Probabilités et applications

transcrit par Emmanuel Hainry
d'après le cours de Pascal Koiran

13 mai 2001

Chapitre 1

Probabilités discrètes

Soit Ω un ensemble fini d'événements élémentaires. Pour chaque $\omega \in \Omega$, $p(\omega) \geq 0$. Soit $A \subseteq \Omega$, on pose

$$P(A) = \sum_{\omega \in A} p(\omega)$$

Exemple 1 *Un dé pipé : $p(6) = \frac{4}{12}, p(1) = p(2) = \frac{1}{12}, p(3) = p(4) = p(5) = \frac{2}{12}$. La probabilité d'obtenir un nombre pair est alors $P(\{2,4,6\}) = \frac{7}{12}$.*

Exemple 2 *tirer au sort un réel appartenant à l'intervalle* $]0;1]$

Exemple 3 *tirage à pile ou face : un événement élémentaire est représenté par une suite de P et de F.*

Revenons à l'exemple 2. $\Omega =]0;1]$. Soit $A = \bigcup_{i=1}^{n}]a_i;b_i] \subseteq \Omega$ avec $\forall i, a_{i+1} > b_i$ (les intervalles sont alors distincts). On pose

$$P(A) = \sum_{i=1}^{n} (b_i - a_i)$$

On a alors $P(\Omega) = 1$ et $P(A \cup B) = P(A) + P(B)$ si A et B sont disjoints.
On peut alors se ramener à l'exemple 3 en ramenant une suite à l'écriture binaire d'un nombre décimal de $]0;1]$.

$$\omega = \sum_{n=1}^{+\infty} \frac{d_n(\omega)}{2^n} \text{ avec } d_n(\omega) \in \{0,1\}$$

Par convention, les nombres dont l'écriture est ambiguë seront représentés par leur écriture se terminant par un nombre infini de 1. Par exemple $\frac{1}{2} = 0,011111111......$ Posons $u_i = d_i(\omega)$ on a alors $\omega \in] \sum_{i=1}^{n} \frac{u_i}{2^i}, \sum_{i=1}^{n} \frac{u_i}{2^i} + \frac{1}{2^n}]$
exemple : $P(\omega; \sum_{i=1}^{n} d_i(\omega) = k) = \frac{C_n^k}{2^n}$ car $\{\omega; \sum_{i=1}^{n} d_i(\omega) = k\}$ est l'union de C_n^k intervalles disjoints de longueur $\frac{1}{2^n}$.

1.1 Loi faible des grands nombres

$$\boxed{\forall \varepsilon > 0, \lim_{n \to +\infty} P \underbrace{\left(\left\{ \omega; \left| \frac{1}{n} \sum_{i=1}^{n} d_i(\omega) - \frac{1}{2} \right| \geq \varepsilon \right\} \right)}_{\text{union finie d'intervalles}} = 0}$$

Démonstration

Soit $f : \Omega \to \mathbb{R}$ une fonction en escalier. c'est à dire telle que $\forall x \in]a_i; a_{i+1}], f(x) = c_i$.
$\{\omega; f(\omega) \geq \alpha\}$ est une réunion finie d'intervalles.
Soit $r_i(\omega) = 2d_i(\omega) - 1$ i.e. $r_i(\omega) = \begin{cases} -1 \text{ si } d_i(\omega) = 0 \\ +1 \text{ si } d_i(\omega) = 1 \end{cases}$
Puis $S_n(\omega) = \sum_{i=1}^{n} r_i(\omega)$. On veut donc démontrer que $\forall \varepsilon > 0, \lim_{n\to+\infty} P\left(\omega; \frac{1}{n}|S_n(\omega)| \geq \varepsilon\right) = 0$

Lemme

Soit $f : \Omega \to \mathbb{R}^+$ une fonction en escalier. Soit $\alpha > 0$.

$$P(\omega; f(\omega) \geq \alpha) \leq \frac{1}{\alpha} \int_0^1 f(\omega) \mathrm{d}\omega$$

Preuve du lemme

$$f(x) = c_i \text{ sur }]a_i; a_{i+1}]$$

$$\begin{aligned} \int_0^1 f(\omega)\mathrm{d}\omega &= \textstyle\sum_{i=1}^{n} c_i(a_{i+1} - a_i) \\ &= \textstyle\sum_{c_i \geq \alpha} c_i(a_{i+1} - a_i) + \sum_{c_i < \alpha} c_i(a_{i+1} - a_i) \\ &\geq \alpha \textstyle\sum_{c_i \geq \alpha} a_{i+1} - a_i \\ &\geq \alpha P(\omega; f(\omega) \geq \alpha) \end{aligned}$$

Suite de la démo

$$\begin{aligned} P\left(\omega; \tfrac{S_n(\omega)}{n} \geq \varepsilon\right) &= P\left(\omega; S_n^2(\omega) \geq n^2\varepsilon^2\right) \\ &\leq \tfrac{1}{n^2\varepsilon^2} \textstyle\int_0^1 S_n^2(\omega)\mathrm{d}\omega \end{aligned}$$

$$\text{Or } S_n(\omega) = \sum_{i=1}^{n} r_i(\omega) \text{ avec } r_i(\omega) \in \{-1; 1\}$$

$$\text{D'où } S_n^2(\omega) = \sum_{i=1}^{n} r_i^2(\omega) + 2\sum_{i<j} r_i(\omega) r_j(\omega)$$

$$\text{Donc } \int_0^1 S_n^2(\omega)\mathrm{d}\omega = n \text{ car } \int_0^1 r_i(\omega)r_j(\omega)\mathrm{d}\omega = 0 \text{ pour } i \neq j$$

$$\text{Et donc } P\left(\omega; \frac{|S_n(\omega)|}{n} \geq \varepsilon\right) \leq \frac{n}{n^2\varepsilon^2} = \frac{1}{n\varepsilon^2} \text{ qui tend vers 0 quand } n \text{ tend vers } +\infty.$$

1.2 Loi forte des grands nombres

Définition

$\omega \in \Omega$ est dit normal si $\lim_{n\to+\infty} \sum_{i=1}^{n} \frac{d_i(\omega)}{n} = \frac{1}{2}$
$A \subseteq \Omega$ est négligeable si $\forall \varepsilon > 0, \exists (I_k)_{k\geq 1}$ suite d'intervalles tels que $A \subseteq \bigcup_{k\geq 1} I_k$ et $\sum_{k\geq 1} P(I_k) \leq \varepsilon$

Exemple

Si $A \subseteq \Omega$ est dénombrable, A est négligeable. En effet, soit $A = \{a_1, a_2, ..., a_n, ...\}$. Soit I_n tel que $a_n \in I_n =]b_n, c_n]$ avec $P(I_n) = \frac{\alpha}{n^2}$. On a $A \subseteq \bigcup I_n$ et $\sum P(I_n) \leq \sum \frac{\alpha}{n^2} \leq \varepsilon$

Théorème

> Soit N l'ensemble des nombres normaux.
> N n'est pas une union finie d'intervalles.
> N^C, complémentaire de N, est négligeable.

$N = \left\{\omega; \lim \frac{S_n(\omega)}{n} = 0\right\}$. Or $P(\omega; |S_n(\omega)| \geq n\varepsilon) \leq \frac{1}{n\varepsilon^2}$

Lemme

$P(\omega; |S_n(\omega)| \geq n\varepsilon) \leq \frac{3}{n^2\varepsilon^4}$

Preuve de la loi forte

Soit $\varepsilon_n > 0$, $\varepsilon_n \longrightarrow 0$ avec $\sum \frac{1}{n^2\varepsilon_n^4}$ convergeant. Par exemple $\varepsilon_n = n^{-\frac{1}{8}}$
$A_n = \{\omega; |S_n(\omega)| \geq n\varepsilon_n\}$
$P(A_n) \leq \frac{3}{n^2\varepsilon^4}$ d'après le lemme.
$\forall m, N^C \subseteq \bigcup_{n=m}^{+\infty} A_n$
Soit ω tel que $\omega \notin A_n$ pour chaque $n \geq m$. Alors $\forall n \geq m, |S_n(\omega)| \leq n\varepsilon_n$ donc $\omega \in N$ Et pour m assez grand, $\sum_{n=m}^{+\infty} P(A_n) \leq \sum_{n=m}^{+\infty} \frac{3}{n^2\varepsilon^4} < \alpha$

Preuve du lemme

$P(\omega; |S_n(\omega)| \geq n\varepsilon) = P(\omega; S_n^4(\omega) \geq n^4\varepsilon^4) \leq \frac{1}{n^4\varepsilon^4} \int_0^1 S_n^4(\omega)\mathrm{d}\omega$
Or $S_n^4(\omega) = (\sum_{i=1}^n r_i(\omega))^4 = \sum_{\alpha,\beta,\gamma,\delta=1}^n r_\alpha(\omega) r_\beta(\omega) r_\gamma(\omega) r_\delta(\omega)$
Apparaissent les termes :

$(*)$	$r_i^4(\omega) = 1$
$(**)$	$r_i^3(\omega) r_j(\omega) = r_i(\omega) r_j(\omega)$
$(***)$	$r_i^2(\omega) r_j(\omega) r_k(\omega) = r_j(\omega) r_k(\omega)$
$(****)$	$r_i^2(\omega) r_j^2(\omega) = 1$
$(*****)$	$r_i(\omega) r_j(\omega) r_k(\omega) r_l(\omega)$

Les termes en $(**)$, $(***)$ et $(*****)$ disparaissent (intégrale nulle) et on obtient

$$\frac{1}{n^4\varepsilon^4} \int_0^1 S_n^4(\omega)\mathrm{d}\omega \leq \frac{3n(n-1)+n}{n^4\varepsilon^4} \leq \frac{3}{n^2\varepsilon^4}$$

Chapitre 2

Mesures de probabilités

On mesure une tour à son ombre et les grands hommes au nombre de leurs détracteurs.

Définition

Une famille $\mathcal{F}$ de parties de Ω est une σ-algèbre (ou "tribu") si

1. $\Omega \in \mathcal{F}$
2. si $A \in \mathcal{F}, \Omega - A \in \mathcal{F}$
3. $\mathcal{F}$ est close par union dénombrable

remarque : Une σ-algèbre est close par intersection dénombrable.

Exemple 4 *$\mathcal{F}$: ensemble des parties de Ω est une σ-algèbre*

Exemple 5 *$\mathcal{F}$: ensemble des unions finies d'intervalles de $\Omega =]0,1]$ n'est pas une σ-algèbre mais une algèbre.*

Exemple 6 *$\mathcal{A}$ famille de parties de Ω. Soit $\sigma(\mathcal{A})$ l'intersection de toutes les σ-algèbres qui contiennent les éléments de $\mathcal{A}$.*
$\sigma(\mathcal{A})$ est une σ-algèbre : c'est la σ-algèbre engendrée par $\mathcal{A}$.
Si $\Omega =]0,1]$, $\mathcal{A}$ l'ensemble des intervalles $]a,b] \subseteq \Omega$, $\sigma(\mathcal{A})$ est la σ-algèbre des boréliens[1] de $]0,1]$.

Exemples de boréliens[2]

1. $[a,b] = \bigcap_{n\geq 1}]a - \frac{1}{n}, b]$ est un borélien.
2. tout ouvert $O \subseteq \Omega =]0,1]$ où $O = \cup]a_n, b_n]$ avec $a_n, b_n \in \mathbb{Q}$ est un borélien.
3. tout fermé $F \subseteq \Omega$ est un borélien.
4. l'ensemble N des nombres normaux est un borélien.
   En effet $N = \bigcap_{k=1}^{\infty} \bigcup_{m=1}^{\infty} \bigcap_{n=m}^{\infty} \underbrace{\left\{\omega; \left|\frac{S_n(\omega)}{n}\right| < \frac{1}{k}\right\}}_{\text{union finie d'intervalles}}$.
   Pour k et m fixés, $\bigcap_{n=m}^{\infty} \left\{\omega; \left|\frac{S_n(\omega)}{n}\right| < \frac{1}{k}\right\}$ est un borélien.

[1] Borel (Emile), mathématicien et homme politique français (Saint-Affrique 1871 - Paris 1956).

[2] borélien : Une partie d'un ensemble topologique E est dite borélienne, ou encore un borélien de E, si elle appartient à la tribu engendrée par les ouverts de E.

2.1 Mesures de probabilité

2.1.1 Définition

Soit $\mathcal{F}$ une algèbre de Ω. Une mesure de probabilité P associe à chaque $A \in \mathcal{F}$ un réel $P(A)$ de sorte que

1. $\forall A \in \mathcal{F}, 0 \leq P(A) \leq 1$
2. $P(\varnothing) = 0, P(\Omega) = 1$
3. additivité dénombrable : soit $(A_n)_{n\geq 1}$ une suite d'éléments de $\mathcal{F}$, alors $P(\bigcup A_n) = \sum_n^{+\infty} P(A_n)$ si les A_n sont disjoints.

2.1.2 Conséquences

Additivité finie

$$P\left(\bigcup_{k=1}^{n} A_k\right) = \sum_{k=1}^{n} P(A_k), \text{ si les } A_k \text{ sont disjoints.}$$

- Si $A \in \mathcal{F}$, $P(\Omega - A) = 1 - P(A)$ car $\Omega = A \cup (\Omega - A)$
- $A \subseteq B \Rightarrow P(A) \leq P(B)$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
  car $A \cup B = (A - B) \cup (B - A) \cup (A \cap B)$
  et donc $\begin{aligned} P(A \cup B) &= P(A - B) + P(B - A) - P(A \cap B) \\ &= P(A) - P(A \cap B) + P(B) - P(A \cap B) + P(A \cap B) \end{aligned}$

Sous-additivité

$$P\left(\bigcup_{k=1}^{n} A_k\right) \leq \sum_{k=1}^{n} P(A_k) \text{ si } \bigcup A_k \in \mathcal{F}$$

En effet $\bigcup_{k=1}^{n} A_k = A_1 \cup (A_2 - A_1) \cup (A_3 - (A_2 \cup A_1)) \cup \ldots \cup (A_n - \bigcup_{i=1}^{n-1} A_i)$
Or $A_2 - A_1 \subseteq A_2$, $(A_3 - (A_2 \cup A_1)) \subseteq A_3$, ... et $(A_n - \bigcup_{i=1}^{n-1} A_i) \subseteq A_n$
D'où le résultat.

2.1.3 Théorème

Soit $(A_n)_{n\geq 1}$ une suite d'éléments de $\mathcal{F}$ telle que $\bigcup_{n\geq 1} A_n \in \mathcal{F}$

1. si $\forall n, A_n \subseteq A_{n+1}$, alors $P(\bigcup_{k\geq 1} A_k) = \lim_{k\to+\infty} P(A_k)$
2. si $\forall n, A_n \supseteq A_{n+1}$, alors $P(\bigcap_{k\geq 1} A_k) = \lim_{k\to+\infty} P(A_k)$

Preuve de 1

soit $A = \bigcup_{n=1}^{+\infty} A_n$
$A = \underbrace{A_1 \cup (A_2 - A_1) \cup (A_3 - A_2) \cup \ldots \cup (A_n - A_{n+1})}_{\text{union disjointe}}$
donc $P(A) = \lim_{n\to+\infty} (P(A_1) + P(A_2 - A_1) + \ldots + P(A_n - A_{n-1})) = \lim_{n\to+\infty} P(A_n)$

Preuve de 2

raisonnement identique

2.1.4 Définition

A est négligeable si $\forall n, \exists A_n \in \mathcal{F}$ tel que $A \subseteq A_n$ et $P(A_n) \leq \frac{1}{n}$. Dans ce cas, si $A \in \mathcal{F}$, alors $P(A) = 0$

2.1.5 Exemple : probabilités discrètes

On prend Ω fini, $\Omega = \{a_1, a_2, ..., a_k\}$. On pose $P(\{a_i\}) = p_i$, avec $0 \leq p_i \leq 1$ et $\sum_{i=1}^{k} p_i = 1$ et pour $A \subseteq \Omega$, $P(A) = \sum_{a \in A} P(\{a\})$

2.2 Mesure de Lebesgue

Il existe une unique mesure de probabilité sur l'ensemble des boréliens de $]0,1]$ telle que pour $]a,b] \subset]0,1]$, $P(]a,b]) = b - a$. On l'appelle mesure de Lebesgue[3].

2.2.1 Théorème

Une mesure de probabilité P sur une algèbre $\mathcal{F}$ a une extension unique sur la σ-algèbre $\sigma(\mathcal{F})$ engendrée par $\mathcal{F}$.

Cas de la mesure de Lebesgue : $\mathcal{F}$ est l'algèbre des unions finies d'intervalles $]a,b] \subseteq]0,1]$. Si $A_n = \bigcup_{i=1}^{n}]a_i, b_i]$, on pose $P(A) = \sum_{i=1}^{n} b_i - a_i$.
Point essentiel : si $]a,b] = \bigcup_{n=1}^{+\infty}]a_n, b_n]$ (union disjointe), a-t-on $b - a = \sum_{k=1}^{+\infty} b_n - a_n$?
Question : Est-ce vrai pour $\mathbb{Q}$?
La réponse est non. Contre-exemple : $\mathbb{Q} = \{q_1, q_2, ..., q_n, ...\}$. On choisit $]a_n, b_n]$ tel que $\forall n q_n \in]a_n, b_n]$ et $b_n - a_n \leq \frac{1}{2^{n+1}}$. On a alors $]0,1] = \bigcup_{n \geq 1}]a_n, b_n]$ et $\sum_{n=1}^{+\infty} b_n - a_n = \frac{1}{2}$!
Conclusion : sur les intervalles de $\mathbb{Q}$, P n'est pas σ-additive.

2.2.2 Lemme

Si $]a,b] \subseteq \bigcup_k]a_k, b_k]$, union finie ou dénombrable d'intervalles, alors $b - a \leq \sum_k b_k - a_k$

Preuve

Le cas fini se démontre par récurrence.
Dans le cas dénombrable, on utilise un argument de compacité : fixons $\varepsilon > 0$. $[a+\varepsilon, b] \subseteq \bigcup_k \left] a_k, b_k + \frac{\varepsilon}{2^k} \right[$ donc $[a+\varepsilon, b] \subseteq \bigcup_{k=1}^{N} \left] a_k, b_k + \frac{\varepsilon}{2^k} \right[$
Et $b - (a + \varepsilon) \leq \sum_{k=1}^{N} (b_k + \frac{\varepsilon}{2^k} - a_k) \leq \left(\sum_{k=1}^{N} b_k - a_k \right) + \varepsilon$

[3] Lebesgue (Henri), mathématicien français (Beauvais 1875 - Paris 1941).

Chapitre 3

Variables Aléatoires

Il n'y a pas de hasard.
Voltaire, *Zadig*

Définition

On se donne Ω, $\mathcal{F}$ σ-algèbre sur Ω et P une mesure de probabilité sur $\mathcal{F}$. Une variable aléatoire est une fonction $X : \Omega \longrightarrow \mathbb{R}$ telle que $\forall x \in \mathbb{R}, \{\omega; X(\omega) \leq x\} \in \mathcal{F}$

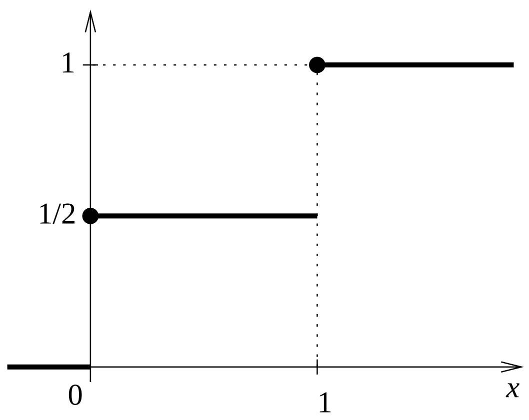
Exemple

$\Omega =]0,1]$, $\mathcal{F}$: le borélien de Ω. $X(\omega) = min\{n; d_n(\omega) = 0\}$. On posera $X(\omega) = 0$ si $\forall n, d_n(\omega) = 1$. On a alors $X : \Omega \longrightarrow \mathbb{N}$

3.1 Fonction de distribution

C'est la fonction $\begin{array}{cccc} F_X : & \mathbb{R} & \longrightarrow & [0,1] \\ & x & \mapsto & P(\{\omega; X(\omega) \leq x\}) \end{array}$

Exemple 7 $\Omega =]0,1]$, $d_n : \Omega \longrightarrow \{0,1\}$
$P(X(\omega) \leq x)$



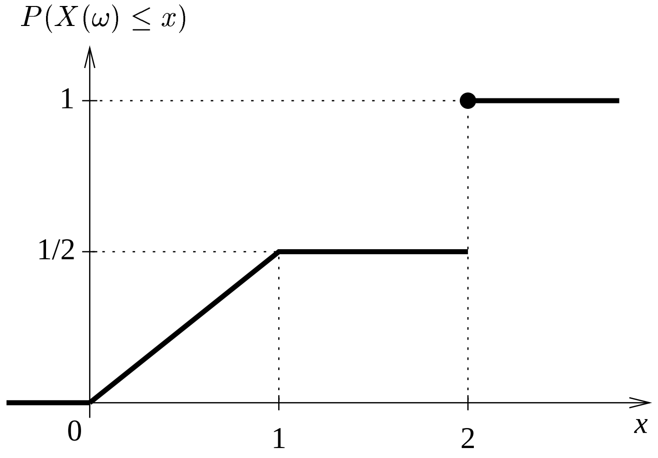
Exemple 8 $\Omega =]0,1] \cup \{2\}$. *Soient $\mathcal{B}$ l'ensemble des boréliens de $]0,1]$ et λ une mesure. On définit la σ-algèbre $\mathcal{F}$ comme suit :*

$$A \in \mathcal{F} \quad \begin{array}{rl} \textit{si} & A \in \mathcal{B} \\ \textit{ou si} & A = B \cup \{2\} \textit{ avec } B \in \mathcal{B} \end{array}$$

Si $A \in \mathcal{B}$, $\quad P(A) = \frac{\lambda(A)}{2}$
Si $A = B \cup \{2\}$, $\quad P(A) = \frac{1}{2} + \frac{\lambda(B)}{2}$
On a bien alors $P(\emptyset) = 0$, *et* $P(\Omega) = \frac{1}{2} + \frac{1}{2} = 1$
Additivité dénombrable : soit $A = \bigcup_n A_n$, *union disjointe finie ou dénombrable d'éléments de* $\mathcal{F}$, *montrons que l'on a* $P(A) = \sum_n P(A_n)$.

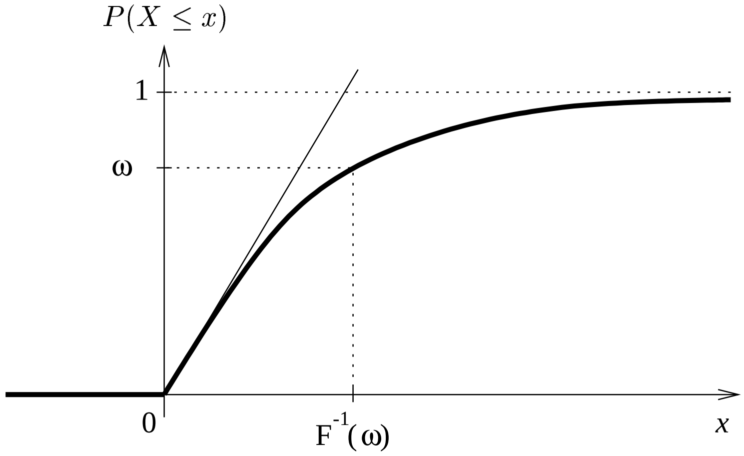
1^{er} *cas :* $\quad \forall n, 2 \notin A_n$
$\quad P(A) = \frac{\lambda(A)}{2} = \frac{1}{2}\sum_n \lambda(A_n) = \sum_n P(A_n)$

$2^{ème}$ *cas :* $\quad 2 \in A_1$
$\quad A_1 = B_1 \cup \{2\}$ *avec* $B_1 \in \mathcal{B}$
$\quad P(A) = \frac{1}{2} + \frac{1}{2}\lambda\left(B_1 \cup \bigcup_{n\geq 2} A_n\right) = \frac{1}{2} + \frac{1}{2}\sum_{n\geq 2}\lambda(A_n) + \frac{1}{2}\lambda(B_1)$
$\quad P(A) = P(A_1) + \frac{1}{2}\sum_{n\geq 2} P(A_n)$



Exemple 9 *distribution exponentielle (modélisation du temps d'attente de l'arrivée du prochain appel dans un central téléphonique).*

$$P(\{\omega; X(\omega) \leq x\}) = \begin{cases} 0 & \textit{si } x < 0 \\ 1 - e^{-\alpha x} & \textit{si } x \geq 0 \end{cases}$$



On prend $\Omega =]0,1[$ *et* $\mathcal{F}$ *le borélien de* Ω. *Soit* λ *la mesure de Lebesgue.*
$X(\omega) = F^{-1}(\omega)$
Donc $P(\{\omega; X(\omega) \leq x\}) = 0$ *si* $x \leq 0$

$$\text{Si } x \geq 0, X(\omega) \leq x \quad \begin{aligned} &\Leftrightarrow F^{-1}(\omega) \leq x \\ &\Leftrightarrow \omega \leq F(x) \end{aligned}$$

Et $P(\{\omega; \omega \leq F(x)\}) = \lambda(]0, F(x)]) = F(x)$

3.1.1 Propriétés

Soit $F : \begin{array}{ccc} \mathbb{R} & \to & [0,1] \\ x & \mapsto & P(\omega; X(\omega) \leq x) \end{array}$

(i) F est croissante

(ii) $\lim_{x\to-\infty} F(x) = 0$ et $\lim_{x\to+\infty} F(x) = 1$
En effet $\lim_{x\to-\infty} P(\omega, X(\omega) \in]-\infty, x]) = P\left(\bigcap_k \{\omega, X(\omega) \in]-\infty, -k]\}\right) = P(\varnothing) = 0$

(iii) F est continue à droite.
démonstration : soit $(x_n) \to x$ avec $\forall n, x_n > x$

$$lim_{n\to+\infty} P(X \in]-\infty, x_n]) = P(X \in \underbrace{\bigcap_n]-\infty, x_n]}_{]-\infty,x]}) = P(X \in]-\infty, x])$$

3.1.2 Théorème

> Une fonction $F : \mathbb{R} \to [0,1]$ est une fonction de distribution si et seulement si elle vérifie les propriétés (i), (ii) et (iii).

Question

$X(\omega) = inf\{x; F(x) \geq \omega\}$. Pour quels $A \subseteq \mathbb{R}$ a-t-on $X^{-1}(A) \in \mathcal{F}$? Pour de tels A, $P(X(\omega) \in A)$ est en effet bien définie. On sait que $\forall a, b, X^{-1}(]a,b]) \in \mathcal{F}$ car $X^{-1}(]a,b]) = X^{-1}(]-\infty, b]) - X^{-1}(]-\infty, a])$

3.2 Proposition

Si $X : \Omega \to \mathbb{R}$ est une variable aléatoire, $X^{-1}(A) \in \mathcal{F}$ pour tout borélien $A \subseteq \mathbb{R}$.

Preuve

Soit $\mathcal{F}'$ l'ensemble des $A \subseteq \mathbb{R}$ tels que $X^{-1}(A) \in \mathcal{F}$. Tous les intervalles de la forme $]a,b]$ sont dans $\mathcal{F}'$. $\mathcal{F}'$ est une σ-algèbre car

- $\varnothing \in \mathcal{F}'$ et $\mathbb{R} \in \mathcal{F}'$
- si $A \in \mathcal{F}'$, alors $X^{-1}(\mathbb{R} - A) = \underbrace{\Omega - \underbrace{X^{-1}(A)}_{\in\mathcal{F}}}_{\in\mathcal{F}}$
- soit $(A_n) \in \mathcal{F}'^{\mathbb{N}}, X^{-1}(\bigcup_{n\geq1} A_n) = \bigcup_{n\geq1} \underbrace{X^{-1}(A_n)}_{\in\mathcal{F}}$ donc $\bigcup_{n\geq1} A_n \in \mathcal{F}'$

Donc $\mathcal{F}'$ contient les boréliens.

3.2.1 Définition

Soient $X : \Omega \to \mathbb{R}$, $A \subseteq \mathbb{R}$ un borélien, on note $Q(A) = P(\omega; X(\omega) \in A)$. Q est une mesure de probabilité sur l'ensemble des boréliens de $\mathbb{R}$, appelée mesure image de P par X.

3.2.2 Remarque

Q est uniquement déterminé par la fonction de distribution F_X de X.
En effet, par définition :

$$
\begin{aligned}
Q(]-\infty, x]) &= F_X(x) \\
Q(]a, b]) &= F_X(b) - F_X(a) \\
Q(]a, b[) &= Q\left(\bigcup_n]a, b_n]\right) \text{ avec } (b_n) \text{ suite croissante tendant vers } b \\
&= \lim_{n \to +\infty} Q(]a, b_n]) \\
&= \lim_{n \to +\infty} F_X(b_n) - F_X(a) \\
&= F_X(b^-) - F_X(a)
\end{aligned}
$$

Chapitre 4

Indépendance

Vive le Québec libre.
Charles de Gaulle

On se donne $(\Omega, \mathcal{F}, P)$.

4.1 Définition

Deux événements A et B de $\mathcal{F}$ sont indépendants si $P(A \cap B) = P(A)P(B)$

Généralisation

Soit (A_n) une suite d'événements de $\mathcal{F}$. Ils sont indépendants si $P(A_{i_1} \cap A_{i_2} \cap ... \cap A_{i_k}) = \prod_{j=1}^{k} P(A_{i_j})$ lorsque les i_j sont distincts.

4.2 Définition

Deux variables aléatoires $X, Y : \Omega \to \mathbb{R}$ sont indépendantes si les événements $X \in A, Y \in B$ sont indépendants quel que soit le choix des boréliens A et B dans $\mathbb{R}$. C'est à dire si

$$P(\omega; X(\omega) \in A \text{ et } Y(\omega) \in B) = P(\omega; X(\omega) \in A)P(\omega; Y(\omega) \in B)$$

Généralisation

Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires. Elles sont indépendantes si $P(\omega; X_1(\omega) \in A_1, X_2(\omega) \in A_2, ..., X_n(\omega) \in A_n) = \prod_{i=1}^{n} P(\omega; X_i(\omega) \in A_i)$

4.3 Exemple

Soient $\Omega =]0,1]$ et $\mathcal{F}$ l'ensemble des boréliens de Ω.
d_n et d_m sont indépendants si $n \neq m$
$P(\underbrace{\{\omega; d_n(\omega) = 1\}}_{\text{union de } 2^{n-1} \text{intervalles de longueur } \frac{1}{2^n}}) = \frac{1}{2} = P(\omega; d_n(\omega) = 0)$
Soient $\varepsilon, \varepsilon' \in \{0,1\}$, $P(\omega; d_m(\omega) = \varepsilon \text{ et } d_n(\omega) = \varepsilon') = \frac{1}{4}$

Chapitre 5

Espérance

Tout le malheur des hommes vient de l'espérance.
Albert Camus

5.1 Définition

Soit $X : \Omega \to \mathbb{R}$ une variable aléatoire discrète. Si $X(\Omega)$ est dénombrable, on pose $A = X(\Omega)$. Si la série $\sum_{a \in A} |a| P(X = a)$ converge, on dit que l'espérance de X existe et on la note

$$E(X) = \sum_{a \in A} aP(X = a)$$

remarque : si $A \subseteq \mathbb{R}^+$, on dit que $E(X) = +\infty$ si la série diverge.

5.2 Espérance d'une somme

Si $E(X)$ et $E(Y)$ existent, alors $E(X + Y) = E(X) + E(Y)$

Preuve

$X(\Omega) = A = \{a_1, a_2, ..., a_n, ...\}$, $Y(\Omega) = B = \{b_1, b_2, ..., b_n, ...\}$

$$\begin{aligned} E(X+Y) &= \textstyle\sum_{i,j} (a_i + b_j) P(X = a_i \text{ et } Y = b_j) \\ &= \textstyle\sum_{i,j} a_i P(X = a_i \text{ et } Y = b_j) + \sum_{i,j} b_j P(X = a_i \text{ et } Y = b_j) \\ &= \textstyle\sum_i a_i P(X = a_i) + \sum_j b_j P(Y = b_j) \\ &= E(X) + E(Y) \end{aligned}$$

5.3 Espérance d'un produit

L'espérance d'un produit n'est pas toujours le produit des espérances.

5.3.1 Exemple

Soient $\Omega = \{1,2,3,4,5,6\}$, $X(\omega) = \omega$ et P la distribution uniforme.

$$E(X)^2 = \left(\frac{1+2+3+4+5+6}{6}\right)^2 = \frac{49}{4}$$

$$E(X^2) = \frac{1+4+9+16+25+36}{6} = \frac{91}{6}$$

5.3.2 Proposition

Si $X, Y : \Omega \to \mathbb{R}$ sont des variables aléatoires indépendantes, alors si $E(X)$ et $E(Y)$ existent, $E(XY) = E(X)E(Y)$

Preuve

$$\begin{aligned} E(XY) &= \sum_{i,j} a_i b_j P(X = a_i \text{ et } Y = b_j) \\ &= \sum_{i,j} a_i b_j P(X = a_i)P(Y = b_j) \\ &= (\sum_i a_i P(X = a_i))(\sum_j b_j P(Y = b_j)) \end{aligned}$$

5.4 Composition des variables aléatoires

Si X et Y sont des variables, $X + Y$ et XY sont-elles des variables aléatoires ?

5.4.1 Lemme

$X : \Omega \to A$ avec A dénombrable est une variable aléatoire si et seulement si $X^{-1}(\{a\}) \in \mathcal{F}$ pour tout $a \in A$.

5.4.2 Application

$X^{-1}(]-\infty, b]) = \bigcup_{x \in A \cap]-\infty,b]} X^{-1}(\{x\})$
Donc $X^{-1}(]-\infty, b]) \in \mathcal{F}$ si on suppose que pour tout $x \in A$, $X^{-1}(\{x\}) \in \mathcal{F}$.
$\Sigma = X + Y$ est donc bien une variable aléatoire.
Soit $a \in A$, $\Sigma^{-1}(\{a\}) = \{\omega; X(\omega) + Y(\omega) = a\} = \bigcup_{b+c=a, b\in A, c\in A} \{\omega; X(\omega) = b \text{ et } Y(\omega) = c\}$. Donc $\Sigma^{-1}(\{a\}) \in \mathcal{F}$.

5.4.3 Généralement

Soient X et Y des variables aléatoires à valeurs dans A dénombrable. $Z : \omega \mapsto \varphi(X(\omega), Y(\omega))$ est une variable aléatoire pour toute fonction $\varphi : \mathbb{R}^2 \to \mathbb{R}$.

5.4.4 Proposition

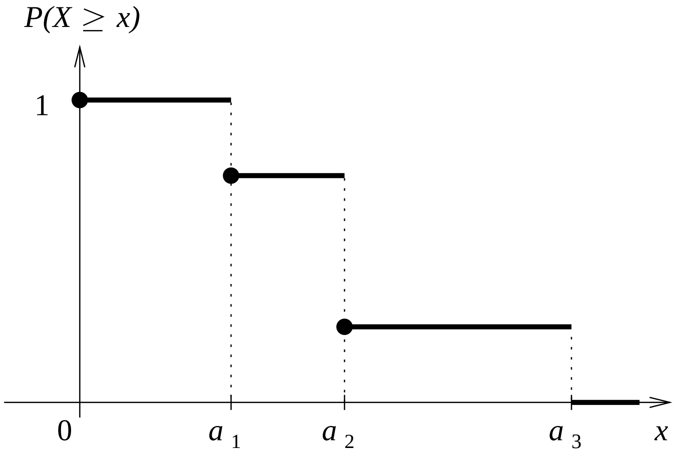
Si $X, Y : \Omega \to \mathbb{R}$ sont des variables aléatoires, $\varphi(X, Y)$ est une variable aléatoire quand φ est continue.

5.5 Autre expression de l'espérance

Si $X : \Omega \to A$ avec $A \subset \mathbb{R}$ fini, alors

$$E(X) = \int_0^{+\infty} P(X \geq x)\mathrm{d}x$$

5.5.1 exemple



5.5.2 Preuve

On suppose $A = \{a_1, a_2, ..., a_n\}$. On a alors $\forall i \leq n-1, P(X = a_i) = P(X \geq a_i) - P(X \geq a_{i+1})$

$$\begin{array}{rcl}
E(X) & = & \sum_{i=1}^{n} a_i P(X = a_i) \\
& = & a_1(P(X \geq a_1) - P(X \geq a_2)) \\
& + & a_2(P(X \geq a_2) - P(X \geq a_3)) \\
& + & a_3(P(X \geq a_3) - P(X \geq a_4)) \\
& \vdots & \\
& + & a_{n-1}(P(X \geq a_{n-1}) - P(X \geq a_n)) \\
& + & a_n P(X \geq a_n) \\
& = & a_1 P(X \geq a_1) \\
& + & (a_2 - a_1)P(X \geq a_2) \\
& + & (a_3 - a_2)P(X \geq a_3) \\
& \vdots & \\
& + & (a_n - a_{n-1})P(X \geq a_n)
\end{array}$$

5.5.3 Cas particulier

Si $A = X(\Omega) \subseteq \mathbb{N}$ et est fini alors $E(X) = \sum_{n \geq 1} P(X \geq n)$.

5.5.4 Proposition

Soit $X : \Omega \to \mathbb{N}$ une variable aléatoire. Si $E(X)$ existe, alors $E(X) = \sum_{n=1}^{+\infty} P(X \geq n)$

Exemple

soient $X_n : \Omega \to \{0,1\}$ des variables aléatoires indépendantes. On suppose que chaque X_n suit la distribution de Bernoulli[1] de paramètre p, ce qui signifie que $P(X_n = 1) = p$ et $P(X_n = 0) = 1 - p$.
Soit T la variable "instant du premier pile". $T(\omega) = min\{n; X_n(\omega) = 1\}$.
On a alors $\forall x \geq 1, P(T = x) = (1-p)^{x-1}p$ (distribution géométrique de paramètre p).

$$\begin{array}{ll} E(T) & = \sum_{x=1}^{+\infty} xp(1-p)^{x-1} \\ E(T) & = \sum_{x=1}^{+\infty} P(T \geq x) \\ & = \sum_{x=1}^{+\infty} (1-p)^{x-1} \\ & = \frac{1}{p} \end{array}$$

Preuve de la proposition

$$E(X) = \lim_{n \to +\infty} E_n \text{ avec } E_n = \sum_{t=1}^{n} tP(X = t)$$

$$\begin{array}{rcl} E_n & = & P(X=1) + 2P(X=2) + ... + nP(X=n) \\ E_n & = & P(X \geq 1) - P(X \geq 2) \\ & + & 2(P(X \geq 2) - P(X \geq 3)) \\ & + & 3(P(X \geq 3) - P(X \geq 4)) \\ & \vdots & \vdots \\ & + & n(P(X \geq n) - P(X \geq n+1)) \\ & = & P(X \geq 1) + P(X \geq 2) + ... + P(X \geq n) \\ & - & nP(X \geq n+1) \end{array}$$

On veut montrer que $\lim_{n \to +\infty} nP(X \geq n+1) = 0$.

$$nP(X \geq n+1) \leq (n+1)P(X \geq n+1) \leq \sum_{t=n+1}^{+\infty} tP(X = t) \longrightarrow_{n \to +\infty} 0$$

Car $E(X) < +\infty$ (en effet, si $\sum_{n=1}^{+\infty} P(X \geq n) < +\infty$ alors $E(X) < +\infty$).

Lemme

Soit $a_n > 0$ une suite décroissante de réels. Si $\sum a_n$ converge, alors $na_n \longrightarrow 0$
On l'applique à $a_n = P(X \geq n)$ et on obtient

$$E(X) = \int_0^{+\infty} P(X \geq x)\mathrm{d}x \text{ si } X \geq 0$$

On en déduit l'espérance d'une variable aléatoire exponentielle : $P(X \leq 0) = 0$ et $P(X \geq x) = e^{-\alpha x}$

$$E(X) = \int_0^{+\infty} e^{-\alpha x}\mathrm{d}x = \frac{1}{\alpha}$$

[1] Bernoulli, famille de mathématiciens originaire d'Anvers et qui, vers la fin du XVI[ème] siècle se réfugia à Bâle.

Chapitre 6

Quicksort

Pour trier un tableau S par l'algorithme de quicksort :

1. choisir un pivot $y \in S$
2. construire $S_1 = \{x \in S; x < y\}$ et $S_2 = \{x \in S; x > y\}$
3. retourner $tri(S_1), y, tri(S_2)$

Il y a plusieurs possibilités pour le choix du pivot :

– on peut prendre pour y le premier élément du tableau. Si le tableau est déjà trié, on fait $\Theta(n^2)$ opérations.

– on peut prendre la médiane comme pivot.
$C(n) = cn + 2C(\frac{n}{2})$
$C(n) = \Theta(n \log(n))$

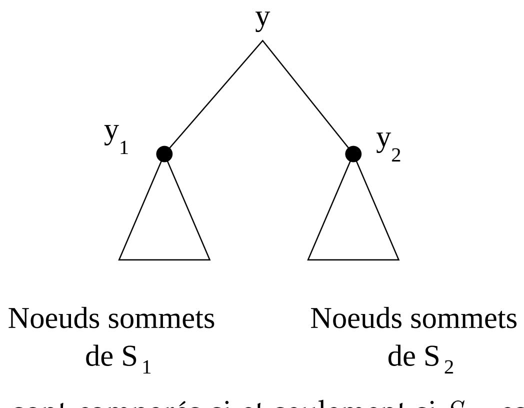
– on peut tirer au sort le pivot (distribution uniforme sur les n entiers du tableau).
$E(C) = O(n \log(n))$

Soit $S_{(i)}$ l'élément d'ordre i de S ($S_{(1)}$ est le plus petit, $S_{(n)}$ le plus grand).
Soit C le nombre de comparaisons.

$C = \sum_{i=1}^{n} \sum_{j>i} X_{i,j}$ avec $X_{i,j} = \begin{cases} 1 \text{ si } S_{(i)} \text{ et } S_{(j)} \text{ sont comparés} \\ 0 \text{ sinon} \end{cases}$

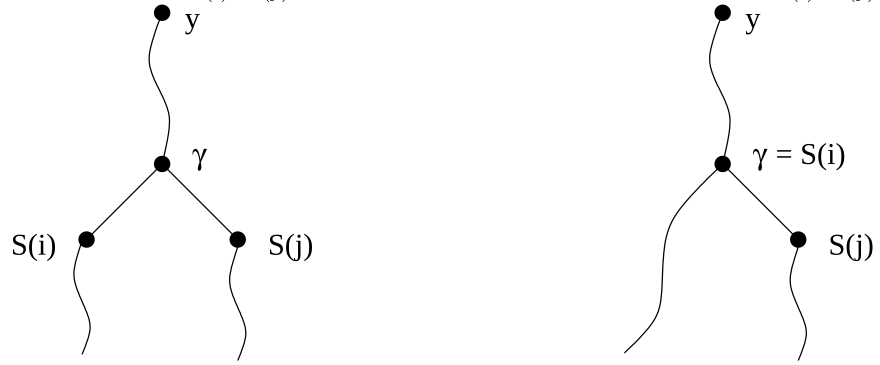
$$\begin{aligned} E(C) &= \sum_{i=1}^{n} \sum_{j=i+1}^{n} E(X_{i,j}) \\ &= \sum_{i=1}^{n} \sum_{j=i+1}^{n} p_{i,j} \text{ avec } p_{i,j} \text{ la probabilité que } S_{(i)} \text{ et } S_{(j)} \text{ soient comparés.} \end{aligned}$$

A chaque déroulement possible de l'algorithme, on associe un arbre. Les sommets sont les éléments de S. La racine est le pivot.



Observation : $S_{(i)}$ et $S_{(j)}$ sont comparés si et seulement si $S_{(i)}$ est un descendant de $S_{(j)}$ (ou vice-versa). Autrement dit, si on appelle γ le sommet le plus haut dans l'arbre étiqueté par un élément x de $[S_{(i)}, ..., S_{(j)}]$, $S_{(i)}$ et $S_{(j)}$ sont comparés si $\gamma \in \{S_{(i)}, S_{(j)}\}$.

Cas où $\gamma \notin \{S_{(i)}, S_{(j)}\}$ Cas où $\gamma \in \{S_{(i)}, S_{(j)}\}$



$p_{i,j}$ est donc la probabilité que $\gamma \in \{S_{(i)}, S_{(j)}\}$, donc $P_{i,j} = \frac{2}{j-i+1}$

$$\begin{aligned}
\text{Donc} \quad E(C) &= \textstyle\sum_{i=1}^{n} \sum_{j=i+1}^{n} \frac{2}{j-i+1} \\
&\leq \textstyle\sum_{i=1}^{n-1} \left(\frac{2}{2} + \frac{2}{3} + ... + \frac{2}{n-i+1} \right) \\
&\leq 2(n-1) \underbrace{\left(\frac{1}{2} + ... + \frac{1}{n} \right)}_{\ln(n)+\Theta(n)} \\
E(C) &\leq 2n\ln(n) + O(n)
\end{aligned}$$

Algorithme

```
Procédure quicksort(l, r : integer) ;
var
      v, t, i, j : integer ;
Begin
      if r>=l then begin
            v := a[r] ;
            i := l-1 ;                              {a[r] est le pivot }
            j := r ;
            repeat
                  repeat i := i+1 until a[i] >= v ;
                  repeat j := j+1 until a[j] <= v ;
                  t := a[i] ;
                  a[i] := a[j] ;
                  a[j] := t ;                       {échange de a[i] et a[j]}
            until j < i
            a[j] := a[i] ;                          {on annule le dernier échange }
            a[i] := a[r] ;
            a[r] := t ;
            quicksort(l,i-1) ;                      {le pivot se retrouve en position i}
            quicksort(i+1, r) ;
      end ;
end ;
```

quicksort(1, n) trie le tableau : $a[0]$ est plus petit que tous les $a[i]$ pour $i \in \{1, ..., n\}$.

Récurrence

Soit C_n le nombre moyen de comparaisons pour un tableau de n éléments.
$C_n = \underbrace{n+1}_{\text{étape de positionnement}} + \frac{1}{n}\sum_{j=1}^{n}(C_{j-1} + C_{n-j})$ pour $n \geq 0$.
Et le pivot a une distribution uniforme sur l'ensemble des n valeurs possibles. Soit j l'indice du pivot. Pour le tableau de gauche, on a une distribution uniforme sur l'ensemble des $(j-i)!$ permutations. De même pour le tableau de droite.

$$C_n = n + 1 + \frac{2}{n}\sum_{j=1}^{n} C_{j-1}$$

$$C_n = (n+1)\left(1 + 2\sum_{k=3}^{n+1}\frac{1}{k}\right)$$

Permutation aléatoire du tableau

```
for i := 2 to N do
begin
        t := randominteger(1, i) ;          {tire un entier au hasard entre 1 et i}
        v := a[t] ;
        a[t] := a[i] ;
        a[i] := v ;
end ;
```

Invariant de boucle

Après l'itération i, $a[1..i]$ contient une permutation aléatoire (distribution uniforme) des i premers éléments du tableau de départ.

Nouvelle version du quicksort probabiliste

On commence par permuter aléatoirement les éléments de a. Puis on effectue un quicksort déterministe. L'espérance est la complexité en moyenne du quicksort déterministe : $O(n\log(n))$.

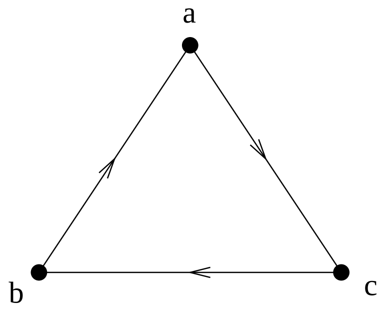
Chapitre 7

Tournois

7.1 Définitions

Un tournoi $T = (V, E)$ est un graphe orienté où V est l'ensemble des sommets (représentant les "joueurs"), T est sans boucle. $\forall (x, y) \in V^2, x \neq y \Rightarrow (x, y) \in E$ ou exclusif $(y, x) \in E$

T a la propriété S_k si pour tout ensemble de k sommets, il existe un sommet qui les bat tous.



T a la propriété S_1 mais pas S_2.

Question

Existe-t-il pour tout k un tournoi ayant la propriété S_k ?

7.2 Théorème

$\boxed{\text{Si } C_n^k \left(1 - \frac{1}{2^k}\right)^{n-k} < 1\text{, il existe un tournoi à } n \text{ joueurs qui a la propriété } S_k.}$

Preuve

Soit $V = \{1, ..., n\}$. Soit Ω l'ensemble des tournois sur V. On a donc $|\Omega| = 2^{\frac{n(n-1)}{2}}$. Soit P une distribution uniforme.

$P(\omega$ a la propriété $S_k) > 0$

Soit K un ensemble de k sommets. Soit $j \notin K$. $P(j$ bat tous les sommets de $K) = \frac{1}{2^k}$

$P(K$ n'est battu par aucun autre sommet $) = \left(1 - \frac{1}{2^k}\right)^{n-k}$ (intersection d'événements indépendants)

$P(\exists K; K$ n'est battu par aucun autre sommet $) \leq C_n^k \left(1 - \frac{1}{2^k}\right)^{n-k}$

7.3 Théorème

Il existe un tournoi de n joueurs avec au moins $\frac{n!}{2^{n-1}}$ chemins hamiltoniens.

Rappel

Un chemin hamiltonien passe une fois et une seule par tous les sommets du graphe. On identifie un chemin hamiltonien à une permutation σ de $\mathbb{N}_n$ telle que $\forall i \in \mathbb{N}_{n-1}, (\sigma(i), \sigma(i+1)) \in E$.

Preuve du théorème

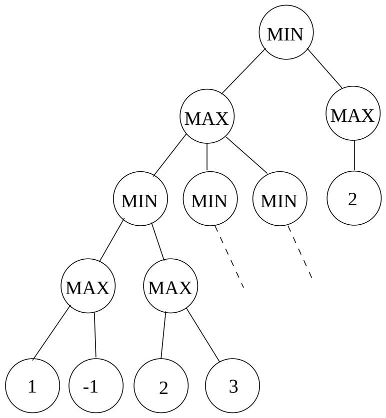
Soit $X : \Omega \to \mathbb{N}$ avec $X(\omega)$ le nombre de chemins hamiltoniens de ω.
$X = \sum_\sigma X_\sigma$ avec $X_\sigma(\omega) = \begin{cases} 1 \text{ si } \sigma \text{ correspond à un chemin hamiltonien} \\ 0 \text{ sinon} \end{cases}$
$E(X) = \sum_\sigma E(X_\sigma) = \sum_\sigma P(X_\sigma = 1) = \frac{n!}{2^{n-1}}$
Il existe ω tel que $X(\omega) \geq E(X)$ car $E(X) = \frac{1}{|\Omega|} \sum_{\omega \in \Omega} X(\omega)$
Il y a donc au moins un $\omega \in \Omega$ tel que $X(\omega) \geq E(X)$

Chapitre 8

Arbres de jeux

8.1 Définition

Les feuilles d'un arbre de jeu sont étiquetées par une valeur réelle. Les nœuds internes à distance paire de la racine sont étiquetés par MIN, les nœuds internes à distance impaire de la racine sont étiquetés par MAX.



La valeur d'un nœud MAX est le max des valeurs de ses fils. La valeur d'un nœud MIN est le min des valeurs de ses fils. On veut évaluer la racine. Le coût de l'algorithme est le nombre de feuilles consultées.

Hypothèses simplificatrices

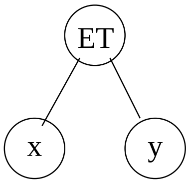
On suppose que l'arbre est un $T_{d,k}$. Ce qui signifie que tous les nœuds internes sont de degré d et que l'arbre est complet de profondeur $2k$ (k nœuds MIN et k nœuds MAX entre la racine et chaque feuille). On prendra $d = 2$. Les nœuds internes seront étiquetés par 0 ou 1 (ainsi MAX se traduit par OU et MIN par ET).

8.2 Théorèmes

Pour tout algorithme déterministe, il existe un $T_{2,k}$ tel que l'algorithme consulte toutes les feuilles (il y en a $2^{2k} = 4^k$).

Il existe un algorithme probabiliste tel que pour tout $T_{2,k}$, l'espérance du nombre de feuilles consultées est inférieur ou égal à 3^k.

8.3 Cas d'un nœud ET

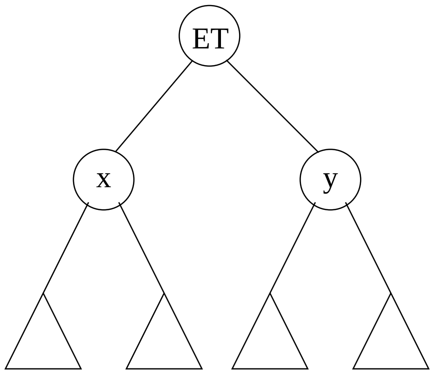


- Si $xy = 11$, tout algorithme (déterministe ou non) doit consulter x et y.
- Si $xy \in \{01, 10\}$, un algorithme déterministe doit consulter x et y.
  si l'algorithme lit en premier x, sur l'entrée $xy = 10$, il devra également lire y.
  si l'algorithme lit en premier y, sur l'entrée $xy = 01$, il devra également lire x.
  algorithme probabiliste : tirer au sort $u \in \{x, y\}$
  si $u = 0$ retourner 0
  sinon retourner l'autre variable
  Espérance du nombre de variables lues pour $xy = 10$: $\frac{1}{2} * 1 + \frac{1}{2} * 2 = \frac{3}{2}$
- Si $xy = 00$, on retourne 1 en ne consultant qu'une variable.

8.4 Algorithme

```
eval(N)
choisir au hasard un fils u de N
Soit v l'autre fils
Si N est un nœud ET
            si eval(u) = 0 retourner 0
            sinon retourner eval(v)
sinon
            si eval(u) = 1 retourner 1
            sinon retourner eval(v)
```

L'espérance du nombre de $T_{2,k-1}$ évalués est ≤ 3. Donc l'espérance du coût pour $T_{2,k}$ est $\leq 3 * 3^{k-1} \leq 3^k$.



Si $x = 0$ et $y = 0$, on évalue au plus un des deux nœuds OU, donc au plus 2 des $T_{2,k-1}$.

Si $x = 1$ et $y = 1$, on évalue les deux nœuds OU. Pour chaque nœud OU, l'espérance du nombre de sous arbres consultés est $\frac{3}{2}$. Donc au total $\frac{3}{2} + \frac{3}{2} = 3$.

Si $x = 0$ et $y = 1$, l'espérance est bornée par $\frac{1}{2} * 2 + \frac{1}{2}\left(2 + \frac{3}{2}\right)$. C'est à dire par $2 + \frac{3}{4} = \frac{11}{4} < 3$.

Chapitre 9

Variance

Souvent femme varie. Et bien fol qui s'y fie.
François Ier

9.1 Inégalités

1. Si $0 \leq X \leq Y$, alors $0 \leq E(X) \leq E(Y)$
   En effet $E(X) = \sum_n a_n P(X = a_n)$ avec $X : \Omega \to A = \{a_1, a_2, ..., a_n\}$
   Et donc $E(X) \geq 0$
   $E(X) = \sum_n a_n \sum_m P(X = a_n \text{ et } Y = a_m) = \sum_n \sum_m a_n P(X = a_n \text{ et } Y = a_m)$
   Or si $a_n > a_m$, $P(X = a_n \text{ et } Y = a_m) = 0$
   $$\begin{aligned} \text{Donc } E(X) &\leq \sum_n \sum_m a_m P(X = a_n \text{ et } Y = a_m) \\ &\leq \sum_m a_m \sum_n P(X = a_n \text{ et } Y = a_m) \\ &\leq \sum_m a_m P(Y = a_m) \\ E(X) &\leq E(Y) \end{aligned}$$
2. Si $E(X)$ et $E(Y)$ existent, alors $X \leq Y \Rightarrow E(X) \leq E(Y)$
3. Si $X \geq 0, P(X \geq \alpha) \leq \frac{1}{\alpha} E(X)$ pour tout $\alpha > 0$.
   En effet, soit Y la variable aléatoire définie par $Y(\omega) = \begin{cases} \alpha \text{ si } X(\omega) \geq \alpha \\ 0 \text{ sinon} \end{cases}$
   $X \geq Y$ donc $E(X) \geq E(Y) = \alpha P(X \geq \alpha)$.

9.2 Définition

On suppose $E(X)$ bien définie.

$$Var(X) = E\left((X - m)^2\right) \text{ avec } m = E(X)$$

9.2.1 Notation

$Var(X) = \sigma^2$. (σ s'appelle l'écart-type)

9.2.2 Remarques

$Var(X) = E(X^2 + m^2 - 2mX) = E(X^2) + m^2 - 2mE(X) = E(X^2) - E(X)^2$
$E(|X|^k)$ s'appelle le moment d'ordre k.
$|X| \leq 1 + X^2$ donc $E(|X|) \leq 1 + E(X^2)$
donc $E(X^2) < +\infty \Rightarrow E(X) < +\infty$

9.3 Autres inégalités

Inégalité de Markov[1] :

$$P(|X| \geq \alpha) \leq \frac{1}{\alpha^k} E(|X|^k)$$

En effet $P(|X| \geq \alpha) = P(|X|^k \geq \alpha^k)$

Inégalité de Tchebychev[2] : obtenue en prenant $k = 2$ et en appliquant l'inégalité de Markov à $X - m$.

$$P(|X - m| \geq \alpha) \leq \frac{\sigma^2}{\alpha^2}$$

9.4 Propriétés de la variance

- Pour tout $\lambda \in \mathbb{R}$, $Var(\lambda X) = \lambda^2 Var(X)$
- Si X et Y sont des variables indépendantes, $Var(X + Y) = Var(X) + Var(Y)$.
  En effet $$\begin{aligned} Var(X+Y) &= E\left((X+Y)^2\right) - E(X+Y)^2 \\ &= E(X^2 + Y^2 + 2XY) - \left(E(X)^2 + E(Y)^2 + 2E(X)E(Y)\right) \\ &= E\left(X^2\right) - E(X)^2 + E\left(Y^2\right) - E(Y)^2 \text{ car } X \text{ et } Y \text{ sont indépendantes} \\ &= Var(x) + Var(Y) \end{aligned}$$
- $Var(X_1 + X_2 + ... + X_n) = \sum_{i=1}^n Var(X_i)$ si les X_i sont deux à deux indépendantes.

Loi faible des grands nombres

Soient $X_1, X_2, ..., X_n$ des variables indépendantes deux à deux avec $E(X_i) = m < +\infty$ et $Var(X_i) = \sigma^2 < +\infty$

$$P\left(\left|\frac{X_1 + X_2 + ... + X_n}{n} - m\right| \geq \alpha\right) \leq \frac{\sigma^2}{m\alpha^2}$$

En effet $Var\left(\frac{X_1+X_2+...+X_n}{n}\right) = \frac{1}{n^2} Var(X_1 + X_2 + ... + X_n) = \frac{\sigma^2}{n}$, Puis on applique l'inégalité de Tchebychev.

Inégalité de Cauchy[3]-Bunyakovskii[4]-Schwarz[5]

$$E(XY)^2 \leq E\left(X^2\right) E\left(Y^2\right)$$

[1] Markov (André Andreïévitch), mathématicien russe (Ryazan 1856 - Petrograd 1922).
[2] Tchebychev (Pafnoutiy Lvovitch), mathématicien russe (Okatovo 1821 - Saint Pétersbourg 1894).
[3] Cauchy (baron Augustin Louis), mathématicien français (Paris 1789 - Sceaux 1857).
[4] Bunyakovskii (Viktor Yakovlevich), mathématicien russe (Podolskaya gubernia 1804 - Saint Pétersbourg 1889).
[5] Schwarz (Karl Hermann Amandus), mathématicien allemand (Hermsdorf-untern-Kynast 1843 - Berlin 1921).

Preuve

Considérons $f(t) = E\left((tX+Y)^2\right)$. $\forall t \in \mathbb{R}, f(t) \geq 0$.

$$\begin{aligned} \text{Or } f(t) &= E\left(t^2X^2 + 2tXY + Y^2\right) \\ &= t^2E\left(X^2\right) + 2tE(XY) + E\left(Y^2\right) \end{aligned}$$

$\Delta \leq 0$ (Le polynôme a au plus une racine réelle)
Or $\Delta = 4E(XY)^2 - 4E\left(X^2\right)E\left(Y^2\right)$
Donc $4E(XY)^2 - 4E\left(X^2\right)E\left(Y^2\right) \leq 0$
Et $E(XY)^2 \leq E\left(X^2\right)E\left(Y^2\right)$

Chapitre 10

Problème des sommes distinctes

10.1 Définitions

10.1.1 Définition

$S = \{x_1, x_2, ..., x_k\}$ ensemble d'entiers ≥ 1 a ses sommes directes distinctes si toutes les sommes de la forme $\sum_{i \in I} x_i$ avec $I \subseteq \mathbb{N}_k$ sont distinctes.

10.1.2 Exemple

$\{1, 2, 3\}$ n'a pas cette propriété.
$\{2^i; 0 \leq i \leq k-1\}$ a ses sommes distinctes.

10.1.3 Définition

$f(n)$ est le plus grand cardinal d'un ensemble $S \subseteq \{1, ..., n\}$ qui a la propriété des sommes distinctes. ($f(n) \geq \lfloor \log(n) \rfloor + 1$).

10.2 Bornes de f

10.2.1 Propriété

$f(n) \leq \log(n) + \log(\log(n)) + O(1)$

Problème d'Erdös[1]

Montrer que $f(n) \leq \log(n) + O(1)$

10.2.2 Théorème

$f(n) \leq \log(n) + \frac{1}{2}\log(\log(n)) + O(1)$

[1] Erdös (Paul), mathématicien hongrois (Budapest 1913 - Varsovie 1996).

Preuve de la propriété

Supposons $S = \{x_1, x_2, ..., x_k\} \subseteq \{1, ..., n\}$ a ses sommes distinctes. $2^k \leq kn$ car $\sum_{i \in I} x_i \leq kn$ pour tout $I \subseteq \{1, ..., k\}$.
D'où $k \leq \lg(k) + \lg(n)$.
Pour k assez grand, $\lg(k) \leq \frac{k}{2}$
$k \leq \frac{k}{2} + \lg(n)$
$k \leq 2\lg(n)$
$k \leq \lg(2\lg(n)) + \lg(n)$
$k \leq \lg(n) + \lg(\lg(n)) + 1$

Démonstration du théorème

Soient $\varepsilon_1, \varepsilon_2, ..., \varepsilon_k$ n variables aléatoires indépendantes avec $P(\varepsilon_i = 0) = P(\varepsilon_i = 1) = \frac{1}{2}$.
$X = \sum_{i=1}^{k} \varepsilon_i x_i$
$Var(X) = \sum_{i=1}^{k} Var(\varepsilon_i x_i) = \sum_{i=1}^{k} x_i^2 Var(\varepsilon_i)$
$Var(\varepsilon_i) = E(\varepsilon_i^2) - E(\varepsilon_i)^2 = \frac{1}{4}$
$Var(X) = \sum_{i=1}^{k} \frac{x_i^2}{4} \leq \frac{kn^2}{4}$
D'après Tchebychev $P(|X - m| \geq \alpha) \leq \frac{Var(X)}{\alpha^2} \leq \frac{kn^2}{4\alpha^2}$
Pour tout a, $P(X = a) = 0$ ou $\frac{1}{2^k}$ Donc $P(|X - m| < \alpha) \leq \frac{2\alpha}{2^k}$.
De ces deux inégalités, on déduit que $\forall \alpha$, $\frac{kn^2}{4\alpha^2} + \frac{\alpha}{2^{k-1}} \geq 1$
Donc $n^2 \geq \frac{4\alpha^2}{k}\left(1 - \frac{\alpha}{2^{k-1}}\right)$
On choisit $\alpha = \frac{2^k}{4}$
$n^2 \geq \left(\frac{2*2^k}{4\sqrt{k}}\sqrt{1 - \frac{1}{2}}\right)^2$ ou encore $n \geq \frac{2^k}{2\sqrt{2k}}$.
D'où $2^k \leq 2n\sqrt{(2k)}$ puis $k \leq 1 + \lg(n) + \frac{1}{2}\lg(2k)$
Or pour k assez grand, $1 + \frac{1}{2}\lg(2k) \leq \frac{k}{2}$ d'où $k \leq \lg(n) + \frac{k}{2}$ et $k \leq 2\lg(n)$
$k \leq 1 + \lg(n) + \frac{1}{2}\lg(4\lg(n))$
$k \leq 2 + \lg(n) + \frac{1}{2}\lg(\lg(n))$

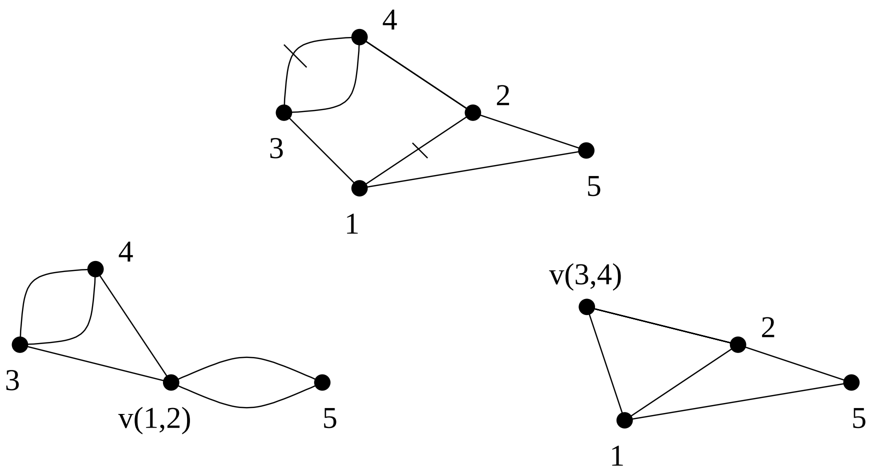
Chapitre 11

Séparateur minimum (min cut)

Que l'homme ne sépare pas ce que Dieu a uni
Evangile selon Saint Matthieu

Soit G est un multigraphe : si $x \neq y$ sont deux sommets, il peut y avoir plusieurs arêtes entre x et y, mais pas de boucle. On suppose G connexe. Une coupe de $G = (V, E)$ (V : ensemble des sommets, E : ensemble des arêtes) est un sous-ensemble $c \subseteq E$ tel que $(V, E - c)$ est non-connexe. On cherche une coupe de cardinal minimal.

Contraction d'arête : soit $e \in E$ une arête reliant x à y. On remplace G par G' tel que l'ensemble des sommets de G' soit $V' = V - \{x, y\} \cup \{v(x, y)\}$. Soit $w(u, v)$ le nombre d'arêtes reliant u à v dans G et $w'(u, v)$ le nombre d'arêtes reliant u à v dans G'. Pour $u \neq v(x, y)$, $w'(u, v(x, y)) = w(u, x) + w(u, y)$. Si $v(x, y) \notin \{t, u\}$, $w'(t, u) = w(t, u)$.



Observation $\forall v \in G, degre(v) \geq$ cardinal de coupe minimale.

11.1 Algorithme

tant que $|V| > 2$
    choisir une arête et la contracter
fin tant que
retourner le nombre d'arêtes restantes

11.1.1 Propriété

L'algorithme retourne toujours une surestimation du cardinal de la coupe minimale.

11.1.2 Lemme

Toute coupe de G est une coupe de G'.

11.1.3 Théorème

Soit C une coupe minimale de G. La probabilité d'obtenir C par l'algorithme est supérieure ou égale à $\frac{2}{n^2}$ avec $n = |V|$.

Preuve

Soit E_i : "à la contraction i, l'algorithme ne contracte pas une arête de C". On cherche $P(E_1 \cap E_2 \cap ... \cap E_{n-2})$. Soit $k = |C|$. Tous les sommets de G sont de degré supérieur ou égal à k donc $|E| \geq \frac{kn}{2}$.

$P(E_1) \geq \frac{k\frac{n}{2} - k}{k\frac{n}{2}} = \frac{n-2}{n} = 1 - \frac{2}{n}$

$P(E_1 \cap E_2 \cap ... \cap E_i) = P(E_1 \cap E_2 \cap ... \cap E_{i-1})P(E_i|E_1 \cap E_2 \cap ... \cap E_{i-1})$

On suppose que pendant les $i - 1$ premières étapes, on n'a contracté aucune arête de C. Dans le graphe couvrant, les sommets sont de degré supérieur ou égal à k et on a $n - (i - 1)$ sommets.

Donc $P(E_i|E_1 \cap E_2 \cap ... \cap E_{i-1}) \geq \frac{k\frac{n-i+1}{2} - k}{k\frac{n-i+1}{2}} = 1 - \frac{2}{n-i+1}$

Et finalement

$$\begin{array}{rcl} P(E_1 \cap E_2 \cap ... \cap E_{n-2}) & \geq & \prod_{i=1}^{n-2} 1 - \frac{2}{n-i+1} \\ & \geq & \frac{2}{n(n-1)} \\ & \geq & \frac{2}{n^2} \end{array}$$

11.1.4 Algorithme amélioré

faire tourner l'algorithme de contraction r fois
et retourner le minimum des résultats obtenus

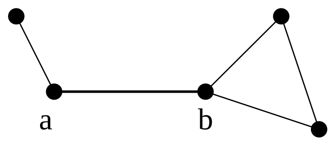
La probabilité d'erreur est majorée par $\left(1 - \frac{2}{n^2}\right)^r$. Donc pour obtenir une probabilité d'erreur inférieure à un ε donné, il suffit d'avoir $r \ln\left(1 - \frac{2}{n^2}\right) \leq \ln(\varepsilon)$. Or $\ln(1+x) \leq x$ Donc il suffit d'avoir

$$r \geq \frac{n^2}{2} \ln\left(\frac{1}{\varepsilon}\right)$$

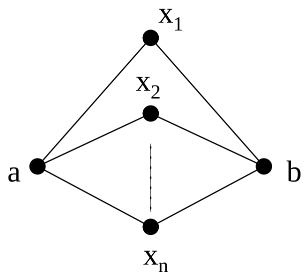
C'est un algorithme de "Monte-Carlo" (la probabilité d'erreur est > 0). Pour un algorithme de "Las Vegas", on obtient toujours le bon résultat (mais le temps de calcul est une variable aléatoire).

remarque Il y a au plus $\frac{n^2}{2}$ séparateurs minimaux dans un graphe à n sommets.

11.2 Définition



Un ab-séparaeur est un ensemble d'arêtes tel que a et b sont déconnectés si on les enlève. On définit également la notion d'ab-séparateur minimum.



Sur cet exemple, la taille d'un séparateur minimum est 2.
Celle d'un ab-séparateur minimum est n.
Et le nombre d'ab-séparateurs minimaux est 2^n (il est exponentiel !).

Chapitre 12

Bornes de Chernoff

Une fois qu'on a dépassé les bornes, il n'y a plus de limites.
Alphonse Allais

12.1 Bornes de Chernoff

Soient $X_1, X_2, ..., X_n$ des variables aléatoires indépendantes telles que $P(X_i = 1) = P(X_i = 0) = \frac{1}{2}$. On pose $S_n = \sum_{i=1}^{n} X_i$ et on cherche $P(S_n > a)$.

12.1.1 Théorème

$$\boxed{\forall a > 0, P(S_n > a) < e^{-\frac{a^2}{2n}}}$$

Preuve

Posons $Y_i = e^{\lambda X_i}$. On a alors $E(Y_i) = \frac{e^{\lambda}+e^{-\lambda}}{2} = \cosh(\lambda) \leq e^{\frac{\lambda^2}{2}}$
$P(S_n > a) = P(e^{\lambda S_n} > e^{\lambda a}) < \frac{E(e^{\lambda S_n})}{e^{\lambda a}}$ (d'après Markov)

$$\begin{aligned}
E(e^{\lambda S_n}) &= E(e^{\lambda(X_1+X_2+...+X_n)}) \\
&= E(e^{\lambda X_1} e^{\lambda X_2} ... e^{\lambda X_n}) \\
&= \prod_{i=1}^{n} E(e^{\lambda X_i}) \\
&= \cosh(\lambda)^n \\
&\leq e^{\frac{\lambda^2 n}{2}}
\end{aligned}$$

D'où $P(S_n > a) < e^{\frac{\lambda^2 n}{2} - \lambda a}$. Et en posant $\lambda = \frac{a}{n}$, on obtient $P(S_n > a) < e^{-\frac{a^2}{2n}}$

Conséquence

$P(S_n > a) \ll 1$ dès que $a \gg \sqrt{n}$

Remarques

– D'après le théorème central limite, $\lim_{n \to +\infty} P(S_n > u\sqrt{n}) \;=\; \int_u^{+\infty} e^{-\frac{t^2}{2}} \mathrm{d}t \;\leq\; e^{-\frac{u^2}{2}}$

- $S_n = 2k - n \Longleftrightarrow \begin{cases} k \text{ des } X_i \text{ valent } 1 \\ n-k \text{ des } X_i \text{ valent } -1 \end{cases}$

  Et donc $P(S_n = 2k-n) = \frac{C_n^k}{2^n}$
- $P(S_n < -a) = P(S_n > a) < e^{-\frac{a^2}{2n}}$

12.2 Discrépance

Soit Ω un ensemble fini et $\mathcal{A}$ une famille de parties de Ω. On veut colorier les points de Ω en B ou R de façon que $\forall A \in \mathcal{A}$, le nombre de points R de A soit égal au nombre de points B. Un coloriage est une fonction $X : \Omega \longrightarrow \{-1, 1\}$. On introduit la discrépance : $\mathrm{disc}(X) = \max_{A\in\mathcal{A}} |\chi(A)|$ avec $\chi(A) = \sum_{a\in A} X(a)$.

Exemple 10 *supposons les éléments de A disjoints. On peut trouver un coloriage X tel que* $\mathrm{disc}(X) \leq 1$

Exemple 11 *Si $\mathcal{A}$ est l'ensemble des parties de Ω, alors* $\forall X, \mathrm{disc}(X) \geq \frac{|\Omega|}{2}$

12.2.1 Théorème

Supposons $|\Omega| = m$, $|\mathcal{A}| = n$. Il existe un coloriage X tel que $\mathrm{disc}(X) \leq \sqrt{2m\ln(2n)}$

12.2.2 Remarque

Pour $m = n$, on peut obtenir $\mathrm{disc}(X) < 6\sqrt{n}$

12.2.3 Preuve du théorème

Soit $\alpha = \sqrt{2m\ln(2n)}$. On considère un coloriage aléatoire C. $\chi(A) = \sum_{a\in A} C(a)$.

D'après Chernoff, $$\begin{array}{c} P(\chi(A) > \alpha) < e^{-\frac{\alpha^2}{2|A|}} \\ P(\chi(A) < -\alpha) < e^{-\frac{\alpha^2}{2|A|}} \end{array}$$

Donc $P(|\chi(A)| > \alpha) < 2e^{-\frac{\alpha^2}{2|A|}}$

Posons $X_A = \begin{cases} 1 & \text{si } |\chi(A)| > \alpha \\ 0 & \text{sinon} \end{cases}$

$E(X_A) < 2e^{-\frac{\alpha^2}{2m}}$

Posons maintenant $X = \sum_{A\in\mathcal{A}} X_A$.

$$\begin{array}{rcl} E(X) & = & \sum_{A\in\mathcal{A}} E(X_A) \\ & < & 2ne^{-\frac{\alpha^2}{2m}} \\ & < & 2ne^{-\ln(2n)} \end{array}$$

On remarque que $E(X) < 1$. Donc il existe un coloriage C avec $X(C) < 1$ c'est à dire $X(C) = 0$. Et $\forall A \in \mathcal{A}, |\chi(A)| \leq \alpha$.

12.3 Variantes de Chernoff

Soit Y une somme de variables aléatoires indicatrices (à valeurs dans $\{0, 1\}$) indépendantes. Soit $\mu = E(Y)$

12.3.1 Lemme 1

$P(Y < (1-\varepsilon)\mu) < e^{-\frac{\mu\varepsilon^2}{2}}$

12.3.2 Lemme 2

$P(Y > (1+\varepsilon)\mu) < \left(\frac{e^\varepsilon}{(1+\varepsilon)(1+\varepsilon)}\right)^\mu$

12.3.3 Remarque

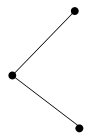
$\forall \hat{\mu} \geq \mu, P(Y > (1+\varepsilon)\hat{\mu}) < \left(\frac{e^\varepsilon}{(1+\varepsilon)(1+\varepsilon)}\right)^{\hat{\mu}}$

12.3.4 Synthèse des lemmes

$P(|Y-\mu| > \varepsilon\mu) < 2e^{-c_\varepsilon \mu}$
avec c_ε une constante qui ne dépend que de ε.

12.4 Graphes aléatoires

Soit $G(n,p)$ un espace de probabilité modélisant les graphes à n sommets et présentant une probabilité p de voir une paire quelconque d'arêtes reliées et tel que les $\frac{n(n-1)}{2}$ événements "ij est une arête" pour $i \neq j$ soient indépendants. $E(\deg(x)) = (n-1)p$

Exemple 12 $G(3,p)$  $P(G) = p^2(1-p)$

12.4.1 Théorème

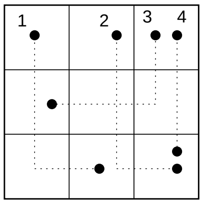
> On pose $p = f(n) = \frac{\ln(n)}{n}\omega(n)$ avec $lim_{n\to+\infty}\omega(n) = +\infty$.
>
> $$\forall \varepsilon > 0, \lim_{n\to+\infty} P(\exists x \in \{1,...,n\}; |\deg(x) - p(n-1)| > \varepsilon(n-1)p) = 0$$

12.4.2 Preuve

Soit $X_{i,j} = \begin{cases} 1 & \text{si } ij \text{ est une arête} \\ 0 & \text{sinon} \end{cases}$
$\deg(i) = \sum_{i\neq j} X_{i,j}$
$P(|\deg(i) - p(n-1)| > \varepsilon(n-1)p) < 2e^{c_\varepsilon p(n-1)}$
$P(\exists x \in \{1,...,n\}; |\deg(x) - p(n-1)| > \varepsilon(n-1)p) \leq 2ne^{-c_\varepsilon p(n-1)}$ (probabilité d'une union)
$P(\exists x \in \{1,...,n\}; |\deg(x) - p(n-1)| > \varepsilon(n-1)p) \leq 2ne^{-c_\varepsilon \frac{\ln(n)}{n}\omega(n)(n-1)}$
Or $2ne^{-c_\varepsilon \frac{\ln(n)}{n}\omega(n)(n-1)} \longrightarrow_{n\to+\infty} 0$

12.5 Conception de circuits

On considère un tableau de cellules logiques. On veut relier certaines paires de cellules par un fil. Un fil relie exactement 2 cellules et ne peut prendre qu'un virage. On a donc au plus 2 chemins possibles pour chaque fil (2 pour les liaisons comportant 1 virage, 1 pour les autres)



12.5.1 Contrainte

La frontière entre 2 cellules ne pourra pas être traversée par plus de w fils.

12.5.2 Exemple

Dans notre exemple, si $w = 2$, la contrainte est respectée. Si $w = 1$, elle ne peut pas l'être, en effet, 4 fils doivent descendre des 3 cellules de la $1^{\text{ère}}$ ligne.

12.5.3 Optimisation

On cherchera à partir de maintenant un tracé S qui minimise le nombre maximal de fils traversant une frontière. Si on pose $w_S(b)$ le nombre de fils traversant la frontière b. On veut donc minimiser $w_S = \max_b w_S(b)$.

12.5.4 Modélisation sous forme de programme linéaire

Définissons les variables $x_{i,0}$ et $x_{i,1}$ de la façon suivante : $x_{i,0} = 1$ si le fil i est horizontal au départ de son extrémité gauche, $x_{i,0} = 0$ sinon, et $x_{i,1} = 1 - x_{i,0}$.

Exemple

$x_{1,0} = 0 \quad x_{1,1} = 1$
$x_{2,0} = 0 \quad x_{2,1} = 1$
$x_{3,0} = 1 \quad x_{3,1} = 0$
$x_{4,0} = 1 \quad x_{4,1} = 0$ ou $x_{4,0} = 0$ $x_{4,1} = 1$

Coût d'une solution

Soit b une frontière. $T_{b,0} = \{i;\ \text{le fil } i \text{ traverse } b \text{ si } x_{i,0} = 1\}$ et $T_{b,1} = \{i;\ \text{le fil } i \text{ traverse } b \text{ si } x_{i,1} = 1\}$.
$w_S(b) = \sum_{i \in T_{b,0}} x_{i,0} + \sum_{i \in T_{b,1}} x_{i,1}$

Enoncé de notre programme linéaire

On veut minimiser w tel que
- $\forall i, x_{i,0} + x_{i,1} = 1$
- $\sum_{i \in T_{b,0}} x_{i,0} + \sum_{i \in T_{b,1}} x_{i,1} \leq w$
- $x_{i,0}, x_{i,1} \in \{0, 1\}$

Malheureusement, un tel problème est NP-complet.

Relaxation

Pour résoudre ce problème, on remplace la contrainte $x_{i,0}, x_{i,1} \in \{0,1\}$ par $x_{i,0}, x_{i,1} \in [0,1]$. L'optimisation de ce programme linéaire se fait en temps polynomial par méthode de l'ellipsoïde ou bien par méthode des points intérieurs. On appelle $\hat{x}_{i,0}$ et $\hat{x}_{i,1}$ la solution du problème relaxé et $\hat{w}$ le minimum obtenu. Puis on remplace $\hat{x}_{i,0}$ et $\hat{x}_{i,1}$ par $\check{x}_{i,0}$ et $\check{x}_{i,1} \in \{0,1\}$.

- On peut arrondir au plus près
  Dans ce cas, $w_S \leq 2w_0$ avec w_0 coût de la solution optimale. En effet, $\sum_{i\in T_{b,0}} \check{x}_{i,0} + \sum_{i\in T_{b,1}} \check{x}_{i,1} \leq \sum_{i\in T_{b,0}} 2\hat{x}_{i,0} + \sum_{i\in T_{b,1}} 2\hat{x}_{i,1} \leq 2\hat{w} \leq 2w_0$
- On peut arrondir aléatoirement : $\check{x}_{i,0} = 1$ avec la proabilité $\hat{x}_{i,0}$
  On a alors $E(\check{x}_{i,0}) = \hat{x}_{i,0}$

$$\begin{array}{rcl} E(w_S(b)) & = & \sum_{i\in T_{b,0}} E(\check{x}_{i,0}) + \sum_{i\in T_{b,1}} E(\check{x}_{i,1}) \\ & = & \sum_{i\in T_{b,0}} \hat{x}_{i,0} + \sum_{i\in T_{b,1}} \hat{x}_{i,1} \\ & \leq & \hat{w} \\ & \leq & w_0 \end{array}$$

Posons $\check{w} = max_b w_S(b)$ et B le nombre de fils.

$$\begin{array}{rcl} P(w_S(b) > (1+\varepsilon)w_0) & < & \left(\frac{e^\varepsilon}{(1+\varepsilon)^{1+\varepsilon}}\right)^{w_0} \\ P(\check{w} > (1+\varepsilon)w_0) & < & B\left(\frac{e^\varepsilon}{(1+\varepsilon)^{1+\varepsilon}}\right)^{w_0} \end{array}$$

Chapitre 13

Méthode des probabilités conditionnelles

13.1 Théorème

Soit $\mathcal{A} = (A_1, A_2, ..., A_n)$ une famille de n parties de Ω ($|\Omega| = m$). Il existe un coloriage $X : \Omega \to \{-1, 1\}$ tel que $|\mathrm{disc}(X)| \leq \sqrt{2m\ln(2n)}$ et ce coloriage peut être obtenu en temps polynomial par un algorithme déterministe.

13.1.1 Rappels

$X(A) = \sum_{a\in A} X(a)$ et $\mathrm{disc}(X) = \max_{A\in\mathcal{A}} |X(A)|$

13.1.2 Construction

On considère un coloriage X aléatoire et on pose $\alpha = \sqrt{2m\ln(2n)}$. On pose $\mathcal{E}_j$ l'événement $|X(A_j)| > \alpha$.

$P\left(\bigcup_{i=1}^n \mathcal{E}_i\right) \leq \sum_{i=1}^n P(\mathcal{E}_i) \leq 2ne^{-\frac{\alpha^2}{2n}} < 1$.

Supposons choisis $X(a_1) = \varepsilon_1, ..., X(a_i) = \varepsilon_i$, comment allons-nous choisir $X(a_{i+1})$?

Posons $P_i = P(\bigcup_{j=1}^n \mathcal{E}_j | X(a_1) = \varepsilon_1, ..., X(a_i) = \varepsilon_i) = P(\mathcal{E}|X_i)$. On va choisir $X(a_{i+1})$ de sorte que $P_{i+1} \leq P_i$. Ceci est possible car $P_i = \frac{1}{2}(A + R)$ où $A = P(\mathcal{E}|X_i \text{ et } X(a_{i+1}) = -1)$ et $R = P(\mathcal{E}|X_i \text{ et } X(a_{i+1}) = 1)$. Et donc $A \leq P_i$ ou $R \leq P_i$ et on peut choisir la valeur correspondante pour $X(a_{i+1})$.

On a alors $p_m \leq \cdots \leq P_{i+1} \leq P_i \leq \cdots \leq P_0 < 1$ donc $P_m < 1$. Or $P_m \in \{0, 1\}$ donc $P_m = 0$

13.1.3 Estimation pessimiste

$\widehat{P_i} = \sum_{j=1}^n P(\mathcal{E}_j | X(a_1) = \varepsilon_1, ..., X(a_i) = \varepsilon_i)$

$P_i \leq \widehat{P_i}$

Propriétés de $\widehat{P_i}$:

- $\widehat{P_0} < 1$
- On peut choisir $X(a_{i+1})$ de telle façon que $\widehat{P_{i+1}} < \widehat{P_i}$
- $\widehat{P_i}$ se calcule en temps polynomial.

A la fin, $\widehat{P_m} < 1$ donc aucun des $\mathcal{E}_j$ n'est réalisé à la fin de l'algorithme.

13.2 Propriété

$$P(\mathcal{E}_j|X_i) = \frac{1}{2}\left(P(\mathcal{E}_j|X_i \text{ et } X(a_{i+1}) = -1) + P(\mathcal{E}_j|X_i \text{ et } X(a_{i+1}) = 1)\right)$$

13.3 Propriété

Il suffit de calculer $P(\mathcal{E}_j|X_i)$

$$X(A_j) = \underbrace{\sum_{k\in\{1,...,i\},a_k\in A_j} X(a_k)}_{\text{valeur connue}} + \underbrace{\sum_{k>i,a_k\in A_j} X(a_k)}_{\text{somme de v. a. indépendantes}}$$

Soit $N = |A_j - \{a_1, ..., a_j\}|$

$$P\left(\sum_{k>i,a_k\in A_i} X(a_k) = 2k - N\right) = \frac{C_N^k}{2^N}$$

Chapitre 14

Chaînes de Markov

Les hommes passent la moitié de leur temps à se forger des chaînes et l'autre à les porter.
Octave Mirbeau

14.1 Présentation

On se donne un ensemble d'éléments S (fini ou dénombrable) et des probabilités de transition $p_{i,j}$ ($i \in S, j \in S$). On nomme X_n l'état au temps n. Le tout vérifiant $P(X_n = i_n | X_0 = i_0, X_1 = i_1, ..., X_{n-1} = i_{n-1}) = P(X_n = i_n | X_{n-1} = i_{n-1}) = p_{n-1,n}$.

Exemple 13 *Modèle de diffusion : on a 2 boîtes contenant chacune r boules, et il y a au total $2r$ boules : r noires et r blanches. L'état du système est déterminé par le nombre de boules blanches dans la première boîte ($S = \{0, 1, ..., r\}$). Une transition consiste à tirer une boule de la première boîte et une de la deuxième et à les échanger.*

- $p_{i,j} = 0$ *si* $|j - i| \geq 2$
- $p_{i,i-1} = \frac{i^2}{r^2}$
- $p_{i,i+1} = \frac{(r-i)^2}{r^2}$
- $p_{i,i} = \frac{2i(r-i)}{r^2}$

Exemple 14 *Marche au hasard dans $\mathbb{Z}^k$ ($S = \mathbb{Z}^k$)*

$(x_1, x_2, ..., x_k)$ a 2^k voisins de la forme $(x_1, ..., x_i \pm 1, ..., x_k)$. Chaque probabilité de transition vaut $\frac{1}{2^k}$

Si $k \leq 2$, on revient toujours au point de départ. Si $k \geq 3$, on revient au point de départ avec une probabilité strictement inférieure à 1.

L'évolution d'une chaîne de Markov est complètement déterminée par les $p_{i,j}$ et la distribution initiale Π^0.

14.1.1 Définition

Un état i est persistant si partant de l'état i, on retourne en i avec une probabilité 1. Sinon, i est transitoire.

Posons $p_i(A) = P(A|X_0 = i)$. La probabilité de se trouver un jour dans l'état j sachant que $X_0 = i$ est $p_i\left(\bigcup_{n=1}^{+\infty} X_n = j\right) = f_{i,j}$. $f_{i,j} = \sum_{n=0}^{+\infty} f_{i,j}^{(n)}$ où $f_{i,j}^{(n)}$ est la probabilité de passer pour la première fois en j à l'instant n sachant que $X_0 = i$. $f_{i,j}^{(n)} = p_i(X_1 \neq j, X_2 \neq j, ..., X_{n-1} \neq j, X_n = j)$

$$i \text{ transitoire } \Leftrightarrow f_{i,i} < 1$$

$$i \text{ persistant } \Leftrightarrow f_{i,i} = 1$$

14.1.2 Propriété

$$p_i(X_n = i \text{ pour une infinité de valeurs de } n) = \begin{cases} 1 & \text{si } i \text{ est persistant} \\ 0 & \text{sinon} \end{cases}$$

$$p_i(|\{n \geq 1; X_n = i\}| \geq k) = f_{i,i}^{(k)} = \begin{cases} 1 & \text{si } i \text{ est persistant} \\ \longrightarrow_{k \to +\infty} 0 & \text{si } i \text{ est transitoire} \end{cases}$$

Remarque

$$p_i(X_n = j \text{ infiniment souvent}) = \begin{cases} 0 & \text{si } j \text{ est transitoire} \\ f_{i,j} & \text{si } j \text{ est persistant} \end{cases}$$

14.1.3 Théorème

$$i \text{ est transitoire} \Leftrightarrow \sum_{n=1}^{+\infty} p_{i,i}^{(n)} < +\infty$$

Preuve

Supposons que $\sum_{n=1}^{+\infty} p_{i,i}^{(n)} < +\infty$

$$p_i(X_n = i \text{ pour une infinité de valeurs de } n) = 0 \quad \text{d'après Borel-Cantelli}$$

donc i est transitoire

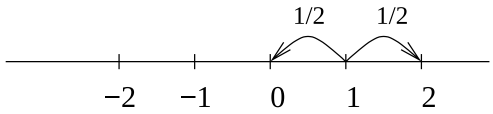
Réciproque : soit i transitoire. On va montrer que $\sum_{n=1}^{+\infty} p_{i,i}^{(n)} < \frac{f_{i,i}}{1-f_{i,i}}$

$$p_{i,i}^{(n)} = \sum_{s=1}^{n} f_{i,i}^s p_{i,i}^{n-s}$$

$$\begin{aligned}
\sum_{n=1}^{+\infty} p_{i,i}^{(n)} &= \sum_{n=1}^{+\infty} \sum_{s=1}^{n} f_{i,i}^s \, p_{i,i}^{n-s} \\
&= \sum_{s=1}^{+\infty} \sum_{n=s}^{+\infty} f_{i,i}^s \, p_{i,i}^{n-s} \\
&= \sum_{s=1}^{+\infty} f_{i,i}^s \sum_{n=0}^{+\infty} p_{i,i}^n \\
&\leq \left(1 + \sum_{n=1}^{+\infty} p_{i,i}^n\right) \underbrace{\left(\sum_{s=1}^{+\infty} f_{i,i}^s\right)}_{f_{i,i}}
\end{aligned}$$

14.1.4 Application aux marches au hasard

Exemple 15 *sur* $\mathbb{Z}$



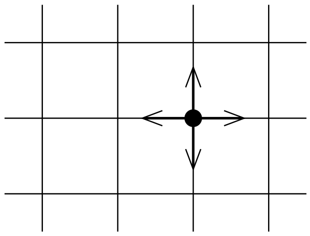
Est-ce que 0 *est transitoire ou non ?*

$$p_{00}^{(n)} = \begin{cases} 0 & \text{si } m \text{ est impair} \\ C_{2n}^n / 2^{2n} & \text{si } m = 2n \end{cases}$$

$$\frac{C_{2n}^n}{2^{2n}} \sim \frac{(2n)^{2n} e^{-2n} \sqrt{4n\pi}}{n^n e^{-n} \sqrt{2\pi n}} \sim \frac{1}{\sqrt{n\pi}}$$

La série $\sum p_{00}^{(n)}$ *diverge, donc l'état* 0 *(comme tout état* i*) est persistant.*

Exemple 16 *sur* $\mathbb{Z}^2$



$$p_{00}^{(n)} = \begin{cases} 0 & \text{si } m \text{ impair} \\ \frac{1}{4^{2n}(\sum_{k=0}^n C_{2k}^k C_{2(n-k)}^{n-k} C_{2n}^{2k})} & \text{si } m \text{ pair} \end{cases}$$

Si m *est pair, on peut également considérer que l'on choisit d'abord les places des mouvements faisant augmenter les coordonnées (il y en à n) puis parmi ceux-ci, ceux concernant l'abscisse, et parmi les* n *déplacements qui font décroître une coordonnée, ceux qui concernent l'abscisse. D'où : (pour* m *pair)*

$$\begin{aligned} p_{00}^{(n)} &= \frac{1}{4^{2n}} (\sum_{u=0}^{n} C_{2n}^n \, C_n^u \, C_n^{n-u}) \\ &= \frac{C_{2n}^n}{4^{2n}} (\sum_{u=0}^{n} C_n^u \, C_n^{n-u}) \\ &= \left(\frac{C_{2n}^n}{2^{2n}} \right)^2 \\ &\sim \frac{1}{n\pi} \end{aligned}$$

Ainsi tous les états sont persistants.

14.1.5 Chaînes irréductibles

Une chaîne est dite irréductible si $\forall (i,j) \; f_{i,j} > 0$. Autrement dit, $\forall (i,j) \, \exists n \geq 1 p_{i,j}^{(n)} > 0$.

14.1.6 Théorème

Pour une chaîne de Markov irréductible, l'une des deux propriétés suivantes est vérifiée :

(i) tous les états sont transitoires :

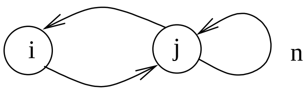
$$\forall (i,j) \quad p_i(X_n = j \text{ infiniment souvent}) = 0 \qquad \sum_{n\geq 1} p_{i,j}^{(n)} < +\infty$$

(ii) tous les états sont persistants :

$$\forall (i,j) \quad p_i(X_n = j \text{ infiniment souvent}) = 1$$

Preuve

Cas (i) Supposons qu'il existe i transitoire. Soit j un autre état. Montrons que $\sum_{n\geq 1} p_{i,j}^{(n)} < +\infty$



$$p_{i,j}^{(r)} > 0 \quad \text{et} \quad p_{j,i}^{s} > 0$$

$$\text{d'où} \quad p_{i,i}^{n+r+s} \quad \geq \quad p_{i,j}^{(r)}\, p_{jj}^{n}\, p_{j,i}^{s}$$

donc $\sum_{n=1}^{+\infty} p_{jj}^{(n)} < +\infty$ puisque $\sum_{n=1}^{+\infty} p_{i,i}^{(n)} < +\infty$, donc j est transitoire.
$pi(X_n = j$ infiniment souvent$) = 0$ car j est transitoire

$$\begin{aligned}
\sum_{n\geq 1} p_{i,j}^{(n)} &= \sum_{n=1}^{+\infty}\sum_{s=1}^{n} f_{i,j}^{(s)}\, p_{j,j}^{(n-s)} \\
&= \sum_{s=1}^{+\infty}\sum_{n=s}^{+\infty} f_{i,j}^{(s)}\, p_{j,j}^{(n-s)} \\
&= \sum_{s=1}^{+\infty}\left(f_{i,j}^{(s)} \sum_{n=0}^{+\infty} p_{j,j}^{(n)}\right) \\
&\leq \underbrace{\left(\sum_{s=1}^{+\infty} f_{i,j}^{(s)}\right)}_{f_{i,j}\leq 1} \underbrace{\left(\sum_{n=0}^{+\infty} p_{j,j}^{(n)}\right)}_{\leq +\infty \text{ car } j \text{ transitoire}} \qquad \leq +\infty
\end{aligned}$$

cas (ii) Soit i persistant. Tous les états sont persistants.

$$p_i\left(X_n = j \text{ infiniment souvent}\right) = \begin{cases} 0 & \text{si } j \text{ transitoire} \\ f_{i,j} & \text{si } j \text{ persistant} \end{cases}$$

On doit prouver que $f_{i,j} = 1$.

$p_{j,i}^{(m)}$ pour un certain $m \geq 1$

$$\begin{aligned} p_{j,i}^{(m)} &= p_j(X_m = i \text{ et } \underbrace{(Xn = j \text{ infiniment souvent})}_{\text{événement de probabilité 1}}) \text{ car j persistant} \\ p_{j,i}^{(m)} &= \sum_{n>m} p_j(X_m = i, X_{m+1} \neq j, X_{m+2} \neq j, \cdots, X_n \neq j) \\ &= \sum_{n>m} p_{j,i}^{(m)} f_{i,j}^{(n-m)} = p_{j,i}^{(m)} \underbrace{\sum_{n>m} f_{i,j}^{(n-m)}}_{\sum_{s\geq 1} f_{i,j}^{(s)}} = p_{j,i}^{(m)} f_{i,j} \end{aligned}$$

donc $f_{j,j} = 1$.

Montrons que $\sum_{n\geq 1} p_{i,j}^{(n)} = +\infty$. Si $\sum_{n\geq 1} p_{i,j}^{(n)} < +\infty$, alors $p_i(Xn = j \text{ infiniment souvent}) = 0$

Exemple 17 *Marche au hasard dans* $\mathbb{Z}^k$

$$\textit{cas (ii)} \Leftrightarrow k \leq 2$$

cas (i) chaîne transitoire

cas (ii) chaîne persistante

Exemple 18 *Chaînes finies : elles sont toutes persistantes*

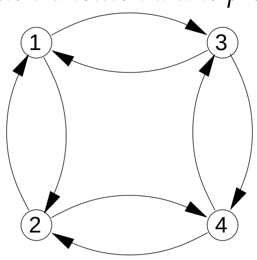
$$S = \{1, \ldots, n\} \qquad \forall t \geq 1 \sum_{j=1}^{n} p_{i,j}^{(t)} = 1$$

Supposons que $\sum_{k\geq 1} p_{i,j}^{(k)} < +\infty$.
Alors $\sum_{j=1}^{n} \sum_{t\geq 1} p_{i,j}^{(t)} < +\infty$.
Or $\sum_{j=1}^{n} \sum_{t\geq 1} p_{i,j}^{(t)} = \sum_{t\geq 1} \sum_{j=1}^{n} p_{i,j}^{(t)} = +\infty$ *ce qui est absurde.*

14.2 Distributions stationnaires

C'est une distribution de probabilités Π sur l'ensemble des états telle que $\mathbf{P}\Pi = \Pi$.

Exemple 19 *Chaque transition a une probabilité de* $1/2$.



$$\mathbf{P} = \begin{pmatrix} 0 & 1/2 & 0 & 1/2 \\ 1/2 & 0 & 1/2 & 0 \\ 0 & 1/2 & 0 & 1/2 \\ 1/2 & 0 & 1/2 & 0 \end{pmatrix}$$

$$\Pi_2 + \Pi_4 = 2\Pi_1 \Pi_1 + \Pi_3 = 2\Pi_2 \Pi_2 + \Pi_4 = 2\Pi_3 \Pi_1 + \Pi_3 = 2\Pi_4$$

D'où $\Pi_1 = \Pi_2 = \Pi_3 = \Pi_4 = 1/4$.

$$\Pi^0 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \quad \Pi^1 = \mathbf{P}\Pi^0 = \begin{pmatrix} 0 \\ 1/2 \\ 0 \\ 1/2 \end{pmatrix} \quad \Pi^2 = \begin{pmatrix} 1/2 \\ 0 \\ 1/2 \\ 0 \end{pmatrix} \quad \Pi^3 = \Pi^1$$

Exemple 20 *Marche aléatoire sur* $\mathbb{Z}$ *Une distribution stationnaire devrait vérifier* $p_i = \frac{1}{2}(p_{i-1} + p_{i+1})$. *Et donc* $p_i = (p_1 - p_0)i + p_0$. *Or* $\forall i, p_i \geq 0$ *d'où* $p_1 - p_0 = 0$ *et* $\sum p_i = 1$ *ce qui est impossible !*

14.2.1 Modélisation d'une file d'attente

Le temps d'attente de chaque client est 2 minutes. Pendant ce temps, il peut arriver 0, 1 ou 2 clients avec des probabilités respectives t_0, t_1, t_2 avec $t_0 + t_1 + t_2 = 1$.

Un état est le nombre de clients. L'espace des états S est alors $\mathbb{N}$. Pour $i \geq 1, p_{i,i-1} = t_0, p_{i,i} = t_1, p_{i,i+1} = t_2$ et pour $i = 0, p_{0,0} = t_0, p_{0,1} = t_1, p_{0,2} = t_2$. La matrice de transition est donc

$$\mathbf{P} = \begin{pmatrix} t_0 & t_0 & 0 & 0 & 0 & \cdots \\ t_1 & t_1 & t_0 & 0 & 0 & \cdots \\ t_2 & t_2 & t_1 & t_0 & 0 & \\ 0 & 0 & t_2 & t_1 & t_0 & \ddots \\ 0 & 0 & 0 & t_2 & t_1 & \ddots \\ \vdots & \vdots & & \ddots & \ddots & \ddots \end{pmatrix}$$

$$\mathbf{P}\Pi = \Pi \Rightarrow \begin{cases} \Pi_0 = t_O\Pi_0 + t_0\Pi_1 \\ \Pi_1 = t_1\Pi_0 + t_1\Pi_1 + t_0\Pi_2 \\ \Pi_2 = t_2\Pi_0 + t_2\Pi_1 + t_1\Pi_2 + t_0\Pi_3 \\ \Pi_k = t_2\Pi_{k-1} + t_1\Pi_k + t_0\Pi_{k+1} \text{ pour } k \geq 3 \end{cases}$$

$$t_0\Pi_{k+1} + (t_1 - 1)\Pi_k + t_2\Pi_{k-1}$$

$$t_0\Pi_{k+1} - (t_0 + t_2)\Pi_k + t_2\Pi_{k-1}$$

$$\begin{aligned} \text{Soit } Q(X) &= t_0X^2 - (t_0 + t_2)X + t_2 \\ &= (X - 1)(t_0X - t_2) \end{aligned}$$

Q a pour racines 1 et $\frac{t_2}{t_0}$. Donc la solution générale est de la forme $a + b\left(\frac{t_2}{t_0}\right)^k$ si $t_2 \neq t_0$, ou $ak + b$ si $t_2 = t_0$.

Si $t_2 = t_0$, on ne peut trouver de distribution stationnaire.

Si $t_2 > t_0$, il n'y a pas non plus de solution.

Si $t_2 < t_0$, il faut $a = 0$ et donc $\Pi_0 = bt_2$, $\Pi_1 = \frac{bt_2(1-t_0)}{t_0}$ et $b = \frac{t_0 - t_2}{t_2}$.

Remarques

- Si $t_0 > t_2$, la chaîne est persistante (cela découle de l'existence d'une distribution stationnaire). La probabilité que la queue ait une longueur k est équivalente à $\left(\frac{t_2}{t_0}\right)^k$
- Si $t_0 = t_2$, la chaîne est toujours persistante.
- Si $t_2 > t_0$, la chaîne n'est pas persistante.

$t_2 - t_0$ est le nombre moyen de clients supplémentaires à chaque étape.

Supposons S fini et cherchons Π tel que $\mathbf{P}\Pi = \Pi$. Π est un vecteur propre associé à la valeur propre 1 ($1 \in \mathrm{Sp}({}^t\mathbf{P})$ car ${}^t\mathbf{P}\begin{pmatrix}1\\1\\\vdots\\1\end{pmatrix} = \begin{pmatrix}1\\1\\\vdots\\1\end{pmatrix}$ et donc $1 \in \mathrm{Sp}(P)$). Montrons tout d'abord que $1 = \rho(P)$, c'est à dire est la valeur propre de plus grand module. On montre que $1 = \rho({}^t\mathbf{P})$: soit λ une valeur propre de ${}^t\mathbf{P}$, soit x un vecteur propre associé. $\lambda x_i = \sum_{j\in S} p_{i,j}x_j$ donc $|\lambda||x_i| \leq \sum_{j\in S}|p_{i,j}||x_j|$. On prend i tel que $|x_i|$ soit maximum. On obtient alors $|\lambda||x_i| \leq |x_i|\sum p_{i,j} = |x_i|$. Donc $|\lambda| \leq 1$.

14.2.2 Théorème de Perron[1]- Frobenius[2]

Soit A une matrice positive. Alors $\rho(A)$ est une valeur propre de A associée à un vecteur propre positif. De plus, si A est irréductible, $\rho(A)$ est une valeur propre simple et le vecteur propre associé est strictement positif.

Définitions

- A positive $\Leftrightarrow \forall i, j, a_{ij} \geq 0$
- A strictement positive $\Leftrightarrow \forall i, j, a_{ij} > 0$
- A est réductible si par permutations de lignes et de colonnes, elle peut se mettre sous la forme $\begin{array}{cc} & \begin{array}{cc} i & j \end{array} \\ \begin{array}{c} i \\ j \end{array} & \begin{pmatrix} * & * \\ 0 & * \end{pmatrix} \end{array}$

14.2.3 Conséquences pour les chaînes de Markov

On obtient Π tel que $\mathbf{P}\Pi = \Pi$. On peut se ramener à $\sum \Pi_j = 1$. On a donc toujours une distribution stationnaire. Si la chaîne de Markov est irréductible, cette distribution stationnaire est unique et $\forall j \in S, \Pi_j > 0$ (car la matrice de transition est irréductible).

14.3 Convergence

On prend S fini ou dénombrable.

14.3.1 Définition

La période de $i \in S$ est le pgcd$\{n \geq 1; p_{i,i}^{(n)} > 0\}$

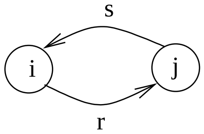
$$\left(\forall n \geq 1, p_{i,i}^{(n)} > 0 \Rightarrow t|n\right) \Leftrightarrow (t|\text{ période}(i))$$

[1]Perron (Oskar), montagnard et mathématicien allemand (Frankenthal 1880 - Munich 1975).
[2]Frobenius (Georg Ferdinand), mathématicien allemand (Berlin 1849 - Charlottenbourg 1917).

Lemme

Pour une chaîne de Markov irréductible, la période est la même pour tous les états. On l'appelle période de la chaîne.

Preuve Soit une chaîne irréductible, i et j deux états de cette chaîne. Ils sont joignables car la chaîne est irréductible.



$$p_{i,j}^{(r)} > 0$$

$$p_{j,i}^{(s)} > 0$$

Soient $t =$ période(i), m tel que $p_{j,j}^{(m)} > 0$. On veut montrer que $t|m$.

$p_{i,i}^{(r+s)} \geq p_{i,j}^{(r)} p_{j,i}^{(s)} > 0$ donc $t|r+s$.

$p_{i,i}^{(m+r+s)} \geq p_{i,j}^{(r)} p_{j,j}^{(m)} p_{j,i}^{(s)} > 0$ donc $t|m+r+s$. Et donc $t|m$ et $t|$période(j). Et par symétrie, période(j) période(i). D'où finalement période$(j) =$ période(i).

Définition

Une chaîne est apériodique si sa période est 1.

Lemme

Soit une chaîne de Markov irréductible et apériodique. Pour tous $i, j \in S$, il existe n_0 tel que $\forall n \geq n_0, p_{i,j}^{(n)} > 0$.

Preuve Soit r tel que $p_{i,j}^{(r)} > 0$

$1 = \text{pgcd}\left\{n \geq 1; p_{i,i}^{(n)} > 0\right\} = \text{pgcd}\left\{n_1, n_2, ..., n_v\right\}$ avec $\forall k \in \{1, 2, ..., v\}, p_{i,i}^{(n_k)} > 0$. Pour tout n assez grand, on peut écrire $n = a_1 n_1 + ... + a_v n_v$ avec tous les $a_k > 0$.

$$p_{i,j}^{(n+r)} \geq p_{i,i}^{(n)} p_{i,j}^{(r)} \geq \left(p_{i,i}^{(n_1)}\right)^{a_1} \left(p_{i,i}^{(n_2)}\right)^{a_2} ... \left(p_{i,i}^{(n_v)}\right)^{a_v} p_{i,j}^{(r)} > 0$$

14.3.2 Théorème de convergence

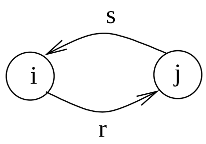
Supposons qu'il existe pour une chaîne de Markov irréductible et apériodique une distribution stationnaire Π, alors Π est unique, la chaîne est persistante, $\forall j \in S, \Pi_j > 0$ et $\lim_{n\to+\infty} p_{i,j}^{(n)} = \Pi_j$.

Remarques

- Si $t_2 > t_0$, il n'y a pas de distribution stationnaire dans la file d'attente bien que la chaîne soit irréductible et apériodique pour $t_2 \neq 0, t_1 \neq 0, t_0 \neq 0$.
- Pour n assez grand, on oublie l'état initial i. $\lim_{n\to+\infty} P(X_n = j) = \Pi_j$ si la chaîne est démarrée dans l'état i. En effet $P(X_n = j) = \sum_{i\in S} P(X_0 = i) p_{i,j}^{(n)} \longrightarrow_{n\to+\infty} \sum_{i\in S} P(X_0 = i)\Pi_j = \Pi_j$

Preuve

- Montrons d'abord que la chaîne est persistante. $\Pi_i = \sum_{j \in S} p_{i,j}^{(n)} \Pi_j$. Si la chaîne est transitoire, on a $\forall i, \Pi_i = 0$, car $\lim_{n \to +\infty} p_{i,j}^{(n)} = 0$, ce qui est contradictoire.
- La convergence des $p_{i,j}$ implique l'unicité de Π : si l'on admet la convergence, on montre que $\forall j, \Pi_j > 0$.



$$p_{i,j}^{(r)} > 0$$

$$p_{j,i}^{(s)} > 0$$

$$p_{i,i}^{(n+r+s)} \geq p_{i,j}^{(r)} p_{j,j}^{(n)} p_{j,i}^{(s)}$$

  Donc $\Pi_i \geq p_{i,j}^{(r)} \Pi_j p_{j,i}^{(s)}$
  Si $\Pi_j > 0$, alors, $\forall i \in S, \Pi_i > 0$
- Si on admet que $\lim_{n \to +\infty} |p_{i,k}^{(n)} - p_{j,k}^{(n)}| = 0$,

$$\sum_{i \in S} \Pi_i (p_{i,k}^{(n)} - p_{j,k}^{(n)}) \longrightarrow_{n \to +\infty} 0$$

$$\sum_{i \in S} \Pi_i p_{i,k}^{(n)} - \sum_{i \in S} \Pi_i p_{j,k}^{(n)} \longrightarrow_{n \to +\infty} 0$$

$$\Pi_k - p_{j,k}^{(n)} \longrightarrow_{n \to +\infty} 0$$

- prouvons maintenant ce qu'on a admis.
  Une chaîne couplée est une chaîne de Markov de la forme $(X_n, Y_n) \in S{*}S$, dont la matrice de transition est telle que $p(ij, kl) = P\left[(X_{n+1}, Y_{n+1}) = (k, l) \text{ et } (X_n, Y_n) = (i, j)\right] = p_{i,k} p_{j,l}$. Une distribution stationnaire pour la chaîne couplée vérifie $\Pi(i, j) = \Pi_i \Pi_j$. La chaîne couplée est donc persistante, et donc irréductible.
  $p^{(n)}(ij, kl) = p_{i,k}^{(n)} p_{j,l}^{(n)}$ or $\begin{cases} p_{i,k}^{(n)} > 0 & \text{pour} \quad n \geq n_0 \\ p_{j,l}^{(n)} > 0 & \text{pour} \quad n \geq n_0 \end{cases}$ car la chaîne de départ est irréductible et apériodique.
  On fixe $i_0 \in S$. $P_{i,j}\left[(X_n, Y_n) = (i_0, i_0) \text{ infiniment souvent}\right] = 1$.
  On définit l'instant de couplage par $\tau = \min\{n; X_n = Y_n = i_0\}$
  $\tau < +\infty$ avec une probabilité 1 car à partir d'un moment les deux chaînes sont couplées alors qu'elles sont parties de valeurs différentes. Donc elles ont oubliées leur passé et vont désormais se comporter identiquement.

$$p_{i,k}^{(n)} = P_{i,j}[X_n = k] \text{ et } p_{j,k}^{(n)} = P_{i,j}[Y_n = k]$$

  On veut $P_{i,j}[X_n = k, \tau \leq n] = P_{i,j}[Y_n = k, \tau \leq k]$

$$\begin{aligned} P_{i,j}[X_n = k, \tau \leq n] &= \sum_{t=0}^{n} P_{i,j}[X_n = k, \tau = t] \\ &= \sum_{t=0}^{n} P_{i,j}[X_n = k | \tau = t] P_{i,j}[\tau = t] \\ &= p_{i_0,k}^{(n-t)} P_{i,j}[\tau = t] \end{aligned}$$

$$\begin{aligned} P_{i,j}[X_n = k] - P_{i,j}[Y_n = k] = &\quad P_{i,j}[X_n = k, \tau \leq n] + P_{i,j}[X_n = k, \tau > t] \\ - &\quad (P_{i,j}[Y_n = k, \tau \leq n] + P_{i,j}[Y_n = k, \tau > t]) \end{aligned}$$

$$\begin{aligned} |p_{i,k}^{(n)} - p_{j,k}^{(n)}| &= |P_{i,j}[X_n = k, \tau > n] - P_{i,j}[Y_n = k, \tau > n]| \\ &\leq P_{i,j}[\tau > n] \\ &\to 0 \end{aligned}$$

qed

Chapitre 15

Algorithmes pour 2-SAT

15.1 Définition

Une instance F de 2-SAT est une conjonction de classes C_i. Chaque classe est une disjonction d'exactement 2 littéraux (x_i ou $\overline{x_i}$).

15.2 Algorithme

$x_1 := \varepsilon_1$, $x_2 := \varepsilon_2$, ..., $x_n := \varepsilon_n$
Où les ε_i sont choisis arbitrairement dans $\{0, 1\}$
Répéter $T(n)$ fois
  si $F(x_1, x_2, ..., x_n)$, alors arrêter
  sinon choisir une classe C telle que $\neg C(x_1, ..., x_n)$
    soit $C = n \vee v$ avec $n \in \{x_i, \overline{x_i}\}$ et $v \in \{x_j, \overline{x_j}\}$
    faire $x_i := \neg x_i$ ou $x_j := \neg x_j$ avec une probabilité $\frac{1}{2}$
déclarer F non satisfiable

Cet algorithme est un algorithme de Monte-Carlo. Il se déroule en temps moyen $O(n^2)$.

15.3 Analyse de l'algorithme

Fixons $a_1, a_2, ..., a_n$ tels que $F(a_1, a_2, ..., a_n)$. On dit que x_i est correct si $x_i = a_i$. Soit X_t le nombre de variables correctes à l'instant t. $X_{t+1} \in \{X_t - 1, X_t + 1\}$

15.3.1 Lemme

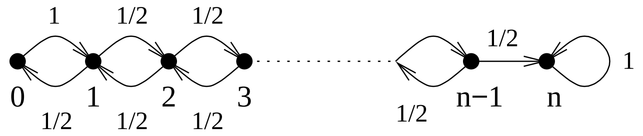
$$P[X_{t+1} = 1 + X_t] \geq \frac{1}{2}$$

Preuve Soit $C = n \vee v$ la classe choisie. $n \in \{x_i, \overline{x_i}\}$, $v \in \{x_j, \overline{x_j}\}$

Si x_i et x_j sont toutes deux incorrectes, alors $X_{t+1} = X_t + 1$.

Si une seule est incorrecte, la probabilité de la choisir est $\frac{1}{2}$.

15.3.2 Lemme



Soit Y_t l'état de la chaîne au temps t. $\forall t, k, P(X_t \geq k) \geq P(Y_t \geq k)$ si $X_0 = Y_0$.

Preuve Raisonnons par récurrence sur t.
Soit $t \geq 1$.
$P(Y_t \geq k) = P(Y_{t-1} \geq k+1) + \frac{1}{2}P(Y_{t-1} = k) + \frac{1}{2}P(Y_{t-1} = k-1)$
$P(X_t \geq k) \geq P(X_{t-1} \geq k+1) + \frac{1}{2}P(X_{t-1} = k) + \frac{1}{2}P(X_{t-1} = k-1)$
Et en appliquant l'hyothèse de récurrence, on obtient $P(X_t \geq k) \geq P(Y_t \geq k)$

15.3.3 Théorème

Soit T le temps de calcul de l'algorithme.
$E(T) = \sum_{t>0} P(T > t) \leq \sum_{t>0} P(X_t \leq n-1) \leq \sum_{t>0} P(Y_t \leq n-1)$
On veut borner $E(T')$ avec T' le temps d'arrêt de la chaîne de Markov. $E(T') = \sum_{t>0} P(T' > t) = \sum P(Y_t \leq n-1)$. On doit obtenir $E(T') = O(n^2)$.

On considère une chaîne de Markov finie, irréductible. Soit $h_{i,j} = \sum_{t=1}^{+\infty} t f_{i,j}^{(t)}$ le temps moyen pour passer en j en partant de i. Alors $\forall i, j, h_{i,j} < +\infty$

Rappel $f_{i,j}^{(t)} = P_i[X_1 \neq j, ...x X_{t-1} \neq j, X_t = j]$

Remarque $h_{i,i}$ s'interprète comme le temps moyen de retour en i.

Preuve Soit h_i le temps moyen pour atteindre l'état n. $h_n = 0$. Soit $i \in \{1, ..., n-1\}$.

$$\begin{aligned}
h_i = h_{i,n} &= \sum_{t=1}^{+\infty} t f_{i,n}^{(t)} \\
&= \sum_{t=1}^{+\infty} t \frac{f_{i-1,n}^{(t-1)} + f_{i+1,n}^{(t-1)}}{2} \\
&= \frac{1}{2}\sum t f_{i-1,n}^{(t-1)} + \frac{1}{2}\sum t f_{i+1,n}^{(t-1)} \\
&= \frac{1}{2}\sum((t-1)+1) f_{i-1,n}^{(t-1)} + \frac{1}{2}\sum((t-1)+1) f_{i+1,n}^{(t-1)} \\
&= 1 + \frac{1}{2}(h_{i-1} + h_{i+1})
\end{aligned}$$

On doit donc résoudre une récurrence linéaire à deux termes avec conditions aux bords : $h_{i+1} - 2h_i + h_{i-1} = -2$ avec $h_n = 0$ et $h_0 = h_1 + 1$.

$h_i = -i^2$est une solution particulière. La solution générale de l'équation sans second membre est $h_i = ai + b$. Donc les solutions de notre récurrence sont de la forme $h_i = ai + b - i^2$. En utilisant les conditions sur h_0 et h_n, on obtient

$$h_i = n^2 - i^2$$

Chapitre 16

Compléments sur les chaînes de Markov

16.1 Théorème

Pour une chaîne de Markov apériodique et irréductible, sans distribution stationnaire, $\forall i, j, \lim_{n\to+\infty} p_{i,j}^{(n)} = 0$.

Rappel Le temps moyen de retour en j est $\mu_j = \sum_{n=1}^{+\infty} n f_{j,j}^{(n)}$.

Lemme Soit j un état persistant. Supposons que $\lim_{n\to+\infty} p_{j,j}^{(n)} = u$. Alors $u > 0 \Leftrightarrow \mu_j > +\infty$ et dans ce cas $u = \frac{1}{\mu_j}$

16.2 Théorème de classification

Pour une chaîne irréductible et apériodique, il y a trois possiblités :

1. La chaîne est transitoire. $\forall i, j, \sum_n p_{i,j}^{(n)} < +\infty$
2. La chaîne est persistante mais il n'y a pas de distribution stationnaire. $\forall i, j, \lim_n p_{i,j}^{(n)} = 0$ mais la convergence est suffisamment lente pour que $\sum_n p_{i,j}^{(n)} = +\infty$. On a $\forall j, \mu_j = +\infty$. La chaîne est dite persistante nulle.
3. La chaîne est persistante et il existe une distribution stationnaire Π. La chaîne est dite persistante positive ou encore ergodique. Dans ce cas, $\lim_n p_{i,j}^{(n)} = \Pi_j > 0$ et $\mu_j = \frac{1}{\Pi_j}$

Exemples Premier cas : file d'attente avec $t_2 > t_0$.
Deuxième cas : file d'attente avec $t_2 = t_0$.
Troisième cas : file d'attente avec $t_2 < t_0$.

16.3 Application aux programmes

16.3.1 Définition

Un programme de longueur K est une suite d'instructions de la forme $x_0 := 1$ ou pour $1 \leq i \leq K$, $x_i := x_j \bot x_k$ où $\bot \in \{+, -, *\}$ et $0 \leq j, k < i$.

On appelle $\tau(m)$ la longueur du plus court programme qui calcule m.

16.3.2 Propriété

$\forall m \geq 1, \tau(m) \leq 2\log(m)$

Preuve $\tau(2m) \leq \tau(m) + 1$; $\tau(2m+1) \leq \tau(m) + 2$

On en déduit que $\tau(m) \leq 2\log(m)$

Remarque $\tau(2^{2^k}) = k + 1$

$\tau(2^{2^k}) \leq k + 1$ car en élevant au carré k fois à partir de 2, on obtient le résultat escompté.

$\tau(2^{2^k}) \geq k+1$ car on peut montrer qu'un programme de longueur k ne peut calculer un nombre plus grand que 2^k.

16.3.3 Problème

Est-ce que $\tau(k!) \leq (\log(k))^\lambda$ pour une certaine constante λ. (on sait que $\tau(k!) \leq 2\log(k!) \sim (\log(k))^\lambda$

16.4 Théorème

Pour presque tous les entiers,

$$\tau(m) \geq \frac{\log(m)}{\log(\log(m))}$$

Une propriété P est vraie pour presque tous les entiers si $\lim_{n \to = \infty} \frac{|\{x \in \{1,\ldots,n\}; x \text{ vérifie } P\}|}{n} = 1$

$\forall \varepsilon > 0, \tau(m) \geq \frac{\log(m)}{\log(\log(m))^{1+\varepsilon}}$

16.5 Problème

Etant donné un programme P, calcule-t-il 0 ?

Etant donnés deux programmes, calculent-ils les mêmes entiers ?

Ces deux problèmes sont équivalents.

16.5.1 Idée de l'algorithme

Remplacer $x_i := x_j \bot x_k$ par $y_i := (y_j \bot y_k) \mod p$

Si $x_0 = 0$, $y_0 = 0$

Sinon, on se trompe si $p | x_k$.

On prendra p au hasard dans $\{2, \ldots, N\}$. Quelle valeur de N faut-il prendre ?

On dira que p est un bon entier s'il ne divise pas x_k (que l'on suppose non nul).

Il y a au moins $\frac{cN}{\log(N)}$ nombres premiers dans $\{2, ..., N\}$. x_k a au plus $\log(x_k)$ diviseurs, donc il y a au moins $\frac{cN}{\log(N)} - \log(x_k)$ bons entiers dans $\{2, ..., N\}$.

Leur proportion α est supérieure à $\frac{1}{N-1}\left[\frac{cN}{\log(N)} - \log(x_k)\right]$.

$$\begin{aligned}\alpha &\geq \frac{c}{\log(N)} - \log(x_k) \\ &\geq \frac{c}{\log(N)} - \frac{2^{k-1}}{N}\end{aligned}$$

Prenons $N = 2^{2k}$. $\alpha \geq \frac{c}{2k} - \frac{2^{k-1}}{2^{2k}} \geq \frac{c'}{2k}$ pour une constante c'.