Banque BCPST Inter-ENS/ENPC - Session 2018

Rapport sur l'épreuve écrite de mathématiques
(épreuve comptant uniquement pour l'admission)

Écoles concernées : ENS (Paris) - ENS de Lyon - ENS de Paris-Saclay - ENPC

Coefficients (en pourcentage du total d'admission) :
ENS de Paris-Saclay : 6.2%
ENS de Lyon : 6.6% pour les deux options
ENS (Paris) : 11.3% pour les deux options
ENPC : 20%

—— o ——

Membre du jury : Gaël Raoul

Le sujet est composé de deux exercices. Le second exercice est composé de deux parties indépendantes. Le premier exercice est basé sur un argument mathématique astucieux introduit dans l'article *Yurtsev et al, Molecular systems biology 2013.* Le second exercice est basé sur le problème du collectionneur de vignettes.

Ce sujet a permis de tester la compétence technique des candidats, faisant intervenir des aspects variés du programme de mathématiques de BCPST : équations différentielles, calcul d'intégrales, probabilités et séries numériques. Les candidats admissibles ont obtenu des notes comprises entre 2.5 et 20, avec un écart type de 3.6.

Le **premier exercice** est une analyse du système d'équations différentielles suivant:

$$\begin{cases} R(0) = r \text{ et } R'(t) = R(t) \text{ pour } t \geq 0, \\ S(0) = 1 - r \text{ et } S'(t) = -S(t)1_{A(t)\geq 1} + 2S(t)1_{A(t)\leq 1} \text{ pour } t \geq 0, \\ A(0) = 2 \text{ et } A'(t) = \frac{-A(t)R(t)}{1+A(t)} \text{ pour } t \geq 0. \end{cases}$$

L'objectif est d'identifier un paramètre r tel que $\frac{R(0)}{R(0)+S(0)} = \frac{R(T)}{R(T)+S(T)}$ à un temps T. Ce temps $T > 0$ est défini implicitement par $R(T) + S(T) = 10$. Les deux paramètres r et T sont donc à déterminer. Biologiquement, $A(\cdot)$ désigne une concentration d'ampicilline, $S(\cdot)$ et $R(\cdot)$ des nombres de bactéries (respectivement sensibles et résistantes à l'ampicilline).

Les trois premières questions ont été très bien traitées par les étudiants. Pour la question 1.4, l'erreur la plus commune a été d'oublier d'invoquer la continuité de la fonction $\tau(\cdot)$ pour décrire $\tau(]0, +\infty[)$. La question suivante demandait des formules explicites pour $S(t)$. Il convenait de traiter le cas où $t \in]0, \tau(r)[$ dans un premier temps, puis de s'intéresser au cas où $t \in]\tau(r), +\infty[$. Dans la moitié des copies la formule obtenue pour le second intervalle n'a pas été initialisée correctement : la *condition initiale* correspondait alors au temps $t = \tau(r)$. La question 1.6 n'a pas posé de problème, mais des erreurs de calcul ont empêché la plupart des étudiants de terminer correctement l'exercice (questions 1.7 à 1.10). Les étudiants aboutissant à la question 1.9 sans erreur ont été récompensés (cela a été le cas d'environ un quart des étudiants).

L'**exercice 2** porte sur le problème de collectionneur de vignettes, qui intervient dans de nombreux domaines, en chimie organique ou en biologie moléculaire par exemple. Ce modèle probabiliste est étudié au moyen d'un calcul de moyenne et de variance dans une première partie, tandis qu'une estimation plus fine des événements rares est proposée dans la seconde partie.

La **première partie de l'exercice 2** commence par un exemple : sur un tirage explicite de X_i pour $i \in \{1, \ldots, 6\}$, il faut reconnaître F_1, F_2, F_3 ainsi que X_{F_1}, X_{F_2}, X_{F_3}. L'objectif de cette question est de familiariser les candidats aux notations de l'exercice. Environ 80% des étudiants ont répondu correctement à cette question. La seconde question porte également sur la définition de F_i donnée dans l'énoncé. Le résultat (croissance de la famille (F_i)) semble trivial, mais rédiger une preuve rigoureuse de cette propriété demande précision et rigueur.

Dans les questions 2.1.3 et 2.1.4, il faut reconnaître des lois de probabilité classiques (les Z_j sont des variables aléatoires de Bernouilli tandis que les F_j sont des variables aléatoires géométriques), et montrer l'indépendance de variables aléatoires. Environ la moitié des candidats ont répondu correctement à ces questions.

La question 2.1.5 n'a posé de soucis qu'à une minorité de candidats. Il faut toutefois faire attention lors de l'estimation de $\left|\sum_{i=1}^{N} \frac{1}{i} - \ln N\right|$ à ne pas faire intervenir l'intégrale $\int_0^1 \frac{1}{t}\, dt$, qui n'est pas finie.

Dans la question 2.1.6, on demande de calculer la moyenne/variance d'une somme de variables aléatoires géométriques indépendantes. De façon surprenante, peu d'étudiants ont abordé cette question. Ceux qui y ont répondu ont également été capables de répondre à la question suivante (en utilisant l'inégalité de Bienaymé–Tchebychev).

La **seconde partie de l'exercice 2** repose sur une utilisation de la formule du crible. Donnée dans l'énoncé, cette formule permet de calculer la probabilité d'une union d'événements non disjoints. La question 2.2.1 nécessite une bonne compréhension de la définition des événements E_i : un tirage (X_l) appartient à $E_{k,i_1} \cap \cdots \cap E_{k,i_l}$ si et seulement si $X_l \notin \{i_1, \ldots, i_l\}$ pour $l \leq k$. Le calcul est alors aisé, les tirages de X_l et $X_{l'}$ étant indépendants si $l \neq l'$. Pour la seconde question il faut utiliser la formule du crible et le résultat de la première question, puis dénombrer le nombre de $l-$uplets $i_1, \ldots, i_l$ satisfaisant $1 \leq i_1 < \cdots < i_l \leq N$. Ces deux questions n'ont été correctement traitées que par 40% des étudiants environ.

La question 2.2.3, indépendante du reste de l'exercice, consiste à établir l'estimation $1 - x \leq e^{-x}$. Presque tous les candidats ont pu établir cette inégalité. Une minorité d'étudiants ont réussi à prouver les deux estimations demandées à la question 2.2.4. Pour le calcul de limite apparaissant dans la question 2.2.5, il faut alors reconnaître

$$\begin{aligned}
\begin{pmatrix} N \\ l \end{pmatrix} \left(1 - \frac{l}{N}\right)^{k_N} &= \frac{1}{l!} e^{\sum_{i=N-l+1}^{N} \log i} e^{k_N \log\left(1 - \frac{l}{N}\right)} \\
&= \frac{1}{l!} e^{-tl} e^{\left(\sum_{i=N-l+1}^{N} \log i - l \log N\right) + \left(l \log N + k_N \log\left(1 - \frac{l}{N}\right) + tl\right)},
\end{aligned}$$

puis tirer parti des deux estimations précédant ce calcul de limite.

La question 2.2.6 doit être traitée en deux temps : il faut montrer que cette série converge absolument, puis de calculer la somme de cette série. Les étudiants, probablement pris par le temps, ont souvent oublié l'une ou l'autre de ces étapes.

Pour obtenir la première estimation énoncée dans la question 2.2.7, il faut utiliser la question 2.2.4 et la convergence absolue de la série (obtenue dans la question 2.2.6). Une petite minorité de candidats a obtenu ce résultat et aucun n'a pu obtenir la seconde estimation, qui était effectivement (et intentionnellement) difficile à obtenir. Pour obtenir cette dernière, il convient tout d'abord utiliser l'écriture de $\mathbb{P}(F_N > k_N)$ obtenue à la question 2.2.2 :

$$\begin{aligned}\mathbb{P}(F_N > k_N) - \sum_{\ell=1}^{\infty}(-1)^{\ell-1}\frac{e^{-t\ell}}{\ell!} &= \sum_{\ell=1}^{N}\left((-1)^{\ell+1}\left(\begin{array}{c}N\\ \ell\end{array}\right)\left(1-\frac{\ell}{N}\right)^{k_N} - (-1)^{\ell-1}\frac{e^{-t\ell}}{\ell!}\right)\\ &\quad + \sum_{\ell=N+1}^{\infty}(-1)^{\ell-1}\frac{e^{-t\ell}}{\ell!}.\end{aligned}$$

La première estimation obtenue dans cette question permet de définir $\kappa > 0$ tel que

$$\sum_{\ell=\kappa}^{N}\left|(-1)^{\ell+1}\left(\begin{array}{c}N\\ \ell\end{array}\right)\left(1-\frac{\ell}{N}\right)^{k_N}\right| + \sum_{\ell=\kappa}^{\infty}\left|(-1)^{\ell-1}\frac{e^{-t\ell}}{\ell!}\right| \leq \varepsilon,$$

lorsque N est assez grand. Pour conclure, il faut utiliser la limite établie dans la question 2.2.5 qui montre que

$$\sum_{\ell=1}^{\kappa}\left((-1)^{\ell+1}\left(\begin{array}{c}N\\ \ell\end{array}\right)\left(1-\frac{\ell}{N}\right)^{k_N} - (-1)^{\ell-1}\frac{e^{-t\ell}}{\ell!}\right)$$

est petit lorsque N est grand.