Structure et algorithmes aléatoires

Introduction

Probabilités discrètes : espaces au plus dénombrables, et application à différents domaines. Complémentaire au cours d'intégration et probabilités

Références :

- Mitzenmacher & Upfal : *Probability and Computing : Randominzed Algorithms and Probabilitic Algorithms*
- Pierre Brémaud : *Discrete Probabilitiy Models and Methods*

Programme :

- Probabilités discrètes et applications
  - Rappels : variables aléatoires, indépendance, espérance et variance, quelques inégalités, fonctions génératrices, boules et urnes
  - Algorithmes aléatoires
  - Méthode probabiliste
  - Graphes aléatoires
- Modèles markoviens
  - Chaînes de Markov, comportement asymptotique
  - Simulation de Monte Carlo et simulation parfaite
  - Extensions et applications : modèles markoviens cachés, modèles markoviens de décision, champs de Gibbs, automates cellulaires probabilistes

Algorithme probabiliste

Algorithme déterministe est tel que pour chaque entrée, il existe une et une seule valeur de sortie.

On veut une réponse rapide et correcte, ce que l'on ne sait pas toujours faire avec un algorithme de déterministe.

On peut ajouter de l'aléa sous forme de **bits aléatoires** : plus une unique sortie pour chaque entrée. La sortie est une variable aléatoire.

On peut modifier l'algorithme de sorte à avoir

- Soit une réponse correcte dans tous les cas et rapide dans la plupart des cas.
- Soit une réponse correcte dans la plupart des cas, mais rapide dans tous les cas.

Exemple :

Problème : deux polynômes de degré d sont-ils égaux ?

Un algo naïf nécessite $\mathcal{O}(d^2)$ opérations pour développer.

Algo probabliliste

Choisir r dans $\{1, ..., 100d\}$ uniformément

Calculer $F(r) et G(r)$ (en $\mathcal{O}(d)$). Si $F(r) = G(r)$ alors $F = G$, sinon $F \neq G$.

On se trompe uniquement si $F \neq G$ et r est une racine de $F - G$.

Au plus d racines dans $\{1, ..., 100d\} \Rightarrow$ probabilité que l'algorithme se trompe est au plus de $1/100$

Attention, pour plus de précision on pourrait vouloir exécuter plusieurs fois l'algo, mais il ne faut alors pas dépasser la complexité de l'algorithme déterministe pour que cela soit intéressant.

Modèle : une file d'attente

X_n clients après le départ du n-ième paquet. $X_0 = 0$. A_n : nombre de clients qui arrivent pendant le service du n-ième client. Alors : $X_{n+1} = max(X_n - 1, 0) + A_n$

(X_n) forme un processus stochastique si A_n est décrit de manière probabiliste. Sous certaines conditions, (X_n) est ce qu'on appelle une chaîne de Markov, qui fait l'objet de la seconde partie du cours.

Chaîne de Markov

Soit $\{X_n, n \in \mathbb{N}\}$ une suite de variables aléatoires (un processus stochastique) à valeurs dans une espace d'états E au plus dénombrable.

Le processus $\{X_n, n \in \mathbb{N}\}$ est appelé une chaîne de Markov à temps discret sur E si pour tout $n \in \mathbb{N}$ pour tous $i, j, i_1, ..., i_{n-1} \in E$,

$P(X_n + 1 = j | X_n = i, ..., X_0 = i_0) = P(X_n + 1 = j | X_n = i)$

Questions :

- comportement asymptotique (stabilité) ?
- Quelle est la taille moyenne de la file d'attente ?
- Et si le nombre de clients double, que faut-il faire pour garder le même temps moyen d'attente qu'avant ?
- Si on a deux files d'attentes, faut-il avoir des salles d'attentes séparées ou une salle commune ?

Plan

- Introduction : Deux exemples d'aplication
- Événements et probabilités
  - Tribus et événements
  - Espaces de probabilités
  - Indépendance, probabilité conditionnelle

- Variables aléatoires
  - Variables aléartoires et discributions
  - Espérance
  - Variance
  - Quelques inégalités

Tribus et événements

Univers : Ω - un ensemble qui décrit toutes les possibilités d'une expérience. Par eemple, pour un dé, on a $\Omega = \{1, 2, 3, 4, 5, 6\}$. Les éléments de Ω sont appelés les événements élémentaires (les éventualités)

Définition : une tribu sur Ω est une famille $\mathcal{F}$ de sous ensemble de Ω qui contient $\emptyset$, Ω, est stable par passage au complémentaire, par réunion dénombrable (intersection dénombrable).

La **tribu grossière** est la plus petite tribu sur Ω.

La **tribu fine** est la plus grosse tribu sur Ω.

Exemples d'événements plus complexes (espaces non dénombrables) :

- $\Omega = \{0, 1\}^{\mathbb{N}}$: une suite infinie de lancers de pièce.
- A=l'événement concerne uniquement les k premiers lancers.
- Les événements de type A ne forment pas une tribu, on considère donc la tribu engendrée (la plus petite tribu contenant les événements de type A). (Ici il s'agit de la tribu engendrée par les cylindres finis)

Espace de probabilités

Définition : une probabilité sur une espace probabilisable (Ω, $\mathcal{F}$) est une application $P : F \to [0, 1]$ telle que $P(\Omega) = 1$, et P est sigma-additive.

Propriétés :

- Passage au complémentaire
- Monotonie
- Union bound ($P(\bigcup A_i) \leq \sum P(A_i)$)
- Continuité séquentielle : La probabilité d'une union croissante des A_n peut être exprimée comme la limite des $P(A_n)$
- Idem pour les intersections décroissantes

Exemple (retour sur les polynômes) : Augmenter la précision, comment

- Augmenter l'espace. Problème: précision sur les grands entiers
- Répéter l'algorithme plusieurs fois
  - On peut choisir avec ou sans remplacement des valeurs déjà tirées. Dans le premier cas, les tirages sont indépendants.

Indépendance

Définition : deux événements A et B sont indépendants si $P(A \cap B) = P(A)P(B)$. Une famille d'événements est mutuellement indépendants si toute sous-famille finie d'évènements est indépendante ($P(\bigcap A_i) = \prod P(A_i)$).

Probabilité conditionnelle

Définition : Probabilité d'un événement A conditionné par B est : $P(A|B) = \frac{P(A \cap B)}{P(B)}$ (uniquement définie si $P(B) > 0$)

Théorème : Formule des probabilités composées : $\mathrm{P}(E_1 \cap \cdots \cap E_n) = P(E_1)P(E_2|E_1)...P(E_n|E_{n-1} \dots E_1)$.

Exemple : Tirage des racines sans remise : un calcul montre que c'est une erreur plus petite qu'avec remplacement.

En pratique il est parfois plus judicieux d'implémenter la version avec remplacement.

Théorème : loi des probabilités totales : $P(A) = \sum_i P(A|E_i)P(E_i)$ pour (E_i) partition de Ω

Théorème : loi de Bayes : $P(A|B) = \frac{P(A)P(B|A)}{P(B)}$

Exercice : vérification d'une multiplication matricielle : Meilleur qu'un algo naïf ($\mathcal{O}(n^3)$) : algo probabiliste. Soit $r \in \{0,1\}^n$.

Complexité de $ABr = Cr$: $\mathcal{O}(n^2)$ (Br puis ABr et Cr. Chaque calcul est en $\mathcal{O}(n^2)$.) Proba d'erreur : $\leq 1/2$ (calcul bourrin par formule des probabilités totales)

Variables aléatoires

Définition : $(\Omega, \mathcal{F}, P)$ espace de probabilité. E un ensemble au plus dénombrable. Une fonction $X : \Omega -> E$ telle que $\forall x \in E, \quad \{\omega | X(\omega) = x\} \in \mathcal{F}$ est une variable aléatoire discète sur E. (techniquement il faut aussi la condition de mesurabilité de X)

Propriété : toute image d'un n-uplet de v.a. est une v.a.

Exemples de distributions :

- Loi constante
- Loi de Bernouilli
- Loi binomiale
- Loi géométrique (sans mémoire)
- Loi de Poisson
- ...

Espérance (cf diapo)

Propriété : linéarité

Propriété : monotonie si $f(X) < g(X)$ p.s. alors $E(f(X)) < E(g(X))$

Propriété : indépendance : $E(XY) = E(X)E(Y)$ si X et Y sont indépendantes

Variance

Propriété : $Var(X, Y) = Var(X) + Var(Y) + 2Cov(X, Y)$

Propriété : Si X et Y sont indépendantes, $Cov(X, Y) = 0$

Propriété : Si X et Y sont indépendantes, $Var(X + Y) = Var(X) + Var(Y)$

Inégalités

Inégalité de Jensen : Si ϕ est une fonction convexe, X et $\phi(X)$ ont une espérance, alors $E[\phi(X)] >= \phi(E[X])$

Inégalité de Markov : X v.a.≥ 0. Alors $P(X \geq a) \leq E[X]/a$.

Exemple : lancer de pièces non biaisées (n fois). Borner la proba d'obtenir au moins $3n/4$ fois face. Markov donne $P(X \geq 3a/4) \leq 2/3$.

Inégalité de Tchebychev : $P(|X - E(X)| \geq a) \leq Var(X)/a^2$ Sur le lancer de n pièces, $P(X \geq 3n/4) \leq 4/n$

Exemple : collectionneur de coupons Chaque boîte contient un et un seul coupon, tiré de manière uniforme parmi toutes les possibilités et indépendamment des autres boîtes. Il y a n coupons au total. Combien de boîtes faut-il acheter en moyenne pour obtenir tous les coupons ? Soit X le nombre de boîtes pour obtenir n coupons. On cherche $E[X]$. Soit X_i le nombre de boîtes ouvertes depuis que le collectionneur a obtenu i-1 coupons jusqu'à l'obtention du i-ème coupon. On a $X = X_1 + ... + X_i$ Distribution des X_i : géométrique de paramètre $(n - (i - 1))/n$. $E[X_i] = n/(n - i + 1)$ Donc $E[X] = nH(n)$ avec $H(n)$ somme harmonique.

Question : quelle est la probabilité que le temps pour collectionner les n coupons soit au moins le double de cette espérance ? Markov : 1/2 Tchebychev : indépendance donc variance linéaire. Tchebychev donne : $\frac{\pi^2}{6H(n)^2}$ Indégalité de Boole (union bound) donne $1/n$

Classification des algorithmes probabilites

Algorithmes de Monte Carlo : complexité déterministe mais peut se tromper avec erreur contrôlée. **Erreur unilatérale :** renvoie `vrai` ou `faux` mais ne se trompe que sur une des valeurs **Erreur bilatérale :** peut se tromper sur les deux valeurs Si la proba d'erreur est <1/2, alors on peut diminuer cette proba en exécutant plusieurs fois l'algorithme et en renvoyant la réponse majoritaire.

Algorithmes de Las Vegas : résout un problème exactement, avec une complexité moyenne finie (que l'on cherche à minimiser) **Exemple :** choix uniforme du pivot dans l'algorithme du tri rapide

Un algorithme Monte Carlo de complexité C_MC qui renvoie soit la réponse correct, soir `erreur` (avec proba au plus p), on peut transforer cet algo en un algorithme de type Las Vega en le répétant tant que la réponse renvoyée est `erreur`. Alors la complexité de l'algorithme Las Vegas est alors E(C_LV) <= C_MC/p

Terminaison :

- Un algo termine avec proba α
- Un algorithme termine presque sûrement
- Un algorithme termine sûrement

Méthode probabiliste

- Argument de comptage
- Méthode du premier moment
- Méthode du second moment

Idée : prouver l'existence d'objets satisfaisant certaines propriétés par des arguments probabilistes. Dans certains cas, on pourra construire effectivement ces objets.

Méthode de comptage

- On dispose d'une collection au plus dénombrables d'objects $a_i, i \in I$.
- Objectif : prouver que l'un d'eux au moins satisfait une propriété $\mathcal{P}$.

Ex : nombre de Ramsey Coloriage des arêtes d'un graphe complet K_n en deux couleurs, rouge et bleu de telle manière qu'il n'y ait pas de grande clique monochrome. (clique : sous-graphe complet)

$R(k)$ est le nombre minimum de sommets tel que pour chaque coloriage des arêtes, il existe une clique de taille k monochrome.

Théorème: si $\binom{n}{k}2^{-\binom{n}{2}+1} < 1$, alors $R(k) > n$ (il est possible de colorier K_n t.q il n'y a pas de clique de tailel k monochrome)

Démo : il y a $2^{\binom{n}{2}}$ coloriage possibles des arêtes K_n avec deux couleurs.

Choix unifiorme de coloriage : on colorie les arêtes iid (indépendante et identiquement distribuées) (i.e. chaque couleur avec proba 1/2)

Soit $i = 1, \ldots, \binom{n}{k}$ une énumération des cliques de taille k.

Soit A_i l'événement "i est une clique monochrome". Alors $P(A_i) = 2^{-\binom{k}{2}+1}$ (deux choix parmi les $2^{\binom{k}{2}}$ coloriages de cette clique possible).

Donc $P(\cup_{i=1}^{\binom{n}{k}} A_i) \leq \sum_{i=1}^{\binom{n}{k}} P(A_i) = \binom{n}{k} 2^{-\binom{k}{2}+1} < 1$ et $P(\cap_{i=1}^{\binom{n}{k}} \bar{A}_i) = 1 - (\cup_{i=1}^{\binom{n}{k}} A_i) > 0$. Il existe donc un tel coloriage.

Construction effective (k constant) On construit un algorithme de type Monte-Carlo. : on colorie chaque arête uniformément et indépendamment. Ensuite, on vérifie qu'il n'y a pas de clique de taille k dans le graphe $\mathcal{O}(n^k)$. Alors, la probabilité d'échec est donc : $p = P(\cup_{i=1}^{\binom{n}{k}} A_i)$

Transformation en algorithme Las-Vegas : répéter tant qu'on trouve une clique monochrome :

$$E(\text{temps d'exécution}) = \mathcal{O}\left(\frac{n^k}{\binom{n}{k}2^{-\binom{k}{2}+1}}\right)$$

Méthode du premier moment

Théorème : si X est une v.a. alors $P(X \geq E[X]) > 0$ et $P(X \leq E[X]) > 0$.

Application 1 : MAXSAT Problème : F formule en FNC (ou CNF en anglais).

Question : qiel est le nombre maximal de cluses satisfiables.

Problème de décision associé (NP-complet) :

Données : F, k.

Question : existe-t-il une affectation des variables telles que k clauses au moins sont satisfiables ?

Théorème : soit F une formule à m clauses, k_i le nombre de littéraux de la i-ème clause et $k = \min k_i$. Il existe une affectation des variables qui satisfait au moins $\sum_{i=1}^{m}(1 - 2^{-k_i}) > m(1 - 2^{-k})$ clauses.

Démo :

On note $x_1 \dots x_n$ les variables de la formule. On affecte de manière uniforme une valeur booléenne à ces variables. Soit E_j "la j-ième clause est satisfaite" et $Y_j = 1_{E_j}$. On a $E[Y_j] = 1 - 2^{-k_j}$. Soit $Y = \sum Y_j$. On a alors : $E[Y] = \sum_{i=1}^{m}(1 - 2^{-k_j}) \geq m(1 - 2^{-k}) > 0$

Comme $P(Y \geq E[Y]) > 0$, cela prouve l'existence d'une affectation qui satisfasse cette clause.

Algorithme de construction :

Utiliser les espérances conditionnelles par rapport à une affectation déjà en partie construite : $E[Y] = E[Y|X_1 = 1]P(X_1 = 1) + E[Y|X_1 = 0]P(X_1 = 0) \leq \max(E[Y|X_1 = 1], E[Y|X_1 = 0])$.

On choisit alors l'affectation qui maximise l'espérance, et on reprend de manière similaire sur les autres valeurs.

Problème: calculer les espérances.

Application 2 : Ensembles indépendants Soit $G = (V, E)$ un graphe. Un ensemble de sommets I et dit indépendant si les différents sommets ne sont pas connectés. On note $n = |V|, m = |E|$.

Théorème : si $2m/n > 1$ alors G possède un ensemble indépendant de tailel au moins $n^2/4m$.

Démo : si $n > m$, alors il existe un indépendant de taille $n - m$ (on peut enlever un sommet extrémité dechaque arête).

Une première phase d'un algorithme de construction consiste donc à enelever des ommets (et les arêtes qui y sont adjacentes).

Algorithme 1 :

$p \in [0, 1]$.

1. Effacer chaque sommet avec probabilité $1 - p$ indépendamment
2. Pour chaque arête restante, la retirer, ainsi qu'un sommet adjacent.

(cela revient à garder des sommets avec probabilité p et considerer le sous-graphe engendré)

Calcul de l'espérance du nombre de sommets et d'arêtes obtenus à la fin.

1. X le nombre de sommets restants après la 1ere arête : $E[X] = np$.
2. Soit Y le nombre d'arêtes à la fin de la première étape. $E[Y] = mp^2$: une arête survit si aucun des sommets qu'elle relit n'est supprimé, ce qui arrive avec probabilité p^2 (par indépendance).

À la deuxième étape, on retire un sommet par arête au plus. Le nombre de sommets restants est donc d'au moins $X - Y$ et $E[X - Y] = np - mp^2$.

On choisit p qui maximise le nombre de sommets restants, ce qui est pour $p = \frac{n}{2m}$ qui est ≤ 1 par hypothèse. Donc $E[X - Y] = \frac{n^2}{2m} - \frac{n^2}{4m} = \frac{n^2}{4m}$.

Remarque : $d = \frac{1}{p} = \frac{2m}{n}$ est le degré moyen du graphe.

Méthode du second moment

Théorème : soit X une variable aléatoire sur $\mathbb{N}$. Alors $P(X \neq X) \leq E[X]$.

Démo : inégalité de Markov avec $a = 1$.

Théorème : dans le même cas, $P(X = 0) \leq \frac{Var(X)}{E[X]^2}$.

Démo : Tchebychev avec $a = E[X]$.

Graphes aléatoires (Erdös-Rényi) Soit $n \in \mathbb{N}$ et $p \in [0,1]$. L'espace $\mathcal{G}(n,p)$ est l'espace des graphes non orientés avec n sommets et où chaque arête a une probabilité p d'exister, indépendamment des autres.

On a $\mathcal{G}(n,p) = (\Omega_n, \mathcal{P}(\Omega_n), P)$ où

- Ω_n est l'ensemble des graphes non orientés avec n sommets $\{1,\ldots,n\}$.
- Si pour $1 \leq u < v \leq n, E_{u,v}$ est l'événement "il y a une arête entre les sommets u et v", $(E_{u,v})$ est une famille d'événements mutuellement indépendants et $P(E_{u,v}) = p$.

Il y a au plus $M = \binom{n}{2}$ arêtes dans un graphe à n sommets et il y a 2^M graphes possibles.

Dans la suite, soit $G_{n,p}$ un tel graphe.

Dans $\mathcal{G}(n,p)$,

- le graphe complet a probabilité p^M,
- le graphe vide a probabilité $(1-p)^M$,
- la probabilité que $G_{n,p}$ ait m arêtes est $\binom{M}{m} p^m (1-p)^{M-m}$.

On veut étudier les comportement de certaines propriétés des graphes quand le nombre de sommets croît vers l'infini et

1. p est fixé
2. p est une fonction de n

Pour la propriété A, une **fonction seuil** est une fonction $g(n)$ telle que

(i) Si $p << g$ $(p = o(g))$ alors $\lim(P(G_{n,p(n)} \text{ satisfait } A)) = 0$

(ii) Si $p >> g$, cette limite vaut 1

Variance d'une somme de v.a. binaires **Lemme :** soient $Y_i \in \{0,1\}, i = 1,\ldots,m$ des variables aléatoires et $Y = \sum_{i=1}^m Y_i$. Alors $Var[Y] \leq E[Y] + \sum_{1 \leq i,j \leq m; i \neq j} Cov(Y_i, Y_j)$

Démonstration : $Var[Y] = \sum_{i=1}^m Var[Y_i] + \sum_{1 \leq i,j \leq m; i \neq j} Cov(Y_i, Y_j)$. Quand $Y_i \in \{0,1\}, E[Y_i^2] = E[Y_i]$ et donc $Var[Y_i] = E[Y_i^2] - E[Y_i]^2 \leq E[Y_i]$. D'où le résultat souhaité.

Exemple: clique de taille 4 **Thoérème :** si A = "contenir une clique de taille 4", alors la fonction seuil est $g(n) = n^{-2/3}$. Plus précisément,

- si $p(n) << n^{-2/3}$, alors $\lim_{n\to\infty} P(G_{n,p(n)} \text{ satisfait } A) = 0$
- si $p(n) >> n^{-2/3}$, alors $\lim_{n\to\infty} P(G_{n,p(n)} \text{ satisfait } A) = 1$

Démonstration : La première assertion se prouve en utilisant l'inégalité de Markov et la seconde en utilisant la méthode du second moment.

Soit $C_1, \ldots, C_{\binom{n}{4}}$ une énumération des ensembles de 4 sommets et définissions les variables aléatoires $X_i \in \{0,1\}, i \in \{1, \ldots, \binom{n}{4}\}$ avec $X_i = 1 \Leftrightarrow C_i$ est une clique de taille 4.

On pose $X = \sum_i X_i$. Nous avons $E[X_i^2] = E[X_i], \forall i$ et

- $E[X] = \sum_i E[X_i] = \binom{n}{4} p(n)^6 = (\frac{1}{24} n^4 + o(n^4)) p(n)^6$
- $Var[X] \leq E[X] + \sum_{1 \leq i,j \leq m; i \neq j} Cov(X_i, X_j)$

Calcul de $Cov(X_i, X_j)$ avec $i \neq j$:

- Si $|C_i \cap C_j| = 0$ alors les deux cliques sont disjointes et X_i et X_j sont indépendants. Donc $Cov(X_i, X_j) = 0$.
- Si $|C_i \cap C_j| = 1$ alors X_i et X_j sont également indépendants et $Cov(X_i, X_j) = 0$.
- Si $|C_i \cap C_j| = 2$ alors les deux cliques partagent une arête. 11 arêtes doivent être dans le graphe (2 cliques de taille 4 qui se superposent sur 2 sommets), d'où $Cov(X_j, X_j) = E[X_i X_j] - E[X_i]E[X_j] \leq E[X_i X_j] \leq p^{11}$.Il y a $\binom{n}{6}$ possibilités de choisir les 6 sommets et $\binom{6}{2;2;2}$ possibilités de les séparer en C_i et C_j.
- Si $|C_i \cap C_j| = 3$, alors les deux cliques partagent 3 arêtes. 9 arêtes doivent être dans le graphe. Donc $Cov(X_i, X_j) \leq E[X_i X_j] \leq p^9$. Il y a $\binom{n}{5}$ possibilités de choisir les 5 sommets et $\binom{5}{3;1;1}$ possibilités de les séparer en C_i et C_j.

Nous avons donc : $Var[X] \leq \binom{n}{4} p(n)^6 + \binom{n}{6}\binom{6}{2;2;2} p(n)^{11} + \binom{n}{5}\binom{5}{3;1;1} p(n)^9$.

Maintenant :

- Si $p(n) = o(n^{-2/3})$, alors par Markov, $P(X \neq 0) \leq E[X] = \binom{n}{4} p(n)^6 = (\frac{1}{24} n^4 + o(n^4)) p(n)^6 = o(1)$.
- Si $n^{-2/3} = o(p(n))$, alors $Var[X] \leq \binom{n}{4} p(n)^6 + \binom{n}{6}\binom{6}{2;2;2} p(n)^{11} + \binom{n}{5}\binom{5}{3;1;1} p(n)^9 = o(n^8 p(n)^{12}) = o(E[X]^2)$, puisque $E[X]^2 = (\binom{n}{4} p(n)^6)^2 = \Theta(n^8 p(n)^{12})$. Ainsi, $P(X = 0) \leq \frac{Var[X]}{E[X]^2} = o(1)$.