Banque BCPST Inter-ENS / ENPC - Session 2021

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 de Paris : 11.3 % pour les deux options
- ENPC : 20 %

Membre du jury : Hélène Leman

— ◦ —

L'épreuve est composée de quatre parties. Elle propose de s'intéresser à des approximations de quantités qui peuvent toutes s'apparenter à des intégrales. La première partie est basée sur une méthode d'analyse numérique pour approcher l'intégrale d'une fonction régulière: méthode de Simpson. Les deux parties suivantes traitent de l'approximation de l'espérance d'une variable aléatoire, ce qui correspond à une intégrale dans le cas d'une variable à densité. Cette approximation passe par la moyenne empirique de variables aléatoires indépendantes et identiquement distribuées : méthode de Monte-Carlo. Une quatrième partie propose de mettre en relation les résultats des parties 1 et 3 et ceux des parties 2 et 3.

Le sujet a permis de tester les candidats sur leur aisance à manipuler les techniques et outils classiques d'analyse et de probabilités au programme des classes préparatoires BCPST. Les notes obtenues par les candidats admissibles sont comprises entre 2 et 18.7 sur 20, avec une moyenne de 9.10 et un écart type de 2.90. La précision de la rédaction et l'honnêteté des candidats dans leurs démonstrations ont été récompensées. Inversement, les rares candidats ayant tenté d'imposer leurs résultats par des affirmations non justifiées ont été sanctionnés.

PARTIE I

Dans la **partie I**, on cherche à quantifier l'erreur commise lorsqu'une intégrale est approchée par la méthode de Simpson, c'est-à-dire lorsqu'on utilise l'approximation de la fonction intégrée par un polynôme quadratique prenant les mêmes valeurs que cette fonction en trois points distincts.

La **question 1** s'intéresse à l'étude de polynômes quadratiques sur l'intervalle $[0, 1]$. Les **questions 1.a et 1.b** ont été bien (voire parfaitement) traitées dans la quasi totalité des copies sélectionnées. De nombreuses méthodes ont été utilisées pour traiter la question 1.b. L'idée retrouvée majoritairement a été de développer les polynômes $Q(x)$ et $Q(0)P_0(x) + Q(1)P_1(x) + Q(1/2)P_2(x)$ pour montrer que les coefficients des monômes étaient identiques. D'autres méthodes ont parfois été utilisées et récompensées identiquement: vérifier que le polynôme $Q(x) - [Q(0)P_0(x) + Q(1)P_1(x) + Q(1/2)P_2(x)]$ admet 3 racines distinctes; montrer que les polynômes P_0, P_1 et P_2 forme une base de $\mathbb{R}_2[X]$ et écrire Q dans cette base.

Le but de la **question 2** est simplement de translater l'intervalle d'étude. Il suffisait d'utiliser le changement de variable indiqué dans l'intégrale $\int_c^d P(x)dx$ pour se retrouver sur l'intervalle $[0, 1]$ et d'appliquer la question précédente en remarquant que l'intégrande était encore un polynôme de degré 2. Comme le changement de variable était indiqué, il s'agissait d'une application directe du cours, et on attendait une rigueur de rédaction qui n'a pas été retrouvée dans de nombreuses copies. Notamment, les nouvelles bornes de l'intervalle d'intégration données ont souvent été fausses.

Le but de la **question 3** est de prouver l'inégalité des accroissements finis, ce qui peut se montrer en utilisant l'égalité des accroissements finis: $\forall x, y \in [a, b], \exists \xi \in [x, y]$ ou $[y, x]$ tel que

$$g(x) - g(y) = (x - y)g'(\xi)$$

puis en majorant g' sur l'intervalle $[a, b]$. L'erreur principale dans cette question a été d'oublier cette dernière étape.

La **question 4** cherche à estimer l'erreur dans l'approximation de Simpson lorsqu'on approche la fonction intégrée par un unique polynôme quadratique sur un petit intervalle $[c, d]$. $\phi(d-c)$ correspond en effet à la différence entre l'intégrale et l'approximation.
La **question 4.a** étudie la fonction ϕ sur l'intervalle $[0, d - c]$. Cette question a été bien traitée dans une majorité de copies. Quelques rares copies n'ont pas réussi à dériver la fonction $h \to \int_c^{c+h} f(y)dy$ ou parfois la fonction $h \to f(c)$. Nous supposons ici qu'il s'agit d'erreurs dans l'interprétation des différentes variables. Parfois, des erreurs dans les premières dérivées ont pu être répercutées dans les dérivées d'ordre supérieur, ce qui n'a alors pas été sanctionné.
La **question 4.b** n'a été traitée que dans les meilleures copies. Il fallait ici appliquer le résultat de la question 3. En effet, pour tout $h \in [0, d - c]$,

$$\phi^{(3)}(h) = \frac{1}{2}\left(f''(c+h) - f''(c+h/2)\right) - \frac{h}{6}\left(f^{(3)}(c+h) + f^{(3)}(c+h/2)\right).$$

Après l'utilisation de l'inégalité triangulaire, la valeur absolue de la différence entre les dérivées secondes peut-être majorée à l'aide de la question 3, tandis que les dérivées troisièmes sont directement majorées sur l'intervalle $[a, b]$. On obtient alors le résultat.
Les **question 4.c et 4.d** se traitent en remarquant que

$$\left|\int_0^h \phi^{(3)}(u)du\right| \leq \int_0^h |\phi^{(3)}(u)|du \leq \int_0^h Mudu \tag{1}$$

puis en calculant les membres extrémaux dans la suite d'inégalités précédentes. Ces questions ont généralement été bien traitées. Cependant, de nombreuses copies ont cherché à calculer le terme du milieu de (1) en intégrant directement la valeur absolue, ou n'ont pas calculé les valeurs des primitives en 0. Ces deux situations ont été sanctionnées.

La **question 5** conclut la première partie et donne l'erreur de l'estimation de l'intégrale par la méthode de Simpson. L'idée est de décomposer l'intégrale sur l'intervalle $[a, b]$ en utilisant la relation de Chasles pour la transformer en somme sur les intervalles $[a_i, a_{i+1}]$, avec $a_i = a + i/n(b-a)$. On peut alors appliquer le résultat de la question précédente. Cette question a été abordée dans un nombre relativement restreint de copies. Certaines copies ont voulu l'aborder à l'aide d'une récurrence ce qui n'était malheureusement pas la bonne approche.

PARTIE II

Dans la **partie II**, on cherche à étudier l'erreur commise lorsqu'on approche l'espérance d'une variable aléatoire par la moyenne empirique d'un échantillon de variables aléatoires indépendantes et identiquement distribuées. Plus précisément, on cherche à évaluer la probabilité:

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)\right| \geq \varepsilon\right).$$

Cette partie traite le cas de variables aléatoires symétriques, admettant au moins un moment d'ordre 2.

Dans la **question a** (question préliminaire), on commence par prouver que cette probabilité tend vers 0 lorsque la taille de l'échantillon n tend vers l'infini. Pour y répondre, on peut faire appel à l'inégalité de Bienaymé-Tchebychev ou à la loi des grands nombres.
La **question b** traite un exemple particulier de variables discrètes. La première partie a généralement été bien traitée et rédigée, sauf quelques copies qui n'ont pas fait appel à l'indépendance des variables aléatoires pour déduire la loi binomiale. La deuxième partie de la question nécessite un calcul exact qui n'a été réalisé que dans un nombre restreint de copies. On peut remarquer que

$$\begin{aligned}\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n}X_i\right| \geq 1\right) &= \mathbb{P}\left(\sum_{i=1}^{n}X_i \geq n\right) + \mathbb{P}\left(\sum_{i=1}^{n}X_i \leq -n\right)\\ &= \mathbb{P}\left(\sum_{i=1}^{n}\frac{X_i+1}{2} \geq n\right) + \mathbb{P}\left(\sum_{i=1}^{n}\frac{X_i+1}{2} \leq 0\right)\\ &= \binom{n}{n}2^n2^0 + \binom{n}{0}2^02^n = 2^{n+1}\end{aligned}$$

puis conclure en prenant le logarithme puis en divisant par n de chaque côté.

Les **questions 1.a et 1.b**, qui faisait surtout appel au cours concernant l'indépendance de variables aléatoires, ont été bien traitées dans une large majorité des copies.

La **question 2** a également été bien traitée dans un grand nombre de copies en appliquant le résultat de la question précédente.

La difficulté de la **question 3** réside dans la preuve que $u_n > 0$ (et pas uniquement $u_n \geq 0$), ce qui peut se prouver par une récurrence directe en utilisant la question 2. Cette question a été traitée dans un nombre très restreint de copies. La deuxième partie est directement obtenue par la composition par la fonction ln.

Les **questions 4 et 5** ont été abordées dans très peu de copies et se sont avérées très délicates. Pour la question 4, on peut utiliser plusieurs fois le résultat de la question précédente, une fois l'entier k décomposé à partir de la division euclidienne $k = pq + r$. La question 5 nécessite de remarquer que la suite S_n est décroissante (par définition) puis de prouver la majoration sur la limite en passant par le résultat de la question 4, notamment en vérifiant que $S_n \leq \alpha_q/q + \beta_q/n$. La difficulté réside dans la manipulation des différents indices.

La **question 6**, qui est la plus difficile de cette partie, n'a été traitée que dans de rares copies qui n'ont pas conclu entièrement. Il faut ici remarquer qu'on peut montrer le résultat de la question 5 quelque soit l'indice q et donc que la limite de la suite S_n est inférieure à $I := \inf\{\alpha_q/q, q \geq 1\}$, puis on conclut avec le théorème des gendarmes puisque $I \leq \alpha_n/n \leq S_n$, pour tout $n \geq 1$.

La **question 7** correspond à la conclusion de cette partie. Il faut appliquer les résultats précédents en remarquant cependant que la probabilité fait intervenir une valeur absolue contrairement aux questions précédentes. On peut utiliser la symétrie des variables aléatoires (hypothèse) pour conclure. Enfin, on pouvait remarquer que la valeur de la limite n'est pas connue et que la technique développée en partie II ne permet pas de vérifier que cette limite est non nulle.

PARTIE III

La **partie III** cherche à quantifier la même quantité que la partie précédente mais sous des hypothèses différentes sur les variables aléatoires, notamment, on suppose qu'elles sont à valeur dans un intervalle fini $[x_1, x_2]$.

Le résultat de la **question 1** peut se prouver en supposant par exemple $y < z$ (même calcul dans l'autre sens) et en étudiant les variations de la fonction $y \mapsto e^{py+(1-p)z} - pe^y - (1-p)e^z$ à l'aide de sa dérivée. Cette question a été traitée dans peu de copies, mais qui l'ont alors bien rédigée de bout en bout.

Pour répondre à la **question 2**, il est possible d'appliquer le résultat précédent avec $p = \frac{x_2 - X}{x_2 - x_1}$, $y = tx_1$ et $z = tx_2$ en considérant l'aléa fixé, puis de prendre l'espérance de chaque côté de l'inégalité. Seules quelques copies ont cherché à prendre l'espérance avant d'appliquer la formule, ce qui ne permettait pas de conclure.

Le but de la **question 3** est clairement indiqué dans l'énoncé. La **question a** a été traitée dans une grande majorité de copies. La **question b** a généré de nombreux problèmes pour vérifier que $g''(t)$ était négative, la seconde partie de la question qui nécessite donc de s'intéresser aux variations de la dérivée g' puis à son signe, a généralement été bien rédigée. Enfin la **question c** n'a généralement pas été comprise: il fallait remarquer que la question b prouvait le résultat pour les t positifs et qu'il fallait argumenter pour le prouver pour tout $t \in \mathbb{R}$. Pour ce faire, il est possible de s'intéresser à $-X$, variable aléatoire qui vérifie les mêmes hypothèses que X, avec une espérance $-\mu$ et qui est à valeur sur $[-x_2, -x_1]$.

La **question 4.a.** est clairement la plus difficile de ce sujet et n'a été abordée qu'un nombre très restreint de fois, très peu de copies ont donné une direction de réponse satisfaisante (cependant ce cas de figure a été récompensée), aucune n'a aboutit. Il faut dans un premier temps prouver que, pour tout $t \geq 0$,

$$\begin{aligned}
\mathbb{P}\left(\left|\frac{1}{n}\sum_{k=1}^{n}(X_k - \mu)\right| \geq \varepsilon\right) &= \mathbb{P}\left(e^{t\sum_{k=1}^{n}(X_k-\mu)} \geq e^{tn\varepsilon}\right) + \mathbb{P}\left(e^{t\sum_{k=1}^{n}(-X_k+\mu)} \geq e^{tn\varepsilon}\right) \\
&\leq e^{-tn\varepsilon}\left(\mathbb{E}[e^{t(X-\mu)}]^n + \mathbb{E}[e^{t(-X+\mu)}]^n\right) \\
&\leq 2e^{-tn\varepsilon}e^{nt^2(x_2-x_1)^2/8},
\end{aligned}$$

puis prendre $t = 4\varepsilon/(x_2 - x_1)^2$ pour conclure.

Enfin, la **question 4.b.** a généralement été bien traitée lorsqu'elle a été abordée.

PARTIE IV

Pour comparer les parties II et III, on pouvait s'intéresser aux différences des hypothèses sur les variables aléatoires, mais aussi aux différences de résultats obtenus: bien que similaires, le résultat de la partie II, qui ne vérifie pas que la limite obtenue est positive, peut s'avérer bien plus faible que le résultat obtenu dans la partie III.

Pour comparer les résultats des parties I et III, il fallait ainsi avoir noté que l'espérance d'une variable aléatoire à densité correspond à une intégrale. Ainsi, ces deux parties donnent un moyen de calculer de manière approchée une intégrale sur un intervalle fini. La première méthode (partie I) est une méthode déterministe tandis que la deuxième méthode (partie III) est une méthode stochastique.

Cette dernière partie plus ouverte n'a été que peu de fois abordée. Cependant quelques remarques intéressantes ont parfois été données.