Probabilités

Basé sur le cours de Guillaume Aubrin
Notes prises par Hugo Salou

29 janvier 2025

Table des matières

0 Échauffement : deux algorithmes probabilistes.

Exemple 0.1 (Vérifier la multiplication de matrices). Soient A, B, C trois matrices carrées à coefficients dans $\mathbb{F}_2 = \{0, 1\}$. On cherche à décider $AB = C$.

Idée 1. On calcule AB et on vérifie l'égalité à C. L'algorithme pour calculer AB avec $(AB)_{i,j} = \sum_{k=1}^{n} A_{i,k}B_{k,j}$ se fait avec une complexité en $\mathrm{O}(n^3)$.

On peut améliorer la complexité en $\mathrm{O}(n^\alpha)$ avec $2 < \alpha < 3$ (actuellement, on peut le faire avec $\alpha \approx 2{,}37$) à l'aide de la méthode de Strassen.

Idée 2. On calcule ABx et Cx pour un vecteur $x \in \mathbb{F}_2^n$. On a des multiplications matrices $\times$ vecteurs, en complexité en $\mathrm{O}(n^2)$. Pour trouver un « bon » vecteur x, on le choisit au hasard.

Lemme 0.1. Si $D \in \mathscr{M}_n(\mathbb{F}_2)$ est non-nulle et $x \in \mathbb{F}_2^n$ est choisi uniformément au hasard, alors on a $\mathrm{P}(Dx \neq 0) \geq \frac{1}{2}$.

Preuve. Au moins un coefficient de D est non-nul et, sans perte de généralité, on peut supposer que $D_{1,n} \neq 0$. Alors,

$$(Dx)_1 = \sum_{i=1}^{n} D_{1,i}x_i = \sum_{i=1}^{n-1} D_{1,i}x_i + x_1.$$

Quels que soient $x_1, \ldots, x_{n-1}$, il y a une probabilité de $\frac{1}{2}$ que

$(Dx)_1 \neq 0$. On en conclut que

$$\mathrm{P}(Dx \neq 0) \geq \mathrm{P}((Dx)_1 \neq 0) = \frac{1}{2}.$$

□

Exemple 0.2 (suite de 0.1). Ainsi, si $AB \neq C$, on a donc

$$\mathrm{P}(ABx \neq Cx) \geq \frac{1}{2}.$$

On choisit $x_1, \ldots, x_{100}$ des vecteurs uniformément dans $\mathbb{F}_2^n$. Si on a $AB \neq C$, alors

$$\mathrm{P}(\forall i \in [\![1, 100]\!], ABx_i = Cx_i) \leq \left(\frac{1}{2}\right)^{100}.$$

On a donc un algorithme ayant une complexité $\mathrm{O}(n^2)$ pour détecter, avec grande probabilité, si $AB = C$.

Exemple 0.3 (Coupe minimale dans un graphe). On considère G un graphe non-orienté sans boucle (éventuellement avec des arêtes multiples). Une *coupe* du graphe est un sous-ensemble $C \subseteq E$ tel que $(V, E \setminus C)$ n'est pas connexe. On cherche une coupe de taille minimale :

$$\mathsf{mincut}(G) = \min\{|C| \mid C \text{ est une coupe}\}.$$

De manière équivalente, on cherche une partition $V = V_1 \sqcup V_2$ (avec $V_1, V_2 \neq \emptyset$) qui minimise le nombre d'arêtes reliant V_1 et V_2.

Étant donné un graphe $G = (V, E)$, et une arête $e = \{x, y\} \in E$, la *contraction de G selon e*, notée G/e, est le graphe où les sommets x et y sont fusionnés en un sommet xy, et les arêtes $\{x, z\}$ ou $\{y, z\}$ sont remplacées en $\{xy, z\}$ si $z \notin \{x, y\}$.

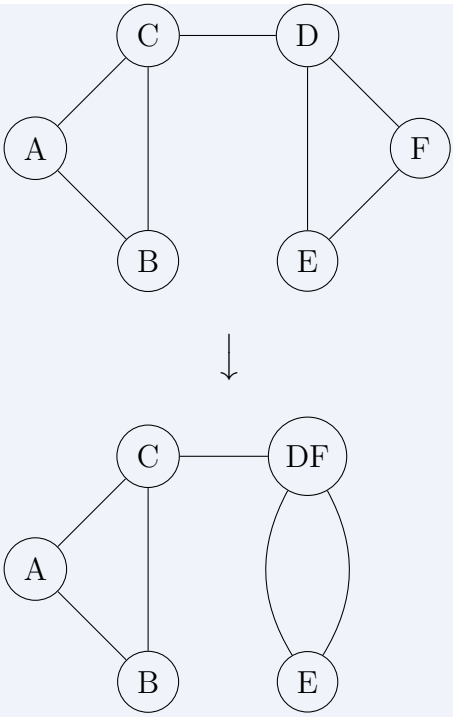

Figure 1 | *Contraction de l'arête* $\{\mathrm{D}, \mathrm{F}\}$.

On a que $\mathsf{mincut}(G/e) \geq \mathsf{mincut}(G)$.

On utilise l'*algorithme de Krager* (1993). On contracte successivement selon des arêtes choisies uniformément au hasard, jusqu'à n'obtenir que 2 sommets, ce qui donne une coupe du graphe initial.

Lemme 0.2. La coupe C produite par l'algorithme de vérifie

$$\mathrm{P}(|C| = \mathsf{mincut}(G)) \geq \frac{2}{n^2},$$

où $n = |V|$. □

Preuve. Soit $k = \mathsf{mincut}(G)$ et C une coupe de taille k. Montrons que $\mathrm{P}(\text{l'algorithme renvoie la coupe } C) \geq 2/n^2$. Notons A_i (pour $i \in [\![1, n-2]\!]$) l'événement « l'arête contractée à la i-ème étape est dans C », et B_i l'événement complémentaire. L'algorithme renvoie la coupe C si et seulement si tous les événements $B_1, \dots, B_{n-2}$ sont vérifiés. On a $\mathrm{P}(A_1) = k/|E| \leq 2/n$. Or, tout sommet a un degré $\geq k$, et donc $|E| \geq nk/2$. Conditionnellement à B_{11}, le graphe obtenu après contraction de la première arête vérifie $\mathsf{mincut}(G/e) = k$ donc $\mathrm{P}(A_2 \mid B_1) \leq 2/(n-1)$. De même, $\mathrm{P}(A_j \mid B_1 \cap \cdots \cap B_{j-1}) \leq 2/(n+1-j)$, pour tout $j \in [\![1, n-2]\!]$. On a donc $\mathrm{P}(A_{n-2} \mid B_1 \cap B_{n-2}) \leq \frac{2}{3}$, et donc

$$\begin{aligned}
\mathrm{P}(B_1 \cap \cdots \cap B_{n-2}) &= \mathrm{P}(B_1)\mathrm{P}(B_2 \mid B_1) \cdots \mathrm{P}(B_{n-2} \mid B_1 \cap \cdots \cap B_{n-1}) \\
&\geq \left(1 - \frac{2}{n}\right)\left(1 - \frac{2}{n-1}\right) \cdots \left(1 - \frac{2}{3}\right) \\
&\geq \frac{n-2}{n}\frac{n-3}{n-1} \times \cdots \times \frac{2}{3} \\
&\geq \frac{2}{n(n-1)} \geq \frac{2}{n^2}.
\end{aligned}$$

□

Exemple 0.4 (suite de 0.3). On répète $N = 50n^2$ fois cet algorithme (tous les choix étant indépendant). On note k_i la taille de la coupe obtenue à la i-ème itération, et alors

$$\mathrm{P}(k_i = \mathsf{mincut}(G)) \geq \frac{2}{n^2},$$

d'où $\mathrm{P}(k_i \neq \mathsf{mincut}(G)) \leq 1 - \frac{2}{n^2}$.

On en conclut que

$$\begin{aligned}\mathrm{P}(\forall i, k_i \neq \mathsf{mincut}(G)) &\leq \left(1 - \frac{2}{n^2}\right)^{50n^2} \\ &\leq \exp\left(-\frac{2}{n^2}50n^2\right) \\ &\leq \exp(-100).\end{aligned}$$

Chaque itération prend un temps en $\mathrm{O}(n^2)$, on obtient donc un algorithme en $\mathrm{O}(n^4)$ qui calcule une coupe minimale avec très grande probabilité.

1 Événements, probabilités, variables aléatoires.

1.1 Espaces de probabilités.

Définition 1.1. Un *espace de probabilité* est la donnée de

- ▷ un ensemble Ω ;
- ▷ un ensemble $\mathcal{F} \subseteq \wp(\Omega)$ de parties de Ω, appelées *événements* ;
- ▷ une fonction $\mathrm{P} : \mathcal{F} \to [0,1]$ qui associe à un événement sa probabilité ;

qui vérifie les axiomes suivants

1. l'ensemble $\mathcal{F}$ est une *tribu* (ou *σ-algèbre*) :
   - ▷ $\Omega \in \mathcal{F}$;
   - ▷ si $A \in \mathcal{F}$ alors $\Omega \setminus A \in \mathcal{F}$;
   - ▷ si $(A_n)_{n\in\mathbb{N}}$ est dans $\mathcal{F}$ alors $\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{F}$;
2. l'application P est une *mesure de probabilité* :
   - ▷ $\mathrm{P}(\Omega) = 1$;
   - ▷ $\mathrm{P}(\emptyset) = 0$;
   - ▷ [*σ-additivité*] si $(A_n)_{n\in\mathbb{N}}$ sont des événements disjoints (*i.e.* $A_n \cap A_m = \emptyset$ si $n \neq m$) alors

$$\mathrm{P}\Big(\bigcup_{n\in\mathbb{N}} A_n\Big) = \sum_{n\in\mathbb{N}} \mathrm{P}(A_n).$$

On supposera donné un espace de probabilité $(\Omega, \mathcal{F}, \mathrm{P})$.

Exemple 1.1. Si Ω est un ensemble fini, on peut choisir $\mathcal{F} = \wp(\Omega)$ et $\mathrm{P}(A) = |A|ă/|\Omega|$. On dit que P est la *probabilité uniforme* sur Ω.

Exemple 1.2. Si Ω est fini ou dénombrable et si $(p_\omega)_{\omega\in\Omega}$ sont des réels positifs tels que $\sum_{\omega\in\Omega} p_\omega = 1$, on peut prendre $\mathcal{F} = \wp(\Omega)$ et poser $\mathrm{P}(A) = \sum_{\omega\in A} p_\omega$. On a alors défini une probabilité à partir de $p_\omega = \mathrm{P}(\{\omega\}) = p_\omega$.

Si A et B sont deux événements avec $A \subseteq B$ alors $\mathrm{P}(A) \leq \mathrm{P}(B)$. En effet, il suffit d'écrire $\mathrm{P}(B) = \mathrm{P}(A) + \mathrm{P}(B \setminus A)$.

Lemme 1.1 (Borne de l'union)**.** Si $(A_n)_{n\in I}$ est une famille finie ou dénombrable d'événements, alors

$$\mathrm{P}\Big(\bigcup_{n\in I} A_n\Big) \leq \sum_{n\in I} \mathrm{P}(A_n).$$

Preuve. On pose $B_n = A_n \setminus \big(\bigcup_{k<n} A_k\big)$. Les (B_n) sont disjoints, et $\bigcup_{n\in I} A_n = \bigcup_{n\in I} B_n$. On a donc

$$\mathrm{P}\Big(\bigcup_{n\in I} A_n\Big) = \mathrm{P}\Big(\bigcup_{n\in I} B_n\Big) = \sum_{n\in I} \mathrm{P}(B_n) \leq \sum_{n\in I} \mathrm{P}(A_n).$$

□

Une question naturelle est : pourquoi ne pas prendre toujours $\mathcal{F} = \wp(\Omega)$?

- Il y a des cas où on ne peut pas, pour des raisons liées à l'infini (en particulier dans le cas non dénombrable).
- Même dans le cas discret, on a parfois intérêt à considérer plusieurs tribus.

1.2 Indépendance.

Définition 1.2. Deux événements A et B sont *indépendants*, noté $A \perp\!\!\!\perp B$, si $\mathrm{P}(A \cap B) = \mathrm{P}(A) \times \mathrm{P}(B)$.

Définition 1.3. Si $\mathrm{P}(B) > 0$, la *probabilité de A selon B* est la probabilité $\mathrm{P}(A \mid B) = \mathrm{P}(A \cap B)/\mathrm{P}(B)$. On a donc $A \perp\!\!\!\perp B \iff \mathrm{P}(A \mid B) = \mathrm{P}(A)$.

Lemme 1.2. Si (A_n) est une partition fini ou dénombrable de Ω en événements et B un événement,

$$\mathrm{P}(B) = \sum_n \mathrm{P}(B \cap A_n) = \sum_n \mathrm{P}(B \mid A_n) \cdot \mathrm{P}(A_n).$$

□

Définition 1.4. Si (A_i) est une famille finie ou infinie d'événements, on dit qu'ils sont *indépendants* si, pour tout $J \subseteq I$ non-vide,

$$\mathrm{P}\Big(\bigcap_{i \in J} A_i\Big) = \prod_{i \in J} \mathrm{P}(A_i).$$

Exemple 1.3. On a que (A, B, C) sont indépendants si et seulement si les quatre conditions sont vérifiées :

- $\mathrm{P}(A \cap B \cap C) = \mathrm{P}(A) \cdot \mathrm{P}(B) \cdot \mathrm{P}(C)$;
- $\mathrm{P}(A \cap B) = \mathrm{P}(A) \cdot \mathrm{P}(B)$;
- $\mathrm{P}(A \cap C) = \mathrm{P}(A) \cdot \mathrm{P}(C)$;
- $\mathrm{P}(B \cap C) = \mathrm{P}(B) \cdot \mathrm{P}(C)$.

Remarque 1.1. On a l'implication « (A_n) indépendant » $\implies$ « (A_n) deux-à-deux indépendant » mais la réciproque est ***fausse***.

1.3 Théorèmes d'existence.

Le théorème suivant justifie l'existence des suites finies ou dénombrables de « *bit*s aléatoires indépendants ».

Théorème 1.1 (Existence de *bits* aléatoires). 1. Pour tout entier $n \in \mathbb{N}$, il existe un espace de probabilité $(\Omega_n, \mathscr{F}_n, \mathrm{P}_n)$ qui contient n événements indépendants de probabilité $\frac{1}{2}$.
2. Il existe un espace de probabilité $(\Omega, \mathscr{F}, \mathrm{P})$ qui contient une suite dénombrable d'événements de probabilité $\frac{1}{2}$.

Preuve. 1. On pose $\Omega_n = \{0,1\}^n$, $\mathscr{F}_n = \wp(\Omega_n)$, et P_n la probabilité uniforme. Si on pose

$$A_k = \{\, \omega = (\omega_1, \dots, \omega_n) \in \{0,1\}^n \mid \omega_k = 1 \,\},$$

alors

$$\mathrm{P}(A_k) = \frac{|A_k|}{|\Omega_n|} = \frac{2^{n-1}}{2^n} = \frac{1}{2}.$$

Si $J \subseteq \{1, \dots, n\}$, en notant $p = |J|$, alors

$$\mathrm{P}\Big(\bigcap_{j\in J} A_j\Big) = \frac{\left|\bigcap_{j\in J} A_j\right|}{|\Omega_n|} = \frac{2^{n-p}}{2^n} = \frac{1}{2^p} = \prod_{j\in J} \mathrm{P}(A_j).$$

On a donc indépendance de $(A_k)_{1\le k\le n}$.

2. On l'admet (▷ existence de la mesure de Lebesgue).

□