Corrigé de l'examen du cours de M2 : Méthodes numériques probabilistes

16 janvier 2008, 09h00 - 12h00.

Algorithme "Coupling From The Past"

1 On a

$$\begin{aligned}\mathbb{P}(X_{n+1} = y | (X_0, \dots, X_n) = (x_0, \dots, x_n)), &= \mathbb{P}(f(\theta_{n+1}, X_n) = y | (X_0, \dots, X_n) = (x_0, \dots, x_n), \\ &= \mathbb{P}(f(\theta_{n+1}, X_n) = y | X_n = x_n),\end{aligned}$$

ce qui montre que $(X_n)_{n\geq 0}$ est une chaîne de Markov. De plus, par indépendance des $(\theta_n)_{n\geq 1}$, on obtient

$$\begin{aligned}\mathbb{P}(X_{n+1} = y | X_n = x) &= \mathbb{P}(f(\theta_{n+1}, X_n) = y | X_n = x), \\ &= \mathbb{P}(f(\theta_{n+1}, x) = y | X_n = x), \\ &= \mathbb{P}(f(\theta_{n+1}, x) = y),\end{aligned}$$

et donc

$$P(x, y) = \mathbb{P}(f(\theta_1, x) = y).$$

2.a La suite $(X_n)_{n\geq 0}$ est bien une chaîne de Markov. On suppose connu à l'instant n la configuration X_n. Pour passer à X_{n+1}, l'algorithme consiste à :

- Tirer trois variables aléatoires : k_{n+1} de manière uniforme sur $\{1, \dots, N\}$, S_{n+1} de manière uniforme dans $\{-1, +1\}$, et U_{n+1} de manière uniforme sur $(0, 1)$,
- Définir la proposition de changement : $\tilde{X}_{n+1,i} = X_{n,i}$ pour tout $i \in \{1, \dots, k_{n+1} - 1, k_{n+1} + 1, \dots, N\}$ et $\tilde{X}_{n+1,k_{n+1}} = S_{n+1}$,
- Calculer le taux d'acceptation $r_{n+1} = \frac{\pi(\tilde{X}_{n+1})}{\pi(X_n)}$ puis :
  - si $U_{n+1} \leq r \wedge 1$, accepter le changement : $X_{n+1} = \tilde{X}_{n+1}$,
  - si $U_{n+1} > r \wedge 1$, rejeter le changement : $X_{n+1} = X_n$.

Le taux d'acceptation ne fait pas intervenir la matrice de transition de la proposition de changement

$$Q(x, y) = \mathbb{P}(\tilde{X}_{n+1} = y | X_n = x)$$

car cette matrice est symétrique.

Calculons de manière un peu plus précise r_{n+1} :

$$\begin{aligned} r_{n+1} &= \frac{\pi(\tilde{X}_{n+1})}{\pi(X_n)}, \\ &= \exp(\beta(V(X_n) - V(\tilde{X}_{n+1})), \\ &= \exp\left(\beta\left(H\sum_{i=1}^{N}(\tilde{X}_{n+1,i} - X_{n,i}) + \sum_{i,j=1}^{N} J_{i,j}(\tilde{X}_{n+1,i}\tilde{X}_{n+1,j} - X_{n,i}X_{n,j})\right)\right), \\ &= \exp\left(\beta(S_{n+1} - X_{n,k_{n+1}})\left(H + 2\sum_{j\in\{1\dots N\}\setminus k_{n+1}} J_{k_{n+1},j}X_{n,j}\right)\right). \end{aligned}$$

2.b La matrice de transition pour la proposition de changement Q est clairement irréductible ($\forall x, y \in M, Q(x,y) > 0$) et apériodique (car $Q(x,x) > 0$) donc fortement irréductible car M est un espace d'états fini. Il en est donc de même pour la chaîne de Markov $(X_n)_{n\geq 0}$. La mesure invariante est bien π par des résultats classiques sur l'algorithme de Metropolis Hasting.

2.c On a bien $X_{n+1} = f(\theta_{n+1}, X_n)$ avec $\theta_{n+1} = (k_{n+1}, S_{n+1}, U_{n+1})$.

3 On a $G_n(x) = X_n^x$ où X_n^x désigne la chaîne de Markov (1), avec $X_0 = x$. Par conséquent, si on suppose que P est une matrice de transition irréductible et apériodique, de mesure invariante π, on a par le théorème ergodique :

$$\lim_{n\to\infty} \mathbb{P}(G_n(x) = y) = \pi(y).$$

4 On suppose que $T < \infty$ presque sûrement, donc $\Phi = D_T$ presque sûrement. Remarquer que comme D_T est une application constante, Φ est bien une variable aléatoire à valeur dans M. On note également que si D_{n_0} est une fonction constante, alors pour tout $n \geq n_0$, $D_n = D_{n_0}$. On a donc, puisque $T < \infty$,

$$\lim_{n\to\infty} D_n = \Phi \text{ p.s.}$$

et on en déduit par convergence dominée que

$$\lim_{n\to\infty} \mathbb{P}(D_n(x) = y) = \mathbb{P}(\Phi = y).$$

Or, $\mathbb{P}(D_n(x) = y) = \mathbb{P}(G_n(x) = y)$ car D_n et G_n ont même loi, et donc, par la question précédente, on obtient

$$\begin{aligned}
\mathbb{P}(\Phi = y) &= \lim_{n\to\infty} \mathbb{P}(D_n(x) = y), \\
&= \lim_{n\to\infty} \mathbb{P}(G_n(x) = y), \\
&= \pi(y).
\end{aligned}$$

5 Dire que D_{d-1} est une application constante est équivalent à dire que $|D_{d-1}(M)| = 1$. Montrons par récurrence que $\mathbb{P}(|f_{\theta_{d-k}} \circ \ldots \circ f_{\theta_{d-1}}(M)| \leq d-k) \geq \alpha^k$. Pour $k = 1$, on a bien $\mathbb{P}(|f_{\theta_{d-1}}(M)| \leq d-1) \geq \alpha$. Ensuite, pour passer de $k-1$ à k (avec $2 \leq k \leq d-1$), on écrit :

$$\begin{aligned}
&\mathbb{P}(|f_{\theta_{d-k}} \circ \ldots \circ f_{\theta_{d-1}}(M)| \leq d-k) \\
&\quad\geq \sum_{A\subset M,\, |A|\leq d-(k-1)} \mathbb{P}(|f_{\theta_{d-k}}(A)| \leq d-k \text{ et } f_{\theta_{d-(k-1)}} \circ \ldots \circ f_{\theta_{d-1}}(M) = A), \\
&\quad= \sum_{A\subset M,\, |A|\leq d-(k-1)} \mathbb{P}(|f_{\theta_{d-k}}(A)| \leq d-k)\mathbb{P}(f_{\theta_{d-(k-1)}} \circ \ldots \circ f_{\theta_{d-1}}(M) = A), \\
&\quad\geq \alpha \sum_{A\subset M,\, |A|\leq d-(k-1)} \mathbb{P}(f_{\theta_{d-(k-1)}} \circ \ldots \circ f_{\theta_{d-1}}(M) = A), \\
&\quad\geq \alpha\, \mathbb{P}(|f_{\theta_{d-(k-1)}} \circ \ldots \circ f_{\theta_{d-1}}(M)| \leq d-(k-1)), \\
&\quad\geq \alpha^k.
\end{aligned}$$

En appliquant le résultat à $k = d-1$, on obtient

$$\mathbb{P}(|f_{\theta_1} \circ \ldots \circ f_{\theta_{d-1}}(M)| \leq 1) \geq \alpha^{d-1}.$$

Remarquer que $T > d-1$ est équivalent D_{d-1} est une application non constante (car si D_{n_0} est une fonction constante, alors pour tout $n \geq n_0$, $D_n = D_{n_0}$), et donc

$$\mathbb{P}(T > d-1) \leq 1 - \alpha^{d-1}.$$

Considérons maintenant

$$\begin{aligned}
\mathbb{P}(T > 2(d-1)) &= \mathbb{P}\left(f_{\theta_1} \circ \ldots \circ f_{\theta_{d-1}} \circ f_{\theta_d} \circ \ldots \circ f_{\theta_{2(d-1)}} \text{ n'est pas constante}\right) \\
&\leq \mathbb{P}\left(f_{\theta_1} \circ \ldots \circ f_{\theta_{d-1}} \text{ n'est pas constante et } f_{\theta_d} \circ \ldots \circ f_{\theta_{2(d-1)}} \text{ n'est pas constante}\right) \\
&\leq \mathbb{P}\left(f_{\theta_1} \circ \ldots \circ f_{\theta_{d-1}} \text{ n'est pas constante}\right) \mathbb{P}\left(f_{\theta_d} \circ \ldots \circ f_{\theta_{2(d-1)}} \text{ n'est pas constante}\right) \\
&\leq (1 - \alpha^{d-1})^2,
\end{aligned}$$

où on a utilisé le fait que la suite $(\theta_n)_{n\geq 1}$ est i.i.d.. Par récurrence, on vérifie que, pour tout $k \in \mathbb{N}$,

$$\mathbb{P}(T > k(d-1)) \leq (1 - \alpha^{d-1})^k$$

et on en déduit que

$$\mathbb{P}(T = \infty) = \lim_{k\to\infty} \mathbb{P}(T > k(d-1)) = 0.$$

6 L'inconvénient principal de cet algorithme est que pour calculer le temps de coalescence T, il faut stocker l'application D_n ce qui coûte généralement très cher, dès que le cardinal de M est grand (ou ce qui est même impossible si M est un espace d'états infini).

7.a Dans ce cas, puisque $\forall x \in M$, $D_n(x_{\min}) \leq D_n(x) \leq D_n(x_{\max})$, le temps de coalescence s'exprime simplement sous la forme :

$$T = \inf\{n \in \mathbb{N}, D_n(x_{\min}) = D_n(x_{\max})\}.$$

Il suffit donc de calculer les deux suites $(D_n(x_{\min}))_{n\geq 0}$ et $(D_n(x_{\max}))_{n\geq 0}$. Remarquer que ces suites ne sont pas des chaînes de Markov. En pratique, il faut donc garder la mémoire de la suite des $(\theta_n)_{n\geq 1}$, et utiliser la même suite $(\theta_n)_{n\geq 1}$ pour le calcul de $(D_n(x_{\min}))_{n\geq 0}$ et de $(D_n(x_{\max}))_{n\geq 0}$.

On peut se poser la question de la complexité de l'algorithme. Si l'on teste à chaque pas de temps si $D_n(x_{\min}) = D_n(x_{\max})$, et que l'on prend comme mesure de la complexité le nombre d'appels à la fonction f, la complexité sera $2T(T+1)/2$ soit d'ordre T^2. En pratique, on préfère tester si $D_n(x_{\min}) = D_n(x_{\max})$ en des temps $n = 2^k$. En effet, dans ce cas, si $2^k < T \leq 2^{k+1}$, on aura fait $2(1 + 2 + 4 + \ldots + 2^{k+1}) = 2(2^{k+2} - 1)$ appels à la fonction f. Dans le pire des cas, $T = 2^k + 1$ et la complexité est donc d'ordre $8T$ (soit seulement 4 fois la complexité optimale $2T$).

7.b Le fait que si D_{n_0} est une fonction constante, alors pour tout $n \geq n_0$, $D_n = D_{n_0}$ montre que $\mathbb{P}(T > n) = \mathbb{P}(D_n \text{ n'est pas une application constante})$. Le fait que D_n et G_n ont même loi montre que

$$\mathbb{P}(D_n \text{ n'est pas une application constante}) = \mathbb{P}(G_n \text{ n'est pas une application constante}).$$

On remarque maintenant que pour tout $k \leq n$,

$$(Y_k = x_{\max} \iff G_k(x_{\min}) = x_{\max}) \implies G_k \text{ est une application constante}$$

et, de plus, si G_k est une application constante, alors G_n aussi, donc

$$\exists k \in \{1, \dots n\}, Y_k = x_{\max} \implies G_n \text{ est une application constante.}$$

Par conséquent

$$\mathbb{P}(G_n \text{ n'est pas une application constante}) \leq \mathbb{P}(\forall k \in \{1, \dots, n\},\, Y_k \neq x_{\max}),$$

et cette probabilité est égale à $P(S > n)$. Comme on suppose que la chaîne $(X_n)_{n\geq 0}$ est irréductible et a une mesure invariante, elle est récurrente positive, et donc $S < \infty$ p.s., *i.e.* $\mathbb{P}(S = \infty) = 0$. On obtient donc

$$\mathbb{P}(T = \infty) = \lim_{n\to\infty} \mathbb{P}(T > n) \leq \lim_{n\to\infty} \mathbb{P}(S > n) = 0.$$

8 Pour montrer que la dynamique est monotone, on considère deux chaînes X_n et Y_n qui évoluent suivant la dynamique (1), on se place à un instant n pour lequel $X_n \leq Y_n$, on fixe $\theta_{n+1} = (k, S, U)$, et on se demande si $X_{n+1} = f(\theta_{n+1}, X_n) \leq Y_{n+1} = f(\theta_{n+1}, Y_n)$. Le seul spin a être modifié est le k-ième, et à l'instant n, on a $X_{n,k} \leq Y_{n,k}$. Supposons que $S = -1$. La seule situation à considérer est $X_{n,k} = 1$ et $Y_{n,k} = 1$ (dans tous les autres cas, quelque soit le résultat de la phase d'acceptation rejet, on aura toujours $X_{n+1} \leq Y_{n+1}$). Plus précisément, la question est : si $S = -1$, $X_{n,k} = Y_{n,k} = 1$ et que la proposition de changement est acceptée pour Y_n (donc $Y_{n+1,k} = -1$), est-ce que la proposition de changement est acceptée pour X_n ? En utilisant la formule pour le taux d'acceptation (cf. question 2.a), on vérifie que dans ce cas, $X_n \leq Y_n$ implique que $r_{n+1}(X_n) \geq r_{n+1}(Y_n)$. Donc si la proposition de changement est acceptée pour Y_n, *i.e.* si $r_{n+1}(Y_n) \geq U$, alors $r_{n+1}(X_n) \geq U$, et donc la proposition de changement est acceptée pour X_n. La situation est symétrique dans le cas $S = 1$. La dynamique est donc monotone.

On dispose donc d'une méthode de simulation exacte du modèle d'Ising dans le cas ferromagnétique ($J_{i,j} \geq 0$), en simulant les suites $D_n(x_{\min})$ et $D_n(x_{\max})$ (cf. question 7.a), et avec $x_{\min} = (-1, \dots, -1)$, $x_{\max} = (1, \dots, 1)$.

Convergence du recuit simulé

1 L'équation de Fokker-Planck associée à la dynamique (2) est

$$\partial_t p = \partial_x((1/2)V' p + (\sigma^2/2)\partial_x p),$$

et on vérifie que

$$\partial_x((1/2)V' p + (\sigma^2/2)\partial_x p) = (\sigma^2/2)\partial_x(q_\sigma \partial_x(p/q_\sigma)).$$

2.a On commence par considérer des fonctions ϕ régulières. Le fait que $\int_{\mathbb{T}} \phi\, q_\sigma = 0$ montre que $\exists x_0$, $\phi(x_0) = 0$. On pose $\psi(x) = \phi(x_0 + x)$. Par Cauchy-Schwarz, on a $\psi(x)^2 =$

$\left(\int_0^x \psi'(y)\,dy\right)^2 \le x \int_0^x \psi'(y)^2\,dy$. Par conséquent,

$$\begin{aligned}
\int_{\mathbb{T}} \psi^2 &= \int_0^1 \psi^2(x)\,dx \\
&\le \int_0^1 x \int_0^x \psi'(y)^2\,dy\,dx, \\
&\le \int_0^1 x \int_0^1 \psi'(y)^2\,dy\,dx, \\
&= \frac{1}{2} \int_{\mathbb{T}} \psi'^2.
\end{aligned}$$

On en déduit que $\int_{\mathbb{T}} \phi^2 = \int_{\mathbb{T}} \psi^2 \le \frac{1}{2} \int_{\mathbb{T}} \psi'^2 = \frac{1}{2} \int_{\mathbb{T}} \phi'^2$. Le résultat est en fait valable pour toute fonction ϕ telle que $\int_{\mathbb{T}} (\phi^2 + \phi'^2) < \infty$ et $\int_{\mathbb{T}} \phi\, q_\sigma = 0$ par densité.

2.b Pour toute fonction ϕ régulière et telle que $\int_{\mathbb{T}} \phi\, q_\sigma = 0$, on a :

$$\begin{aligned}
\int_{\mathbb{T}} \phi'^2 q_\sigma &\ge Z_\sigma^{-1} \exp(-\max V/\sigma^2) \int_{\mathbb{T}} \phi'^2, \\
&\ge 2 Z_\sigma^{-1} \exp(-\max V/\sigma^2) \int_{\mathbb{T}} \phi^2, \\
&\ge 2 Z_\sigma^{-1} \exp(-\max V/\sigma^2) Z_\sigma \exp(\min V/\sigma^2) \int_{\mathbb{T}} \phi^2 q_\sigma, \\
&\ge 2 \exp(-\mathrm{osc}(V)/\sigma^2) \int_{\mathbb{T}} \phi^2 q_\sigma.
\end{aligned}$$

Cette inégalité est en fait valable pour toute fonction ϕ telle que $\int_{\mathbb{T}} (\phi^2 + \phi'^2)\, q_\sigma < \infty$ et $\int_{\mathbb{T}} \phi\, q_\sigma = 0$ par densité. On en déduit que $\lambda_1(\sigma) \ge \sigma^2 \exp(-\mathrm{osc}(V)/\sigma^2)$.

3 On a

$$\begin{aligned}
\frac{df}{dt} &= \int_{\mathbb{T}} \partial_t p \,(p/q_{\sigma_0} - 1)\,, \\
&= \int_{\mathbb{T}} (\sigma_0^2/2) \partial_x (q_{\sigma_0} \partial_x (p/q_{\sigma_0}))\,(p/q_{\sigma_0} - 1)\,, \\
&= -(\sigma_0^2/2) \int_{\mathbb{T}} (\partial_x (p/q_{\sigma_0}))^2 q_{\sigma_0}.
\end{aligned}$$

Or, par définition du trou spectral, puisque $\int (p/q_{\sigma_0} - 1)\, q_{\sigma_0} = 0$, on a

$$(\sigma_0^2/2) \int_{\mathbb{T}} (\partial_x (p/q_{\sigma_0}))^2 q_{\sigma_0} \ge \lambda_1(\sigma_0) \int_{\mathbb{T}} (p/q_{\sigma_0} - 1)^2\, q_{\sigma_0}.$$

On en déduit que

$$\frac{df}{dt} \le -2\lambda_1(\sigma_0) f$$

et donc $f(t) \leq f(0)\exp(-2\lambda_1(\sigma_0)t)$. On remarque que par Cauchy-Schwarz

$$\begin{aligned}\int_{\mathbb{T}} |p - q_{\sigma_0}| &= \int_{\mathbb{T}} |p/q_{\sigma_0} - 1| q_{\sigma_0}, \\ &\leq \left(\int_{\mathbb{T}} |p/q_{\sigma_0} - 1|^2 q_{\sigma_0}\right)^{1/2}, \\ &\leq (2f(t))^{1/2}, \\ &\leq (2f(0))^{1/2}\exp(-\lambda_1(\sigma_0)t).\end{aligned}$$

4 En développant le carré, on a

$$\begin{aligned}e(t) &= \frac{1}{2}\int_{\mathbb{T}} \left(\left(\frac{p(t,\cdot)}{q_{\sigma(t)}}\right)^2 - 2\frac{p(t,\cdot)}{q_{\sigma(t)}} + 1\right) q_{\sigma(t)}, \\ &= \frac{1}{2}\left(\int_{\mathbb{T}} \frac{p(t,\cdot)^2}{q_{\sigma(t)}} - 2\int_{\mathbb{T}} p(t,\cdot) + \int_{\mathbb{T}} q_{\sigma(t)}\right), \\ &= \frac{1}{2}\left(\int_{\mathbb{T}} \frac{p(t,\cdot)^2}{q_{\sigma(t)}} - 1\right).\end{aligned}$$

On a donc

$$\frac{de}{dt} = \int_{\mathbb{T}} \partial_t p \frac{p}{q_\sigma} - \frac{1}{2}\int_{\mathbb{T}} \partial_t q_\sigma \left(\frac{p}{q_\sigma}\right)^2.$$

Pour le premier terme, on a

$$\begin{aligned}\int_{\mathbb{T}} \partial_t p \frac{p}{q_\sigma} &= \int_{\mathbb{T}} (\sigma^2/2)\partial_x(q_\sigma \partial_x(p/q_\sigma))\frac{p}{q_\sigma}, \\ &= -(\sigma^2/2)\int_{\mathbb{T}} (\partial_x(p/q_\sigma))^2 q_\sigma, \\ &\leq -\lambda_1(\sigma)\int_{\mathbb{T}} (p/q_\sigma - 1)^2 q_\sigma, \\ &\leq -2\lambda_1(\sigma)e(t).\end{aligned}$$

Pour le second terme, on a

$$\begin{aligned}\partial_t q_\sigma &= \partial_t(Z_\sigma^{-1}\exp(-V\sigma^{-2})), \\ &= -\partial_t(Z_\sigma)Z_\sigma^{-2}\exp(-V\sigma^{-2}) - Z_\sigma^{-1}V\partial_t(\sigma^{-2})\exp(-V\sigma^{-2}), \\ &= \left(\int_{\mathbb{T}} V\partial_t(\sigma^{-2})\exp(-V\sigma^{-2})\right) Z_\sigma^{-1}q_\sigma - V\partial_t(\sigma^{-2})q_\sigma, \\ &= \partial_t(\sigma^{-2})q_\sigma\left(\int_{\mathbb{T}} Vq_\sigma - V\right).\end{aligned}$$

On en déduit que

$$\begin{aligned}
-\frac{1}{2}\int_{\mathbb{T}} \partial_t q_\sigma \left(\frac{p}{q_\sigma}\right)^2 &= -\frac{1}{2}\int_{\mathbb{T}} \partial_t(\sigma^{-2}) \left(\int_{\mathbb{T}} V q_\sigma - V\right)\left(\frac{p^2}{q_\sigma}\right), \\
&= -\frac{1}{2a(t+2)}\int_{\mathbb{T}} \left(\int_{\mathbb{T}} V q_\sigma - V\right)\left(\frac{p^2}{q_\sigma}\right), \\
&\leq -\frac{1}{2a(t+2)}\int_{\mathbb{T}} (\min V - \max V)\left(\frac{p^2}{q_\sigma}\right), \\
&\leq \frac{\mathrm{osc} V}{2a(t+2)}(2e(t)+1).
\end{aligned}$$

On obtient ensuite facilement le résultat annoncé.

5 En utilisant la question 2.b, on a

$$\begin{aligned}
\alpha(t) &\geq 2\sigma^2 \exp(-\mathrm{osc}(V)\sigma^{-2}) - \frac{\mathrm{osc}(V)}{a(t+2)} \\
&\geq \frac{2a}{\ln(t+2)}(t+2)^{-\mathrm{osc}(V)/a} - \frac{\mathrm{osc}(V)}{a(t+2)}.
\end{aligned}$$

On voit donc que dans la limite $t \to \infty$, $\alpha(t) \geq \frac{C}{t^{\mathrm{oscV}/a} \ln t}$, pour une constante $C > 0$. On en déduit que $\int_0^\infty \alpha(t)\, dt = \infty$, et que $\lim_{t\to\infty} \frac{\beta(t)}{\alpha(t)} = 0$.

On a clairement $\frac{d}{dt}\left(\exp\left(\int_0^t \alpha(s)\, ds\right) e(t)\right) \leq \exp\left(\int_0^t \alpha(s)\, ds\right)\beta(t)$ et donc, pour $0 \leq t_0 \leq t$,

$$\begin{aligned}
e(t) &\leq \exp\left(-\int_0^t \alpha(s)\, ds\right)\int_0^t \exp\left(\int_0^s \alpha(r)\, dr\right)\beta(s)\, ds \\
&\leq \exp\left(-\int_0^t \alpha(s)\, ds\right)\left(\int_0^{t_0} \exp\left(\int_0^s \alpha(r)\, dr\right)\beta(s)\, ds + \int_{t_0}^t \exp\left(\int_0^s \alpha(r)\, dr\right)\beta(s)\, ds\right) \\
&\leq \exp\left(-\int_0^t \alpha(s)\, ds\right)\int_0^{t_0} \exp\left(\int_0^s \alpha(r)\, dr\right)\beta(s)\, ds + \int_{t_0}^t \exp\left(-\int_s^t \alpha(r)\, dr\right)\frac{\beta(s)}{\alpha(s)}\alpha(s)\, ds.
\end{aligned}$$

Pour t_0 assez grand, $\alpha(t)$ est positif pour tout $t \geq t_0$ et on peut donc majorer le dernier terme de la façon suivante :

$$\begin{aligned}
\int_{t_0}^t \exp\left(-\int_s^t \alpha(r)\, dr\right)\frac{\beta(s)}{\alpha(s)}\alpha(s)\, ds &\leq \sup\left(\frac{\beta(t)}{\alpha(t)}, t \geq t_0\right)\int_{t_0}^t \exp\left(-\int_s^t \alpha(r)\, dr\right)\alpha(s)\, ds \\
&= \sup\left(\frac{\beta(t)}{\alpha(t)}, t \geq t_0\right)\left(1 - \exp\left(-\int_{t_0}^t \alpha(r)\, dr\right)\right) \\
&\leq \sup\left(\frac{\beta(t)}{\alpha(t)}, t \geq t_0\right).
\end{aligned}$$

On obtient donc

$$e(t) \leq \exp\left(-\int_0^t \alpha(s)\, ds\right)\int_0^{t_0} \exp\left(\int_0^s \alpha(r)\, dr\right)\beta(s)\, ds + \sup\left(\frac{\beta(t)}{\alpha(t)}, t \geq t_0\right).$$

Pour t_0 assez grand, le second terme est aussi petit que l'on veut, et on fait tendre ensuite le premier terme vers 0 quand $t \to \infty$. Ceci montre que $\lim_{t\to\infty} e(t) = 0$.

6 Par Cauchy-Schwarz, on obtient facilement

$$\begin{aligned} e(t) &\geq \frac{1}{2}\int_{\mathcal{M}^\delta}\left(\frac{p(t,\cdot)}{q_{\sigma(t)}}-1\right)^2 q_{\sigma(t)}, \\ &\geq \frac{1}{2}\left(\int_{\mathcal{M}^\delta}|p(t,\cdot)-q_{\sigma(t)}|\right)^2, \\ &\geq \frac{1}{2}\left|\mathbb{P}(X_t\in\mathcal{M}^\delta)-\int_{\mathcal{M}^\delta} q_{\sigma(t)}\right|^2. \end{aligned}$$

Comme $\sigma(t)$ tend vers 0 quand $t\to\infty$, on a clairement $\lim_{t\to\infty}\int_{\mathcal{M}^\delta} q_{\sigma(t)}=1$. Par conséquent, $\lim_{t\to\infty}\mathbb{P}(X_t\in\mathcal{M}^\delta)=1$. Autrement dit, $\lim_{t\to\infty}\mathbb{P}(\mathrm{dist}(X_t,\mathcal{M})\geq\delta)=0$, ce qui termine la preuve.