Université de Lille 1
U.F.R. de Mathématiques Pures et Appliquées

IS Math 314 **Année 2008–2009**

Fiche n° 4

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 $\mathbf{E}\,[f(U_1)f(1-U_1)] \leq \mathbf{E}\,[f(U_1)]^2$.

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

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

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

Ex 2. *Test sur les intentions de vote*

1) Soit S_n une variable aléatoire sur $(\Omega, \mathcal{F}, (P_\theta)_{\theta\in[0,1]})$ de loi binomiale $\mathrm{Bin}(n,\theta)$ sous P_θ.

a) Montrer que, pour toute constante a, l'application

$$\theta \mapsto P_\theta(S_n \geq a)$$

est croissante sur $[0,1]$.
Indication : on utilisera une méthode de simulation de la loi binomiale.

b) Soit $\theta_0 \in [0,1]$. En déduire la construction d'un évènement de la forme

$$R = \{S_n \geq a\}$$

de telle sorte que $\sup_{\theta \leq \theta_0} P_\theta(R) \leq 0,05$.

2) Application à la construction d'un test : avant une élection, un candidat A affirme qu'il est sûr de dépasser le score des 30%. Pour savoir s'il a raison, on interroge un nombre n suffisant de personnes et on note S_n le nombre de personnes parmi les n interrogées qui ont l'intention de voter pour le candidat A. Si le nombre S_n dépasse un certain seuil (à déterminer), le candidat A affirmera que son score dépasse 30% en souhaitant que le risque de se tromper n'excède pas 5%.

On introduit alors le paramètre θ représentant la probabilité (inconnue) de voter pour A sur l'ensemble des électeurs et un modèle statistique $(\Omega, \mathcal{F}, (P_\theta)_{\theta \in]0,1[})$ dans lequel S_n suit une loi binomiale $\mathrm{Bin}(n, \theta)$ sous P_θ.

On observe une réalisation de S_n et à l'aide de cette observation, on veut décider l'une des deux hypothèses suivantes :

$$H_0 \quad : \quad \theta \leq 0,3 \qquad \text{ou bien} \qquad H_1 \quad : \quad \theta > 0,3.$$

La règle de décision sera donc du type

- si on observe $S_n(\omega) \geq a$, alors on rejette H_0 (et on accepte H_1)
- sinon, on accepte l'hypothèse H_0,

a étant un seuil à déterminer. Il est déterminé de telle sorte que la probabilité de rejeter H_0 à tort soit contrôlée par un niveau donné (ici 5%). Par exemple, c'est la probabilité pour le candidat A de conclure que $\theta > 0,3$ alors qu'en réalité $\theta \leq 0,3$.

a) Déterminer le seuil a (en fonction de n).

b) On recueille 32% d'intentions de votes pour A auprès de 500 personnes, que pouvez-vous en conclure ?

Ex 3. *Les ex-aequo d'un échantillon*

1) On se propose de montrer que si X et Y sont deux variables aléatoires réelles *indépendantes* de fonctions de répartition respectives F et G, et si F est *continue*, alors $P(X = Y) = 0$. Voici quelques indications. On note pour $M \in \mathbb{N}^*$,

$$\Delta_M := \{(x,y) \in [-M,M]^2;\ x = y\} \quad \text{et} \quad \Delta := \{(x,y) \in \mathbb{R}^2;\ x = y\}.$$

Il s'agit de prouver que $P_{(X,Y)}(\Delta) = 0$ (pourquoi ?). En remarquant [1] que pour tout $k \in \mathbb{N}^*$,

$$\Delta_M \subset \bigcup_{-Mk \leq j < Mk} \left[\frac{j}{k}, \frac{j+1}{k}\right]^2,$$

utiliser la continuité *uniforme* de F sur le *compact* $[-M, M]$ pour montrer que

$$\forall \varepsilon > 0, \quad P_{(X,Y)}(\Delta_M) \leq \varepsilon\big(G(M) - G(-M)\big).$$

En déduire que $P_{(X,Y)}(\Delta_M) = 0$, puis que $P_{(X,Y)}(\Delta) = 0$.

2) Montrer que si $X_1, \ldots, X_n$ est un échantillon de la loi de f.d.r. F continue, presque sûrement il n'y a pas d'*ex-aequo* dans l'échantillon.

[1] Dessin impératif !

3) On suppose que la f.d.r. F est discontinue au point x_0 et on note

$$N_n(x_0) := \operatorname{card}\{i = 1, \dots, n;\ X_i = x_0\}.$$

Montrer que $N_n(x_0) \xrightarrow[n\to+\infty]{\text{p.s.}} +\infty$. On donnera un résultat plus précis que cette affirmation.

Ex 4. *Calculs sur la f.d.r. empirique*

On note $X_1, \dots, X_n$ un échantillon de la loi de f.d.r. F, *i.e.* les X_i sont indépendantes et de même loi de f.d.r. F. On note F_n la fonction de répartition empirique associée à l'échantillon :

$$\forall x \in \mathbb{R}, \quad F_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{\{X_i \le x\}}.$$

1) Tracer le graphe de F_n pour une réalisation de l'échantillon $X_1, \dots, X_n$ de votre choix (en prenant $n = 5$ par exemple).

2) On suppose dans cette question que les X_i sont des v.a. positives (ou plus généralement que $P(X_1 < 0) = 0$). Calculez $\int_0^{+\infty}(1 - F_n(t))\,\mathrm{d}t$ et interprétez le résultat.

Entraînement supplémentaire facultatif

Ex 5. *Polygone de fréquences cumulées croissantes*

Soit $X_1, \dots, X_n$ un échantillon, F la f.d.r. de chaque X_i et F_n la fonction de répartition empirique associée à cet échantillon. On note $X_{n:1}, \dots, X_{n:n}$ les statistiques d'ordre de l'échantillon, obtenues par le réordonnement croissant de l'échantillon, *i.e.* pour tout $\omega \in \Omega$, $\{X_1(\omega), \dots, X_n(\omega)\} = \{X_{n:1}(\omega), \dots, X_{n:n}(\omega)\}$ et $X_{n:i}(\omega) \le X_{n:i+1}(\omega)$ pour $i = 1, \dots, n-1$. En particulier $X_{n:1} = \min_{1\le i\le n} X_i$ et $X_{n:n} = \max_{1\le i\le n} X_i$. On posera par commodité $X_{n:0} := X_{n:1} - 1$ (convention non standard). On définit le lissage polygonal G_n de F_n, comme la fonction continue affine par morceaux valant 1 à droite de $X_{n:n}$, valant 0 à gauche de $X_{n:1} - 1$, coïncidant avec F_n en chaque $X_{n:i}$ (donc en chaque point de saut de F_n) et interpolant linéairement entre deux sauts consécutifs de F_n.

1) Représentez F_n et G_n pour les deux échantillons observés suivants

$$X_1(\omega) = 2;\ X_2(\omega) = 1{,}5;\ X_3(\omega) = 4;\ X_4(\omega) = 0{,}5.$$

$$X_1(\omega') = 3;\ X_2(\omega') = 2;\ X_3(\omega') = 2;\ X_4(\omega') = 1{,}5;\ X_5(\omega') = 4.$$

2) On suppose dans cette question F continue sur $\mathbb{R}$. En utilisant le théorème de Glivenko-Cantelli, montrez que presque-sûrement G_n converge vers F uniformément sur $\mathbb{R}$ quand n tend vers $+\infty$.

3) Donnez un argument général montrant que cette convergence uniforme presque sûre ne peut avoir lieu lorsque F n'est pas continue sur $\mathbb{R}$. Proposez une autre explication de ce résultat négatif en vous basant sur l'exercice 3.