Génération, Comptage et Approximation
La méthode MC2

Richard Lassaigne
Logique mathématique,
CNRS-Université Paris Diderot

ANR Sécurité et Sûreté Informatique
Projet VERAP
http ://www.lri.fr/∼mdr/verap

Complexité des problèmes

- Décision : Existe-t-il une solution ?
- Recherche : Trouver une solution
- Comptage : Compter les solutions
- Génération : Engendrer les solutions (de manière uniforme)

Question : L'approximation permet-elle d'améliorer
l'efficacité des algorithmes ?

Pb	SAT	VC	CLIQUE	TSP	KNAPSACK
Déci.	NP-com.	NP-com.	NP-com.	NP-com.	NP-com.
Opt.					Pseudo-Poly
Approx. (ε-appr.)	MaxSAT $\varepsilon = 1/2$	MinVC $\varepsilon = 1/2$	MaxCLIQ.	MinTSP	FPTAS
Non Approx.	FPTAS $\Rightarrow$ P=NP		ε-approx. $\Rightarrow$P=NP	ε-approx. $\Rightarrow$P=NP	

Fully Polynomial-Time Approximation Scheme (FPTAS) : poly($|x|$, $1/\varepsilon$)

- Existe-t-il un rapport entre échantillonnage uniforme, comptage et approximation ?
- Comment utiliser la méthode de Monte-Carlo pour la génération (presque) uniforme ?
- Que peut-on dire sur le temps de convergence ?

- Problèmes de comptage et
  schémas probabilistes d'approximation

- De la génération presque uniforme à
  l'approximation du comptage

- La méthode MC^2 (Monte-Carlo Markov Chain)

- Couplage et temps de mélange

p-Prédicat : R relation binaire sur Σ^* t.q.

- pour tout $x, y \in \Sigma^*$, $R(x, y)$ entraîne $|y| = \text{poly}(|x|)$
- pour tout $x, y \in \Sigma^*$, $R(x, y)$ est décidable
  en temps poly $(|x|)$

Fonction F_A associée à un pb A de la classe NP :
fonction de comptage $F_A(x) = |\{y \in \Sigma^* \ / \ R(x, y)\}|$

Classe $\#P$: Fonctions $F : \Sigma^* \longrightarrow \mathbb{N}$ t.q.
il existe un p-Prédicat R avec $F(x) = |\{y \in \Sigma^* \ / \ R(x, y)\}|$

Exemples : $\#SAT$, $\#IS$, $\#COL$, $\#HAM$

Réduction entre fonctions $\#P$: $f, g : \Sigma^* \longrightarrow \mathbb{N}$
$g \prec_P f$ s'il existe 2 fonctions calculables en temps poly
$h : \Sigma^* \longrightarrow \Sigma^*$ et $\alpha : \mathbb{N} \longrightarrow \mathbb{N}$ t.q. $g(x) = \alpha(f(h(x)))$

Soit f une fonction de la classe $\#P$
Schéma Probabiliste d'Approximation (RAS) pour f :

Algorithme probabiliste A t.q. pour toute entrée x,
pour tout $\varepsilon > 0$ (paramètre d'approximation),
pour tout $\delta > 0$ (paramètre de confiance),
A fournit une valeur $A(x, \varepsilon, \delta)$ satisfaisant

$$Prob_{\Omega}(|A(x, \varepsilon, \delta) - f(x)| \leq \varepsilon . f(x)) \geq 1 - \delta$$

Ω est l'espace probabiliste des tirages de l'algorithme

Pleinement polynomial (FPRAS) :
Le temps de calcul est poly$(|x|, 1/\varepsilon, log(1/\delta))$

Problème $\#DNF$: $\#P$-complet

Théorème : (R. Karp, M. Luby et N. Madras, 89)

La méthode de Monte-Carlo avec borne (Chernoff-Hoeffding) sur l'échantillon est un FPRAS pour le problème $\#DNF$

F formule DNF : $\bigvee_{i=1}^{m} G_i$

S_i ensemble des valuations satisfaisant G_i $(i = 1, \ldots, n)$

$c(v) =$ nombre de formules G_i que la valuation v satisfait

Idée : On échantillonne dans le multi-ensemble $M = \biguplus_{i=1}^{m} S_i$

- choisir une formule G_i avec probabilité $|S_i|/|M|$
- choisir une valuation v satisfaisant G_i de manière uniforme
- tester si la valuation v ne satisfait aucune des G_j pour $j < i$

Lemme 1 : $X(v) = |M|/c(v)$ est un bon estimateur pour $\#F$

Lemme 2 : $\mu = \#F/|M| \geq 1/m$

Taille de l'échantillon (polynomiale) : $N = \frac{4m}{\varepsilon^2} ln(2/\delta)$

Problème Network Reliability :

- Entrée : Un graphe $G = (E, R)$ connexe
  avec probabilité de disparition p pour chaque arête
- Sortie : La probabilité que le graphe G devienne non connexe

Résultat (D. Karger, 94) : Conception d'un FPRAS

Méthode :

- Dénombrement des coupures α-minimales dans un graphe
- Utilisation du FPRAS pour la version probabiliste de $\#DNF$

Théorème : S'il existe un FPRAS pour le problème $\#SAT$ alors $RP = NP$

Idée :

- Un FPRAS pour $\#SAT$ permettrait de distinguer, en temps polynomial, pour l'entrée x, entre $\#(x) = 0$ et $\#(x) > 0$
- Alors il existerait un algorithme probabiliste en temps polynomial pour SAT et $NP \subseteq BPP$

Remarque : M. Jerrum et A. Sinclair

$$NP \subseteq BPP \Rightarrow RP = NP$$

Soit f une fonction de la classe $\#P$:
$f(x) = |\{y \in \Sigma^* \ / \ R(x,y)\}|$
Générateur presque uniforme (AUG) pour f :

Algorithme probabiliste G t.q.
pour tout x, $\varepsilon > 0$
si $f(x) > 0$ alors pour tout y t.q. $(x,y) \in R$,

$$|Prob_\Omega(G \text{ sort } y) - 1/f(x)| \leq \varepsilon/f(x)$$

Ω est l'espace probabiliste des tirages de l'algorithme

Pleinement polynomial (FPAUG) :
Le temps de calcul est poly($|x|, 1/\varepsilon$))

Méthode :

Estimation (Monte-Carlo) + borne de Chernoff-Hoeffding

X variable de Bernouilli $(0, 1)$ avec probabilité de success p

- Faire M tirages aléatoires indépendants $X_1, X_2, \ldots, X_N$
- Estimer p par $Y = \sum_{i=1}^{N} X_i/N$ avec erreur ε
- La taille de l'échantillon M est telle que la probabilité d'erreur (de l'algorithme) $< \delta$

Borne de Chernoff-Hoeffding :

$$Pr[Y < p - \varepsilon] + Pr[Y > p + \varepsilon] < 2e^{-3N\varepsilon^2}$$

Si $M \geq \ln(\frac{2}{\delta})/3\varepsilon^2$, alors

$$Pr[p - \varepsilon \leq Y \leq p + \varepsilon)] \geq 1 - \delta$$

Exemple : Ensembles indépendants dans un graphe $G = (V, E)$
Ordre sur les arêtes : $e_1, \ldots, e_m$, $E_i = \{e_1, \ldots, e_i\}$, $G_i = (V, E_i)$
$\Omega(G_i)$ espace des ensembles indépendants de G_i
$r_i = |\Omega(G_i)|/|\Omega(G_{i-1})|$ $(i = 1, \ldots, m)$
Problème : Estimer $|\Omega(G)| = 2^n \prod_{i=1}^{m} r_i$

Algorithme d'estimation de r_i :

- X := 0
- Faire $M = \lceil 1296 m^2 ln(2m/\delta)/\varepsilon^2 \rceil$ tirages indépendants :
  engendrer un élément $(\epsilon/6m)$-uniforme dans $\Omega(G_{i-1})$
  si c'est un ensemble indépendant dans G_i, alors $X := X + 1$
- $r'_i := X/M$

Lemme : C'est une $(\epsilon/2m, \delta/m)$-approximation de r_i

Conséquence : Borne sur l'erreur d'estimation de $|\Omega(G)|$

Soit $R = \prod_{i=1}^{m} r'_i/r_i$

Alors $Prob(|R-1| \leq \varepsilon) \geq 1-\delta$

Théorème :

Etant donné un FPAUG pour les ensembles indépendants dans un graphe, on peut construire un FPRAS pour $\#IS$

Théorème : [M. Jerrum, A. Sinclair]

Conditions équivalentes pour une relation auto-réductible R :

- existence d'un FPAUG
- existence d'un FPRAS

Chaîne de Markov (finie) : matrice de transition $P = (p_{ij})$

Théorème 1 : Toute chaîne de Markov finie,
irréductible et apériodique (i.e. $pgcd\{t/P^t(i,j) > 0\} = 1$)
a une unique distribution stationnaire $\bar{\pi} = (\pi_0, \ldots, \pi_n)$

Théorème 2 : Si $\mathcal{M}$ est finie, irréductible et apériodique,
et si $\bar{\pi}$ est une distribution t.q. pour tout i, j,

$$\pi_i P_{ij} = \pi_j P_{ji}$$

alors $\bar{\pi}$ est la distribution stationnaire de $\mathcal{M}$

Conséquence : $\bar{\pi}$ est uniforme ssi P est symétrique

Remarque : Une marche aléatoire ne mène pas
à une distribution uniforme

Lemme 1 : Espace probabiliste Ω (fini)
muni d'une structure de voisinages $\{N(x)/x \in \Omega\}$
$N = max_{x\in\Omega}|N(x)|$
Soit $M \geq N$ et $\mathcal{M}$ la chaîne de Markov définie par

- $p_{xy} = 1/M$ si $x \neq y$ et $y \in N(x)$
- $p_{xy} = 0$ si $x \neq y$ et $y \notin N(x)$
- $p_{xy} = 1 - |N(x)|/M$ si $x = y$

Si cette chaîne est irréductible et apériodique,
alors la distribution stationnaire obtenue est uniforme

Exemple : On désire une chaîne de Markov où,
pour la distribution stationnaire, chaque ensemble indépendant I a une probabilité proportionnelle à $\lambda^{|I|}$ pour un certain λ

Idée : Généraliser l'introduction de probabilités positives
pour les boucles

Lemme 2 : Soit $M \geq N$ et pour tout $x \in \Omega$, $\pi_x > 0$
la probabilité désirée pour x dans la distribution stationnaire
On considère la chaîne de Markov définie par

- $p_{xy} = (1/M)min(1, \pi_y/\pi_x)$ si $x \neq y$ et $y \in N(x)$
- $p_{xy} = 0$ si $x \neq y$ et $y \notin N(x)$
- $p_{xy} = 1 - \sum_{y \neq x} p_{xy}$ si $x = y$

Si cette chaîne est irréductible et apériodique,
alors la distribution stationnaire obtenue est $\bar{\pi} = (\pi_x)_{x \in \Omega}$

Distance entre 2 distributions μ et ν sur S :

$$||\mu - \nu|| = 1/2 \sum_{x \in S} |\mu(x) - \nu(x)|$$

Lemme : $||\mu - \nu|| = max_{A \subseteq S} |\mu(A) - \nu(A)|$

Temps de mélange :

- $\bar{\pi}$ distribution stationnaire d'une chaîne de Markov sur S
- p_x^t distribution des états de la chaîne issue en x à l'instant t

Soit $\Delta_x(t) = ||p_x^t - \bar{\pi}||$ et $\Delta(t) = max_{x \in S} \Delta_x(t)$

Temps de mélange : $\tau_x(\varepsilon) = min\{t / \Delta_x(t) \leq \varepsilon\}$

$$\tau(\varepsilon) = max_{x \in S} \tau_x(\varepsilon)$$

Une chaîne de Markov mélange rapidement :

$\tau(\varepsilon)$ polynomial en $log(1/\varepsilon)$ et en la taille du problème

Technique générale pour borner le temps de mélange

Couplage d'une chaîne de Markov $\mathcal{M}$ sur un espace S :
Chaîne de Markov $Z_t = (X_t, Y_t)$ sur l'espace S^2 t.q.
$Prob(X_{t+1} = x'/Z_t = (x, y)) = Prob(M_{t+1} = x'/M_t = x)$ et
$Prob(Y_{t+1} = y'/Z_t = (x, y)) = Prob(M_{t+1} = y'/M_t = y)$ et

Lemme (couplage) : $Z_t = (X_t, Y_t)$ couplage pour la chaîne $\mathcal{M}$
S'il existe un temps T t.q. pour tout $x, y \in S$

$$Prob(X_T \neq Y_T/X_0 = x, Y_0 = y) \leq \varepsilon$$

alors $\tau(\varepsilon) \leq T$
c.à.d. pour tout état initial, la distance entre la distribution de la chaîne après T étapes et la distribution stationnaire est $\leq \varepsilon$

Ensembles indépendants de taille fixée k

Mouvement $m(v, w, X_t)$: choix uniforme de $v \in X_t$ et de $w \in V$

- si $w \notin X_t$ et $(X_t - v) \cup \{w\}$ est indépendant,
  alors $X_{t+1} = (X_t - v) \cup \{w\}$,
- sinon $X_{t+1} = X_t$

Chaîne de Markov ergodique convergeant
vers une distribution stationnaire uniforme

Couplage $Z_t = (X_t, Y_t)$:

- $X_t \longrightarrow X_{t+1}$ suivant le mouvement $m(v, w, X_t)$
- $Y_t \longrightarrow Y_{t+1}$ si $v \in Y_t$, $m(v, w, Y_t)$
  sinon choix uniforme de v' dans Y_t et $m(v', w, Y_t)$

Temps de mélange : $\tau_x(\varepsilon) \leq \frac{knln(k/\varepsilon)}{n-3(k-1)(\Delta+1)}$ lorsque $k \leq \frac{n}{3(\Delta+1)}$
Δ est le degré maximum du graphe

Lemme : Distributions μ_X et μ_Y sur un espace d'états S

Soit $Z = (X, Y)$ une variable aléatoire sur S^2

X (resp. Y) distribuée suivant μ_X (resp. μ_Y)

Alors $Prob(X \neq Y) \geq ||\mu_X - \mu_Y||$

De plus, il existe une distribution jointe $Z = (X, Y)$,

X (resp. Y) distribuée suivant μ_X (resp. μ_Y),

pour laquelle l'égalité est réalisée

Théorème : Pour toute chaîne de Markov ergodique ,

$$\Delta(T+1) \leq \Delta(T)$$

Rappel : $\Delta(t) = max_{x \in S} ||p_x^t - \bar{\pi}||$

Coloriages propres

Remarque : Sur tout graphe de degré maximum Δ,
il existe un coloriage propre avec $c = \Delta + 1$ couleurs

Chaîne de Markov sur les coloriages :

- choisir uniformément un sommet v et une couleur l
  recolorier v avec l si le nouveau coloriage est propre
  (i.e. si v n'a pas de voisin de couleur l)
- sinon pas de mouvement

Chaîne de Markov apériodique et irréductible si $c \geq \Delta + 2$
convergeant vers une distribution stationnaire uniforme
Couplage : mouvement identique sur les 2 copies de la chaîne

Temps de mélange : $\tau_x(\varepsilon) \leq \frac{ncln(n/\varepsilon)}{c-4\Delta}$ lorsque $c \geq 4\Delta + 1$

- Sur les classes de complexité :
  Christos Papadimitriou
  Computational Complexity. Addison-Wessley, 1994.
- Sur les rapports entre Logique et Complexité :
  R. Lassaigne et M. de Rougemont
  Logic and Complexity. Springer-Verlag, 2004.
- Sur l'analyse probabiliste des algorithmes :
  M. Mitzenmacher and E. Upfal
  Probability and Computing : Randomized Algorithms and Probabilistic Analysis.
  Cambridge University Press, New York, 2005.