ÉPREUVE ÉCRITE de MATHÉMATIQUES 2006
ENS : PARIS LYON CACHAN

Durée : 4 heures
Coefficients/Total général : Paris 16/145 ; Lyon 4/58.5 ; Cachan 4/65

MEMBRES du JURY : Arnaud BASSON, Amaury LAMBERT

La question de la *richesse spécifique*, c'est-à-dire de la distribution des abondances des espèces, a été abordée à plusieurs reprises au cours du XX$^{\text{e}}$ siècle. Les trois modèles mathématiques les plus célèbres formaient le sujet de l'épreuve 2006.

Les notes vont de 0,45 à 20. La moyenne est de 8,38 et l'écart-type de 4,12. 53% des copies obtiennent une note inférieure à 8, 29% ont la moyenne, 10% ont 15 ou plus.

Survol du sujet, corrigé succinct, perspectives

Il s'agit en général de savoir ce que prédit un *modèle neutre*, c'est-à-dire un modèle où ne sont pas prises en compte les différences d'adaptation et les interactions entre espèces, de façon à pouvoir détecter ces dernières en comparant les observations des distributions d'abondances aux prédictions du modèle. En un mot, la question se réduit à : « quelle serait la richesse spécifique si celle-ci était le fruit du hasard ? »

De tels modèles s'appliquent également à la question de la diversité génétique des populations (partie II) ou de l'estimation du nombre d'espèces inconnues (partie III).

La première partie du sujet était la plus difficile. Le modèle qui y est exposé est celui du « *stick-breaking* » du célèbre écologiste MacArthur, que l'on a traduit par « partage aléatoire ». L'hypothèse de MacArthur est que l'abondance d'une espèce est proportionnelle à la quantité de ressources dont elle dispose, et que les ressources disponibles sont réparties au hasard entre $n+1$ espèces, exactement comme si l'on jetait uniformément n points dans l'intervalle $[\,0\,;1\,]$ et que chaque intervalle représente la part des ressources allouée à chaque espèce. Malgré la simplicité du problème ainsi posé, la loi de variables aléatoires telles que le plus petit ou le plus grand fragment, est difficile à calculer.

Dans la question 1, on utilise le fait que le k-ième point X_k tombe dans $[\,x\,;x+dx\,[$ ssi un des n points tombe dans cet intervalle (n possibilités), et parmi les $n-1$ restants, $k-1$ points tombent dans $[\,0\,;x\,[$ et $n-k$ dans $[\,x\,;1\,]$, ce qui donne

$$\mathbb{P}_{n+1}(\mathrm{X}_k \in dx) = n\, \mathrm{C}_{n-1}^{k-1}\, x^{k-1}(1-x)^{n-k}\, dx.$$

En intégrant cette équation, on en déduit notamment que pour tous entiers i et j

$$\int_0^1 x^i(1-x)^j\, dx = \frac{i\,!\; j\,!}{(i+j+1)\,!}\ .$$

De plus, comme les $n+1$ fragments ont la même espérance et que leur somme vaut 1, on en déduit que cette espérance vaut $(n+1)^{-1}$.

Dans la question 2, on s'intéresse à la loi du plus petit fragment, soit Y_1. Sous $\mathbb{P}_2$, c'est-à-dire lorsqu'un seul point est jeté (deux espèces), $\mathrm{Y}_1 > y$ si et seulement si le point jeté est dans l'intervalle $]\,y\,;1-y\,[$, ce qui prouve que $\mathbb{P}_2(\mathrm{Y}_1 > y) = 1-2y$. L'hypothèse de récurrence au rang $n \geq 2$ est alors

$$\mathbb{P}_{n+1}(\mathrm{Y}_1 > y) = (1-(n+1)y)^n \qquad y \leq (n+1)^{-1}.$$

La question 2.b-i avait pour but de mettre les candidats sur la piste de la renormalisation. En effet, pour répondre à la question 2.b-ii, on peut montrer que sous $\mathbb{P}_{n+1}$, quitte à placer le plus petit intervalle à l'extrémité gauche de $[\,0\,;1\,]$ ($n+1$ possibilités), $\mathrm{Y}_1 \in dy$ si et seulement si un des n points tombe dans $[\,y\,;y+dy\,[$ (n possibilités), les $n-1$ points restants tombent dans l'intervalle $]\,y\,;1\,]$, et dans cet intervalle de largeur $1-y$, le plus petit fragment est de taille supérieure à y. Par un argument de renormalisation, la probabilité de ce dernier événement est égale à $\mathbb{P}_n((1-y)\mathrm{Y}_1 > y)$.

Pour vérifier l'hypothèse de récurrence au rang $n+1$, il fallait donc écrire

$$\begin{aligned}\mathbb{P}_{n+1}(\mathrm{Y}_1 > y) &= \int_y^{1/(n+1)} \mathbb{P}_{n+1}(\mathrm{Y}_1 \in dx) \\ &= \int_y^{1/(n+1)} n(n+1)(1-x)^{n-1}\mathbb{P}_n(\mathrm{Y}_1 > x/(1-x))\,dx\end{aligned}$$

et appliquer l'hypothèse de rang n à $y = x/(1-x)$. Pour la question 2.d, une intégration par parties très classique fournit l'égalité $\mathbb{E}_{n+1}(\mathrm{Y}_1) = \int_0^{1/(n+1)} \mathbb{P}_{n+1}(\mathrm{Y}_1 > x)\,dx$. Ceci permet de montrer que cette espérance vaut $(n+1)^{-2}$, à comparer avec l'espérance $(n+1)^{-1}$ d'un fragment pris au hasard.

L'objet de la question 3 était d'établir une relation entre $\mathbb{E}_{n+1}(\mathrm{Y}_{k+1})$ et $\mathbb{E}_n(\mathrm{Y}_k)$ afin de trouver une expression générale pour ces espérances. Conditionnellement au fait que le plus petit fragment mesure y, la taille de tous les autres fragments est supérieure à y. Si l'on retranche alors un intervalle de largeur y à chacun de ces fragments, la configuration obtenue a même loi que celle où l'on a jeté uniformément $n-1$ points dans un intervalle de largeur totale $1-(n+1)y$. Grâce à la question 3.a-i, on trouve la loi conditionnelle du n-uplet $(\mathrm{Y}_2, \dots, \mathrm{Y}_{n+1})$, avec $a = (n+1)y$. On obtient par conséquent

$$\mathbb{E}_{n+1}(\mathrm{Y}_{k+1} \mid \mathrm{Y}_1 = y) = \mathbb{E}_n(y + (1-(n+1)y)\mathrm{Y}_k) = y + (1-(n+1)y)\mathbb{E}_n(\mathrm{Y}_k),$$

ce qui donne, en intégrant par rapport à la loi de Y_1,

$$\mathbb{E}_{n+1}(\mathrm{Y}_{k+1}) = \mathbb{E}_{n+1}(\mathrm{Y}_1) + (1-(n+1)\mathbb{E}_{n+1}(\mathrm{Y}_1))\mathbb{E}_n(\mathrm{Y}_k) = \frac{1}{(n+1)^2} + \frac{n}{n+1}\mathbb{E}_n(\mathrm{Y}_k).$$

Il est alors extrêmement intéressant de voir que l'on peut calculer explicitement $\mathbb{E}_n(\mathrm{Y}_k)$. Notamment, la taille moyenne du plus grand fragment est équivalente, lorsque $n \to \infty$, à $\ln(n)/n$ (non demandé), et le quotient de l'abondance moyenne de l'espèce la plus commune à celle de l'espèce la plus rare croît asymptotiquement comme $n\ln(n)$, lorsqu'il y a n espèces en présence. On peut également obtenir un équivalent de l'abondance relative moyenne de la k-ième espèce la plus rare sous la forme

$$\mathbb{E}_n(\mathrm{Y}_k) \sim -\frac{1}{n}\ln\left(1-\frac{k}{n}\right) \sim \frac{k}{n^2}\ ,$$

lorsque k est fixé et que $n \to \infty$.

Dans la deuxième partie, il s'agissait d'analyser le modèle que l'on obtient après renormalisation du modèle d'extinction-spéciation suivant, dit *modèle à somme nulle de Hubbell* (non mentionné dans le sujet). Dans une population de taille constante N, chaque individu qui meurt est immédiatement remplacé par un nouvel individu dont l'espèce est :

- une toute nouvelle espèce, avec la *probabilité de mutation* u
- choisie au hasard, au prorata des espèces en présence, avec la probabilité $1-u$.

Si l'on échantillonne au hasard n individus dans une telle population prise dans son état stationnaire, lorsque N est grand et que u est petit, équivalent à $\theta/(2\mathrm{N})$, on obtient alors le modèle décrit dans la deuxième partie (mais cette relation n'est pas immédiate). Ce modèle appelé « la métaphore de la cantine » dans le sujet est plus connu sous le nom de « *processus du restaurant chinois* ».

Dans cette partie, n désigne le nombre d'invidus, de sorte que le nombre d'espèces K_n représentées dans cet échantillon est inférieur à n (dans la partie précédente, n était le nombre d'espèces, et l'on s'intéressait à leurs abondances renormalisées). En utilisant le fait que la probabilité que le convive $k+1$ s'asseye à la seule table occupée sachant qu'une seule table est occupée vaut $k(k+\theta)^{-1}$, on montre par une récurrence immédiate l'expression demandée pour la probabilité $q_{n+1,1}$ qu'une seule table soit occupée. La formule des probabilités totales (deux événements...) permet ensuite d'établir que

$$q_{n+1,i} = \frac{n}{n+\theta}\ q_{n,i} + \frac{\theta}{n+\theta}\ q_{n,i-1}\ ,$$

et d'en déduire (question 2) que

$$\mathrm{P}_{n+1}(\mathrm{X}) = \frac{n+\theta\mathrm{X}}{n+\theta}\ \mathrm{P}_n(\mathrm{X}).$$

La question 3 guidait les candidats vers les formules suivantes

$$\mathbb{E}(\mathrm{K}_n) = \sum_{k=0}^{n-1}\frac{\theta}{k+\theta} \qquad \text{et} \qquad \mathrm{Var}(\mathrm{K}_n) = \sum_{k=0}^{n-1}\frac{k\ \theta}{(k+\theta)^2}\ .$$

Dans la question 4, on voit qu'il est beaucoup plus astucieux d'écrire K_n sous la forme

$$\mathrm{K}_n = \sum_{i=1}^{n}\varepsilon_i,$$

où les (ε_i) sont des variables de Bernoulli indépendantes d'espérance $\theta/(i-1+\theta)$, pour $i = 1, \ldots, n$. On retrouve alors immédiatement les formules de l'espérance et de la variance données plus haut (en précisant comme il se doit que l'espérance est linéaire, et que la variance d'une somme de variables indépendantes égale la somme de leurs variances). Enfin, les questions 5 et 6 proposaient un encadrement intégral pour ces quantités, permettant de fournir les équivalents suivants

$$\mathbb{E}(\mathrm{K}_n) \sim \theta \ln(n) \quad \text{et également} \quad \mathrm{Var}(\mathrm{K}_n) \sim \theta \ln(n),$$

lorsque la taille de l'échantillon n tend vers $+\infty$.

La troisième et dernière partie expose en détail les résultats anciens du statisticien et généticien R.A. Fisher sur la loi de l'abondance d'une seule espèce isolée.

Les sous-parties A et B sont préliminaires. Dans la sous-partie A, on définit la fonction Γ, qui est le prolongement aux nombres réels positifs de la fonction factorielle, et on montre que si X est une variable aléatoire suivant la loi Gamma(θ), alors

$$\mathbb{E}(\mathrm{X}) = \mathrm{Var}(\mathrm{X}) = \theta.$$

Dans la sous-partie B, il s'agit de montrer que pour tout réel x de l'intervalle $]-1\,;1[$,

$$\sum_{n=1}^{\infty} \frac{x^n}{n} = -\ln(1-x).$$

La sous-partie C se concentre sur l'étude du couple (D, X), où D désigne la *densité géographique* de l'espèce considérée et X le *nombre d'individus* qui appartiennent à cette espèce *et* qui ont été *observés*. On suppose qu'il existe un réel positif ρ tel que D/ρ suive la loi Gamma(θ), et que conditionnellement à $\mathrm{D} = d$, X suive la loi de Poisson de paramètre d. On peut alors en déduire facilement (question 9) la loi du couple, ainsi que la loi de X. De plus (question 10), la formule des probabilités conditionnelles donne

$$\mathbb{P}(\mathrm{G} \in dy \mid \mathrm{X} = k) = y^{k+\theta-1} e^{-(\rho+1)y} \frac{(\rho+1)^{k+\theta}}{\Gamma(k+\theta)}\, dy,$$

de telle sorte que

$$\mathbb{P}((\rho+1)\mathrm{D}/\rho \in dx \mid \mathrm{X} = k) = \frac{1}{\Gamma(k+\theta)}\, e^{-x} x^{k+\theta-1}\, dx,$$

c'est-à-dire que conditionnellement à $\mathrm{X} = k$, $(\rho+1)\mathrm{D}/\rho$ suit la loi Gamma$(k+\theta)$.

Dans la question 11, on s'intéresse à la limite lorsque $\theta \to 0^+$ de la loi du couple (D, X) conditionnée par le fait que l'on s'intéresse à une espèce qui est connue (c'est-à-dire $\mathrm{X} \geq 1$). On peut montrer que

$$\mathbb{P}^\star(\mathrm{X} = k, \mathrm{G} \in dy) := \lim_{\theta \to 0^+} \mathbb{P}(\mathrm{X} = k, \mathrm{G} \in dy \mid \mathrm{X} \geqslant 1) = c\, \frac{\rho^k}{k!}\, y^{k-1}\, \mathrm{e}^{-(\rho+1)y}\, dy,$$

où le calcul de la constante c nécessite d'utiliser les résultats de la sous-partie A sur la fonction Γ, et donne $c = 1/\ln(1+\rho)$. En intégrant par rapport à la variable y, on obtient notamment

$$\mathbb{P}^\star(\mathrm{X} = k) = c\,\frac{x^k}{k} \qquad k \geqslant 1,$$

où $x = \rho/(\rho+1)$, ce qui justifie le nom de *série logarithmique* (voir sous-partie B). Des calculs classiques mais légèrement fastidieux montrent que

$$\begin{aligned}
\mathbb{E}^\star(\mathrm{D}) &= c\rho x \\
\mathbb{E}^\star(\mathrm{X}) &= c\rho \\
\mathbb{E}^\star(\mathrm{DX}) &= c\rho^2 \\
\mathrm{Cov}^\star(\mathrm{D},\mathrm{X}) &= c^2\rho^2(-\ln(1-x) - x).
\end{aligned}$$

Cette dernière covariance est toujours positive, ce qui n'est pas surprenant, dans la mesure où plus la densité est grande, plus on s'attend à observer d'individus.

La dernière question donne un aperçu des conclusions que l'on peut tirer de ces résultats. Si les S espèces d'un écosystème ont des abondances indépendantes et de même loi donnée par $\mathbb{P}^\star$, le nombre attendu d'espèces singletons vaut

$$\mathrm{M} = \mathrm{S}\,\mathbb{P}^\star(\mathrm{X} = 1) = \mathrm{S}cx,$$

et le nombre attendu d'individus observés au total vaut

$$\mathrm{N} = \mathrm{S}\,\mathbb{E}^\star(\mathrm{X}) = \mathrm{S}c\rho.$$

En particulier, si l'on oberve N *individus dont* M *sont les seuls représentants de leur espèce*, on peut estimer ρ par $(\mathrm{N}-\mathrm{M})/\mathrm{M}$. Enfin si $f(x)$ est le quotient S/N du nombre d'espèces par le nombre d'individus observés, on a

$$f(x) = -\frac{1-x}{x}\,\ln(1-x),$$

qui est une fonction décroissante de $x = \rho/(\rho+1)$, dont la limite vaut 1 lorsque $x \to 0^+$ (beaucoup d'espèces très peu abondantes), et 0 lorsque $x \to 1^-$ (peu d'espèces très abondantes).

Commentaire détaillé du travail des candidats

De façon générale, nous avons constaté que les candidats savent faire preuve d'une certaine initiative pour mener des raisonnements probabilistes en s'aidant de leur intuition des situations modélisées (parties I et II). Au contraire, ils laissent souvent apparaître de graves lacunes lorsqu'il s'agit de mettre en œuvre des arguments d'analyse (aspect calculatoire mis à part) : convergence d'une série, estimation d'un équivalent d'intégrale, etc. (partie III principalement).

Les questions appelant un raisonnement de nature probabiliste et qui ont été traitées de manière satisfaisante incluent notamment les questions I.1.a, I.2.a-i, I.3.a-i, II.1.a, II.1.b, II.4.a. Il fallait cependant prendre garde à ne pas s'enliser dans des explications confuses et peu convaincantes.

À la question I.1.c, nous attendions une preuve que $\mathbb{E}(\mathrm{X}_k - \mathrm{X}_{k-1}) = (n+1)^{-1}$ par le calcul (les deux questions précédentes fournissaient les outils nécessaires) ; quelques points ont néanmoins été accordés aux candidats qui ont justifié cette égalité par un raisonnement qualitatif (« tous les intervalles ont la même longueur en moyenne »), mais il a été constaté à cette occasion que certains candidats confondent *l'espérance* de la longueur des intervalles (les $\mathbb{E}(\mathrm{X}_k - \mathrm{X}_{k-1})$) avec la *moyenne* des longueurs des $n+1$ intervalles ($(n+1)^{-1}\sum_{k=1}^{n+1}(\mathrm{X}_k - \mathrm{X}_{k-1})$), qui vaut effectivement $(n+1)^{-1}$ également, quelles que soient les valeurs des X_k...

La question I.1.d a donné lieu à des réponses surprenantes : de nombreux candidats sont incapables de justifier que généralement $\mathbb{E}(\mathrm{Y}_k) \neq (n+1)^{-1}$, mais voient aisément que $\mathbb{E}(\mathrm{Y}_{n+1}) > (n+1)^{-1}$. Ce cas particulier ($k = n+1$) mettait pourtant bien en lumière la situation générale.

Aux questions I.1.d et I.2.a-i, certaines copies confondent espérance et probabilité (confusion entre $\mathbb{E}(\mathrm{Y}_1) < (n+1)^{-1}$ et $\mathbb{P}(\mathrm{Y}_1 < (n+1)^{-1}) = 1$). Il semblerait que des candidats aient entendu parler d'événement presque sûr ou presque impossible (il en est ainsi de $\{\mathrm{Y}_1 = (n+1)^{-1}\}$), mais la notion n'est bien entendu pas maîtrisée ; à moins que cet usage du mot « presque » ne révèle une conception précaire des probabilités.

La question I.2.a-ii était triviale, puisque $\mathrm{Y}_1 > y \Leftrightarrow \mathrm{X}_1 \in]y, 1-y[$. Pourtant nombreuses sont les copies qui affirment comme une évidence que Y_1 est une v.a. uniforme sur $[0, \frac{1}{2}]$ – ce qui rend d'ailleurs la question tautologique – et en déduisent ensuite que $\mathbb{P}_2(\mathrm{Y}_1 > y) = 1 - 2y$ – ce qui est l'inverse du raisonnement attendu. Quelques copies confondent le nombre de segments et le nombre de points jetés X_k, sans que cela ne les empêche toujours de « montrer » le résultat demandé...

Le changement de variable linéaire $(u_1, ..., u_n) = a(u'_1, ..., u'_n)$ suffit à résoudre la question I.2.b-i, mais très peu de copies y pensent. Les réponses s'aidant d'un raisonnement qualitatif sans calcul ont été acceptées.

Les questions I.2.b-ii et I.3.a-ii étaient les des deux questions les plus difficiles du problème. Chacune d'elles a été résolue par un seul candidat (différent).

En I.2.c, il suffisait d'intégrer la relation de la question précédente. De bien trop nombreux étourdis confondent $\mathbb{P}(\mathrm{Y} > y)\,dy$ et $\mathbb{P}(\mathrm{Y} \in dy)$.

La fin de cette partie (à partir de I.3.a-ii) n'a quasiment pas été abordée.

La seconde partie débutait par une question facile, que l'on pouvait traiter par récurrence ou à l'aide de la formule des probabilités composées, ou encore à l'aide de l'indépendance des choix des convives ; mais dans ce dernier cas, il fallait formuler correctement l'argument d'indépendance. En effet, les événements {le convive i s'asseoit à la seule table occupée} et {le convive $i+1$ s'asseoit à la seule table occupée} ne sont bien sûr pas indépendants.

La question II.2.b. appelait un raisonnement par récurrence qui a été le plus souvent réussi. De façon générale, la technique du raisonnement par récurrence est bien maîtrisée (de même aux questions III.2.b et III.5). Les questions II.3.a et II.3.b ont été bien traitées aussi, la technique de calcul de l'espérance et de la variance d'une variable aléatoire discrète X à l'aide des dérivées du polynôme $\mathrm{P}(\lambda) = \sum_{k=0}^{n} \mathbb{P}(\mathrm{X} = k)\lambda^k$ semble familière aux candidats.

En II.4.b, ceux qui obtiennent un résultat différent de celui obtenu en II.3.a ou b sans avoir l'air de s'en émouvoir manquent de présence d'esprit.

L'énoncé comportait une coquille dans la question II.5.a, où il fallait en fait montrer

$$1 + \int_1^n \frac{\theta}{\theta + x}\,dx \leqslant \mathbb{E}(\mathrm{K}_n) \leqslant 1 + \int_0^{n-1} \frac{\theta}{\theta + x}\,dx.$$

Il a été tenu compte de cette erreur dans le barème, et certains candidats ont d'ailleurs su rectifier l'erreur. Il est tout de même décevant de voir des copies parvenir à « prouver » l'encadrement tel qu'il était écrit, dans la mesure où le majorant est clairement inférieur au minorant. En tous les cas, la technique de comparaison série-intégrale n'est pas inconnue des candidats, mais malheureusement sa mise en œuvre est fort laborieuse et beaucoup se fourvoient en cours de route.

La coquille de la question II.5.a n'avait aucune influence sur la question II.5.b, où il est apparu que les candidats ne maîtrisent pas du tout le calcul des équivalents (dont la définition leur est néanmoins connue). De plus, les simplifications d'équivalents sont très rarement faites, et nous avons souvent lu des équivalents tels que $\theta\ln((\theta+n)/\theta)$, voire carrément $1 + \theta\ln((\theta+n)/\theta)$ sans qu'il soit vu que ces quantités sont équivalentes à $\theta\ln(n)$.

La question II.6, où il fallait invoquer la convergence de la série $\sum_i \theta^2(i+\theta)^{-2}$ et le fait qu'une constante est un $o(\ln n)$ (seules deux copies y sont parvenues), a donné lieu à des réponses fantaisistes.

La troisième et dernière partie du sujet débutait par une question apparemment fort simple : justifier la définition de la fonction $\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}\,dt$. Nous avons malheureusement été extrêmement surpris de constater que la question de la convergence d'une intégrale généralisée est tout simplement étrangère à la majorité des candidats, qui ne voient pas où se situe le problème, et jugent que l'intégrale est bien définie puisque l'intégrande l'est elle-même. Nous en avons été réduits à valoriser le simple fait d'avoir vu que la justification était nécessaire. Dans tous les cas, un grand nombre de ceux qui traitent la convergence au voisinage de $+\infty$ oublient la convergence en 0^+, et réciproquement. Enfin, un certain nombre de copies pensent qu'il suffit de montrer que $t^{x-1}e^{-t} \to 0$ quand $t \to +\infty$ pour prouver la convergence au voisinage de $+\infty$, tandis que d'autre sont bloquées par le simple fait que $t^{x-1}e^{-t}$ diverge en 0^+. On voit aussi des copies qui pensent s'en tirer à l'aide du fait que l'intégrale $\int_0^\infty t^{x-1}\,dt$ converge... Bref, cette question a été traitée de manière catastrophique. Seule une dizaine de copies savent en fournir une preuve à peu près complète.

En III.2.c il était nécessaire d'invoquer la continuité de la fonction Γ en $x = 1$ pour montrer que $\Gamma(x) = \Gamma(x+1)/x \sim_{x\to 0} \Gamma(1)/x = 1/x$. Les copies qui ont fait mention de ce problème de continuité ont été valorisées, mais aucune démonstration n'a été exigée.

La question III.4 a elle aussi été fort mal réussie. À l'instar de la convergence des intégrales, la convergence des séries pose d'insurmontables problèmes à bien des candidats. Si la convergence de la série $\sum n^{-1}x^n$ lorsque $x \in [0,1[$ paraît abordable (quoique certains candidats pensent encore qu'il suffit de majorer par la série $\sum n^{-1}$...), le cas où $x \in]-1,0[$ donne lieu à beaucoup d'inventions. Cette série a beau être de signe alterné, de nombreuses personnes pensent qu'il suffit de majorer le terme général (et fournissent d'emblée la majoration grossièrement fausse $n^{-1}x^n \leqslant x^n$). Pour celles qui pensent à majorer le module $\left|n^{-1}x^n\right|$, il serait bon de rappeler sur la copie qu'une série absolument convergente est convergente. Mais la notion de convergence absolue est apparemment peu connue du plus grand nombre.

La question III.6, qui n'est autre qu'une formule de Taylor avec reste intégral et dont on attendait une preuve par récurrence, a été paradoxalement délaissée.

Les candidats qui traitent la question III.7 voient bien que $(x-t)^n(1-t)^{-n-1}$ tend vers 0 pour tout $t \in [0,x]$, mais la convergence simple ne suffit pas pour effectuer rigoureusement le passage à la limite sous l'intégrale... Rares sont les copies qui y parviennent, en montrant par exemple que $\sup_t \left|(x-t)/(1-t)\right| < 1$ pour justifier la convergence uniforme de l'intégrande.

Les questions III.8.a et III.8.b sont des questions de cours, et leur réponse semble connue par cœur de la plupart des candidats. Il est tout de même bon de préciser ce que l'on entend par des notations telles que $\mathcal{B}(n,p)$ et $\mathcal{P}(d)$, car il ne nous semble pas que ces notations soient canoniques. À noter toutefois que certains candidats confondent l'approximation de la loi binômiale de paramètres N et p par une loi de Poisson lorsque $\mathrm{N} \to \infty$, $p \sim d/\mathrm{N}$, avec l'approximation par une loi normale (une poignée de candidats pense que le résultat est tantôt une loi de Poisson, tantôt une loi normale suivant la valeur du paramètre d...)

On introduisait ensuite deux variables X (suivant une loi de Poisson) et D (suivant une loi Gamma) ; la formulation de l'énoncé contenait une imprécision, mais qui n'intervenait qu'à la question III.9.a, où la majorité des candidats a heureusement compris qu'il fallait écrire $\mathbb{P}(\mathrm{X}=k, \mathrm{G} \in dy) = \mathbb{P}(\mathrm{X}=k \mid \mathrm{G}=y)\,\mathbb{P}(\mathrm{G} \in dy)$, et le résultat suivait aussitôt.

En III.9.b, il suffisait d'intégrer la relation fournie à la question précédente, puis de poser $t = (\rho+1)y$ dans l'intégrale ; de façon étonnante, la grande majorité des candidats n'a pas vu ce changement de variable simple. On est d'autant plus surpris que la même faiblesse se répète à la question III.10.b, qui se limitait à une réécriture du résultat de la question III.10.a avec ce même changement de variable linéaire. Un nombre non négligeable de candidats s'y est embrouillé ; de même encore en III.11.c.

À la question III.10.c, le calcul demandé éait immédiat (on trouve tout de même de temps à autre des erreurs dans les copies). Quant au commentaire qui était demandé, la fatigue de fin d'épreuve a donné lieu à des réponses qui souvent étaient soit absurdes soit tautologiques.

En III.11.b, un réflexe conditionné a poussé les candidats à calculer la limite $\rho \to 0$ alors que l'énoncé demandait de faire tendre θ vers 0, ce qui était moins évident.

Les valeurs des constantes c et x de la question 11 n'étaient pas données dans l'énoncé et il appartenait aux candidats de les déterminer. Ceci a nous a fait constater combien il est difficile pour certains de mener correctement des calculs lorsque le résultat n'est pas fourni...

En pratique, les questions III.11.b, c, d n'ont vraiment été abordées que par de rares candidats, et la III.12 n'a essentiellement pas été traitée.

La première partie était la plus difficile (une petite moitié des copies l'esquive d'ailleurs après deux ou trois questions, voire dans sa totalité), et la troisième la plus facile. Néanmoins c'est la seconde qui a nettement été la mieux réussie : ceci s'explique par les grosses difficultés rencontrées par les candidats en analyse, qui les ont sérieusement pénalisés dans la partie III.

La stratégie consistant à se concentrer sur une seule des trois parties n'était pas payante, et il valait mieux répartir son effort sur au moins deux parties : à deux ou trois exceptions près, toutes les copies ayant obtenu une note au-dessus de la moyenne ont abordé plusieurs parties. Les candidats ayant obtenu une (très) bonne note ont généralement traité au moins deux parties à fond.

Il apparaît à travers ce sujet que les probabilités sont un point fort des candidats. On pourra encore progresser sur ce point en soignant davantage la rigueur des raisonnements et en veillant à éviter les explications imprécises (qui cherchent souvent à noyer le poisson).

Rappelons aux candidats que les réponses malhonnêtes et autres tentatives de mystifier l'examinateur sont très facilement décelées lors de la correction, et sévèrement sanctionnées. L'exigence d'une stricte honnêteté intellectuelle ne souffre pas d'entorse.

Les techniques de base telles que la récurrence et les manipulations élémentaires des intégrales étudiées en première année de classe préparatoire sont heureusement maîtrisées (sauf dans les pires copies). Les capacités calculatoires des candidats sont généralement satisfaisantes.

À l'opposé, il apparaît qu'un certain nombre de techniques d'analyse vues en deuxième année ne sont que rarement maîtrisées par les candidats, et c'est là un constat alarmant. Il convient de consacrer un effort important pendant l'année à cet aspect du programme, car il s'agit de techniques tout à fait basiques sans lesquelles l'outil mathématique est inefficient.

On constate aussi, et ce n'est pas sans lien avec la remarque précédente, un manque de recul évident face aux résultats des calculs ; un exemple significatif est celui de la question II.3.b, où certains candidats trouvent sans ciller que $\mathrm{Var}(\mathrm{K}_n) = 0$. Le sentiment dominant est que les candidats ne comprennent pas vraiment la signification des calculs qu'ils effectuent (comme s'il s'agissait de recettes). Il faut donc s'entraîner à prendre de la hauteur par rapport à ce que l'on écrit et s'habituer à en vérifier systématiquement la pertinence. Cette nécessaire lucidité et cet esprit critique ne peuvent s'acquérir que par une pratique assidue tout au long de l'année ; c'est une tournure d'esprit indispensable à quiconque se destine à une carrière dans la recherche ou l'enseignement, et qui est donc appréciée et valorisée au concours d'entrée des Écoles Normales.

Naturellement, le manque de recul des candidats se traduit aussi par de grandes difficultés à faire le lien entre les résultats obtenus et leur interprétation dans le cadre des modèles étudiés (ainsi aux questions I.1.d, III.10.c, III.11.d). Là encore il faut s'exercer davantage : c'est la capacité d'utiliser des outils mathématiques pour traiter des problèmes issus de la biologie qui est en jeu ici.

À ce propos, nous redisons comme chaque année que les Écoles Normales ne pourront pas longtemps faire l'économie, à tous les sens du terme, de repenser la stature des mathématiques au concours BCPST. La nécessité croissante pour la recherche contemporaine de biologistes rompus à l'usage des mathématiques, et le rôle prépondérant des Écoles Normales dans leur formation, contrastent de manière criante avec les coefficients dont les mathématiques sont pourvues au concours d'entrée.