ENS LYON
L3 Probabilités

Probabilités, L3 Informatique Fondamentale

Examen final
26 mai 2025

Exercice 1 Méthode probabiliste

Soit $G = (V, E)$ un graphe (non orienté, sans boucle ni arête multiple). A l'aide de la méthode probabiliste, montrer qu'il existe une partition $V = V_1 \cup V_2$ telle que le nombre d'arêtes reliant un sommet de V_1 à un sommet de V_2 est $\geqslant |E|/2$.

On considère une partition aléatoire obtenue en plaçant chaque sommet dans V_1 ou V_2 avec probabilité $1/2$, indépendamment. La probabilité qu'une arête (x, y) soit transverse (c'est-à-dire qu'un des sommets est dans V_1 et l'autre dans V_2) est

$$\mathbf{P}(x \in V_1,\ y \in V_2) + \mathbf{P}(x \in V_2,\ y \in V_1) = \frac{1}{2} \times \frac{1}{2} + \frac{1}{2} \times \frac{1}{2} = \frac{1}{2}$$

Soit $E' \subset E$ le sous-ensemble des arêtes transverses. Par linéarité de l'espérance,

$$\mathbf{E}|E'| = \sum_{e \in E} \mathbf{P}(e \text{ est transvserse}) = \frac{|E|}{2}.$$

Il existe donc une partition telle que $|E|' \geqslant \frac{|E|}{2}$.

Exercice 2 Chiffre des unités de la somme de $n \gg 1$ dés

Soit (X_n) une suite de variables aléatoires i.i.d. de loi uniforme sur $\{1, \dots, 6\}$. On note Y_n la valeur de $X_1 + \dots + X_n$ modulo 10. A l'aide d'un théorème sur les chaînes de Markov, déterminer pour tout $k \in \{0, \dots, 9\}$ la valeur de

$$\lim_{n \to \infty} \mathbf{P}(Y_n = k).$$

La suite (Y_n) est une chaîne de Markov d'espace d'états $S = \mathbf{Z}/10\mathbf{Z}$ et de matrice de transition donnée par $Q(x, x+i) = \frac{1}{6}$ pour $x \in S$ et $i \in \{1, \dots, 6\}$, les autres termes étant nuls. Il est facile de voir que cette chaîne de Markov est irréductible, apériodique (par exemple car $Q^2(0,0)$ et $Q^3(0,0)$ sont non nuls) et de probabilité invariante la probabilité uniforme sur S. Par le théorème de convergence sur les chaînes de Markov, pour tout $k \in S$ on a $\lim_{n \to \infty} \mathbf{P}(Y_n = k) = \frac{1}{10}$.

Exercice 3 Valeurs extrêmes pour le temps de retour

Dans cet exercice les graphes sont supposés connexes, non orientés, sans boucle ni arête multiple. On fixe un entier $n \geqslant 1$.

Étant donné un graphe $G = (V, E)$ à n sommets et $x \in V$, on s'intéresse à la quantité $\mathbf{E}_x[T_x]$, le temps moyen de retour en x, pour la marche aléatoire sur le graphe G.

1. Quel choix de G et de x minimise la quantité $\mathbf{E}_x[T_x]$?
2. Quel choix de G et de x maximise la quantité $\mathbf{E}_x[T_x]$?

1. Pour la marche aléatoire issue de x, le temps de retour en x est $\geqslant 2$. La valeur 2 est atteinte avec probabilité 1 (donc $\mathbf{E}_x[T_x] = 2$) pour le graphe en étoile lorsqu'on part du sommet relié à tous les autres.
2. On utilise la formule du cours $\mathbf{E}_x[T_x] = \frac{2|E|}{k}$ où k est le degré de x. Sous la contrainte que x est de degré k, la quantité $|E|$ est maximale et vaut $k + \binom{n-1}{2}$ lorsque tous les sommets distincts de x sont reliés entre eux. On alors $\mathbf{E}_x[T_x] = 1 + \frac{\binom{n-1}{2}}{k}$, maximal pour $k = 1$. Le graphe recherché est le graphe complet à $n - 1$ sommet auquel on a ajouté un sommet supplémentaire de degré 1.

Exercice 4 Lecture

Vous décidez de lire les livres de la bibliothèque de l'ENS selon l'algorithme suivant : après avoir lu un livre, ce qui vous prend une semaine, vous lisez le suivant dans l'ordre alphabétique (sans revenir en arrière si jamais des livres ont été ajoutés entre temps). La bibliothèque achète un nombre fixe $k \in \mathbf{N}^*$ de livres par semaine. Le but de cet exercice est de calculer la probabilité de lire le dernier livre dans l'ordre alphabétique.

On modélise le problème en identifiant l'ensemble des livres à $[0,1]$ et en supposant que les achats sont indépendants et de loi uniforme. Soient $(Z_n^j)_{n\geqslant 1, 1\leqslant j\leqslant k}$ une famille de variables aléatoires i.i.d. de loi uniforme dans $[0,1]$. L'ensemble $L_n = \{Z_m^j : m \leqslant n, 1 \leqslant j \leqslant k\}$ décrit la bibliothèque à la $n$ème semaine. On définit par récurrence une suite (X_n) par $X_0 = 0$ et

$$X_{n+1} = \begin{cases} \text{le minimum de l'ensemble }]X_n, 1] \cap L_{n+1} & \text{si ce dernier est non vide} \\ 1 & \text{sinon} \end{cases}$$

La v.a. $X_n \in [0,1]$ correspond au livre lu à la $n$ème semaine. On veut calculer $\mathbf{P}(\exists n : X_n = 1)$, la probabilité d'avoir lu le dernier livre.

1. Montrer que si $Z_1, \dots, Z_N$ sont des variables aléatoires i.i.d. de loi uniforme dans $[0,1]$ alors $\mathbf{E}[\min(Z_1, \dots, Z_N)] = \frac{1}{N+1}$. On pourra utiliser le fait que si X est une variable aléatoire positive, alors $\mathbf{E}[X] = \int_0^\infty \mathbf{P}(X > t)\,\mathrm{d}t$.
   Avec la formule proposée, on a
$$\begin{aligned} \mathbf{E}[\min(Z_1, \dots, Z_N)] &= \int_0^1 \mathbf{P}(\min(Z_1, \dots, Z_N) > t)\mathrm{d}t \\ &= \int_0^1 \mathbf{P}(\forall i \in \{1, \dots, N\}, Z_i > t)\,\mathrm{d}t \\ &= \int_0^1 (1-t)^N\,\mathrm{d}t \\ &= \frac{1}{N+1} \end{aligned}$$

2. On pose $Y_n = 1 - X_n$. Montrer que pour tout $n \geqslant 1$, $\mathbf{E}[Y_n] \leqslant \left(1 - \frac{1}{kn+1}\right)\mathbf{E}[Y_{n-1}]$. Soient $a \leqslant b$ des réels. Les variables aléatoires $\tilde{Z}_i = a + (b-a)Z_i$ sont i.i.d. de loi unfirome sur $[a,b]$; par linéarité de l'espérance
$$\mathbf{E}[\min(\tilde{Z}_1, \dots, \tilde{Z}_N)] = a + \frac{b-a}{N+1}$$
   Fixons un entier n. Étant donnés $X_0, \dots, X_{n-1}$, soit $\tilde{L}_n$ un ensemble de kn v.a.i.i.d. de loi uniforme sur $[X_{n-1}, 1]$, indépendantes de $(X_1, \dots, X_{n-1})$; on note $\tilde{E}$ l'espérance par rapport à ces variables aléatoires. On peut coupler L_n et $\tilde{L}_n$ de sorte que $L_n \subset \tilde{L}_n$ et donc $\min \tilde{L}_n \leqslant X_n$. On a donc
$$\begin{aligned} \mathbf{E}[Y_n] &= 1 - \mathbf{E}[X_n] \\ &\leqslant 1 - \mathbf{E}\tilde{\mathbf{E}}[\min \tilde{L}_n] \\ &= 1 - \mathbf{E}\left[X_{n-1} + \frac{1 - X_{n-1}}{kn+1}\right] \\ &= \left(1 - \frac{1}{kn+1}\right)\mathbf{E}[Y_{n-1}] \end{aligned}$$

3. En déduire que $\lim_{n\to\infty} \mathbf{E}[Y_n] = 0$. On admet que, à l'aide du théorème de convergence dominée, cela implique que (Y_n) tend presque sûrement vers 0. On obtient par récurrence sur n que
$$\mathbf{E}[Y_n] \leqslant \prod_{j=1}^n \left(1 - \frac{1}{kj+1}\right)$$
   Avec l'inégalité $1 + x \leqslant \exp(x)$, cela donne
$$\log \mathbf{E}[Y_n] \leqslant -\sum_{j=1}^n \frac{1}{kj+1}.$$
   Comme cette série diverge, on a donc $\lim \mathbf{E}[Y_n] = 0$.

4. Soit $n \in \mathbf{N}$. On considère l'événement $A_n = \{Y_n < 1/k\}$. Quel théorème implique que, conditionnellement à A_n, on a presque sûrement $Y_m = 0$ pour m assez grand ? Fixons $\alpha < 1/k$. Soit $M_n = |\{j : Z_n^j \geqslant \alpha\}|$ le nombre de livres ajoutés dans l'intervalle $[1-\alpha, 1]$ à la $n$ème semaine. Les v.a. (M_n) sont i.i.d. de loi binomiale $\mathsf{B}(k, \alpha)$ et leur espérance vaut $\mathbf{E}[M_n] = k\alpha < 1$. Par la loi forte des grands nombres, on a presque sûrement

$$\lim_{n\to\infty} (M_1 + \cdots + M_n - n) = -\infty$$

Soit $T_n = |]X_{n-1}, 1] \cap L_n|$, le nombre de livres restants à lire dans la bibliothèque à la $n$ème semaine. Si $X_n > 1 - \alpha$ alors pour tout $m \geqslant n$

$$T_{m+1} \leqslant \max(0, T_m - 1 + M_m)$$

et donc $T_m = 0$ pour assez grand. On a donc montré que conditionnellement à $X_n > 1 - \alpha$, le dernier livre sera lu avec probabilité 1.

5. Conclure Par la question 3, presque sûrement il existe un n tel que $Y_n < \frac{1}{2k}$, et donc (par la question 4) $Y_m = 0$ pour m assez grand.

Exercice 5 Algorithme probabiliste pour 3-SAT

0. Rappeler le principe de l'algorithme probabiliste pour 2-SAT vu en cours (on demande uniquement les grandes lignes du raisonnement, pas de refaire tous les calculs en détail).

Cf cours

On va adapter cet algorithme au problème 3-SAT. On considère une formule 3-SAT en N variables booléennes $x_1, \ldots, x_N$, c'est-à-dire une conjonction de disjonctions (appelées clauses) dans lesquelles 3 variables exactement (ou leur négation) apparaissent. On suppose qu'une clause ne contient jamais deux fois la même variable, ni une variable et sa négation. Un exemple de telle formule est

$$(x_1 \vee \overline{x_2} \vee \overline{x_6}) \wedge (x_3 \vee \overline{x_1} \vee \overline{x_6}) \wedge (\overline{x_4} \vee x_2 \vee \overline{x_6}) \wedge (x_5 \vee \overline{x_2} \vee x_1)$$

On considère l'algorithme probabiliste suivant, où $m \in \mathbf{N}$ est un paramètre, afin de décider si une formule 3-SAT est satisfiable.

`RANDOM 3-SAT`

1 Initialiser avec une affectation des variables choisie uniformément au hasard.

2 Répéter m fois, en s'arrêtant si toutes les clauses sont satisfaites

  (a) Choisir arbitrairement une clause non satisfaite.

  (b) Choisir uniformément au hasard une des 3 variables de cette clause et la remplacer par sa négation.

3 Si une affectation valide a été trouvée, la retourner.
Sinon, répondre que la formule n'est pas satisfiable.

1. Supposons qu'il existe une affectation S des variables satisfaisant la formule. On note $X_n \in \{0, \ldots, N\}$ le nombre de variables ayant la même valeur que dans S au $n$ème tour de boucle de l'algorithme. Étant donnés $i \in \{0, \ldots, N-1\}$ et $j \in \{0, \ldots, N\}$, que pouvez-vous dire de

$$\mathbf{P}(X_{n+1} = j | X_n = i) \ ?$$

Tant que l'algorithme ne s'arrête pas, on a $X_{n+1} = X_n \pm 1$. Dans une clause non satisfaite, au moins une des 3 variables a une valeur différente de celle de S, et donc

$$\mathbf{P}(X_{n+1} = j | X_n = i) \begin{cases} \geqslant 1/3 & \text{si } j = i+1 \\ \leqslant 2/3 & \text{si } j = i-1 \\ = 0 & \text{si } |j-i| \neq 1 \end{cases}$$

2. On considère la chaîne de Markov $(Y_n)_{n \geqslant 0}$ d'espace d'états $\{0, \ldots, N\}$ donnée par la matrice de transition

$$Q(0,1) = Q(N,N) = 1$$
$$Q(i,i-1) = 2/3, \ Q(i,i+1) = 1/3 \qquad \text{si } 0 < i < N,$$

les autres coefficients étant nuls.

(a) Quels sont les états récurrents ou transitoires ? L'étant N est récurrent, les autres états sont stationnaires.

(b) On note $T_N = \inf\{n \geqslant 0 : Y_n = N\}$ le temps d'atteinte de l'état N. Pour $0 \leqslant j \leqslant N$, on note $h_j = \mathbf{E}_j[T_N]$ (où $\mathbf{E}_j$ dénote l'espérance pour la chaîne issue de j). Donner une relation de récurrence pour h_j. On a si $0 < j < N$,

$$h_j = 1 + \frac{2}{3} h_{j-1} + \frac{1}{3} h_{j+1}$$

ainsi que $h_N = 0$ et $h_0 = 1 + h_1$.

(c) En déduire que $h_j = 2^{N+2} - 2^{j+2} - 3(N-j)$. Un calcul direct (qu'il faut faire dans la copie) montre que ce choix est solution du système. Ce système linéaire à $N+1$ inconnues et $N+1$ équations admet une unique solution. On peut le vérifier par un calcul de déterminant ; sinon voici un argument élémentaire. Supposons que h'_j est une autre solution et posons $u_j = h_j - h'_j$. Alors on a $u_N = 0$, $u_0 = u_1$ et $u_j = \frac{2}{3} u_{j-1} + \frac{1}{3} u_{j+1}$. Soit M la valeur maximale de u_j pour $0 \leqslant j \leqslant N$. Si $u_{j_0} = M$, alors nécessairement $u_{j_0-1} = u_{j_0+1} = M$; en itérant ce raisonnement et on arrive à la conclusion que $M = 0$ et $u_j = 0$ pour tout j.

3. Comment faut-il choisir le paramètre m (en fonction de N) de l'algorithme `RANDOM 3-SAT` pour que la probabilité d'erreur soit inférieure à 2^{-100} ? La chaîne de Markov de la question précédente est une version pessimiste de l'algorithme : on peut définir un couplage de (X_n) et (Y_n) de sorte que $X_n \geqslant Y_n$. En particulier, si $S_N = \inf\{n \geqslant 0 : X_n = N\}$ alors $S_N \leqslant T_N$ et donc $\mathbf{E}[S_N] \leqslant \mathbf{E}[T_N] \leqslant \mathbf{E}_0[T_N] = h_0$ (puisque (h_j) est décroissante. On a donc $\mathbf{E}[S_N] \leqslant 2^{N+2}$ et par l'inégalité de MARKOV $\mathbf{P}(S_N \geqslant 2^{N+3}) \leqslant \frac{1}{2}$ et le choix $m = 8 \cdot 2^N$ garantit une probabilité d'erreur $\leqslant 1/2$ — cette analyse est correcte pour le pire choix d'initialisation dans l'étape 1 de l'algorithme (donc a fortiori pour une initialisation aléatoire). Par conséquent, en choisissant $m = 800 \cdot 2^N$ et en analysant ces m boucles comme 100 blocs de $8 \cdot 2^N$ boucles, on obtient une probabilité d'erreur $\leqslant 2^{-100}$. Cet algorithme n'a aucun intérêt dans la mesure où la recherche exhaustive prend un temps $O(2^N)$.
4. Pour $0 \leqslant j \leqslant N$, on pose $q(j) = \mathbf{P}_j(T_N \leqslant 3N)$. Montrer que

$$q(N-j) \geqslant \binom{3j}{j}\left(\frac{2}{3}\right)^j\left(\frac{1}{3}\right)^{2j} \geqslant \frac{1}{(3j+1)2^j}$$

Soient $(Z_1, \ldots, Z_{3j})$ des variables aléatoires de loi de BERNOULLI $\mathsf{B}(1/3)$. Si on définit $(Y_n)_{0\leqslant n\leqslant 3j}$ par $Y_0 = N-j$ puis en demandant que $Y_{n+1} = Y_n + 1$ si $Z_n = 1$ et que $Y_{n+1} = Y_n - 1$ si $Z_n = 0$ (sauf dans le cas où $Y_n = 0$, où on pose $Y_{n+1} = 1$ quelle que soit la valeur de Z_n), alors (Y_n) suit bien la loi de la chaîne de MARKOV issue de $N-j$, et de plus

$$\text{au moins } 2j \text{ parmi les v.a. } Z_1, \ldots, Z_{3j} \text{ valent } 1 \Longrightarrow T_N \leqslant 3j \Longrightarrow T_N \leqslant 3N$$

et donc

$$q(N-j) \geqslant \mathbf{P}(2j \text{ parmi les v.a. } Z_1, \ldots, Z_{3j} \text{ valent } 1) = \binom{3j}{j}\left(\frac{2}{3}\right)^j\left(\frac{1}{3}\right)^{2j}$$

On utilise l'inégalité (où h est la fonction d'entropie binaire)

$$\binom{3j}{j} \geqslant \frac{2^{3jh(1/3)}}{3j+1}$$

comme $2^{h(1/3)} = 3^{1/3} \cdot \left(\frac{3}{2}\right)^{2/3}$, on a

$$2^{3jh(1/3)} = \left(\frac{27}{4}\right)^j$$

d'où le résultat voulu.
5. Montrer que si $m = 3N$, la probabilité que l'algorithme `RANDOM 3-SAT` (appelé sur une formule satisfiable) trouve une affectation valide des variables est $\geqslant \frac{1}{3N+1}\left(\frac{3}{4}\right)^N$. Comme à la question 3, on utilise l'inégalité $S_N \leqslant T_N$. Par l'étape 1 de l'algorithme, l'état initial de la chaîne (Y_n) suit la loi binomiale $\mathsf{B}(N, 1/2)$. On a donc

$$\begin{aligned}\mathbf{P}(S_N \leqslant 3N) &\geqslant \mathbf{P}(T_N \leqslant 3N) \\ &\geqslant \sum_{j=0}^{N} q(j)2^{-N}\binom{N}{j} \\ &\geqslant \frac{1}{2^N(3N+1)}\sum_{j=0}^{N} 2^{-j}\binom{N}{j} \\ &= \frac{1}{2^N(3N+1)}\left(1+\frac{1}{2}\right)^N\end{aligned}$$

d'où le résultat

6. En déduire un algorithme probabiliste pour 3-SAT de meilleure complexité que celle obtenue à la question 3. On lance M fois indépendemment l'algorithme pour la valeur $m = 3N$. La probabilité d'erreur lorsque l'algorithme est appelé sur une formule satisfiable est

$$p_{\text{erreur}} \leqslant \left(1 - \frac{1}{3N+1}(3/4)^N\right)^M \leqslant \exp\left(-\frac{M}{3N+1}(3/4)^N\right).$$

   En choisissant $M = 100(4/3)^N * (3N+1)$, la probabilité d'erreur est $\leqslant \exp(-100)$. La complexité, bien qu'exponentielle, est meilleure que celle de la question 3.

7. Comment cela ce généralise-t-il au problème k-SAT pour $k > 3$? Cf l'article d'origine `https://homepages.cwi.nl/~rdewolf/schoning99.pdf`