Composition de Mathématiques – B, Filière MP (X)

Cette année, le concours d'admission de l'École Polytechnique s'est déroulé pendant l'épidémie de Covid-19 et était soumis aux nombreuses mesures restrictives qui l'accompagnaient. De ce fait, le concours était plus difficile qu'à la normale, tant pour la préparation des candidats que pour l'organisation d'épreuves.

1. Présentation du sujet

Le sujet se compose de trois parties indépendantes. La Partie I du sujet contient une démonstration du principe de grandes déviations pour une variable aléatoire X discrète symétrique (à nombre fini de valeurs),

$$P[X = 1] = P[X = -1] = \frac{1}{2}.$$

Dans ce cas simple, la démonstration du principe est assez élémentaire. La Question 3 est une inégalité de grande déviation, et la Question 8b est une version asymptotique de ce principe. Notons que l'expression qui apparaît à la Question 3,

$$\Psi(t) = \inf_{\lambda \geq 0} (\psi(\lambda) - \lambda t), \quad \psi(\lambda) = \log E(e^{\lambda X}),$$

est la transformée de Legendre de la variable aléatoire X; cette quantité joue un rôle important dans la théorie de grandes déviations d'une manière générale.

D'une façon assez prévisible, cette partie du sujet requiert la maîtrise solide de bases de probabilités selon le programme MPSI. Les premières questions (Questions 1-4) de la partie I étaient traitées par un grand nombre de candidats. La maîtrise des Questions 5a et 5b était souhaitable pour la bonne compréhension des questions suivants. Ainsi, ces questions qui ont mis en difficulté une partie considérable de candidats et ont marqué un coup d'arrêt dans leur travail sur la Partie I.

La Partie II du sujet traite principalement du comportement asymptotique de l'intégrale de Laplace

$$I_t = \int_a^b e^{tf(x)}\, dx, \quad t \to +\infty,$$

où la fonction réelle f satisfait un nombre d'hypothèses convenables. Les résultats sur l'intégrale I_t (Questions 9-11) sont ensuite utilisés pour démontrer la formule de Stirling (Questions 12a, 12b).

Les questions 9, 10, 11 et 12b ont mis en difficulté certains candidats.

La Partie III du sujet s'apparente à la précédente et étudie l'intégrale oscillatoire

$$J_t = \int_0^1 g(x) \sin(tf(x))\, dx, \quad t \to +\infty,$$

où les fonctions réelles f et g satisfont des conditions spécifiques. Il convient d'observer que les problèmes des Parties II et III sont les faces différentes (réelle et imaginaire) d'un problème sur l'asymptotique d'une intégrale de Fourier *complexe* de la forme

$$H_t = \int_a^b g(x) e^{tf(x)}\, dx, \quad t \to +\infty,$$

où $f = f_1 + if_2$ et g, f_1, f_2 sont des fonctions réelles satisfaisant certaines conditions.

Malgré certaines difficultés, les Questions 13-16, 18a, et 18b de la Partie III ont été abordées par la majorité de candidats. Les Questions 17 et 19 sont plus complexes et elles contiennent l'essentiel de cette partie du sujet. Cependant, elles n'étaient traitées que par un très petit nombre d'élèves participant au concours.

Les Parties II, III font naturellement l'appel à des notions clés d'analyse classique du programme MPSI (notion d'équivalence de quantités, passage à la limite sous le signe d'intégrale, changement d'ordre de limites, développement asymptotique d'une quantité (une intégrale), changement de variables sous le signe d'intégrale, intégration par parties, etc.)

2. Conseils généraux pour les candidats.es

Traditionnellement, le sujet de l'épreuve de Mathématiques - B porte sur des points importants du programme de MPSI. La première partie du sujet met à l'honneur les probabilités et les concepts correspondants (notion de variables aléatoires indépendantes, formule de transfert, les techniques standard de calcul de probabilités via espérances, etc.) Les Parties II et III requièrent la bonne maîtrise du calcul intégral classique (changement de variables convenables, intégrales propres, impropres, majoration de restes d'intégrales, équivalences de quantités intégrales, les asymptotiques d'intégrales, etc.)

Pour traiter le sujet d'une manière satisfaisante, il était nécessaire de manipuler avec soin et précision ces nombreuses notions. Le sujet a révélé des lacunes importantes chez certains candidats dont on cite quelques exemples ci-dessous:

I. Les calculs des Questions 3-5b (et même de la Question 1) se font aisément à l'aide de la formule de transfert, le résultat de base d'une très grande efficacité pour les variables aléatoires discrètes. Malheureusement, plusieurs candidats ne l'évoquaient pas ou n'arrivaient pas à l'appliquer, ce qui compromettait grandement leurs chances de répondre à ces questions.

II. L'une des méthodes de base en probabilité est de démontrer une inégalité entre des variables aléatoires "point par point" sur l'espace probabilisé et ensuite de l'intégrer pour obtenir une inégalité correspondante entre les espérances de ces variables. Bien des candidats n'ont pas su utiliser cette astuce qui permettait de répondre aux Questions 6 et 7.

III. Avant d'écrire le signe d'une limite, il est de bonne aloi de s'assurer que la limite existe (cf. la Question 8b). En parlant de cette question, mentionnons que l'on ne pouvait pas passer "directement" à la limite $\varepsilon \to 0+$ dans l'inégalité de la Question 8a

$$\frac{1}{n} \log P[S_n \geq m(\lambda) - \varepsilon] \geq \psi(\lambda) - \lambda m(\lambda) - \lambda\varepsilon + u_n(\varepsilon),$$

car $u_n(\varepsilon) = \log(1 - 4/(n\varepsilon^2))/n$ et la fonction $u_n(\varepsilon)$ n'est pas définie pour $\varepsilon > 0$ suffisamment petit. La Question 8b est assez délicate à traiter et elle est discutée en détail en p. 5 de ce rapport.

IV. Concernant la Question 9, notons qu'il est facile de construire une fonction $f : [a, b] \to \mathbb{R},\ x_0 \in]a, b[$ avec les propriétés suivantes:

- la fonction f est de classe C^1 sur $[a, b]$,
- le point x_0 est un maximum local (ou global) de la fonction,
- la dérivée $f'(x)$ change de signe un nombre infini de fois sur tout intervalle de forme $]x_0 - \delta, x_0[$ ou $]x_0, x_0 + \delta[$, $\forall \delta > 0$ suffisamment petit. En particulier, la fonction $f(x)$ n'est ni strictement croissante, ni strictement décroissante sur ces intervalles.

Bien évidemment, ce phénomène est impossible pour une fonction de classe C^2 (ou C^∞) sur $[a,b]$. En répondant à cette question, bien des candidats ont donné des raisonnements faux utilisant seulement la condition (H) et sans faire appel à la condition C^∞, cruciale pour ce point.

V. On dit que deux fonctions strictement positives $A(x)$ et $B(x)$ sont équivalentes au voisinage du point x_0 (notation: $A(x) \sim_{x_0} B(x)$) ssi

$$\lim_{x \to x_0} \frac{A(x)}{B(x)} = 1.$$

Cette relation n'était pas vérifiée correctement par de nombreux candidats à la Question 10.

VI. C'est une énormité de dire que la fonction $F(x) = \sin(x^2)$ est $\sqrt{2\pi}$-périodique. Idem de dire que

$$|\sin(x^2)| \geq \frac{2}{\pi} x^2$$

pour tout x réel, car cette inégalité est valable seulement sur un voisinage convenable du point $x_0 = 0$, cf. la Question 13.

VII. Soit f une fonction continue et positive sur $[0, +\infty[$. Il doit être *absolument clair* pour tous les candidats qu'aucune implication n'est valide entre les deux propriétés suivantes

1. $\lim_{x \to +\infty} f(x) = 0$;
2. $\int_0^\infty f(x)\,dx$ converge.

Voir la Question 13 encore une fois.

VIII. Il est impossible de déduire le résultat de la Question 15 à partir de celui de la Question 14. En effet, le rayon de convergence de la série entière de la Question 14 est $R = +\infty$. Dans la Question 15, on considère $a \to +\infty$; le paramètre a "sort" donc sur le bord de l'intervalle de convergence de la série, et on ne peux plus utiliser le changement d'ordre d'intégration et de la sommation dues à la convergence normale (uniforme) à l'intérieur de l'intervalle de convergence.

Toutes ces erreurs montrent de la part de trop nombreux candidats des bases en Analyse réelle bien fragiles. Ces notions sont pourtant abordées dès la première année de CPGE en MPSI, et plus d'un an après, elles devraient être bien assimilées par les candidats.

Comme chaque année, il est préférable de s'attacher à traiter correctement plusieurs questions consécutives et parmi elles des questions plus difficiles, plutôt que d'essayer de survoler toutes les parties et de tenter de "grappiller" des points sur les questions les plus faciles. Le barème est établi de sorte qu'une telle stratégie est forcément vouée à l'échec. A l'opposé, passer du temps pour, par exemple, réussir à traiter correctement les questions comme la 5a, 5b donnait la clé pour réussir les questions suivantes de la Partie I et s'assurer un nombre de points suffisant pour atteindre une note supérieure à 12/20, note qui sépare les 470 premiers candidats des autres. Il est par contre tout à fait autorisé de sauter une question que l'on ne parviendrait pas à résoudre, puis d'en utiliser le résultat plus tard. Il faut alors veiller à ne pas oublier de vérifier soigneusement toutes les hypothèses requises pour appliquer ces "boîtes noires".

Le soin apporté aux copies pose toujours problème. Il reste encore trop de candidats qui ne mettent pas en avant, dans la rédaction de leurs réponses, les arguments clés de la démonstration et qui présentent dans leur copie des calculs ou des raisonnements qui n'aboutissent pas. Nous devons donc encore une fois rappeler aux candidats que l'usage d'un brouillon est indispensable afin de ne présenter sur sa copie que les étapes essentielles d'un raisonnement ou d'un calcul et de ne pas y faire figurer des arguments faux ou trop incomplets.

Il est aussi demandé aux candidats une lecture plus attentive des énoncés et des questions posées. Certains candidats ne répondent qu'à la "moitié" de la question posée sans dire un seul mot a propos du reste. Préciser que l'on admet le résultat correspondant au reste de la question facilite la lecture des correcteurs. De même, les phrases du type "... et à partir de là il est facile de voir que..." alors que justement, le point délicat reste à traiter, n'apportent pas de points. Il est inutile de compter sur un manque de vigilance de la part du correcteur.

Enfin, la lisibilité des copies peut parfois poser un réel problème aux correcteurs. Nous rencontrons encore trop de copies remplies de râtures et/ou parfaitement illisibles du fait d'une graphie microscopique ou indéchiffrable. Dans les cas où malgré tous nos efforts de déchiffrage, certaines parties du texte restent incompréhensibles pour le correcteur, et dans le doute, les points ne sont pas attribués. Dans l'autre sens, il est évident qu'une copie bien présentée met le correcteur dans de bien meilleures dispositions au moment d'attribuer des points à une question.

3. Statistiques générales

Les résultats de l'épreuve sont en accord avec les directives statistiques générales du concours, ce qui garantit l'influence respective convenable des différentes épreuves.

Nombre de copies :	1580
Note moyenne :	9.47
Écart-type :	3.81

Les notes des candidats (français et étrangers) se répartissent selon le tableau suivant :

$0 \leq \mathrm{N} < 4$	106	6.71 %
$4 \leq \mathrm{N} < 8$	516	32.66 %
$8 \leq \mathrm{N} < 12$	487	30.82 %
$12 \leq \mathrm{N} < 16$	415	26.27 %
$16 \leq \mathrm{N} \leq 20$	56	3.54 %
Total :	1580	100%

Nombre de candidats :	1580
Note moyenne :	9.47
Écart-type :	3.81

4. Examen détaillé des questions

Ce qui suit n'est pas un corrigé de l'épreuve mais une liste de commentaires et des remarques de correcteurs question par question.

Partie I

Question 1 Cette question facile a été traitée correctement par 90% de candidats. Elle se fait ou bien à l'aide de formule de transfert, ou bien à l'aide de l'inégalité de Markov. Dans ce dernier cas il est indispensable de rappeler la positivité de la variable aléatoire discirète (= *vad* par la suite) $e^{\lambda X},\ \lambda \in \mathbb{R}$.

Question 2 Cette question a été correctement faite par 60% de candidats. La façon simple et rapide était de déduire le résultat de la symétrie de la vad

$$S_n = \frac{1}{n}\sum_{k=1}^{n} X_k.$$

Une autre manière (plus technique) était d'y reconnaître la loi binomiale. Il y a eu un nombre de preuves fausses prétendant par exemple que

$$\text{“}P[S_n \geq 0] \geq nP[X_1 \geq 0] = \frac{n}{2}\text{ “},$$

etc., dont nous ne parlerons pas ici.

Question 3 La question a été bien traitée par 80% de candidats. La solution consistait à appliquer le résultat de la Question 1 à la vad $e^{\lambda n S_n}$ et de passer à l'infimum par rapport à $\lambda > 0$. L'appel à la continuité de la fonction $(\psi(\lambda) - \lambda t)$ au point $\lambda_0 = 0$ a été souvent oublié.

Question 4 La question a été bien faite par 80% de candidats. Une fois de plus, le calcul de la fonction $m(\lambda)$ se fait aisément par la formule de transfert. Pour montrer que la fonction m effectue une bijection de $\mathbb{R}$ sur l'intervalle $]-1,1[$, certains candidats évoquaient le TVI sans parler de la continuité de la fonction m, ce qui est inacceptable.

Question 5a Les Questions 5a et 5b sont plus complexes que les Questions 1-4 et elles marquent un coup d'arrêt du travail sur la Partie I pour un nombre de candidats. La Question 5a a été réussie par 40% d'élèves participant au concours.

La solution de la question est assez rapide et élégante si on applique d'emblée le lemme des coalitions et on utilise la re-écriture judicieuse de l'expression $E[(X_1 - m(\lambda))(X_2 - m(\lambda))D_n(\lambda)]$. Plusieurs candidats rataient cette voie et donnaient une solution beaucoup plus technique et calculatoire (y compris par récurrence).

Question 5b La question est réussie par 23% des candidats. La clé de la solution consistait à écrire la quantité $E[(S_n - m(\lambda))^2 D_n(\lambda)]$ comme une somme de termes de la Question 5a et de l'utiliser pour faire la majoration souhaitée. Certains candidats ont su affiner la solution demandée en obtenant l'inégalité plus forte

$$E[(S_n - m(\lambda))^2 D_n(\lambda)] \leq \frac{1}{n}.$$

L'idée de ce raffinement consistait à remplacer l'inégalité de Cauchy-Schwarz par un calcul explicite d'espérance.

Question 6 Étrangement, cette question était relativement peu abordée. Elle a été bien faite par les candidats qui l'ont traitée (40% de réussite). La bonne façon de faire consistait à considérer la fonction indicatrice de l'événement $\{|S_n - m(\lambda)| \leq \varepsilon\}$, la majorer (ponctuellement) par la vad qui apparaît du coté droit de l'inégalité, et d'intégrer cette inégalité.

Question 7 Les remarques faites à la question précédente s'appliquent à cette question également. Le taux de réussite à cette question est de 10% seulement attestant de la plus grande technicité de cette question comparativement aux précédentes. L'appel à la Question 5b est essentiel à sa solution.

Question 8a La question est bien réussie par les candidats qui ont tenté de la résoudre; le taux de réussite est de 24%. Malgré une certaine technicité et l'appel aux Questions 6 et 7, cette question ne contient aucune piège.

Question 8b C'est probablement la question la plus difficile de tout le sujet; le taux de réussite sur cette question est de 2-3%. Étonnamment, de nombreux candidats ont identifié cette question comme pouvant être “grappillée”. La première erreur commune sur cette question était de croire que sa solution découle simplement (et mécaniquement) des Questions 4 et 7. La deuxième erreur consistait à passer à la limite dans la minoration de la quantité

$$\frac{1}{n} \log P[S_n \geq t],$$

alors que l'existence de cette limite n'était pas démontrée, cf. le point **III** à la page 2. Superflu de dire que l'existence de la limite en question constitue le cœur de la question et ne doit aucunement être négligé. Enfin, la solution ne se fait pas par le théorème d'encadrement (ou de gendarmes) non plus, car $\lim_{\varepsilon \to 0^+} u_n(\varepsilon)$ n'est pas définie pour n fixé.

La bonne façon de faire consistait à vérifier la définition même de la limite de la suite $(\frac{1}{n} \log P[S_n \geq t])_n$ en utilisant les questions précédentes de la Partie I du sujet.

Question 8c Cette question a été négligée par beaucoup de candidats, et pourtant elle représentait des points "faciles à prendre". Elle a été bien traitée par 23% d'élèves participant au concours. La solution (triviale) consistait à calculer les parties de gauche et de droite de la relation en question et de s'assurer de leur égalité.

Partie II

Question 9 Cette question classique d'analyse élémentaire a posée de grandes difficultés à bien des candidats. Bien que la question a été abordée par la quasi-totalité des élèves, seulement 53% ont réussi à la traiter de manière convenable.

Comme mentionné au point **IV** page 2, l'hypothèse (H) ne suffit pas pour démontrer la propriété demandée, et il est nécessaire de faire l'appel au fait que la fonction f est de classe C^∞ (C^2 suffirait) au voisinage du point x_0. A peu près un tiers des candidats part "de propriétés imaginaires" des maxima sans rien démontrer. Une autre partie de candidats fait la démonstration en passant par la formule de Taylor; dans ce cas, il convenait de bien gérer le reste de Taylor-Young de cette formule.

Cette question montre encore une fois que beaucoup de candidats ont une vision très (trop?) graphique et simpliste des concepts en analyse et n'ont pas la culture de contre-exemples fins même dans des situations relativement simples.

Question 10 C'est encore une question très classique relevant de la méthode de Laplace sur les propriétés asymptotiques d'intégrales. Elle a été bien traitée par 23% des candidats. La vérification de la définition de l'équivalence de deux intégrales

$$\int_a^b e^{tf(x)}\, dx \quad \text{et} \quad \int_{x_0-\delta}^{x_0+\delta} e^{tf(x)}\, dx, \quad t \to +\infty$$

a posé un grand problème. En particulier, il fallait minorer la deuxième intégrale par $\int_{x_0-\delta'}^{x_0+\delta'} e^{tf(x)}\, dx$ avec δ', $0 < \delta' < \delta$, bien choisi.

Question 11 Encore une question bien connue de certains candidats. Dans cette question est plus technique que la précédente, il s'agissait d'approcher la fonction $f(x)$ par son polynôme de Taylor d'ordre 2 (avec le reste de Taylor-Young), de la ramener à la gaussienne par un changement de variables approprié et de conclure le calcul. La partie très délicate était la bonne gestion du reste et des approximations introduites au long de ce procédé.

Question 12a C'est une question standard et élémentaire, qui était traitée et réussie par presque tous les candidats. Le taux de réussite là-dessus est de 85%.

Question 12b Cette question est une application directe de la Question 11. Néanmoins, il fallait découper l'intégrale

$$\int_0^{+\infty} e^{-t} t^n\, dt$$

en trois intégrales sur les segments $[0, a[, [a, A[$ et $[A, +\infty[$ avec les paramètres $0 < a < A < +\infty$ convenablement choisis et les traiter séparément. Ce piège a été vu par un très petit nombre de candidats.

Partie III

Question 13 On peut trouver cette question dans bon nombre de manuels d'analyse et d'annales d'examens. Malgré son caractère élémentaire, elle révèle bien souvent des lacunes importantes dans la compréhension des concepts de convergence d'une intégrale impropre et d'une série numérique par les candidats. Ainsi, elle a été abordée par la quasi-totalité de candidats mais réussie par seulement 33% d'entre eux.

Les différents aspects de la réponse à cette question ont été discutés aux point **VI, VII** à la page 3. Rappelons que la fonction $f(x) = \sin(x^2)$ n'est pas périodique; en particulier, pour traiter cette question on ne peut pas utiliser la formule de Taylor pour $f(x)$ car cette dernière donne une approximation de la fonction locale, tandis que l'intégrale

$$\lim_{a\to+\infty} \int_0^a \sin(x^2)\, dx$$

s'étend sur $\mathbb{R}_+$ tout entier.

Question 14 C'est une question simple qui porte principalement sur le changement d'ordre de sommation et d'une intégration d'une série entière. On peut répondre à cette question ou bien en passant par les propriétés de convergence d'une série entière (de rayon de convergence $R = +\infty$ dans ce cas) ou bien par le résultat de commutation de sommation et d'intégration sous une hypothèse de convergence uniforme (ou normale).

La question a été bien traitée par 78% de candidats.

Question 15 Cette question, une fois de plus, fait partie de nombreux manuels d'analyse et d'annales d'épreuves. Néanmoins, il a été réussi seulement par 22% de candidats. L'existence de ces limites se démontre très facilement a l'aide d'intégration par parties (= *IPP*); des nombreux élèves ont tenté à répondre à cette question en se référant à la Question 14. Cela ne marche pas, car il s'agit d'intégrales

$$\int_0^a \sin(x^2)\, dx, \qquad \int_0^a \cos(x^2)\, dx,$$

nous avons $a \to +\infty$ et le rayon de convergence de séries entières correspondantes $R = +\infty$. Généralement parlant, le passage à la limite sur le bord de l'intervalle de convergence n'est pas valable et il demande des justifications supplémentaires (souvent assez sophistiquées).

Question 16 Dans cette question, il s'agit de donner le développement asymptotique de l'intégrale

$$\int_0^a \sin(x^2)\, dx$$

pour $a \to +\infty$. Comme dans la question précédente, on procède par IPP, mais on doit l'effectuer trois fois de suite pour contrôler les termes demandés. Cette question a été réussie par 10% de candidats.

Question 17 La Questions 17 a été très peu abordée. Le taux de réussite de cette question est de 1%. La première étape de la réponse à cette question consiste à écrire

$$\int_{x_0}^1 g(x) \sin(tf(x))\, dx = g(x_0) \int_{x_0}^1 \sin(tf(x))\, dx + \int_{x_0}^1 (g(x) - g(x_0)) \sin(tf(x))\, dx.$$

Après l'avoir bien justifié, on montre par IPP que la deuxième intégrale est un $O(1/t)$ quand $t \to +\infty$. Il découle ensuite de propriétés de la fonction f et d'un changement de variable convenable que

$$\int_{x_0}^{1} \sin(tf(x))\,dx = \int_{x_0}^{x_0+\delta} \sin(tf(x))\,dx + O(1/t), \quad \delta \in]0, 1 - x_0].$$

Question 18a Cette question fait appel aux propriétés de la dérivabilité de la fonction f et au TVI. Une fois de plus, ce dernier a été souvent cité sans mentionner la continuité de la fonction f.

La question a été réussi par 43% de candidats.

Question 18b Il s'agit de vérifier la dérivabilité concernée pour appliquer la formule de Taylor à la fonction f. Pour ainsi dire, elle ne contient aucune piège.

Elle a été bien traité par 22% de candidats.

Question 19 Avec la Question 17, cette question, assez technique contient l'essentiel de la Partie III. Après avoir vérifié que le changement de variable suivant est licite (cf. les questions précédentes), l'astuce consiste à poser

$$f(x) = f(x_0) + u^2,$$

pour ramener le calcul de l'intégrale à

$$\int_{0}^{+\infty} \sin(x^2)\,dx = \int_{0}^{+\infty} \cos(x^2)\,dx = \frac{\sqrt{2\pi}}{4}.$$

Cette question a été réussie par 1% de candidats.

Question 20 Ici, l'astuce consistait à observer que les intégrales

$$\int_{0}^{x_0} g(x)\sin(tf(x))\,dx, \quad \int_{x_0}^{1} g(x)\sin(tf(x))\,dx$$

se traitent de la même manière et ont le même comportement asymptotique.

Cette question a été bien traitée par 1% de candidats.