrm{e}^{-\frac{c}{2}}.$$

Idée de preuve. La distance en variation totale $d_{\mathrm{VT}}(\pi_t, \nu)$ est la moitié de la norme L^1 de la fonction $m \mapsto 2^N \pi_t(m) - 1$ dans l'espace $\mathrm{L}^1(M, \nu)$. Par Cauchy-Schwarz, on a donc

$$\begin{aligned}(d_{\mathrm{VT}}(\pi_t, \nu))^2 &\leq \frac{1}{4} \left\| 2^N \pi_t(\cdot) - 1 \right\|^2_{\mathrm{L}^2(M,\nu)} \\ &\leq \frac{1}{4} \sum_{w \neq 00\cdots0} (2^N \langle \pi_t \mid \varepsilon_w \rangle)^2 = \frac{1}{4} \sum_{w \neq 00\cdots0} \left(1 - \frac{2\rho|w|}{N}\right)^{2t}.\end{aligned}$$

En supposant $\rho = \frac{1}{2}$ et en regroupant les termes de la somme selon le poids $|w| = k \in [\![1, N]\!]$, on obtient la majoration

$$(d_{\mathrm{VT}}(\pi_t, \nu))^2 \leq \frac{1}{4} \sum_{k=1}^{N} \frac{1}{k!}\, \mathrm{e}^{k(\log N - \frac{2t}{N})} \leq \frac{1}{4} \left(\mathrm{e}^{\mathrm{e}^{(\log N - \frac{2t}{N})}} - 1 \right).$$

Si $t = \frac{N}{2}(\log N + c)$ avec $c > 0$, alors on obtient bien $d_{\mathrm{VT}}(\pi_t, \nu) \leq \frac{1}{2}\sqrt{\mathrm{e}^{\mathrm{e}^{-c}} - 1} = O(\mathrm{e}^{-\frac{c}{2}})$. □

Supposons qu'on dispose de nombreuses copies indépendantes $m^1(t), \ldots, m^k(t)$ issues du même mot initial $m(0)$. On peut à partir de ces copies estimer la distribution π_t, et pour t petit cette distribution dépend beaucoup de $m(0)$. Le résultat précédent montre que si $t \geq \frac{N \log N}{2}$, alors π_t est très proche de la loi stationnaire ν, et on ne peut donc plus rien dire de $m(0)$ en observant les messages dégradés $m^1(t), \ldots, m^k(t)$. On peut par ailleurs montrer que, si $t \leq \frac{N \log N}{2}$, alors on a un résultat inverse de celui donné par le théorème 6 :

Proposition 7. *Supposons toujours $\rho = \frac{1}{2}$. Il existe une constante K telle que, si $t = \frac{N}{2}(\log N - c)$ avec $c > 0$, alors*

$$d_{\mathrm{VT}}(\pi_t, \nu) \geq 1 - \frac{K}{c^2}.$$

Ainsi, aux alentours du temps $T = \frac{N \log N}{2}$, la loi π_t du message $m(t)$ s'homogénéise subitement, et après ce temps on ne peut plus rien dire du message initial $m(0)$.

Questions

Pour la rédaction des programmes, on pourra utilise n'importe quel langage de programmation, ou éventuellement donner une description détaillée de l'algorithme (pseudo-code).

I.1 Vérifier que la matrice de transition P est apériodique et bistochastique. Montrer que si P est une matrice bistochastique sur un espace fini X (c'est-à-dire que pour tout $x \in X$, $\sum_{y\in X} P(x,y) = \sum_{y\in X} P(y,x) = 1$), alors la mesure de probabilité uniforme $\nu(x) = \frac{1}{\operatorname{card} X}$ est invariante par P.

I.2 Écrire un programme `MarkovMessage(m,t)` qui prend en argument un mot initial `m` de taille N arbitraire et un temps `t`, et qui calcule les états $m(1), m(2), \ldots, m(t)$ de la chaîne de Markov. On pourra fixer $\rho = \frac{1}{2}$ dans tous les programmes.

I.3 Démontrer la proposition 2, et calculer la mesure invariante de la chaîne $(D(t))_{t\in\mathbb{N}}$ (on pourra chercher une mesure réversible). Le théorème ergodique s'applique-t-il à cette chaîne ? Si c'est le cas, expliquer pourquoi et énoncer le résultat.

I.4 Vérifier que la solution de l'équation de récurrence $\mathbb{E}[D(t+1)] = \rho + (1 - \frac{2\rho}{N})\,\mathbb{E}[D(t)]$ est bien $\mathbb{E}[D(t)] = \frac{N}{2}(1 - (1 - \frac{2\rho}{N})^t)$. Démontrer la formule donnée dans le texte pour $\operatorname{var}(D(t))$, ainsi que l'inégalité $\operatorname{var}(D(t)) \leq \frac{N}{4}$. Pour la variance, on pourra d'abord établir une relation de récurrence satisfaite par $\mathbb{E}[C(t)]$, où $C(t) = D(t)(D(t) - N)$.

I.5 Écrire des programmes `MoyenneEmpiriqueD(t,N)` et `VarianceEmpiriqueD(t,N)` qui permettent de vérifier par l'expérience les formules théoriques pour $\mathbb{E}[D(t)]$ et $\operatorname{var}(D(t))$.

I.6 Si $w \in M$ est un mot fixé (pas forcément $m(0)$), calculer $\mathbb{E}[d(m(t), w)]$. Pour tout temps t, quel est le mot w qui rend $\mathbb{E}[d(m(t), w)]$ minimal ? En déduire un algorithme qui permet de retrouver avec grande probabilité $m(0)$ à partir de copies indépendantes $m^1(t), m^2(t), \ldots, m^k(t)$. Discuter de l'efficacité de cet algorithme en fonction de k et de t. On pourra éventuellement programmer l'algorithme pour des mots de petite taille (par exemple $N = 5$).

II.1 Démontrer entièrement la proposition 4.

II.2 La matrice de transition P agit sur l'espace de fonctions $V = \mathbb{R}^M$ par $f \mapsto Pf$, où les fonctions f sont représentées par des vecteurs colonnes $(f(m))_{m\in M}$. Calculer $P\,\varepsilon_w$. Montrer que P est diagonalisable et donner ses valeurs propres, et ses vecteurs propres à droite.

II.3 Montrer que pour toutes mesures de probabilité ρ_1 et ρ_2 sur M, si $d_{\mathrm{VT}}(\rho_1, \rho_2) = \max_{A\subset M} |\rho_1(A) - \rho_2(A)|$, alors le maximum est atteint en la partie

$$B = \{m \in M \mid \rho_1(m) \geq \rho_2(m)\}.$$

En déduire l'identité $d_{\mathrm{VT}}(\rho_1, \rho_2) = \frac{1}{2}\sum_{m\in M} |\rho_1(m) - \rho_2(m)|$.

II.4 On suppose dans tout ce qui suit $\rho = \frac{1}{2}$. Détailler les deux séries d'inégalités dans la preuve du théorème 6.

II.5 Écrire un programme `VariationTotale(t,N)` qui dessine un graphe de $d_{\mathrm{VT}}(\pi_t, \nu)$ en fonction de t et de la taille N des mots. Produire un tel graphe lorsque $N = 8$, et commenter ce graphe.

II.6 (Facultatif) En considérant des événements A de la forme $\{D(t) \leq M\}$ avec M convenablement choisi, et en utilisant les estimées de $\mathbb{E}[D(t)]$ et de $\operatorname{var}(D(t))$, trouver pour tout temps $t = \frac{N}{2}(\log N - c)$ avec c suffisamment grand un événement A avec $\pi_t(A)$ proche de 1 et $\nu(A)$ proche de 0. En déduire une preuve de la proposition 7.