ÉPREUVE ÉCRITE de MATHÉMATIQUES – BCPST 2011

ENS : PARIS – LYON - CACHAN

Durée : 4 heures
Coefficients/Total général : Paris 16/145 ; Lyon 4/58,5 ; Cachan 4/65

MEMBRES du JURY : Nicolas CHAMPAGNAT

L'épreuve 2011 portait sur l'étude de la propagation d'épidémies dans des graphes aléatoires. L'approche de modélisation épidémiologique par des graphes a connu un développement récent notable, conjointement aux autres domaines d'applications concernés par les graphes aléatoires, comme par exemple les réseaux de communication. Nous renvoyons à la référence ci-dessous pour un récent panorama des travaux dans ce domaine. Dans le cas des applications en épidémiologie, le graphe représente un réseau de transmission d'un pathogène, et les questions biologiques qui se posent se traduisent en terme de propriétés des composantes connexes du graphe. La virulence du pathogène peut par exemple se mesurer par la taille de la plus grande composante connexe du graphe.

Le cadre le plus simple où les réponses à de telles questions ont pu être données est celui du graphe étudié par Erdös et Rényi à la fin des années 50, où chaque arête reliant une paire de sommets est présente dans le graphe avec une probabilité fixée, indépendamment des autres arêtes. C'est sur ce modèle que portait l'épreuve de cette année, dont l'objectif était de démontrer que, lorsque la taille du graphe tend vers l'infini, la plus grande composante du graphe est soit d'une taille de l'ordre de , soit de l'ordre de , selon que le nombre moyen de voisins d'un sommet dans le graphe est plus grand ou plus petit que 1. Le premier cas correspond à une épidémie qui touche une fraction importante de la population, le second à une épidémie qui touche un fraction négligeable de la population.

Le sujet était composé de 5 parties de difficulté croissante. La première portait sur une caractérisation algorithmique de la composante connexe d'un sommet donné, utilisée dans toute la suite de l'épreuve. La seconde partie portait sur une approximation déterministe de la taille de cette composante connexe, destinée à donner une idée des résultats étudiés dans la suite. La troisième partie rassemblait diverses propriétés des marches aléatoires apparaissant dans la caractérisation algorithmique de la première partie, utilisées dans les deux dernières parties. La quatrième étudiait la taille de la plus grande composante connexe du graphe lorsque le nombre moyen de voisins d'un sommet est plus petit que 1. Enfin, la cinquième partie s'intéressait au cas où le nombre moyen de voisins d'un sommet est plus grand que 1.

Ce dernier cas, où le nombre moyen de voisins d'un sommet est plus grand que 1, est le plus difficile à étudier, et a conduit à un sujet particulièrement long cette année. Comme nous l'avons déjà fait par le passé, nous rappelons que le barème de l'épreuve est adapté aux résultats des candidats, de telle sorte que l'étendue des notes permette de discriminer au mieux entre eux. Face à une épreuve longue, le candidat a particulièrement intérêt à lire attentivement l'ensemble du sujet pour choisir sur quelles questions ou parties se concentrer. Même si traiter le sujet dans l'ordre peut permettre au candidat de mieux en saisir la progression et de ne pas passer à côté de questions dont les résultats sont utiles dans la suite, la notation ne tient pas compte de l'ordre dans lequel les questions sont traitées, ni de l'absence de réponse donnée aux questions utilisées dans d'autres.

La dernière partie s'est avérée trop difficile pour les candidats cette année. Une version moins technique, ne montrant que des résultats partiels, aurait été plus appropriée. Seule une faible proportion (13%) des copies corrigées ont pu prendre quelques point sur cette partie, en répondant aux questions les plus faciles, presque uniquement 18)a) et b), et 20)a). Le reste du sujet a été très inégalement traité. La grande majorité des candidats a abordé avec un succès variable l'ensemble des parties 1 et 2, et la moitié des questions de la partie 3. Les questions plus difficiles de la partie 3, ainsi que la partie 4, ont servi à départager les meilleures copies.

La partie 1 a dans l'ensemble été bien traitée par les candidats. La principale source d'erreurs provient d'une mauvaise compréhension de l'algorithme décrit dans la question 1) : il y est bien spécifié qu'un seul sommet devient inactif à chaque étape, et non plusieurs. Le sujet étant clair sur ce point, nous ne pouvons mettre ces erreurs que sur le compte d'une lecture trop rapide ou pas assez précise de la question par les candidats. La question 3)c) pouvait poser problème si l'on cherchait à prouver le résultat de façon précise à l'aide de probabilités conditionnelles et des variable aléatoires . Aucun candidat n'a cherché à exprimer le résultat de cette façon, et nous n'attendions pas de réponse plus précise que celle consistant à compter le nombre de voisins neutres du sommet actif, et à dire que, conditionnellement à la valeur de ce nombre, le nombre de nouveaux sommets actifs est une somme de variables aléatoires de Bernoulli, donc suit une loi binomiale. Ce point du programme est globalement bien maîtrisé des candidats.

La partie 2 a également été traitée de façon globalement satisfaisante par les candidats, à l'exception des manipulations de développement limités et d'équivalents dans la question 7)b). Il est vrai que l'utilisation des logarithmes rendait cette question moins proches des calculs classique, mais le calcul pouvait se faire sans employer d'outil nouveau, en appliquant le développement limité à l'ordre 1 de l'exponentielle en 0 à la suite et en étudiant chacun des termes. Cette application des développements limités aurait dû être mieux traitée. Les questions 7)a) et c), bien que dans une moindre mesure, ont également été traitées de façon décevante par beaucoup de candidats : la première à cause d'erreurs de calcul ou de signe, et la seconde parce que beaucoup de candidats n'ont pas su faire le lien avec l'équivalent de et le développement limité de en 0. A cause de ces multiples erreurs, la question 8), qui éclairait l'ensemble du sujet, a été rarement complètement traitée. A noter que, dans la question 5), beaucoup de candidats n'ont pas saisi que la suite ne donne qu'une approximation de l'espérance de parce que la relation de récurrence sur dans la question 3)c) n'est valide que sur l'événement . On n'attendait pas plus de justification que le fait que la relation de récurrence sur aurait été exacte si l'égalité de la question 3)c) était vraie sans condition sur et . La question 4) faisait intervenir une interversion de sommation, qui n'a pas été indiquée dans le sujet par erreur.

La partie 3 contenait à la fois des questions faciles et des questions difficiles (particulièrement certaines sous-questions des questions 11) et 13)). Les calculs basés sur l'indépendance ou les preuves d'inégalités faisant intervenir la fonction exponentielle (questions 10), 13)a), 13)d)) ont était globalement correctement traités. Parmi les déceptions, un nombre beaucoup trop faible de candidats a vu le lien entre la loi des grands nombres et la question 11)a), le lien entre la question 11)c) et l'indépendance des variables aléatoires a été très rarement fait, et les calculs de variations de fonctions des questions 11)b) et 12)a) ont posé autant de difficultés que ceux de la question 7)a). Certains étudiants ne font apparemment pas la différence entre la loi des grands nombres et le théorème « central limit ». Les questions visant à établir la majoration de la probabilité d'un événement par une espérance, comme les questions 13)b) et e), n'ont presque pas été traitées et auraient sans doute mérité d'être plus détaillées dans le sujet.

La partie 4 n'était pas plus difficile que la partie précédente, mais elle reposait sur ce qui précédait dans le sujet. Ceci explique sans doute pourquoi si peu de candidats on su la traiter de façon satisfaisante. Si 41% des candidats ont obtenu des points sur cette partie, seuls 9% ont fait le lien, même partiellement, entre la question 15) et le problème initial de la taille de la plus grande composante connexe. L’expression de la variable aléatoire de la question 14) a été donnée par beaucoup de candidats, qui ont eu l’intuition de cette formule sans être capable de la justifier de façon mathématique rigoureuse. La moitié des points de cette question leur a été donnée. La question 15)a) avait un énoncé technique mais pouvait être résolue en vérifiant les propriétés de façon « mécanique », ce que beaucoup de candidats ont fait au moins partiellement. Il était plus difficile de faire le lien entre cette question et la suite. La question 16)b) nécessitait simplement de majorer la probabilité d’une réunion d’événements par la somme de leurs probabilités. Un seul candidat a résolu cette question, mais ceci est vraisemblablement à mettre sur le compte de la longueur du sujet. Certains candidats ont répondu correctement à la question 17), qui portait sur la propagation d’une épidémie dans le graphe, sans avoir répondu aux questions précédentes. Les points de la question leur ont été alloués. Globalement, cette partie destinée à départager les meilleurs candidats a joué ce rôle de façon satisfaisante. A cause de la longueur du sujet et de la difficulté des questions, la partie 5 n’a pas pu remplir ce même rôle.

Les notes vont de 1,75 à 20. La moyenne est de 8,87, le mode de 8,63 et l'écart-type de 3,52. 8% des copies ont une note inférieure à 4, 28% inférieure à 7, 23% supérieure à 11, 5% supérieure à 15.

Référence :

Durrett, R. *Random Graph Dynamics.* Cambridge University Press, 2007.