Université des Sciences et Technologies de Lille
U.F.R. de Mathématiques Pures et Appliquées

IS-Math314 **Année 2007-2008**

Fiche n° 5

Ex 1. *La méthode de Monte-Carlo pour le calcul d'intégrales*

Soit $f : [0,1] \to \mathbb{R}$ une fonction continue. On se propose de donner une valeur approchée de

$$m := \int_0^1 f(x)\,\mathrm{d}x,$$

par une méthode probabiliste appelée « méthode de Monte-Carlo ». Pour cela on utilise la simulation informatique d'une suite $(U_i)_{i\geq 1}$ de variables aléatoires indépendantes et de même loi uniforme sur $[0,1]$. On pose

$$M_{2n} := \frac{1}{2n}\sum_{i=1}^{2n} f(U_i) \quad \text{et} \quad \tilde{M}_{2n} := \frac{1}{2n}\sum_{i=1}^{n}\left(f(U_i)+f(1-U_i)\right)$$

1) Expliquer pourquoi $X_1 := f(U_1)$ et $Y_1 := f(U_1)+f(1-U_1)$ sont intégrables et exprimer leur espérance à l'aide de l'intégrale m.

2) En vous appuyant sur un théorème du cours, montrer que les suites $(M_{2n})_{n\in\mathbb{N}^*}$ et $\left(\tilde{M}_{2n}\right)_{n\in\mathbb{N}^*}$ convergent presque sûrement vers m quand n tend vers $+\infty$.

Ce résultat légitime pour n « grand » l'approximation de m par la valeur $M_{2n}(\omega)$ (ou $\tilde{M}_{2n}(\omega)$) calculée à partir de l'échantillon généré par l'ordinateur.

3) Expliquer pourquoi X_1 et Y_1 sont de carré intégrable et exprimer leur variance à l'aide de la fonction f.

4) On suppose de plus que f est croissante sur $[0,1]$.

a) Montrer que pour tout $(x,y)\in[0,1]^2$, $(f(x)-f(y))\,(f(1-x)-f(1-y))\leq 0$.

b) En déduire que $\mathbf{E}\left[f(U_1)f(1-U_1)\right]\leq \mathbf{E}\left[f(U_1)\right]^2$.

c) Comparer les variances de M_{2n} et de $\tilde{M}_{2n}$.

d) Finalement, quelle valeur choisiriez-vous entre $M_{2n}(\omega)$ ou $\tilde{M}_{2n}(\omega)$ pour approximer m ?

5) En utilisant le théorème limite central, proposer un intervalle de confiance pour m de niveau 95% (en négligeant l'erreur due à l'approximation gaussienne) en supposant que l'on connait un majorant M de $\sup_{x\in[0,1]}|f(x)|$.

Ex 2. Un produit commercialisé est présenté dans des boites sur lesquelles on peut lire : *contenance 500 grammes*. La quantité contenue dans une boite choisie au hasard est en réalité une variable aléatoire gaussienne X.
A partir des données fournies, construire un intervalle de confiance pour la moyenne de X, au niveau de confiance 90%. Construire également un intervalle de confiance pour l'écart-type de X, avec le même niveau de confiance.

Données (en grammes) : 490 ; 490 ; 490 ; 492 ; 492 ; 495 ; 497 ; 497 ; 502 ; 505.

Ex 3. Dans un laboratoire, on a mis au point un nouvel appareil de mesure du rayonnement émis par un objet porté à haute température. De par la construction de l'appareil, on sait que les résultats des mesures sont gaussiens, avec pour moyenne la température de l'objet. Leur écart-type σ est lié à la qualité de l'appareil, et pour l'évaluer, on effectue avec six mesures sur un objet dont on connait la température : 1600 degrés.

1) On note $X_1,\dots,X_6$ les résultats des 6 mesures effectuées. Ecrire l'expression de la moyenne empirique et de la variance empirique des X_i. En utilisant le théorème de Student, déterminer un intervalle de confiance pour σ au niveau de confiance 99%. Quel est le défaut de cette méthode ?

2) A partir de la variable aléatoire $W = \sum_{i=1}^{6} \left(\frac{X_i - 1600}{\sigma} \right)^2$ construire un meilleur intervalle de confiance pour σ au niveau de confiance 99%.

Données : 1597.43 , 1597.96 , 1590.26 , 1590.46 , 1612.31 , 1610.40
Pour simplifier les calculs, on a calculé la somme de ces valeurs : 9598.82
et la somme de leurs carrés : 15356680.34

Ex 4. *Estimation de la durée d'une panne*

Des étudiants font un TP en utilisant n ordinateurs reliés à un serveur [1]. Chaque ordinateur envoie au serveur, à des instants aléatoires, des requêtes variées (exécution de commande, transmission de données, etc).

A l'instant $t = 0$ ce serveur tombe en panne. Il redémarre au bout d'une durée τ.

Les étudiants ne savent pas quelle est la durée τ de la panne. Mais pour i de 1 à n, ils constatent au bout de quel temps T_i après le début de panne la première requête de l'ordinateur i est acceptée par le serveur. Pour estimer la durée τ de la panne, ils disposent donc de n données $T_1(\omega), \cdots, T_n(\omega)$.

Les requêtes ayant lieu à des instants aléatoires, il est naturel de supposer que la durée $X_i = T_i - \tau$ entre le redémarrage du serveur et la première requête de l'ordinateur i suit une loi exponentielle. Le paramètre a de cette loi exponentielle sera supposé connu et identique pour toutes les machines (que représente-t-il ?). On supposera également que les v.a. $X_1, \cdots, X_n$ sont indépendantes.

1) Quelle est la loi de la v.a. $V = \inf(X_1, \cdots, X_n)$?

2) Estimer la durée inconnue τ à partir de $T_1, \cdots, T_n$ en utilisant la méthode du maximum de vraisemblance. On note W l'estimateur obtenu.

3) Calculer le biais de W. En déduire un estimateur sans biais Z de la durée de panne. Quelle est la variance de ce nouvel estimateur ?

4) Fabriquer un autre estimateur sans biais de τ, en utilisant cette fois la moyenne empirique $\overline{T}$ des T_i. Quelle est sa variance ?

1. Concrètement, un gros ordinateur sans écran installé dans un local verrouillé, climatisé, et sous alarme. Le seul point important ici est que les étudiants ont la possibilité d'utiliser le serveur, mais pas de le voir.

Ex 5. *Un mélange de gaussiennes (issu de l'examen du 30 juin 2007)*

On considère le modèle statistique $\big(\Omega, \mathcal{F}, (P_\theta)_{\theta\in[0,1]}\big)$ et une variable aléatoire $X : \Omega \to \mathbb{R}$, dont la loi sous P_θ a pour fonction de répartition F_θ définie par :

$$\forall \theta \in]0,1[,\ \forall x \in \mathbb{R}, \quad F_\theta(x) = (1-\theta)\Phi\Big(\frac{x-\mu}{\sigma}\Big) + \theta\Phi\Big(\frac{x-\mu}{5\sigma}\Big),$$

où Φ est la fonction de répartition de la loi $\mathfrak{N}(0,1)$. Dans tout l'exercice, les paramètres $\mu \in \mathbb{R}$ et $\sigma \in]0,+\infty[$ sont supposés connus.

1) Quelle est la loi de X sous P_0 ? Sous P_1 ? *Indication :* calculez $P_0\big((X-\mu)/\sigma \leq x\big)$.

2) Expliquez pourquoi la loi de X sous P_θ admet une densité f_θ et calculez la.

3) Calculez $\mathbf{E}_\theta X$.

4) Calculez $\operatorname{Var}_\theta X$.

5) Au vu du calcul précédent, proposez un estimateur T_n fortement consistant de θ, basé sur un échantillon $X_1, \dots, X_n$ de v.a. indépendantes de même loi que X sous P_θ.

6) Lors d'une séance de T.D. sur machine à laquelle vous avez échappé, l'enseignante a demandé à son groupe de proposer une méthode de simulation de la loi de X sous P_θ. L'étudiant G. Duflair a proposé de définir une variable aléatoire Y par

$$Y := (\sigma Z + \mu)\mathbf{1}_{\{U>\theta\}} + (5\sigma Z + \mu)\mathbf{1}_{\{U\leq\theta\}},$$

où U est une v.a. de loi uniforme sur $[0,1]$ et Z *indépendante* de U est gaussienne $\mathfrak{N}(0,1)$. Les variables U et Z sont fournies facilement par le générateur de nombre aléatoires utilisé. G. Duflair s'apprête à en tirer un programme Scilab, quand l'enseignante lui demande de justifier sa réponse. Pouvez vous l'aider ?

Ex 6. *Estimation de médiane*

Sur l'espace probabilisé $(\Omega, \mathcal{F}, P)$, soit $(X_i)_{i\geq 1}$ une suite de variables aléatoires réelles indépendantes de même loi de f.d.r. F. On suppose que F^{-1} définie par

$$\forall u \in]0,1[, \quad F^{-1}(u) := \inf\{t \in \mathbb{R};\ F(t) \geq u\}.$$

est continue au point $\frac{1}{2}$. Donner une condition suffisante sur F pour qu'il en soit ainsi.

On se propose d'estimer la plus petite médiane de la loi des X_i, donc $F^{-1}(\frac{1}{2})$ à partir de l'observation d'un échantillon de grande taille. On note F_n la f.d.r. empirique construite sur l'échantillon $X_1, \dots, X_n$ et on pose

$$M_n := F_n^{-1}\Big(\frac{1}{2}\Big) = \inf\Big\{t \in \mathbb{R};\ F_n(t) \geq \frac{1}{2}\Big\}.$$

Le but de cet exercice est de prouver que M_n converge presque-sûrement vers $F^{-1}(\frac{1}{2})$ lorsque n tend vers $+\infty$.

1) En écrivant $F(M_n) = F_n(M_n) + \big(F(M_n) - F_n(M_n)\big)$ et en comparant $F_n(M_n)$ avec $\frac{1}{2}$, montrer grâce au théorème de Glivenko-Cantelli, qu'il existe un $\Omega_1 \in \mathcal{F}$ tel que $P(\Omega_1) = 1$ et

$$\forall \omega \in \Omega_1,\ \forall \varepsilon > 0,\ \exists N_1 = N_1(\omega, \varepsilon),\ \forall n \geq N_1, \quad F(M_n(\omega)) > \frac{1}{2} - \varepsilon. \qquad (1)$$

2) De même en écrivant $F(M_n-\varepsilon)=F_n(M_n-\varepsilon)+\big(F(M_n-\varepsilon)-F_n(M_n-\varepsilon)\big)$, pour tout $\varepsilon>0$, et en comparant $F_n(M_n-\varepsilon)$ avec $\frac{1}{2}$, montrer qu'il existe un $\Omega_2\in\mathcal{F}$ tel que $P(\Omega_2)=1$ et

$$\forall\omega\in\Omega_2,\ \forall\varepsilon>0,\ \exists N_2=N_2(\omega,\varepsilon),\ \forall n\geq N_2,\quad F(M_n(\omega)-\varepsilon)<\frac{1}{2}+\varepsilon. \qquad (2)$$

3) Déduire de ce qui précède l'existence d'un $\Omega_0\in\mathcal{F}$ de probabilité 1 tel que

$$\forall\omega\in\Omega_0,\ \forall\varepsilon>0,\ \exists N_0=N_0(\omega,\varepsilon),\ \forall n\geq N_0,\quad F^{-1}\Big(\frac{1}{2}-\varepsilon\Big)\leq M_n(\omega)\leq F^{-1}\Big(\frac{1}{2}+\varepsilon\Big).$$

Conclure sur la convergence p.s. de M_n.

4) À quel endroit a-t-on vraiment besoin de la convergence *uniforme* presque sûre du théorème de Glivenko-Cantelli ?

Entraînement supplémentaire facultatif :

Ex 7. *Contrôle de bouteilles (issu de l'examen du 29 mai 2007)*

Dans un contrôle de production, on mesure la masse de 10 bouteilles vides du même modèle. On considère cette masse comme une variable aléatoire *gaussienne* d'espérance m et d'écart-type σ inconnus. Les données relevées (en grammes) sont les suivantes.

248,282	250,354	249,934	249,548	250,710
252,005	251,630	247,752	251,107	249,494

Les intervalles de confiance demandés ci-dessous devront être justifiés par un résultat du cours. La somme des valeurs observées est 2 500,816 et celle de leurs carrés est 625 425.

1) Proposez un intervalle de confiance au niveau 95% pour m.

2) Proposez un intervalle de confiance au niveau 90% pour σ.