EXAMEN MAO – 10 JUIN 2004

GILLES STOLTZ

Déroulement de l'examen

La durée de l'examen est de 4 heures. Je vous rappelle que comme tout examen public, les tentatives de triches sont passibles de 5 ans d'interdiction d'examens et de concours de la fonction publique (y compris l'agrégation, le CAPES et tout DEA). Le seul document autorisé est le résumé que vous avez pu écrire sur un recto d'une feuille A4. Vous me rendrez :

– sur une copie d'examen, les réponses aux questions théoriques, ainsi que les résultats numériques indexés par `num` dans cet énoncé,
– le listing *commenté* des fonctions ou des scripts dont vous aurez eus besoin (indexés par `fun`),
– les graphiques (`grp`) ; portez-leur une attention toute particulière : ils doivent être lisibles. N'oubliez pas qu'au lieu de différencier les courbes par leur couleur, vous pouvez changer la matière du trait (options `'dash'`, `'solid'` pour `plot`, tracé de points symbolisés par des diamants ou des triangles plutôt que de courbes, *etc*).

NB : Vous avez le droit de traiter différemment les trois problèmes de modélisation proposés, d'ajouter des questions avec vos propres réponses, de me démontrer que mes questions sont sans intérêt (ayez de bons arguments, quand même), *etc.* C'est la rigueur de la démarche de modélisation et d'implémentation effective que je note. Voyez les questions ci-dessous comme des pistes, et sentez-vous libres d'en faire plus que demandé et d'explorer ces problèmes avec plus de profondeur.

1. Deux exercices...

Exercice 1 [Calculs de π]. On veut calculer une approximation de π par méthodes de Monte-Carlo. **5 pts**

(1) Etant donnée une fonction $f : [0,1] \to \mathbb{R}$, intégrable, rappeler comment calculer une estimation de son intégrale sur $[0,1]$, à l'aide d'une suite auxiliaire de variables aléatoires i.i.d. selon une loi uniforme sur $[0,1]$, notée $U_1,U_2,\ldots$.

(2) Préciser les garanties théoriques que l'on *peut* avoir sur le résultat du calcul. Plus précisément, je veux deux types de garanties non-asymptotiques, une garantie faisant intervenir un terme de variance et une autre ne dépendant que des bornes $\mathbb{L}^\infty$ des variables aléatoires considérées. Pour chacune, vous préciserez bien les hypothèses d'application.

(3) L'élève cosinus propose la fonction

$$f(x) = 2\frac{1}{\sqrt{1-x^2}} \, .$$

En fonction de votre réponse à la question précédente, ce choix vous semble-t-il bien adapté ? (Rappel : la primitive de f est $2\mathrm{Arc}\sin$.)

(4) Que dire de l'élève lambda qui propose quant à lui

$$g(x) = 4\sqrt{1-x^2} \ ?$$

(5) Fixez-vous un nombre N de variables uniformes auxquelles vous avez accès pour que les approximations soient toutes deux, si cela est possible, garanties à $\varepsilon = 10^{-2}$ près avec probabilité $1-\delta = 95\%$, et écrivez un fichier script [1] réalisant deux séries de 10 approximations, la première série avec f, l'autre avec g. **fun**

(6) Indiquez pour chaque série l'écart de la plus mauvaise approximation à π et donnez en une phrase la morale de cet exercice (qui de f et g est la meilleure en théorie et qui est celle en pratique, et surtout pourquoi)... **num**

Exercice 2 [Approximation binômiale-Poisson]. Il est connu qu'une suite de variables aléatoires de lois binômiales $\mathcal{B}(n,p_n)$ converge en loi vers une Poisson de paramètre λ dès que $np_n \to \lambda$. Ainsi, lorsque n est grand mais que $n\,p$ est petit (entre 1 et 10), on approche volontiers une binômiale de paramètres n et p par une Poisson de paramètre $n\,p$. C'est ce que l'on appelle l'approximation binômiale par Poisson. **5 pts**

On a en plus la vitesse de convergence suivante. Soit B_n une suite de variables aléatoires de lois $\mathcal{B}(n,\lambda/n)$. Alors :

$$\sup_{k\geq 0} \left| \mathbb{P}[B_n = k] - e^{-\lambda}\frac{\lambda^k}{k\,!} \right| \leq \frac{2\lambda^2}{n} \, .$$

(1) Donner un exemple de situation pratique où l'on peut faire cette approximation (cela revient à me donner un exemple de $\mathcal{B}(n,\lambda/n)$ avec n grand et λ petit).

(2) Ecrire un fichier script [2] simulant des N-échantillon de $\mathcal{B}(n,\lambda/n)$, pour n décrivant `[10 25 50 100]`, et $\lambda = 5$ (vous choisirez vous-même une valeur convenable pour N, en la justifiant). Pour chaque valeur de n, vous montrerez d'abord graphiquement [3] la proximité éventuelle de B_n à une Poisson de paramètre λ, et ce, en utilisant vos N-échantillons. N'oubliez pas de *justifier* le choix du type de représentation graphique. **fun** **grp**

(3) Enfin, toujours dans le même fichier script, calculez quatre valeurs numériques, une pour chaque n, me quantifiant la dite proximité. Comparez ces valeurs à celles proposées par la garantie théorique. **num**

2. … et un problème

Exercice 3 [Processus de Galton-Watson]. Voici un exemple classique de processus de population. Une cellule se divise, et donne naissance à D cellules, où $D \in \{0,1,2\}$: $D = 2$ signifie que la mitose cellulaire s'est bien déroulée, ce qui arrive avec probabilité p_2. Le cas $D =$ **10 pts**

---

1. et non une fonction !
2. ici non plus, je ne veux pas voir de fonction, mais une suite de commandes dans un fichier script !
3. Attention, la qualité visuelle des graphiques entrera en compte dans la notation. Faites attention à bien être centrés sur les endroits d'intérêt, je ne veux pas voir de grands blancs et une concentration d'informations dans un espace réduit : sachez gérer, si possible automatiquement (dans le script), l'affichage.

1 correspond à l'obtention d'une cellule non viable parmi les deux cellules-filles (probabilité p_1), et $D = 0$ désigne les cas où la cellule mère meurt pendant la division (probabilité p_0). On note m la moyenne de la loi de D, et σ^2 sa variance.

Soit $(D_{n,k})_{n\geq 1,k\geq 1}$ des variables aléatoires i.i.d. de loi celle de D. On part de $Z_0 = 1$ cellule, dont on étudie la descendance. Pour $n \geq 1$, le nombre de cellules à la génération n est ainsi donné par

$$Z_n = \sum_{k=1}^{Z_{n-1}} D_{n,k} \ .$$

(1) Justifier que $(Z_n)_{n\geq 0}$ forme une chaîne de Markov.

On note q la probabilité d'extinction, *id est*,

$$q = \mathbb{P}\left[\exists n \in \mathbb{N}^* \,|Z_n = 0\right] \ ,$$

et on admet que $\mathbb{E}[Z_n] = m^n$.

(2) Soit q_n la probabilité que $Z_n = 0$. Montrer que $(q_n)_{n\geq 0}$ est croissante et préciser sa limite.

(3) Dans le cas (dit sous-critique) $m < 1$, par exemple pour

$$(p_0, p_1, p_2) = (1/5, 7/10, 1/10) \ ,$$

écrire une fonctions calculant un N-échantillon de $(Z_0,Z_1,\ldots,Z_n)$, où $n = 100$ et N est assez grand (justifier sa valeur), et illustrant graphiquement la convergence de la suite des probabilités d'extinction $(q_n)_n$ vers sa valeur limite. Que semble valoir q ? Enfin, tracez sur un même graphique une suite d'estimations empiriques des $(\mathbb{E}[Z_n])_{n\geq 0}$ contre les valeurs théoriques $(m^n)_{n\geq 0}$. `fun` `grp` `grp`

Dans le cas sur-critique $m > 1$, on a convergence presque-sûre et dans $\mathbb{L}^2$ de Z_n/m^n vers une variable aléatoire W, de moyenne 1 et de variance égale à $\sigma^2/(m^2 - m)$.

(4) On fixe

$$(p_0, p_1, p_2) = (1/10, 7/10, 1/5) \ .$$

Pour $n = 100$ et N bien choisi, écrire une fonction qui simule une approximation empirique de la loi de W : cette fonction estimera la probabilité que $W = 0$, ainsi que sa variance et sa moyenne, et les comparera aux valeurs proposées par la théorie. Ensuite, vous représenterez graphiquement la loi de W par un histogramme correctement calibré. `fun` `num` `grp`

Dans le cas critique $m = 1$, il y a extinction presque-sûre, à la vitesse :

$$\mathbb{P}\left[Z_n = 0\right] \sim \frac{2}{\sigma^2 n} \ .$$

De plus, les lois des Z_n/n, conditionnés à la non-extinction (i.e. à $Z_n > 0$), convergent vers une loi exponentielle de paramètre $\sigma^2/2$.

(5) Illustrez ces deux faits selon votre convenance, en justifiant votre démarche et vos choix de représentation. Pour cela, considérez `fun, grp`

$$(p_0, p_1, p_2) = (3/20, 7/10, 3/20) \ .$$