Second concours ENS Saclay/Rennes 2018

Sujet 1 : Probabilités et Statistique - Rapport

Le sujet porte sur l'étude de processus d'urnes avec renforcement, qui sont utilisés dans de nombreux domaines (marches aléatoires renforcées, modèles de graphes aléatoires, etc.). Un résultat prouvé dans ce sujet qui a un intérêt intrinsèque est une version du Théorème de de Finnetti portant sur les suites de variables échangeables. Les outils utilisés dans ce sujet sont des arguments de martingale (inégalités maximales et convergence) et de méthode de moments.

Les notes vont de 4,4 à 18,2 ; moyenne : 10,8, écart-type : 3,87.

Le sujet comprend 4 parties. Le sujet était long et il n'était pas nécessaire d'aborder toutes les parties pour avoir une très bonne note. Chacune d'entre elles comportait des questions élémentaires (parfois injustement maltraitées dans beaucoup de copies) qu'il était possible d'aborder en admettant les questions/parties précédentes. Cependant, le barème a favorisé l'effort manifeste de certains candidats de traiter des parties dans leur intégralité (quitte à en faire moins que d'autres), vis-à-vis des stratégies visant à glaner des points ici et là.

Quelques remarques générales

— De manière générale, les candidats gagneraient à expliciter précisément quelles hypothèses ils utilisent et à quel moment. La stratégie visant à accumuler les hypothèses dans l'espoir que l'hypothèse correcte se trouve au milieu d'entre elles n'est pas acceptable et a été sanctionnée.

  Un exemple : quand on trouve dans une copie "car les variables sont i.i.d.", utilise-t-on à ce moment le caractère indépendant des variables, le fait qu'elles ont même loi ou a-t-on vraiment besoin des deux ?

— On aimerait voir dans les énoncés d'égalité entre deux variables $X = Y$, la mention "presque sûrement" qui est souvent oubliée dans la plupart des copies.

  Plus généralement, à la lecture de certaines copies, l'absence de cette mention "p.s." fait qu'on a l'impression de ne plus manipuler des variables aléatoires mais seulement des nombres déterministes, ce qui a été à l'origine d'erreurs de raisonnement. Ainsi, quand certains candidats écrivent que, pour N temps d'arrêt, le contraire de "$N < \infty$" est "$N = \infty$", c'est au mieux source de confusion, au pire grossièrement faux et à l'origine de raisonnements erronés. On méditera à cet effet sur la validité des deux affirmations suivantes :

$$\text{Pour presque tout } \omega \in \Omega, \text{ le contraire de "}N(\omega) < \infty\text{" est "}N(\omega) = \infty\text{"}$$

et

$$\text{Le contraire de "}N < \infty, \text{ p.s." est "}N = \infty, \text{ p.s."}$$

Partie 1

Cette partie introductive a été globalement correctement traitée par l'ensemble des candidats. Quelques commentaires cependant :

— Il est regrettable de constater que beaucoup de candidats semblent croire que la définition de martingale consiste seulement à vérifier que $\mathbb{E}(V_{n+1}|\mathcal{F}_n) = V_n$, p.s. pour tout n, en oubliant de mentionner au préalable que $(V_n)_n$ est un processus intégrable et adapté. En particulier, le correcteur avoue sa perplexité devant des raisonnements (souvent vus dans les copies) du genre :

1. $\mathbb{E}(V_{n+1}|\mathcal{F}_n) =$ (calcul d'une page) $= V_n$, p.s. pour tout n.
2. autres considérations annexes ...
3. (V_n) est intégrable.

Si le candidat a passé une page à manipuler une espérance conditionnelle, on espère bien qu'il a **au préalable** vérifié qu'il a le droit de le faire ! Mentionner a posteriori que (V_n) est intégrable ne peut que questionner le correcteur quant à la (mé)connaissance du candidat à propos de la construction et des conditions d'existence de l'espérance conditionnelle d'une variable aléatoire. Une autre possibilité aurait été de mentionner au préalable que V_n est p.s. positive par construction, mais cet argument n'a jamais été vu dans les copies.

— Le calcul de $\mathbb{E}(\mathbf{1}_{U_{n+1} \leqslant V_n}|\mathcal{F}_n)$ a été l'occasion de beaucoup d'approximations dans les justifications (dans le meilleur des cas) et de beaucoup d'erreurs manifestes et d'énoncés incorrects (dans le pire des cas). L'égalité fondamentale ici est la suivante

$$\mathbb{E}(\mathbf{1}_{U_{n+1} \leqslant V_n}|\mathcal{F}_n) = \int \mathbf{1}_{u \leqslant V_n} \mathbb{P}_U(\mathrm{d}u), \ p.s.$$

Ce calcul repose sur deux hypothèses : celle que (V_n) est $\mathcal{F}_n$-mesurable et celle que U_{n+1} est **indépendante** de $\mathcal{F}_n$. L'hypothèse d'indépendance a été très souvent oubliée dans les justifications, menant à des énoncés incorrects.

De multiples variations autour de l'énoncé type

$$\mathbb{E}(\mathbf{1}_{U_{n+1} \leqslant V_n}|\mathcal{F}_n) = \mathbb{P}(U_{n+1} \leqslant V_n) = V_n$$

ont été (malheureusement) rencontrées dans les copies et ont suscité l'incompréhension du correcteur. Au-delà de l'incorrection de l'énoncé et de l'absence de justifications, $\mathbb{P}(U_{n+1} \leqslant V_n)$, qui est un nombre déterministe, serait égal à V_n, une variable aléatoire ?

Partie 2

Cette section a été abordée par la majorité des candidats et plutôt bien traitée dans l'ensemble. Cependant, certaines questions méritent des commentaires :

- Partie 2.1, question 1)e) : cette question a été particulièrement maltraitée. Dans la phrase, “pour tout $C > 0$, la martingale arrêtée $X_{n \wedge N_C}$ converge presque-sûrement”, le “presque-sûrement” dépend de C. Il y a donc là un problème classique d'échange de quantificateur et de p.s. qui ne peut se faire que si on se restreint à un ensemble dénombrable de C. Il est un peu regrettable que seules quelques rares copies aient vu ce problème et qu'un nombre encore plus restreint d'entre elles ait traité correctement cette question.
- Partie 2.2, question 5 : dans beaucoup de copies, il y a confusion entre sup et lim sup. En particulier, on a vu souvent des affirmations étranges du type $a_p \leqslant \limsup_n a_n$ pour tout p. On méditera à cet effet l'exemple de la suite $a_n = (-1)^n$ si $n \neq 2018$ et $a_{2018} = 2$. Très peu de candidats ont traité correctement cette question.

Seules les meilleures copies ont trouvé l'argument d'itération $x_{k+1} = \rho(x_k)$ et de point fixe de ρ.

Partie 3

Le début de cette partie a été abordé par la majorité des candidats. La fin de la partie 3.1 et la partie 3.2 n'ont été abordées que par les meilleures copies. Quelques commentaires :

- Question 1 : de façon surprenante, cette question a été incomprise de la grande majorité des candidats. En particulier, il faut affirmer une bonne fois pour toute que les variables X_i **ne sont pas** indépendantes ! Elles sont indépendantes **conditionnellement à** Θ, ce qui n'est pas la même chose, sauf cas particulier sur Θ. Cette question méritait un soin particulier dans sa rédaction et la plupart des candidats sont passés à côté de cette question.
- Question 3.a) : avant d'utiliser l'inégalité de Chebychev, on aurait apprécié de voir les hypothèses de cette inégalité...
- Question 8) c) : peu de candidats (parmi les rares l'ayant abordée) ont vu la difficulté de cette question : l'argument de passage par densité des polynômes est la plupart du temps correctement énoncé, mais la suite laisse souvent à désirer.

Partie 4

Cette dernière partie, relativement élémentaire, a été traitée (la plupart du temps de façon honorable) par de très rares copies.