Examen du cours de M2 :
Méthodes numériques probabilistes

6 janvier 2012, 09h00 - 12h00.

Les notes de cours sont autorisées.

Exercice 1 : Grandes déviations et échantillonnage d'importance

Première partie : grandes déviations

1 Soit $(X_i)_{i\geq 1}$ des variables aléatoires réelles i.i.d. de carré intégrable. On note $\overline{X}_n = \sum_{i=1}^n X_i$. Pour $x > \mathbb{E}(X_1)$ fixé, on s'intéresse à la probabilité :

$$p_n(x) = \mathbb{P}(\overline{X}_n/n \geq x). \tag{1}$$

Que vaut $\lim_{n\to\infty} p_n(x)$? Pour $y \in \mathbb{R}$, que vaut $\lim_{n\to\infty} p_n(\mathbb{E}(X_1) + y/\sqrt{n})$?

2 On suppose dans cette question que $X_i = G_i$ est une suite i.i.d. de variables aléatoires réelles gaussiennes centrées réduites et on considère $\overline{G}_n = \sum_{i=1}^n G_i$. Quelle est la loi de $\overline{G}_n$? Montrer que dans ce cas,

$$\lim_{n\to\infty} \frac{1}{n} \ln p_n(x) = -\frac{x^2}{2}.$$

3 Pour $\theta \in \mathbb{R}$, soit $\Gamma(\theta) = \ln \mathbb{E}(\exp(\theta X_1))$ (qui peut éventuellement valoir $+\infty$). On introduit également $\Gamma^*(x) = \sup_{\theta\in\mathbb{R}}(\theta x - \Gamma(\theta))$. Calculer Γ et Γ^* dans le cas particulier où X_1 est une gaussienne centrée réduite. Généralement, pourquoi Γ^* est une fonction convexe ? Vérifier que $\Gamma(\theta) \geq \theta\mathbb{E}(X_1)$. En déduire que (i) $\Gamma^*(\mathbb{E}(X_1)) = 0$, (ii) pour $x > \mathbb{E}(X_1)$, $\Gamma^*(x) = \sup_{\theta\geq 0}(\theta x - \Gamma(\theta))$ et (iii) $x \mapsto \Gamma^*(x)$ est croissante sur $[\mathbb{E}(X_1), \infty)$.

4 En utilisant l'inégalité, pour tout $z \in \mathbb{R}$, $1_{z\geq 0} \leq \exp(z)$, montrer que, pour $x > \mathbb{E}(X_1)$,

$$\frac{1}{n} \ln p_n(x) \leq -\Gamma^*(x).$$

On a donc $\limsup_{n\to\infty} \frac{1}{n} \ln p_n(x) \leq -\Gamma^*(x)$. On admettra dans la suite le résultat dû à Cramer :

$$\lim_{n\to\infty} \frac{1}{n} \ln p_n(x) = -\Gamma^*(x). \tag{2}$$

De plus, on supposera que le suprémum est atteint en un point θ^* (pour le point x considéré) : $\Gamma^*(x) = \sup_{\theta\in\mathbb{R}}(\theta x - \Gamma(\theta)) = \theta^* x - \Gamma(\theta^*)$.

Seconde partie : échantillonnage d'importance

5 Soit Y une variable aléatoire réelle continue de densité p. On veut estimer $\mathbb{P}(Y \geq x)$, pour x fixé. Pour cela, on introduit une variable aléatoire continue Z, de densité q, et on l'utilise comme fonction d'importance :

$$\mathbb{P}(Y \geq x) = \mathbb{E}\left(1_{Z\geq x} \frac{p}{q}(Z)\right).$$

Vérifier l'identité ci-dessus puis calculer et identifier la densité q optimale en terme de variance.

6 Désormais, on est intéressé par estimer la probabilité $p_n(x)$ définie par (1) par une méthode de Monte Carlo qui reste efficace dans la limite n grand. On considère tout d'abord l'estimateur naïf

$$I_K(x) = \frac{1}{K} \sum_{k=1}^{K} 1_{\overline{X}_n^k/n \geq x}$$

où les $(\overline{X}_n^k)_{k\geq 1}$ désignent des réalisations i.i.d. de $\overline{X}_n$. En considérant l'erreur relative

$$\sqrt{\mathrm{Var}\left(1_{\overline{X}_n/n\geq x}\right)}/p_n(x)$$

et en utilisant (2), expliquer pourquoi, dans la limite $n \to \infty$, il faut choisir K de plus en plus grand pour obtenir une estimation de précision fixée.

7 On utilise une méthode d'échantillonage d'importance, en écrivant

$$p_n(x) = \mathbb{E}\left(1_{\overline{X}_n/n\geq x}\right) = \mathbb{E}\left(1_{\overline{Y}_n/n\geq x}\, L_n(\overline{Y}_n)\right)$$

où la loi de $\overline{Y}_n$ et la fonction $L_n : \mathbb{R} \to \mathbb{R}$ sont tels que, pour toute fonction φ mesurable bornée, $\mathbb{E}(\varphi(\overline{X}_n)) = \mathbb{E}(\varphi(\overline{Y}_n)L_n(\overline{Y}_n))$. Montrer que $\mathbb{E}(1/L_n(\overline{X}_n)) = 1$. On utilise donc comme nouvel estimateur

$$\tilde{I}_K = \frac{1}{K} \sum_{k=1}^{K} 1_{\overline{Y}_n^k/n\geq x} L_n(\overline{Y}_n^k),$$

où les $(\overline{Y}_n^k)_{k\geq 1}$ désignent des réalisations i.i.d. de $\overline{Y}_n$. On note $M_n(x) = \mathbb{E}\left(1_{\overline{Y}_n/n\geq x}\, L_n^2(\overline{Y}_n)\right)$. Montrer que $M_n(x) \geq (p_n(x))^2$. En utilisant (2), discuter le comportement asymptotique de $M_n(x)$ qui permettrait d'avoir un bon estimateur dans la limite $n \to \infty$ (c'est-à-dire un estimateur d'erreur relative bornée, pour K fixé).

8 Dans cette question, on suppose que

$$L_n(y) = \exp\left(-\theta y + n\Gamma(\theta)\right).$$

Vérifier que $\mathbb{E}(1/L_n(\overline{X}_n)) = 1$. Montrer que $\overline{Y}_n$ a même loi que $\sum_{i=1}^{n} Y_i$ avec les $(Y_i)_{i\geq 1}$ i.i.d. de loi définie par : pour toute fonction φ mesurable bornée,

$$\mathbb{E}\left(\varphi(X_1)\right) = \mathbb{E}\left(\varphi(Y_1)\exp(-\theta Y_1 + \Gamma(\theta))\right).$$

Prouver que $M_n(x) \leq \exp\left(-2n(\theta x - \Gamma(\theta))\right)$. Comment choisir θ de manière à obtenir un bon estimateur dans la limite $n \to \infty$?

9 Discuter comment réaliser l'échantillonnage d'importance présenté à la question 8 en pratique, dans le cas gaussien de la question 2.

Exercice 2 : Dynamique *overdamped Langevin* et quelques propriétés de la distribution quasi-stationnaire

On considère une dynamique de type *overdamped Langevin*

$$dX_t = -\nabla V(X_t)\,dt + \sqrt{2\beta^{-1}}dB_t, \tag{3}$$

où X_t est un processus stochastique à valeur dans $\mathbb{R}^n$, $V : \mathbb{R}^n \to \mathbb{R}$ est une fonction régulière, β est une constante réelle positive et B_t un Brownien n-dimensionnel standard. On suppose que V est tel qu'il existe une unique solution forte à (3), et que $Z = \int_{\mathbb{R}^n} \exp(-\beta V) < \infty$. On note $d\mu = Z^{-1}\exp(-\beta V(x))\,dx$ la mesure de Boltzmann-Gibbs associée à V. On introduit l'opérateur différentiel associé à l'équation différentielle stochastique (3) :

$$L = -\nabla V \cdot \nabla + \beta^{-1}\Delta \tag{4}$$

On se donne une partition de l'espace $\mathbb{R}^n$ sous la forme d'une application $\mathcal{S} : \mathbb{R}^n \to \mathbb{N}$ qui à une configuration dans $\mathbb{R}^n$ associe un entier (l'indice d'un état). On suppose que les états $W_k = \{x \in \mathbb{R}^n, \mathcal{S}(x) = k\}$ (pour $k \in \mathbb{N}$) sont des sous-ensembles de $\mathbb{R}^n$, bornés, connexes et réguliers. Dans la suite, on considère un état W_k fixé. On note W l'intérieur (supposé non vide) de W_k et on définit $T_W = \inf\{t \geq 0, X_t \notin W\}$ le premier temps de sortie de W.

1 Pour $t > 0$ fixé, on considère $(u(s,x))_{s\in[0,t],x\in W}$ une solution régulière de l'équation aux dérivées partielles :

$$\begin{cases} \dfrac{\partial u}{\partial s}(s,x) = -Lu(s,x) \text{ pour } s \in (0,t) \text{ et } x \in W, \\ u(s,x) = \varphi(x) \text{ pour } s \in (0,t) \text{ et } x \in \partial W, \\ u(t,x) = v_0(x) \text{ pour } x \in W. \end{cases} \tag{5}$$

Montrer que

$$u(0,x) = \mathbb{E}\left(1_{T_W^x < t}\,\varphi(X^x_{T_W^x})\right) + \mathbb{E}\left(1_{T_W^x \geq t}\,v_0(X_t^x)\right), \tag{6}$$

où X_t^x désigne la solution de (3) telle que $X_0^x = x$, et $T_W^x = \inf\{t \geq 0, X_t^x \notin W\}$. En déduire une formule de représentation pour une fonction $(v(t,x))_{t\geq 0, x\in W}$ solution régulière de

$$\begin{cases} \dfrac{\partial v}{\partial t}(t,x) = Lv(t,x) \text{ pour } t \geq 0 \text{ et } x \in W, \\ v(t,x) = \varphi(x) \text{ pour } t \geq 0 \text{ et } x \in \partial W, \\ v(0,x) = v_0(x) \text{ pour } x \in W. \end{cases} \tag{7}$$

2.1 Pour f et g deux fonctions $\mathcal{C}_c^\infty(W)$ (*i.e.* $\mathcal{C}^\infty(W)$ et à support compact inclus dans W), montrer que

$$\int_W fLg\,d\mu = \int_W gLf\,d\mu = -\beta^{-1}\int \nabla f \cdot \nabla g\,d\mu.$$

On introduit

$$\lambda_1 = \inf_{f\in H^1_{\mu,0}(W)} \frac{\beta^{-1}\int_W |\nabla f|^2\,d\mu}{\int_W f^2\,d\mu}, \tag{8}$$

où $H^1_{\mu,0}(W) = \left\{f : W \to \mathbb{R}, \int_W \left(f^2 + |\nabla f|^2\right)\,d\mu < \infty,\ f = 0 \text{ sur } \partial W\right\}$ est un espace de Hilbert. On utilisera aussi l'espace de Hilbert $L^2_\mu(W) = \left\{f : W \to \mathbb{R}, \int_W f^2\,d\mu < \infty\right\}$.

2.2 Montrer que $\lambda_1 > 0$, et que l'infimum dans (8) est atteint en une fonction $u_1 \in H^1_{\mu,0}(W)$. *Indication : on pourra utiliser le fait que les normes sur $H^1_{\mu,0}(W)$ et sur $H^1_0(W)$ sont équivalentes, et l'inégalité de Poincaré sur $H^1_0(W)$.*

2.3 *Question facultative, dont on peut admettre les résultats pour la suite.* Montrer qu'on peut supposer que le miniseur u_1 est tel que $u_1 \geq 0$. Montrer que u_1 est une solution faible de

$$\begin{cases} Lu_1 = -\lambda_1 u_1 \text{ sur } W, \\ u_1 = 0 \text{ sur } \partial W. \end{cases}$$

En utilisant l'inégalité de Harnack, on montre que $u_1 > 0$ sur W, et que c'est l'unique minimiseur associé à (8) à une constante multiplicative près (admis pour la suite). On introduit alors la mesure de probabilité (appelée mesure quasi-stationnaire)

$$d\nu = \frac{u_1 d\mu}{\int_W u_1 d\mu} = \frac{u_1(x) \exp(-\beta V(x))\, dx}{\int_W u_1(x) \exp(-\beta V(x))\, dx}.$$

3 Soit $t > 0$ et $f : W \to \mathbb{R}$ une fonction test mesurable bornée. Montrer que

$$\int_W \mathbb{E}\left(f(X^x_t) 1_{t \leq T^x_W}\right) \nu(dx) = \int_W f\, d\nu \int_W \mathbb{P}\left(t \leq T^x_W\right) \nu(dx).$$

En remarquant que si $(X_t)_{t\geq 0}$ est solution de (3) avec X_0 distribué suivant ν, on a $\mathbb{E}\left(f(X_t) 1_{t\leq T_W}\right) = \int_W \mathbb{E}\left(f(X^x_t) 1_{t\leq T^x_W}\right) \nu(dx)$ et $\mathbb{P}\left(t \leq T_W\right) = \int_W \mathbb{P}\left(t \leq T^x_W\right) \nu(dx)$, donner une interprétation probabiliste de cette propriété. *Indication : On pourra introduire $v(t,x) = \mathbb{E}\left(f(X^x_t) 1_{t\leq T^x_W}\right)$ et $\bar{v}(t,x) = \mathbb{P}\left(t \leq T^x_W\right)$, et considérer les dérivées en temps de $\int_W v(t,x)\, \nu(dx)$ et de $\int_W \bar{v}(t,x)\, \nu(dx)$.*

4 Soit $\varphi : \partial W \to \mathbb{R}$ une fonction test mesurable bornée et soit

$$g(t) = \mathbb{E}\left(1_{T_W < t}\, \varphi(X_{T_W})\right) = \int_W \mathbb{E}\left(1_{T^x_W < t}\, \varphi(X^x_{T_W})\right) \nu(dx).$$

Montrer que

$$g'(t) = -\lambda_1 g(t) - \beta^{-1} \int_{\partial W} \varphi \frac{\partial q}{\partial n} d\sigma_{\partial W},$$

où q est la densité de ν par rapport à la mesure de Lebesgue, n désigne la normale sortante à W et $\sigma_{\partial W}$ la mesure de Lebesgue sur le bord ∂W. En déduire une expression analytique pour $g(t)$.

5 On suppose que $(X_t)_{t\geq 0}$ est solution de (3), avec X_0 distribué suivant ν. Déduire de la question précédente la loi du couple (T_W, X_{T_W}).

Dans notre cadre d'hypothèses, il est standard de montrer (et on admettra) que l'opérateur L défini par (4) sur W avec des conditions aux limites de Dirichlet nulle sur ∂W a un spectre discret de valeurs propres négatives $(-\lambda_1, -\lambda_2, \ldots, -\lambda_n, \ldots)$ que l'on suppose ordonnées et comptées avec leur multiplicité :

$$0 > -\lambda_1 > -\lambda_2 \geq \ldots \geq -\lambda_n \geq \ldots$$

avec des fonctions propres associées $(u_1, u_2, \dots, u_n, \dots)$ qui forment une base de $H^1_{\mu,0}(W)$ et de $L^2_\mu(W)$. On suppose que pour tout $k \geq 1$, $\int_W (u_k)^2 \, d\mu = 1$.

6 On considère $v(t,x) = \mathbb{E}\left(v_0(X_t^x) 1_{t \leq T_W^x}\right)$. On suppose que $v_0 \in L^2_\mu$. Donner l'équation aux dérivées partielles satisfaite par v et écrire $v(t,x)$ sous forme de série, en fonction des $(-\lambda_k, u_k)_{k \geq 1}$.

7 On suppose que $(X_t)_{t \geq 0}$ est solution de (3), avec X_0 distribué selon μ_0, la mesure μ_0 étant à support dans W. On suppose que μ_0 admet une densité $d\mu_0/d\mu = q_0$ par rapport à μ et telle que $q_0 \in L^2_\mu$. Soit $f : W \to \mathbb{R}$ une fonction test. On considère la quantité

$$h(t) = \mathbb{E}(f(X_t) | t \leq T_W) = \frac{\mathbb{E}\left(f(X_t) 1_{t \leq T_W}\right)}{\mathbb{P}(t \leq T_W)}.$$

En utilisant la question précédente, montrer que $\lim_{t \to \infty} h(t) = \int_W f d\nu$.

8 Donner l'interprétation probabiliste de la propriété démontrée à la question précédente. En déduire que si le processus X_t reste "très longtemps" dans un état W, on peut supposer que le temps qu'il reste à y passer est exponentiel, et que le prochain état visité est indépendant de ce temps. Faire un lien avec un type de modèle en dynamique moléculaire vu en cours.

9 *Question subsidiaire.* On se place à nouveau dans le cadre d'hypothèse de la question 7. On a vu que si la condition initiale est distribuée suivant la distribution quasi-stationnaire $(\mu_0 = \nu)$ le temps de sortie est exponentiellement distribué. Que dire de la réciproque ?