

ÉPREUVE ÉCRITE de MATHÉMATIQUES ENS : PARIS LYON CACHAN

Durée : 4 heures

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

MEMBRE du JURY : Amaury LAMBERT

Contrairement à une croyance généreusement répandue, les mathématicien(ne)s n'ont pas attendu l'émergence de la génomique, de la protéomique, ou autre bioinformatique, pour apporter leur concours aux sciences du vivant, ni pour s'en nourrir. Il suffit, pour s'en convaincre, d'aborder brièvement le simple exemple de l'écologie et de l'évolution, qui formaient le cadre d'étude du sujet 2005.

Dès avant 1850, les équations différentielles étaient utilisées en démographie et en biologie des populations (Euler, Malthus, Verhulst). À la fin du XIX^{ème} siècle, la théorie des probabilités et des statistiques amorçait avec la génétique naissante et avec la dynamique des populations un dialogue qui ne se démentira plus (Mendel, Quételet, Galton, Pearson). Les premières décennies du XX^{ème} siècle verront les débuts d'une épopée triomphante, celle de la génétique des populations et de l'écologie, principalement grâce à l'alliance des probabilités et des équations aux dérivées partielles (Fisher, Wright, Haldane, Lotka, Volterra, McKendrick, Kolmogorov). Et cette histoire, évidemment, se poursuit jusqu'à nos jours, notamment à travers la théorie neutraliste de l'évolution (Kimura), l'évolution de la coopération et la théorie des jeux (Maynard-Smith), la génétique mathématique (Ewens), l'évolution moléculaire et la coalescence (Kingman), la dynamique adaptative...

Une des questions que pose l'étude de la biodiversité est de savoir comment, à partir d'une population homogène, peuvent émerger des sous-espèces génétiquement distinctes, voire deux espèces différentes. Cette divergence nécessite en effet que deux populations échangent suffisamment peu de gènes pour permettre l'accumulation des différences génétiques par mutation et sélection. La possibilité d'une île ou d'une barrière géographique en est une illustration limpide. Néanmoins, même si une telle barrière n'existe pas, la seule répartition spatiale des individus peut donner lieu à ce que Sewall Wright appelait « l'isolement par la distance ». Le problème de la **décroissance de l'apparentement avec la distance** a d'abord été étudié par Sewall Wright, puis par Gustave Malécot. En 1953, Motoo Kimura a posé à nouveau (puis a traité) ce problème, en faisant usage du modèle du jardin japonais (« stepping stone model »). Le but du sujet 2005 était de formaliser rigoureusement ces travaux et de mettre en évidence les principales propriétés du modèle à partir d'hypothèses minimales -il faut pour cela considérer que les effectifs considérés sont les effectifs des porteurs d'un certain allèle dans une population plus grande.

Dans la première partie, le problème est traité de façon déterministe par un système infini d'équations différentielles. Il s'agit de déterminer la densité de population $u_n(t)$

présente sur le territoire \mathcal{T}_n au temps t .

Le but de la section A est de montrer que lorsque les migrations se font **uniquement vers la droite** (à taux m), la population au temps t se répartit sur les territoires $(\mathcal{T}_n)_n$ suivant la **distribution de Poisson de paramètre mt** (propagation d'une onde).

Dans la section B, les migrations se font vers la gauche (taux m_g) **et** vers la droite (taux m_d), avec une barrière étanche en \mathcal{T}_0 . On montre qu'au voisinage de l'**équilibre dynamique**, la population se répartit suivant la **distribution géométrique de paramètre m_d/m_g** si $m_d < m_g$, et est localement éteinte sinon.

Dans la deuxième partie, les migrations locales bidirectionnelles de la première partie sont combinées à des migrations à longue distance, à taux m_∞ , vers des territoires choisis « uniformément au hasard ». Les guillemets sont dus au fait qu'il y a une infinité de territoires : rigoureusement, le taux d'immigration à longue distance est $m_\infty \bar{u}$, où \bar{u} est la limite lorsque $n \rightarrow \infty$ des moyennes de Césaro $\bar{u}_n = n^{-1} \sum_{k=0}^{n-1} u_k$, notée dans la suite MCL.

La section A est ainsi consacrée à l'étude des suites de Césaro (suites admettant une MCL). On montre que toute suite périodique est de Césaro (question 3), puis on exhibe **une suite de Césaro non bornée** (question 4) et **une suite bornée qui n'est pas de Césaro** (question 5). Enfin, il est demandé à la question 6 de démontrer le théorème de Césaro : **toute suite convergente est de Césaro, sa MCL égalant sa limite**.

La section B étudie la dynamique de la migration combinant migration locale et migration à longue distance. On montre d'abord que la MCL \bar{u} **reste constante en fonction du temps**, puis on étudie la différence $\tilde{u}_n := u_n - \bar{u}$ à l'équilibre. **Cette différence décroît géométriquement avec n** , la raison étant égale à $\mu_1 \in]0; 1[$, où

$$\mu_1 = \frac{1}{2} \left(1 + \lambda + x - \sqrt{(1 + \lambda + x)^2 - 4\lambda} \right),$$

avec les notations du sujet

$$\lambda = \frac{m_d}{m_g} \quad x = \frac{m_\infty}{m_g}.$$

Ceci prouve incidemment la **stabilisation spatiale** de la suite $(u_n)_n$, puisqu'à l'**équilibre dynamique elle converge vers $\bar{u}(t=0)$** lorsque $n \rightarrow \infty$. La question 11 consiste à calculer des développements limités de μ_1 lorsque les migrations à longue distance sont rares devant les migrations locales ($x \ll 1$). En particulier,

$$\mu_1 = \begin{cases} 1 - x/(\lambda - 1) + o(x) & \text{si } \lambda > 1 \\ 1 - \sqrt{x} + o(\sqrt{x}) & \text{si } \lambda = 1 \\ \lambda - \lambda x/(1 - \lambda) + o(x) & \text{si } \lambda < 1 \end{cases}$$

À la question 12, on montre que **la suite des densités de population reste au-dessus de sa limite si $m_d < m_g$ et en-dessous si $m_d > m_g$** (dans ce dernier cas les migrations à longue distance permettent néanmoins à la population de persister). En

effet, $u_n = \bar{u} + c_1 \mu_1^n$, avec

$$c_1 = \frac{2(1 - \lambda)\bar{u}}{\lambda - 1 + x + \sqrt{(\lambda - 1 - x)^2 + 4x}}.$$

La troisième et dernière partie traite le problème des migrations locales bidirectionnelles sous l'angle probabiliste : les déplacements individuels sont des **marches aléatoires indépendantes** (et de même loi), un pas à gauche (resp. à droite) se faisant avec probabilité p_g (resp. p_d).

Dans la section A, il est demandé de relever le paradoxe suivant ($p_d \neq p_g$) : si tous les territoires sont initialement peuplés, toute demie-droite le reste indéfiniment (question 2), alors qu'avec probabilité 1 toutes les trajectoires individuelles sont transientes (question 3). Il est ainsi permis de donner un sens aux résultats de la section suivante.

Dans la section B, on désigne par \mathbf{c}_k la **covariance à l'équilibre entre les effectifs de deux territoires distants de k** . Cette suite vérifie une relation de récurrence d'ordre 4 (question 6), et puisque $c_k = c_{-k}$ (question 5), l'inverse de toute racine du polynôme P associé est aussi racine de P (question 7). Comme de plus 1 est une racine évidente de P, il est possible de calculer toutes ses racines. En particulier, avec la notation

$$\sigma := \frac{(p_g + p_d)(1 - pg - pd)}{p_g p_d},$$

on montre que $P(X) = (X - 1)^2(X^2 + (\sigma + 2)X + 1)$. Les deux racines de P différentes de 1 sont donc négatives, ce que l'on interprète comme un phénomène de vases communicants : **deux populations adjacentes sont corrélées négativement** puisqu'une migration entre les deux provoque simultanément la croissance de l'une et la décroissance de l'autre. **La suite des covariances entre deux territoires distants de $k \geq 1$ est donc une suite géométrique alternée** de raison $\lambda_2 \in]-1; 0[$ (question 8). Lorsque $p_g = p_d = p$, alors

$$\lambda_2 = -\frac{1}{p} \left(1 - \sqrt{1 - p}\right)^2.$$

En général, les candidats 2005 savent très bien appliquer les théorèmes d'encadrement et manipuler les opérateurs d'espérance et de variance.

En revanche, ils ont beaucoup de difficulté avec les quantificateurs \exists, \forall et les expressions littérales associées. Il faudrait par exemple bien comprendre qu'il n'y a pas de différence entre une suite bornée et une suite « bornée à partir d'un certain rang »... La question II.A.6 (théorème de Cesàro), conformément au programme, a été divisée en sous-questions qui permettent de guider le candidat. De plus, il est manifeste (d'après certaines copies qui traitent la question incomplètement) que ce théorème est traité en exercice au cours de l'année. Néanmoins, il est sidérant de voir que seuls 3% des candidats ont eu le maximum des points à cette question et qu'une bonne moitié (des admissibles, faut-il le rappeler...) n'en a obtenu aucun.

On attendrait au concours des Écoles Normales que les candidats fissent usage d'une certaine finesse d'esprit. Ils devraient faire plus souvent preuve de sang-froid, et moins appel à leurs réflexes qu'à leur réflexion. Ainsi, sans même parler des quelque 15% de candidats qui pensent que toute suite bornée est convergente (ou qui préfèrent ne pas répondre à cette question II.A.1), on ne compte pas le nombre de ceux qui ont cru appuyer leur (bonne) réponse en donnant un exemple de suite bornée non convergente...

La mauvaise maîtrise des développements limités de Taylor est difficile à accepter, mais c'est une réalité : près de la moitié des étudiants qui ont pu obtenir (question II.B.9) l'expression de μ_1 , n'ont pas su correctement calculer un seul des DL demandés à la question II.B.11.

Il a été très apprécié quand les candidats interprétaient élégamment les résultats et les équations, notamment à l'aide de concepts empruntés à la physique (flux, conservation,...). Il est malheureusement fréquent qu'aucun soin ne soit apporté à la description littérale et à l'interprétation. Selon beaucoup de candidats, dans la première partie, les termes de la suite u_n sont des « nombres d'individus », ou, plus souvent, considérés comme inférieurs à 1 (ce sont des densités...), les divers coefficients m également (ce sont des taux...), et il est très rarement tenu compte du fait que m est un taux de migration *par unité de temps*.

Enfin, mentionnons qu'il est très imprudent de vouloir traiter linéairement le sujet question après question, dans l'ordre où elles sont posées. Une variante consiste bien sûr à aller chercher les questions que l'on se sent capable de résoudre le jour du concours (la technique dite du « grappillage »). Toutefois, il s'avère souvent plus productif de traiter exhaustivement et sans erreur un nombre réduit de questions difficiles, surtout cette année où le sujet présentait un grand nombre de questions élémentaires. Pour cette même raison, la meilleure stratégie cette année (c'est le cas notamment de la meilleure copie) consistait à répondre très succinctement aux questions faciles et à apporter soin et rigueur à la démonstration des questions difficiles. Un exemple intéressant est celui d'un(e) candidat(e) qui n'a pas traité la première partie (dont il était précisé qu'elle était la plus facile) et a obtenu une note finale de 14 sur 20. De manière générale, les questions les mieux pondérées étaient : II3, II4a, II5a, II11b, II12b, II12c, III2b, III6a, III8b, III8c. Ces questions ont été abordées par à peu près 4 ou 5 % des candidats.

À l'instar de mes prédécesseurs, je voudrais terminer en soulignant l'incohérence qu'il y a pour les Écoles Normales à réduire les mathématiques à une épreuve d'admission (à l'aide d'un argument comptable -on ne s'en étonne plus), à l'heure où les politiques scientifiques nationale et européenne destinent les mathématiques à un rôle grandissant dans le développement de la biologie contemporaine. Ce qui, bien sûr, est une tendance internationale (on pourra consulter par exemple les deux références ci-dessous).

Hastings, A., Palmer, M.A. 2003. A Bright Future for Biologists and Mathematicians? *Science*. **299** 2003-2004.

Cohen, J.E. 2004. Mathematics Is Biology's Next Microscope, Only Better ; Biology Is Mathematics' Next Physics, Only Better. *PLoS*. **2** (12). Disponible en ligne.