

Probabilités

Yves Robert et ses étudiants-scribes

2008-2009

Table des matières

1	Espaces de probabilités	5
1.1	Exemple	5
1.2	Définitions	5
1.3	Exemples (cf. polycopié)	6
1.3.1	Monty Hall	6
1.3.2	Paradoxes	7
1.4	Un algo min-cut	7
2	Variables aléatoires, espérance, variance	11
2.1	Définitions	11
2.2	Quicksort	11
2.3	Espérance conditionnelle	12
2.4	Variance	13
2.5	Quelques lois classiques	14
2.5.1	Bernoulli	14
2.5.2	Loi binomiale	14
2.5.3	Loi géométrique	14
2.5.4	Application : jeu des figurines (<i>Coupon collector problem</i>)	16
3	Balls and Bins (ou Boules, urnes et poissons)	17
3.1	Boules et urnes	17
3.1.1	Nombre maximum de boules dans une urne pour $m = n$	18
3.1.2	Poisson	18
3.1.3	Limite de la loi binomiale	19
3.1.4	Approximation de Poisson	19
4	Bornes et approximation de Poisson	21
4.1	Bornes	21
4.1.1	Markov	21
4.1.2	Chebyshev	21
4.1.3	Chernov	22
4.2	Algorithme pour trouver la médiane	23
5	Lois continues	27
5.1	Back to basics	27
5.2	Quelques lois continues	28
5.3	Convergences	29
5.4	Le théorème central limite	29
6	Un algorithme probabiliste	31
6.1	Rappels	31
6.1.1	Loi Binomiale	31
6.1.2	Loi de Poisson	31
6.1.3	Dans la vraie vie	31
6.2	Un joli algo de chemin hamiltonien	32
6.2.1	Graphes aléatoires	32

6.2.2	Algorithme	32
6.2.3	Algorithme Modifié	33
6.2.4	Validité de l'algorithme	34
7	Chaînes de Markov	37
7.1	Exemple introductif	37
7.2	Définition et premières propriétés	37
7.3	Chaînes de Markov irréductibles	38
7.4	Convergence	39
7.4.1	Définition	39
7.4.2	Théorème de convergence sur les chaînes de Markov régulières	39
7.4.3	Autre preuve	40
7.5	Chaînes de Markov réversibles et Monte-Carlo	41
7.5.1	Random Walk	41
7.5.2	Birth and death	42
7.5.3	Des 0 et des 1	42
7.6	2-SAT et 3-SAT	43
7.6.1	2-SAT	43
7.6.2	3-SAT	45
7.7	Jeux et paradoxes de Parrondo	47
7.7.1	Présentation des deux jeux	47
7.7.2	Analyse du jeu B	47
7.7.3	Troisième jeu	48
7.8	La Méthode probabiliste	48
7.8.1	Introduction	48
7.8.2	Lemme local de Lovasz	48
7.8.3	Exemple d'application	49
7.8.4	Preuve du théorème	50
7.8.5	Algorithme probabiliste pour k-SAT	51
8	Processus de Poisson	53
8.1	Rappels sur la loi exponentielle	53
8.2	Processus de Poisson	53
8.3	Chaînes de Markov	54
8.3.1	Distribution limite	55
8.4	Les files d'attentes M/M/1	55
9	Files d'attente	57
9.1	Processus de comptage	57
9.1.1	Binomial	57
9.1.2	Poisson	57
9.2	Loi de Little	58
9.3	Serveur Bernoulli	59
9.4	Serveur M/M/1	60
9.4.1	Régime permanent	60
9.4.2	Performances	60

Chapitre 1

Espaces de probabilités

Transcription: Chantal Keller.

1.1 Exemple

Considérons deux polynômes F et G de degrés inférieurs ou égaux à d à coefficients entiers. Comment savoir s'ils sont égaux ?

Méthode déterministe On les met tous les deux sous forme développée pour comparer les coefficients, ce qui est en $O(d^2)$ dans le pire des cas.

Méthode probabiliste On choisit un entier r uniformément au hasard entre 0 et $100d$ et on calcule $F(r) - G(r)$.

- Si $F(r) \neq G(r)$, $F \neq G$.
- Sinon, on remarque que :

$$\mathbb{P}(F(r) = G(r) \text{ et } F \neq G) \leq \frac{1}{100}$$

car $F - G$ a au plus d racines.

Si on prend r_1, \dots, r_k uniformément au hasard entre 0 et $100d$ avec remplacement (c'est-à-dire en oubliant les choix précédents), on a :

$$\mathbb{P}(\text{erreur}) \leq \left(\frac{1}{100}\right)^k$$

1.2 Définitions

Définition. Un espace de probabilités est :

- Ω un ensemble de référence ;
- \mathcal{F} une famille de sous-ensemble de Ω , dont les éléments sont appelés *événements* ;
- $\mathbb{P}() : \mathcal{F} \rightarrow \mathbb{R}$
- $E \rightarrow \mathbb{P}(E)$

tels que :

- $\forall E \in \mathcal{F}, 0 \leq \mathbb{P}(E) \leq 1$;
- $\mathbb{P}(\Omega) = 1$;
- pour toute union finie (ou dénombrable) d'évènements E_i deux à deux disjoints, $\mathbb{P}(\bigcup E_i) = \sum \mathbb{P}(E_i)$.

Lemme 1. Pour tout E_1, E_2 évènements :

$$\mathbb{P}(E_1 \cap E_2) = \mathbb{P}(E_1) + \mathbb{P}(E_2) - \mathbb{P}(E_1 \cup E_2)$$

Lemme 2. Pour tout E_i un ensemble d'évènements :

$$\begin{aligned} \mathbb{P}\left(\bigcup E_i\right) &= \sum \mathbb{P}(E_i) - \sum_{i < j} \mathbb{P}(E_i \cap E_j) + \sum_{i < j < k} \mathbb{P}(E_i \cap E_j \cap E_k) - \dots \\ \mathbb{P}\left(\bigcup E_i\right) &\leq \sum \mathbb{P}(E_i) \end{aligned}$$

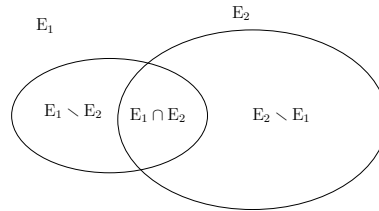


FIG. 1.1 – Union de deux évènements

Définition (Indépendance). E et F sont indépendants si et seulement si $\mathbb{P}(E \cap F) = \mathbb{P}(E)\mathbb{P}(F)$.

n évènements sont mutuellement indépendants si et seulement si $\mathbb{P}\left(\bigcap_{i \in I} E_i\right) = \prod_{i \in I} \mathbb{P}(E_i)$ pour tout sous-ensemble $I \subset \{1, 2, \dots, n\}$.

Définition (Dépendance).

$$\mathbb{P}(E|F) = \frac{\mathbb{P}(E \cap F)}{\mathbb{P}(F)}$$

Retour sur l'exemple E_i : « $F \neq G$ et on a choisi une racine r de $F - G$ à la i -ème itération». On cherche la probabilité d'erreur après k choix : $\mathbb{P}(E_1 \cap \dots \cap E_k)$.

– Avec remplacement :

$$\begin{aligned} \mathbb{P}(E_i) &= \frac{\text{nombre de racines de } F - G}{100d} \\ \mathbb{P}(E_i) &\leq \frac{d}{100d} \end{aligned}$$

– Sans remplacement :

$$\begin{aligned} \mathbb{P}(E_1) &= \frac{\text{nombre de racines de } F - G}{100d} \\ \mathbb{P}(E_2|E_1) &= \frac{(\text{nombre de racines de } F - G) - 1}{100d - 1} \\ \mathbb{P}(E_2|E_1) &\leq \frac{d - 1}{100d - 1} \\ &\vdots \\ \mathbb{P}(E_i|E_1 \cap \dots \cap E_{i-1}) &= \mathbb{P}(E_1)\mathbb{P}(E_2|E_1)\dots\mathbb{P}(E_k|E_1 \cap \dots \cap E_{k-1})\dots \end{aligned}$$

Théorème 3 (Loi de probabilité totale). Soit E_1, \dots, E_n une partition de Ω . Pour tout évènement B :

$$\mathbb{P}(B) = \sum_{i=1}^n \mathbb{P}(B \cap E_i) = \sum_{i=1}^n \mathbb{P}(B|E_i)\mathbb{P}(E_i)$$

1.3 Exemples (cf. AMS book)

1.3.1 Monty Hall

On présente trois portes à un candidat : derrière l'une d'entre elles se trouve une belle voiture à gagner, et il n'y a rien derrière les deux autres. Le candidat choisit l'une des trois portes ; Monty ouvre une des deux autres portes, derrière laquelle il n'y a rien ; le candidat peut alors choisir de garder la même porte ou d'en changer. Qu'a-t-il intérêt à faire ?

↔ Un arbre de décision montre que le candidat a deux chances sur trois de gagner s'il change de porte.

1.3.2 Paradoxes

1. (a) Il y a deux enfants dans une famille, l'un est un garçon. Quelle est la probabilité qu'il y ait deux garçons ?
- (b) Il y a deux enfants dans une famille, on rencontre le père avec un garçon dans la rue. Quelle ait la probabilité qu'il y ait deux garçons ?

↔

- (a) Au départ, chacun des évènements (garçon,garçon), (garçon,filles), (filles,garçon), (filles,filles) a une probabilité de 1/4. Comme on sait que le dernier n'est pas possible, on en déduit la probabilité cherchée :

$$p = \frac{1/4}{3/4} = \frac{1}{3}$$

- (b) Cette fois le père part se promener avec l'un de ses deux enfants, que l'on suppose choisi au hasard. Un arbre de décision tenant compte des personnes avec qui le père peut se promener montre que la probabilité cherchée est :

$$p = \frac{1}{2}$$

2. Deux enveloppes contiennent chacune une certaine somme d'argent $X \neq Y$. X et Y sont des valeurs aléatoires indépendantes. On ouvre la première enveloppe et on découvre X . Peut-on décider si $X < Y$ avec une probabilité strictement supérieure à 0.5 (ce qui semble impossible) ? Et pourtant : on tire à pile ou face et on note Z le nombre de tirages avant d'obtenir face, auquel on ajoute 1/2. Si $X < Z$, on décide que $X < Y$. Augmente-t-on ses chances de gain ?

↔ Si $X < Z < Y$ ou $Y < Z < X$ (Z est entre X et Y), ce qui est un événement de probabilité non nulle puisque $X \neq Y$, on a raison. Sinon, on a une chance sur deux d'avoir raison puisque X et Y sont tous deux soit inférieurs soit supérieurs à Z , et on a raison avec probabilité exactement 0.5 dans ce cas..

On a donc un peu plus d'une chance sur deux de gagner en appliquant cette méthode.

1.4 Produit de matrices

Soient A, B, C trois matrices à coefficients dans $\mathbb{Z}/2\mathbb{Z}$, de taille n^2 . Comment déterminer si $C = AB$?

Algorithme déterministe On calcule AB , ce qui est en $O(n^\omega)$.

Algorithme probabiliste On choisit $r = (r_1 \dots r_n) \in \{0, 1\}^n$ uniformément au hasard et on vérifie si $Cr = AB r$, ce qui s'effectue en $O(n^2)$. On remarque que choisir r uniformément revient à choisir chaque r_i uniformément et indépendamment.

Supposons que $D = C - AB \neq 0$ mais $Dr = 0$.

- $D \neq 0$, par exemple $d_{1,1} \neq 0$.

- $Dr = 0$, c'est-à-dire $\sum_{j=1}^n d_{1,j} r_j = 0$.

Donc : $r_1 = -\frac{1}{d_{1,1}} \sum_{j=2}^n d_{1,j} r_j$.

Cela arrive avec une probabilité 1/2, donc

$$\mathbb{P}(Dr = 0 \cap D \neq 0) \leq \frac{1}{2}$$

(Il est possible que seul $d_{1,1}$ soit non nul.)

Loi de Bayes

Théorème 4. *Loi de Bayes*

Soit E_1, \dots, E_n une partition

$$\mathbb{P}(E_i | B) = \frac{\mathbb{P}(E_i \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B | E_i) \mathbb{P}(E_i)}{\sum_{j=1}^n \mathbb{P}(B | E_j) \mathbb{P}(E_j)}$$

Exemple des pièces

On a 3 pièces : 2 normales et 1 faussée (qui donne face les $\frac{2}{3}$ du temps)

On tire FFP. On pose E_i « la i^e est fausse » et B « on obtient FFP ».

On veut \mathbb{P} (première pièce fausse) = $\mathbb{P}(E_1 | B)$.

On a

$$\mathbb{P}(B | E_1) = \frac{2}{3} \cdot \frac{1}{2} = \frac{1}{3}$$

$$\mathbb{P}(B | E_2) = \mathbb{P}(B | E_1) = \frac{1}{6}$$

$$\mathbb{P}(B | E_3) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$$

d'où

$$\mathbb{P}(E_1 | B) = \frac{\mathbb{P}(B | E_1) \mathbb{P}(E_1)}{\text{somme}} = \frac{\frac{1}{3}}{\frac{1}{3} + \frac{1}{6} + \frac{1}{4}} = \frac{2}{5} > \frac{1}{3}$$

Retour sur $C = AB$

Notons E : « $C = AB$ » et B : « le test renvoie OK (on a bien $Cr = ABr$) »

Sans information du tout, au départ on suppose $\mathbb{P}(E) = \mathbb{P}(\bar{E}) = \frac{1}{2}$

Or $\mathbb{P}(B | E) = 1$ et $\mathbb{P}(B | \bar{E}) \leq \frac{1}{2}$

Donc

$$\mathbb{P}(E | B) = \frac{\mathbb{P}(B | E) \mathbb{P}(E)}{\mathbb{P}(B | E) \mathbb{P}(E) + \mathbb{P}(B | \bar{E}) \mathbb{P}(\bar{E})} \geq \frac{1 \times 1/2}{1 \times 1/2 + 1/2 \times 1/2} = \frac{2}{3}$$

Deuxième passe : à la lumière du premier test, $\mathbb{P}(E) = \frac{2}{3}$, $\mathbb{P}(\bar{E}) = \frac{1}{3}$. D'où $\mathbb{P}(E | B) = \frac{4}{5}$.

Après i tests on aura $\mathbb{P}(E | B) = \frac{2^{i+1}}{2^{i+1} + 1}$.

1.5 Un algo min-cut

Transcription: Olivier Schwander.

Définition (Min-cut). Ensemble d'arêtes de cardinal minimal dont le retrait déconnecte le graphe.

Algorithme randomisé à chaque étape :

- choisir une arête au hasard
- la «contracter» (fusionner les 2 sommets)

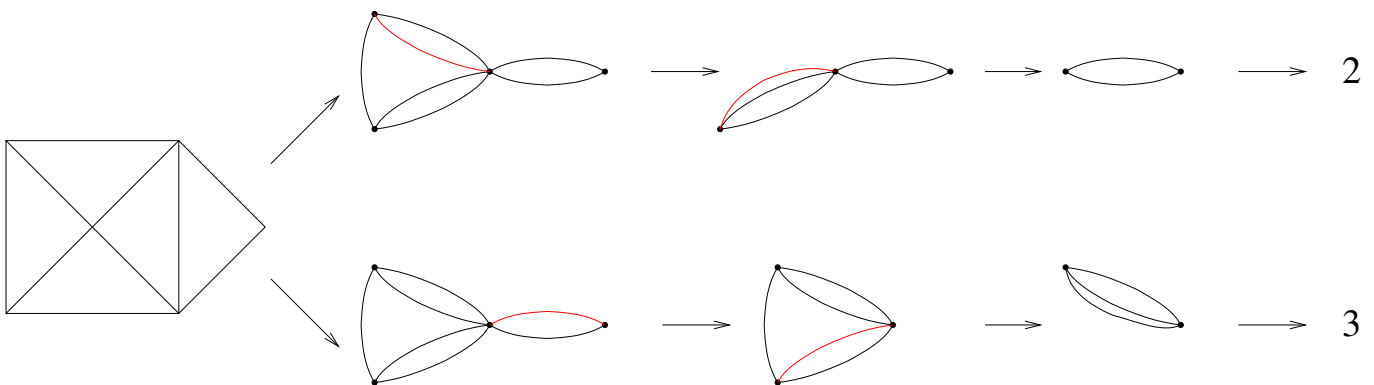


FIG. 1.2 – Déroulement de l'algorithme

Théorème 5. La probabilité de succès est supérieure à $\frac{2}{n(n-1)}$

Démonstration. Soit k la taille d'un min-cut.

Soit C un min-cut donné (il peut y en avoir plusieurs).

Soit E_i l'évènement «l'arête contractée à l'itération i n'est pas dans C ».

Soit $F_i = \bigcap_{j=1}^i E_j$.

Combien vaut $\mathbb{P}(F_{n-2})$?

Il y a au moins $\frac{kn}{2}$ arêtes dans le graphe (car $\deg v \geq k$ pour tout $v \in V$).

$$\mathbb{P}(E_1) = 1 - \frac{k}{\text{nb d'arêtes}} \geq 1 - \frac{2}{n}$$

$$\mathbb{P}(E_2|E_1) \geq 1 - \frac{2}{n-1} \quad \text{car nb d'arêtes} \geq \frac{k(n-1)}{2}$$

$$\mathbb{P}(E_i|F_{i-1}) \geq 1 - \frac{2}{n-i+1}$$

$$\begin{aligned} \mathbb{P}(F_{n-2}) &= \mathbb{P}(E_{n-2} \cap F_{n-3}) \\ &= \mathbb{P}(E_{n-2} | F_{n-3}) \mathbb{P}(F_{n-3}) \\ &= \dots \\ &= \mathbb{P}(E_{n-2} | F_{n-3}) \mathbb{P}(E_{n-3} | F_{n-4}) \dots \mathbb{P}(E_1) \\ &= \frac{2}{n(n-1)} \end{aligned}$$

■

Si on répète x fois l'algorithme et qu'on renvoie la plus petite valeur trouvée, on trouve un mauvais résultat avec un probabilité

$$p = \left(1 - \frac{2}{n(n-1)}\right)^x$$

Avec $x \geq n(n-1) \ln(n)$, on a $p \leq \frac{1}{n^2}$.

Chapitre 2

Variables aléatoires, espérance, variance

2.1 Définitions

Définition (Variable aléatoire). Une *variable aléatoire* X est une fonction de $\Omega \rightarrow \mathbb{R}$.

On note $\mathbb{P}(X = a) = \sum_{s \in \Omega, X(s)=a} \mathbb{P}(s)$.

Exemple 1. X somme de deux dés.

$$\mathbb{P}(X = 4) = \frac{3}{36}$$

Définition (Indépendance). X et Y sont deux variables *indépendantes* si et seulement si

$$\mathbb{P}(X = x \cap Y = y) = \mathbb{P}(X = x) \times \mathbb{P}(Y = y)$$

Définition (Espérance).

$$\mathbb{E}(X) = \sum_{i \in X(\Omega)} i \mathbb{P}(X = i)$$

C'est la moyenne pondérée des probabilités.

Exemple 2. Somme de deux dés? $\mathbb{E}(X) = 7$

Proposition 6.

$$\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$$

$$\mathbb{E}(cX) = c\mathbb{E}(X) \text{ pour } c \text{ constant (linéarité de l'espérance)}$$

Démonstration.

$$\begin{aligned} \mathbb{E}(X + Y) &= \sum_i \sum_j (i + j) \mathbb{P}(X = i \cap Y = j) \\ &= \sum_i i \sum_j \mathbb{P}(X = i \cap Y = j) + \sum_j j \sum_i \mathbb{P}(X = i \cap Y = j) \\ &= \sum_i i \mathbb{P}(X = i) + \sum_j j \mathbb{P}(X = j) \text{ par la loi de probabilité totale} \\ &= \mathbb{E}(x) + \mathbb{E}(y) \end{aligned}$$

Note vrai même si X et Y ne sont pas indépendantes :

$$\mathbb{E}(X + 2X^2) = \mathbb{E}(X) + 2\mathbb{E}(X^2)$$

■

2.2 Quicksort

Quicksort classique ENTRÉE : un ensemble S de n éléments

SORTIE : S trié

Choisir un élément de S comme pivot : y

Partitionner autour du pivot : S_1 et S_2

Appel récursif sur S_1 et S_2

Quicksort randomisé (RandQS) ENTRÉE : un ensemble S de n éléments

SORTIE : S trié

Choisir *au hasard* (chaque élément a autant de chances d'être choisi) un élément de S comme pivot : y

Partitionner autour du pivot : S_1 y S_2

Appel récursif sur S_1 et S_2

Complexité soit X une variable aléatoire donnant le nombre de comparaisons de RandQS.

On cherche $\mathbb{E}(X)$.

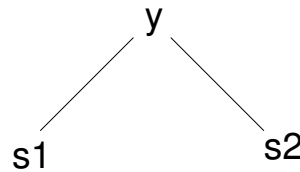
Pour $1 \leq i \leq n$, $s(i)$ est le i^e élément de la liste triée ($s(1)$ est le plus petit, $s(n)$ le plus grand).

$X_{i,j}$ est la VA qui vaut 1 si on a comparé $s(i)$ et $s(j)$ dans l'algorithme et 0 sinon.

$$\mathbb{E}(X_{i,j}) = \mathbb{P}(\text{comparaison de } s(i) \text{ et } s(j)) = p_{i,j}$$

$$\mathbb{E}(X) = \sum_{i=1}^n \sum_{j=i+1}^n \mathbb{E}(X_{i,j}) \text{ par linéarité de l'espérance}$$

Exécution de l'algorithme :



Le parcours infixe de l'arbre donne la liste triée.

Lemme 7. On compare $s(i)$ et $s(j)$ si et seulement si l'un des deux apparaît dans le parcours en largeur de l'arbre avant tous les $s(\ell), i < \ell < j$.

Démonstration. Une comparaison se fait toujours entre un pivot et un autre élément. On compare donc i et j si et seulement si l'un des deux est pris comme pivot avant qu'un autre pivot ℓ ne les ai séparés dans deux sous-tableaux triés indépendamment.

Donc

$$p_{ij} = \frac{i \text{ ou } j}{\text{tous les cas}} = \frac{2}{j - i + 1}$$

$$\begin{aligned} \mathbb{E}(X) &= \sum_{i=1}^n \sum_{j=i+1}^n \frac{2}{j - i + 1} \\ &\leq \sum_{i=1}^n \sum_{k=1}^n \frac{2}{k} \\ &\leq 2n \sum_{i=1}^n \frac{1}{i} \end{aligned}$$

or $\sum_{i=1}^n \frac{1}{i} = H_n = \ln(n) + \gamma + o(1)$ où H_n est le n -ème nombre harmonique et γ la constante d'Euler ■

Transcription: Florian Hatat.

2.3 Espérance conditionnelle

Définition (Espérance conditionnelle).

$$\mathbb{E}(Y | Z = z) = \sum_y y \mathbb{P}(Y = y | Z = z)$$

Exercice 1. On considère deux dés, et on note X_1 la valeur du premier, X_2 celle du second, et X la somme des deux. Que valent $\mathbb{E}(X | X_1 = 2)$ et $\mathbb{E}(X_1 | X = 5)$?

$$\begin{aligned} \mathbb{E}(X | X_1 = 2) &= \sum_{2 \leq x \leq 12} x \mathbb{P}(X = x | X_1 = 2) \\ &= \sum_{3 \leq x \leq 8} x \mathbb{P}(X = x | X_1 = 2) \\ &= 3 \times \frac{1}{6} + 4 \times \frac{1}{6} + 5 \times \frac{1}{6} + 6 \times \frac{1}{6} + 7 \times \frac{1}{6} + 8 \times \frac{1}{6} = \frac{11}{2} \end{aligned}$$

$$\begin{aligned} \mathbb{E}(X_1 | X = 5) &= \sum_{1 \leq x \leq 4} x \mathbb{P}(X_1 = x | X = 5) \\ \mathbb{P}(X_1 = x | X = 5) &= \frac{\mathbb{P}(X_1 = x \cap X = 5)}{\mathbb{P}(X = 5)} = \frac{1/36}{4/36} = \frac{1}{4} \\ \mathbb{E}(X_1 | X = 5) &= 1 \times \frac{1}{4} + 2 \times \frac{1}{4} + 3 \times \frac{1}{4} + 4 \times \frac{1}{4} = \frac{5}{2} \end{aligned}$$

Proposition 8. Pour toutes variables aléatoires X et Y :

$$\mathbb{E}(X) = \sum_y \mathbb{P}(Y = y) \mathbb{E}(X | Y = y)$$

Démonstration.

$$\begin{aligned} \sum_y \mathbb{P}(Y = y) \mathbb{E}(X | Y = y) &= \sum_y \sum_x x \mathbb{P}(Y = y) \mathbb{P}(X = x | Y = y) \\ &= \sum_x \sum_y x \mathbb{P}(X = x \cap Y = y) \\ &= \sum_x x \mathbb{P}(x) = \mathbb{E}(X) \quad \blacksquare \end{aligned}$$

2.4 Variance

Définition (Variance).

$$\text{Var}(X) = \mathbb{E}((X - \mathbb{E}(X))^2) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$$

La variance caractérise la distribution des points par rapport à la moyenne.

$$\text{Var}(X) = \sum_x (x - \mu)^2 \mathbb{P}(X = x) \quad \text{où } \mu = \mathbb{E}(X)$$

Si a et b sont deux constantes :

$$\text{Var}(aX + b) = a^2 \text{Var}(X)$$

(b « disparaît », car la variance est centrée par rapport à la moyenne.)

Définition. Écart-type

$$\sigma(X) = \sqrt{\text{Var}(X)}$$

L'écart-type est homogène à X .

Proposition 9. Les égalités suivantes sont vérifiées si X et Y sont indépendantes :

$$\begin{aligned} \mathbb{E}(XY) &= \mathbb{E}(X) \mathbb{E}(Y) \\ \text{Var}(X + Y) &= \text{Var}(X) + \text{Var}(Y) \end{aligned}$$

La variance est linéaire pour des variables indépendantes (ce n'est pas vrai dans le cas général).

Démonstration.

$$\begin{aligned}\mathbb{E}(XY) &= \sum_{(x,y)} xy\mathbb{P}(X = x \cap Y = y) \\ &= \sum_{(x,y)} xy\mathbb{P}(X = x)\mathbb{P}(Y = y) \quad \text{car } X \text{ et } Y \text{ sont indépendantes} \\ &= \mathbb{E}(X)\mathbb{E}(Y)\end{aligned}$$

$$\begin{aligned}\text{Var}(X + Y) &= \mathbb{E}((X + Y)^2) - (\mathbb{E}(X + Y))^2 \\ &= \mathbb{E}(X^2) + 2\mathbb{E}(XY) + \mathbb{E}(Y^2) - \mathbb{E}(X)^2 - \mathbb{E}(Y)^2 - 2\mathbb{E}(X)\mathbb{E}(Y) \\ &= (\mathbb{E}(X^2) - \mathbb{E}(X)^2) + (\mathbb{E}(Y^2) - \mathbb{E}(Y)^2) \\ &= \text{Var}(X) + \text{Var}(Y) \quad \blacksquare\end{aligned}$$

2.5 Quelques lois classiques

2.5.1 Bernoulli

On tire une pièce, qui tombe sur pile (succès, $Y = 1$) avec une probabilité p , et sur face (échec, $Y = 0$) avec une probabilité $1 - p$. Alors :

$$\begin{aligned}\mathbb{E}(Y) &= p \\ \text{Var}(Y) &= p(1 - p)\end{aligned}$$

2.5.2 Loi binomiale

On fait n tirages avec la pièce, avec une probabilité p de succès, et $1 - p$ d'échec, et on note X le nombre de succès.

$$\mathbb{P}(X = x) = C_n^x p^x (1 - p)^{n-x}$$

D'après la formule du binôme, la somme des probabilités vaut bien 1.

Espérance (première méthode)

$$\begin{aligned}\mathbb{E}(X) &= \sum_{x=0}^n x C_n^x p^x (1 - p)^{n-x} \\ &= np \sum_{x=0}^n \frac{(n-1)!}{(x-1)!(n-1-(x-1))!} p^{x-1} (1-p)^{n-1-(x-1)} \\ &= np\end{aligned}$$

Espérance (seconde méthode) On note X_i la variable qui vaut 1 si l'on a obtenu un succès au i^{e} tirage, 0 sinon. Alors $X = X_1 + \dots + X_n$, et $\mathbb{E}(X_1) = \dots = \mathbb{E}(X_n)$.

$$\mathbb{E}(X) = n\mathbb{E}(X_1) = np$$

Variance Les X_i sont indépendantes, donc :

$$\text{Var}(X) = \text{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \text{Var}(X_i) = np(1 - p)$$

2.5.3 Loi géométrique

On tire une pièce avec une probabilité $p > 0$ de succès, et on note X le nombre de tirages avant d'obtenir un premier succès.

$$\mathbb{P}(X = x) = (1 - p)^{x-1} p \quad x = 1, 2, \dots, \infty$$

On a bien $\sum_{x=1}^{\infty} \mathbb{P}(X = x) = 1$ (somme d'une série géométrique).

Espérance (méthode laborieuse)

$$\mathbb{E}(X) = \sum_{x=1}^{\infty} x(1-p)^{x-1}p$$

On reconnaît la dérivée d'une série formelle $f(q) = \sum_{z=0}^{\infty} q^z$:

$$f'(q) = \sum_{z=1}^{\infty} zq^{z-1} = \frac{1}{(1-q)^2}$$

Donc :

$$\mathbb{E}(X) = \frac{1}{p}$$

Espérance (méthode astucieuse)

Lemme 10. *Si X est une variable aléatoire sur \mathbb{N} , alors :*

$$\mathbb{E}(X) = \sum_{i=1}^{\infty} \mathbb{P}(X \geq i)$$

Démonstration.

$$\begin{aligned} \sum_{i=1}^{\infty} \mathbb{P}(X \geq i) &= \sum_{i=1}^{\infty} \sum_{j=i}^{\infty} \mathbb{P}(X = j) \\ &= \sum_{j=1}^{\infty} \sum_{i=1}^j \mathbb{P}(X = j) \quad (\text{échange possible car tout est positif}) \\ &= \sum_{j=1}^{\infty} j \mathbb{P}(X = j) = \mathbb{E}(X) \end{aligned}$$

■

Donc si X est géométrique :

$$\begin{aligned} \mathbb{P}(X \geq i) &= \sum_{n=i}^{\infty} (1-p)^{n-1}p \\ &= (1-p)^{i-1} \underbrace{(1 + (1-p) + \dots)}_{=1} p \\ &= (1-p)^{i-1} \\ \sum_{i=1}^{\infty} \mathbb{P}(X \geq i) &= \sum_{i=1}^{\infty} (1-p)^{i-1} \\ \mathbb{E}(X) &= \frac{1}{p} \end{aligned}$$

Espérance (weird) X est le nombre de tirages avant d'obtenir un premier succès. On note Y la variable aléatoire (Bernoulli) :

$$Y = \begin{cases} 0 & \text{si le premier tirage est un échec} \\ 1 & \text{sinon} \end{cases}$$

Alors, avec la loi de l'espérance conditionnelle :

$$\mathbb{E}(X) = \mathbb{P}(Y = 0) \mathbb{E}(X | Y = 0) + \underbrace{\mathbb{P}(Y = 1) \mathbb{E}(X | Y = 1)}_{=p}$$

– Si $Y = 1$, alors $X = 1$ et $\mathbb{E}(X | Y = 1) = 1$.

- Si $Y = 0$, soit Z le nombre de tirages restant avant le premier succès : intuitivement, « $X = Z + 1$ si le premier tirage est un échec ».

$$\begin{aligned}\mathbb{P}(Y = 0) \mathbb{E}(X | Y = 0) &= (1 - p) \mathbb{E}(Z + 1) \\ &= (1 - p) \mathbb{E}(Z) + (1 - p)\end{aligned}$$

Lemme 11 (Propriété de sans-mémoire). *Si X est géométrique de paramètre p :*

$$\forall n > 0, \mathbb{P}(X = n + k | X > k) = \mathbb{P}(X = n)$$

Démonstration.

$$\begin{aligned}\mathbb{P}(X = n + k | X > k) &= \frac{\mathbb{P}(X = n + k \cap X > k)}{\mathbb{P}(X > k)} = \frac{\mathbb{P}(X = n + k)}{\mathbb{P}(X > k)} \\ &= \frac{(1 - p)^{n+k-1} p}{(1 - p)^k} = (1 - p)^{n-1} p \\ &= \mathbb{P}(X = n)\end{aligned}$$

■

Donc $\mathbb{E}(Z) = \mathbb{E}(X)$, d'où $\mathbb{E}(X) = (1 - p)\mathbb{E}(X) + 1$, c'est-à-dire $\mathbb{E}(X) = 1/p$.

Variance En exo ! Lire le poly de l'ACM.

$$\text{Var}(X) = \frac{1 - p}{p^2}$$

2.5.4 Application : jeu des figurines (*Coupon collector problem*)

Dans chaque boîte de vache qui rit, il y a un coupon parmi n possibles. Les coupons sont tous équiprobables, distribués indépendamment, et on n'échange pas de coupons. Soit X le nombre de boîtes à acheter pour avoir les n coupons différents. Que vaut $\mathbb{E}(X)$?

Soit X_i le nombre de boîtes achetées entre le moment où l'on a exactement $i - 1$ coupons différents, et le moment où l'on obtient le i^{e} . Par exemple, $X_1 = 1$.

$$X = \sum_{i=1}^n X_i$$

Chaque X_i est géométrique, de paramètre $p_i = \frac{n-i+1}{n}$.

$$\mathbb{E}(X) = \sum_{i=1}^n \mathbb{E}(X_i) = \sum_{i=1}^n \frac{1}{p_i} = \sum_{i=1}^n \frac{n}{n - (i - 1)} = n \underbrace{\sum_{j=1}^n \frac{1}{j}}_{=H_n}$$

H_n est le n^{e} nombre harmonique : $H_n = \ln n + \gamma + o(1)$

$$\mathbb{E}(X) = nH_n = n \ln n + O(n)$$

Chapitre 3

Balls and Bins (ou Boules, urnes et poissons)

Transcription: Lionel Rieg.

Échauffement : les anniversaires m personnes se trouvent dans une salle. Y en a-t-il deux nées le même jour ? Notons D l'évènement « Tous les anniversaires sont distincts ».

$$\mathbb{P}(D) = \frac{\binom{365}{m} m!}{365^m} = 1 \left(1 - \frac{1}{365}\right) \left(1 - \frac{2}{365}\right) \dots \left(1 - \frac{m-1}{365}\right)$$

Pour $m \geq 23$, $\mathbb{P}(D) > \frac{1}{2}$.

Généralisons avec m boules (les personnes) qu'on veut placer dans n urnes (les dates d'anniversaires). On note E l'évènement « toutes les boules sont dans des urnes distinctes ».

$$\mathbb{P}(E) = 1 \left(1 - \frac{1}{n}\right) \left(1 - \frac{2}{n}\right) \dots \left(1 - \frac{m-1}{n}\right) = \prod_{j=1}^{m-1} \underbrace{\left(1 - \frac{j}{n}\right)}_{\approx e^{-\frac{j}{n}} \text{ si } j \ll n} \approx e^{-\frac{m(m-1)}{2n}} \approx e^{-\frac{m^2}{2n}}$$

Pour avoir $\mathbb{P}(E) < 1/2$ avec $n = 365$, on trouve $m = \sqrt{2n \ln 2} \approx 22,49$. Pour le problème des anniversaires, si on pose

- E_k « la k^{e} personne a son anniversaire différent des $k-1$ premières »
- F_k « les k premières personnes ont un anniversaire en commun »,

on trouve

$$\mathbb{P}(F_k) = \mathbb{P}(\overline{E_1} \cup \dots \cup \overline{E_k}) \leq \sum_{i=1}^k \underbrace{\mathbb{P}(\overline{E_i})}_{\leq \frac{i-1}{n}} \leq \frac{k(k-1)}{2n}$$

On atteint l'égalité si les $i-1$ premiers anniversaires sont tous distincts. Si $k \leq \lfloor \sqrt{n} \rfloor$, $\mathbb{P}(F_k) < 1/2$.

Supposons $\overline{F_{\lfloor \sqrt{n} \rfloor}}$, i.e. les $\lfloor \sqrt{n} \rfloor$ premières personnes ont des dates d'anniversaires différentes. On regarde les $\lfloor \sqrt{n} \rfloor$ personnes suivantes :

$$\mathbb{P}(\overline{F_{\lfloor 2\sqrt{n} \rfloor}}) < \mathbb{P}(\overline{F_{\lfloor 2\sqrt{n} \rfloor}} \mid \overline{F_{\lfloor \sqrt{n} \rfloor}}) \leq \left(1 - \frac{\lfloor \sqrt{n} \rfloor}{n}\right)^{\lfloor \sqrt{n} \rfloor} < \frac{1}{e}$$

Ainsi $\mathbb{P}(F_k) < \frac{1}{2}$ pour $k \leq \sqrt{n}$ et $\mathbb{P}(F_k) > 1 - \frac{1}{e} > \frac{1}{2}$ pour $k > 2\sqrt{n}$.

3.1 Boules et urnes

On se donne dans toute cette partie m boules et n urnes. D'après le résultat précédent, si $m = \Omega(\sqrt{n})$, on a une « bonne probabilité » qu'une urne ait au moins deux boules.

3.1.1 Nombre maximum de boules dans une urne pour $m = n$

Le nombre moyen de boules attendu dans une urne est évidemment m/n . Mais quel est le nombre maximum de boules attendu dans une urne? Notons B_i le nombre de boules dans l'urne i .

Théorème 12. $\mathbb{P}(\max B_i \geq \frac{3 \ln n}{\ln \ln n}) \leq \frac{1}{n}$ pour n assez grand.

Démonstration. Établissons tout d'abord une inégalité : $\int_{x-1}^x \ln t \, dt \leq \ln x$ d'où par sommation $\int_1^x \ln t \, dt \leq \ln(x!)$. Mais $\int_1^x \ln t \, dt = x \ln x - x + 1 = x \ln \frac{x}{e} + 1$. Finalement, $x \ln \frac{x}{e} \leq \ln \frac{x!}{e}$ donc $\frac{1}{x!} \leq \frac{e}{x!} \leq (\frac{e}{x})^x$.

Considérons maintenant la première urne.

$$\mathbb{P}(B_1 \geq M) \leq \binom{m}{M} \left(\frac{1}{n}\right)^M = \binom{n}{M} \left(\frac{1}{n}\right)^M \leq \frac{n^M}{M!} \left(\frac{1}{n}\right)^M = \frac{1}{M!} \leq \left(\frac{e}{M}\right)^M$$

On a donc $\mathbb{P}(\max B_i \geq M) \leq n \left(\frac{e}{M}\right)^M$.

Pour n assez grand pour que $\frac{\ln \ln \ln n}{\ln \ln n} \leq \frac{1}{3}$ et $M \geq \frac{3 \ln n}{\ln \ln n}$,

$$\begin{aligned} n \left(\frac{e}{M}\right)^M &\leq n \left(\frac{e \ln \ln n}{3 \ln n}\right)^{\frac{3 \ln n}{\ln \ln n}} \leq n \left(\frac{\ln \ln n}{\ln n}\right)^{\frac{3 \ln n}{\ln \ln n}} \\ &\leq n e^{(\ln \ln \ln n - \ln \ln n) \frac{3 \ln n}{\ln \ln n}} \leq e^{-2 \ln n + 3 \ln n \frac{\ln \ln \ln n}{\ln \ln n}} \leq e^{-\ln n} = \frac{1}{n} \quad \blacksquare \end{aligned}$$

3.1.2 Poisson

Quel(le) est le nombre (ou la la fraction) d'urnes vides?

$\mathbb{P}(B_1 = 0) = \left(1 - \frac{1}{n}\right)^m \approx e^{-\frac{m}{n}}$ Si X est la variable aléatoire représentant le nombre d'urnes vides, $\mathbb{E}(X) = n \left(1 - \frac{1}{n}\right)^m \approx n e^{-\frac{m}{n}}$ et la fraction d'urnes vides est donc $e^{-\frac{m}{n}}$.

Quelle est la probabilité d'avoir deux boules dans une même urne?

$$\mathbb{P}(B_1 = 2) = \binom{m}{2} \left(\frac{1}{n}\right)^2 \left(1 - \frac{1}{n}\right)^{m-2} \approx \frac{m^2}{2!} \frac{1}{n^2} \underbrace{\left(1 - \frac{1}{n}\right)^m}_{\approx e^{-\frac{m}{n}}} \approx \frac{1}{2} \left(\frac{m}{n}\right)^2 e^{-\frac{m}{n}}$$

Définition (loi de Poisson). La loi de Poisson de paramètre μ est définie par $\mathbb{P}(X = j) = e^{-\mu} \frac{\mu^j}{j!}$.

Propriétés.

- $\sum_{j=0}^{+\infty} \mathbb{P}(X = j) = 1$
- $\mathbb{E}(X) = \mu$
- La somme de deux variables aléatoires indépendantes suivant une loi de Poisson suit une loi de Poisson.

Démonstration 1 (par le calcul). Soit X et Y des variables aléatoires suivant des lois de Poisson de paramètres respectifs μ_1 et μ_2 .

$$\begin{aligned} \mathbb{P}(X + Y = j) &= \sum_{k=0}^j \mathbb{P}(X = k \text{ et } Y = j - k) \\ &= \sum_{k=0}^j \mathbb{P}(X = k) \mathbb{P}(Y = j - k) \quad \text{par indépendance} \\ &= \sum_{k=0}^j e^{-\mu_1} e^{-\mu_2} \frac{\mu_1^k}{k!} \frac{\mu_2^{j-k}}{(j-k)!} \\ &= \frac{e^{-(\mu_1 + \mu_2)}}{j!} \underbrace{\sum_{k=0}^j \frac{j!}{k!(j-k)!} \mu_1^k \mu_2^{j-k}}_{(\mu_1 + \mu_2)^j} \quad \blacksquare \end{aligned}$$

Démonstration 2 (par les fonctions caractéristiques). Lorsque X et Y sont indépendantes, e^{tX} et e^{tY} sont également indépendantes. D'où (c'est général) :

$$g_{X+Y}(t) = \mathbb{E}(e^{t(X+Y)}) = \mathbb{E}(e^{tX} e^{tY}) = \mathbb{E}(e^{tX}) \cdot \mathbb{E}(e^{tY}) = g_X(t) \cdot g_Y(t)$$

Or si x suit une loi de Poisson de paramètre μ ,

$$g_X(t) = \mathbb{E}(e^{tX}) = \sum_{k=0}^{+\infty} \frac{e^{-\mu} \mu^k}{k!} e^{tk} = e^{-\mu} \sum_{k=0}^{+\infty} \frac{(\mu e^t)^k}{k!} = e^{-\mu} e^{\mu e^t} = e^{\mu(e^t - 1)}$$

Finalement, $g_X(t)g_Y(t) = e^{\mu_1(e^t - 1)} e^{\mu_2(e^t - 1)} = e^{(\mu_1 + \mu_2)(e^t - 1)}$. ■

3.1.3 Limite de la loi binomiale

Théorème 13. Si $(X_n)_{n \in \mathbb{N}}$ est une suite de variables aléatoires suivant des lois binomiales de paramètres n et p_n et si $\lim_{n \rightarrow +\infty} n \cdot p_n = \mu \in \mathbb{R}$, alors $\forall k \geq 0$, $\lim_{n \rightarrow +\infty} \mathbb{P}(X_n = k) = e^{-\mu} \frac{\mu^k}{k!}$.

Autrement dit, $X_n \xrightarrow{\text{loi}} \text{Poisson}(\mu)$.

Démonstration. Notons tout d'abord que

- $\lim_{n \rightarrow +\infty} p_n = 0$
- $1 - x = e^{-x} + o(x)$
- $(a + o(\frac{1}{n}))^n \xrightarrow{n \rightarrow \infty} a^n$

Pour k fixé,

$$\begin{aligned} \mathbb{P}(X_n = k) &= \frac{n!}{k!(n-k)!} p_n^k (1-p_n)^{n-k} \\ &= \frac{1}{k!} \underbrace{n(n-1)\dots(n-k+1)}_{\rightarrow 1} \frac{1}{n^k} \underbrace{\left(\frac{np_n}{\mu}\right)^k}_{\rightarrow 1} \mu^k \underbrace{\left(1 - \frac{\mu}{n} \frac{np_n}{\mu}\right)^n}_{1+o(1)} \underbrace{(1-p_n)^{-k}}_{\rightarrow 1} \\ &\quad \left(1 - \frac{\mu}{n} + o\left(\frac{1}{n}\right)\right)^n = \left(e^{-\frac{\mu}{n} + o\left(\frac{1}{n}\right)}\right)^n \rightarrow e^{-\mu} \\ &\rightarrow e^{-\mu} \frac{\mu^k}{k!} \end{aligned}$$

■

3.1.4 Approximation de Poisson

La distribution des boules dans une urne donnée peut être approximée par une loi de Poisson de paramètre $\mu = m/n$. Mais nous allons être plus précis. On se donne toujours m boules et n urnes. On pose

- $X_i^{(m)}$ une variable aléatoire qui compte le nombre de boules dans l'urne i ;
- $Y_i^{(m)}$ une variable aléatoire indépendante des $Y_j^{(m)}$ ($j \neq i$) et qui suit une loi de Poisson de paramètre m/n .

Individuellement, chaque Y_i approxime X_i . Mais on connaît aussi le nombre total de boules, on peut en tirer parti :

Théorème 14. Pour tout m , la distribution de $(Y_1^{(m)}, \dots, Y_n^{(m)})$ conditionnée à $\sum_{i=1}^n Y_i^{(m)} = m$ est celle de $(X_1^{(m)}, \dots, X_n^{(m)})$.

Démonstration. On a d'une part, si $\sum k_i = m$,

$$\mathbb{P}\left(X_1^{(m)} = k_1, X_2^{(m)} = k_2, \dots, X_n^{(m)} = k_n\right) = \frac{m!}{k_1! k_2! \dots k_n! n^m}$$

et d'autre part,

$$\begin{aligned} &\mathbb{P}\left(Y_1^{(m)} = k_1, \dots, Y_n^{(m)} = k_n \mid \sum_{i=1}^n Y_i^{(m)} = m\right) \\ &= \frac{\mathbb{P}\left((Y_1^{(m)} = k_1) \cap \dots \cap (Y_n^{(m)} = k_n)\right)}{\mathbb{P}\left(\sum_{i=1}^n Y_i^{(m)} = m\right)} \\ &= \frac{\prod_{i=1}^n \frac{e^{-\frac{m}{n}} \left(\frac{m}{n}\right)^{k_i}}{k_i!}}{e^{-m} \frac{m^m}{m!}} \\ &= \frac{e^{-m} \left(\frac{m}{n}\right)^m}{k_1! k_2! \dots k_n!} \bigg/ e^{-m} \frac{m^m}{m!} = \frac{m!}{k_1! k_2! \dots k_n! n^m} \end{aligned}$$

car les $Y_i^{(m)}$ sont indépendantes et $\sum_{i=1}^n Y_i^{(m)}$ suit une loi de Poisson de paramètre m . ■

Théorème 15. Si f est une fonction croissante, alors

$$\mathbb{E} \left(f(X_1^{(m)}, \dots, X_n^{(m)}) \right) \leq e\sqrt{m} \mathbb{E} \left(f(Y_1^{(m)}, \dots, Y_n^{(m)}) \right)$$

« Un événement qui va arriver avec une probabilité (petite) p^* pour les variables de Poisson va arriver avec une probabilité inférieure à $p^*e\sqrt{m}$ en vrai ».

Lemme 16. $\forall x \in \mathbb{R}, x! \leq e\sqrt{x} \left(\frac{x}{e}\right)^x$

Démonstration. Par concavité de \ln , $\frac{\ln(i-1)+\ln i}{2} \leq \int_{i-1}^i \ln t \, dt$

Par sommation, $\sum_{i=2}^x \ln i - \frac{\ln x}{2} \leq \int_1^x \ln t \, dt$ i.e. $\ln(x!) - \frac{\ln x}{2} \leq x \ln x - x + 1$.

D'où $x! \leq e^{x \ln x - x + 1 + \frac{\ln x}{2}} = e^{1+x(\ln x - 1)} e^{\frac{\ln x}{2}} = e \left(\frac{x}{e}\right)^x \sqrt{x}$ ■

Démonstration du théorème.

$$\begin{aligned} & \mathbb{E} \left(f(Y_1^{(m)}, \dots, Y_n^{(m)}) \right) \\ &= \sum_{k=0}^{+\infty} \mathbb{E} \left(f(Y_1^{(m)}, \dots, Y_n^{(m)}) \mid \sum_{i=1}^n Y_i^{(m)} = k \right) \cdot \mathbb{P} \left(\sum_{i=1}^n Y_i^{(m)} = k \right) \\ &\geq_{(k=m)} \mathbb{E} \left(f(Y_1^{(m)}, \dots, Y_n^{(m)}) \mid \sum_{i=1}^n Y_i^{(m)} = m \right) \cdot \mathbb{P} \left(\sum_{i=1}^n Y_i^{(m)} = m \right) \\ &\geq \mathbb{E} \left(f(X_1^{(m)}, \dots, X_n^{(m)}) \right) \cdot e^{-m} \frac{m^m}{m!} \end{aligned}$$

Or $e^{-m} \frac{m^m}{m!} \geq \frac{e^{-m} m^m}{e\sqrt{m} \left(\frac{m}{e}\right)^m} = \frac{1}{e\sqrt{m}}$, d'où le résultat. ■

Chapitre 4

Bornes et approximation de Poisson

Transcription: Mathilde Noual.

4.1 Bornes

4.1.1 Markov

Proposition 17. Soit une variable aléatoire $X \geq 0$. $\forall a > 0$,

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}(X)}{a}$$

Démonstration. Soit $I = \begin{cases} 1 & \text{si } X \geq a \\ 0 & \text{sinon} \end{cases}$

On a $I \leq \frac{X}{a}$ et $\mathbb{E}(I) = \mathbb{P}(I = 1) = \mathbb{P}(X \geq a) \leq \mathbb{E}\left(\frac{X}{a}\right) = \frac{\mathbb{E}(X)}{a}$. ■

4.1.2 Chebyshev

Proposition 18. $\forall a > 0$,

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq a) \leq \frac{\text{Var}(X)}{a^2}$$

Démonstration. On pose $Y = (X - \mathbb{E}(X))^2$. Comme $\mathbb{E}(Y) = \text{Var}(X)$, on applique Markov sur $Y \geq 0$ et on obtient :

$$\mathbb{P}(|X - \mathbb{E}(X)| \geq a) = \mathbb{P}((X - \mathbb{E}(X))^2 \geq a^2) = \mathbb{P}(Y \geq a^2) \leq \frac{\mathbb{E}(Y)}{a^2} = \frac{\text{Var}(X)}{a^2}$$

■

Exemple 3. On tire une pièce n fois.

X = nombre de fois que la pièce tombe sur face. C'est une variable aléatoire binomiale de paramètres $n, \frac{1}{2}$. $\mathbb{E}(X) = \frac{n}{2}$ et $\text{Var}(X) = \frac{n}{4}$.

– Avec Markov : $\mathbb{P}(X \geq \frac{3n}{4}) \leq \frac{2}{3}$

– Avec Chebyshev : $\mathbb{P}(X \geq \frac{3n}{4}) \leq \mathbb{P}(|X - \mathbb{E}(X)| \geq \frac{n}{4}) \leq \frac{4}{n}$

Exemple 4. “Coupon Collector problem”

Dans chaque boîte de biip il y a un coupon parmi n possibles. Soit X le nombre de boîtes à acheter avant de les avoir tous.

$$\mathbb{E}(X) = nH_n \text{ et } X = \sum_{i=1}^n X_i$$

où X_i est le nombre de boîtes achetées entre le moment où on obtient le $(i - 1)^e$ coupon et le moment où on obtient le i^e (cf. cours précédent).

$$\begin{aligned} \text{Var}(X) &= \sum_{i=1}^n \text{Var}(X_i) \\ &= \sum_{i=1}^n \frac{1-p_i}{p_i^2} \\ &\leq \sum_{i=1}^n \frac{1}{p_i^2} \\ &\leq n^2 \sum_{i=1}^n \frac{1}{i^2} \\ &\leq n^2 \sum_{i=1}^{\infty} \frac{1}{i^2} \\ &= n^2 \frac{\pi^2}{6} \end{aligned}$$

- Avec Markov : $\mathbb{P}(X \geq 2nH_n) \leq \frac{nH_n}{2nH_n} = 1/2$
- Avec Chebyshev : $\mathbb{P}(X \geq 2nH_n) \leq \mathbb{P}(|X - \mathbb{E}(X)| \geq nH_n) \leq \frac{\text{Var}(X)}{(nH_n)^2} \leq \frac{\pi^2}{6H_n^2} = \mathcal{O}\left(\frac{1}{\ln(n)^2}\right)$
- Autre méthode :
 Probabilité de ne pas avoir obtenu le i^e coupon après x étapes : $p_{i,x} = (1 - \frac{1}{n})^x \leq e^{-\frac{x}{n}}$.

$$\mathbb{P}(X \geq x) \leq \sum_{i=1}^n p_{i,x} \leq \sum_{i=1}^n e^{-\frac{x}{n}} = ne^{-\frac{x}{n}}$$

Soit si $x = nH_n + cn \leq 2n \ln(n)$ pour n assez grand (avec c constante positive),

$$\mathbb{P}(X \geq x) \leq \frac{e^{-c}}{n}$$

4.1.3 Chernov

Fonction génératrice

Soit X une variable aléatoire et

$$\begin{aligned} g(t) &= \mathbb{E}(e^{tX}) \\ &= \sum_{j=0}^{\infty} e^{x_j t} \mathbb{P}(X = x_j). \end{aligned}$$

Les sommes à l'infini ont un sens si le domaine de X est fini car, dans ce cas, les séries sont bornées.

$$g(t) = \mathbb{E}\left(\sum_{j=0}^{\infty} \frac{X^k t^k}{k!}\right) = \sum_{j=0}^{\infty} \frac{\mathbb{E}(X^k) t^k}{k!}$$

On pose $\mu_0 = 1$ pour toute variable aléatoire.

$$\mu_1 = \mathbb{E}(X),$$

$\mu_k = \mathbb{E}(X^k)$ est le **moment d'ordre k** de X ,

Dérivée k^e en 0 : $g^{(k)}(0) = \mu_k$,

$$\text{Var}(X) = \mu_2 - \mu_1^2 = g''(0) - g'(0)^2.$$

Exemple 5. Soit X une variable aléatoire binomiale de paramètres n et p .

$$\begin{aligned} g(t) &= \sum_{j=0}^{\infty} e^{tj} \binom{n}{j} p^j (1-p)^{n-j} = (pe^t + 1 - p)^n \\ g'(t) &= npe^t (pe^t + 1 - p)^{n-1} \\ g''(t) &= np(e^t(pe^t + 1 - p)^{n-1} + pe^{2t}(n-1)(pe^t + 1 - p)^{n-2}) \end{aligned}$$

$$\text{Var}(X) = g''(0) - g'(0)^2 = (np + np^2(n-1)) - n^2p^2 = np(1-p)$$

Exemple 6. Soit X une variable aléatoire géométrique de paramètre p .

$$g(t) = \sum_{j=1}^{\infty} e^{tj} (1-p)^{j-1} p = pe^t \sum_{j=0}^{\infty} ((1-p)e^t)^j = \frac{pe^t}{1 - (1-p)e^t}$$

(valable à condition que $(1-p)e^t < 1$ i.e. $t < -\ln(1-p)$)

Bornes de Chernov

$\forall t > 0$ arbitraire, on a, en appliquant Markov :

$$\mathbb{P}(X \geq a) = \mathbb{P}(e^{tX} \geq e^{ta}) \leq \frac{\mathbb{E}(e^{tX})}{e^{ta}}$$

donc,

$$\mathbb{P}(X \geq a) \leq \min_{t>0} \left(\frac{\mathbb{E}(e^{tX})}{e^{ta}} \right)$$

4.2 Algorithme pour trouver la médiane

Définition. La médiane d'un ensemble \mathcal{S} totalement ordonné de n éléments est l'élément m de \mathcal{S} tel que au moins $\lfloor \frac{n}{2} \rfloor$ éléments de \mathcal{S} lui sont inférieurs ou égaux et au moins $\lfloor \frac{n}{2} \rfloor + 1$ éléments de \mathcal{S} lui sont supérieurs ou égaux.

Elle se trouve en $\mathcal{O}(n \log n)$ ou en $\mathcal{O}(n)$ par un algorithme déterministe.

Principe : trouver 2 éléments d et u de \mathcal{S} tels que $d \leq m \leq u$ et tels que l'ensemble $\{s \in \mathcal{S} \mid d \leq s \leq u\}$ soit petit.

Entrée : un ensemble \mathcal{S} de n éléments distincts. (On suppose que n est impair et que $n^{\frac{1}{4}}$ est un entier)

Sortie : *mediane*(\mathcal{S}) ou ÉCHEC.

Algorithme :

1. \mathcal{R} = ensemble des "échantillons" $\leftarrow n^{\frac{3}{4}}$ éléments de \mathcal{S} choisis avec replacement (i.e. ces choix d'éléments sont indépendants).
2. Trier \mathcal{R} .
3. Soient d le $\frac{n^{\frac{3}{4}}}{2} - \sqrt{n}$ plus petit élément de \mathcal{R} et u le $\frac{n^{\frac{3}{4}}}{2} + \sqrt{n}$ plus grand élément de \mathcal{R} .
4. $C \leftarrow \{s \in \mathcal{S} \mid d \leq s \leq u\}$
 $l_d \leftarrow rg_{\mathcal{S}}^-(d) = |\{s \in \mathcal{S} \mid s < d\}|$
 $l_u \leftarrow rg_{\mathcal{S}}^+(u) = |\{s \in \mathcal{S} \mid s > u\}|$
5. Si $l_d > \frac{n}{2}$ ou $l_u > \frac{n}{2}$ alors ÉCHEC.
6. Si $|C| \geq 4n^{\frac{3}{4}}$ alors ÉCHEC sinon trier C .
7. Renvoyer le $(\frac{n}{2} - l_d + 1)^e$ élément de C trié.

Lemme 19. Si l'algorithme rend quelque chose alors c'est la médiane et elle est obtenue en temps linéaire.

Démonstration. – la taille de \mathcal{R} a été choisie pour que l'étape 2. se fasse en temps $\mathcal{O}(|\mathcal{R}| \log(|\mathcal{R}|)) = \mathcal{O}(n)$.
– C se trouve en $\mathcal{O}(n)$, l_d et l_u aussi puisqu'il suffit de parcourir \mathcal{S} et comparer ses éléments avec d et u .
– la taille de C permet son tri $\mathcal{O}(n^{\frac{3}{4}} \log(n^{\frac{3}{4}})) = \mathcal{O}(n)$
donc toutes les étapes se font en $\mathcal{O}(n)$.
– Grâce à 5., la médiane $m \in [d, u]$. ■

Proposition 20. $\mathbb{P}(\text{ÉCHEC}) \leq \frac{1}{n^{\frac{1}{4}}}$

Démonstration. Soient les évènements :

- $\mathcal{E}_1 : Y_1 = |\{r \in \mathcal{R} \mid r \leq m\}| \leq \frac{n^{\frac{3}{4}}}{2} - \sqrt{n}$
 - $\mathcal{E}_2 : Y_2 = |\{r \in \mathcal{R} \mid r \geq m\}| \leq \frac{n^{\frac{3}{4}}}{2} - \sqrt{n}$
 - $\mathcal{E}_3 : |C| \geq 4n^{\frac{3}{4}}$
-

Lemme 21. L'algorithme renvoie ÉCHEC si et seulement si au moins un des trois évènements $\mathcal{E}_1, \mathcal{E}_2, \mathcal{E}_3$ s'est produit.

Démonstration. $\mathcal{E}_3 \rightarrow$ échec à l'étape 6. $\mathcal{E}_1 \Leftrightarrow l_d > \frac{n}{2}$, $\mathcal{E}_2 \Leftrightarrow l_u > \frac{n}{2} \rightarrow$ échec à l'étape 5. ■

Lemme 22. $\mathbb{P}(\mathcal{E}_1) \leq \frac{1}{4n^{\frac{1}{4}}}$

Démonstration. Soit la variable aléatoire

$$X_i = \begin{cases} 1 & \text{si le } i^{\text{e}} \text{ échantillon est inférieur ou égal à } m \\ 0 & \text{sinon} \end{cases}$$

X_i suit une loi de Bernoulli de paramètre $p_i = \frac{\frac{n-1}{2}+1}{n} = \frac{1}{2} + \frac{1}{2n}$.

Comme les éléments de \mathcal{R} sont choisis "avec remplacement", les X_i sont indépendants. \mathcal{E}_1 est équivalent à

$$Y = \sum_{i=1}^{\frac{n^{\frac{3}{4}}}{2}} X_i \leq \frac{n^{\frac{3}{4}}}{2} - \sqrt{n}$$

Y est une variable aléatoire binomiale de paramètres $n^{\frac{3}{4}}$ et $p = p_i = \frac{1}{2} + \frac{1}{2n}$.

$$\mathbb{E}(Y) = \frac{n^{\frac{3}{4}}}{2} + \frac{1}{2n^{\frac{1}{4}}}$$

$$\begin{aligned} \text{Var}(Y) &= n^{\frac{3}{4}}p(1-p) \\ &= n^{\frac{3}{4}}\left(\frac{1}{2} + \frac{1}{2n}\right)\left(\frac{1}{2} - \frac{1}{2n}\right) \\ &= \frac{n^{\frac{3}{4}}}{4} - \frac{1}{4n^{\frac{5}{4}}} \\ &< \frac{n^{\frac{3}{4}}}{4} \end{aligned}$$

En appliquant l'inégalité de Chebyshev (avec $a = \sqrt{n}$), on a :

$$\begin{aligned} \mathbb{P}(\mathcal{E}_1) &= \mathbb{P}\left(Y < \frac{n^{\frac{3}{4}}}{2} - \sqrt{n}\right) \\ &\leq \mathbb{P}\left(Y < \frac{n^{\frac{3}{4}}}{2} + \frac{1}{2n^{\frac{1}{4}}} - \sqrt{n}\right) \\ &= \mathbb{P}(Y < \mathbb{E}(Y) - \sqrt{n}) \\ &\leq \frac{\text{Var}(Y)}{n} \end{aligned}$$
■

\mathcal{E}_2 se traite de façon symétrique.

Lemme 23. $\mathbb{P}(\mathcal{E}_3) \leq \frac{1}{2n^{\frac{1}{4}}}$

Démonstration. Si \mathcal{E}_3 se produit, c'est qu'au moins un des deux évènements suivants se produit :

$\mathcal{E}_{3,1}$: au moins $2n^{\frac{3}{4}}$ éléments de C sont supérieurs à m ,

$\mathcal{E}_{3,2}$: au moins $2n^{\frac{3}{4}}$ éléments de C sont inférieurs à m .

Si $\mathcal{E}_{3,1}$ se produit alors il y a au moins $\frac{n}{2} + 2n^{\frac{3}{4}}$ éléments plus petits que u dans \mathcal{S} . Il y a donc $n^{\frac{3}{4}} - \sqrt{n}$ éléments de \mathcal{R} qui font partie des $\frac{n}{2} - 2n^{\frac{3}{4}}$ plus grands éléments de \mathcal{S} .

Soit $X = \sum_{i=1}^{n^{\frac{3}{4}}} X_i$, où

$$X_i = \begin{cases} 1 & \text{si le } i^{\text{e}} \text{ échantillon est parmi les } \frac{n}{2} - 2n^{\frac{3}{4}} \text{ plus grands éléments de } \mathcal{S} \\ 0 & \text{sinon} \end{cases}$$

X est binomiale de paramètres $n^{\frac{3}{4}}$ et $p = \frac{\frac{n}{2}-2n^{\frac{3}{4}}}{n} = \frac{1}{2} - \frac{2}{n^{\frac{1}{4}}}$.

$$\mathbb{E}(X) = \frac{n^{\frac{3}{4}}}{2} - 2\sqrt{n}$$

$$\text{Var}(X) = \frac{n^{\frac{3}{4}}}{4} - 4n^{\frac{1}{4}} < \frac{n^{\frac{3}{4}}}{4}$$

En appliquant l'inégalité de Chebychev, on a :

$$\begin{aligned} \mathbb{P}(\mathcal{E}_{3,1}) &= \mathbb{P}\left(X \geq \frac{n^{\frac{3}{4}}}{2} - \sqrt{n}\right) \\ &\leq \mathbb{P}\left(|X - \mathbb{E}(X)| \geq \sqrt{n}\right) \\ &\leq \frac{\text{Var}(X)}{n} \\ &< \frac{1}{4n^{\frac{1}{4}}} \end{aligned}$$

De même, $\mathbb{P}(\mathcal{E}_{3,2}) \leq \frac{1}{4n^{\frac{1}{4}}}$ et donc

$$\mathbb{P}(\mathcal{E}_3) \leq \mathbb{P}(\mathcal{E}_{3,1}) + \mathbb{P}(\mathcal{E}_{3,2}) \leq \frac{1}{2n^{\frac{1}{4}}} \quad \blacksquare$$

Conclusion : $\mathbb{P}(\text{ÉCHEC}) \leq \mathbb{P}(\mathcal{E}_1) + \mathbb{P}(\mathcal{E}_2) + \mathbb{P}(\mathcal{E}_3) \leq \frac{1}{n^{\frac{1}{4}}}$

Borne sur le nombre d'essais nécessaires

X = nombre d'appels à l'algorithme avant d'obtenir la valeur de la médiane. C'est une variable aléatoire géométrique de paramètre $p \geq 1 - \frac{1}{n^{\frac{1}{4}}}$.

$$\mathbb{E}(X) = \frac{1}{p} \leq \frac{1}{1 - \frac{1}{n^{\frac{1}{4}}}} \leq 1 + (n^{\frac{-1}{4}})^4 = 1 + \frac{1}{n} \rightarrow 1 \text{ si } n \rightarrow \infty$$

Chapitre 5

Lois continues

Transcription: personne.

5.1 Back to basics

Définition (espace probabilisé).

- Ω espace des possibles, $\omega \in \Omega$ est une expérience élémentaire ;
- $\mathcal{A} \subseteq \mathcal{P}(\Omega)$ l'ensemble des événements (ce que l'on peut observer). \mathcal{A} est une σ -*algèbre*, ou *tribu*, c'est-à-dire :
 - $\Omega \in \mathcal{A}$
 - stable par complémentaire
 - stable par union dénombrable
- $\mathbb{P} : \mathcal{A} \rightarrow [0; 1]$ une *probabilité*, c'est-à-dire
 - $\mathbb{P}(\Omega) = 1$
 - Pour toute suite $(A_n)_{n \in \mathbb{N}}$ d'événements deux à deux disjoints, $\mathbb{P}(\bigcup A_n) = \sum \mathbb{P}(A_n)$.

En d'autres termes, \mathbb{P} est compatible avec les opérations sur \mathcal{A} .

Sous-additivité : pour toute suite $(A_n)_{n \in \mathbb{N}}$, $\mathbb{P}(\bigcup A_n) \leq \sum \mathbb{P}(A_n)$.

Si $\Omega = \mathbb{R}$ (ou un intervalle de \mathbb{R}), on utilise en général la tribu des boréliens $\mathcal{B}(\mathbb{R})$, c'est la plus petite tribu contenant les intervalles ouverts de \mathbb{R} . Elle contient en particulier $\{x\}$ pour tout $x \in \mathbb{R}$ car $\{x\} = \bigcap_{n \in \mathbb{N}}]x - \frac{1}{n}; x + \frac{1}{n}[$. Attention, un élément de $\mathcal{B}(\mathbb{R})$ ne s'obtient pas toujours par un nombre fini d'union ou intersections dénombrables.

Si $\Omega = [0; 1]$, on peut poser $\mathbb{P}([a; b]) := b - a$ pour tout $a, b \in [0; 1]$ (c'est la mesure de Lebesgue).

Définition.

- Une *variable aléatoire* (v.a.) à valeur dans un ensemble E est une fonction $X : \Omega \rightarrow E$.
- $\{X \in B\}$ est l'événement $\{\omega \mid X(\omega) \in B\} = X^{-1}(B) \in \mathcal{A}$.

Note : on exige que X soit mesurable, c'est-à-dire que l'on munit E d'une tribu \mathcal{B} et l'on exige $\forall B \in \mathcal{B} \quad X^{-1}(B) \in \mathcal{A}$. Ce qui veut simplement dire que $\mathbb{P}(B)$ existe et que l'on ne considère pas n'importe quel événement.

Exemple 7. $\Omega := [0; 1]$, $\mathbb{P}([a; b]) := b - a$, $X(\omega) := 2\omega + 1$.

Alors $\mathbb{P}(1 \leq X \leq 2) = \mathbb{P}(X^{-1}([1; 2])) = \mathbb{P}([0; \frac{1}{2}]) = \frac{1}{2}$.

Avantage d'une σ -algèbre

Cela permet de définir certains événements comme « $\forall n \quad X_n \geq \frac{1}{n}$ ». En effet, c'est l'événement

$$\bigcap_n \left\{ X_n \geq \frac{1}{n} \right\}$$

On peut aussi définir « Il y a une infinité de X_n plus grands que 0 » :

$$\bigcap_k \bigcup_{n \geq k} \{X_n \geq 0\}$$

Hérésie mathématique

En pratique, on ne justifiera pas l'existence de $(\Omega, \mathcal{A}, \mathbb{P})$.

On admet donc que l'on peut construire des variables aléatoires *indépendantes* et *identiquement distribuées* (i.i.d.) sur un même espace Ω .

5.2 Quelques lois continues

Fonction de répartition

Soit X une v.a. réelle (i.e. à valeurs dans \mathbb{R}).

Définition (fonction de répartition). C'est $F : x \rightarrow \mathbb{P}(X \leq x)$.

Ainsi $\mathbb{P}(X \in]a; b]) = F(b) - F(a)$.

Proposition 24. F est la fonction de répartition d'un v.a. réelle si et seulement si F est croissante, continue à droite, $\lim_{-\infty} F = 0$ et $\lim_{+\infty} F = 1$.

On dit que X a une loi continue si sa fonction de répartition est continue. $\mathbb{P}(X = x) = F(x) - F(x^-)$ donc F est continue si et seulement si $\forall x \in \mathbb{R} \quad \mathbb{P}(X = x) = 0$.

Densité

Dans ce qui nous intéresse, F sera continue et C^1 par morceaux, on peut donc poser $f := F'$ sur les morceaux, c'est la *densité de probabilité* de X .

Propriétés.

- $\mathbb{P}(X \leq x) = \int_{-\infty}^x f$
- $\int_{-\infty}^{+\infty} f = \mathbb{P}(X \in \mathbb{R}) = 1$

Soit $h : \mathbb{R} \rightarrow \mathbb{R}$, on définit $\mathbb{E}(h(X)) = \int_{\mathbb{R}} h(x)f(x)dx$.

En particulier, $\mathbb{E}(X) = \int_{\mathbb{R}} x f(x)dx$ et le *moment d'ordre* k (pour $k \in \mathbb{N}$) est $\mathbb{E}(X^k) = \int_{\mathbb{R}} x^k f(x)dx$ s'il existe.

Presque sûr

On dit qu'un ensemble A quelconque de Ω est *négligeable* s'il existe $B \in \mathcal{A}$ tel que $A \subseteq B$ et $\mathbb{P}(B) = 0$. En particulier si $\mathbb{P}(A) = 0$.

Une propriété vraie en dehors de A est dite vraie *presque sûrement* (p.s.).

Loi uniforme sur $[a; b]$

C'est la loi notée $U([a; b])$ de densité

$$\begin{cases} \frac{1}{b-a} & \text{si } x \in [a; b] \\ 0 & \text{sinon} \end{cases}$$

Sur $[0; 1]$, $\mathbb{E}(X) = \frac{1}{2}$, $\text{Var}(X) = \frac{1}{12}$, $F(x) = x$.

Loi exponentielle

C'est la loi de densité $\lambda e^{-\lambda x}$ pour x positif, 0 sinon.

$\mathbb{E}(X) = 1/\lambda$, $\text{Var}(X) = 1/\lambda^2$.

$F(X) = P(X \leq x) = \int_0^x \lambda e^{-\lambda t} dt = 1 - e^{-\lambda x}$

Exemple 8. Soit U une v.a. uniforme sur $[0; 1]$. Soit $Y := F^{-1}(U) = \frac{-\ln(1-U)}{\lambda}$.

$\mathbb{P}(Y \leq y) = \mathbb{P}(F^{-1}(U) \leq y) = \mathbb{P}(U \leq F(y)) = F(y)$. Donc Y et X ont même loi.

Loi normale

En gros, c'est la loi de densité (symétrique) e^{-x^2} .

On veut choisir la moyenne m : on corrige en $e^{-(x-m)^2}$.

On veut aussi choisir la variance σ^2 : $e^{-\frac{(x-m)^2}{2\sigma^2}}$.

Enfin, il faut que $\int_{\mathbb{R}} f = 1$: $\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}$.

On la note $\mathcal{N}(m, \sigma)$. Sa densité est la « courbe en cloche » ou gaussienne.

On l'a conçue pour : $\mathbb{E}(X) = m$, $\text{Var}(X) = \sigma^2$.

Version centrée réduite $\mathcal{N}(0, 1)$: $\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$.

Loi naturelle pour un bruit (non borné) autour d'une valeur moyenne.

Proposition 25. Si $X_1 \sim \mathcal{N}(m_1, \sigma_1)$ et $X_2 \sim \mathcal{N}(m_2, \sigma_2)$ alors $X_1 + X_2 \sim \mathcal{N}(m_1 + m_2, \sqrt{\sigma_1^2 + \sigma_2^2})$.

5.3 Convergences

Soient $X : \Omega \rightarrow \mathbb{R}$ une v.a. réelle et $(X_n)_{n \in \mathbb{N}}$ une suite de v.a. réelles.

On note F_X et $(F_{X_n})_{n \in \mathbb{N}}$ leurs fonctions de répartition.

Définition (convergence en loi). X_n converge *en loi* vers X , noté $X_n \xrightarrow{\text{loi}} X$ si $F_{X_n}(x) \rightarrow F_X(x)$ en tout point x de continuité de F_X .

Remarque. Il n'est pas nécessaire que les X_n soient définies sur le même espace probabilisé que X .

Exercice 2 (importance de la continuité). Donner un exemple de $(X_n)_{n \in \mathbb{N}}$, X et x tels que $X_n \xrightarrow{\text{loi}} X$ mais $F_{X_n}(x) \not\rightarrow F_X(x)$.

Exercice 3. Donner un exemple de (X_n) et X telles que $X_n \xrightarrow{\text{loi}} X$ mais $\mathbb{P}(|X_n - X| > \varepsilon) > \eta$ pour $\varepsilon, \eta > 0$.

On cherche donc une notion plus forte.

Définition (convergence en probabilité). $X_n \xrightarrow{\mathbb{P}} X \iff \mathbb{P}(|X_n - X| > \varepsilon) \rightarrow 0$

Définition (convergence presque sûre). $\mathbb{P}(\{\omega \mid X_n(\omega) \not\rightarrow X(\omega)\}) = 0$

Théorème 26. $X_n \xrightarrow{\text{p.s.}} X \implies X_n \xrightarrow{\mathbb{P}} X \implies X_n \xrightarrow{\text{loi}} X$

Exercice 4. Donner un exemple de X_n et X telles que $X_n \xrightarrow{\mathbb{P}} X$ mais $X_n \not\xrightarrow{\text{p.s.}} X$

Notes.

- En revanche, on peut extraire une suite qui converge p.s.
- Pour être complet, il faudrait parler de $\mathbb{E}(|X_n - X|^r) \rightarrow 0$.
- Il y a des réciproques partielles au théorème.

5.4 Le théorème central limite

C'est le théorème qui est central.

On veut justifier l'intuition selon laquelle, si on fait la moyenne (empirique) des résultats de plusieurs réalisations d'une même expérience, on s'approche de la moyenne (l'espérance) de la v.a. décrivant cette expérience.

Théorème 27 (loi (forte) des grands nombres). Soit X_n une suite de v.a. réelles i.i.d. telles que $m := \mathbb{E}(X) < \infty$. Soit $S_n := \sum_{i=1}^n X_i$. Alors

$$\frac{S_n}{n} \xrightarrow{\text{p.s.}} m$$

Théorème 28 (central limite). Soit X_n une suite de v.a. réelles i.i.d. telles que $m := \mathbb{E}(X) < \infty$ et $\sigma^2 := \text{Var}(X) < \infty$. Soit $S_n := \sum_{i=1}^n X_i$. Alors

$$\frac{S_n - nm}{\sigma\sqrt{n}} \xrightarrow{\text{loi}} \mathcal{N}(0, 1)$$

En d'autres termes, si F est la fonction de répartition de $\mathcal{N}(0, 1)$,

$$\forall x \in \mathbb{R} \quad \mathbb{P} \left(\frac{S_n - m}{\sigma/\sqrt{n}} \leq x \right) \rightarrow F(x)$$

Exemple 9. Si X_n suit une loi de Bernoulli de paramètre p (S_n suit donc une loi binomiale de paramètres (n, p)), alors

$$\frac{S_n - np}{\sqrt{np(1-p)}} \xrightarrow{\text{loi}} \mathcal{N}(0, 1)$$

Pièges de la convergence en loi

Exercice 5. Donner un exemple de $(X_n)_{n \in \mathbb{N}}$ et X telles que $X_n \xrightarrow{\text{loi}} X$ mais $X_n - X \not\xrightarrow{\text{loi}} 0$.

Exercice 6. Donner un exemple de X, Y, Z telles que X et Y ont même loi mais XZ et YZ n'ont pas la même loi.

Exercice 7. Donner un exemple de $(F_n)_{n \in \mathbb{N}}$ fonctions de répartition telles que F_n converge simplement vers F mais F n'est pas une fonction de répartition.

Chapitre 6

Un algorithme probabiliste

Transcription: Pascal Vanier.

6.1 Rappels

6.1.1 Loi Binomiale

Si X est une variable aléatoire binomiale de paramètres n et p_n , telle que $\lim_{n \rightarrow +\infty} np_n = \mu$ (par exemple $p_n = \frac{\mu}{n}$), alors :

$$\lim_{n \rightarrow \infty} \mathbb{P}(X = k) = \frac{e^{-\mu} \mu^k}{k!}$$

6.1.2 Loi de Poisson

Si X_1 et X_2 sont deux variables aléatoires de Poisson de paramètres μ_1 et μ_2 respectivement et si elles sont indépendantes, alors $X_1 + X_2$ est une variable aléatoire qui suit la loi de Poisson de paramètre $\mu_1 + \mu_2$.

Propriété. La transformée de Laplace d'une v.a. X de Poisson de paramètre μ est $\mathbb{E}(e^{tX}) = e^{\mu(e^t-1)}$.

Lemme 29. Soit X une variable aléatoire de Poisson de paramètre μ , alors

$$\mathbb{P}(X \geq x) \leq \frac{e^{-\mu} (e\mu)^x}{x^x} = e^{x-\mu-x \ln \frac{x}{\mu}}$$

Démonstration. C'est une application de la borne de Chernoff.

$$\mathbb{P}(X \geq x) = \mathbb{P}(e^{tX} \geq e^{tx}) \leq \frac{\mathbb{E}(e^{tx})}{e^{tx}} = e^{\mu(e^t-1)-tx}$$

Il suffit alors de choisir $t := \ln\left(\frac{x}{\mu}\right)$ et de substituer dans l'équation précédente. ■

6.1.3 Dans la vraie vie

Considérons m boules distribuées au hasard dans n urnes. La distribution des boules dans une urne donnée peut être approximée par une loi de Poisson de paramètre $\mu = \frac{m}{n}$. On pose :

- $X_i^{(m)}$ la variable aléatoire qui compte le nombre de boules dans l'urne i ;
- $Y_i^{(m)}$ la variable aléatoire indépendante des $Y_j^{(m)}$ ($j \neq i$) et qui suit une loi de Poisson de paramètre $\frac{m}{n}$.

Individuellement, chaque Y_i approxime X_i . Mais on connaît aussi le nombre total de boules, on exploite ce fait. Dans l'approximation de Poisson, si on conditionne par $\sum Y_i^{(p)} = m$, alors on obtient une vraie distribution à m boules.

Le théorème 15 du Chapitre Balls and Bins dit que

$$\mathbb{P}(\text{un événement dans la vraie vie}) \leq e\sqrt{m} \mathbb{P}(\text{cet événement dans l'approximation de Poisson})$$

On retourne au coupon collector avec n coupons différents, X est la variable aléatoire représentant le nombre d'achats avant d'avoir un coupon de chaque sorte. Ce problème est similaire à Balls and Bins quand on recherche le nombre minimum de boules à lancer pour en avoir une dans chaque urne. $\mathbb{E}(X) = nH_n \sim n \ln n$ où H_n est la série harmonique tronquée à n . De plus, $\forall y \geq 0, \mathbb{P}(X \geq n \ln n + yn) \leq e^{-y}$.

Théorème 30. $\forall y > 0 \quad \lim_{n \rightarrow \infty} \mathbb{P}(X \geq n \ln n + yn) = 1 - e^{-e^{-y}}$

Lorsque $y = 4$, cette limite vaut environ 0,02.

Démonstration. On fait une approximation de Poisson, avec $m = n \ln n + yn$, Y_i est de poisson de paramètre u avec $u = \ln n + y$.

$$\mathbb{P}(\text{1re urne vide}) = e^{-u} = e^{-(\ln n + y)} = \frac{e^{-y}}{n}$$

$$\mathbb{P}(\text{aucune urne vide}) = (1 - \frac{e^{-y}}{n})^n \rightarrow e^{-e^{-y}}$$

On note U l'événement « Aucune urne n'est vide ». Si l'on pose Y le nombre de coupons achetés, $U_1 = U \cap (|Y - m| \leq \sqrt{2m \ln m})$, $U_2 = U \cap (|Y - m| \geq \sqrt{2m \ln m})$ et $\mathbb{P}(U) = \mathbb{P}(U_1) + \mathbb{P}(U_2)$, on a alors :

Lemme 31. 1. $\mathbb{P}(U_2) \rightarrow 0$

$$2. \mathbb{P}(U_1) \rightarrow \mathbb{P}(U \cap (Y = m)) = \mathbb{P}(X = m)$$

Démonstration.

1. C'est une application du lemme 29

$$\mathbb{P}(X \geq x) \leq e^{x - m - x \ln \frac{x}{m}}. \text{ On choisit alors } x := m + \sqrt{2m \ln m}.$$

$$\text{On a } \ln(1+z) \geq z - \frac{z^2}{2} \text{ si } z \geq 0, \text{ donc } \mathbb{P}(X \geq x) \leq e^{-\ln m + \sqrt{2 \frac{(\ln m)^3}{m}}} \rightarrow 0$$

2. $\mathbb{P}(U | X = k)$ est croissant avec k , on a

$$|\mathbb{P}(U_1) - \mathbb{P}(U) X = m| \leq \mathbb{P}(U | X = m + \sqrt{2m \ln m}) - \mathbb{P}(U | X = m - \sqrt{2m \ln m}) = o(1) \text{ car } 2 \frac{\sqrt{2m \ln m}}{n} \rightarrow 0$$

qui représente le nombre de nouvelles boules entre la première urne et la deuxième.

■

■

6.2 Un joli algo de chemin hamiltonien

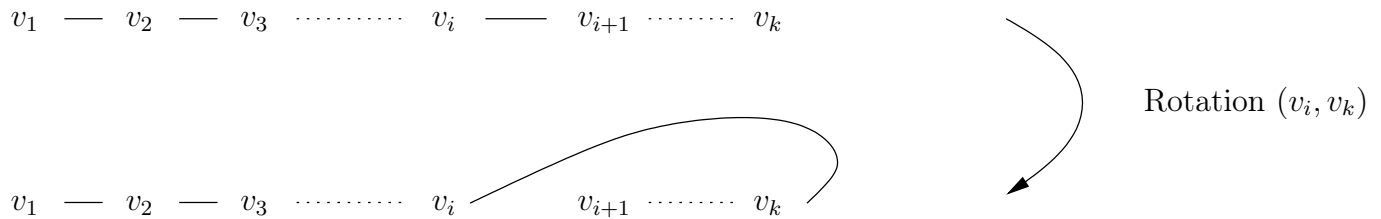
6.2.1 Graphes aléatoires

Voici deux modèles de graphes aléatoires. Le premier est celui proposé par Erdős et Rényi.

- Pour un graphe à n sommets noté $G_{n,p}$. Chaque arête (parmi les C_n^2 possibles) existe avec probabilité p . Le nombre d'arêtes suit donc une loi binomiale, son espérance est $C_n^2 \cdot p$.
- Pour un graphe à n sommets et m arêtes noté $G_{n,m}$. On prend une arête au hasard parmi les C_n^2 puis une autre parmi celles restantes et ainsi de suite, jusqu'à avoir choisi m arêtes. Si $m = n \ln n + yn$ alors $\mathbb{P}(\text{sommet isolé}) \rightarrow e^{-e^{-y}}$ (la démonstration se fait comme avant).

6.2.2 Algorithme

Pour expliquer une rotation (v_i, v_k) pour un chemin (v_1, \dots, v_k) , rien ne vaut une bonne figure :



Input : graphe $G = (V, E)$ à N sommets

Output : cycle hamiltonien ou **Échec**.

foreach $v \in V$ **do**

$\text{used}(v) := \emptyset$

$\text{unused}(v) := \{(v, u) \mid (v, u) \in E\}$

Prendre un sommet au hasard, en faire la tête du chemin.

repeat

 Soit (v_1, \dots, v_k) le chemin courant (v_k la tête)

 Soit (v, u) la 1^{re} arête de $\text{unused}(v_k)$ (si cette liste est non vide, sinon **return Échec**).

$\text{unused}(v_k) \simeq \{(v_k, u)\}$; $\text{used}(v_k) \cup = \{(v_k, u)\}$

 Pareil pour u

if $u \notin \{v_1, \dots, v_{k-1}\}$ **then**

 ajouter $u = v_{k+1}$ au chemin et la tête devient u

else

 Soit i tel que $u = v_i$

 Faire une rotation entre v_k, v_i et la tête, qui devient alors v_{i+1} .

until

– Soit on a trouvé une solution, i.e. $k = n$ et $v_i = v_1$;

– Soit la liste des arêtes inutilisées de la tête est vide, dans ce cas on a renvoyé **Échec**.

Théorème 32. Si $p \geq \frac{40 \ln n}{n}$, $\mathbb{P}(\text{trouver un chemin pour } G_{n,p}) = 1 - \mathcal{O}(\frac{1}{n})$.

On va utiliser un algorithme modifié pour établir ce théorème.

6.2.3 Algorithme Modifié

Input : graphe $G = (V, E)$ à N sommets

Output : cycle hamiltonien ou **Échec**.

foreach $v \in V$ **do**

$\text{used}(v) := \emptyset$

$\text{unused}(v) := \{(v, u) \mid (v, u) \in E\}$

Prendre un sommet au hasard, en faire la tête du chemin.

repeat

 Soit (v_1, \dots, v_k) le chemin courant (v_k la tête).

 On effectue l'une des actions suivantes aléatoirement :

with proba $\frac{1}{n}$ **do**

 renverser le chemin, la tête devient donc v_1 ;

else with proba $\frac{1}{n} |\text{used}(v_k)|$ **do**

1 choisir une arête (v_k, v_i) au hasard dans $\text{used}(v_k)$

 et faire une rotation (v_k, v_i) , la tête devient donc v_{i+1} ;

else (avec probabilité $1 - \frac{1}{n} - \frac{1}{n} |\text{used}(v_k)|$)

2 Soit (v_k, u) la première arête de $\text{unused}(v_k)$ (si cette liste est non vide, sinon **return Échec**).

if $u \notin \{v_1, \dots, v_k\}$ **then** $v_{k+1} := u$, la tête devient donc u

if $u = v_i$ **then** effectuer une rotation (v_k, v_i) , la tête devient donc v_{i+1}

 Mettre à jour unused et used .

until

– Soit on a trouvé une solution, i.e. $k = n$ et $v_i = v_1$;

– soit la liste des arêtes inutilisées de la tête est vide, dans ce cas on a renvoyé **Échec**.

Transcription: Nathalie Aubrun. g

6.2.4 Validité de l'algorithme

Lemme 33. *Supposons que les listes `used` et `unused` sont construites en insérant les sommets de manière équiprobable.*

Tant que l'algorithme peut progresser (c'est-à-dire qu'à la ligne 2 on trouve v_k tel que $\text{unused}(v_k) \neq \emptyset$) alors tous les sommets ont la même probabilité de devenir la prochaine tête.

Démonstration. Notons (v_1, \dots, v_k) le chemin en cours de calcul par l'algorithme. Quelle peut-être la prochaine tête ?

- le sommet v_1 ?

Pour cela il faut que la liste soit renversée, ce qui arrive avec probabilité $\frac{1}{n}$.

- un sommet u du chemin ?

On a deux cas possibles :

- soit $(v_k, u) \in \text{used}(v_k)$ et dans ce cas on est forcément passé par la ligne 1 de l'algorithme, avec probabilité :

$$\frac{|\text{used}(v_k)|}{n} \times \frac{1}{|\text{used}(v_k)|} = \frac{1}{n}$$

(produit de la probabilité d'être passé par cette étape et de la probabilité de choisir u) ;

- soit $(v_k, u) \in \text{unused}(v_k)$. La probabilité que $(v_k, u) \in \text{unused}(v_k)$ vaut $\frac{n-1-|\text{used}(v_k)|}{n}$ (on peut choisir tous les sommets sauf les sommets qu'on a déjà vu et le sommet en cours de traitement par l'algorithme). La probabilité de choisir précisément u parmi ces sommets (équiprobables par construction des listes) est :

$$\frac{n-1-|\text{used}(v_k)|}{n} \times \frac{1}{n-1-|\text{used}(v_k)|} = \frac{1}{n}$$

- un sommet u n'apparaissant pas dans le chemin ?

Le raisonnement est le même que dans le cas où u apparaît dans le chemin. On obtient la même probabilité :

$$\frac{|\text{used}(v_k)|}{n} \times \frac{1}{|\text{used}(v_k)|} = \frac{1}{n}$$

Tous les sommets sont donc équiprobables, à condition d'avoir construit les listes `used` et `unused` convenablement. ■

Lemme 34. *Si on construit les listes `unused` en insérant les arêtes avec une probabilité $q \geq \frac{20 \ln n}{n}$, alors l'algorithme modifié trouve un chemin hamiltonien en $\mathcal{O}(n \ln n)$ itérations et avec probabilité $1 - \mathcal{O}(\frac{1}{n})$.*

Démonstration. On définit deux évènements :

- E_1 : « L'algorithme a fonctionné $3n \ln n$ étapes sans rencontrer de sommet v tel que $\text{unused}(v) = \emptyset$, mais n'a pas trouvé de chemin hamiltonien. »
- E_2 : « Au moins une liste $\text{unused}(v)$ est s'est vidée durant les $3n \ln n$ premières étapes. »

Si on le stoppe après $3n \ln n$ étapes, l'algorithme échoue si et seulement si on a E_1 ou E_2 . On va montrer que $\mathbb{P}(E_1) \leq \frac{2}{n} = \mathcal{O}(\frac{1}{n})$ et $\mathbb{P}(E_2) \leq \frac{2}{n} = \mathcal{O}(\frac{1}{n})$, ce qui prouvera le lemme.

- $\mathbb{P}(E_1) \leq \frac{2}{n}$:

La probabilité de ne pas avoir visité un sommet u au cours des $2n \ln n$ premières étapes de l'algorithme est $\leq (1 - \frac{1}{n})^{2n \ln n} \leq \frac{1}{n^2}$ (même principe que pour le coupons collector). Donc la probabilité de ne pas avoir visité tous les sommets au cours des $2n \ln n$ premières étapes de l'algorithme est $\leq \frac{1}{n}$.

Il reste $n \ln n$ étapes, durant chacune desquelles la probabilité de fermer le chemin est $\frac{1}{n}$. La probabilité de ne pas boucler pendant ces $n \ln n$ étapes est donc $\leq (1 - \frac{1}{n})^{n \ln n} \leq \frac{1}{n}$. En sommant on trouve

- $\mathbb{P}(E_1) \leq \frac{2}{n}$.
- $\mathbb{P}(E_2) \leq \frac{2}{n}$:

On découpe E_2 en :

- E_{2a} : « Au moins $9 \ln n$ arêtes sont retirées de $\text{unused}(v_k)$ pour un certain k au cours des $3n \ln n$ premières étapes ».
- E_{2b} : « Il y avait au départ un v_k avec $|\text{unused}(v_k)| \leq 10 \ln n$ ».

Les évènements E_{2a} et E_{2b} sont tels que $\overline{E_{2a}} \wedge \overline{E_{2b}} \Rightarrow \overline{E_2}$, i.e. $\overline{E_{2a}} \cap \overline{E_{2b}} \subseteq \overline{E_2}$, donc

$\mathbb{P}(\overline{E_{2a}} \cup \overline{E_{2b}}) \leq \mathbb{P}(\overline{E_2}) = 1 - \mathbb{P}(E_2)$, i.e. $\mathbb{P}(E_2) \leq \mathbb{P}(E_{2a}) + \mathbb{P}(E_{2b})$.

Remarque. Les évènements E_{2a} et E_{2b} ne sont pas indépendants.

On montre qu'on peut majorer $\mathbb{P}(E_{2a})$ et $\mathbb{P}(E_{2b})$ par $\frac{1}{n}$.

– $\mathbb{P}(E_{2a}) \leq \frac{1}{n}$:

Que peut-on dire du nombre de fois où un sommet v_k est en tête de chemin ? D'après le lemme 33, on a une loi binomiale de coefficient $p = \frac{1}{n}$. On cherche donc à majorer $\mathbb{P}(E_{2a}) = \mathbb{P}(X \geq 9 \ln n)$. Pour cela on utilise la borne de Chernoff qui nous donne :

$$\mathbb{P}(X \geq 9 \ln n) \leq \min_{t>0} \frac{(\frac{1}{n}e^t + (1 - \frac{1}{n}))^{3n \ln n}}{e^{9t \ln n}}$$

En choisissant $t = \frac{\ln n}{3}$ on obtient $\mathbb{P}(E_{2a}) \leq \frac{1}{n^2}$ d'où $\mathbb{P}(E_{2a}) \leq \frac{1}{n}$.

– $\mathbb{P}(E_{2b}) \leq \frac{1}{n}$:

Le nombre d'arêtes dans une liste $\text{unused}(v_k)$ suit une loi binomiale $B(n-1, q)$ puisqu'il y a $n-1$ arêtes possibles que l'on insère avec probabilité q . On applique une nouvelle fois la borne de Chernoff pour majorer $\mathbb{P}(E_{2b}) = \mathbb{P}(X \leq 10 \ln n)$. Il suffit de choisir $t = -\ln n$ pour obtenir $\mathbb{P}(E_{2b}) \leq \frac{1}{n}$. ■

Corrolaire 35. *En initialisant les $\text{unused}(v_k)$ comme il faut, si on fait tourner l'algorithme modifié sur un graphe aléatoire de $G_{n,p}$ ($p \geq \frac{40 \ln n}{n}$) alors il renvoie un chemin hamiltonien avec probabilité $1 - \mathcal{O}(\frac{1}{n})$.*

Démonstration. Soit (u, v) une arête de $G_{n,p}$. Il suffit de choisir des bonnes probabilités avec lesquelles on place cette arête dans les listes $\text{unused}(u)$ et $\text{unused}(v)$. On pose $p = 2q - q^2$. On insère l'arête (u, v) dans :

- $\text{unused}(u)$ avec probabilité $p \frac{q(1-q)}{2q-q^2}$
- $\text{unused}(v)$ avec probabilité $p \frac{q(1-q)}{2q-q^2}$
- les deux avec probabilité $p \frac{q^2}{2q-q^2}$

De cette façon :

$$\begin{aligned} \mathbb{P}((u, v) \in \text{unused}(u)) &= \mathbb{P}((u, v) \in \text{unused}(u) \cap \text{unused}(v)) + \mathbb{P}((u, v) \in \text{unused}(u) \setminus \text{unused}(v)) \\ &= p \frac{q(1-q)}{2q-q^2} + p \frac{q^2}{2q-q^2} \\ &= q \end{aligned}$$

et de la même manière, $\mathbb{P}((u, v) \in \text{unused}(v)) = q$.

On a donc :

$$\begin{aligned} \mathbb{P}((u, v) \in \text{unused}(u) \cap \text{unused}(v)) &= p \frac{q^2}{2q-q^2} = q^2 \\ &= \mathbb{P}((u, v) \in \text{unused}(u)) \times \mathbb{P}((u, v) \in \text{unused}(v)) \end{aligned}$$

Donc l'appartenance de l'arête (u, v) à l'une des listes en dépend pas de son appartenance à l'autre, c'est-à-dire qu'on est dans les hypothèses du lemme 33. ■

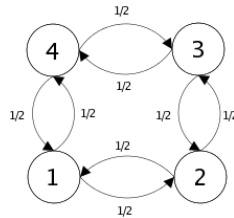
Chapitre 7

Chaînes de Markov

Transcription: Nathalie Aubrun.

7.1 Exemple introductif

Exemple 10. On veut déplacer un jeton sur le graphe suivant. Pour cela à chaque étape on lance une pièce : si elle tombe sur pile on le déplace dans le sens horaire, sinon dans le sens inverse. On part du sommet numéro 1 : $\mathbb{P}(X_0 = v_1) = 1$. Où sera le jeton à l'étape n ?



La loi de probabilité vérifie les propriétés suivantes :

- la position suivante n'est pas indépendante de la précédente ;
- la position suivante ne dépend que de la précédente, indépendamment de tout l'historique qu'on a suivi pour aller jusqu'à la précédente (X_{n+1} dépend de X_0, X_1, \dots, X_n mais cette dépendance est capturée par X_n seul).

On a par exemple :

$$\mathbb{P}(x_{n+1} = v_1 | X_n = v_2) = \frac{1}{2}$$

$$\mathbb{P}(x_{n+1} = v_3 | X_n = v_2) = \frac{1}{2}$$

$$\mathbb{P}(x_{n+1} = v_1 | X_n = v_2, X_{n-1} = \dots, \dots, X_0 = \dots) = \mathbb{P}(x_{n+1} = v_1 | X_n = v_2) = \frac{1}{2}$$

Un tel système s'appelle une chaîne de Markov.

7.2 Définition et premières propriétés

Définition. Une chaîne de Markov à k états s_1, \dots, s_k est caractérisée par la matrice de transition $P = (p_{i,j})_{1 \leq i,j \leq k}$ où $p_{i,j} = \mathbb{P}(X_{n+1} = s_j | X_n = s_i)$.

Remarque. Une telle chaîne de Markov est dite finie (nombre fini d'états) et homogène (les probabilités $p_{i,j}$ ne dépendent pas du temps).

Exemple 11. Pour l'exemple introductif, la matrice vaut :

$$\begin{pmatrix} 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \end{pmatrix}$$

Remarque. Dans le cas général, la matrice P n'a aucune raison d'être symétrique.

Proposition 36. Si P est la matrice d'une chaîne de Markov, alors :

- $\forall (i, j), 0 \leq p_{i,j} \leq 1$
- la somme des coefficients d'une ligne vaut 1 : $\forall i, \sum_{j=1}^n p_{i,j} = 1$

Proposition 37. Si $\mu_0 = (P(X_0 = s_1), \dots, P(X_0 = s_k))$ est la distribution de probabilité initiale (à l'instant $t = 0$), alors la distribution à l'instant $t = n$ vaut $\mu_n = \mu_0 P^n$.

Exemple 12. Si dans l'exemple introductif on pose au départ le jeton sur l'état 1, on a $\mu_0 = (1, 0, 0, 0)$.

Dans cet exemple, on a $P = P^{2n+1} = \begin{pmatrix} 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \end{pmatrix}$ et $P^2 = P^{2n} = \begin{pmatrix} \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix}$

Définition. Soit $P = (p_{i,j})_{1 \leq i,j \leq k}$ une chaîne de Markov. On peut lui associer un automate à k états où les transitions sont de la forme :

$$s_i \xrightarrow{p_{i,j}} s_j$$

Exemple 13. On peut modéliser la navigation sur internet par une chaîne de Markov. Soit X_n la page sur laquelle je me trouve après n clics. On construit l'automate comme suit :

- si la page i n'a pas de lien, on pose $p_{i,i} = 1$ et donc l'état s_i boucle sur lui-même avec probabilité 1 ;
- si la page i comporte d_i liens, dont un vers la page j , on ajoute la transition $s_i \xrightarrow{\frac{1}{d_i}} s_j$.

7.3 Chaînes de Markov irréductibles

Définition.

- On dit que s_i communique avec s_j (noté $s_i \rightarrow s_j$) s'il existe $n \in \mathbb{N}$ tel que $\mathbb{P}(X_n = s_j | X_0 = s_i) \neq 0$.
- On dit qu'une chaîne de Markov est irréductible si tous les états communiquent entre eux (autrement dit le graphe n'a qu'une seule composante connexe).

Définition. La période d'un état s_i est $d(s_i) = \text{pgcd} \left\{ n \mid p_{i,i}^n \neq 0 \right\}$. On dit qu'une chaîne de Markov est apériodique si $\forall i, d(s_i) = 1$.

Exemple 14. La chaîne de Markov de l'exemple introductif est de période 2 car les coefficients diagonaux des matrices P^{2n+1} sont tous nuls, alors que ceux P^2 sont tous non nuls.

Une autre façon de retrouver la période est de remarquer que le graphe est biparti entre $\{v_1, v_3\}$ et $\{v_2, v_4\}$. Tout les cycles sont donc de longueur paire, donc la période est au moins 2. Comme les $p_{i,i}^2$ sont non nuls, c'est exactement 2.

Théorème 38. Soit une chaîne de Markov à k états $\{s_1, \dots, s_k\}$ de matrice P . Si la chaîne est apériodique, il existe un entier N tel que $\forall i, \forall n \geq N, p_{i,i}^n \neq 0$.

Démonstration. Soit $A_i = \left\{ n \geq 1 \mid p_{i,i}^n \neq 0 \right\}$. Alors :

- $\text{pgcd}(A_i) = 1$ car la chaîne est apériodique.
- A_i est stable par addition. En effet s'il existe un chemin de s_i à s_i en a étapes et un autre en a' étapes, alors il existe un chemin de s_i à s_i en $a + a'$ étapes, et sa probabilité est $p_{i,i}^{a+a'} \geq p_{i,i}^a \times p_{i,i}^{a'} > 0$.

Le lemme qui suit permettra de conclure.

Lemme 39. Soit $A = \{a_1, \dots, a_n\}$ une partie de \mathbb{N} stable par addition et telle que $\text{pgcd}(a_1, \dots, a_n) = 1$. Alors il existe un entier N tel que $\forall n \geq N, n \in A$.

Démonstration. Comme $\bigwedge \{a_i, i = 1 \dots n\} = 1$ il existe des entiers k_i tels que $1 = \sum_{i=1}^n k_i a_i$ (théorème de Bezout). A priori ces entiers peuvent être négatifs. On découpe donc la somme en regroupant les k_i positifs et négatifs : $1 = M - P$ avec $M, P \in A$.

Il suffit de choisir $N = P(P - 1)$. Pour $n \geq P(P - 1)$ on effectue la division euclidienne de n par P : $n = Pq + r$. On a donc $n = Pq + r(M - P) = (q - r)P + rM$. Comme $q - r \geq 0$ ceci montre que $n \in A$. ■

Pour chaque ensemble A_i le lemme nous donne un entier N_i . Il suffit de poser $N = \max N_i$. ■

Transcription: Bertrand Marc.

7.4 Convergence

7.4.1 Définition

Définition (Chaîne de Markov régulière). On dit qu'une chaîne de Markov est *régulière* si

$$\exists N \quad \forall n \geq N \quad \forall i, j \quad P_{i,j}^n > 0$$

Proposition 40. Une chaîne de Markov irréductible et apériodique est régulière.

7.4.2 Théorème de convergence sur les chaînes de Markov régulières

Théorème 41. Si P est la matrice d'une chaîne de Markov régulière, alors il existe w_1, \dots, w_n tels que

$$\lim_{n \rightarrow +\infty} P^n = \begin{pmatrix} w_1 & \dots & w_n \\ w_1 & \dots & w_n \\ \vdots & & \vdots \\ w_1 & \dots & w_n \end{pmatrix} \quad \text{avec} \quad \sum w_i = 1$$

Lemme 42. Soit P une matrice de transition de taille $k \geq 2$ à coefficients strictement positifs et $d := \min_{i,j} P_{i,j}$.

On remarque que $0 < d \leq \frac{1}{2}$ car $k \geq 2$ et $\sum_j P_{i,j} = 1 \quad \forall i$. Soit y un vecteur arbitraire à coefficients positifs et

- $m_0 := \min_i y_i$
- $M_0 := \max_i y_i$
- $m_1 := \min_i (Py)_i$
- $M_1 := \max_i (Py)_i$

Alors
$$M_1 - m_1 \leq (1 - 2d)(M_0 - m_0)$$

Démonstration. $M_1 \leq dm_0 + (1 - d)M_0$ et $m_1 \geq dM_0 + (1 - d)m_0$, d'où le résultat. ■

Preuve du théorème. Supposons pour l'instant $P > 0$ pour être dans le cas du lemme. Soit y un vecteur arbitraire à coefficients positifs, $m_n := \min_i (P^n y)_i$ et $M_n := \max_i (P^n y)_i$.

On a $M_0 \geq M_1 \geq M_2 \dots$ et $m_0 \leq m_1 \leq \dots$. D'après le lemme 42, $M_n - m_n \rightarrow 0$ (c'est une suite géométrique de raison $1 - 2d < 1$) donc il existe $u := \lim M_n = \lim m_n$. Ainsi,

$$\forall y \quad P^n y \rightarrow \begin{pmatrix} u \\ u \\ \vdots \\ u \end{pmatrix}$$

On choisit alors $y := e_j$ ($j^{\text{ième}}$ vecteur de la base canonique) et l'on obtient la limite de la j^{e} colonne de P^n :

$$P^n y \rightarrow \begin{pmatrix} w_j \\ w_j \\ \dots \\ w_j \end{pmatrix}$$

Finalement, P^n tend vers une matrice dont toutes les lignes sont identiques.

Dans le cas général, remarquons qu'il existe un N tel que $P^N > 0$. Le raisonnement précédent montre que $M_{(nN)} - m_{(nN)} \rightarrow 0$. Or on a toujours $M_0 \geq M_1 \geq M_2 \dots$ et $m_0 \leq m_1 \leq \dots$, d'où $M_n - m_n \rightarrow 0$. On conclut comme ci-dessus. ■

7.4.3 Autre preuve

Voir le livre de l'ACM.

Lemme 43. Soit une chaîne de Markov régulière de matrice P .

1. $\exists! w \quad w = Pw$.
2. Soit $c := \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$. $Pc = c$ et $Py = y \Rightarrow y = \lambda c$.

Démonstration.

1. $P^n \rightarrow W$ donc $PP^n \rightarrow PW$. $P^{n+1} \rightarrow W$ donc $W = PW$. $w = Pw$.
2. L'espace propre est de dimension 1 (car il est en somme directe avec le noyau de dimension $n - 1$). Il n'y a donc qu'une seule valeur propre. ■

Théorème 44. Pour tout vecteur de distribution initiale u , $uP^n \rightarrow w$.

Démonstration. Soit X_0, X_1, \dots la chaîne de matrice P et distribution initiale u , et Y_0, Y_1, \dots de distribution initiale w . On fait évoluer X et Y indépendamment.

On définit une nouvelle chaîne de matrice $P^*((i, j), (k, l)) = P_{ij}P_{kl}$. Les états de la chaîne de Markov sont les couples (s_i, s_j) . Il est clair que P^* est régulière.

Soit T la première étape où P^* est dans un état (s_i, s_i) . $\lim_{n \rightarrow +\infty} \mathbb{P}(T > n) = 0$.

Puisqu'on a une chaîne de Markov régulière, il existe N tel que $P_{ij}^N > 0$. On pose $d := \min_{ij} P_{ij}^N$, et $T_j := \min \{n \geq 1 \mid X_n = s_j\}$.

Lemme 45. Pour tout j ,

$$\lim_{n \rightarrow +\infty} \mathbb{P}(T_j > n) = 0 \quad \text{et} \quad \mathbb{E}(T_j) < +\infty$$

Preuve du lemme.

$$\begin{aligned} \mathbb{P}(T_j \leq N) &\geq \mathbb{P}(X_N = s_j) = \sum_{i=1}^k \mathbb{P}(X_0 = s_i, X_N = s_j) \\ &= \sum_{i=1}^k \mathbb{P}(X_0 = s_i) \underbrace{\mathbb{P}(X_N = s_j \mid X_0 = s_i)}_{P_{ij}^N} \\ \mathbb{P}(T_j \leq N) &\geq d \\ \mathbb{P}(T_j > N) &\leq 1 - d \end{aligned}$$

$\mathbb{P}(X_{2N} = s_j \mid T > N) \geq d$ par définition de d . Ainsi $\mathbb{P}(X_{2N} \neq s_j \mid T > N) \leq 1 - d$.

$$\begin{aligned} \mathbb{P}(T_j > 2N) &= \mathbb{P}(T_j > N) \mathbb{P}(T_j > 2N \mid T_j > N) \\ &\leq (1 - d) \mathbb{P}(T_j > 2N \mid T_j > N) \\ &\leq (1 - d) \mathbb{P}(X_{2N} \neq s_j \mid T_j > N) \\ &\leq (1 - d)^2 \end{aligned}$$

De même, $\mathbb{P}(T_j > kN) \leq (1 - d)^k$. Donc $\lim_{n \rightarrow +\infty} \mathbb{P}(T_j > n) = 0$.

De même pour l'espérance :

$$\begin{aligned} \mathbb{E}(T_j) &= \sum_{n=1}^{\infty} n\mathbb{P}(T_j = n) = \sum_{n=1}^{\infty} \mathbb{P}(T_j \geq n) = \sum_{n=0}^{\infty} \mathbb{P}(T_j > n) \\ &= \sum_{k=0}^{\infty} \sum_{n=kN}^{kN+N-1} \mathbb{P}(T_j > n) \\ &\leq \sum_{k=0}^{\infty} N\mathbb{P}(T_j > kN) \\ &\leq N \sum_{k=0}^{\infty} (1-d)^k = \frac{N}{d} \end{aligned}$$



Transcription: Alexandre Derouet-Jourdan.

7.5 Chaînes de Markov réversibles et Monte-Carlo

Définition. Soit $X_0 \dots X_n$ une chaîne de Markov sur $S = \{s_1, \dots, s_k\}$, de matrice de transition P . Elle est dite réversible s'il existe un vecteur π tel que :

- $\forall i \pi_i \geq 0$ et $\sum \pi_i = 1$
- $\forall i, j \pi_i P_{i,j} = \pi_j P_{j,i}$

Un vecteur vérifiant la première propriété est dit vecteur de probabilité.

Proposition 46. Si une chaîne de Markov de matrice de transition P est réversible avec π , alors

$$\pi P = \pi$$

π est dit stationnaire.

Démonstration. Soit x le vecteur πP . On a :

$$\forall i \ x_i = \sum_j \pi_j P_{j,i} = \sum_j \pi_i P_{i,j} = \pi_i \sum_j P_{i,j} = \pi_i$$



7.5.1 Random Walk

Sur un graphe $G = (V, E)$, on considère la marche aléatoire de sommet en sommet telle que la probabilité d'aller d'un sommet i à un sommet j soit $\frac{1}{deg(i)}$ si j est voisin de i , 0 sinon. On a donc une chaîne de Markov de matrice de transition :

$$P_{i,j} = \begin{cases} \frac{1}{deg(i)} & \text{si } i \text{ et } j \text{ sont voisins} \\ 0 & \text{sinon} \end{cases}$$

On construit le vecteur π par $\pi_i = \frac{deg(i)}{\sum_k deg(k)}$ et on vérifie que P est réversible avec π : pour tout i et j ,

$$\begin{aligned} \pi_i P_{i,j} &= \frac{deg(i)}{\sum_k deg(k)} \times \frac{1}{deg(i)} = \frac{1}{\sum_k deg(k)} \\ \pi_j P_{j,i} &= \frac{deg(j)}{\sum_k deg(k)} \times \frac{1}{deg(j)} = \frac{1}{\sum_k deg(k)} \end{aligned}$$

On vérifie de plus que π est un vecteur de probabilité.

De plus, cette chaîne de Markov est irréductible. En effet, le graphe associé à la chaîne de Markov et le graphe G sont à peu près les mêmes. Comme G est connexe, le graphe associé est connexe est la chaîne de Markov est irréductible.

Elle est aussi apériodique. Enfin, tout dépend du graphe. On a l'équivalence suivante : La chaîne de Markov est apériodique si et seulement si le graphe G n'est pas biparti. Si G est biparti, alors tous les cycles

sont de longueur paire et la chaîne est apériodique. Si le graphe n'est pas biparti, alors il existe un cycle de longueur impaire qui contient un sommet s_k et pour tout sommet s_i , en considérant un chemin entre s_i et s_k puis le cycle de longueur impaire et le retour du chemin entre s_i et s_k , on a pour tout sommet s_i , un cycle de longueur impaire le contenant. Comme tous les sommets appartiennent à des cycles de longueur 2, on en déduit que la chaîne de Markov est apériodique.

7.5.2 Birth and death

Dans ce jeu on gagne ou on perd un dollar en fonction de ce qu'on a. Si on a i dollars, on gagne un dollar avec probabilité $P_{i,j}$. On a donc $P_{i,j} = 0$ si $|i - j| \geq 2$ et $P_{i,j} \neq 0$ si $|i - j| = 1$. Pour $|i - j| = 1$, on peut choisir librement $P_{i,j}$ et $P_{i,i}$ tant que P reste une matrice stochastique.

On construit alors le vecteur π de cette manière :

- $\pi_1 = a$
- $\pi_2 = \frac{aP_{1,2}}{P_{2,1}}$
- $\pi_3 = \frac{\pi_2 P_{2,3}}{P_{3,2}} = \frac{aP_{1,2}P_{2,3}}{P_{3,2}P_{2,1}}$
- ...
- $\pi_n = \frac{\pi_{n-1}P_{n-1,n}}{P_{n,n-1}} = a \frac{\prod_{i=1}^{n-1} P_{i,i+1}}{\prod_{i=1}^{n-1} P_{i+1,i}}$

Ensuite, on fixe a de manière à avoir $\sum_{i=1}^n \pi_i = 1$.

Ces conditions nécessaires à ce que la chaîne de Markov modélisant les gains soit réversible sont aussi suffisantes : pour tout i et j ,

- si $i = j$, $\pi_i P_{i,j} = \pi_j P_{j,i}$
- si $|i - j| \geq 2$, $P_{i,j} = P_{j,i} = 0$
- si $|i - j| = 1$, π a été construit pour.

Remarque. Avec tout ce qu'on vient de faire, on pourrait penser que toute chaîne de Markov est réversible. Ce n'est pas le cas, par exemple pour une marche aléatoire sur un carré, avec probabilité 1/4 de tourner dans le sens horaire et probabilité 3/4 de tourner dans le sens anti-horaire.

7.5.3 Des 0 et des 1

On se place sur une grille.

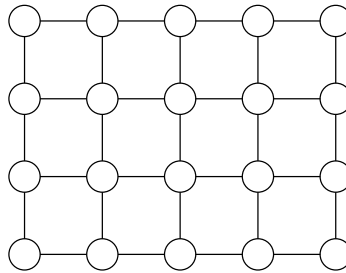


FIG. 7.1 – une grille...

Une configuration valide est une fonction de V dans $\{0, 1\}$ telle que deux sommets adjacents ne sont pas tous les deux à 1. Ce problème vient de la physique.

- Combien de configurations valides ?
- Si on prend une configuration au hasard, combien contient-elle de 1 ?

Assignons la même probabilité à toutes les configurations C :

$$\mu_G(C) = \begin{cases} \frac{1}{Z_G} & \text{si } C \text{ est valide} \\ 0 & \text{sinon} \end{cases}$$

où Z_G est le nombre de configurations valides sur la grille G .

Une bonne façon de résoudre ces problèmes est de simuler une chaîne de Markov qui converge vers la distribution stationnaire. On définit la chaîne de Markov suivante :

- L'ensemble des états est l'ensemble des configurations valides.
- On passe d'une configuration à la suivante de cette façon :
 - On tire un sommet v au hasard.

- On jette une pièce. Si c'est face et que tous les voisins de v sont à 0, on le passe à 1. Si c'est pile, on le passe à 0.
- Les autres sommets sont inchangés.

On a bien défini une chaîne de Markov car l'état X_n ne dépend que de l'état X_{n-1} .

Elle est régulière :

- Elle est irréductible. Pour aller de l'état i à l'état j ,
 - on ne tire que des piles, jusqu'à ce que la configuration soit la configuration nulle.
 - On ne tire que les sommets de j et on ne tire que des faces pour les faire passer à 1. Toutes les configurations sur le chemin ont moins de 1 que la configuration d'arrivée, donc sont valides.
- Elle est apériodique car il y a des boucles de longueurs 1.

Est-elle réversible ? On va regarder si

$$\mu_G(C)P_{C,C'} = \mu_G(C')P_{C',C}$$

On a $\mu_G(C) = \mu_G(C') = \frac{1}{Z_G}$

- Si C et C' diffèrent d'au moins 2 sommets, on a $P_{C,C'} = P_{C',C} = 0$
- Si $C = C'$, le résultat est clair.
- Si C et C' diffèrent d'exactement 1 sommet, on a $P_{C,C'} = \frac{1}{2|V|} = P_{C',C}$

On a donc bien $\mu_G(C)P_{C,C'} = \mu_G(C')P_{C',C}$.

Pour obtenir des informations sur μ_G , on part de la configuration nulle et on simule ; la loi de X_n converge vers μ_G . On obtient aussi de manière expérimentale l'espérance du nombre de 1. Pour un graphe général, il y a peu de chances qu'on puisse dire des choses non triviales sur ce nombre de 1. Si on est sur une grille 2D ou 3D, il doit y avoir quelques formules.

On peut s'interroger sur la rapidité de convergence, ce qui est traité dans le poly de l'ACM.

Transcription: Xavier Pujol.

7.6 2-SAT et 3-SAT

7.6.1 2-SAT

On considère une instance de 2-SAT à n variables x_1, \dots, x_n et k clauses, c'est à dire une formule :

$$F = (x_{u_1}^* \vee x_{v_1}^*) \wedge \dots \wedge (x_{u_k}^* \vee x_{v_k}^*) \quad (\text{où } x^* \text{ représente } x \text{ ou } \bar{x})$$

On va décrire un algorithme probabiliste pour 2-SAT. On choisit un entier m (la probabilité de succès dépendra de cet entier). L'algorithme s'exécute en temps polynomial en n et m .

Algorithme :

- Partir d'une instanciation des x_i arbitraire.
- Répéter tant que F n'est pas satisfaite, et au plus $2mn^2$ fois
 - Prendre une clause non satisfaite
 - Inverser la valeur d'un des deux littéraux (pris au hasard)
- Si F est satisfaite, renvoyer OUI, sinon renvoyer NON

Exemple 15. On prend la formule :

$$F = (x_1 \vee \bar{x}_2) \wedge (\bar{x}_1 \vee \bar{x}_3) \wedge (x_1 \vee x_2) \wedge (x_4 \vee \bar{x}_3) \wedge (\bar{x}_4 \vee \bar{x}_1)$$

et on choisit une instanciation, par exemple $(x_1, x_2, x_3, x_4) = (1, 0, 1, 1)$.

- La dernière clause $(\bar{x}_4 \vee \bar{x}_1)$ n'est pas satisfaite. On prend une de ses variables, x_4 , et on l'inverse : $x_4 = 0$.
- La clause $(x_4 \vee \bar{x}_3)$ n'est plus satisfaite. On inverse x_3 : $x_3 = 0$.
- On a trouvé une solution : $(x_1, x_2, x_3, x_4) = (1, 0, 0, 0)$.

Remarque.

- Le nombre de clauses est au plus $k = O(n^2)$.
- Une étape de la boucle s'exécute en temps polynomial.

On s'intéresse maintenant au nombre d'étapes nécessaires.

Théorème 47.

- Si F n'est pas satisfaisable, l'algorithme répond toujours NON.
- Si F est satisfaisable, l'algorithme répond OUI avec une probabilité plus grande que $1 - \frac{1}{2^m}$.

Démonstration. Il est clair que si F n'est pas satisfaisable, la réponse de l'algorithme est NON. Supposons que F soit satisfaite par une instantiation S des variables. On pose :

$$X_i = \text{nombre de variables qui concordent avec } S \text{ à l'étape } i$$

Si à une étape, $X_i = n$, alors l'algorithme s'arrête et renvoie la bonne réponse (ce n'est qu'une borne supérieure : l'algorithme peut réussir avant, car S n'est pas nécessairement la seule solution).

Comment X_i varie-t-il d'une étape à l'autre ? Si $X_i = 0$, alors $X_{i+1} = 1$. Sinon, $1 \leq X_i \leq n - 1$. A cette étape, l'algorithme choisit une clause qui n'est pas satisfaite. Au moins une de ses deux variables ne concorde pas avec S .

- Si dans la clause choisie, aucune des deux variables ne concorde avec S , alors on progresse toujours : $\mathbb{P}(X_{i+1} = j + 1 \mid X_i = j) = 1$
- Sinon, la probabilité d'inverser la bonne variable est de $\frac{1}{2}$, autrement dit $\mathbb{P}(X_{i+1} = j + 1 \mid X_i = j) = \mathbb{P}(X_{i+1} = j - 1 \mid X_i = j) = \frac{1}{2}$

Dans tous les cas,

$$\mathbb{P}(X_{i+1} = j + 1 \mid X_i = j) \geq \frac{1}{2}$$

$$\mathbb{P}(X_{i+1} = j - 1 \mid X_i = j) \leq \frac{1}{2}$$

On considère la nouvelle chaîne de Markov (Y_i) caractérisée par :

$$\begin{cases} Y_0 = 0 \\ \mathbb{P}(Y_{i+1} = 1 \mid Y_i = 0) = 1 \\ \forall j \in \llbracket 1, n - 1 \rrbracket \quad \mathbb{P}(Y_{i+1} = j + 1 \mid Y_i = j) = \mathbb{P}(Y_{i+1} = j - 1 \mid Y_i = j) = \frac{1}{2} \end{cases}$$

Intuitivement, la chaîne Y_i progresse moins bien vers n que X_i . Par récurrence sur k , on démontre que $\forall j \in \llbracket 1, n \rrbracket, \mathbb{P}(X_k \geq j) \geq \mathbb{P}(Y_k \geq j)$. C'est vrai pour $k = 0$ car $Y_0 = 0$. Soit $j \in \llbracket 1, n - 1 \rrbracket$:

$$\begin{aligned} \mathbb{P}(X_{k+1} \geq j) &\geq \mathbb{P}(X_k \geq j + 1) + \frac{1}{2}\mathbb{P}(X_k = j) + \frac{1}{2}\mathbb{P}(X_k = j - 1) \\ &= \frac{1}{2}\mathbb{P}(X_k \geq j + 1) + \frac{1}{2}\mathbb{P}(X_k \geq j - 1) \\ &\geq \frac{1}{2}\mathbb{P}(Y_k \geq j + 1) + \frac{1}{2}\mathbb{P}(Y_k \geq j - 1) \\ &= \mathbb{P}(Y_k \geq j + 1) + \frac{1}{2}\mathbb{P}(Y_k = j) + \frac{1}{2}\mathbb{P}(Y_k = j - 1) = \mathbb{P}(Y_{k+1} \geq j) \\ \mathbb{P}(X_{k+1} = n) &\geq \mathbb{P}(X_k = n) + \frac{1}{2}\mathbb{P}(X_k = n - 1) \\ &\geq \mathbb{P}(Y_k = n) + \frac{1}{2}\mathbb{P}(Y_k = n - 1) = \mathbb{P}(Y_{k+1} = n) \end{aligned}$$

En particulier, on a la formule suivante, valide pour toute distribution de probabilité sur l'état X_0 :

$$\mathbb{P}(X_k < n) = 1 - \mathbb{P}(X_k = n) \leq 1 - \mathbb{P}(Y_k = n) = \mathbb{P}(Y_k < n)$$

Démontrons maintenant que Y_i arrive assez rapidement à n . On pose :

$$h_j := \mathbb{E}(\text{nombre } a_j \text{ d'étapes qu'il faut à } Y \text{ pour atteindre } n \text{ à partir de } j)$$

Les h_j sont liés par la relation :

$$\begin{cases} h_0 = 1 + h_1 \\ \forall j \in \llbracket 1, n - 1 \rrbracket, h_j = \frac{h_{j-1}+1}{2} + \frac{h_{j+1}+1}{2} \\ h_n = 0 \end{cases}$$

On vérifie aisément que $h_j = n^2 - j^2$ est la solution, en particulier $h_0 = n^2$. Il n'y a plus qu'à relier l'espérance du nombre d'étapes à la probabilité de réussite.

On découpe les $2mn^2$ itérations de boucle en m tranches de taille $2n^2$. Sur chaque tranche $0 \leq t < m$, on introduit la variable $X_i^{(t)} = X_{2n^2t+i}$. On note $b^{(t)}$ le nombre d'étapes pour atteindre une solution à partir du début de cette tranche. On majore maintenant la probabilité d'échec de l'algorithme :

$$\begin{aligned}
\mathbb{P}(\forall t, b^{(t)} > 2n^2) &= \mathbb{P}(b^{(0)} > 2n^2) \prod_{t=1}^m \mathbb{P}(b^{(t)} > 2n^2 \mid b^{(t-1)} > 2n^2) \\
&\leq \mathbb{P}(X_{2n^2} < n) \prod_{t=1}^m \mathbb{P}(X_{2n^2}^{(t)} < n \mid X_{2n^2}^{(t-1)} < n) \\
&= \mathbb{P}(X_{2n^2} < n) \prod_{t=1}^m \mathbb{P}(X_{2n^2}^{(t)} < n \mid X_0^{(t)} < n) \\
&\leq \mathbb{P}(Y_{2n^2} < n) \prod_{t=1}^m \mathbb{P}(Y_{2n^2} < n) \\
&= (\mathbb{P}(a_0 > 2n^2))^m
\end{aligned}$$

On termine grâce à l'inégalité de Markov :

$$\mathbb{P}(a_0 > 2n^2) \leq \frac{h_0}{2n^2} = \frac{n^2}{2n^2} = \frac{1}{2}$$

Dont on déduit que :

$$\mathbb{P}(\forall t, b^{(t)} > 2n^2) \leq \left(\frac{1}{2}\right)^m \quad \blacksquare$$

Remarque. Il existe un algorithme déterministe polynomial résolvant 2-SAT.

7.6.2 3-SAT

On se donne une formule à n variables et k clauses de la forme :

$$F = (x_{u_1}^* \vee x_{v_1}^* \vee x_{w_1}^*) \wedge \cdots \wedge (x_{u_k}^* \vee x_{v_k}^* \vee x_{w_k}^*)$$

Le principe de l'algorithme est le même que pour 2-SAT, mais le nombre C d'itérations nécessaires ne sera plus polynomial.

Algorithme :

- Partir d'une instanciation des x_i arbitraire.
- Répéter tant que F n'est pas satisfaite, et au plus C fois.
 - Prendre une clause non satisfaite.
 - Inverser la valeur d'un des trois littéraux (pris au hasard).
- Si F est satisfaite, renvoyer OUI, sinon renvoyer NON.

Comme pour 2-SAT, on considère une formule F satisfaite par une instanciation S et on introduit :

$$X_i = \text{nombre de variables qui concordent avec } S \text{ à l'étape } i$$

On démontre alors que :

$$\begin{aligned}
\mathbb{P}(X_{i+1} = j + 1 \mid X_i = j) &\geq \frac{1}{3} \\
\mathbb{P}(X_{i+1} = j - 1 \mid X_i = j) &\leq \frac{2}{3}
\end{aligned}$$

Pour estimer le nombre d'étapes nécessaires, on définit (Y_i) :

$$\begin{cases} Y_0 = 0 \\ \mathbb{P}(Y_{i+1} = 1 \mid Y_i = 0) = 1 \\ \forall j \in \llbracket 1, n-1 \rrbracket, \begin{cases} \mathbb{P}(Y_{i+1} = j + 1 \mid Y_i = j) = \frac{1}{3} \\ \mathbb{P}(Y_{i+1} = j - 1 \mid Y_i = j) = \frac{2}{3} \end{cases} \end{cases}$$

Avec les mêmes notations que précédemment, on établit la relation :

$$\begin{cases} h_0 = h_1 + 1, h_n = 0 \\ h_j = \frac{2h_{j-1}}{3} + \frac{h_{j+1}}{3} + 1 \text{ pour } 1 \leq j \leq n-1 \end{cases}$$

et la solution est $h_j = 2^{n+1} - 2^{j+1} - 3(n-j)$. Autrement dit, l'algorithme est *a priori* aussi lent que l'énumération exhaustive des instanciations. Mais il existe une variante de l'algorithme qui diminue le temps d'exécution d'un facteur exponentiel :

Algorithme Répéter au plus m fois :

- Partir d'une instanciation des x_i aléatoire.
- Appliquer l'algorithme sur $C = 3n$ étapes.

Pour analyser cet algorithme, on introduit la chaîne de Markov (Z_i) définie par $Z_0 = X_0$ et :

$$\forall j \in \mathbb{Z} \quad \begin{cases} \mathbb{P}(Z_{i+1} = j+1 | Z_i = j) = \frac{1}{3} \\ \mathbb{P}(Z_{i+1} = j-1 | Z_i = j) = \frac{2}{3} \end{cases}$$

On pourrait montrer que $\forall i, \forall k \leq n, \mathbb{P}(X_i \geq k) \geq \mathbb{P}(Z_i \geq k)$.

On note q_j la probabilité que l'algorithme arrive à S en au plus $3n$ étapes en partant d'une configuration avec j variables en désaccord.

$$\begin{aligned} q_j &\geq \sum_{l \leq 3n} \mathbb{P}(Z_l = n) \\ &= \sum_{j+2k \leq 3n} \mathbb{P}(Z \text{ diminue } k \text{ fois et augmente } j+2k \text{ fois avant d'atteindre } n) \\ &\geq \mathbb{P}(Z \text{ diminue } j \text{ fois et augmente } 3j \text{ fois avant d'atteindre } n) \\ &= \binom{3j}{j} \left(\frac{2}{3}\right)^j \left(\frac{1}{3}\right)^{2j} \end{aligned}$$

Par la formule de Stirling, on trouve :

$$q_0 = 1 \quad q_j \geq \frac{\text{constante}}{\sqrt{j}} \left(\frac{1}{2}\right)^j$$

Pour conclure, on somme sur tous les j possibles :

$$\begin{aligned} q &= \mathbb{P}(\text{l'algorithme converge en } 3n \text{ étapes}) \\ &= \sum_{j=0}^n \mathbb{P}(\text{l'instanciation initiale diffère de } S \text{ pour } j \text{ variables}) \times q_j \\ &= \underbrace{\frac{1}{2^n}}_{q_0=1} + \sum_{j=1}^n \binom{n}{j} \frac{1}{2^n} q_j \\ &\geq \frac{1}{2^n} \left(1 + \frac{1}{\sqrt{n}} \sum_{j=1}^n \binom{n}{j} \left(\frac{1}{2}\right)^j \right) \\ &\geq \frac{1}{2^n} \left(1 + \frac{1}{\sqrt{n}} \left(1 + \frac{1}{2}\right)^n \right) \\ &\geq \frac{\text{constante}'}{\sqrt{n}} \left(\frac{3}{4}\right)^n \end{aligned}$$

Comme on réinitialise les variables toutes les $3n$ étapes, les applications de l'algorithme sont indépendantes. Le nombre d'applications suit donc une loi géométrique de paramètre q . On en déduit :

$$\mathbb{E}(\text{nombre d'étapes}) \leq \frac{3n}{q} = O\left(n^{\frac{3}{2}} \left(\frac{4}{3}\right)^n\right)$$

7.7 Jeux et paradoxes de Parrondo

7.7.1 Présentation des deux jeux

Jeu A : on tire une pièce biaisée

- Avec probabilité $p_A < \frac{1}{2}$, le joueur obtient face et gagne un dollar.
- Avec probabilité $1 - p_A > \frac{1}{2}$, le joueur obtient pile et perd un dollar.

C'est un jeu perdant.

Jeu B : ce jeu utilise deux pièces B et C biaisées. On note p_B et p_C les probabilités de faire face avec ces pièces. Exemple : $p_B = 0,09$ et $p_C = 0,74$. g est le gain courant du joueur.

- Si $g \equiv 0 \pmod{3}$, on tire B .
- Si $g \not\equiv 0 \pmod{3}$, on tire C .

Dans les deux cas, face rapporte un dollar et pile fait perdre un dollar.

Si $g \pmod{3}$ est uniformément distribué dans $\{1, 2, 3\}$, l'espérance du gain à cette étape est :

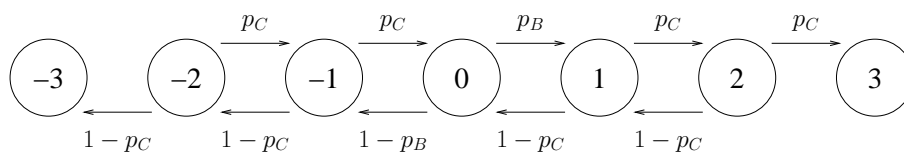
$$\frac{1}{3}p_B + \frac{2}{3}p_C = \frac{157}{300} > \frac{1}{2}$$

Le jeu est-il gagnant ? Non, car en fait, on va voir que $g \equiv 0 \pmod{3}$ plus souvent qu'une fois sur 3.

7.7.2 Analyse du jeu B

Première méthode

On étudie la chaîne de Markov X_i définie par l'automate suivant :



Le jeu B est perdant si $\mathbb{P}(\text{arriver à } -3 \mid X_0 = 0) > \mathbb{P}(\text{arriver à } 3 \mid X_0 = 0)$. Pour le déterminer, on pose :

$$z_i = \mathbb{P}(\text{arriver à } -3 \mid X_0 = i)$$

Les z_i sont liés par les relations suivantes :

$$\begin{cases} z_{-3} = 1, z_3 = 0 \\ z_i = p_C z_{i+1} + (1 - p_C) z_{i-1} \text{ pour } i \in \{-2, -1, 1, 2\} \\ z_0 = p_B z_1 + (1 - p_B) z_{-1} \end{cases}$$

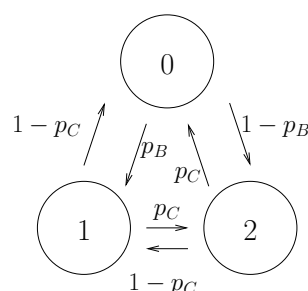
On résout ce système et on trouve :

$$z_0 = \frac{(1 - p_B)(1 - p_C)^2}{(1 - p_B)(1 - p_C)^2 + p_B p_C^2} = 0,555 > \frac{1}{2} \text{ dans l'exemple}$$

Donc le jeu B est perdant.

Deuxième méthode

On calcule la distribution limite Π de la chaîne de Markov suivante :



La matrice de transition de cette chaîne est :

$$P = \begin{pmatrix} 0 & p_B & 1 - p_B \\ 1 - p_C & 0 & p_C \\ p_C & 1 - p_C & 0 \end{pmatrix}$$

La probabilité de gagner après avoir joué suffisamment longtemps est :

$$\Pi_0 p_B + \Pi_1 p_C + \Pi_2 p_C = \Pi_0 (p_B - p_C) + p_C = \frac{86421}{175900} < \frac{1}{2}$$

Donc on démontre à nouveau que le jeu B est perdant.

7.7.3 Troisième jeu

Jeu C : on tire de manière répétée une pièce équilibrée D .

- Si D donne face, on joue au jeu A une fois.
- Si D donne pile, on joue au jeu B une fois.

Jouer à C revient à jouer au jeu B en changeant les caractéristiques des pièces :

$$\begin{cases} p_B^* = \frac{1}{2}(p_A + p_B) \\ p_C^* = \frac{1}{2}(p_A + p_C) \end{cases}$$

En calculant la distribution stationnaire pour cette nouvelle version du jeu B, on trouve que :

$$\Pi_0 (p_B - p_C) + p_C > \frac{1}{2}$$

donc le jeu C est gagnant.

Appelons X_A , X_B et X_C les variables aléatoires représentant les gains aux jeux A, B et C. N'a-t-on pas $E(X_C) = \frac{1}{2}E(X_A) + \frac{1}{2}E(X_B) < 0$, en contradiction avec le résultat précédent? Non, car $E(X_B)$ et $E(X_C)$ dépendent de l'état (c'est à dire le gain courant). Lorsqu'on joue au jeu C un grand nombre de fois, $E(X_B) > 0$ car le gain n'a pas la même distribution que lorsqu'on joue au jeu B . On a toujours $\mathbb{E}(X_C | s) = \frac{1}{2}\mathbb{E}(X_A | s) + \frac{1}{2}\mathbb{E}(X_B | s)$ pour tout état s , la linéarité de l'espérance est inexorable.

Transcription: Chaddaï Fouché.

7.8 La Méthode probabiliste

La méthode probabiliste s'appuie sur les probabilités pour démontrer l'existence de certains objets.

7.8.1 Introduction

Considérons une formule de k -SAT à m clauses. Si les variables sont instanciées aléatoirement, la probabilité pour chaque clause d'être satisfaite est $1 - \frac{1}{2^k}$. Par linéarité de l'espérance, l'espérance du nombre de clauses satisfaites est $m(1 - \frac{1}{2^k})$. Il existe donc une instantiation qui satisfait au moins $m(1 - \frac{1}{2^k})$ clauses (cf. TD12 pour d'autres exemples).

7.8.2 Lemme local de Lovasz

Motivation On a E_1, \dots, E_n évènements « à éviter, ou mauvais », et on veut montrer qu'il existe un évènement non-inclus dans ceux-là. Hélas les E_i ne sont pas indépendants.

Lemme 48. Si E et F sont indépendants, \overline{E} et \overline{F} le sont aussi.

Démonstration.

$$\begin{aligned} \mathbb{P}(\overline{E} \cap \overline{F}) &= 1 - \mathbb{P}(E \cup F) \\ &= 1 - (\mathbb{P}(E) + \mathbb{P}(F) - \mathbb{P}(E \cap F)) \\ &= 1 - (\mathbb{P}(E) + \mathbb{P}(F) - \mathbb{P}(E)\mathbb{P}(F)) \\ &= (1 - \mathbb{P}(E))(1 - \mathbb{P}(F)) \\ &= \mathbb{P}(\overline{E})\mathbb{P}(\overline{F}) \end{aligned}$$

■

Ce lemme montre que si les E_i étaient indépendants deux à deux, notre problème serait réglé, en effet les $\overline{E_i}$ seraient alors également indépendants :

$$\mathbb{P}\left(\bigcup_{i=1}^n E_i\right) = \mathbb{P}\left(\bigcap_{i=1}^n \overline{E_i}\right) = \prod_{i=1}^n (1 - \mathbb{P}(E_i))$$

Et ce produit est non-nul pourvu qu'aucun des E_i ne soit de probabilité 1.

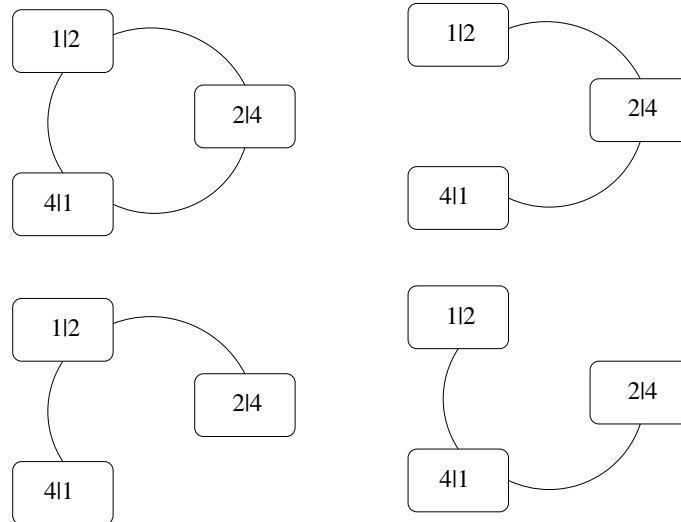
Définition (Mutuelle indépendance). On dit que E est mutuellement indépendant de l'ensemble d'évènements $\{E_1 \dots E_n\}$ si :

$$\forall I \subset \{1, \dots, n\}, \quad \mathbb{P}\left(E \mid \bigcap_{i \in I} E_i\right) = \mathbb{P}(E)$$

Définition (Graphe de dépendance). Un graphe de dépendance pour les évènements E_1, \dots, E_n est un graphe $G = (V, E)$ tel que $V = \{1, \dots, n\}$ et pour tout i , E_i est mutuellement indépendant de l'ensemble $\{E_j \mid (i, j) \notin E\}$.

Noter que cette définition peut correspondre à plusieurs graphes différents pour un même ensemble d'évènements. De plus, l'ensemble des graphes possibles n'est pas nécessairement totalement ordonné par la relation d'inclusion, autrement dit, on peut avoir plusieurs minima.

Exemple 16. Lançons un dé, et appelons X la variable aléatoire résultat. On note $a|b$ l'évènement ($X = a \cup X = b$). L'ensemble $\{1|2, 2|4, 4|1\}$ a les graphes de dépendance suivant :



Les trois derniers en particulier sont des minima : par exemple $1|2$ et $2|4$ sont indépendants, de même que $2|4$ et $4|1$, mais $2|4$ n'est pas mutuellement indépendant de $\{1|2, 4|1\}$. En effet, $\mathbb{P}((2|4) \mid (1|2 \cap 4|1 = 1)) = 0$.

Toutefois, il existe parfois un graphe de dépendance minimum, c'est le cas où E_i indépendant de E_j et de E_k implique que E_i est mutuellement indépendant de $\{E_j, E_k\}$. Il se trouve justement que c'est le cas pour k -SAT : les satisfactions de deux clauses ne sont indépendantes que si les deux clauses n'ont aucune variables en commun, et une clause est mutuellement indépendante de l'ensemble des clauses avec lesquelles elle est indépendante, car elle ne partage aucune variable avec elles.

Théorème 49 (de Lovasz, dit Local Lemma). Soit E_1, \dots, E_n un ensemble d'évènements, $p \in [0, \frac{1}{4}]$ et $d \in \mathbb{N}$ tels que :

- $\forall i \in [1, n] \quad \mathbb{P}(E_i) \leq p$
- Il existe un graphe de dépendance dont le degré maximal est inférieur ou égal à d .
- $pd \leq \frac{1}{4}$

Alors $\mathbb{P}\left(\bigcap_{i=1}^n \overline{E_i}\right) \neq 0$.

Ce théorème nous permet de répondre à notre question initiale.

7.8.3 Exemple d'application

Propriété. Si aucune variable d'une formule de k -SAT n'apparaît plus de $T = \frac{2^{k-2}}{k}$ fois, alors cette formule est satisfiable.

Démonstration. Soit E_i l'évènement « la i^e clause n'est pas satisfaite ». $\mathbb{P}(\bigcap_{i=1}^n \overline{E_i}) \neq 0$ si et seulement si la formule est satisfiable.

Posons $p = \mathbb{P}(E_i) = \frac{1}{2^k}$. La première condition du théorème est ainsi trivialement vérifiée.

Utilisons maintenant la remarque faite plus tôt sur les graphes de dépendance pour k -SAT et prenons le minimum.

$$d \leq k \left(\frac{2^{k-2}}{k} - 1 \right)$$

car il y a k variables dans une clauses et chacune de ces variables apparaît dans au plus $T - 1$ autres clauses. La clause est donc dépendante d'au plus $k(T - 1)$ clauses, et mutuellement indépendante de toutes les autres. Ce qui donne

$$pd \leq k \frac{1}{2^k} \frac{2^{k-2}}{k} \leq \frac{1}{4}$$

On peut donc utiliser le théorème de Lovasz pour conclure à la satisfiabilité de la formule. ■

7.8.4 Preuve du théorème

Démonstration. Pour démontrer le local lemma, on va faire une récurrence sur le cardinal s d'un sous-ensemble S de $\llbracket 1, n \rrbracket$. On va montrer les propriétés suivantes :

- si $|S| < n$ et $k \notin S$, alors $\mathbb{P}(E_k \mid \bigcap_{j \in S} \overline{E_j}) \leq 2p$.
- si $S \neq \emptyset$ alors $\mathbb{P}(\bigcap_{j \in S} \overline{E_j}) \neq 0$.

Soit $G = (V, E)$ un graphe de dépendance satisfaisant les conditions du théorème. Si $s = 0$, $\mathbb{P}(E_k) \leq p \leq 2p$. Sinon $s > 0$, on partage alors S par rapport à un $k \notin S$ donné :

- $S_1 = \{j \in S \mid (k, j) \in E\}$
- $S_2 = S - S_1$

Il y a alors deux cas :

1. $S_2 = S$, E_k est mutuellement indépendant des $(E_i)_{i \in S}$ et donc des $\overline{E_i}$:

$$\mathbb{P}\left(E_k \mid \bigcap_{j \in S} \overline{E_j}\right) = \mathbb{P}(E_k) \leq p \leq 2p$$

2. $|S_2| < s$, posons alors :

- $F_S = \bigcap_{j \in S} \overline{E_j}$
- $F_{S_1} = \bigcap_{j \in S_1} \overline{E_j}$
- $F_{S_2} = \bigcap_{j \in S_2} \overline{E_j}$

Notons que $F_S = F_{S_1} \cap F_{S_2}$.

$$\mathbb{P}(E_k \mid F_S) = \frac{\mathbb{P}(E_k \cap F_S)}{\mathbb{P}(F_S)}$$

$$\mathbb{P}(E_k \cap F_S) = \mathbb{P}(E_k \text{ cap } F_{S_1} \mid F_{S_2}) \mathbb{P}(F_{S_2})$$

$$\mathbb{P}(F_S) = \mathbb{P}(F_{S_1} \mid F_{S_2}) \mathbb{P}(F_{S_2})$$

Ceci établi, majorons le numérateur et minorons le dénominateur :

$$\mathbb{P}(E_k \text{ cap } F_{S_1} \mid F_{S_2}) \leq \mathbb{P}(E_k \mid F_{S_2}) = \mathbb{P}(E_k) \leq p$$

Sachant que $\mathbb{P}(F_{S_2}) \leq 1$, nous avons majoré le numérateur par p .

$$\begin{aligned}
\mathbb{P}(F_{S_1} | F_{S_2}) &= \mathbb{P}\left(\bigcap_{j \in S_1} \overline{E_j} \mid \bigcap_{j \in S_2} \overline{E_j}\right) \\
&\geq 1 - \sum_{i \in S_1} \mathbb{P}\left(E_i \mid \bigcap_{j \in S_2} \overline{E_j}\right) \\
&\geq 1 - |S_1| \times 2p \\
&\geq 1 - 2pd \\
&\geq \frac{1}{2}
\end{aligned}$$

En utilisant l'hypothèse de récurrence pour $-\mathbb{P}\left(E_i \mid \bigcap_{j \in S_2} \overline{E_j}\right) \geq -2p$.

On a donc bien $\mathbb{P}\left(E_k \mid \bigcap_{j \in S} \overline{E_j}\right) \leq 2p$.

Montrons maintenant la deuxième propriété pour $s \neq 0$: Pour $s = 1$, $\mathbb{P}(\overline{E_j}) = 1 - \mathbb{P}(E_j) \geq 1 - p > 0$. Pour $s > 1$, on peut supposer sans perdre en généralité que $S = \{1, 2, \dots, s\}$, on a alors :

$$\begin{aligned}
\mathbb{P}\left(\bigcap_{i=1}^s \overline{E_i}\right) &= \prod_{i=1}^s \mathbb{P}\left(\overline{E_i} \mid \bigcap_{j=1}^{i-1} \overline{E_j}\right) \\
&= \prod_{i=1}^s (1 - \mathbb{P}\left(E_i \mid \bigcap_{j=1}^{i-1} \overline{E_j}\right)) \\
&\geq \prod_{i=1}^s (1 - 2p) \\
&> 0
\end{aligned}$$

En utilisant la première propriété qu'on vient de démontrer.

La deuxième propriété appliquée pour $s = n$ permet de conclure. ■

7.8.5 Algorithme probabiliste pour k-SAT

Soit une formule de k-SAT, k constante paire. On note x_1, \dots, x_l les variables et c_1, \dots, c_m les clauses. On suppose que chaque variable n'apparaît au plus que $T = 2^{\alpha k}$ fois, avec α petit.

Définition (Clause Dangereuse). Une clause est dite dangereuse si la moitié ($k/2$) de ses variables sont déjà instanciées mais qu'elle est encore fausse.

On peut alors appliquer un algorithme en deux phases :

1. On instancie aléatoirement des variables tant que cela n'est pas « dangereux »
2. On instancie les variables restantes avec une stratégie.

Première phase Pour i de 1 à l , instancier x_i si elle ne figure pas dans une clause dangereuse, sinon retarder son instanciation.

Deuxième phase Soit E_i l'évènement « la i^e clause n'est pas satisfaite ». Créer le graphe de dépendance minimal des évènements E_i pour les c_i qui ne sont pas encore satisfaites. Les nœuds sont donc les clauses dangereuses à la fin de la première phase et les arêtes se trouvent entre les clauses qui partagent des variables. On fait une énumération exhaustive des composantes connexes du graphe.

Lemme 50. *Il existe une instanciation des variables retardées qui satisfait la formule.*

Démonstration. Soit c_i une clause qui est encore fausse après la première phase. $\mathbb{P}(E_i) \leq 2^{-k/2}$ car il reste au moins $k/2$ variables retardées dans chaque clause du graphe.

$d \leq kT$ et $4pd \leq 4 \times k \times 2^{-k/2} \times 2^{\alpha k} \leq 1$ pourvu que α soit suffisamment petit. ■

Chapitre 8

Processus de Poisson

Transcription: Fanny Dufossé.

8.1 Rappels sur la loi exponentielle

Densité $f(x) = \theta e^{-\theta x}$ pour $x \geq 0$

Fonction de distribution $F(x) = P(X \leq x) = \begin{cases} 1 - e^{-\theta x} & \text{si } x \geq 0 \\ 0 & \text{sinon} \end{cases}$

Moments $\mathbb{E}(X) = \frac{1}{\theta}$ $\text{Var}(X) = \frac{1}{\theta^2}$

Propriété de « sans mémoire » $\mathbb{P}(X < s + t | X > t) = \mathbb{P}(X > s)$

Théorème 51. Si X_1, \dots, X_n sont des variables indépendantes suivant des lois exponentielles de paramètres $\theta_1, \dots, \theta_n$, alors $\min(X_1, \dots, X_n)$ est une loi exponentielle de paramètre $\sum_{i=1}^n \theta_i$ et $\mathbb{P}(\min(X_1, \dots, X_n) = X_i) = \frac{\theta_i}{\sum_{i=1}^n \theta_i}$

Démonstration. On montre le cas $n = 2$, le reste suit par récurrence.

$$\begin{aligned} \mathbb{P}(\min(X_1, X_2) > x) &= \mathbb{P}((X_1 > x) \wedge (X_2 > x)) \\ &= \mathbb{P}(X_1 > x) \mathbb{P}(X_2 > x) \\ &= e^{-\theta_1 x} e^{-\theta_2 x} = e^{-(\theta_1 + \theta_2)x} \end{aligned}$$

$$\begin{aligned} \mathbb{P}(X_1 < X_2) &= \int_{X_2=0}^{\infty} \int_{X_1=0}^{X_2} f(x_1, x_2) dx_1 dx_2 \quad \text{avec } f(x_1, x_2) = \theta_1 e^{-\theta_1 x_1} \theta_2 e^{-\theta_2 x_2} \\ &= \frac{\theta_1}{\theta_1 + \theta_2} \end{aligned}$$

Exemple 17. À la gare, poste ou aéroport, quel guichet me servira ?

8.2 Processus de Poisson

Définition. Un processus de Poisson de paramètre (ou taux) λ est un processus de comptage $(N_t)_{t \geq 0}$ tel que

- $N_0 = 0$
- le processus de Poisson a des incréments indépendants et stationnaires :
 - $\forall t, s > 0$, la distribution de $N_{t+s} - N_s$ est la même que celle de N_t
 - $\forall [t_1, t_2], [t_3, t_4]$ intervalles disjoints, la distribution de $N_{t_2} - N_{t_1}$ est indépendante de la distribution de $N_{t_4} - N_{t_3}$
- $\lim_{t \rightarrow 0} \frac{\mathbb{P}(N_t=1)}{t} = \lambda$ ce qui signifie que la probabilité d'un seul évènement dans un temps court t tend vers λt
- $\lim_{t \rightarrow 0} \frac{\mathbb{P}(N_t=2)}{t} = 0$ ce qui signifie que la probabilité d'un seul évènement dans un temps court t tend vers 0

λ est donc le taux, ou vitesse d'arrivée des évènements.

Théorème 52. Soit $(N_t)_{t \geq 0}$ un processus de Poisson de paramètre λ .

$$\forall t, s > 0 \quad \forall n \geq 0 \quad P_n(t) := \mathbb{P}(N_{s+t} - N_s = n) = e^{-\lambda t} \frac{(\lambda t)^n}{n!}$$

suit une loi de Poisson

Théorème 53. Soit $(N_t)_{t \geq 0}$ un processus tel que

- $N_0 = 0$
- les incréments sont indépendants (intervalles)
- le nombre d'évènements sur tout intervalle de temps de longueur t suit une loi de Poisson de paramètre λt .

Alors N_t est un processus de Poisson de paramètre λ .

Démonstration.

$$\lim_{t \rightarrow 0} \frac{\mathbb{P}(N_t = 1)}{t} = \lim_{t \rightarrow 0} \frac{e^{-\lambda t} \lambda t}{1! \cdot t} = \lambda$$

$$\lim_{t \rightarrow 0} \frac{\mathbb{P}(N_t = 2)}{t} = \lim_{t \rightarrow 0} \frac{e^{-\lambda t} \lambda^2 t^2}{2t} = 0 \quad \blacksquare$$

Définition (temps entre deux évènements, ou *inter-arrival times*). On note X_1 la date d'arrivée du premier évènement, X_n le temps entre le $(n - 1)^e$ évènement et le n^e évènement.

Théorème 54. Les X_i sont des variables indépendantes suivant des lois exponentielles de paramètre λ .

Démonstration.

$$\mathbb{P}(X_1 > t) = \mathbb{P}(N_t = 0) = e^{-\lambda t}$$

Donc X_1 est de paramètre λ .

$$\begin{aligned} & \mathbb{P}(X_i > t_i \mid (X_1 = t_1) \wedge (X_2 = t_2) \wedge \dots \wedge (X_{i-1} = t_{i-1})) \\ &= \mathbb{P}\left(N_{\sum_{k=0}^i t_k} - N_{\sum_{k=0}^{i-1} t_k} = 0\right) \\ &= e^{-\lambda t} \end{aligned} \quad \blacksquare$$

Théorème 55 (réciproque). Soit $(N_t)_{t \geq 0}$ tel que

- $N_0 = 0$
- les temps entre deux évènements sont des variables indépendantes qui suivent des lois exponentielles de paramètre λ .

Alors, N_t est un processus de Poisson de paramètre λ .

Théorème 56 (Combinaison de Poisson). Soient $(N_t)_{t \geq 0}$ et $(M_t)_{t \geq 0}$ deux processus de Poisson de paramètres respectifs λ et μ . Alors

- $N_t + M_t$ est un processus de Poisson de paramètre $\lambda + \mu$;
- chaque évènement de $N_t + M_t$ vient de N_t avec probabilité $\frac{\lambda}{\lambda + \mu}$.

8.3 Chaînes de Markov

Définition. $(X_t)_{t \geq 0}$ est un processus de Markov si seul le présent influe sur l'avenir :

$$\forall s, t \quad \mathbb{P}(X_{s+t} = x \mid X_u, 0 \leq u \leq t) = \mathbb{P}(X_{s+t} = x \mid X_t)$$

et $\mathbb{P}(X_{s+t} = x \mid X_t)$ est indépendant de t .

Définition (opérateur). On définit une chaîne de Markov par

- une matrice de transition P ; p_{ij} est la probabilité que le prochain état soit j si l'état courant est i
- $(\theta_1, \dots, \theta_n)$; la distribution du temps passé dans l'état i suit une loi exponentielles de paramètre θ_i .

8.3.1 Distribution limite

On admet l'existence et l'unicité. On note $P_{i,j}(t)$ la probabilité d'être dans l'état i au temps t si on était dans l'état j au temps 0. Soit $\Pi_i := \lim_{t \rightarrow +\infty} P_{i,j}(t)$.

Théorème 57. *S'il y a existence des Π_i , alors*

$$\Pi_i \theta_i = \sum_k \Pi_k \theta_k p_{k,i}$$

De plus, si $\forall i, \theta_i = \theta$, alors

$$\Pi = \Pi P$$

(c'est la même équation qu'en temps discret).

Ce théorème correspond à la loi de conservation du taux des transitions vers et depuis l'état i .

8.4 Les files d'attentes M/M/1

Dans la notation « M/M/1 »,

- La première lettre correspond à la nature du processus d'arrivée (M pour Markov, G pour général). Ici on prend un processus de Poisson de paramètre λ .
- La seconde lettre correspond à la nature du processus de service. Ici on prend un processus de Poisson de paramètre μ .
- Le nombre correspond au nombre de serveurs (ou guichets).

Soit M_t le nombre de clients dans la file au temps t . C'est un processus de Markov.

$$P_k(t) := \mathbb{P}(M_t = k) \xrightarrow[t \rightarrow \infty]{} \Pi_k$$

Appliquons la loi de conservation à l'état 1 : $\lambda \Pi_0 = \mu \Pi_1$

À l'état k , on obtient

$$\lambda \Pi_{k-1} = \mu \Pi_k$$

Car

$$\Pi_k(\lambda + \mu) = \lambda \Pi_{k-1} + \mu \Pi_{k+1}$$

De plus

$$\sum_k \Pi_k = 1$$

Chapitre 9

Files d'attente

Transcription: Adrien Panhaleux. Le but de cette section est de trouver une loi pour modéliser des files d'attentes où des clients arrivent avec une certaine probabilité et leur service est traité selon une autre probabilité. Le but est bien sûr d'arriver à un régime permanent où la liste de clients en attente (ou processus) est borné.

9.1 Processus de comptage

On va tout d'abord chercher à modéliser l'arrivée de clients dans la file d'attente, sans s'occuper de traiter leur demande.

9.1.1 Binomial

On rappelle que la loi binomiale $Binom(n, p)$, notée ici $X(n)$, est le nombre de succès en n essais, avec une probabilité p de succès. La loi géométrique $Geom(p)$ notée ici Y est le nombre d'essais entre deux succès consécutifs.

On définit un intervalle de temps (*frame* en anglais) Δ pendant lequel on suppose qu'un seul client peut arriver. $t = n\Delta$ est alors le temps pendant lequel il peut arriver n clients. Soit p la probabilité qu'un client arrive pendant l'intervalle de temps Δ . Le nombre de clients arrivant sur une plage de temps t suit donc une loi binomiale, et on peut calculer son espérance :

$$\lambda := \mathbb{E}(X(n)) = \mathbb{E}(X(t/\Delta)) = \frac{p}{\Delta}$$

λ est appelé le taux d'arrivée pour une file d'attente (*rate*). On note le temps entre deux arrivées par $T := \Delta Y$, avec Y la loi géométrique de paramètre p .

La loi $X(n)$ est une chaîne de Markov. En effet, on a :

$$p_{i,j} = \begin{cases} p & \text{si } j = i + 1 \\ 1 - p & \text{si } j = i \\ 0 & \text{sinon} \end{cases}$$
$$P = \begin{pmatrix} 1-p & p & 0 & \dots & 0 \\ 0 & 1-p & p & \dots & 0 \\ \vdots & 0 & \ddots & \ddots & \vdots \\ 0 & \vdots & \ddots & 1-p & p \\ 0 & 0 & \dots & 0 & 1 \end{pmatrix}$$

On a donc $p_{i,j}^k = \mathbb{P}(j - i \text{ succès en } k \text{ essais}) = C_k^{i-j} p^{j-i} (1-p)^{k-(j-i)}$.

9.1.2 Poisson

Si $\Delta \rightarrow 0$ et $n \rightarrow +\infty$, avec $\lambda t = np$ constant, alors la loi binomiale $X(n) = Binom(n, p)$ tend vers une loi de Poisson de paramètre λ :

$$\lim_{\substack{n \rightarrow +\infty \\ p \rightarrow 0 \\ np = \lambda t}} C_n^k p^k (1-p)^{n-k} = e^{-\lambda} \frac{\lambda^k}{k!}$$

On peut alors calculer la loi du temps écoulé entre l'arrivée de deux clients (*inter-arrival time*) :

$$\begin{aligned}\mathbb{P}(T \leq t) &= \mathbb{P}(Y \leq n) \\ &= 1 - (1 - p)^n \quad (\text{car } Y \text{ est la loi géométrique}) \\ &= 1 - \left(1 - \frac{\lambda t}{n}\right)^n \\ &\rightarrow 1 - e^{-\lambda t}\end{aligned}$$

Ce qui montre que le temps inter-arrivées suit une loi exponentielle.

Remarque. Si l'on oublie l'hypothèse d'un seul événement par frame, on sait que la probabilité d'avoir plus d'un événement est en $o(\Delta)$, et donc tend vers 0. Ceci justifie que l'on peut supposer qu'il y a effectivement un seul événement.

9.2 Loi de Little

On a une file d'attente desservie par plusieurs serveurs. La stratégie pour vider la file d'attente est quelconque. On peut cependant établir un théorème, après avoir défini beaucoup de variables. Accrochez-vous :

$A(t)$: nombre d'arrivées au temps t ,

$\lambda_A = \frac{\mathbb{E}(A(t))}{t}$ taux d'arrivée,

$p_\lambda = \frac{1}{\lambda_A}$ espérance du temps inter-arrivées,

μ_S : temps moyen de service,

$\lambda_S = \frac{1}{\mu_S}$ taux moyen de service,

$r = \frac{\lambda_A}{\lambda_S} = \frac{\mu_S}{\mu_A}$ taux d'utilisation ($r \leq$ nombre de serveurs),

$X_S(t)$: nombre de tâches (ou clients) en cours de traitement au temps t ,

$X_W(t)$: nombre de tâches en attente,

$X(t) = X_S(t) + X_W(t)$: nombre de tâches,

S_k : temps de service de la tâche k ,

W_k : temps d'attente de la tâche k ,

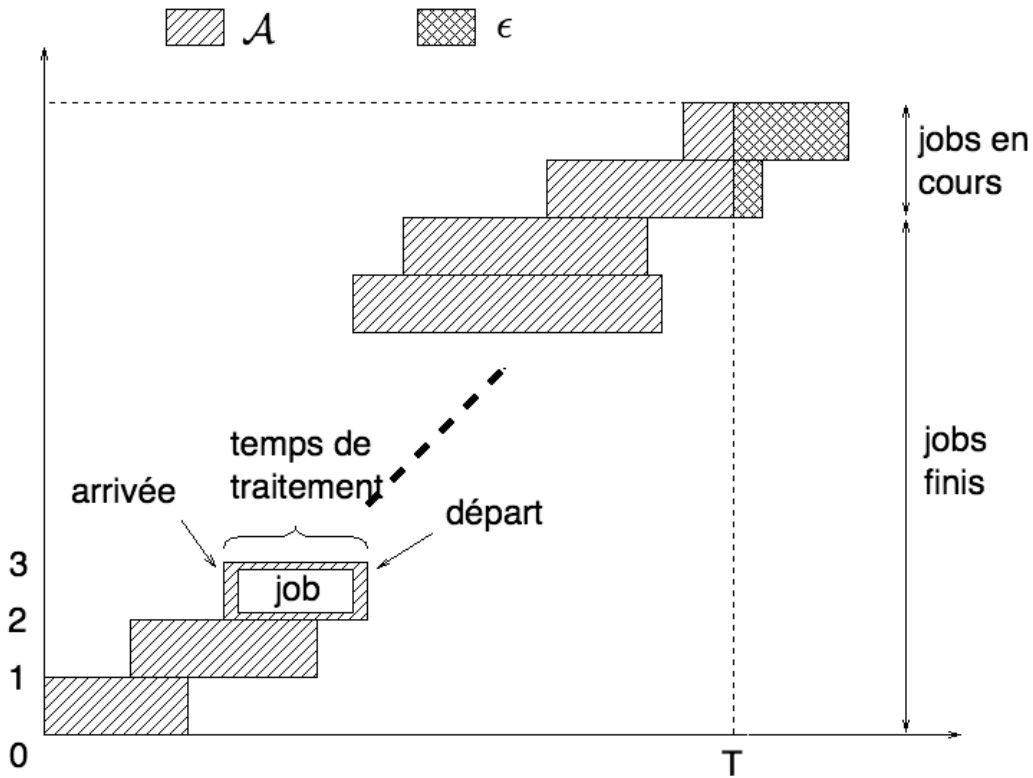
$R_k = S_k + W_k$: temps de réponse de la tâche k ,

S, W, R : S_k, W_k, R_k en régime permanent.

Théorème 58 (Little).

$$\lambda_A \mathbb{E}(R) = \mathbb{E}(X)$$

Démonstration.



$$\begin{aligned}
 \text{L'aire de } \mathcal{A} \text{ dans la figure} &= \sum_{k=1}^{A(T)} R_k - \epsilon \\
 &= \int_0^T X(t) dt \\
 \underbrace{\frac{1}{T} \sum_{k=1}^{A(T)} R_k}_{\frac{\mathbb{E}(A(T))\mathbb{E}(R)}{T} \text{ (admis)}} &= \frac{1}{T} \int_0^T X(t) dt = \mathbb{E}(X)
 \end{aligned}$$

Remarque : ϵ disparaît car T tend vers l'infini (on se place en régime permanent). En admettant que $\lambda_A = \frac{\mathbb{E}(A(T))}{T}$, on a le résultat. ■

En adaptant la démonstration du théorème de Little, on en déduit immédiatement le corollaire suivant :

Corrolaire 59.

$$\mathbb{E}(X_W) = \lambda_A \mathbb{E}(W)$$

$$\mathbb{E}(X_S) = \lambda_A \mathbb{E}(S)$$

9.3 Serveur Bernoulli

On a ici un unique serveur discret, à capacité infinie. L'arrivée des clients se fait toujours par une loi binomiale : la probabilité qu'un client arrive pendant un intervalle de temps Δ vaut $p_A = \lambda_A \Delta$. De même, pendant un intervalle de temps Δ , le serveur a une probabilité $p_S = \lambda_S \Delta$ de servir un client (s'il existe).

On suppose que le temps inter-arrivées et le temps de service sont indépendants. Cette supposition est essentielle : sans elle, on ne peut presque rien faire, et c'est pourquoi tous les travaux sur les files d'attente probabilistes font cette supposition.

Comme il y a au maximum un événement de chaque type possible par frame, on observe comme dans la partie 9.1.1 que l'état de la file d'attente forme une chaîne de Markov :

$$\begin{aligned}
 p_{i,i-1} &= \mathbb{P}(\text{pas d'arrivée et un départ}) = (1 - p_A)p_S \\
 p_{i,i} &= \mathbb{P}(\text{pas d'arrivée et pas de départ}) + \mathbb{P}(\text{une arrivée et un départ}) = p_A p_S + (1 - p_A)(1 - p_S) \\
 p_{i,i+1} &= \mathbb{P}(\text{une arrivée et pas de départ}) = p_A(1 - p_S)
 \end{aligned}$$

On obtient donc une matrice stochastique tri-diagonale :

$$P = \begin{pmatrix} p_{APS} + 1 - p_A & p_A(1 - p_S) & 0 & 0 & \dots & 0 \\ (1 - p_A)p_S & p_{APS} + (1 - p_A)(1 - p_S) & p_A(1 - p_S) & 0 & \dots & 0 \\ 0 & \ddots & \ddots & \ddots & \ddots & \vdots \end{pmatrix}$$

9.4 Serveur M/M/1

Le serveur M/M/1 est une approximation à l'ordre 1 en Δ du serveur Bernoulli :

$$\begin{aligned} p_{i,i-1} &= \lambda_S \Delta \\ p_{i,i} &= 1 - \lambda_A \Delta - \lambda_S \Delta \\ p_{i,i+1} &= \lambda_A \Delta \end{aligned}$$

9.4.1 Régime permanent

En régime permanent, on peut donc calculer le vecteur stable :

$$\pi P = \pi$$

$$\sum_{i=0}^{\infty} \pi_i = 1$$

On en déduit

$$\lambda_A \pi_0 = \lambda_S \pi_1$$

$$\lambda_A \pi_1 = \lambda_S \pi_2$$

Par récurrence

$$\pi_i = \frac{\lambda_A}{\lambda_S} \pi_{i-1} = r \pi_{i-1}$$

où r est le taux d'utilisation du système. On voit alors que l'on a un régime permanent si et seulement si $r < 1$. Si c'est le cas, on a alors

$$\sum_{i=0}^{\infty} \pi_i = \frac{\pi_0}{1 - r} = 1$$

9.4.2 Performances

Étudions les performances du serveur M/M/1, dans le cas $r < 1$.

Utilisation Le taux d'utilisation du système est $r = 1 - \pi_0$. La probabilité que le serveur soit occupé est r , tandis que la probabilité que le serveur soit inactif est $1 - r$.

Temps d'attente Un travail arrive alors qu'il y a déjà X travaux dans la file d'attente. Calculons tout d'abord l'espérance de X . Pour cela, on introduit la variable $Y = X + 1$. On a donc

$$\mathbb{P}(Y = y) = \mathbb{P}(X = y - 1) = \pi_{y-1} = r^{y-1} (1 - r) \quad (\text{Geom}(1 - r))$$

$$\mathbb{E}(X) = \mathbb{E}(Y) - 1 = \frac{1}{1 - r} - 1 = \frac{r}{1 - r}$$

De plus, on sait que $\mathbb{E}(S) = 1/\lambda_S$, car S suit une loi exponentielle de paramètre λ_S . Par indépendance sur les S_i , on en déduit que

$$\begin{aligned} \mathbb{E}(W) &= \mathbb{E}(S_1 + S_2 + \dots + S_X) \\ &= \mathbb{E}(S) \mathbb{E}(X) \\ &= \frac{r}{\lambda_S(1 - r)} \end{aligned}$$

Temps de réponse

$$\mathbb{E}(R) = \mathbb{E}(W) + \mathbb{E}(S) = \frac{r}{\lambda_S(1 - r)} + \frac{1}{\lambda_S} = \frac{1}{\lambda_S(1 - r)}$$

Longueur de la file La longueur de la file est donnée par $X_W = X - X_S$. Puisque X_S est une loi de Bernoulli de paramètre r ,

$$\mathbb{E}(X_W) = \frac{r^2}{1 - r}$$

Little Un lecteur attentif se sera rendu compte que notre étude des performances du serveur M/M/1 vérifie effectivement le théorème de Little, à savoir :

$$\lambda_A \mathbb{E}(R) = \mathbb{E}(X), \quad \lambda_A \mathbb{E}(S) = \mathbb{E}(X_S), \quad \lambda_A \mathbb{E}(W) = \mathbb{E}(X_W)$$