Méthodes classiques de dérandomisation

Nicolas Schabanel*

16 novembre 2001

Remarques bibliographiques :

- Méthode des probabilités conditionnelles implicite dans Erdös et Selfridge, 1973.
- Méthode des espaces de probabilités 2 à 2 indépendantes originnellement dûes à Joffe (en mathématiques), 1974, redécouvertes en informatique indépendamment par Chor et Goldreich, 1985, et Luby et Widgerson, 1985.
- Méthode à base de matrices de Toeplitz dûe à Carter et Wedman, 1979.

Notes : Las Vegas produit toujours la bonne réponse mais tourne en temps aléatoire - Monte Carlo tourne en temps déterministe mais produit une réponse aléatoire

PLAN DE L'EXPOSÉ :

1 L'aléatoire permet (souvent) d'aller nettement plus vite

Un exemple :

Game Tree Evaluation : (voir le livre "Randomized Algorithm", de Motwani et Raghavan, pp 28–37). Un algorithme aléatoire optimal tourne en temps $n^{\log_4 3} \leqslant n^{0.793} \ll n$ le temps du meilleur algorithme déterministe.

Cette borne est optimale (principe de Yao : le temps du meilleur algorithme randomisé est minoré par le temps du meilleur algorithme déterministe sur la pire distribution)

*Ce résumé est largement inspiré du polycopié de cours de Oded Goldreich, téléchargeable sur sa page: http://www.wisdom.weizmann.ac.il/~oded/homepage.html.

2 Une approximation randomisée pour Max-Cut et Max-Sat

2.1 Max-Cut

2.1.1 Le problème

Max-Cut. Soit $G = (V, E)$ un graphe non-orienté avec des poids $w_e \geqslant 0$ sur les arêtes. Trouver une partition $(S, \bar{S})$ de V qui maximise le poids total des arêtes de cette coupe, i.e. les arêtes ayant une extremité dans S et l'autre dans $\bar{S}$.

Complexité de Max-Cut. Max-Cut est $MaxSNP$-difficile : pas de PTAS ; on ne peut approximer à mieux que 83/84. Max-Cut sans poids est déjà NP-difficile.

Algorithmes 1976 : 1/2 par Sahni et Gonzales. La meilleure approximation connue : 0.87856 à base de programmation semi-définie, par Goemans et Williamson, 1995. 1998 : Un PTRAS pour le cas métrique.

2.1.2 Un algorithme "trivial"

Algorithme randomisé "trivial". Tirer indépendamment avec probabilité 1/2 l'appartenance de chaque sommet x à S ou $\bar{S}$.

Analyse de l'algorithme. C'est une 1/2-approximation : $OUTPUT \geqslant \frac{1}{2}OPT$.

Points-Clés : Linéarité de la valeur moyenne. Indépendance de tirages 2 à 2.

Définitions/Propriétés.
– *Linéarité de la valeur moyenne :* Pour **toutes** variables aléatoires X et Y :

$$\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y]$$

– *Indépendances :* Deux évennements A et B sont indépendants si :

$$\Pr\{A \cap B\} = \Pr\{A\} \cdot \Pr\{B\}$$

- *Variables indépendances :* Deux variables aléatoires X et Y à valeurs discrètes sont indépendantes si : pour tout x et y,

$$\Pr\{X = x \cap Y = y\} = \Pr\{X = x\} \cdot \Pr\{Y = y\}$$

- *Conséquence :* Si X et Y sont deux variables aléatoires indépendantes, alors :

$$\mathbb{E}[XY] = \mathbb{E}[X] \cdot \mathbb{E}[Y]$$

2.2 Max-SAT

Max-SAT. Soit $f = \bigwedge_{c \in \mathcal{C}} c$ une formule sous forme conjonctive normale sur les variables booléennes $x_1, \ldots, x_n$ et des poids $w_c \geqslant 0$ sur les clauses c de f. Trouver une affectation des variables $x_1, \ldots, x_n$ qui maximise la somme des poids des clauses satisfaites : $\sum_{c \text{ satisfaite}} w_c$.

Complexité de Max-SAT. Max-2SAT est NP-difficile (2SAT est dans P) et est $MaxSNP$-difficile ; on ne peut approximer à mieux que 95/96. La meilleure approximation possible pour Max-3SAT est 7/8 (sauf si $P = NP$).

Algorithmes pour Max-SAT. L'algorithme randomisé trivial donne le 7/8 pour Max-3SAT. Algorithmes pour Max-SAT donne 3/4 à base de LP (Yannakakis, 1994) ; mais un meilleur à base de SP donne une 0.87856 pour Max-2SAT (Goemans et Williamson, 1995). Idem, pour Max-SAT, .7584-approximation par Goemans et Williamson, 1995

2.2.1 Algorithme trivial

Pas le temps

3 Introduction à la dérandomisation

3.1 Pour quoi faire ?

- L'aléatoire est-il nécessaire ? $ZPP = RP = BPP = P$? ($ZPP = RP \cap coRP$ = Las Vegas, zéro erreur, temps en moyenne polynomial ; RP = temps déterministe polynomial, résultat probabiliste uniquement pour les réponses positives ; BPP = Monte Carlo, temps déterministe polynomial, erreurs des deux côtés).
- On souhaite parfois une réponse sûre (on en dispose pas de véritable source aléatoire).

– Construction et analyse simple élégante d'algorithme déterministe (parfois) originaux. Pas toujours de sacrifice du temps de calcul

3.2 Différentes méthodes

– *Probabilités conditionnelles :* Construire un algorithme glouton qui effectue les "meilleurs choix probabilistes".
– *Réduire l'espace des probabilités :* Utiliser un espace de probabilités plus petit (typiquement de taille log) en remarquant que les propriétés requises par l'analyse de l'algorithme sont faibles. Deux variantes :
  – Les espaces de variables k à k indépendantes. (dans ce talk)
  – Les espaces de bias faible (pas dans ce talk)

4 La méthode des probabilités conditionnelles

Principe. Remplacer les choix aléatoire par le meilleur choix probabiliste.

Définition. La probabilité d'un évennement B conditionnée par l'évenement A est :

$$\Pr\{B|A\} = \frac{\Pr\{A \cap B\}}{\Pr\{A\}}$$

Si A est B sont des évennements indépendants, alors :

$$\Pr\{B|A\} = \Pr\{B\}$$

La valeur moyenne d'une variable X étant donnée un évennement A est :

$$\mathbb{E}[X|A] = \sum_x x \Pr\{X = x|A\}$$

Propriété fondamentale. Soient deux variables aléatoires X et Y :

$$\mathbb{E}[Y] = \sum_x \Pr\{X = x\}\mathbb{E}[Y|X = x] \leqslant \max_x \mathbb{E}[Y|X = x]$$

Application : $Y =$ Valeur de la sortie de l'algorithme et $X =$ Un des choix aléatoires.
Plus généralement : Soient $A_1, \ldots, A_n$ des évennements mutuellement exclusifs dont l'union est l'espace tout entier, alors :

$$\mathbb{E}[Y] = \sum_i \Pr\{A_i\}\mathbb{E}[Y|A_i] \leqslant \max_i \mathbb{E}[Y|A_i]$$

Détail de la méthode. Remplacer tous les choix aléatoires de la valeur x d'une variable aléatoire X par le choix x^* qui maximise :

$$\mathbb{E}[OUTPUT|X = x, \text{choix précédents}]$$

Mise en œuvre. Il faut que le calcul de la probabilité conditionnelle soit possible (en temps polynomial).

4.1 Application à Max-Cut

Certains sommets sont dans S, d'autres dans S'. On sélectionne le sommet v, quel est le meilleur choix pour v ? i.e. quel est le choix qui va maximiser la valeur moyenne de la sortie conditionné au choix de v ?

Calcul de la probabilité conditionnelle. Quelque soit le choix pour v :
- Toutes les arêtes internes à S ou S' ne contribuent plus à la valeur finale.
- Toutes les arêtes entre S et S' contribuent au coût.
- Toutes les arêtes qui n'ont pas v pour extrémité contribuent pour moitié au coût dans l'algorithme aléatoire.
- Toutes les arêtes qui ont pour extrémité v mais l'autre extrémité ni dans S ni dans S', contribuent pour moitié au coût dans l'algorithme aléatoire.

Reste donc :
- Les arêtes qui ont pour extrémité v et l'autre dans S, qui contriburont au coût pour totalité si $v \in S'$, mais pour 0 si $v \in S$.
- Les arêtes qui ont pour extrémité v et l'autre dans S', qui contriburont au coût pour totalité si $v \in S$, mais pour 0 si $v \in S'$.

Maximiser la valeur moyenne conditionnée au choix de v revient donc à choisir l'ensemble S ou S' vers lequel v est l'extrémités d'arêtes dont poids total est maximal !

On retrouve l'algorithme glouton le plus naturel qui soit.

L'algorithme randomisé est une 1/2-approximation, cet algorithme est donc lui aussi une 1/2-approximation ! (L'analyse est exacte, cf exemple).

4.2 Application à Max-SAT

Laisser en exercice.

5 La méthode des variables k à k indépendantes

5.1 Principe

Principe. Pour générer q variables aléatoires uniformes sur r bits, on a besoin de qr bits aléatoire.
Pour générer des q variables aléatoires uniformes seulement 2 à 2 indépendantes sur r bits, nous n'avons pas besoin de qr bits, seulement $r + \lceil \log q \rceil$ suffisent !

Applications. Si $r = O(\log n)$ et $q = O(n)$, il suffit de tout essayer ! On obtient un algorithme (parallèle) dans NC.
C'est le cas pour notre algorithme de Max-Cut : nous avons besoin $q = |V|$ fois de $r = 1$ bits aléatoires 2 à 2 indépendants seulement.

5.2 Construction de tels espaces réduits

Définition. 2 à 2 indépendance. Ce n'est pas pareil qu'indépendance, ex : soit X et Y deux variables indépendantes sur $\{0,1\}^r$, alors X, Y et $X \oplus Y$ sont 2 à 2 indépendantes, mais pas indépendantes.

Propriété Pour deux variables aléatoires X_1 et X_2 sur $\mathcal{X}$, (X_1, X_2) est uniforme sur $\mathcal{X} \times \mathcal{X}$ $\Leftrightarrow$ X_1 et X_2 sont uniformes sur $\mathcal{X}$ et mutuellement indépendantes.

Corollaire. Si X_1 et X_2 sont des variables aléatoires à dans $\mathcal{X}$ de fini : X_1 et X_2 sont uniformes sur $\mathcal{X}$ et mutuellement indépendantes $\Leftrightarrow$ pour tout $(x_1, x_2) \in \mathcal{X} \times \mathcal{X}$,

$$\Pr\{X_1 = x_1 \wedge X_2 = x_2\} = \frac{1}{|\mathcal{X}|^2}$$

Espace de probabilité pour chaque variable : $\mathcal{S} = \{0,1\}^r$. Taille de l'espace pour q variables uniformes sur $\mathcal{S}$, a priori S^q. On va montrer que l'on peut se ramener à un espace de probabilités Ω tel que : $|\Omega| \approx (\max(|\mathcal{S}|, q))^2$ (précisément : $|\Omega| = \max(2^r, 2^{\lceil \log_2 q \rceil})^2$. Ainsi le nombre de bits aléatoires nécessaire devient $2\max(\lceil \log q \rceil, 2r)$.

Réduction. Sans perte de généralité, on suppose $q \leq |\mathcal{S}| = 2^r$: il suffit de prendre un espace plus grand avec $r' = \lceil \log q \rceil$; si on obtient une distribution uniforme 2 à 2 indépendantes de q variables sur r' bits, il suffit de ne garder que les r premiers pour obtenir q variables 2 à 2 indépendantes sur r bits.

Construction de Ω. Mettons sur $\mathcal{S}$ la structure de corps $GF(2^r)$ (elle existe d'ailleurs uniquement pour les ensembles dont le cardinal est une puissance de nombre premier). Soit :

$$\Omega = \{\text{The set of all linear polynomials on } \mathcal{S}\} = \{ax + b : (a, b) \in \mathcal{S}^2\}$$

De façon équivalente ($q \geq 2$) :

$$\Omega = \{s_a, b = (a \cdot 1 + b, a \cdot 2 + b, \ldots, a \cdot q + b) : (a, b) \in \mathcal{S}^2\}$$

Ici i désigne le "i-ème" élément du corps donné par un ordre quelconque (NB : rien à voir avec les sommes itérées de l'unité et donc pas de limitation dûes à la caractéristique du corps).

Lemme/Remarque. Si X est une variable aléatoire *quelconque* sur un corps $\mathcal{K}$ et U est une variable aléatoire uniforme sur $\mathcal{K}$ et indépendante de X, alors $X + U$ est une variable aléatoire uniforme sur $\mathcal{K}$.
Ainsi, chaque $ai + b$ suit bien une distribution uniforme sur $\mathcal{S}$.

On veut démontrer que s'il on tire uniformément $(a, b) \in \mathcal{S}^2$, alors les q variables de s_a, b sont bien uniformément réparties sur $\mathcal{S}$ et deux à deux indépendantes. Il suffit poure celà de prouver que :

Lemme. Pour tout $i \neq j$ de $\{1, \ldots, q\} \in \mathcal{S}$, et pour tout x et y dans $\mathcal{S}$:

$$\Pr\{a \cdot i + b = x \wedge a \cdot j + b = y\} = \frac{1}{|\mathcal{S}|^2}$$

Preuve.

$$\Pr\{a \cdot i + b = x \wedge a \cdot j + b = y\} = \frac{|\{(a, b) : ai + b = x \wedge aj + b = y\}}{|\mathcal{S}|^2}$$

Or s'il on regarde ces deux équations comme un système d'équation en (a, b), il n'existe qu'une unique solution puisque la matrice

$$\begin{pmatrix} i & 1 \\ j & 2 \end{pmatrix}$$

est inversible dès que $i \neq j$.

Génération de variables k à k indépendantes. Les q-uplets générés précédemment ne sont pas 3 à 3 indépendants (le système à 3 équations n'admet pas toujours une unique solution).

Mais une transformation très simple permet de régler le problème, il suffit de définir :

$$\Omega_k = \{\text{polynômes de degré} < k \text{ sur } \mathcal{S}\} = \{(\sum_{l<k} a_l 1^l, \sum_{l<k} a_l 2^l, \ldots, \sum_{l<k} a_l q^l) : (a_0, \ldots, a_{q-1}) \in \mathcal{S}^k\}$$

La matrice est alors une matrice de Van der Monde, elle est inversible, et l'on obtient la propriété de k à k indépendance souhaitée pour un coût économique de $k \max(r, \lceil \log_2 q \rceil)$ bits.

5.3 Mise en œuvre sur Max-Cut

Pour Max-Cut : $r = 1$ et $q = |V|$. L'espace de probabilités Ω nécessite donc de tirer exactement $2\lceil \log_2 q \rceil \leq 2 \log_2 n + 2$ bits (pour sélectionner a et b). On peut donc *tous* les essayer (en parallèle) et garder le meilleur résultat !

5.4 Autres méthodes de construction d'espaces réduits de variables 2 à 2 indépendante

5.4.1 Échantillonnage à base de transformations affines

Cette méthode ne se généralise pas au cas de variables k à k indépendantes.

On ne suppose plus que $q \leq r$. En revanche, on suppose sans perte de généralité que $q = 2^\ell$ (il suffit de prendre $q' = 2^{\lceil \log_2 q \rceil}$, avoir plus de variables, ne peut pas faire de mal).

Posons

$$\Omega = \{\text{transformation affines de } [q] = \{0,1\}^\ell \text{ dans } \mathcal{S} = \{0,1\}^r\}$$

Soit $u_1, \ldots, u_q$ sont tous les vecteurs de $\{0,1\}^\ell$, alors :

$$\Omega = \left\{ (Au_1 + b, \cdots, Au_q + b) : \begin{array}{l} A \text{ est une } r \times \ell\text{-matrice booléenne} \\ \text{et } b \text{ un vecteur de } \{0,1\}^r \end{array} \right\}$$

La taille de Ω est alors :

$$|\Omega| = 2^{r\ell + r} \approx 2^{r(1+\log_2 q)} \ll 2^{rq}$$

En fait on verra comment mieux choisir les matrices A afin de réduire encore la taille de Ω à $2^{2r+\ell-1}$ et donc être comparable avec la méthode précédente.

On propose alors la distribution suivante sur Ω : tirage uniforme des coefficients de A et de b.

Remarque. En utilisant le lemme précédent, comme b a une distribution uniforme, on retrouve bien que chacun des $Au_i + b$ suit une distribution uniforme sur $\{0,1\}^r$.

Uniformité et 2 à 2 indépendance. Comme précédemment, il s'agit de prouver que : pour tout $u \neq v$ dans $\{0,1\}^\ell$, pour tout x et y dans $\mathcal{S} = \{0,1\}^r$,

$$\Pr\{Au + b = x \wedge Av + b = y\} = \frac{1}{|\mathcal{S}|^2}$$

Preuve.

Étape 1. On commence par réécrire :

$$\Pr\{Au + b = x \wedge Av + b = y\} = \Pr_{(A,b)}\{Au + b = x\} \cdot \Pr_{(A,b)}\{Av + b = y | Au + b = x\}$$

En utilisant le lemme précédent comme A et b sont tirés indépendamment on sait que :

$$\Pr_{(A,b)}\{Au + b = x\} = \frac{1}{|\mathcal{S}|}$$

Ainsi on est ramené à montrer que :

$$\Pr_{(A,b)}\{Av + b = y | Au + b = x\} = \frac{1}{|\mathcal{S}|}$$

Comme l'événement $Av + b = y$ est considéré uniquement lorsque l'événement $Au + b = x$ se produit, on peut substituer b dans $Av + b = y$ par $b = x - Au$, et donc :

$$\Pr_{(A,b)}\{Av + b = y | Au + b = x\} = \Pr_{(A,b)}\{A(v - u) = y - x | b = x - Au\}$$

Comme on cherche pour tout x et y et pour tout u et v, posons $z = y - x$ et $w = u - v$, on obtient :

$$\Pr_{(A,b)}\{Av + b = y | Au + b = x\} = \Pr_{(A,b)}\{Aw = z | b = x - Au\}$$

Or $b = x - Au$ est indépendant de $Aw = z$ puisque b est indépendant de A. Ainsi :

$$\Pr_{(A,b)}\{Av + b = y | Au + b = x\} = Pr_{(A,b)}\{Aw = z\}$$

Et donc :

$$\Pr\{Au + b = x \wedge Av + b = y\} = Pr_{(A,b)}\{Aw = z\}$$

où z et w vérifient en particulier que $w \neq 0$. *Nous n'avons pour l'instant utiliser à aucun moment la structure de A.*

Étape 2. Montrons que $Pr_{(A,b)}\{Aw = z\} = \frac{1}{|\mathcal{S}|}$.

Puisque $w \neq 0$, soit i tel que $w_i = 1$. Regardons la $i$éme colonne de A, elle est aléatoire uniforme et indépendantes des autres colonnes, par le lemme de tout à l'heure sur l'ajout d'une variable aléatoire uniforme à une autre variable, on en conclut que Aw est aléatoire uniforme sur $\mathcal{S} = \{0,1\}^r$ et donc vaut z avec probabilité $\frac{1}{|\mathcal{S}|}$.

Nous avons donc ramener un espace de taille $|\mathcal{S}|^q = 2^{rq}$ à un espace de taille $|\Omega| = 2^{\lceil \log_2 q \rceil r + r}$ où seulement $r(1 + \lceil \log_2 q \rceil)$ bits aléatoires sont nécessaires.

Nous allons voir comment améliorer cette méthode pour réduire le nombre de bits nécessaires à $2r + \lceil \log_2 q \rceil$.

5.4.2 Matrices de Toeplitz

Nous allons montrer ici que l'on peut réduire drastiquement le nombre de bits aléatoires utilisés dans la matrice A en restreignant A à être "de Toeplitz".

De même que précédemment, on suppose que $q = 2^\ell$ ($\ell = \lceil \log_2 q \rceil$).

Une matrice de Toeplitz est une matrice dont tous les éléments d'une même diagonale ont la même valeur.

Définition. Une $r \times \ell$-matrice T est une matrice de Toeplitz si pour tout i, j, $T_{i,j} = T_{i+1,j+1}$.

Une matrice de Toeplitz est complétement spécifiée par $r + \ell - 1$ bits.

Comme précédemment, on numérote les vecteurs On pose donc :

$$\Omega = \left\{ (Tu_1 + b, \cdots, Tu_q + b) : \begin{array}{l} t \text{ est une } r \times \ell\text{-matrice de Toeplitz} \\ \text{et } b \text{ un vecteur de } \{0,1\}^r \end{array} \right\}$$

Propriété. Uniformité et indépendance 2 à 2. pour tout $u \neq v$ dans $\{0,1\}^\ell$, pour tout x et y dans $\mathcal{S} = \{0,1\}^r$,

$$\Pr\{Tu + b = x \wedge Tv + b = y\} = \frac{1}{|\mathcal{S}|^2}$$

Preuve. Comme précédemment, en deux étapes :

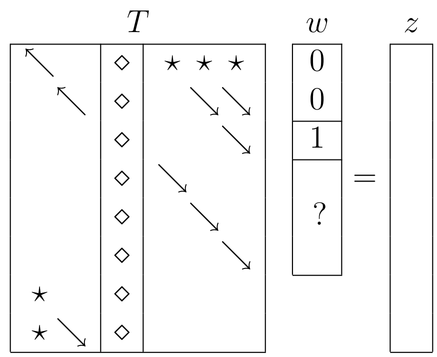

FIG. 1 – T est complétement déterminée par les $i-1$ derniers éléments de la première colonne, tous ceux de la i-ème, et les $\ell - i$ derniers de la première ligne.

Étape 1. Comme nous l'avons remarqué précédemment, l'étape 1 de la preuve précédente n'utilise pas la structure interne de A, ainsi on également l'existence de $z \in \{0,1\}^r$ et $w \in \{0,1\}^\ell \setminus 0$ tels que :

$$\Pr\{Tu + b = x \wedge Tv + b = y\} = \frac{1}{|\mathcal{S}|} \Pr_T\{Tw = z\}$$

Étape 2. Montrons donc que $\Pr_T\{Tw = z\} = \frac{1}{|\mathcal{S}|}$ lorsque $w \neq 0$.

Il faut affiner l'argumentation précédente : au lieu de choisir un indice i tel que $w_i = 1$ quelconque, choisissons le plus petit tel i. Et regardons la $i$ème colonne de T (il y en a ℓ, tout va bien). En regardant la figure 1, on constate que T est complétement déterminée par les $i-1$ derniers éléments de la première colonne, tous ceux de la i-ème, et les $\ell - i$ derniers de la première ligne (soit $i - 1 + r + \ell - i = r + \ell - 1$).

Nous pouvons maintenant décrire T uniquement par les éléments de "type $\star$" et ceux de "type $\diamond$". Comme ils sont sélectionnés indépendamment, il suffit de montrer que :

$$\Pr\{Tw = z | \star\} = \frac{1}{|S|}$$

où $\Pr\{Tw = z|\star\}$ désigne la probabilité de l'évennement en ayant fixer les coefficients de "type $\star$" à une constante arbitraire.

Il suffit pour celà de montrer que $T(\diamond, \star)w = z$ admet une solution $\diamond$ unique.

A TERMINER LA PROCHAINE FOIS.