Ecole Normale Supérieure

2017

Temps de mélange

Exercices 2

1 Distance en variation totale

On rappelle que si μ, ν sont deux mesures de probabilité sur Ω fini ou dénombrable, on définit

$$||\mu - \nu||_{\mathrm{TV}} = \sup_{A \subset \Omega} \mu(A) - \nu(A).$$

On a vu en cours des caractérisations alternatives de la distance en variation totale.

Par ailleurs, si P désigne le noyau de transition d'une chaîne X irréductible sur Ω, et qui possède une unique mesure stationnaire π, on définit pour $t \geq 0$,

$$d(t) = \max_{x \in \Omega} ||P^t(x, \cdot) - \pi||_{TV},$$

et enfin pour $\varepsilon > 0$,

$$t_{\mathrm{mix}}(\varepsilon) = \inf\{t \geq 0 : d(t) \leq \varepsilon\}, \qquad t_{\mathrm{mix}} = t_{\mathrm{mix}}(1/4).$$

Exercice 1 Vérifier que $d_{\mathrm{TV}}(\mu, \nu) = ||\mu - \nu||_{\mathrm{TV}}$ définit une distance sur l'espace $\mathbf{P}(\Omega)$ des mesures de probabilités sur Ω.

Il suffit de remarquer que puisque μ et ν sont des probas, pour tout $A \subset \Omega$, $\mu(A) - \nu(A) = -(\mu(A^c) - \nu(A^c))$, ce qui permet de voir que

$$||\mu - \nu||_{\mathrm{TV}} = \sup_{A \subset \Omega} \mu(A) - \nu(A) = \sup_{A \subset \Omega} |\mu(A) - \nu(A)|.$$

Symétrie, positivité, séparation, et inégalité triangulaire découlent alors facilement.

Exercice 2 Montrer que

$$||\mu - \nu||_{\mathrm{TV}} = 1 - \sum_{x \in \Omega} \min\{\mu(x), \nu(x)\}.$$

Notons que $\min\{\mu(x), \nu(x)\}$ vaut $\nu(x)$ lorsque $x \in B := \{y \in \Omega : \mu(y) \geq \nu(y)\}$ tandis que ce minimum vaut $\mu(x)$ lorsque $x \in B^c$. $||\mu - \nu||_{\mathrm{TV}} = \mu(B) - \nu(B)$. On a donc

$$\sum_{x \in \Omega} \mu(x) \wedge \nu(x) = \nu(B) + \mu(B^c) = \nu(B) + 1 - \mu(B) = 1 - ||\mu - \nu||_{\mathrm{TV}}.$$

Exercice 3 Soient $(X_n, n \geq 0)$ et X_∞ des variables aléatoires à valeurs dans $\mathbb{Z}$, $(\nu_n, n \geq 0)$ et ν_∞ leurs lois respectives. Montrer que

$$X_n \xrightarrow[n \to \infty]{\text{(loi)}} X_\infty \quad \Leftrightarrow \quad ||\nu_n - \nu_\infty||_{\mathrm{TV}} \xrightarrow[n \to \infty]{} 0.$$

Si $||\nu_n - \nu_\infty||_{\mathrm{TV}} = \frac{1}{2}\sum_{x\in\mathbb{Z}} |\nu_n(x) - \nu_\infty(x)|$ tend vers 0 alors pour tout $x \in \mathbb{Z}$, $\nu_n(x) \to \nu_\infty(x)$ et on a bien la convergence en loi de X_n vers X.
Inversement si X_n tend en loi vers X_∞, fixons $\varepsilon > 0$. Comme $\{X_n, X_\infty\}$ est tendue, on peut trouver K_ε tel que

$$\sup\{\mathbb{P}(|X_n| > K), \mathbb{P}(|X_\infty| > K), n \in \mathbb{N}\} \le \varepsilon/3.$$

Comme par ailleurs $\mathbb{P}(X_n = x) \to \mathbb{P}(X_\infty = x)$ pour tout entier x, on peut trouver un entier n_0 tel que pour tout $n \ge n_0$,

$$\sum_{|x|\le K} |\mathbb{P}(X_n = x) - \mathbb{P}(X_\infty = x)| \le \varepsilon/3,$$

et il découle que

$$2||\nu_n - \nu_\infty||_{\mathrm{TV}} = \sum_{x\in\mathbb{Z}} |\mathbb{P}(X_n = x) - \mathbb{P}(X_\infty = x)| \le \varepsilon,$$

comme souhaité.

Exercice 4 Soit $\Omega = \{1, 2\}$ et $P = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}$. Calculer $d(t)$ pour tout $t \ge 0$.
$d(0) = 1/2$, $d(t) = 0, t \ge 1$.

Exercice 5 Que vaut $||\mu - \nu||_{\mathrm{TV}}$ lorsque μ et ν sont des lois de Bernoulli de paramètres respectifs p, q ?
$2|p - q|$.

Exercice 6 On fixe $p \in [0, 1]$. Soit μ la loi d'une variable Bernoulli de paramètre p et ν celle d'une variable de Poisson, également de paramètre p.
Montrer que

$$||\mu - \nu||_{\mathrm{TV}} = p(1 - \exp(-p)).$$

En remarquant que $p + \exp(-p) - 1 \ge 0$, on obtient que $B := \{x \in \mathbb{N} : \mu(x) > \nu(x)\} = \{1\}$, et donc

$$||\mu - \nu||_{\mathrm{TV}} = \mu(1) - \nu(1) = p - p\exp(-p),$$

comme souhaité.

Exercice 7 Si μ_1, μ_2 sont des mesures sur $\mathbb{Z}$ on peut définir le produit de convolution

$$\mu_1 * \mu_2(x) = \sum_{y\in\mathbb{Z}} \mu_1(y)\mu_2(x - y), \quad x \in \mathbb{Z}$$

et la mesure produit (sur $\mathbb{Z}^2$)

$$\mu_1 \times \mu_2(x, y) = \mu_1(x)\mu_2(y), \quad (x, y) \in \mathbb{Z}^2.$$

Montrer alors que si $\mu_1, \mu_2, \nu_1, \nu_2$ sont des mesures sur $\mathbb{Z}$ on a

$$||\mu_1 * \mu_2 - \nu_1 * \nu_2||_{\mathrm{TV}} \le ||\mu_1 \times \mu_2 - \nu_1 \times \nu_2||_{\mathrm{TV}} \le ||\mu_1 - \nu_1||_{\mathrm{TV}} + ||\mu_2 - \nu_2||_{\mathrm{TV}}.$$

$$
\begin{aligned}
2||\mu_1 * \mu_2 - \nu_1 * \nu_2||_{\mathrm{TV}} &= \sum_{x\in\mathbb{Z}} \left| \sum_{y\in\mathbb{Z}} \mu_1(y)\mu_2(x-y) - \sum_{y\in\mathbb{Z}} \nu_1(y)\nu_2(x-y) \right| \\
&\le \sum_{x\in\mathbb{Z}} \sum_{y\in\mathbb{Z}} |\mu_1(y)\mu_2(x-y) - \nu_1(y)\nu_2(x-y)| \\
&= \sum_{y\in\mathbb{Z}} \sum_{z\in\mathbb{Z}} |\mu_1(y)\mu_2(z) - \nu_1(y)\nu_2(z)| \\
&= 2||\mu_1 \times \mu_2 - \nu_1 \times \nu_2||_{\mathrm{TV}} \\
&\le \sum_{x,y\in\mathbb{Z}} |\mu_1(x)\mu_2(y) - \nu_1(x)\mu_2(y)| + \sum_{x,y\in\mathbb{Z}} |\nu_1(x)\mu_2(y) - \nu_1(x)\nu_2(y)| \\
&= 2||\mu_1 - \mu_2||_{\mathrm{TV}} + 2||\nu_1 - \nu_2||_{\mathrm{TV}},
\end{aligned}
$$

où à la dernière ligne on a utilisé que μ_2, ν_1 sont des probas.

Exercice 8 On considère des nombres $\{p_{n,m} \in [0,1], n \in \mathbb{N}, 0 \le m \le n\}$ tels que

$$\lambda_n := \sum_{m=0}^{n} p_{n,m} \underset{n\to\infty}{\longrightarrow} \lambda > 0, \qquad \max_{0\le m\le n} p_{n,m} \underset{n\to\infty}{\longrightarrow} 0$$

On définit alors μ_n la loi de $\sum_{m=1}^{n} \xi_{n,m}$, où les variables $\xi_{n,m} \sim \mathrm{Ber}(p_{n,m})$ sont indépendantes. On note par ailleurs ν_n la loi d'une variable de Poisson de paramètre λ_n, et ν celle d'une variable de Poisson de paramètre λ.

1. A l'aide des exercices précédents établir que

$$||\mu_n - \nu_n||_{\mathrm{TV}} \le 2 \sum_{m=0}^{n} p_{n,m}(1 - \exp(-p_{n,m})).$$

2. En déduire que lorsque $n \to \infty$, $||\mu_n - \nu||_{\mathrm{TV}} \to 0$
3. Quel résultat classique cet exercice permet-il de généraliser ?

1. μ_n est la convolée des lois de n lois de Bernoulli de paramètres respectifs $p_{n,m}, 0 \le m \le n$; ν_n est la convolée de n lois de Poisson, également de paramètres respectifs $p_{n,m}$. L'exercice 7 permet d'affirmer que la distance entre les convolées est bornée par la somme sur $m \in \{0, ..., n\}$ des distances entre Bernoulli et Poisson de paramètre $p_{n,m}$. Or l'exercice 6 permet justement de borner la distance entre une loi de Bernoulli et une Poisson de même paramètre, ce qui amène au résultat souhaité.
2. En utilisant par exemple que $1 - \exp(-p_{n,m}) \le p_{n,m}$, on obtient

$$\sum_{m=0}^{n} p_{n,m}(1 - \exp(-p_{n,m})) \le \max_{0\le m\le n} p_{n,m} \sum_{m=0}^{n} p_{n,m} = \lambda_n \max_{0\le m\le n} p_{n,m},$$

qui tend clairement vers 0 d'après nos hypothèses.

3. Lorsque $p_{n,0} = 0, p_{n,m} = \lambda/n, 1 \le m \le n$, on retrouve le résultat bien connu qu'une variable Binômiale de paramètres $n, \lambda/n$ converge en loi vers une variable de Poisson de paramètre λ. Mais le résultat qu'on vient de démontrer est beaucoup plus général.

Exercice 9 Soit X, Y des variables binômiales de parmètres respectifs n, p et n, q avec $n \in \mathbb{N}^*$ et $0 \leq p \leq q \leq 1$. On note μ la loi de X, ν la loi de Y.

1. Montrer que si $k_0 = \left\lfloor n \frac{\log\left(\frac{1-p}{1-q}\right)}{\log\left(\frac{q(1-p)}{p(1-q)}\right)} \right\rfloor$, on a
$$||\mu - \nu||_{\mathrm{TV}} = \mathbb{P}(X \leq k_0) - \mathbb{P}(Y \leq k_0).$$
2. Dans cette question on suppose que $p = 1/2$ et que pour un $c \in \mathbb{R}_+$, $q = q(n) = \frac{1}{2} + \frac{c}{2\sqrt{n}}$. On note μ et ν_n les lois respectives de X, Y_n
Montrer alors que
$$k_0 = \frac{n}{2} + \frac{c\sqrt{n}}{4} + O(1),$$
puis que
$$\lim_{n\to\infty} ||\mu - \nu_n||_{\mathrm{TV}} = \mathbb{P}(Z \in [-c/2, c/2]),$$
où Z est une variable dont on précisera la loi.
3. Sans faire les calculs, se convaincre qu'on pourrait faire un raisonnement similaire et obtenir des limites non dégénérées (ni 0 ni 1) lorsque $a \in (0,1)$ est fixé, $p_n = a + \frac{c_1}{2\sqrt{n}}$ est fixé et $q_n = a + \frac{c_2}{2\sqrt{n}}$ (avec $c_1 < c_2$).
4. Qu'implique les questions précédentes lorsque p_n, q_n approchent la même valeur $a \in [0,1]$ avec $|q_n - p_n| << \sqrt{n}$? Et lorsque p_n, q_n approchent la même valeur a avec $|q_n - p_n| >> \sqrt{n}$?

1. Puisque $p < q$, il est clair que la fonction $x \to p^x(1-p)^{n-x} - q^x(1-q)^{n-x}$ décroît sur $[0, n]$, et s'annule en
$$x_0 = n \frac{\log\left(\frac{1-p}{1-q}\right)}{\log\left(\frac{q(1-p)}{p(1-q)}\right)}.$$
Ainsi $\mathbb{P}(X = k) - \mathbb{P}(Y = k)$ est positive ssi $k \leq k_0$, et le résultat souhaité découle alors de l'une des caractérisations de la distance en variation totale.
2. Le développement de k_0 est obtenu grâce à l'expression de la question précédente. Par ailleurs, le TCL permet d'affirmer que
$$\mathbb{P}(X \leq k_0) = \mathbb{P}\left(\frac{2X - n}{\sqrt{n}} \leq \frac{c}{2} + O(1/\sqrt{n})\right) \to \mathbb{P}(Z \leq c/2),$$
où $Z \sim \mathcal{N}(0,1)$. Par une généralisation du TCL (e.g. le TCL pour les martingales), en remarquant que $\mathbb{E}[Y_n] = \frac{n}{2} + \frac{c\sqrt{n}}{2}$, $\mathrm{Var}[Y_n] = \frac{n}{4} - \frac{c^2}{4}$, on a
$$\mathbb{P}(Y_n \leq k_0) = \mathbb{P}\left(\frac{2Y_n - n - c\sqrt{n}}{\sqrt{n}} \leq -\frac{c}{2} + O(1/\sqrt{n})\right) \to \mathbb{P}(Z \leq -c/2),$$
ce qui conduit au résultat souhaité.
3. Il est clair qu'on peut obtenir de manière similaire un DL pour le $k_0(n)$ correspondant. En utilisant par exemple le TCL pour les martingales, on va obtenir un résultat du même genre lorsque $X_n \sim \mathrm{Bin}(n, p_n), Y_n \sim \mathrm{Bin}(n, q_n)$, à savoir une limite non dégénérée de la forme $\mathbb{P}(Z \in [f(c_1), g(c_2)])$, avec $f(c_1), g(c_2) \in \mathbb{R}$ (en fait f et g sont même des fonctions affines).

4. Lorsque $|q_n - p_n| << \sqrt{n}$ la distance est majorée par le cas précédent avec c_1 arbitrairement proche de c_2, et on obtient donc une limite nulle pour $||\mu_n - \nu_n||_{\mathrm{TV}}$. A l'inverse lorsque $|q_n - p_n| >> \sqrt{n}$, la distance est minorée par le cas précédent avec c_1 arbitrairement éloigné de c_2, et on obtient donc que $||\mu_n - \nu_n||_{\mathrm{TV}} \to 1$.

Exercice 10 Soit X, Y des variables de Poisson de paramètres respectifs $\lambda, \lambda + \ell$ avec $\lambda > 0, \ell \geq 0$. On note μ la loi de X, ν la loi de Y.

1. Montrer que si $k_0 = \left\lfloor \frac{\ell}{\log\left(1+\frac{\ell}{\lambda}\right)} \right\rfloor$, on a
$$||\mu - \nu||_{\mathrm{TV}} = \mathbb{P}(X \leq k_0) - \mathbb{P}(Y \leq k_0).$$
2. Dans cette question on va considérer l'asymptotique $\lambda \to \infty$. Montrer que si $\ell = c\sqrt{\lambda} + o(\lambda^{1/4})$, avec $c > 0$, on a
$$k_0 = \lambda + \left(\frac{c}{2}\right)\sqrt{\lambda} + o(\sqrt{\lambda}),$$
et en déduire que
$$\lim_{\lambda\to\infty} ||\mu - \nu||_{\mathrm{TV}} = \mathbb{P}(Z \in I(c)),$$
où on précisera la loi de Z, et l'intervalle $I(c)$.

Par un raisonnement similaire à celui de l'exercice précédent on obtient le résultat souhaité avec $Z \sim \mathcal{N}(0,1)$ et $I(c) = [-c/2, c/2]$

Exercice 11

1. Montrer que pour tout $t \geq 0$
$$d(t) = \sup_{\mu\in\mathbf{P}(\Omega)} ||\mu P^t - \pi||_{\mathrm{TV}},$$
puis que
$$d(t) = \sup_{\mu\in\mathbf{P}(\Omega), A\subset\Omega} \sum_{x\in\Omega, y\in A} (\mu(x) - \pi(x))P^t(x,y).$$
2. Montrer que $\overline{d}(t) := \max_{x,y\in\Omega} ||P^t(x,\cdot) - P^t(y,\cdot)||_{\mathrm{TV}}$ vérifie, pour tout $t \geq 0$,
$$\overline{d}(t) = \sup_{\mu\in\mathbf{P}(\Omega)\nu\in\mathbf{P}(\Omega)} ||\mu P^t - \nu P^t||_{\mathrm{TV}},$$
et en déduire une égalité similaire à celle obtenue pour $d(t)$ dans la question précédente.

1. L'inégalité $\leq$ et évidente puisque $\mathbf{P}(\Omega) \supset \{\delta_x, x \in \Omega\}$.
Par ailleurs, pour tout $\mu \in \mathbf{P}(\Omega)$, on a
$$\begin{aligned} ||\mu P^t - \pi||_{\mathrm{TV}} &= \frac{1}{2}\sum_{y\in\Omega}\left|\sum_{x\in\Omega}\mu(x)P^t(x,y) - \pi(y)\right| \\ &\leq \frac{1}{2}\sum_{y\in\Omega}\sum_{x\in\Omega}\mu(x)|P^t(x,y) - \pi(y)| \leq d(t). \end{aligned}$$

Puisque π est stationnaire on a $\pi = \pi P^t$ et donc par ce qui précède et la définition de la variation totale, on obtient

$$d(t) = \sup_{\mu \in \mathbf{P}(\Omega), A \subset \Omega} \mu P^t(A) - \pi P^t(A),$$

ce qui est le résultat souhaité.

2. Comme dans la première question l'inégalité $\leq$ est évidente. Par ailleurs pour tous $\mu, \nu \in \mathbf{P}(\Omega)$, on a

$$\begin{aligned} ||\mu P^t - \nu P^t||_{\mathrm{TV}} &= \frac{1}{2} \sum_{y \in \Omega} \left| \sum_{x \in \Omega} \mu(x) P^t(x,y) - \sum_{z \in \Omega} \nu(z) P^t(z,y) \right| \\ &\leq \frac{1}{2} \sum_{y \in \Omega} \sum_{z \in \Omega} \mu(x)\nu(y) \sum_{y \in \Omega} \left| P^t(x,y) - P^t(z,y) \right| \leq \overline{d}(t). \end{aligned}$$

On obtient alors que

$$\overline{d}(t) = \sup_{\mu,\nu \in \mathbf{P}(\Omega), A \subset \Omega} \sum_{x \in \Omega, y \in A} (\mu(x) - \nu(x)) P^t(x,y).$$

2 Autres distances

Exercice 12 Soit, pour un $t \geq 1$, $\overline{d}_t^P(\mu,\nu) = ||\mu P^t - \nu P^t||_{TV}$.

1. Montrer que $\overline{d}_t^P$ est une pseudo-distance sur $\mathcal{P}(\Omega)$.
2. Donner un exemple de chaîne et d'un $t \geq 1$ pour laquelle $\overline{d}_t^P$ n'est pas une distance.
3. Donner un exemple de chaîne pour laquelle $\overline{d}_t$ est une distance pour tout $t \geq 1$.
4. Montrer que $t \to \overline{d}_t^P(\mu,\nu)$ est décroissante.
5. Montrer que $t \to d(t)$ est décroissante.

1. La positivité, la symétrie, l'inégalité triangulaire découlent de ces propriétés pour d_{TV}.
2. Pour $P = \begin{pmatrix} 1-p & p \\ 1-p & p \end{pmatrix}$ on a vu que quelque soit μ, la loi de X_1 est toujours la même, de sorte que dans ce cas, quelque soient μ, ν, et $t \geq 1$, $d_t^P(\mu,\nu) = 0$, et on n'a donc dans ce cas évidemment affaire qu'à une pseudo-distance, pas une distance.
3. Pour $P = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix}$ avec $p + q \neq 1$, on peut facilement montrer que $\mu \neq \nu \Rightarrow \mu P \neq \nu P$, et donc $\mu \neq \nu \Rightarrow \mu P^t \neq \nu P^t$, dans ce cas d_t^P vérifie la séparation et est donc une distance.
4. Supposons que (X_t, Y_t) réalise un couplage optimal de μP^t et νP^t. Sachant X_t, Y_t, on peut alors définir X_{t+1}, Y_{t+1} de la façon suivante :
   – Si $X_t \neq Y_t$, $X_{t+1} \sim P(X_t, \cdot)$ et indépendamment $Y_{t+1} \sim P(Y_t, \cdot)$.
   – Si $X_t = Y_t$, $X_{t+1} = Y_{t+1} \sim P(X_t, \cdot)$.

Les variables (X_{t+1}, Y_{t+1}) réalisent un couplage de μP^{t+1} et de νP^{t+1}, et on obtient

$$d^P_{t+1}(\mu,\nu) \leq \mathbb{P}(X_{t+1} \neq Y_{t+1}) \leq \mathbb{P}(X_t \neq Y_t) = d^P_t(\mu,\nu).$$

5. On a
$$d(t) = \max_{x\in\Omega} d^P_t(\delta_x, \pi).$$
Le maximum pour $d(t+1)$ est réalisé, disons, en x_{t+1}, et la question précédente entraîne alors que
$$d(t+1) = d^P_{t+1}(\delta_{x_{t+1}}, \pi) \leq d^P_t(\delta_{x_{t+1}}, \pi) \leq d(t).$$

Exercice 13 Si π est la distribution stationnaire d'une chaîne X sur Ω, pour une fonction $f : \Omega \to \mathbb{R}$, on peut définir, pour $p > 0$, sa norme $\ell^p(\pi)$:

$$||f||_p = \left[\sum_{x\in\Omega} |f(x)|^p\, \pi(x)\right]^{1/p}.$$

sa norme $\ell^\infty(\pi)$ étant $||f||_\infty = \max_{x\in\Omega} |f(x)|$.
Le cas $p = 2$ est particulièrement intéressant puisque

$$\langle f, g\rangle_\pi := \sum_{x\in\Omega} f(x)g(x)\pi(x)$$

définit un produit scalaire sur l'espace $\{f : \Omega \to \mathbb{R}\}$.

1. Si ν et μ sont deux mesures de probabilité sur Ω que vaut
$$\left\|\frac{\nu}{\pi} - \frac{\mu}{\pi}\right\|_1 \quad ?$$
2. Soit $f : \Omega \to \mathbb{R}$. Etablir que $||f||_p$ croît avec $p > 0$.
3. Montrer que
$$4\left\|P^t(x,\cdot) - \pi\right\|^2_{\mathrm{TV}} \leq \left\|\frac{P^t(x,\cdot)}{\pi} - 1\right\|^2_2.$$
4. Montrer que
$$d(t) \leq \frac{1}{2}\max_{x,y\in\Omega}\left|\frac{P^t(x,y)}{\pi(y)} - 1\right|.$$

1. Clairement
$$\left\|\frac{\nu}{\pi} - \frac{\mu}{\pi}\right\|_1 = 2||\mu - \nu||_{TV}.$$
2. Supposons $q > p \geq 1$, et notons P_π la loi de $X \sim \pi$:
$$\begin{aligned} ||f||_q^p &= \left(\sum_{x\in\Omega} |f(x)|^q \pi(x)\right)^{p/q} \\ &= \left(E_\pi\left[|f(X)|^q\right]\right)^{p/q} \\ &\geq E_\pi\left[|f(X)|^p\right] = ||f||_p^p \end{aligned}$$
où l'inégalité de la dernière ligne ci-dessus est due à Jensen avec $\phi : x \to x^{p/q}$

3. D'après la première question

$$4\left\|P^t(x,\cdot)-\pi\right\|_{\mathrm{TV}}^2=\left\|\frac{P^t(x,\cdot)}{\pi}-1\right\|_1^2,$$

et d'après la deuxième ceci est borné par $\left\|\frac{P^t(x,\cdot)}{\pi}-1\right\|_2^2$, comme souhaité.

4. En notant que $||f||_\infty=\lim_{p\to\infty}||f||_p$ on voit qu'on peut étendre le résultat de la deuxième question à $p\in\overline{\mathbb{R}}_+^*$. On en déduit

$$2d(t)=\max_{x\in\Omega}\left\|\frac{P^t(x,\cdot)}{\pi}-1\right\|_1\leq\max_{x\in\Omega}\left\|\frac{P^t(x,\cdot)}{\pi}-1\right\|_\infty=\max_{x,y\in\Omega}\left|\frac{P^t(x,y)}{\pi(y)}-1\right|,$$

comme souhaité.

Exercice 14 La distance de Hellinger entre deux probabilités sur Ω, μ et ν, est donnée par

$$d_H(\mu,\nu)=\sqrt{\sum_{x\in\Omega}(\sqrt{\mu(x)}-\sqrt{\nu(x)})^2}$$

1. Fixons $x_0\in\Omega$ et introduisons $g_{x_0,t}(x)=\frac{P^t(x_0,x)}{\pi(x)}, x\in\Omega$. Etablir que

$$d_H(P^t(x_0,\cdot),\pi)=||\sqrt{g_{x_0,t}}-1||_2.$$

2. Montrer que

$$||\mu-\nu||_{\mathrm{TV}}\leq d_H(\mu,\nu).$$

3. En déduire que

$$t_{\mathrm{mix}}(\varepsilon)\leq\inf\{t\geq 0:\quad\forall x_0\in\Omega\quad d_H(P^t(x_0,\cdot),\pi)\leq\varepsilon\}.$$

1. D'après la définition de $g_{x_0,t}$, on a

$$\begin{aligned}||\sqrt{g_{x_0,t}}-1||_2 &= \sqrt{\sum_{x\in\Omega}\left(\sqrt{\frac{P^t(x_0,x)}{\pi(x)}}-1\right)^2\pi(x)}\\ &= \sqrt{\sum_{x\in\Omega}\left(\sqrt{P^t(x_0,x)}-\sqrt{\pi(x)}\right)^2}=d_H(P^t(x_0,\cdot),\pi).\end{aligned}$$

2. On a

$$2||P^t(x_0,\cdot)-\pi||_{\mathrm{TV}}=||g_{x_0,t}-1||_1=\sum_{x\in\Omega}\pi(x)\left|\sqrt{g_{x_0,t}(x)}-1\right|\left(\sqrt{g_{x_0,t}}+1\right).$$

Par Cauchy-Schwartz, ceci est borné par

$$\sqrt{\sum_{x\in\Omega}\pi(x)\left(\sqrt{g_{x_0,t}(x)}-1\right)^2}\sqrt{\sum_{x\in\Omega}\pi(x)\left(\sqrt{g_{x_0,t}(x)}+1\right)^2}.$$

Le premier terme du produit ci-dessus est exactement $||\sqrt{g_{x_0,t}} - 1||_2 = d_H(P^t(x_0,\cdot),\pi)$ d'après la première question. Quant au second terme du produit, on a

$$\sum_{x\in\Omega} \pi(x) \left(\sqrt{g_{x_0,t}(x)} - 1\right)^2 = 2 + 2\sum_{x\in\Omega} \pi(x)\sqrt{g_{x_0,t}}.$$

A nouveau par Cauchy-Schwartz,

$$\sum_{x\in\Omega} \pi(x)\sqrt{g_{x_0,t}} \leq 1,$$

on conclut donc que

$$2||\mu - \nu||_{\mathrm{TV}} \leq 2d_H(P^t(x_0,\cdot),\pi).$$

3. évident d'après la question précédente et la définition de $t_{\mathrm{mix}}(\varepsilon)$.

3 Chaîne retournée dans le temps

Soit X est une chaîne de Markov irréductible, de distribution stationnaire π. On définit pour $x, y \in \Omega$,

$$\widehat{P}(x,y) := \frac{\pi(y)P(y,x)}{\pi(x)}.$$

Exercice 15

1. Montrer que $\widehat{P}$ est bien définie, et que c'est le noyau de transition d'une chaîne $\widehat{X}$ sur Ω, également irréductible. On note $\widehat{\mathbb{P}}_\mu$ la loi de cette chaîne issue de μ.
2. Montrer que π est également une distribution stationnaire de $\widehat{X}$.
3. Montrer que pour tous $t \geq 0, x_0, ..., x_t \in \Omega$, on a

$$\mathbb{P}_\pi(X_0 = x_0, ..., X_t = x_t) = \widehat{\mathbb{P}}_\pi\left(\widehat{X}_0 = x_t, ..., \widehat{X}_t = x_0\right).$$

4. Que se passe-t-il lorsque X est réversible ?

1. X irréductible sur Ω fini ou dénombrable implique que $\pi(x) > 0$ pour tout $x \in \Omega$, comme on l'a vu dans la feuille 1, ceci permet d'assurer que $\widehat{P}(x,y)$ est bien défini pour tous $x, y \in \Omega$.
   Pour vérifier que c'est un noyau de transition, fixons $x \in \Omega$ et calculons

$$\begin{aligned}\sum_{y\in\Omega} \widehat{P}(x,y) &= \sum_{y\in\Omega} \frac{\pi(y)P(y,x)}{\pi(x)} \\ &= \frac{1}{\pi(x)}\pi P(x) = 1,\end{aligned}$$

   où on a utilisé que π est stationnaire.
   Enfin remarquons que puisque $\pi(x) > 0$ pour tout $x \in \Omega$, $P^r(x,y) > 0$ implique $\widehat{P}^r(y,x) > 0$ également, de sorte que l'irréductibilité de X entraîne celle de $\widehat{X}$.

2. On a

$$\begin{aligned}\pi\widehat{P}(y) &= \sum_{x\in\Omega}\pi(x)\widehat{P}(x,y)\\ &= \sum_{x\in\Omega}\pi(y)P(y,x)=\pi(y),\end{aligned}$$

et donc π est également stationnaire pour $\widehat{P}$.

3. On a, en utilisant la définition de $\widehat{P}$,

$$\begin{aligned}\widehat{\mathbb{P}}_\pi\left(\widehat{X}_0=x_t,...,\widehat{X}_t=x_0\right) &= \pi(x_t)\widehat{P}(x_t,x_{t-1})...\widehat{P}(x_1,x_0)\\ &= P(x_{t-1},x_t)...P(x_0,x_1)\pi(x_0)\\ &= \mathbb{P}_\pi\left(X_0=x_0,...,X_t=x_t\right)\end{aligned}$$

4. Lorsque X est réversible, $\widehat{P}=P$, de sorte que les lois de X sous P_μ et de $\widehat{X}$ sous $\widehat{\mathbb{P}}_\mu$ coïncident.

Exercice 16 Dans cet exercice on suppose que X est une marche sur le groupe G (fini) dont la distribution d'un pas est dictée par la distribution μ. Précisément, le noyau P de X est tel que

$$P(g,hg)=\mu(h)\qquad \forall g,h\in G.$$

On suppose que $\mathcal{H}=\{h:\mu(h)>0\}$ engendre G. On note $\widehat{P}$ le noyau de la chaîne retourné, $\widehat{d}(t)=\max_{x\in\Omega}||\widehat{P}^t(x,\cdot)-\pi||_{TV}$ et $\widehat{t_{\text{mix}}}(\varepsilon):=\inf\{t\geq 0:\widehat{d}(t)\leq\varepsilon\}$.

1. Montrer que la chaîne retournée reste une marche sur le groupe G, et décrire la distribution $\widehat{\mu}$ correspondante.
2. Montrer que pour tout $g\in G, t\geq 0$ on a

$$P^t(id,g)=\widehat{P}^t(id,g^{-1}).$$

En déduire que pour tout $\varepsilon>0$, $t_{\text{mix}}(\varepsilon)=\widehat{t_{\text{mix}}}(\varepsilon)$.

1. D'après un exercice de la feuille 1, on a affaire à une marche irréductible sur G, l'unique distribution stationnaire est uniforme sur G. La chaîne retournée est donc une marche sur G dont la distribution de pas est dictée par $\widehat{\mu}$ telle que $\widehat{\mu}(h)=\mu(h^{-1}), h\in G$, en effet

$$\widehat{P}(g,hg)=P(hg,g)=\mu(h^{-1})\qquad\forall g,h\in G.$$

2. Pour $g\in G$, notons $A_t(g):=\{(x_1,...,x_t)\in G:x_1...x_t=g\}$. Observons que $(x_1,...,x_t)\in A_t(g)$ est équivalent à $(x_t^{-1},...,x_1^{-1})\in A_t(g^{-1})$, et donc

$$\begin{aligned}P^t(id,g) &= \sum_{(x_1,...,x_t)\in A_t(g)}\prod_{i=1}^t\mu(x_i)\\ &= \sum_{(x_1,...,x_t)\in A_t(g)}\prod_{i=1}^t\widehat{\mu}(x_i^{-1})\\ &= \sum_{(y_1,...,y_t)\in A_t(g^{-1})}\prod_{i=1}^t\widehat{\mu}(y_i)=\widehat{P}(id,g^{-1})\end{aligned}$$

Reste alors à voir que

$$\begin{aligned} d(t) = 2\|P^t(id,\cdot) - \pi\|_{\mathrm{TV}} &= \sum_{g\in G}\left|P^t(id,g) - \frac{1}{|G|}\right| \\ &= \sum_{g\in G}\left|\widehat{P}^t(id,g^{-1}) - \frac{1}{|G|}\right| \\ &= \sum_{g'\in G}\left|\widehat{P}^t(id,g') - \frac{1}{|G|}\right| = 2\|\widehat{P}^t(id,\cdot) - \pi\|_{\mathrm{TV}} = \widehat{d}(t). \end{aligned}$$

4 Couplages : deux exemples

Un couplage $\{(X_t, Y_t), t \geq 0\}$ avec noyau de transition P est tel que sous $\mathbb{P}_{x,y}$, pour tout $t \geq 0$, $X_t \sim P^t(x,\cdot), Y_t \sim P^t(y\cdot)$, on note $\tau_{\text{couple}} := \inf\{t \geq 0 : X_t = Y_t\}$. On verra en cours que

$$d(t) \leq \overline{d}(t) \leq \max_{x,y\in\Omega} \mathbb{P}_{x,y}(\tau_{\text{couple}} > t).$$

Exercice 17 On considère le couplage suivant de deux marches paresseuses sur le tore discret $(\mathbb{Z}/n\mathbb{Z})^d$. Au temps $t+1$:

– Indépendamment, et indépendamment des étapes précédentes, on tire $i_{t+1} \sim \mathrm{Unif}\{1,...,d\}$, $\xi_{t+1} = \begin{cases} 1 & \text{avec proba } 1/4 \\ -1 & \text{avec proba } 1/4 \\ 0 & \text{avec proba } 1/2 \end{cases}$, $B_{t+1} \sim \mathrm{Ber}(1/2)$, et enfin $\zeta_{t+1} = \begin{cases} 1 & \text{avec proba } 1/2 \\ -1 & \text{avec proba } 1/2 \end{cases}$.

– Si $X_t^{i_{t+1}} = Y_t^{i_{t+1}}$, on pose $X_{t+1}^{i_{t+1}} = Y_{t+1}^{i_{t+1}} = X_t^{i_{t+1}} + \xi_{t+1}$.

– Si $X_t^{i_{t+1}} \neq Y_t^{i_{t+1}}$, on pose

$$X_{t+1}^{i_{t+1}} = X_t^{i_{t+1}} + B_{t+1}\zeta_{t+1}, \qquad Y_{t+1}^{i_{t+1}} = Y_t^{i_{t+1}} + (1 - B_{t+1})\zeta_{t+1}.$$

– On laisse les autres coordonnées des deux chaînes inchangées.

1. Montrer que $\{(X_t, Y_t), t \geq 0\}$ est un couplage de marches paresseuses sur le tore discret de dimension d.
2. Soit $i \in \{1,...,d\}$ et $Z_t^i := X_t^i - Y_t^i, t \geq 0$. Montrer que $(Z_t^i, t \geq 0)$ est une chaîne de Markov sur $\mathbb{Z}/n\mathbb{Z}$ dont on précisera le noyau de transition.
3. Estimer $\max_{x,y\in\Omega} \mathbb{E}_{x,y}[\tau_i]$ où $\tau_i = \inf\{t \geq 0 : Z_t^i = 0\}$.
4. En déduire finalement que pour la marche paresseuse sur $(\mathbb{Z}/n\mathbb{Z})^d$ on a

$$t_{\text{mix}}(\varepsilon) \leq d^2 n^2 \left\lceil \log_2\left(\frac{1}{\varepsilon}\right)\right\rceil.$$

1. Conditionnellement à $(X_t, Y_t) = (x, y)$, on a

$$\begin{aligned} X_{t+1} &= \sum_{i=1}^{d} \mathbb{1}_{i_t=i} X_t \left(\mathbb{1}_{\{x^i=y^i,\xi_{t+1}=0\}} + \mathbb{1}_{\{B_{t+1}=0,x^i\neq y^i\}}\right) \\ &+ \sum_{i=1}^{d} \mathbb{1}_{i_t=i}(X_t+e_i) \left(\mathbb{1}_{\{\xi_{t+1}=1,x^i=y^i\}} + \mathbb{1}_{\{B_{t+1}=1,\zeta_{t+1}=1,x^i\neq y^i\}}\right) \\ &+ \sum_{i=1}^{d} \mathbb{1}_{i_t=i}(X_t-e_i) \left(\mathbb{1}_{\{\xi_{t+1}=-1,x^i=y^i\}} + \mathbb{1}_{\{B_{t+1}=1,\zeta_{t+1}=-1,x^i\neq y^i\}}\right) \end{aligned}$$

   Pour déterminer la loi de X_{t+1}, notons que $i_{t+1}, \xi_{t+1}, B_{t+1}, \zeta_{t+1}$ sont indépendantes, et indépendantes des étapes précédentes, donc leur lois respectives ne sont pas affectées par le choix de x, y.
   Or $\mathbb{P}(\xi_{t+1} = 0) = \mathbb{P}(B_{t+1} = 0) = 1/2$ de sorte que $\mathbb{P}(X_{t+1} = X_t) = 1/2$. Similairement $\mathbb{P}(i_t = i, \xi_{t+1} = 1) = \mathbb{P}(i_t = i, B_{t+1} = 1, \zeta_{t+1} = 1) = \frac{1}{4d}$ de sorte que $\mathbb{P}(X_{t+1} = X_t + e_i) = \frac{1}{4d}, i = 1, ..., d$, et de même $\mathbb{P}(X_{t+1} = X_t - e_i) = \frac{1}{4d}, i = 1, ..., d$.
   Ainsi $(X_t)_{t\geq 0}$ est bien une marche paresseuse sur $(\mathbb{Z}/n\mathbb{Z})^d$. La preuve pour $(Y_t)_{t\geq 0}$ est en tous points similaire.
2. Pour $k \in \mathbb{Z}/n\mathbb{Z}$, travaillons conditionnellement à $\{Z^i_t = k\}$ et notons $\mathbb{Q}^i_k$ la mesure correspondante. Si $k = 0 = n$, les deux marches évoluent ensemble et donc $\mathbb{Q}^i_0(Z^i_{t+1} = Z^i_t = 0) = 1$. En revanche, si $k \in \{1, ..., n-1\}$

$$\begin{aligned} &\mathbb{Q}^i_k(Z^i_{t+1} = Z^i_t) = \frac{d-1}{d}, \\ &\mathbb{Q}^i_k(Z^i_{t+1} = Z^i_t + 1) = \mathbb{Q}^i_k(X^i_{t+1} = X^i_t + 1, Y^i_{t+1} = Y^i_t) + \mathbb{Q}^i_k(X^i_{t+1} = X^i_t, Y^i_{t+1} = Y^i_t - 1) = \frac{1}{2d} \\ &\mathbb{Q}^i_k(Z^i_{t+1} = Z^i_t - 1) = \frac{1}{2d}. \end{aligned}$$

   Ainsi $(Z^i_t)_{t\geq 0}$ est une marche simple symétrique, $\frac{d-1}{d}$-paresseuse sur $\{0, ..., n\}$, absorbée en $0, n$.
3. On a affaire au problème classique de la ruine du joueur, et on sait que $\mathbb{E}_{x,y}[\tau_i] = d(x_i - y_i)^2$ (ceci provient du fait que pour la marche simple S sur $\mathbb{Z}$, $(S^2_t - t)_{t\geq 0}$ est une martingale) qui est maximum pour $(x_i - y_i)^2 = \frac{n^2}{4}$. On conclut que $\max_{x,y\in\Omega} \mathbb{E}_{x,y}[\tau_i] = \frac{n^2 d}{2}$.
4. De la question précédente on déduit que

$$\max_{i\in\{1,...,d\}} \max_{x,y\in\Omega} \mathbb{E}_{x,y}[\tau_i] \leq \sum_{i=1}^{d} \max_{x,y\in\Omega} \mathbb{E}_{x,y}[\tau_i] \leq \frac{d^2 n^2}{4}.$$

   Remarque : Un peu plus de travail permettrait d'obtenir une borne en $Cd\log(d)n^2$.

   Comme $\tau_{\text{couple}} = \max_{i\in\{1,...,d\}} \tau_i$, la formule du cours et l'inégalité de Markov permettent de conclure que

$$d(t) \leq \max_{x,y\in\Omega} \mathbb{P}_{x,y}(\tau_{\text{couple}} > t) \leq \max_{x,y\in\Omega} \max_{1\leq i\leq d} \frac{\mathbb{E}_{x,y}[\tau_i]}{t} \leq \frac{d^2 n^2}{4t},$$

d'où $t_{\text{mix}} \leq d^2n^2$, et finalement grâce au résultat découlant de la sous-multiplicativité de $\overline{d}(t)$, on conclut au résultat souhaité.

Exercice 18 On considère $(X_t, t \geq 0)$ la marche simple, paresseuse, sur l'arbre binaire enraciné de profondeur k. On note $|X_t|$ la profondeur de X_t.

1. Montrer que $(|X_t|, t \geq 0)$ reste une chaîne de Markov dont on déterminera le noyau de transition.
2. Déterminer les mesures stationnaires respectives de X, $|X|$.
3. Soit $\tau_j = \inf\{t \geq 0 : |X_t| = k\}$. Estimer $E = \mathbb{E}_0[\tau_k] + \mathbb{E}_k[\tau_0]$.
4. On définit $(X_t, Y_t)_{t\geq 0}$ itérativement, comme suit :
   – Si $|X_t| = |Y_t|$, alors X et Y effectuent au temps $t+1$ un mouvement similaire dans l'arbre.
   Par exemple si elles se trouvent à une profondeur $i \in \{1, ..., k-1\}$, alors avec proba $1/2$, les deux chaînes restent sur place, avec proba $1/6$, X_{t+1} est le parent de X_t et Y_{t+1} celui de Y_t, avec proba $1/6$, X_{t+1} est le premier descendant de X_t, Y_{t+1} celui de Y_t, enfin avec proba $1/6$, X_{t+1} est le deuxième descendant de X_t, Y_{t+1} celui de Y_t.
   Si elles se trouvent toutes deux à profondeur k, avec proba $1/2$ elles restent toutes deux sur place, et avec proba $1/2$ elles remontent au parent correspondant.
   Enfin si elles se trouvent toutes deux à profondeur 0, on peut les faire évoluer ensemble.
   – Si $|X_t| \neq |Y_t|$, on tire $B_{t+1} \sim \text{Ber}(1/2)$. Si $B_{t+1} = 0$, X_{t+1} est l'un des voisins de X_t choisi uniformément, tandis que $Y_{t+1} = Y_t$, et si $B_{t+1} = 1$, $X_{t+1} = X_t$ tandis que Y_{t+1} est choisi uniformément parmi les voisins de Y_t.

   Montrer que $\{(X_t, Y_t), t \geq 0\}$ est un couplage de deux marches simples paresseuses sur l'arbre binaire.
5. Expliquer pourquoi $\mathbb{E}[\tau_{\text{couple}}] \leq E$. En déduire une borne supérieure sur $t_{\text{mix}}(\varepsilon)$ pour la marche simple paresseuse sur l'arbre binaire de profondeur k.

1. $(|X_t|)_{t\geq 0}$ est une chaîne projetée de X, à valeurs dans $\{0, ..., k\}$. Son noyau de transition Q vérifie
$$Q(i,i) = 1/2, i \in \{0, ..., k\}, \quad Q(0,1) = 1/2, Q(k,k-1) = 1/2,$$
$$Q(i,i-1) = 1/6, i \in \{1, ..., k-1\}, \quad Q(i,i+1) = 1/3, i \in \{1, ..., k-1\}$$
   Ainsi, $|X|$ est une marche paresseuse, asymétrique (biais $1/3$ vers k), sur $\{0, ..., k\}$, réfléchie en $0, k$.
2. La chaîne X est une marche simple (paresseuse) sur l'arbre binaire $\mathcal{T}$, si d_x désigne le degré d'un noeud $x \in \mathcal{T}$, notons que $d = \sum_{x\in\mathcal{T}} d_x = 2 + 3\sum_{i=1}^{k-1} 2^i + 2^k = 2^{k+2} - 4$, et la mesure stationnaire π est donc donnée par $\pi(x) = \frac{d_x}{2^{k+2}-4}$.
   Quant à la chaîne $|X|$, sa mesure stationnaire λ peut être calculée directement, ou bien à partir de π. On trouve que
$$\lambda(0) = \pi(0) = \frac{1}{2^{k+1}-2}, \quad \lambda(i) = \frac{3 \times 2^{i-2}}{2^k - 1}, i \in \{1, ..., k-1\} \quad \lambda(k) = \frac{2^{k-2}}{2^k - 1}.$$

3. Posons $u_i = \mathbb{E}_{i-1}[\tau_k] - \mathbb{E}_i[\tau_k], i = 1, ..., k$. Par Markov simple au temps 1, on trouve

$$\mathbb{E}_0[\tau_k] = 1 + \frac{1}{2}\mathbb{E}_1[\tau_k] + \frac{1}{2}\mathbb{E}_0[\tau_k], \text{ et donc } u_1 = 2,$$
$$\mathbb{E}_i[\tau_k] = 1 + \frac{1}{6}\mathbb{E}_{i-1}[\tau_k] + \frac{1}{3}\mathbb{E}_{i+1}[\tau_k] + \frac{1}{2}\mathbb{E}_i[\tau_k], \text{ et donc } u_{i+1} = \frac{1}{2}u_i + 3, i \in \{1, ..., k-1\}$$

On en déduit que

$$u_i = -2^{3-i} + 6, \text{ et donc } \mathbb{E}_0[\tau_k] = \sum_{i=1}^{k} u_i = 6k - 2^3(1 - 2^{-k}).$$

De même en posant $v_i = \mathbb{E}_i[\tau_0] - \mathbb{E}_{i-1}[\tau_0]$, il vient

$$v_k = 2, \; v_i = 2v_{i+1} + 6, i \in \{1, ..., k-1\}$$

et donc $v_{k-i} = 2^{3+i} - 6$,

$$\mathbb{E}_k[\tau_0] = \sum_{i=0}^{k-1} v_{k-i} = 2^{3+k} - 8 - 6k.$$

Finalement

$$E = -2^3(1 - 2^{-k}) + 2^{3+k} - 8 = 2^{3+k} - 16 + 2^{3-k}.$$

On notera au passage que $n = 2^{k+1} - 1$ est le nombre de noeuds de notre arbre, on a donc trouvé que $E = 4n - 12 + O(1/n)$.

4. Considérons X :

$$\begin{aligned}\mathbb{P}(X_{t+1} = X_t) &= \mathbb{P}(X_{t+1} = X_t \mid |X_t| = |Y_t|)\mathbb{P}(|X_t| = |Y_t|) + \mathbb{P}(X_{t+1} = X_t \mid |X_t| \neq |Y_t|)\mathbb{P}(|X_t| \neq |Y_t|) \\ &= \frac{1}{2}\mathbb{P}(|X_t| = |Y_t|) + \mathbb{P}(B_{t+1} = 1)\mathbb{P}(|X_t| \neq |Y_t|) = \frac{1}{2}.\end{aligned}$$

Par ailleurs si $X_{t+1} \neq X_t$, il est facile de voir qu'il est choisi uniformément parmi les voisins de X_t. Ainsi X est bien une marche paresseuse sur $\mathcal{T}$. Le raisonnement pour Y est identique.

5. Supposons que X, Y sont respectivement issus de x, y. Sans perte de généralité (X et Y jouant des rôles symétriques) on peut supposer que $|x| \leq |y|$. Notons que notre couplage impose que $|X_t| \leq |Y_t|$ pour tout $t \geq 0$.
Lorsque X atteint une des feuilles de l'arbre (i.e. lorsque $|X| = k$) on a forcément $|X_t| \geq |Y_t|$ et donc $|X_t| = |Y_t|$. Les chaînes projetées sont donc couplées à partir de τ_k.
Notons alors $\sigma_0 = \inf\{t \geq \tau_k : |X_t| = 0\}$. Puisque $\sigma_0 \geq \tau_k$, on a $|X_{\sigma_0}| = |Y_{\sigma_0}| = 0$, et donc les chaînes se trouvent toutes deux à la racine au temps σ_0. On a donc $\tau_{\text{couple}} \leq \sigma_0$, et donc quelque soient x, y, $\mathbb{E}_{x,y}[\tau_{\text{couple}}] \leq E$ comme souhaité.
D'après la formule du cours et l'inégalité de Markov on en déduit que

$$d(t) \leq \frac{E}{t} \leq \frac{2^{3+k} - 16 + 2^{3-k}}{t} \leq \frac{4n}{t},$$

où on rappelle que $n = 2^{k+1} - 1$ est le nombre de noeuds de notre arbre.
On en déduit $t_{\text{mix}} \leq 16n$, et

$$t_{\text{mix}}(\varepsilon) \leq 16n \left\lceil \log_2 \left(\frac{1}{\varepsilon}\right) \right\rceil$$