Banque BCPST Inter-ENS / ENPC / Mines/ Centrale - Session 2025

Rapport relatif à l'épreuve écrite de mathématiques

Écoles partageant cette épreuve :
ENS Paris-Saclay, ENS (Paris), ENS de Lyon, ENPC, Mines de Paris, Centrale

Épreuve comptant uniquement pour l'admission.

Coefficients (en pourcentage du total d'admission de chaque concours) :
- ENS Paris-Saclay : 6,2%
- ENS de Lyon : 6,6%
- ENS (Paris) : 13,7%
- ENPC/Mines/Centrale : 17,5%

Membres du jury : Paul Bastide, François Bienvenu, Miraine Davila Felipe

Présentation générale du sujet

L'objectif du sujet était de reproduire une partie des résultats d'un article scientifique récent [1] portant sur l'estimation de la moyenne d'un trait quantitatif au sein d'une espèce ancestrale à partir de la mesure du trait de ses descendants. La **partie I** demandait aux candidat·es de re-démontrer quelques résultats de cours portant sur les sommes de variables gaussiennes, et de les étendre à un formalisme matriciel. Tous les résultats utiles dans le reste du sujet étaient clairement identifiés et pouvaient être admis. La **partie II** consistait à démontrer l'identité matricielle de Woodbury, et à l'appliquer à quelques cas particuliers qui se retrouvent dans la suite. La **partie III** partait d'un résultat du programme sur l'estimation de la moyenne d'un échantillon indépendant et identiquement distribué par la moyenne empirique et de la démonstration de la loi faible des grands nombres pour progressivement complexifier la structure des données, et voir que le résultat de convergence valable dans le cas simple n'est plus vérifié lorsque l'on ajoute une certaine structure de dépendance entre les données. Dans la **partie IV**, cette structure de dépendance résultait d'un modèle d'évolution d'un trait continu sur un arbre phylogénétique. Ceci permettait de montrer que, dans ce modèle, il n'est pas possible d'estimer le trait de l'espèce ancestrale avec une précision arbitraire, et ce quel que soit le nombre d'espèces descendantes observées.

Remarques générales

Nous rappelons aux candidat·es l'importance de lire le sujet dans son intégralité pour en saisir la cohérence dans son ensemble. Ici, plusieurs résultats pouvaient être admis et ré-utilisés dans les questions suivantes. De nombreuses copies ont perdu du temps à re-démontrer des résultats dans un cas particulier, alors qu'elles l'avaient déjà démontré – souvent avec succès – dans un cas général quelques questions plus haut.

Nous insistons également sur l'importance de la rigueur des raisonnements et des notations. Certaines copies présentant de nombreuses ratures, ou noyant leur raisonnement sous des assertions hors sujet (ou fausses ou n'ayant pas de sens) ont été pénalisées comparées aux copies présentant des arguments clairs et concis. Par ailleurs, il est demandé aux candidat·es de respecter les notations de l'énoncé, ou à défaut d'expliciter clairement leurs choix de notations.

Il est demandé aux candidat·es de numéroter correctement les questions. Des copies qui sautaient d'une partie à l'autre et numérotaient leurs questions de façon incorrecte ont été pénalisées. Par ailleurs, les candidat·es sont libres de traiter les parties dans l'ordre de leur choix et de sauter des questions, mais il est conseillé, au sein d'une partie, de traiter les questions dans l'ordre du sujet et d'éviter de sauter d'une partie à l'autre dans la rédaction.

Enfin, certaines copies se sont pénalisées par leur écriture : lorsqu'une expression mathématique est absolument illisible, il est impossible de supposer qu'elle est correcte lors de la correction.

Commentaires détaillés sur chaque question

Partie I

La **partie I** consistait essentiellement à redémontrer des résultats du programme. Ces démonstrations, qui faisaient appel à divers points de probabilités et d'analyse du programme, étaient guidées, et les questions étaient pensées pour ne pas être bloquantes : chaque résultat intermédiaire pouvait être admis pour traiter le reste de la partie. La question **I.2.(a)** reposait notamment sur l'application du théorème de la bijection. La question **I.3** proposait de montrer qu'une variable gaussienne avec une plus grande variance est "moins concentrée" autour de sa variance. Elle ne pouvait pas être résolue par l'appel à l'inégalité de Bienaymé–Tchebychev, qui dans ce cas donnait des bornes supérieures non exploitables, mais par le recours à des variables centrées réduites telles que mises en avant par la question **I.2.(b)**. Elle permettait également de tester la compréhension par les candidat·es du comportement des variables aléatoires gaussiennes. La question **I.4** était assez guidée et plutôt calculatoire. Le corollaire de la question **I.4.(c)** nécessitait une bonne compréhension et maîtrise du calcul matriciel.

I.1.(a). Il existait plusieurs façons de traiter cette question, qui consistaient toutes à expliciter la densité de $aX + b$ et à reconnaître la densité de la loi normale. Ceci pouvait se faire soit en exprimant la fonction de répartition comme une intégrale, puis en effectuant un changement de variable ; soit en dérivant la fonction de répartition.

De nombreuses de copies n'ont pas identifié le problème posé par le signe de a dans l'inégalité $aX + b \leq x$. Parmi les copies ayant distingué les cas $a > 0$ et $a < 0$, certaines ont eu du mal à traiter le cas $a < 0$.

Un petit nombre de copies ont utilisé − correctement mais sans la justifier − la formule hors programme $f_Y(x) = f_X((x - b)/a)/|a|$ pour $Y = aX + b$. Ces copies se sont vu attribuer une partie des points.

Enfin, un nombre non négligeable de copies se sont contentées de calculer l'espérance et la variance $aX + b$ pour répondre à la question, ce qui est insuffisant en absence de l'une des justifications mentionnées précédemment permettant d'identifier la loi normale.

I.1.(b). De très nombreuses copies ont recopié la preuve de la question précédente, plutôt que d'appliquer le résultat de cette question.

I.2.(a). Cette question a été assez mal réussie : de nombreuses copies ont tenté de montrer séparément l'existence et l'unicité, sans penser à utiliser le théorème de la bijection. Peu de copies ont justifié correctement la continuité et la croissance de u.

I.2.(b). Cette question pouvait être traitée soit directement, par application de la relation de Chasles et un changement de variable, soit en exploitant la formule $\Phi(x) + \Phi(-x) = 1$ pour tout x réel et Φ la fonction de répartition de la loi normale centrée réduite.

I.3. Cette question pouvait être résolue en revenant à des variables centrées réduites afin de les comparer. La quasi-totalité des copies ont eu recours à l'inégalité de Bienaymé–Tchebychev, qui ne menait pourtant ici à rien – ce dont très peu de copies se sont rendu compte. Les rares copies ayant clairement mentionné que leur approche n'aboutissait pas, en expliquant pourquoi, ont pu, en fonction de la pertinence des explications, se voir récompenser.

De façon surprenante, la deuxième partie de la question a également pu poser problème. Un nombre non-négligeable de copies ont représenté les densités de $\mathcal{N}(\mu_1, \sigma_1^2)$ et de $\mathcal{N}(\mu_2, \sigma_2^2)$ de façon qualitativement incorrecte (par exemple en faisant coïncider le maximum de ces deux densités), et l'interprétation a parfois révélé une mauvaise compréhension des probabilités (comme l'affirmation erronée « *par définition de la variance, on a toujours* $\mathbb{P}(|X_1 - \mu_1| > \delta) \leq \mathbb{P}(|X_2 - \mu_2| > \delta)$ *lorsque* $\sigma_1^2 \leq \sigma_2^2$ »).

On note qu'ici l'hypothèse d'indépendance prise par l'énoncé n'était pas utile. Cela n'a pas semblé gêner les candidat·es ayant traité la question.

I.4.(a).i. Cette question a été traitée par la plupart des copies, et généralement réussie. Toutefois, les copies ayant fourni le bon résultat mais en utilisant de mauvais arguments – par exemple, en invoquant la quadraticité de la variance plutôt que l'indépendance entre X_1 et X_2 pour justifier $\mathbb{V}(X_1 + X_2) = \mathbb{V}(X_1) + \mathbb{V}(X_2)$ – ont été pénalisées.

I.4.(a).ii-iii. Ces questions, assez calculatoires, ont été plutôt bien réussies. Les copies les ayant traitées efficacement sont généralement celles qui sont parties de l'expression donnée par l'énoncé pour retomber sur l'expression de définition.

De nombreuses copies semblent avoir perdu beaucoup de temps sur ce calcul, alors que sa résolution n'était pas nécessaire pour poursuivre le sujet car le résultat était donné. Admettre un tel résultat peut éventuellement être une stratégie payante, puisqu'elle permet de gagner du temps et de traiter des questions plus intéressantes dans la suite du sujet.

Il est également conseillé d'utiliser un brouillon plutôt que de se lancer directement dans le calcul, afin d'éviter les trop nombreuses ratures qui peuvent pénaliser la copie lors de la correction.

I.4.(a).iv. Parmi les copies ayant traité cette question, la grande majorité se sont contentées de montrer que si $X_1 \sim \mathcal{N}(0, \sigma_1^2)$ et $X_2 \sim \mathcal{N}(0, \sigma_2^2)$ sont indépendantes, alors $X_1 + X_2 \sim N(0, \sigma_1^2 + \sigma_2^2)$, ce qui n'était pas ce qu'il était demandé de prouver.

Par ailleurs, si plusieurs copies ont correctement invoqué le produit de convolution pour calculer la densité de $X_1 + X_2$, beaucoup ont omis de préciser que ce résultat ne s'applique qu'aux variables indépendantes.

I.4.(b). La quasi-totalité des copies ayant traité cette question ont vu qu'il fallait procéder par récurrence. En revanche, un nombre significatif de copies ont mal fait l'initialisation, notamment en oubliant de considérer $a_i \neq 1$; ou n'ont pas justifié l'indépendance entre $\sum_{i=1}^{n} a_i X_i$ et X_{n+1} (par le lemme des coalitions).

Encore une fois, plutôt que d'utiliser le résultat de la question **4.(a)**, un nombre non-négligeable de copies ont choisi de recopier sa preuve. Ces copies n'ont pas été pénalisées, mais se sont pénalisées elles-mêmes en perdant du temps.

I.4.(c). Peu de copies ont traité cette question. La plupart des copies l'ayant traitée ont oublié de justifier que $a^T M$ était non-nul.

Partie II

La **partie II** portait sur l'identité matricielle de Woodbury, qui est relativement classique et très utilisée en statistique computationnelle notamment, car elle permet de calculer une inverse de manière récursive. La question **II.1** testait les candidat·es sur des notions d'algèbre linéaire et consistait à démontrer de manière guidée que deux espaces étaient liés par un isomorphisme. La question **II.3** consistait en l'application de la formule dans des cas particuliers, et testait l'aisance des candidat·es dans la manipulation de matrices (il fallait en particulier reconnaître qu'un certain produit matriciel donnait un scalaire).

II.1.(a)-(b). Ces questions ont été bien réussies. Toutefois, quasiment aucune copie n'a remarqué qu'il était possible d'utiliser **I.(a)** pour prouver **I.(b)**, en échangeant les rôles de (A, U_1) et de (C, U_2). Il a été observé dans plusieurs copies des tentatives d'inversion de matrices non carrées (comme $\mathbb{1}$ ou U_2), ce qui a été fortement pénalisé.

II.1.(c). Cette question a posé problème dans de nombreuses copies, qui n'ont pas justifié la validité de $f \circ g(x) = g \circ f(x) = x$. Certaines copies ont tenté de justifier séparément l'injectivité et la surjectivité de la restriction de f, par exemple en s'intéressant à $\ker(f)$; mais elles sont rarement parvenues à une preuve complète.

On note également que l'énoncé manquait de précision ici, et il aurait été plus correct d'écrire « Montrer que f *induit* un isomorphisme de $\ker(A + U_1 C U_2)$ dans $\ker(C^{-1} + U_2 A^{-1} U_1)$ ».

II.2. Cette question a été mal réussie, car peu de copies ont fait le lien avec la question précédente. Plusieurs copies ont tenté de "bluffer" en écrivant directement (ou avec seulement une étape intermédiaire) $(A + U_1 C U_2)B = I$, où B était l'expression donnée dans l'énoncé ; or il fallait plutôt calculer $B(A + U_1 C U_2)$. Ces copies ne se sont donc pas vu attribuer de points.

II.3.(a). La plupart des copies ayant traité cette question ont réalisé qu'il fallait utiliser le Lemme 2, mais peu de copies ont pensé à prendre une matrice C de taille 1×1 et ont à la place pris $C = I_n$, ce qui posait ensuite problème pour justifier l'inversibilité de $C^{-1} + U_2 A^{-1} U_1$ et pour simplifier l'expression finale.

Un nombre non-négligeable de copies ont tenté de justifier l'inversibilité de V en invoquant, comme prétendus résultats de cours, des énoncés incorrects d'algèbre linéaire (tels que « *ajouter une même constante à toutes les entrées d'une matrice ne change pas son rang* »).

II.3.(b)-(c). Ces questions n'ont été traitées que par un très petit nombre de copies, et demandaient une bonne maîtrise du calcul matriciel.

Partie III

La **partie III** était la partie la plus longue, et avec le plus de points au barème. Les questions allaient dans un ordre de difficulté croissant, et étaient de moins en moins guidées. Très peu de candidat·es en sont venu·es complètement à bout, mais beaucoup ont gagné des points sur les parties les plus proches du cours. La question **III.1**, indépendante des autres, consistait à démontrer des inégalités utiles par la suite, et qui pouvaient être admises. Elle faisait appel à des outils d'analyse classique (**III.1.(a)-(b)**), ainsi qu'à des résultats sur les matrices symétriques réelles (**III.1.(c)**). La question **III.2** reprenait l'estimateur empirique de la moyenne dans les conditions du programme, en considérant des variables indépendantes et identiquement distribuées, et invitait les candidat·es à re-démontrer la loi faible des grands nombre dans ce cas, ainsi qu'à réaliser que cet estimateur était le "meilleur" (au sens de la variance) au sein d'une classe d'estimateurs linéaires sans biais définie dans l'énoncé. La question **III.3** présentait un premier raffinement de ce résultat, en considérant des variables toujours indépendantes mais non identiquement distribuées, pour lesquelles une variante de la loi faible des grands nombres était établie. La question **III.4** introduisait de la dépendance entre les variables, ce qui rendait caduque la loi faible des grands nombres. Seules les meilleures copies se sont rendu compte du problème, et parmi elles aucune n'a proposé de démonstration convaincante de la non convergence de la probabilité de déviation étudiée. La question **III.5** consistait en une re-formulation du problème en termes matriciels, et la question **III.6** exploitait ce formalisme pour exhiber un estimateur de variance minimale. Ces deux questions n'ont été presque jamais abordées.

III.1.(a)-(b). Les indications données à ces questions avaient pour but de les rendre faisables sans avoir recours à l'inégalité de Cauchy-Schwarz, hors programme. Toutefois, il était bien entendu possible d'utiliser l'inégalité de Cauchy-Schwarz (et, de fait, un certain nombre de copies y ont eu recours).

Il existait une imprécision dans la formulation de la question : il n'était pas explicitement demandé de montrer que $b = (1/n)\mathbb{1}_n$ était l'*unique* vecteur pour lequel l'égalité était vérifiée, ce qui était pourtant l'intention du sujet. De ce fait, les copies s'étant contenté de montrer que b était un vecteur pour lequel l'inégalité était vérifiée n'ont pas été pénalisées.

III.1.(c).i. La plupart des copies ont réussi à montrer que V était symétrique et que $x^T V x \geq 0$. Cependant, très peu de copies ont correctement justifié que $x^T V x > 0$ lorsque $x \neq 0$, ce qui nécessitait d'utiliser l'hypothèse $\mathrm{rg}(M) = n$.

III.1.(c).ii. Cette question a été dans l'ensemble bien réussie ; toutefois, des rédactions regrettables telles que « une matrice symétrique réelle est *par définition* diagonalisable » ont été relevées. De nombreuses copies ont également omis de rappeler que les vecteurs propres étaient par définition non nuls, ce qui était ici essentiel dans la démonstration.

III.1.(c).iii. Peu de copies ont su justifier $x^T V^{-1} x > 0$ pour $x \neq 0$, ce qui pouvait se faire en explicitant la décomposition spectrale.

III.1.(c).iv-v. Ces questions n'ont quasiment pas été traitées. Elles demandaient principalement de la dextérité dans le calcul matriciel.

III.2.(a)-(e). Ces questions ont été fréquemment traitées, souvent avec succès. Toutefois, plusieurs copies n'ont pas su faire le lien avec les questions précédentes, et ont reproduit dans un cas particulier les démonstrations (parfois fausses) de résultats plus généraux déjà abordés plus haut.

III.2.(f). Plutôt que d'utiliser le résultat de la question **I.3**, de nombreuses copies ont de nouveau invoqué l'inégalité de Bienaymé–Tchebychev, qui n'était d'aucune aide ici.

III.2.(g). Cette question a été bien réussie dans l'ensemble. La réponse attendue consistait à utiliser l'inégalité de Bienaymé–Tchebychev pour montrer que $\mathbb{P}(|M_n - \mu| > \delta) \to 0$. Certaines copies ont à la place directement invoqué la loi faible des grands nombres, qui est bien au programme. Lorsque la justification était rigoureuse et qu'il était bien précisé qu'il s'agissait de la loi *faible* des grands nombres, ces copies se sont vu attribuer tous les points.

Certaines copies ont interprété $\mathbb{P}(|M_n - \mu| > \delta) \to 0$ en disant que M_n converge presque sûrement vers μ. Bien qu'ici on ait effectivement la convergence presque sûre de M_n vers μ (loi forte des grands nombres), la notion de convergence presque sûre est hors programme, et n'est en général pas impliquée par $\mathbb{P}(|M_n - \mu| > \delta) \to 0$. Ainsi, en fonction de la précision de la rédaction, les copies ayant évoqué la convergence presque sûre ont pu se voir pénalisées.

III.3. Question peu traitée ; beaucoup d'erreurs dans le calcul de l'espérance et de la variance de $Q_n(r)$.

III.4. Peu de copies ont traité cette question, et parmi le peu de copies l'ayant traitée, quasiment aucune copie n'a réalisé que les variables Y_i n'étaient pas indépendantes (certaines copies ont même prétendu qu'elles l'étaient en invoquant – de manière vague et nécessairement incorrecte – le lemme des coalitions) et n'est parvenue à calculer correctement la variance de M_n.

III.5-6. Questions très peu traitées, avec de nouveau beaucoup d'erreurs dans le calcul de l'espérance et de la variance de M_n, et des invocations incorrectes de la loi faible des grands nombres.

Partie IV

La **partie IV** présentait une modélisation de l'évolution d'un trait quantitatif le long d'un arbre phylogénétique dans un cas particulier, qui est le modèle ayant motivé tous les développements précédents. Un traitement complet de cette partie nécessitait une bonne compréhension générale du sujet. La question **IV.3.**, qui faisait appel à la question **III.6** (elle même peu traitée), n'a été que très peu abordée. En revanche, la question **IV.1** était très abordable, et la question **IV.2** ne présentait pas de difficulté technique, puisqu'elle consistait uniquement à prendre du recul sur le modèle présenté. Ces deux questions ont été correctement résolues dans de nombreuses copies.

IV. L'énoncé comportait ici aussi une imprécision, en évoquant une "indépendance deux à deux", hors programme, à la place d'une "indépendance mutuelle", en accord avec les résultats des parties précédentes. Cet abus de langage n'a pas semblé gêner les candidat·es ayant traité cette partie.

IV.1.(a)-(b). Ces questions ont été plutôt réussies par la quasi-totalité des copies les ayant traitées. Toutefois, l'hypothèse d'indépendance entre les incréments, nécessaire pour conclure, a rarement été rappelée explicitement.

IV.1.(c). Cette question a posé problème à la plupart des copies, qui n'ont là encore pas réalisé que les variables X_A et X_B n'étaient pas indépendantes. Parmi celles l'ayant réalisé et ayant correctement appliqué le Corollaire 1, peu de copies ont pris la peine de justifier son applicabilité (en vérifiant que la matrice M_2 est bien de rang 2).

IV.1.(d). Comme les questions **IV.1.(a)-(b)**, cette question a été bien réussie.

IV.2. Cette question a été bien réussie dans l'ensemble, mais peu de copies ont relevé/commenté l'hypothèse faite dans la modélisation que, suite à une spéciation, les deux espèces filles évoluent indépendamment.

IV.3. Cette question n'a quasiment pas été traitée ; et, lorsque cela a été le cas, elle n'a pas été réussie : presque aucune copie n'est parvenue à la conclusion qu'il était impossible de remonter à la valeur ancestrale μ du trait, et ce quelle que soit la quantité de données disponibles.

Références

[1] Ho, L. S. T. and Ané, C. (2014). A Linear-Time Algorithm for Gaussian and Non-Gaussian Trait Evolution Models. *Systematic Biology*, 63(3) :397–408.