Séminaire TEST

Andrés SÁNCHEZ PÉREZ

4 avril 2014

1 Présentation du sujet

Les méthodes de Monte-Carlo par chaînes de Markov sont une classe de technique d'échantillonnage. Un algorithme MCMC repose sur le parcours d'une chaîne de Markov qui a pour loi stationnaire la distribution à échantillonner.

1.1 Notation

Soient $(X, \mathcal{X})$, $(Y, \mathcal{Y})$, $(Z, \mathcal{Z})$ des espaces mesurables. Nous denotons par
$\mathbb{F}(X, \mathcal{X})$ l'ensemble des fonctions mesurables de $(X, \mathcal{X})$ dans $[-\infty, \infty]$.
$\mathbb{F}_+(X, \mathcal{X})$ l'ensemble des fonctions mesurables de $(X, \mathcal{X})$ dans $[0, \infty]$.
$\mathbb{F}_b(X, \mathcal{X})$ l'ensemble des fonctions mesurables et bornées de $(X, \mathcal{X})$ dans $[0, \infty)$.
$\mathbb{M}_\pm(\mathcal{X})$ l'ensemble des mesures signées et finies dans $(X, \mathcal{X})$.
$\mathbb{M}_+(\mathcal{X})$ l'ensemble des mesures dans $(X, \mathcal{X})$.
$\mathbb{M}_1(\mathcal{X})$ l'ensemble des mesures de probabilité dans $(X, \mathcal{X})$.

Définition 1 *$M : X \times \mathcal{Y} \to [0, \infty]$ est un noyau si*
- *Pour tout $x \in \mathcal{X}$, $M(x, \cdot)$ est une mesure sur $\mathcal{Y}$.*
- *Pour tout $A \in \mathcal{Y}$, $M(\cdot, A)$ est une fonction mesurable.*
  M est dit Markovien si pour tout $x \in X$, $M(x, Y) = 1$.
  m est la densité de M par rapport à la mesure λ si $M(x, A) = \int_A m(x, y)\lambda(\mathrm{d}y)$.

Définition 2 *Pour $f \in \mathbb{F}(Y, \mathcal{Y})$, $\mu \in \mathbb{M}_+(\mathcal{X})$, et les noyaux $M : X \times \mathcal{Y} \to [0, \infty]$ et $N : Y \times \mathcal{Z} \to [0, \infty]$, nous définons*
- *$Mf : x \to \int M(x, \mathrm{d}y) f(y)$. Mf est mesurable : convergence monotone sur $\mathbb{F}_+$ et étendre aux réelles.*
- *$\mu M : A \to \int M(x, A)\mu(\mathrm{d}x)$. $\mu M \in \mathbb{M}_+(\mathcal{Y})$: convergence monotone pour montrer l'additivité dénombrable.*
- *$MN(x, A) = \int M(x, \mathrm{d}y) N(y, A)$.*

Supposez par la suite que $(X, \mathcal{X})$ est un espace polonais, i.e. métrisable, à base dénombrable dont la topologie peut être définie par une distance qui en fait un espace complet.

Définition 3 *Nous disons que le noyau $M : X \times \mathcal{X} \to [0, \infty]$ est réversible par rapport à la mesure μ si pour toutes $f, g \in \mathbb{F}_+(X, \mathcal{X})$*

$$\int \int \mu(\mathrm{d}x) M(x, \mathrm{d}y) f(x) g(y) = \int \int \mu(\mathrm{d}x) M(x, \mathrm{d}y) f(y) g(x) \, .$$

Définition 4 *La mesure μ est stationnaire par rapport à M si $\mu M = \mu$.*

Proposition 1 *Si M est réversible par rapport à la mesure μ alors μ est stationnaire.*

Preuve
Soient f quelconque dans $\mathbb{F}_+(X, \mathcal{X})$ et $g = 1$

$$\begin{aligned} \int\int \mu(\mathrm{d}x)N(x,\mathrm{d}y)f(x) &= \int\int \mu(\mathrm{d}x)N(x,\mathrm{d}y)f(y) \Leftrightarrow \\ \mu(f) &= \mu N(f)\ . \end{aligned}$$

■

Pout toute chaîne de Markov $X = (X_n)_{n\in\mathbb{N}}$ homogene de loi initiale π_0 et noyau de transition P, nous avons $X_n \sim \mu_0 P^n$.

2 Algorithme de Metropolis-Hastings

$X = (X_t)_{t\in\mathbb{N}}$ est la chaîne que nous voulons generer avec π comme loi stationnaire. Supposons que la loi π a une densité π (par un abus de notation) par rapport à la mesure λ. Nous considérons le noyau Q sur $(X, \mathcal{X})$ qui a une densité q aussi par rapport à la mesure λ. Pour chaque $x \in Q$ nous savons tirer selon la distribution $Q(x, \cdot)$.

Algorithm 1: Metropolis-Hastings

Input: Une densité π et un noyau Q de densité q
Output: Une chaîne de Markov X

$X_0 = x_0$;
for ***each*** t **do**
  $Y_t \sim Q(X_{t-1}, \cdot)$;
  tirer $U \sim \mathcal{U}(0,1)$;
  if $U \leq \alpha(X_{t-1}, Y_t)$ **then**
    $X_t = Y_t$ **else**
      $X_t = X_{t-1}$
return $X = (X_0, X_1, \ldots)$

Où

$$\alpha(x,y) = \begin{cases} \min\left\{\dfrac{\pi(y)\,q(y,x)}{\pi(x)\,q(x,y)}, 1\right\} & \text{si } \pi(x)\,q(x,y) > 0\ , \\ 1 & \text{si } \pi(x)\,q(x,y) = 0\ . \end{cases}$$

Nous remarquons que

$$X_n = 1_{\{U \leq \alpha(X_{n-1}, Y_n)\}} Y_n + 1_{\{U > \alpha(X_{n-1}, Y_n)\}} X_{n-1}\ .$$

$$
\begin{aligned}
\mathbb{P}\left(X_1 \in A \,|X_0 = x\right) =& \mathbb{P}\left(1_{\{U\leq\alpha(X_0,Y_1)\}}Y_1 + 1_{\{U>\alpha(X_0,Y_1)\}}X_0 \in A \,|X_0 = x\right) \\
=& \mathbb{P}\left(\{U \leq \alpha\left(X_0, Y_1\right)\} \cap \{Y_1 \in A\} \,|X_0 = x\right) \\
&+ \mathbb{P}\left(\{U > \alpha\left(X_0, Y_1\right)\} \cap \{X_0 \in A\} \,|X_0 = x\right) \\
=& \int\int \mathbb{1}_{u\leq\alpha(x,y)}\mathbb{1}_{y\in A}\mathbb{1}_{u\in[0,1]}q(x,y)\mathrm{d}u\lambda\left(\mathrm{d}y\right) \\
&+ \int\int \mathbb{1}_{u>\alpha(x,y)}\mathbb{1}_{x\in A}\mathbb{1}_{u\in[0,1]}q(x,y)\mathrm{d}u\lambda\left(\mathrm{d}y\right) \\
=& \int_A \alpha(x,y)q(x,y)\lambda\left(\mathrm{d}y\right) + \mathbb{1}_{x\in A}\int \left(1-\alpha\left(x,z\right)\right)q(x,z)\mathrm{d}u\lambda\left(\mathrm{d}z\right) \ .
\end{aligned}
$$

X est une chaîne de Markov de noyau

$$
P\left(x,\mathrm{d}y\right) \quad = \quad \alpha\left(x,y\right)q\left(x,y\right)\lambda\left(\mathrm{d}y\right) + \left[\int \left(1-\alpha\left(x,z\right)\right)q\left(x,z\right)\lambda\left(\mathrm{d}z\right)\right]\delta_x\left(\mathrm{d}y\right) \ .
$$

Proposition 2 *P est réversible par rapport à π.*

Preuve

$$
\begin{aligned}
\pi(\mathrm{d}x)P\left(x,\mathrm{d}y\right) \quad = \quad & \alpha\left(x,y\right)q\left(x,y\right)\pi\left(x\right)\lambda\left(\mathrm{d}x\right)\lambda\left(\mathrm{d}y\right) \\
&+ \left[\int \left(1-\alpha\left(x,z\right)\right)q\left(x,z\right)\lambda\left(\mathrm{d}z\right)\right]\pi\left(x\right)\lambda\left(\mathrm{d}x\right)\delta_x\left(\mathrm{d}y\right) \ .
\end{aligned}
$$

$$
\alpha\left(x,y\right)q\left(x,y\right)\pi\left(x\right) = \begin{cases} \min\left\{\pi\left(y\right)q\left(y,x\right), \pi\left(x\right)q\left(x,y\right)\right\} & \text{si } \ q\left(x,y\right)\pi\left(x\right) \neq 0 \ , \\ 0 & \text{sinon .} \end{cases}
$$

Soient $f, g \in \mathbb{F}_+\left(X, \mathcal{X}\right)$

$$
\begin{aligned}
\int\int f(x)g(y)\pi(\mathrm{d}x)P\left(x,\mathrm{d}y\right) =& \int\int f(x)g(y)\alpha\left(x,y\right)q\left(x,y\right)\pi\left(x\right)\lambda\left(\mathrm{d}x\right)\lambda\left(\mathrm{d}y\right) \\
&+ \int\int f(x)g(y)\left[\int \left(1-\alpha\left(x,z\right)\right)q\left(x,z\right)\lambda\left(\mathrm{d}z\right)\right]\pi\left(x\right)\lambda\left(\mathrm{d}x\right)\delta_x\left(\mathrm{d}y\right) \\
=& \int\int f(y)g(x)\alpha\left(x,y\right)q\left(x,y\right)\pi\left(x\right)\lambda\left(\mathrm{d}x\right)\lambda\left(\mathrm{d}y\right) \\
&+ \int f(x)g(x)\left[\int \left(1-\alpha\left(x,z\right)\right)q\left(x,z\right)\lambda\left(\mathrm{d}z\right)\right]\pi\left(x\right)\lambda\left(\mathrm{d}x\right) \\
=& \int\int f(y)g(x)\pi(\mathrm{d}x)P\left(x,\mathrm{d}y\right) \ .
\end{aligned}
$$

■

3 Ergodicité

3.1 Systèmes dynamiques

Théorème 3.1 *Soit (F, D) un espace métrique complet quelconque. Soit $T : F \to F$ un opérateur continu tel que, pour un certain $m \in \mathbb{N}^*$, $\alpha \in (0, 1)$ et $u, v \in F$*

$$
D\left(T^m u, T^m v\right) \leq \alpha D\left(u, v\right) \ .
$$

Il existe donc un unique point fixe $a \in F$ et pour tout $u \in F$

$$D(T^n u, a) \leq \left(1-\alpha^{1/m}\right)^{-1} \max_{0\leq i<m} \alpha^{-i/m} D\left(T^i u, T^{i+1}u\right)\alpha^{n/m} .$$

En outre, s'il existe $A \geq 1$ telle que $D(Tu, Tv) \leq AD(u, v)$ pour tout $u, v \in F$, alors

$$D(T^n u, a) \leq \left(\alpha^{-1/m}A\right)^{m-1} D(u, a)\alpha^{n/m} .$$

Preuve
Unicité. Soient a, b des points fixes de T, cela veut dire que $a = Ta = T^m a$ et de même pour b.

$$D(a, b) = D(T^m a, T^m b) \leq \alpha D(a, b) \Rightarrow D(a, b) = 0 .$$

Existence. Considère $u, v \in F$ et $n \in \mathbb{N}$. $n = km + r$ avec $0 \leq r < m$.

$$D(T^n u, T^n v) \leq \alpha^k D(T^r u, T^r v) .$$

En prenant $v = Tu$

$$\begin{aligned} D(T^n u, T^n v) &\leq \alpha^k D\left(T^r u, T^{r+1}u\right) \leq \alpha^{n/m}\alpha^{-r/m} D\left(T^r u, T^{r+1}u\right) \\ &\leq \alpha^{n/m} \max_{0\leq r<m} \alpha^{-r/m} D\left(T^r u, T^{r+1}u\right) . \end{aligned}$$

En conséquence $\{T^n u\}_{n\geq 0}$ est de Cauchy. Appelons a la limite

$$\begin{aligned} D(T^n u, a) &\leq \max_{0\leq r<m} \alpha^{-r/m} D\left(T^r u, T^{r+1}u\right) \sum_{q=n}^{\infty} \alpha^{n/m} \\ &= \left(1-\alpha^{1/m}\right)^{-1} \max_{0\leq i<m} \alpha^{-i/m} D\left(T^i u, T^{i+1}u\right)\alpha^{n/m} . \end{aligned}$$

Comme T est continu $Ta = T \lim T^n u = \lim T^{n+1}u = a$.
Si maintenant $D(Tu, Tv) \leq AD(u, v)$

$$\begin{aligned} D(T^n u, a) = D(T^n u, T^n a) &\leq \alpha^{\lfloor n/m\rfloor} D\left(T^{n-m\lfloor n/m\rfloor}u, T^{n-m\lfloor n/m\rfloor}a\right) \\ &\leq \alpha^{\lfloor n/m\rfloor} A^{n-m\lfloor n/m\rfloor} D(u, a) . \end{aligned}$$

Pour finir nous utilisons que $\lfloor n/m\rfloor \geq n/m - (m-1)/m$ et que $A \geq 1$ et $\alpha \leq 1$.

■

3.2 Espaces $M_V(\mathcal{X})$

Théorème 3.2 *Soit F un sous-espace de $\mathbb{M}_1(\mathcal{X})$ et D une métrique sur F telle que $\delta_x \in F$ pour tout $x \in X$ et (F, D) est complet. Soit P un noyau markovien continu pour D et tel que $\xi P \in F$ pour tout $\xi \in F$. Supposez qu'il existe un certain $m \in \mathbb{N}^*$, $\alpha \in (0, 1)$ et $A > 0$ telles que pour tous $\xi, \xi' \in F$*

$$D(\xi P, \xi' P) \leq AD(\xi, \xi') \qquad D(\xi P^m, \xi' P^m) \leq \alpha D(\xi, \xi') .$$

Alors, il existe une unique mesure invariante $\pi \in F$ et pour tout $\xi \in F$

$$\begin{aligned} D(\xi P^n, \pi) &\leq \left(1-\alpha^{1/m}\right)^{-1}\left(\alpha^{-1/m}A\right)^{m-1} D(\xi, \xi P)\alpha^{n/m} , \\ D(\xi P^n, \pi) &\leq \left(\alpha^{-1/m}A\right)^{m-1} D(\xi, \pi)\alpha^{n/m} . \end{aligned}$$

En outre, si la convergence d'une suite de mesures de probabilité dans F par rapport à D implique la convergence faible, π est donc la seule measure de probabilité P invariante.

Preuve
Soit π la unique mesure invariante dans F et $\tilde{\pi}$ une mesure invariante dans $\mathbb{M}_1(\mathcal{X})$. Pour toute fonction f continue et bornée

$$\tilde{\pi}(f) = \tilde{\pi}P^n(f) = \int \tilde{\pi}(\mathrm{d}x) P^n f(x) \ .$$

$\delta_x P^n \Rightarrow \pi$ (converge faiblement).

$$\delta_x P^n(f) = \int P^n(x, \mathrm{d}y) f(y) = P^n f(x) \to \pi(f)$$

En plus $|P^n f(x)| \leq |f|_\infty$. Pour le théorème de convergence dominée $\int P^n f(x) \tilde{\pi}(\mathrm{d}x) = \pi(f)$.

■

Soit $V : X \to [1, \infty)$ une fonction mesurable. Nous allons étudier des sous-espaces de $\mathbb{M}_\pm(\mathcal{X})$ munis de la norme $\|\cdot\|_V$:

$$\|\xi\|_V = \sup\{\xi(f) : f \in \mathbb{F}_b(X, \mathcal{X}), |f/V|_\infty \leq 1\} \ .$$

Nous denotons $\mathbb{M}_V(\mathcal{X}) = \{\xi \in \mathbb{M}_\pm(\mathcal{X}) : \|\xi\|_V < \infty\}$.

Proposition 3 *$(\mathbb{M}_V(\mathcal{X}), \|\cdot\|_V)$ est complet.*

Dans le cas de la variation totale $V \equiv 1$.

Proposition 4 *La convergence en norme $\|\cdot\|_V$ entraîne la convergence faible.*

Définition 5 *Le noyau P est dit $V-$ géometriquement ergodique si il existe deux constantes $C < \infty$ et $\rho \in (0, 1)$ telles que*

$$\sup_{x \in X} \frac{\|P^n(x, \cdot) - \pi\|_V}{V(x)} \leq C\rho^n \ .$$

Définition 6 *Le V-coefficient de Dobrushin est le coefficient de Lipschitz de P par rapport à la distance $\|\cdot\|_V$.*

$$\Delta_V(P) = \sup_{\xi \neq \xi'} \frac{d_V(\xi P, \xi' P)}{d_V(\xi, \xi')},$$

Lemme 1

$$\Delta_V(P) = \sup_{x,y} \frac{\|P(x, \cdot) - P(y, \cdot)\|_V}{V(x) + V(y)}$$

Définition 7 *Définitions*

– *$X = (X_t)_{t \in \mathbb{N}}$ est dites $\varphi-$ irréductible si il existe une mesure φ sur $\mathcal{X}$ telle que pour tout $A \in \mathcal{X}$ vérifiant $\varphi(A) > 0$ et pour tout $x \in X$, il existe $n \in \mathbb{N}$ tel que $P^n(x, A) > 0$.*

– *Un ensemble mesurable $C \in \mathcal{X}$ est dit $\mu-$ small si il existe une mesure non triviale μ sur $\mathcal{X}$ telle que pour tout $x \in C$, et tout $B \in \mathcal{X}$, $P(x, B) \geq \mu(B)$.*
– *Si il existe une mesure μ et un ensemble $\mu-$ small A tel que $\mu(A) > 0$, alors la chaîne est dite fortement apériodique.*

Définition 8 *Condition de Doeblin : $\exists m \geq 1$, $\varepsilon > 0$ et $\nu \in \mathbb{M}_1(\mathcal{X})$ tels que pour tout $x \in X$ et $A \in \mathcal{X}$, $P^m(x, A) \geq \varepsilon\nu(A)$.*

Proposition 5 *La condition de Doeblin implique que $\Delta_1(P^m) \leq 1 - \varepsilon$.*

Preuve
Soit $Q(x, \cdot) = (1 - \varepsilon)^{-1} (P^m(x, A) - \varepsilon\nu(A))$ est un noyau et

$$\|P^m(x, \cdot) - P^m(y, \cdot)\|_{TV} = (1 - \varepsilon)\|Q(x, \cdot) - Q(y, \cdot)\|_{TV} \leq 2(1 - \varepsilon) .$$

■

Proposition 6 *Si la chaîne $X = (X_t)_{t\in\mathbb{N}}$ est $\varphi-$ irréductible et apériodique, de plus, il existe un ensemble small C et une fonction $V : X \to [1, \infty)$ tels que l'on ait la condition de dérive suivante : il existe deux constantes $\lambda \in (0, 1)$ et $b < \infty$ telles que*

$$PV(x) \leq \lambda V(x) + b\mathbb{1}_C(x), \forall x \in X ,$$

alors la chaîne est $V-$ géometriquement ergodique.

Pour une fonction $f \in \mathbb{F}(X, \mathcal{X})$ nous considerons

$$S_n(f) = \frac{1}{n}\sum_{t=0}^{n-1} f(X_t) .$$

Définition 9 *Si π est la unique probabilité invariante de P, pour $f \in \mathbb{F}(X, \mathcal{X})$ telle que $\pi|f| < \infty$ l'équation de Poisson associée à f est définie par*

$$\hat{f} - P\hat{f} = f - \pi(f) .$$

Nous disons que $\hat{f}$ est une solution si elle satisfait l'équation et $P\left|\hat{f}\right|(x) < \infty$.

Lemme 2 *Si le noyau est $V-$ géometriquement ergodique pour toute f telle que $|f/V|_\infty < \infty$ nous avons*

$$\begin{aligned} \sum_{n=0}^{\infty} |P^n f(x) - \pi(f)| &< \infty \\ \hat{f}(x) &= \sum_{n=0}^{\infty} (P^n f(x) - \pi(f)) \end{aligned}$$

est une solution de l'équation de Poisson associée à f

Preuve
Soit $\hat{f}_n(x) = \sum_{t=0}^{n-1} (P^t f(x) - \pi(f))$

$$\left|\hat{f}_{n+1}(x) - \hat{f}_n(x)\right| = |P^n f(x) - \pi(f)| \leq C\rho^n V(x) |f/V|_\infty , \quad (1)$$

et $\hat{f}_n(x)$ converge.

$$\left|\frac{\hat{f}_n(x)}{V(x)}\right| \leq \frac{\sum_{n=0}^{\infty} |P^n f(x) - \pi(f)|}{V(x)} \leq C\,(1-\rho)^{-1} \left|\frac{f}{V}\right|_\infty$$

Sous conditions d'integrabilité de V par rapport à $P(x,\cdot)$ on retrouve que $P\hat{f}_n(x) \to P\hat{f}$.

$$\hat{f}(x) - P\hat{f}(x) = \lim\left(\hat{f}_n(x) - P\hat{f}_n(x)\right) = f(x) - \pi(f) + \lim\left(P^n f(x) - \pi(f)\right) = f(x) - \pi(f)\ .$$

■

Proposition 7 *Soit f telle que $\pi|f| < \infty$ et $\hat{f}$ une solution de l'équation de Poisson. Donc,*

$$S_n(f) - \pi(f) = \frac{M_n\left(\hat{f}\right)}{n} + \frac{\hat{f}(X_0) - \hat{f}(X_n)}{n}$$

$$M_n\left(\hat{f}\right) = \sum_{t=1}^{n}\left\{\hat{f}(X_t) - \mathbb{E}\left[\hat{f}(X_t)\,|\mathcal{F}_{t-1}\right]\right\} = \sum_{t=1}^{n}\left\{\hat{f}(X_t) - P\hat{f}(X_{t-1})\right\}$$

où $\mathcal{F}_t = \sigma(X_0, \ldots, X_t)$. En outre $\left(M_n\left(\hat{f}\right), \mathcal{F}_n\right)$ est une martingale.

Preuve

$$S_n(f) - \pi(f) = \frac{1}{n}\sum_{t=0}^{n-1}\left(f(X_t) - \pi(f)\right) = \frac{1}{n}\sum_{t=0}^{n-1}\left(\hat{f}(X_t) - P\hat{f}(X_t)\right)$$

■

Théorème 3.3 *Si la chaîne $X = (X_t)_{t\in\mathbb{N}}$ est $\varphi-$ irréductible et apériodique, de plus, il existe un ensemble small C et une fonction $V : \mathcal{X} \to [1,\infty)$ tels que l'on ait la condition de dérive suivante : il existe deux constantes $\lambda \in (0,1)$ et $b < \infty$ telles que*

$$PV(x) \leq \lambda V(x) + b\mathbb{1}_C(x), \forall x \in \mathcal{X}\ ,$$

et que $\pi(V) < \infty$ alors pour tout $a > 1$ et toute fonction f telle que $|f|_{V^{1/a}} < \infty$,

$$\lim \frac{1}{n}\sum_{k=1}^{n} f(X_k) = \pi(f)\ \text{ p.s.}$$