TEMPS DE MÉLANGE DE LA MARCHE ALÉATOIRE SUR L'HYPERCUBE - CORRIGÉ

On a rédigé les programmes demandés en pseudo-code, les variables apparaissant en gras.

I.1 Pour tout mot m, $P(m,m) = 1-\rho > 0$, donc $\mathrm{pgcd}(\{t \in \mathbb{N} \,|\, P^t(m,m) > 0\}) = 1$ et la matrice de transition est apériodique. D'autre part, la matrice est bien bistochastique, car elle est symétrique et stochastique, donc

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

Ceci implique que la loi uniforme est invariante :

$$(\nu P)(m) = \sum_{m' \in M} \nu(m')P(m',m) = \frac{1}{\mathrm{card}\, M} \sum_{m' \in M} P(m',m) = \frac{1}{\mathrm{card}\, M} = \nu(m).$$

I.2 On peut utiliser le programme suivant, qui considère un mot comme une liste :

```
MarkovMessage(m,t):
   résultat = liste vide []
   mot = m
   N = longueur de m
   pour s entre 1 et t:
       aléa = variable aléatoire uniforme dans [0,1]
       si aléa < 0.5:
           i = partie entière de 2*N*aléa
           mot[i] = 1 - mot[i]
       ajouter le mot à la liste résultat
   retourner le résultat
```

I.3 Pour montrer que $D(t)$ est une chaîne de Markov, on va utiliser le théorème de représentation. Fixons un mot initial $m(0)$, et pour tout mot $m \in M$, notons $\{i_1 < i_2 < \cdots < i_k\}$ les indices tels que $m_{i_a} \neq (m(0))_{i_a}$, et $\{j_1 < j_2 < \cdots < j_{N-k}\}$ les indices tels que $m_{j_b} = (m(0))_{j_b}$. Notons que ces familles d'indices sont des fonctions bien définies de m. Considérons alors l'application

$$g : M \times [0,1] \to M$$

$$m, u \mapsto \begin{cases} m & \text{si } u \geq \rho, \\ m' = m + (0, \ldots, 1_{i_a}, \ldots, 0) & \text{si } \frac{\rho\,(a-1)}{N} \leq u < \frac{\rho\, a}{N}, \\ m' = m + (0, \ldots, 1_{j_b}, \ldots, 0) & \text{si } \frac{\rho\,(d(m,m(0))+b-1)}{N} \leq u < \frac{\rho\,(d(m,m(0))+b)}{N}, \end{cases}$$

l'addition ayant lieu dans $M = (\mathbb{Z}/2\mathbb{Z})^N$. Si $(U_t)_{t \in \mathbb{N}}$ est une suite de variables aléatoires uniformes sur $[0,1]$ et indépendantes, alors le processus $(m(t))_{t \in \mathbb{N}}$ défini par $m(t+1) = g(m(t), U_t)$ est une chaîne de Markov sur M de matrice de transition P. Or, avec cette construction particulière de la chaîne de Markov $(m(t))_{t \in \mathbb{N}}$, on peut facilement calculer

$D(t)$, car

$$D(t+1) = d(g(m(t), U_t), m(0)) = \begin{cases} D(t) & \text{si } U_t \geq \rho, \\ D(t) - 1 & \text{si } U_t < \rho\, \frac{D(t)}{N}, \\ D(t) + 1 & \text{si } \rho\, \frac{D(t)}{N} \leq U_t < \rho. \end{cases}$$

Autrement dit, si l'on considère l'application

$$f : [\![0, N]\!] \times [0,1] \to [\![0, N]\!]$$
$$k, u \mapsto \begin{cases} k & \text{si } u \geq \rho, \\ k-1 & \text{si } u < \frac{\rho\, k}{N}, \\ k+1 & \text{si } \frac{\rho\, k}{N} \leq u < \rho, \end{cases}$$

alors $D(t)$ vérifie l'équation de récurrence $D(t+1) = f(D(t), U_t)$, avec $D(0) = 0$. Par le théorème de représentation des chaînes de Markov, $(D(t))_{t \in \mathbb{N}}$ est une chaîne de Markov sur $[\![0, N]\!]$, de matrice de transition :

$$Q(k,k) = \mathbb{P}[U \geq \rho] = 1 - \rho$$
$$Q(k,k-1) = \mathbb{P}\left[U < \frac{\rho\, k}{N}\right] = \frac{\rho\, k}{N}$$
$$Q(k,k+1) = \mathbb{P}\left[\frac{\rho\, k}{N} \leq U < \rho\right] = \frac{\rho\,(N-k)}{N}.$$

Pour $\rho = 1$, on reconnaît la chaîne de Markov des urnes d'Ehrenfest, et il est bien connu que la loi binomiale $\mathcal{B}(N, \frac{1}{2})$ en est la mesure invariante. C'est encore vrai pour $\rho \in (0,1)$ arbitraire, car

$$\frac{1}{2^N}\binom{N}{k} Q(k,k+1) = \frac{\rho}{2^N}\frac{N-k}{N}\binom{N}{k}$$
$$= \frac{\rho}{2^N}\frac{k+1}{N}\binom{N}{k+1} = \frac{1}{2^N}\binom{N}{k+1} Q(k+1,k).$$

La loi invariante est donc toujours $p(k) = \frac{1}{2^N}\binom{N}{k}$. Pour $\rho \in (0,1)$, la matrice de transition vérifie $Q(k,k) = 1 - \rho > 0$ pour tout k, et d'autre part elle est irréductible (on peut passer de proche en proche de tout état $k \in [\![0, N]\!]$ à tout autre état l). On a donc une chaîne de Markov irréductible apériodique, et comme l'espace d'états est fini, elle est récurrente positive. Le théorème ergodique s'applique, et il assure que

$$\lim_{t \to \infty} \mathbb{P}[D(t) = k] = \frac{1}{2^N}\binom{N}{k}$$

pour tout $k \in [\![0, N]\!]$.

I.4 Posons $\delta_t = \mathbb{E}[D(t)]$, et résolvons l'équation de récurrence $\delta_{t+1} = \rho + (1 - \frac{2\rho}{N})\delta_t$ avec condition initiale $\delta_0 = 0$. Le point fixe de l'équation est $\frac{N}{2}$, et en le soustrayant à gauche et à droite de l'équation de récurrence, on obtient :

$$\left(\delta_{t+1} - \frac{N}{2}\right) = \left(1 - \frac{2\rho}{N}\right)\left(\delta_t - \frac{N}{2}\right),$$

donc

$$\delta_t - \frac{N}{2} = \left(1 - \frac{2\rho}{N}\right)^t \left(\delta_0 - \frac{N}{2}\right) = -\frac{N}{2}\left(1 - \frac{2\rho}{N}\right)^t,$$

et finalement $\delta_t = \frac{N}{2}(1 - (1 - \frac{2\rho}{N})^t)$.

Examinons maintenant $\mathbb{E}[(D(t))^2]$. Cette quantité vérifie l'équation de récurrence :

$$\begin{aligned}
\mathbb{E}[(D(t+1))^2] &= \sum_{k=0}^{N} \mathbb{E}[(D(t+1))^2 \mid D(t)=k]\,\mathbb{P}[D(t)=k] \\
&= \sum_{k=0}^{N} \left((1-\rho)k^2 + \frac{\rho\,k}{N}(k-1)^2 + \frac{\rho\,(N-k)}{N}(k+1)^2 \right) \mathbb{P}[D(t)=k] \\
&= \sum_{k=0}^{N} \left(\left(1-\frac{4\rho}{N}\right) k^2 + 2\rho\,k + \rho \right) \mathbb{P}[D(t)=k] \\
&= \left(1-\frac{4\rho}{N}\right) \mathbb{E}[(D(t))^2] + 2\rho\,\mathbb{E}[D(t)] + \rho.
\end{aligned}$$

Avec $C(t) = (D(t))^2 - N\,D(t)$, on obtient alors :

$$\begin{aligned}
\mathbb{E}[C(t+1)] &= \left(1-\frac{4\rho}{N}\right) \mathbb{E}[(D(t))^2] + 2\rho\,\mathbb{E}[D(t)] + \rho - N\rho - N\left(1-\frac{2\rho}{N}\right)\mathbb{E}[D(t)] \\
&= \left(1-\frac{4\rho}{N}\right) \mathbb{E}[C(t)] + \rho(1-N).
\end{aligned}$$

Le point fixe de l'équation est $-\frac{N(N-1)}{4}$, et comme $C(0)=0$, on en déduit

$$\mathbb{E}[C(t)] = \frac{N(N-1)}{4}\left(\left(1-\frac{4\rho}{N}\right)^t - 1\right).$$

Finalement,

$$\begin{aligned}
\operatorname{var}(D(t)) &= \mathbb{E}[(D(t))^2] - \mathbb{E}[D(t)]^2 = \mathbb{E}[C(t)] + \mathbb{E}[D(t)](N-\mathbb{E}[D(t)]) \\
&= \frac{N(N-1)}{4}\left(\left(1-\frac{4\rho}{N}\right)^t - 1\right) + \frac{N^2}{4}\left(1-\left(1-\frac{2\rho}{N}\right)^t\right)\left(1+\left(1-\frac{2\rho}{N}\right)^t\right) \\
&= \frac{N}{4}\left(1 + (N-1)\left(1-\frac{4\rho}{N}\right)^t - N\left(1-\frac{2\rho}{N}\right)^{2t}\right).
\end{aligned}$$

On peut réécrire la formule sous la forme

$$\frac{N}{4}\left(1-\left(1-\frac{4\rho}{N}\right)^t + N\left(\left(1-\frac{4\rho}{N}\right)^t - \left(1-\frac{2\rho}{N}\right)^{2t}\right)\right),$$

et comme $1-\frac{4\rho}{N} \le (1-\frac{2\rho}{N})^2$, le terme en facteur de N est négatif, tout comme $-(1-\frac{4\rho}{N})^t$. On conclut que $\operatorname{var}(D(t)) \le \frac{N}{4}$.

I.5 On programme la chaîne de Markov $D(t)$, dépendant également du paramètre N :

```
D(t,N):
    résultat = 0
    pour s entre 1 et t:
        aléa = variable aléatoire uniforme dans [0,1]
        si aléa < 0.5*résultat/N:
            résultat = résultat-1
        sinon, si aléa < 0.5:
            résultat = résultat+1
    retourner le résultat
```

La moyenne empirique et la variance empirique (sans biais) de $D(t)$ sont alors données par les formules

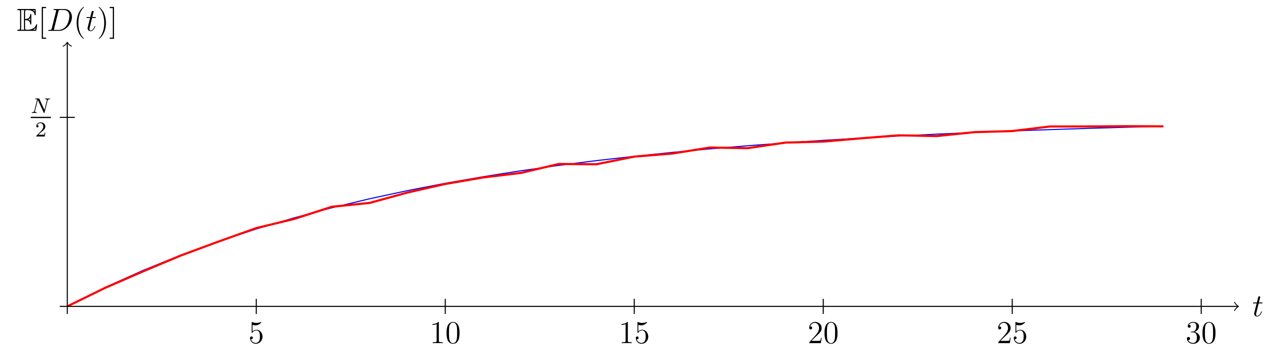
$$\overline{m}(t) = \frac{1}{n}\sum_{i=1}^{n} D^i(t) \quad ; \quad \overline{V}(t) = \frac{1}{n-1}\sum_{i=1}^{n}(D^i(t) - \overline{m}(t))^2$$

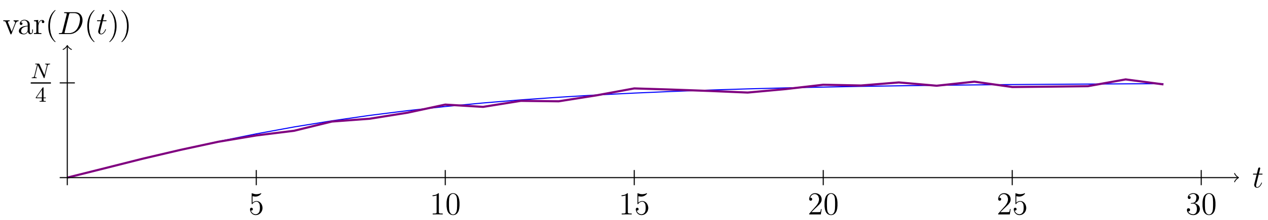
avec des copies indépendantes $D^1(t), D^2(t), \ldots, D^n(t)$ de $D(t)$. Ces formules correspondent aux programmes suivants :

```
MoyenneEmpiriqueD(t,N,n):
    L = [n copies de D(t,N)]
    retourner la moyenne de L

VarianceEmpiriqueD(t,N,n):
    L = [n copies de D(t,N)]
    m = moyenne de L
    L2 = [(L[i]-m)^2 pour i entre 1 et n]
    retourner somme(L2)/(n-1)
```

On a dessiné ci-dessous les graphes de $\mathbb{E}[D(t)]$ et $\mathrm{var}(D(t))$ pour $N = 10$ et t entre 1 et 30, avec $n = 1000$ essais pour chaque calcul de moyenne ou variance empirique :





On constate bien l'accord avec les valeurs théoriques.

I.6 Si m est un mot fixé, on voit facilement que $d(m(t), m)$ est la même chaîne de Markov que $d(m(t), m(0))$, mais avec un état initial $d(m(0), m)$ au lieu de 0. En particulier, $\mathbb{E}[d(m(t), m)]$ vérifie la même équation de récurrence que $\mathbb{E}[D(t)]$, et on obtient :

$$\mathbb{E}[d(m(t), m)] = \frac{N}{2} + \left(d(m(0), m) - \frac{N}{2}\right)\left(1 - \frac{2\rho}{N}\right)^t.$$

En particulier, on a toujours

$$\mathbb{E}[d(m(t), m)] \geq \frac{N}{2}\left(1 - \left(1 - \frac{2\rho}{N}\right)^t\right) = \mathbb{E}[d(m(t), m(0))],$$

donc $m(0)$ est l'unique mot qui minimise la quantité $\mathbb{E}[d(m(t), m)]$. Supposons qu'on dispose de k copies de $m(t)$, et qu'on souhaite retrouver $m(0)$. Si k est assez grand, alors par la loi des grands nombres appliquée aux quantités

$$\frac{1}{k} \sum_{i=1}^{k} d(m^i(t), m),$$

avec grande probabilité, $m = m(0)$ sera le mot qui donne la quantité minimale. On peut donc "deviner" $m(0)$ avec les programmes suivants :

```
DistanceHamming(m,W):
    N = longueur de m
    résultat = 0
    pour i entre 1 et N:
        si m[i] est différent de w[i]:
            ajouter 1 au résultat
    retourner le résultat

DevinerMessage(L):
    k = longueur de la liste L
    N = longueur d'un mot dans L
    M = liste des 2^N mots de {0,1}^N
    distances = liste vide []
    pour m dans M:
        ajouter à la fin de distances la quantité
        somme([liste des DistanceHamming(w,m) avec w dans L])/k
    retourner M[i], où i est l'indice du minimum de distances
```

Le problème de cet algorithme est qu'il n'est efficace que si k est très grand. Plus précisément, lorsque m parcourt M, les valeurs théoriques des quantités $\mathbb{E}[d(m(t), m)]$ sont séparées les unes des autres par $(1 - \frac{2\rho}{N})^t = \left(1 - \frac{1}{N}\right)^t$ si l'on suppose $\rho = \frac{1}{2}$. Or, étant données k copies indépendantes de $m(t)$, la variance de la moyenne empirique $\frac{1}{k} \sum_{i=1}^{k} d(m^i(t), m(0))$ est $\mathrm{var}(D(t))/k$, qu'on peut approcher (ou au moins majorer) par $\frac{N}{4k}$. Pour que la moyenne empirique $\frac{1}{k} \sum_{i=1}^{k} d(m^i(t), m(0))$ ne se retrouve pas régulièrement supérieure à la seconde plus petite valeur théorique $\mathbb{E}[D(t)] + \left(1 - \frac{1}{N}\right)^t$ à cause de cette variance, il faut donc au moins

$$\sqrt{\frac{N}{4k}} \ll \left(1 - \frac{1}{N}\right)^t \qquad ; \qquad k \gg \frac{N}{\left(1 - \frac{1}{N}\right)^{2t}}.$$

En remplaçant $1 - \frac{1}{N}$ par e^{-N}, on arrive à l'estimée suivante : la méthode proposée pour deviner $m(0)$ a une chance de marcher correctement si k est au moins de l'ordre de $N \, \mathrm{e}^{\frac{2t}{N}}$. En particulier, dès que $t \geq N$, il faut énormément de copies k pour avoir une chance de retrouver le message.

En pratique, sur un message de taille $N = 10$, si $t = 10$, on retrouve correctement le message $m(0)$ avec 50 copies $m^k(t)$ (avec probabilité estimée à 96%), mais si $t = 20$, il faut déjà environ 400 copies pour avoir la même probabilité.

II.1 Si m et m' sont deux mots et w est un autre mot, alors

$$\varepsilon_w(m+m') = (-1)^{\sum_{i=1}^m (w_i)(m_i+m'_i)} = (-1)^{(\sum_{i=1}^N w_i m_i + \sum_{i=1}^N w_i m'_i)} = \varepsilon_w(m)\,\varepsilon_w(m'),$$

donc les ε_w sont bien des morphismes de groupe de $((\mathbb{Z}/2\mathbb{Z})^N,+)$ vers $(\{\pm 1\},\times)$. Le même calcul montre par ailleurs que $\varepsilon_w\,\varepsilon'_w = \varepsilon_{w+w'}$. Ceci implique l'orthonormalité de ces fonctions, car

$$\begin{aligned}\langle \varepsilon_w \mid \varepsilon_{w'}\rangle &= \frac{1}{2^N}\sum_{m\in\{0,1\}^N} \varepsilon_w(m)\,\varepsilon_{w'}(m) = \frac{1}{2^N}\sum_{m\in\{0,1\}^N}\varepsilon_{w+w'}(m)\\ &= \frac{1}{2^N}\prod_{i=1}^N\left(\sum_{m_i\in\{0,1\}}(-1)^{m_i(w_i+w'_i)}\right).\end{aligned}$$

Si $w=w'$, alors toutes les sommes valent 2 et on obtient $\langle \varepsilon_w \mid \varepsilon_w\rangle = 1$. Au contraire, si $w\neq w'$, alors pour tout indice i tel que $w_i \neq w'_i$, la somme correspondante s'annule, et on a donc $\langle \varepsilon_w \mid \varepsilon_{w'}\rangle = 0$. Comme $\dim V = \operatorname{card} M = 2^N$, on conclut que $(\varepsilon_w)_{w\in M}$ est une base orthonormée de V.

Pour la seconde partie de la proposition, on écrit grâce aux propriétés de morphismes, et au fait que $-m_1 = m_1$ dans $\mathbb{Z}/2\mathbb{Z}$:

$$\begin{aligned}\frac{1}{2^N}\sum_{m_1+m_2=m}\varepsilon_{w_1}(m_1)\,\varepsilon_{w_2}(m_2) &= \left(\frac{1}{2^N}\sum_{m_1\in M}\varepsilon_{w_1}(m_1)\,\varepsilon_{w_2}(m_1)\,\varepsilon_{w_2}(m_2-m_1)\right)\\ &= \left(\frac{1}{2^N}\sum_{m_1\in M}\varepsilon_{w_1}(m_1)\,\varepsilon_{w_2}(m_1)\,\varepsilon_{w_2}(m_2+m_1)\right)\\ &= \langle \varepsilon_{w_1} \mid \varepsilon_{w_2}\rangle\;\varepsilon_{w_2}(m) = \delta_{w_1,w_2}\,\varepsilon_{w_1}(m).\end{aligned}$$

II.2 On calcule

$$\begin{aligned}(P\cdot\varepsilon_w)(m) &= \sum_{m'\in M} P(m,m')\varepsilon_w(m') = \sum_{m'\in M} p(m-m')\varepsilon_w(m')\\ &= \sum_{m',w'\in M}\langle p \mid \varepsilon_{w'}\rangle\,\varepsilon'_w(m-m')\varepsilon_w(m') = 2^N\sum_{w'\in M}\langle p\mid\varepsilon_{w'}\rangle\;\delta_{w',w}\,\varepsilon_w(m)\\ &= \big(2^N\,\langle p \mid \varepsilon_w\rangle\big)\;\varepsilon_w(m) = \left(1-\frac{2\rho|w|}{N}\right)\;\varepsilon_w(m).\end{aligned}$$

On en déduit que $(\varepsilon_w)_{w\in M}$ est une base de diagonalisation de P, avec la valeur propre $(1-\frac{2\rho|w|}{M})$ associée à chaque vecteur ε_w.

II.3 Soit $A\subset M$ une partie qui maximise la différence $|\rho_1(A)-\rho_2(A)|$. Notons $A_+ = \{m\in A\,|\,\rho_1(m) > \rho_2(m)\}$ et $A_- = \{m\in A\,|\,\rho_2(m)>\rho_1(m)\}$: on a

$$|\rho_1(A)-\rho_2(A)| = |(\rho_1(A_+)-\rho_2(A_+)) - (\rho_2(A_-)-\rho_1(A_-))|.$$

Si A_+ et A_- sont non vides, alors les deux quantités $(\rho_1(A_+)-\rho_2(A_+))$ et $(\rho_2(A_-)-\rho_1(A_-))$ sont strictement positives, et $|\rho_1(A)-\rho_2(A)| < |\rho_1(A_+)-\rho_2(A_+)|$, ce qui contredit la maximalité de la partie A. On conclut que soit A_+, soit A_- est vide, et comme la différence $|\rho_1(A)-\rho_2(A)|$ est invariante par $A\mapsto A^c$, on peut supposer sans perte de généralité que $A_- = \emptyset$. Alors, A est incluse dans $B = \{m\in M\,|\,\rho_1(m)\geq \rho_2(m)\}$, et si $m\in B\backslash A$, alors on augmente la quantité $|\rho_1(A)-\rho_2(A)| = \rho_1(A)-\rho_2(A)$

en adjoignant m à A. On conclut que B est une partie qui maximise $|\rho_1(\cdot)-\rho_2(\cdot)|$. Finalement,

$$\begin{aligned}
d_{\mathrm{VT}}(\rho_1,\rho_2) = |\rho_1(B)-\rho_2(B)| &= \frac{|\rho_1(B)-\rho_2(B)|+|\rho_1(B^{\mathrm{c}})-\rho_2(B^{\mathrm{c}})|}{2} \\
&= \frac{1}{2}\sum_{m\in B}(\rho_1(m)-\rho_2(m)) + \frac{1}{2}\sum_{m\in B^{\mathrm{c}}}(\rho_2(m)-\rho_1(m)) \\
&= \frac{1}{2}\sum_{m\in B}|\rho_1(m)-\rho_2(m)| + \frac{1}{2}\sum_{m\in B^{\mathrm{c}}}|\rho_1(m)-\rho_2(m)| = \frac{1}{2}\sum_{m\in M}|\rho_1(m)-\rho_2(m)|.
\end{aligned}$$

II.4 On a

$$\begin{aligned}
(d_{\mathrm{VT}}(\pi_t,\nu))^2 &= \frac{1}{4}\left(\sum_{m\in M}\left|\pi_t(m)-\frac{1}{2^N}\right|\right)^2 \quad \text{d'après la question précédente,} \\
&= \frac{1}{4}\left(\sum_{m\in M}\nu(m)\,\left|2^N\pi_t(m)-1\right|\right)^2 = \frac{1}{4}\left(\left\|2^N\,\pi_t(\cdot)-1\right\|_{\mathrm{L}^1(M,\nu)}\right)^2, \\
&\leq \frac{1}{4}(\|2^N\,\pi_t(\cdot)-1\|_{\mathrm{L}^2(M,\nu)})^2 \quad \text{par l'inégalité de Cauchy-Schwarz,} \\
&\leq \frac{1}{4}\sum_{w\in M}\left\langle 2^N\pi_t(\cdot)-1\,\middle|\,\varepsilon_w\right\rangle^2 \quad \text{car } (\varepsilon_w)_{w\in M} \text{ est une base orthonormée.}
\end{aligned}$$

Notons que $\varepsilon_{00\cdots0}$ est la fonction constante égale à 1. Par conséquent,

$$\left\langle 2^N\pi_t\,\middle|\,\varepsilon_{00\cdots0}\right\rangle = \sum_{m\in M}\pi_t(m) = 1 = \langle 1\mid\varepsilon_{00\cdots0}\rangle$$

et le terme correspondant au mot nul ne contribue pas à la somme. Pour les autres mots, on a $\langle 1\mid\varepsilon_w\rangle = \langle\varepsilon_{00\cdots0}\mid\varepsilon_w\rangle = 0$, donc

$$\begin{aligned}
\left\langle 2^N\pi_t(\cdot)-1\,\middle|\,\varepsilon_w\right\rangle = 2^N\,\langle\pi_t\mid\varepsilon_w\rangle &= \sum_{m\in M}\pi_t(m)\,\varepsilon_w(m) \\
&= \frac{1}{2^N}\sum_{w',m\in M}\left(1-\frac{2\rho|w'|}{N}\right)^t\varepsilon_{w'}(m(0)-m)\,\varepsilon_w(m) \\
&= \sum_{w'\in M}\left(1-\frac{2\rho|w'|}{N}\right)^t\delta_{w,w'}\varepsilon_{w'}(m(0)) = \left(1-\frac{2\rho|w|}{N}\right)^t\varepsilon_w(m(0)).
\end{aligned}$$

On obtient donc bien $(d_{\mathrm{VT}}(\pi_t,\nu))^2 \leq \frac{1}{4}\sum_{w\neq 00\cdots0}\left(1-\frac{2\rho|w|}{N}\right)^{2t}$. Pour tout $k\in[\![1,N]\!]$, il y a $\binom{N}{k}$ mots qui ont un poids k, donc, avec $\rho=\frac{1}{2}$,

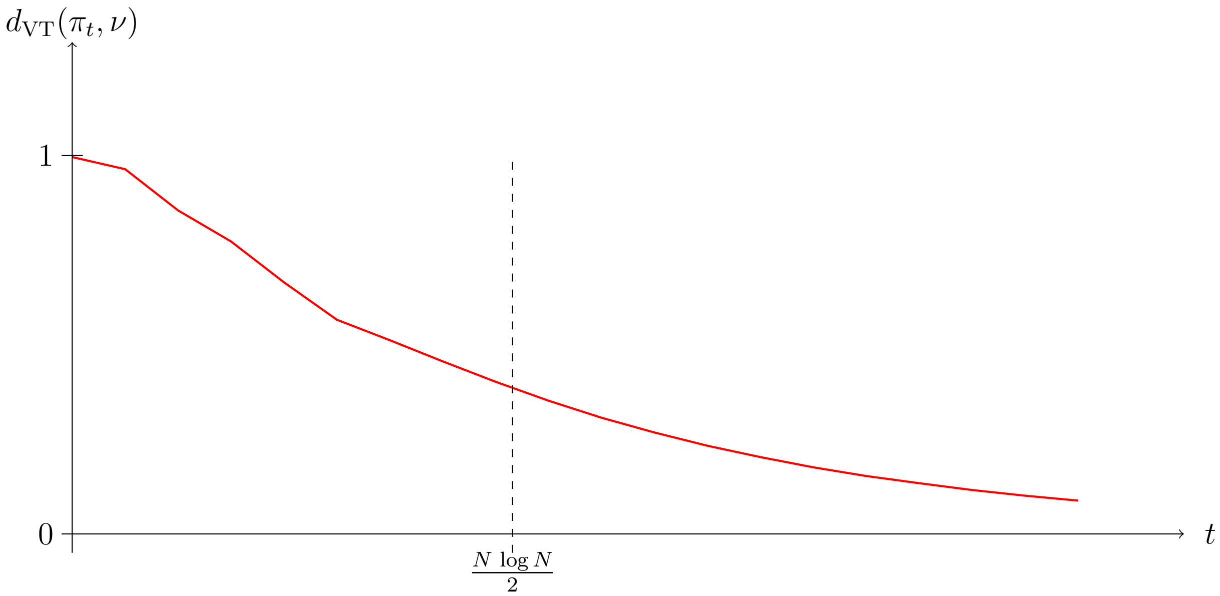
$$\begin{aligned}
(d_{\mathrm{VT}}(\pi_t,\nu))^2 &\leq \frac{1}{4}\sum_{k=1}^{N}\binom{N}{k}\left(1-\frac{k}{N}\right)^{2t}, \\
&\leq \frac{1}{4}\sum_{k=1}^{N}\frac{N^k}{k!}\,\mathrm{e}^{-\frac{2kt}{N}} \qquad \text{car } 1-\frac{k}{N}\leq \mathrm{e}^{-\frac{k}{N}} \text{ et } \binom{N}{k}\leq\frac{N^k}{k!}, \\
&\leq \frac{1}{4}\sum_{k=1}^{\infty}\frac{1}{k!}\left(\mathrm{e}^{(\log N-\frac{2t}{N})}\right)^k \quad \text{en rajoutant les termes } k>N.
\end{aligned}$$

II.5 On utilisera la formule exacte pour π_t :

```
epsilon(w,m):
    N = longueur de w
    retourner (-1)^(somme([w[i]*m[i] pour i entre 1 et N]))

VariationTotale(t,N):
    M = liste des 2^N mots de {0,1}^N
    résultat = 0
    pour m dans M:
        S = 0
        pour w dans M:
            ajouter à S la quantité (1-somme(w)/N)^t * epsilon(w,m)
        ajouter à résultat la valeur absolue de (S-1)
    retourner résultat/2^(N+1)
```

Pour $N = 8$, on obtient le graphe suivant :



La distance en variation totale est effectivement assez grande pour $t < \frac{N \log N}{2}$, et petite pour $t > \frac{N \log N}{2}$ (le phénomène serait plus visible avec des plus grandes valeurs de N).