Échauffement : deux algorithmes probabilistes

Les meilleurs algorithmes probabilistes (connus) sont souvent plus simples et/ou plus efficaces que les meilleurs algorithmes déterministes (connus). On va illustrer ce principe sur deux exemples, en utilisant le langage probabiliste («indépendance», «probabilité conditionnelle») qui sera introduit rigoureusement dans le prochain chapitre.

0.1 Vérifier la multiplication matricielle

Soient A, B, C trois matrices $n \times n$ à coefficients dans le corps $\mathbf{F}_2 = \{0, 1\}$. Le problème est de déterminer si l'équation $AB = C$ est vraie ou fausse.

Une première idée est de calculer le produit $A \cdot B$ et de vérifier si les coefficients sont les mêmes que ceux de C. L'algorithme naïf qui utilise la formule

$$(AB)_{ij} = \sum_k A_{ik} B_{kj}$$

a une complexité $\Theta(n^3)$. Des algorithmes plus sophistiqués basés sur une idée de Strassen améliorent la complexité en $\Theta(n^\alpha)$ pour $2 < \alpha < 3$ (le record actuel est $\alpha \approx 2,37$ et on conjecture que la valeur optimale est $\alpha = 2$).

Une autre idée est de vérifier la formule à travers le prisme probabiliste, c'est-à-dire de vérifier si l'équation

$$ABx = Cx$$

est satisfaite pour un vecteur $x \in \mathbf{F}_2^n$ choisi au hasard. Une telle vérification s'effectue en $\Theta(n^2)$, qui est clairement la complexité optimale de la multiplication matrice $\times$ vecteur. La clé est le lemme suivant.

Lemme. *Soit $D \in \mathsf{M}_n(\mathbf{F}_2)$ une matrice non nulle et $x \in \mathbf{F}_2^n$ choisi uniformément au hasard. Alors*

$$\mathbf{P}(Dx \neq 0) \geqslant 1/2.$$

Démonstration. Il existe un coefficient non nul dans la matrice D ; sans perte de généralité supposons que c'est le coefficient D_{1n}. On a alors

$$(Dx)_1 = \sum_{j=1}^{n} d_{1j} x_j = \sum_{j=1}^{n-1} d_{1j} x_j + x_n.$$

On remarque alors que quels que soient $(x_1, \ldots, x_{n-1})$ fixés, il y a probabilité $\frac{1}{2}$ (sur le choix de x_n) que $(Dx)_1 \neq 0$. $\square$

Comme conséquence du lemme, on a le résultat suivant : si $AB \neq C$ et si $x \in \mathbf{F}_2^n$ est choisi au hasard, alors

$$\mathbf{P}(ABx = Cx) \leqslant \frac{1}{2}.$$

Si on répète 100 fois cette vérification pour des vecteurs $x_1, \ldots, x_{100}$ choisis indépendemment, on a

$$\mathbf{P}(ABx_i = Cx_i \text{ pour tout } i) \leqslant 2^{-100} = 0 \text{ en pratique}$$

et on obtient donc un algorithme probabiliste qui permet de vérifier la multiplication matricielle en temps $\Theta(n^2)$.

Cet argument repose implicitement sur le concept *d'indépendance* que l'on étudiera formellement plus tard.

0.2 Coupe minimale dans un graphe

Soit $G = (V, E)$ un graphe non orienté sans boucle, ayant possiblement des arêtes multiples. On pose $n = |V|$.

Une *coupe* de G est un sous-ensemble $C \subset E$ tel que $(V, E \setminus C)$ n'est pas connexe. Le problème est de déterminer le cardinal minimal d'une coupe de G, que l'on note $\mathrm{mincut}(G)$. Autrement dit, on cherche une partition $V = V_1 \cup V_2$ (avec V_1 et V_2 non vides) qui minimise le nombre d'arêtes joignant un sommet de V_1 à un sommet de V_2. Il existe des algorithmes déterministes efficaces pour résoudre ce problème. Mais il y a plus simple : l'algorithme probabiliste de Karger (1993).

L'algorithme de Karger repose sur la notion de contraction d'un graphe selon une arête. Étant donnée une arête $e = \{x, y\} \in E$, la contraction de G selon e, notée G/e, est le graphe obtenu en identifiant les sommets x et y (pour obtenir un nouveau sommet noté xy), en remplaçant les arêtes $\{x, z\}$ ou $\{y, z\}$ par $\{xy, z\}$ et en effaçant les boucles éventuellement créées. Une contraction d'un graphe à n sommets peut être implémentée en temps $O(n)$, par exemple en représentant le graphe par sa matrice d'adjacence.

Figure 1 – Un graphe G (à gauche) et sa contraction G/e pour $e = \{a, b\}$ (à droite)

Pour toute arête e de G, on a $\mathrm{mincut}(G) \leqslant \mathrm{mincut}(G/e)$ puisque les coupes de G/e correspondent aux coupes de G qui n'utilisent pas l'arête e. L'algorithme de Karger consiste à effectuer des contractions au hasard.

Algorithme (Algorithme de Karger)**.** Tant que G contient > 2 sommets, répéter la procédure suivante : choisir uniformément au hasard une arête e de G et remplacer G par G/e. On obtient ainsi un graphe à 2 sommets qui correspond à une partition $V = V_1 \cup V_2$ et donc à une coupe du graphe initial.

Il est important de conserver les arêtes multiples : par exemple, si on l'applique l'algorithme au graphe qui est à droite de la figure 1, l'arête $\{ab, d\}$ est contractée avec probabilité $\frac{1}{5}$ puisque le graphe comprend 5 arêtes.

Dans la description de l'algorithme donnée ci-dessus, on considère implicitement que les différents choix aléatoires effectués par l'algorithme sont indépendants. Cette remarque vaut pour tous les algorithmes probabilistes étudiés dans ce cours.

Il est clair que l'algorithme termine puisque le nombre de sommets diminue de 1 à chaque étape. Le lemme-clé est le suivant.

Lemme. *La coupe C produite par l'algorithme de* Karger *vérifie*

$$\mathbf{P}(|C| = \mathrm{mincut}(G)) \geqslant \frac{2}{n^2}.$$

Si on répète $N = 50n^2$ fois cet algorithme (tous les choix étant indépendants), et si on note k_i la coupe obtenue à la $i$ème exécution de l'algorithme, alors

$$\begin{aligned}
\mathbf{P}\left(\min_{1\leqslant i\leqslant N} k_i \neq \mathrm{mincut}(G)\right) &\leqslant \left(1-\frac{2}{n^2}\right)^N \\
&\leqslant \exp\left(-\frac{2N}{n^2}\right) \\
&= \exp(-100) \approx 0.
\end{aligned}$$

On a donc un algorithme probabiliste de complexité $O(n^2T)$ pour trouver la coupe minimale d'un graphe, où $T = O(n^2)$ est la complexité d'une itération.

Preuve du lemme. Soit $k = \mathrm{mincut}(G)$ et C une coupe de taille k. Pour $1 \leqslant i \leqslant n-2$, considérons les événements

$$A_i = \text{« l'arête choisie à la } i\text{ème étape est dans } C \text{ »}$$

et soit B_i l'événement complémentaire de A_i. On a $\mathbf{P}(A_1) = \frac{k}{|E|}$. Mais tout sommet a degré $\geqslant k$ et donc $|E| \geqslant \frac{kn}{2}$; on a donc $\mathbf{P}(A_1) \leqslant \frac{2}{n}$.

Conditionnellement à B_1, le graphe obtenu après contraction de la première arête a aussi une coupe minimale égale à k. Ce graphe a $n-1$ sommets et on a donc par le même argument

$$\mathbf{P}(A_2|B_1) \leqslant \frac{2}{n-1}.$$

De la même manière, on a

$$\mathbf{P}(A_3|B_1 \cap B_2) \leqslant \frac{2}{n-2}$$

$$\vdots$$

$$\mathbf{P}(A_{n-2}|B_1 \cap B_2 \cap \ldots B_{n-3}) \leqslant \frac{2}{3}$$

On a donc

$$\begin{aligned}
\mathbf{P}(B_1 \cap B_2 \cap \cdots \cap B_{n-2}) &= \mathbf{P}(B_1)\mathbf{P}(B_2|B_1)\mathbf{P}(B_3|B_1 \cap B_2)\ldots\mathbf{P}(B_{n-2}|B_1 \cap \cdots \cap B_{n-3}) \\
&\geqslant \left(1-\frac{2}{n}\right)\left(1-\frac{2}{n-1}\right)\cdots\left(1-\frac{2}{3}\right) \\
&= \frac{2}{n(n-1)} \\
&\geqslant \frac{2}{n^2}
\end{aligned}$$

Lorsque les événements A_1, A_2, …, A_n sont réalisés, la coupe produite par l'algorithme de Karger est la coupe C. Ceci conclut la preuve du lemme. □

Chapitre 1

Événements, probabilités, variables aléatoires

1.1 Espaces de probabilité

Définition. Un *espace de probabilité* est la donnée de
— un ensemble Ω,
— une famille $\mathcal{F}$ de parties de Ω (c'est-à-dire $\mathcal{F} \subset \mathcal{P}(\Omega)$), l'ensemble des *événements*,
— une fonction $\mathbf{P} : \mathcal{F} \to [0,1]$ qui a un événement associe sa *probabilité*,
qui vérifie les axiomes suivants :

1. La famille $\mathcal{F}$ est une *tribu* (en anglais : σ-algebra), c'est-à-dire telle que
   — Ω est un événement,
   — Si A est un événement, alors $\Omega \setminus A$ est événement,
   — si $(A_n)_{n\in\mathbf{N}}$ est une suite d'événements, alors $\bigcup A_n$ est un événement.
2. $\mathbf{P}$ est une *mesure de probabilité*, c'est-à-dire que
   — on a $\mathbf{P}(\Omega) = 1$ et $\mathbf{P}(\emptyset) = 0$,
   — si $(A_n)_{n\in\mathbf{N}}$ est une suite d'événements deux à deux disjoints (c'est à dire que $A_m \cap A_n = \emptyset$ si $m \neq n$), alors
   $$\mathbf{P}\left(\bigcup_{n\in\mathbf{N}} A_n\right) = \sum_{n\in\mathbf{N}} \mathbf{P}(A_n).$$
   Cette propriété s'appelle la *σ-additivité*.

Dans tout le cours, on suppose donné un espace de probabilité $(\Omega, \mathcal{F}, \mathbf{P})$.

Exemple. Si Ω est un ensemble fini, on peut prendre $\mathcal{F} = \mathcal{P}(\Omega)$ et définir pour $A \subset \Omega$
$$\mathbf{P}(A) = \frac{|A|}{|\Omega|}.$$
On dit que $\mathbf{P}$ est la probabilité uniforme sur Ω.

Exemple (généralise le précédent). Si Ω est un ensemble fini ou dénombrable et si $(p_\omega)_{\omega\in\Omega}$ est une famille de réels $\geqslant 0$ vérifiant $\sum p_\omega = 1$, on peut prendre $\mathcal{F} = \mathcal{P}(\Omega)$ et définir pour $A \subset \Omega$
$$\mathbf{P}(A) = \sum_{\omega\in A} p_\omega.$$
Un espace de probabilité de ce type est appelé un espace de probabilité discret.

Remarquons que si A et B sont des événements tels que $A \subset B$, alors $\mathbf{P}(A) \leqslant \mathbf{P}(B)$. En effet, par σ-additivité (appliquée à une suite d'événements dont tous sauf deux sont vides) on a $\mathbf{P}(B) = \mathbf{P}(A) + \mathbf{P}(B \setminus A) \geqslant \mathbf{P}(A)$. Le lemme suivant est à la fois trivial et fondamental.

Lemme (Borne de l'union)**.** *Si (A_n) est une suite finie ou dénombrable d'événements, alors*

$$\mathbf{P}\left(\bigcup_n A_n\right) \leqslant \sum_n \mathbf{P}(A_n).$$

Démonstration. On définit $B_n = A_n \setminus \bigcup_{k<n} A_k$. On a alors $B_n \subset A_n$ et $\bigcup B_n = \bigcup A_n$. Puisque les événements B_n sont deux à deux disjoints, on a par σ-additivité,

$$\mathbf{P}\left(\bigcup_n A_n\right) = \mathbf{P}\left(\bigcup_n B_n\right) = \sum_n \mathbf{P}(B_n) \leqslant \sum_n \mathbf{P}(A_n)$$

d'où le résultat. □

Une question naturelle : pourquoi ne pas toujours prendre $\mathcal{F} = \mathcal{P}(\Omega)$? Quel intérêt y a-t-il a exclure des parties de l'ensemble des événements ? Il y a deux raisons sur lesquelles on reviendra

— il y a des cas où on ne peut pas, pour des raisons liées à l'infini.
— même dans le cas discret, il y a parfois intérêt à considérer plusieurs tribus différentes.

1.2 Événements

Définition. Deux événements A et B sont *indépendants* ($A \perp\!\!\!\perp B$) si

$$\mathbf{P}(A \cap B) = \mathbf{P}(A)\mathbf{P}(B).$$

Si $\mathbf{P}(B) > 0$, la probabilité conditionnelle de A sachant B est définie par $\mathbf{P}(A|B) = \mathbf{P}(A \cap B)/\mathbf{P}(B)$. On a donc

$$A \perp\!\!\!\perp B \iff \mathbf{P}(A|B) = \mathbf{P}(A)$$

et donc la probabilité de A «ne dépend pas» de B. Voila un autre lemme trivial.

Lemme. *Soit (A_n) une partition finie ou dénombrable de Ω en événements tels que $\mathbf{P}(A_n) > 0$ pour tout n. Alors pour tout événement B*

$$\mathbf{P}(B) = \sum_n \mathbf{P}(B \cap A_n) = \sum_n \mathbf{P}(B|A_n)\mathbf{P}(A_n).$$

Définition. Soit (A_n) une famille finie ou infinie d'événements. On dit que les événements (A_n) sont *indépendants* si pour tout ensemble I fini, on a

$$\mathbf{P}\left(\bigcap_{i \in I} A_i\right) = \prod_{i \in I} \mathbf{P}(A_i).$$

Attention : soient trois événements A_1, A_2 , A_3. On a l'implication

$$A_1, A_2, A_3 \text{ indépendants } \Longrightarrow \mathbf{P}(A_1 \cap A_2 \cap A_3) = \mathbf{P}(A_1)\mathbf{P}(A_2)\mathbf{P}(A_3)$$

mais la réciproque est fausse en général, comme on s'en convainc en considérant par exemple $A_3 = \emptyset$. De même, si (A_n) sont des événements, alors

$$(A_n) \text{ indépendants } \Longrightarrow (A_n) \text{ 2 à 2 indépendants}$$

et la réciproque est fausse en général.

Exercice. Montrer que des événements (A_n) sont indépendants si et seulement si les événements $(\Omega \setminus A_n)$ sont indépendants.

Exercice. L'indépendance de n événements requiert de vérifier 2^n équations. Donner, pour tout n, un exemple où toutes ces équations sont vérifiées sauf une.

Fin cours # 1 du 12 septembre