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

IS Math 314 **Année 2009–2010**

Fiche n° 4

Ex 1. *Intervalles de confiance pour une probabilité inconnue (Première session 2009)*

Dans cet exercice qui est une question de cours à peine déguisée, on dispose d'un 400-échantillon observé $(X_i(\omega))_{1\leq i\leq 400}$ d'une loi de Bernoulli. Le paramètre p de cette loi est inconnu, mais on sait avec certitude que $0 < p < 0{,}3$. Le nombre de valeurs 1 dans l'échantillon observé est 72.

1) Compte-tenu de ces informations, quel est le meilleur majorant que l'on peut proposer pour $\operatorname{Var} X_1$?

2) Donnez un intervalle de confiance I au niveau 95% pour p en utilisant la méthode avec variance *majorée* et en indiquant avec précision sur quel théorème du cours vous vous appuyez.

3) On note $\overline{X}$ et S^2 la moyenne et la variance empiriques de l'échantillon $(X_i)_{1\leq i\leq 400}$. Vérifiez que

$$S^2 = \overline{X}(1-\overline{X}).$$

Indication : comparez X_i et X_i^2 lorsque X_i est à valeurs dans $\{0,1\}$.

4) Donnez un intervalle de confiance J au niveau 95% pour p en utilisant la méthode avec variance *estimée* et en indiquant avec précision sur quel théorème du cours vous vous appuyez.

Ex 2. *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

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

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 $\big(\tilde{M}_{2n}\big)_{n\in\mathbb{N}^*}$ 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 3. *Échauffement gaussien (Première session 2008)*

Soient X et Y, deux variables aléatoires définies sur le même espace probabilisé, indépendantes et de même loi gaussienne standard $\mathfrak{N}(0,1)$. On pose $Z := X + Y$.

1) Calculez $\operatorname{Cov}(Z,Y)$ après avoir justifié son existence.

2) Z et Y sont-elles indépendantes ?

3) Le vecteur (Z,Y) est-il gaussien ?

Seules les réponses argumentées seront prises en compte.

Ex 4. *Vous reprendrez bien un peu de vecteurs gaussiens ? (Seconde session 2007)*

Soient X_1 et X_2 deux variables aléatoires définies sur le même espace probabilisé $(\Omega,\mathcal{F},P)$, indépendantes et de même loi gaussienne $\mathcal{N}(0,1)$. Soient a et b des réels vérifiant $a^2+b^2=1$. On définit les variables aléatoires Y_1 et Y_2 par l'égalité matricielle :

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} a & -b \\ b & a \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}.$$

1) Expliquez *sans aucun calcul* pourquoi le vecteur aléatoire $V = (Y_1, Y_2)$ est un vecteur gaussien de $\mathbb{R}^2$.

2) Calculez le vecteur espérance et la matrice de covariance de (Y_1,Y_2). Comparez les lois des vecteurs (X_1,X_2) et (Y_1,Y_2).

3) Que peut-on dire du vecteur aléatoire image de (X_1,X_2) par une rotation de centre $(0,0)$ et d'angle 38° ?

Ex 5. *(Partiel 2007)*

Pour tout $n \geq 1$, soit $N_n = (N_{n,1},\dots,N_{n,d})$ un vecteur aléatoire suivant la loi multinomiale de paramètres n et $p = (p_1,\dots,p_d)$. Utilisez le théorème limite central vectoriel pour montrer que pour toutes constantes a et b, la suite $(Y_n)_{n\geq 1}$ de variables aléatoires réelles définies par

$$Y_n := \sqrt{n}\Big(a\frac{N_{n,1}}{n} + b\frac{N_{n,2}}{n} - (ap_1+bp_2)\Big)$$

converge en loi vers une gaussienne Y dont vous donnerez l'espérance et la variance.

Ex 6. *Une marche aléatoire dans $\mathbb{R}^2$ (Partiel 2009)*

On se propose d'étudier le comportement asymptotique de la trajectoire d'une particule qui partie de O, choisit à chaque instant $k \in \mathbb{N}$ une direction au hasard et parcourt une distance 1 dans cette direction, avant de choisir une nouvelle direction à l'instant $k+1$. Les questions 1) et 2) sont des préliminaires qui peuvent se traiter comme des exercices indépendants. Commencez par lire l'énoncé jusqu'au bout et par repérer les questions qui peuvent se traiter facilement en admettant le résultat de la question précédente.

1) Soit U une variable aléatoire de loi uniforme sur $[0, 2\pi]$. On pose $X = \cos U$, $Y = \sin U$. Calculez l'espérance et la matrice de covariance du vecteur aléatoire (X, Y). X et Y sont-elles indépendantes ?

2) Soit $(U_k)_{k\geq 1}$ une suite de variables aléatoires réelles indépendantes et de même loi et h une fonction continue bornée sur $\mathbb{R}$. Montrez que

$$\frac{1}{n}\sum_{k=1}^{n} h(U_k) \xrightarrow[n\to+\infty]{\text{p.s.}} \mathbf{E}h(U_1). \tag{1}$$

3) Revenant à la marche aléatoire décrite en introduction, on note M_n la position occupée par la particule à l'instant n, avec $M_0 = O = (0, 0)$. En notant S_n et T_n les coordonnées de M_n dans le repère canonique de $\mathbb{R}^2$, on a donc

$$M_n = (S_n, T_n), \qquad S_n = \sum_{k=1}^{n} \cos U_k, \qquad T_n = \sum_{k=1}^{n} \sin U_k,$$

où $(U_k)_{k\geq 1}$ est une suite de variables aléatoires indépendantes de même loi uniforme sur $[0, 2\pi]$ représentant la succession des choix de direction « au hasard ». En vous appuyant sur les questions préliminaires, expliquez pourquoi presque-sûrement, la distance OM_n est un $o(n)$ quand n tend vers $+\infty$.

4) Montrez en utilisant le TLC vectoriel que

$$\frac{1}{\sqrt{n}}(S_n, T_n) \xrightarrow[n\to+\infty]{\text{loi}} W, \tag{2}$$

où W est un vecteur aléatoire de loi gaussienne $\mathfrak{N}(\mu, K)$ avec :

$$\mu = (0, 0), \qquad K = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix}.$$

Expliquez pourquoi W est à composantes indépendantes et à densité

$$f : (x, y) \longmapsto \frac{1}{\pi}\exp\big(-(x^2 + y^2)\big).$$

5) Soit D_r le disque de centre O et de rayon r. Calculez explicitement $P(W \in D_r)$.

6) La signification pratique du résultat de la question 3 est que pour n « grand », le point M_n va se trouver dans un disque de centre O et de rayon *très inférieur* à n. Utilisez les deux questions précédentes pour préciser ce résultat en montrant que $P(OM_n \leq r\sqrt{n})$ converge vers $1 - \exp(-r^2)$ quand n tend vers l'infini.

7) Voici une petite application numérique en guise d'illustration. On prend comme unité 1 cm (distance parcourue par la particule entre deux changements de direction). À quelle distance maximale de O peut se trouver théoriquement le point $M_{10\,000}$? Dans quel disque minimal allez vous le chercher si vous souhaitez le localiser avec une probabilité de succès de 99% ?