UMLV
Modélisation stochastique (M2)

Mardi 19 Décembre 2006 (14h-17h)

Exercice 1 (Dernier temps de passage).
Soit Y une chaîne de Markov irréductible transiente à valeurs dans un espace d'état au plus dénombrable. On note $\tau_y = \sup\{n \geq 0; Y_n = y\}$ le dernier temps de passage en y.

1. Montrer que $\sum_{n\geq 0} \mathbb{P}(\tau_y = n | Y_0 = y) = 1$. En déduire que
$$\mathbb{P}(Y_n \neq y \quad \forall n \geq 1 | Y_0 = y) = \frac{1}{1 + \sum_{n\geq 1} \mathbb{P}(Y_n = y | Y_0 = y)} .$$
2. Montrer que $N = \text{Card}\ \{n \geq 0; Y_n = y\}$ suit, conditionnellement à $\{Y_0 = y\}$, une loi géométrique dont on précisera le paramètre.

Soit $(X_n, n \geq 1)$ une suite de variables aléatoires indépendantes de même loi de Bernoulli de paramètre $p \in]0,1[$: $\mathbb{P}(X_n = 1) = p = 1 - \mathbb{P}(X_n = -1)$. On suppose $p \neq 1/2$. On définit $S_0 = 0$, et $S_n = X_1 + \cdots + X_n$ pour $n \geq 1$.

3. Vérifier que $S = (S_n, n \geq 0)$ est une chaîne de Markov à valeurs dans $\mathbb{Z}$. Montrer que S est irréductible transiente (on pourra étudier $\lim_{n\to\infty} S_n/n$).
4. Évaluer $\mathbb{P}(S_n = 0)$. Écrire la fonction $f(x) = (1-x)^{-1/2}$ sous forme d'une série entière (pour $x \in (-1, 1)$). Déduire de la question 1. que $\mathbb{P}(S_n \neq 0 \quad \forall n \geq 1) = |1 - 2p|$.

$\triangle$

Exercice 2 ("Waste recycling Monte Carlo").
L'algorithme "Waste recycling Monte Carlo" [1] (WR) est une variante de l'algorithme de Metropolis Hasting qui tient compte de toutes les propositions (et pas seulement des propositions acceptées comme dans Metropolis-Hastings). Soit

- E un espace au plus dénombrable,
- π une probabilité sur E telle que $\pi(x) > 0$ pour tout $x \in E$,
- Q une matrice de transition irréductible telle que si $Q(x, y) = 0$ alors $Q(y, x) = 0$,
- ρ une fonction définie sur $E \times E$ à valeurs dans $(0, 1]$ telle que pour tout $x, y \in E$,
$$\rho(x, y)\pi(x)Q(x, y) = \rho(y, x)\pi(y)Q(y, x).$$

Soit X_0 une variable aléatoire à valeurs dans E. On suppose construit $X_0, \ldots, X_n$. La proposition à l'étape $n+1$, $\tilde{X}_{n+1}$ est distribuée suivant $Q(X_n, \cdot)$. La proposition est acceptée avec probabilité $\rho(X_n, \tilde{X}_{n+1})$ et alors on pose $X_{n+1} = \tilde{X}_{n+1}$. Si elle est rejetée, alors on pose $X_{n+1} = X_n$.

Pour ν une probabilité sur E et f une fonction définie sur E, on pose $\langle \nu, f\rangle = \sum_{x\in E} \nu(x) f(x)$, dès que le second membre a un sens.

[1] D. Frenkel. Waste Recycling Monte Carlo, *To appear* (2006).

1. Vérifier que $X = (X_n, n \geq 0)$ est une chaîne de Markov. Donner sa matrice de transition, P. Vérifier que P est réversible par rapport à π. Montrer que P est irréductible.
2. Soit f telle que $\langle \pi, |f| \rangle < \infty$. Montrer, en utilisant un théorème du cours, que $I_n = \frac{1}{n}\sum_{k=1}^{n} f(X_k)$ est un estimateur de $\langle \pi, f \rangle$ fortement convergent (i.e. $\lim_{n\to\infty} I_n = \langle \pi, f \rangle$ p.s.).
3. On pose $f^c(x, \tilde{x}) = \mathbb{E}[f(X_1)|X_0 = x, \tilde{X}_1 = \tilde{x}]$ et $I_n^{WR}(f) = \frac{1}{n}\sum_{k=0}^{n-1} f^c(X_k, \tilde{X}_{k+1})$. Montrer que $I_n^{WR}(f)$ et $I_n(f)$ sont des estimateurs $\langle \pi, f \rangle$ ayant même biais (i.e. $\mathbb{E}[I_n^{WR}(f)] = \mathbb{E}[I_n(f)]$).
4. Montrer que $X^c = (X_n^c = (X_n, \tilde{X}_{n+1}), n \geq 0)$ est une chaîne de Markov. Donner sa matrice de transition, P^c. Est-elle irréductible ? Montrer que π^c définie par $\pi^c(x, \tilde{x}) = \pi(x)Q(x, \tilde{x})$ est une probabilité invariante pour P^c. Est-ce la seule ?
5. Calculer $\langle \pi^c, f^c \rangle$. Montrer que $I_n^{WR}(f)$ converge p.s. vers $\langle \pi, f \rangle$.
6. Montrer que $\mathrm{Var}_\pi(f(X_k)) \geq \mathrm{Var}_\pi(f^c(X_k, \tilde{X}_{k+1}))$, ou l'indice π indique que la loi de X_0 est π. Peut-on en déduire que la variance de $I_n^{WR}(f)$ est plus faible que celle de I_n ?
7. On suppose qu'il existe F solution de l'équation de Poisson $F - PF = f - \langle \pi, f \rangle$ telle que $\langle \pi, F^2 \rangle < \infty$. Montrer, en utilisant un théorème du cours, que, quand n tend vers l'infini, $n^{-1/2}(\sum_{k=1}^n F(X_k) - PF(X_{k-1}) - \langle \pi, f \rangle)$ converge en loi vers une loi gaussienne, $\mathcal{N}(0, \sigma^2)$, de variance $\sigma^2 = \langle \pi, F^2 \rangle - \langle \pi, (PF)^2 \rangle$. En déduire que I_n est un estimateur de $\langle \pi, f \rangle$ asymptotiquement normal de variance asymptotique σ^2 (i.e. on a la convergence en loi de $\sqrt{n}(I_n - \langle \pi, f \rangle)$ vers la loi gaussienne centrée de variance σ^2 quand n tend vers l'infini).
8. (Facultatif) Montrer que $P^c F^c = (PF)^c$. Vérifier que F^c est solution de l'équation de Poisson $F^c - P^c F^c = f^c - \langle \pi^c, f^c \rangle$. En déduire que I_n^{WR} est un estimateur asymptotiquement normal de $\langle \pi, f \rangle$ de variance asymptotique $\sigma_{WR}^2 = \langle \pi^c, (F^c)^2 \rangle - \langle \pi^c, ((PF)^c)^2 \rangle$.
9. (Facultatif) Montrer que
$$\sigma^2 - \sigma_{WR}^2 = \mathbb{E}_\pi\left[(1 - \rho(X_0, X_1))\Big[(F(X_1) - F(X_0))^2 - (PF(X_1) - PF(X_0))^2\Big]\right].$$

On peut, à partir de l'expression de $\sigma^2 - \sigma_{WR}^2$ fournie par la dernière question, montrer que l'algorithme WR améliore dans certains cas particuliers la variance asymptotique pour l'estimation de $\langle \pi, f \rangle$. △

Exercice 3 (Fonction de contraste pour chaîne de Markov cachée)**.**
On considère un processus à temps discret $Z = (Z_n = (X_n, Y_n), n \geq 0)$ à valeurs dans $\{1, 2\} \times \{1, 2\}$. Le processus $X = (X_n, n \geq 0)$ est une chaîne de Markov stationnaire (la loi de X_0 est une loi invariante) de matrice de transition

$$P_a = \begin{pmatrix} 1-a & a \\ a & 1-a \end{pmatrix},$$

où $a \in [\delta, 1-\delta]$, et δ désigne un réel positif arbitrairement petit. On considère de plus, que pour tout $n \geq 0$, et tout $x_k, y_k \in \{1, 2\}$, avec $0 \leq k \leq n$:

$$\begin{aligned}\mathbb{P}(Y_n = y_n|X_0 = x_0, ..., X_n = x_n, Y_0 = y_0, ..., Y_{n-1} = y_{n-1}) &= \mathbb{P}(Y_n = y_n|X_n = x_n) \\ &= \mathbb{P}(Y_0 = y_n|X_0 = x_n).\end{aligned}$$

On rappelle que lorsque la chaîne de Markov X n'est pas observable, le processus Y "seul" est appelé chaîne de Markov cachée. On note $\mathbb{P}(Y_0 = j|X_0 = i) = \gamma(j|i)$ pour tout $(i,j) \in \{1,2\}^2$ et $\beta = \gamma(1|2)$. On suppose $\beta \in [\delta, 1-\delta]$ et $\gamma(1|1) = \lambda\beta$ où $0 < \lambda < 1$ est un facteur multiplicatif supposé **connu**.

Le modèle de Markov caché ainsi défini est entièrement paramétré par $\theta = (a, \beta) \in \Theta_\delta = [\delta, 1-\delta]^2$. Le but de cet exercice est de proposer un estimateur convergent de θ.

1. Écrire, à l'aide de P et γ, la vraisemblance, $p_\theta(y_0, y_1, ..., y_n)$, du modèle de Markov caché défini ci-dessus associée à une trajectoire observée $(y_0, y_1, ..., y_n)$. Donner la définition de l'estimateur du maximum de vraisemblance dans ce cadre, en indiquant quelles difficultés nous serions amenés à rencontrer (*vs.* une suite de variables aléatoires indépendantes de même loi ou une suite markovienne) lors de son étude.
2. Montrer que $p_\theta(1) = p_{\theta'}(1)$ et $p_\theta(1,1) = p_{\theta'}(1,1)$ impliquent $\theta = \theta'$. En déduire que le modèle est identifiable.

On considère l'estimateur suivant du vrai paramètre θ_0 inconnu :

$$\hat{\theta}_n = \arg\max_{\theta \in \Theta_\delta} L_n(\theta; Y_0, ..., Y_{2n+1}), \quad \text{où} \quad L_n(\theta; Y_0, ..., Y_{2n+1}) = \log \prod_{k=0}^{n-1} p_\theta(Y_{2k}, Y_{2k+1}). \tag{1}$$

On note $\mathcal{H}_{\theta_0}(\theta) = \mathbb{E}_{\theta_0}[\log p_\theta(Y_0, Y_1)]$.

3. Montrer que le processus $((X_{2n}, Y_{2n}, X_{2n+1}, Y_{2n+1}), n \geq 0)$ est markovien et irréductible. Montrer, à l'aide du théorème ergodique, que $\mathbb{P}_{\theta_0}$-p.s. $\lim_{n\to\infty} \frac{1}{n} L_n(\theta; Y_0, ..., Y_{2n+1}) = \mathcal{H}_{\theta_0}(\theta)$.
4. Montrer à l'aide de l'inégalité de Jensen que pour tout $\theta \in \Theta_\delta$,

$$\mathcal{H}_{\theta_0}(\theta) - \mathcal{H}_{\theta_0}(\theta_0) = \mathbb{E}_{\theta_0}\left[\log \frac{p_\theta(Y_0, Y_1)}{p_{\theta_0}(Y_0, Y_1)}\right] \leq 0.$$

   En déduire à l'aide de la question précèdente et de l'identifiabilité du modèle que la fonction $\mathcal{H}_{\theta_0}(\theta)$ vérifie la propriété dite de *contraste* :

$$\forall \theta \in \Theta_\delta \quad : \quad \mathcal{H}_{\theta_0}(\theta) \leq \mathcal{H}_{\theta_0}(\theta_0), \quad \text{et} \quad \mathcal{H}_{\theta_0}(\theta) = \mathcal{H}_{\theta_0}(\theta_0) \Rightarrow \theta = \theta_0.$$

5. Conclure sur la pertinence de la méthode d'estimation décrite en (1). On pourra s'inspirer de la preuve faite en cours sur la consistance de l'estimateur du maximum de vraisemblance dans le cas d'une modèle paramétrique avec des variables aléatoires indépendantes de même loi.

$\triangle$