Structures et algorithmes aléatoires

Anne Bouillard

Notes de cours pour l'année 2015-2016

Table des matières

Cours de probabilités discrètes et applications à différents domaines de l'informatique.

Probabilités discrètes. On travaille sur des espaces au plus dénombrables par opposition aux probabilités continues (sur $\mathbb{R}$ ou $\mathbb{R}^n$) qui impliquent au préalable des notions d'intégrabilité et constitue un cours à part entière, complémentaire à celui-ci, en licence de mathématiques.

Applications en informatique

- *algorithmique :* à titre d'exemple, voici quelques cours proposés au MPRI qui utilisent des notions de probabilités
  - algorithmique avancée
  - complexité randomisée
  - analyse d'algorithmes
  - aspects algorithmiques de la combinatoire
- *Calcul quantique*
- *Théorie des réseaux de communication :* files d'attentes, protocoles de communication...
- *Cours de L3/M1 du DI ayant ce cours comme pré-requis :* Théorie de l'information et du codage, algorithmique des réseaux, apprentissage statistique...

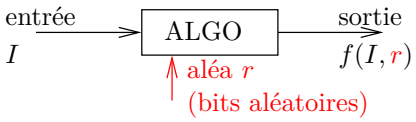
Deux exemples d'application

1- Algorithme probabiliste

Un algorithme *déterministe* est tel que pour chaque entrée, il existe une et une seule valeur de sortie, qui sera toujours renvoyée.

On veut une réponse rapide et correcte, ce que l'on ne sait pas toujours faire avec un algorithme déterministe. Même un algorithme en $\mathcal{O}(n^3)$ par exemple peut n'être pas assez rapide pour des données de grande taille. Sous réserve que $P \neq NP$, ce n'est pas même pas toujours possible de trouver un algorithme polynomial.

On peut ajouter de l'aléa sous forme de bits aléatoires. Il n'existe donc plus a priori une unique sortie pour chaque entrée.



On peut modifier l'algorithme de sorte à avoir :

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

Exemple (vérifier l'égalité de deux polynômes). Considérons deux polynômes de degré d, par exemple F et G donnés ci-dessous. Sont-ils égaux ?

$$F(x) = (x+1)(x-2)(x+3)(x-4)(x+5)(x-6) \stackrel{?}{=} x^6 - 7x^3 + 25 = G(x)$$

Algorithme 1 : Algorithme déterministe

début
mettre les deux polynômes sous forme canonique ($\sum_{i=0}^{d} c_i x^i$);
vérifier que les coefficients sont égaux.

On peut vérifier cela de plusieurs manières.

Un algorithme naïf nécessite $\mathcal{O}(d^2)$ opérations pour développer un polynôme en supposant que les opérations arithmétiques s'effectuent en temps constant (on pourrait aussi faire en $\mathcal{O}(d \log d)$ par « diviser pour régner », mais pour ce premier exemple, on ne compare que des approches naïves).

Si, par exemple, quelqu'un a écrit un programme qui développe un polynôme et veut vérifier son algorithme. Ici, l'algorithme donné passe par le même calcul et nécessite le même temps de calcul. Il présente donc un double inconvénient : le temps de calcul, et la reproduction d'une erreur similaire, en cas d'algorithme erroné.

On peut utiliser l'algorithme aléatoire suivant :

Algorithme 2 : Algorithme probabiliste

début
Choisir r dans l'ensemble $\{1, \ldots, 100d\}$ uniformément (chaque nombre a la même chance d'être choisi);
calculer $F(r)$ et $G(r)$ (en $\mathcal{O}(d)$);
si $F(r) = G(r)$ **alors**
  `Retourner` $F = G$
sinon
  `Retourner` $F \neq G$

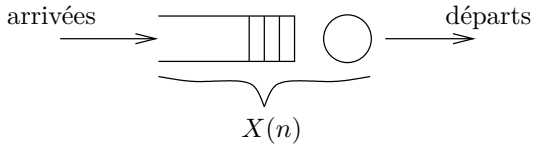
— Si $F = G$, alors l'algorithme renvoie la bonne réponse
— Si $F \neq G$
  — Si $F(r) \neq G(r)$, alors l'algorithme renvoie la bonne réponse
  — si $F(r) = G(r)$, alors l'algorithme se trompe

L'algorithme se trompe uniquement si $F \neq G$ et r est une racine de $F - G$. Or ce polynôme a au plus d racines dans $\{1, \ldots, 100d\}$. La probabilité que l'algorithme se trompe est alors au plus $1/100$.

2- Canal de communication

On considère un système régit par une horloge qui produit des *tics* à intervalles réguliers et où la transmission d'un paquet monopolise l'intervalle de temps entre deux *tics*. Soient $X(n)$ nombre de paquets dans la file juste après le n-ième *tic* et $a(n)$ nombre de paquets qui arrivent entre le n-ième *tic* et le $n+1$-ième. Alors

$$X(n+1) = \max(X(n) - 1, 0) + a(n).$$



$(X(n))_{n\in\mathbb{N}}$ forme un processus stochastique si $a(n)$ est décrit de manière probabiliste. Sous certaines conditions, $(X(n))_{n\in\mathbb{N}}$ est ce qu'on appelle une chaîne de Markov, qui fait l'objet de la seconde partie du cours.

Bibliographie

Michael Mitzenmacher et Eli Upfal. *Probability and Computing : Randomized Algorithms and Probabilistic Analysis*, Cambridge University Press, 2005.

Pierre Brémaud. *Initiation aux probabilités et aux chaînes de Markov*, Springer, 2009.

Pierre Brémaud. *Markov Chains : Gibbs Fields, Monte Carlo Simulation, and Queues*, Springer, 1999.

Olle Häggström. *Finite Markov Chains and Algorithmic Applications*, Cambridge University Press. 2002.

Chapitre 1

Probabilités discrètes

1.1 Événements et probabilités

Exemple (lancer de dés). On dit « une chance sur 6 d'obtenir un 3 » ou « une chance sur deux d'obtenir un chiffre pair ». Informellement, « obtenir une 3 » et « obtenir un chiffre pair » sont des *événements* et leurs probabilités respectives sont 1/6 et 1/2.

L'exercice suivant montre qu'une définition informelle peut donner lieu à des ambiguïtés.

> **Exercice 1** *Tirer une corde au hasard*
> Quelle est la probabilité qu'une corde d'un cercle *choisie au hasard* soit plus grande que le côté du triangle équilatéral inscrit dans le même cercle ?
>
> Trouver plusieurs manières de « choisir au hasard » qui donnent des probabilités différentes.

Il est donc nécessaire de définir rigoureusement la notion de « choisie au hasard ». Pour ce faire, on introduit la notion d'espace probabiliste.

1.1.1 Tribus et événements

On se donne
- Ω un ensemble qui décrit toutes les possibilités d'une expérience. Par exemple, pour un dé, on a $\Omega = \{1, 2, 3, 4, 5, 6\}$. Les éléments de Ω sont appelés les *épreuves* ou les *réalisations* ;
- intuitivement, un *événement* est un sous-ensemble de Ω. Pour un dé par exemple, « obtenir un 3 » correspond à l'événement $\{3\}$ et « obtenir un chiffre pair » correspond à l'événement $\{2, 4, 6\}$.

Définissons maintenant ces concepts de manière formelle.

Définition 1. *Une* tribu *sur Ω est une famille $\mathcal{F}$ de sous-ensembles de Ω telle que*

(α) *Ω et $\emptyset$ sont dans $\mathcal{F}$ ($\Omega, \emptyset \in \mathcal{F}$).*

(β) *si $A \in \mathcal{F}$, alors $A^c \triangleq \bar{A} \triangleq \Omega \setminus A \in \mathcal{F}$ (stabilité par complémentaire).*

(γ) *si $(A_n)_{n\in\mathbb{N}} \in \mathcal{F}^{\mathbb{N}}$ alors $\bigcup_{n=0}^{\infty} A_n \in \mathcal{F}$ (stabilité par union dénombrable).*

La tribu grossière est la plus petite tribu sur Ω et c'est $\{\Omega, \emptyset\}$. La tribu fine est la plus grosse tribu sur Ω et c'est $\{A \mid A \subseteq \Omega\}$.

Les événements doivent former une tribu. En général, dans le cadre de ce cours, on considérera la tribu fine, car l'on se place dans le cas où Ω est au plus dénombrable, et dans ce cas, $\mathcal{P}(\Omega)$ est engendré par les singletons.

Notons que $(\beta) + (\gamma)$ entraîne la stabilité par intersection dénombrable : $\bigcap_{n\in\mathbb{N}} A_n = (\bigcup_{n\in\mathbb{N}} A_n^c)^c$.

Exemple (événements plus complexes : petite incursion dans les espaces non dénombrables). Soit $\Omega = \{0,1\}^{\mathbb{N}}$. Les épreuves sont les éléments $\omega = (\omega_0, \omega_1, \omega_2, \ldots)$, $\omega_i \in \{0,1\}$. Soit $A = \{\omega \mid w_i \in A_i, i \leq k\}$ avec $A_i \subseteq \{0,1\}$. En d'autres termes, l'espace des épreuves peut se voir comme une suite infinie de lancers de pièce et l'événement A concerne uniquement les k premiers lancers. Les événements de type A ne sont pas stables par union dénombrable. Par exemple, l'événement $\{\omega \mid \lim_{n\to\infty} \omega_n = 0\}$ représente les épreuves ayant un nombre fini de 1 mais appartient à la tribu engendrée par ces événements :

$$A = \bigcup_{n\in\mathbb{N}} \bigcap_{m\geq n} \{\omega \mid \omega_m = 0\}.$$

1.1.2 Espace de probabilités, axiomes des probabilités

Définition 2. *Soit Ω un espace d'épreuves et $\mathcal{F}$ une tribu sur Ω. Une* probabilité *sur $(\Omega, \mathcal{F})$ est une application $\mathbf{P} : \mathcal{F} \to [0,1]$ telle que*

(α) $\mathbf{P}(\Omega) = 1$;

(β) (σ-additivité) *si $(A_n)_{n\in\mathbb{N}}$ est une suite d'événements* deux à deux disjoints *alors*

$$\mathbf{P}(\cup_{n\in\mathbb{N}} A_n) = \sum_{n\in\mathbb{N}} \mathbf{P}(A_n).$$

On appelle $(\Omega, \mathcal{F}, \mathbf{P})$ un *espace de probabilités*.

Soit A un événement. Si $\mathbf{P}(A) = 1$, alors on dit que A est presque sûr. Si $\mathbf{P}(A) = 0$, alors on dit que A est presque impossible.

Exemple (construction de probabilité). On interprète 0 comme pile et 1 comme face. On reprend l'exemple $\Omega = \{0,1\}^{\mathbb{N}}$ et on note $E_n = \{w \mid w_n = 1\}$ l'événement « le i-ème lancer est face ».

On veut trouver une probabilité telle que

— $\mathbf{P}(E_n) = p$, et donc $\mathbf{P}(E_n^c) = 1 - p$ et

— on a « indépendance des lancers » : $\mathbf{P}(\cap_{i\leq n} E_i^{(c)}) = \prod_{i\leq n} \mathbf{P}(E_i^{(c)})$, où $E_i^{(c)}$ représente soit E_i, soit E_i^c.

Alors, on doit nécessairement choisir

$$\mathbf{P}(\cap_{i\leq n} E_i^{(c)}) = p^{\sum_{i=0}^n a_i} (1-p)^{n-\sum_{i=0}^n a_i}$$

où $a_i = 1$ si E_i apparaît ou $= 0$ si c'est E_i^c qui est dans la formule.

On peut montrer (hors programme) que ceci suffit à bien définir une probabilité.

Proposition 1. *Soient $(\Omega, \mathcal{F}, \mathbf{P})$ un espace de probabilités et A, B et A_n, $n \in \mathbb{N}$ des événements (de $\mathcal{F}$).*

1. $\mathbf{P}(A^c) = 1 - \mathbf{P}(A)$.
2. $\mathbf{P}(\emptyset) = 0$.
3. $A \subseteq B \Rightarrow \mathbf{P}(A) \leq \mathbf{P}(B)$ (monotonie).
4. $\mathbf{P}(\cup_{n\in\mathbb{N}} A_n) \leq \sum_{n\in\mathbb{N}} \mathbf{P}(A_n)$ (union-bound).
5. $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cap B)$.

Démonstration.

1. $\mathbf{P}(A) + \mathbf{P}(A^c) = \mathbf{P}(\Omega) = 1$.
2. $\mathbf{P}(\emptyset) = \mathbf{P}(\Omega^c) = 1 - \mathbf{P}(\Omega) = 1 - 1 = 0$.
3. $\mathbf{P}(B) = \mathbf{P}(B \setminus A) + \mathbf{P}(A)$. Mais $\mathbf{P}(B \setminus A) \geq 0$ donc $\mathbf{P}(B) \geq \mathbf{P}(A)$.
4. Posons $B_n = A_n - \cup_{k<n} A_k \subseteq A_n$. Les B_n sont deux à deux disjoints donc $\mathbf{P}(\cup_{n\in\mathbb{N}} A_n) = \mathbf{P}(\cup_{n\in\mathbb{N}} B_n) = \sum_{n\in\mathbb{N}} \mathbf{P}(B_n) \leq \sum_{n\in\mathbb{N}} \mathbf{P}(A_n)$.
5. $\mathbf{P}(A) = \mathbf{P}(A \setminus (A \cap B)) + \mathbf{P}(A \cap B)$. De même, $\mathbf{P}(B) = \mathbf{P}(B \setminus (A \cap B)) + \mathbf{P}(A \cap B)$ et $\mathbf{P}(A \cup B) = \mathbf{P}(A \setminus (A \cap B)) + \mathbf{P}(B \setminus (A \cap B)) + \mathbf{P}(A \cap B)$. Au final, on a donc bien $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cap B)$.

□

Continuité séquentielle

Théorème 1 (Continuité séquentielle)**.** *Soit $(A_n)_{n\in\mathbb{N}}$ une suite croissante d'événements ($\forall n \in \mathbb{N}$, $A_n \subseteq A_{n+1}$). Alors*

$$\mathbf{P}(\cup_{n\in\mathbb{N}} A_n) = \lim_{n\to\infty} \mathbf{P}(A_n).$$

Démonstration. Posons $B_n = A_n \setminus A_{n-1}$ (avec la convention $A_{-1} = \emptyset$). Alors $\cup_{n\in\mathbb{N}} A_n = \cup_{n\in\mathbb{N}} B_n$ et les B_n sont deux à deux disjoints, et $A_n = \cup_{k\leq n} B_k$. Donc,

$$\mathbf{P}(\cup_{n\in\mathbb{N}} A_n) = \mathbf{P}(\cup_{n\in\mathbb{N}} B_n) = \sum_{n\in\mathbb{N}} \mathbf{P}(B_n) = \lim_{n\to\infty} \sum_{k\leq n} \mathbf{P}(B_k) = \lim_{n\to\infty} \mathbf{P}(\cup_{k\leq n} B_k) = \lim_{n\to\infty} \mathbf{P}(A_n).$$

□

Corollaire 1. *Soit $(B_n)_{n\in\mathbb{N}}$ une suite décroissante d'événements ($\forall n \in \mathbb{N}$, $B_{n+1} \subseteq B_n$). Alors*

$$\mathbf{P}(\cap_{n\in\mathbb{N}} B_n) = \lim_{n\to\infty} \mathbf{P}(B_n).$$

Démonstration.

$$\mathbf{P}(\cap_{n\in\mathbb{N}} B_n) = 1 - \mathbf{P}(\overline{\cap_{n\in\mathbb{N}} B_n}) = 1 - \mathbf{P}(\cup_{n\in\mathbb{N}} B_n^c) = 1 - \lim_{n\to\infty} \mathbf{P}(B_n^c) = \lim_{n\to\infty} (1 - \mathbf{P}(B_n^c)) = \lim_{n\to\infty} \mathbf{P}(B_n).$$

□

1.1.3 Indépendance, probabilité conditionnelle

Exemple (égalité de deux polynômes.). Reprenons l'exemple de la vérification de l'égalité de deux polynômes présenté en introduction. Soient F et G deux polynôme degré d. La probabilité que l'algorithme échoue est $\mathbf{P}(\text{algorithme échoue}) \leq 1/100$.

Que faire si l'on veut une plus grande précision ?

1. Augmenter l'espace des entiers dans lequel on choisit la racine potentielle. Ce n'est pas satisfaisant car pour de grands entiers, cela mènerait à des problèmes de précision.
2. Répéter l'algorithme plusieurs fois. Si l'on a une sortie r telle que $F(r) \neq G(r)$, alors on sait que $F \neq G$. On peut choisir avec ou sans remise des valeurs déjà tirées. La première solution ne dépend alors pas des résultats déjà obtenus alors que la seconde en dépend.

Avec remise, intuitivement, faire k essais entraîne une probabilité de se tromper d'au plus $(1/100)^k$, sans remise, cette probabilité sera plus petite.

Indépendance d'évenements

Définition 3. *Deux événements A et B sont indépendants si $\mathbf{P}(A \cap B) = \mathbf{P}(A) \cdot \mathbf{P}(B)$. $(A_n)_{n\in\mathbb{N}}$ est une famille d'événements mutuellement indépendants si pour toute famille finie $(A_{i_0}, \ldots, A_{i_n})$, $\mathbf{P}(A_{i_0} \cap \cdots \cap A_{i_n}) = \mathbf{P}(A_{i_0}) \cdots \mathbf{P}(A_{i_n})$.*

Exercice 2 *cf. TD*
$(A_n)_{n\in\mathbb{N}}$ est une famille d'événements mutuellement indépendants si et seulement si la famille $(B_n)_{n\in\mathbb{N}}$ telle que $\forall n \in \mathbb{N}$, $B_n = A_n$ ou A_n^c l'est.

Exercice 3 *cf. TD*
L'indépendance mutuelle n'est pas l'indépendance deux à deux.

Théorème 2. *Soit $(C_n)_{n\in\mathbb{N}}$ une suite d'événements mutuellement indépendants. Alors*

$$\mathbf{P}(\cap_{n\in\mathbb{N}} C_n) = \prod_{n\in\mathbb{N}} \mathbf{P}(C_n).$$

Démonstration. On utilise la continuité séquentielle. Soit $B_n = \cap_{k\leq n} C_k$. Alors $(B_n)_{n\in\mathbb{N}}$ est une suite décroissante d'événements et d'une part,

$$\mathbf{P}(B_n) = \mathbf{P}(\cap_{k=0}^{n} C_k) = \prod_{k=0}^{n} \mathbf{P}(C_k),$$

et d'autre part,

$$\lim_{n\to\infty} \mathbf{P}(B_n) = \mathbf{P}(\cap_{n\in\mathbb{N}} B_n) = \mathbf{P}(\cap_{n\in\mathbb{N}} C_n) \text{ et } \lim_{n\to\infty} \prod_{k=0}^{n} \mathbf{P}(C_k) = \prod_{k=0}^{\infty} \mathbf{P}(C_k).$$

Le résultat s'en déduit immédiatement. □

Dans l'algorithme précédemment décrit, on tire à chaque fois une valeur au hasard entre 1 et 100d. Ces tirages sont effectués indépendamment. Soit E_i l'événement « je tire une racine pour le i-ème tirage ». La probabilité que l'algorithme se trompe est alors $\mathbf{P}(\cap_{i=1}^{k} E_i) = \prod_{i=1}^{k} \mathbf{P}(E_i) \leq (1/100)^k$. L'erreur décroît donc de manière exponentielle.

Probabilité conditionnelle

Supposons maintenant que l'on ne tire pas une valeur déjà tirée. Les événements E_i (la i-ème valeur tirée est une racine) ne sont donc plus indépendants. On introduit la notion de *probabilité conditionnelle*.

Définition 4. *La probabilité d'un événement A étant donné (ou conditionné à) un événement B est définie par*

$$\mathbf{P}(A \mid B) = \frac{\mathbf{P}(A \cap B)}{\mathbf{P}(B)}.$$

Cette probabilité est bien définie uniquement si $\mathbf{P}(B) \neq 0$. Dans le cas contraire, on peut définir $\mathbf{P}(A \mid B)$ de manière arbitraire.

Remarque 1. *Si deux événements sont indépendants, alors* $\mathbf{P}(A \mid B) = \frac{\mathbf{P}(A \cap B)}{\mathbf{P}(B)} = \frac{\mathbf{P}(A)\mathbf{P}(B)}{\mathbf{P}(B)} = \mathbf{P}(A)$.

Les probabilités sans remise deviennent (notons que l'on omet le signe $\cap$ dans les formules, cette omission deviendra quasiment automatique dans la suite) :

$$\begin{aligned}
\mathbf{P}(E_1 \cap \cdots \cap E_k) &= \mathbf{P}(E_k \mid E_1 \cap \cdots \cap E_{k-1})\mathbf{P}(E_1 \cap \cdots \cap E_{k-1}) \\
&= \mathbf{P}(E_k \mid E_1 \cap \cdots \cap E_{k-1})\mathbf{P}(E_{k-1} \mid E_1 \cap \cdots \cap E_{k-2})\mathbf{P}(E_1 \cap \cdots \cap E_{k-2}) \\
&= \mathbf{P}(E_1)\mathbf{P}(E_2 \mid E_1)\mathbf{P}(E_3 \mid E_1, E_2) \cdots \mathbf{P}(E_k \mid E_1, \ldots, E_{k-1})
\end{aligned}$$

Or $\mathbf{P}(E_i \mid E_1 \cdots E_{i-1}) \leq \frac{d-(j-1)}{100d-(j-1)}$ (avec égalité si les d racines sont distinctes), donc pour $k \leq d$,

$$\mathbf{P}(E_1 \cap \cdots \cap E_k) \leq \prod_{j=1}^{k} \frac{d-(j-1)}{100d-(j-1)} \leq \left(\frac{1}{100}\right)^k.$$

On obtient une erreur plus petite qu'avec remise, mais en pratique il est parfois plus judicieux algorithmiquement d'utiliser la méthode avec remise.

La formule utilisée est le théorème de Bayes :

Théorème 3 (Loi de Bayes séquentielle)**.** *Soient $E_1, \ldots, E_n$ des événements. La probabilité de l'intersection de ces événements peut se calculer comme*

$$\mathbf{P}(E_1 \cap \cdots \cap E_k) = \mathbf{P}(E_1)\mathbf{P}(E_2 \mid E_1)\mathbf{P}(E_3 \mid E_1 \cap E_2) \cdots \mathbf{P}(E_k \mid E_1 \cap \cdots \cap E_{k-1}).$$

La preuve se fait par induction en utilisant le principe évoqué plus haut.

On peut aussi déduire de la définition des probabilités conditionnelles les lois (de Bayes) suivantes.

Théorème 4 (Loi des probabilités totales)**.** *Soit* $(E_1, \ldots, E_n)$ *une partition de* Ω *(les événements* $E_1, \ldots, E_n$ *sont* 2 à 2 disjoints *tels que* $\cup_{i=1}^n E_i = \Omega$*). Alors pour tout événement* A*,*

$$\mathbf{P}(A) = \sum_{i=1}^{n} \mathbf{P}(A \mid E_i)\mathbf{P}(E_i).$$

Démonstration.

$$\begin{aligned} \mathbf{P}(A) &= \mathbf{P}(\cup_{i=1}^n (A \cap E_i)) \\ &= \sum_{i=1}^{n} \mathbf{P}(A \cap E_i) = \sum_{i=1}^{n} \mathbf{P}(A \mid E_i)\mathbf{P}(E_i). \end{aligned}$$

□

Proposition 2 (Loi de rétrodiction)**.** *Pour tous événements* A *et* B *avec* $\mathbf{P}(B) > 0$*,*

$$\mathbf{P}(A \mid B) = \frac{\mathbf{P}(A)}{\mathbf{P}(B)}\mathbf{P}(B \mid A).$$

Démonstration.

$$\mathbf{P}(A \mid B) = \frac{\mathbf{P}(A \cap B)}{\mathbf{P}(B)} = \frac{\mathbf{P}(A)}{\mathbf{P}(B)}\frac{\mathbf{P}(A \cap B)}{\mathbf{P}(A)} = \frac{\mathbf{P}(A)}{\mathbf{P}(B)}\mathbf{P}(B \mid A).$$

□

Exercice 4
On se donne trois pièces dont une et une seule est biaisée de telle sorte que $\mathbf{P}(\texttt{Face}) = 2/3$. On lance les pièces et on obtient `Face Face Pile`.

Quelle est la probabilité que la première pièce soit la pièce biaisée ?

1.2 Variables aléatoires

1.2.1 Variables aléatoires et distribution

Définition 5. *Soit $(\Omega, \mathcal{F}, \mathbf{P})$ un espace de probabilité. Soit E un ensemble au plus dénombrable. Une fonction $X : \Omega \to E$ telle que pour tout $x \in E$, $\{\omega \mid X(\omega) = x\} \in \mathcal{F}$ est une* variable aléatoire (v.a.) discrète sur E.

On note dans la suite $\{X = x\}$ ou plus simplement $X = x$ l'événement $\{\omega \mid X(\omega) = x\}$ et $\{X \in A\}$ ou plus simplement $X \in A$ l'événement $\{\omega \mid X(\omega) \in A\}$.

Si $E \subseteq \mathbb{R}$, on parle de variable aléatoire réelle (v.a.r.).

Définition 6. *Soit E un ensemble au plus dénombrable. Soit $\{p(x),\ x \in E\}$ une suite de réels tels que $\forall x \in E$, $0 \leq p(x) \leq 1$ et $\sum_{x \in E} p(x) = 1$. Cette suite est appelée* distribution de probabilité *sur E. Si une variable aléatoire (v.a.) X vérifie $\mathbf{P}(X \in A) = \sum_{x \in A} p(x)$ pour tout $A \subseteq E$, on dit que $(p(x),\ x \in E)$ est la distribution (ou loi) de probabilité de X.*

On peut adapter les définitions vues pour les événements aux variables aléatoires :

Définition 7.
- *Deux variables aléatoires X et Y sont* indépendantes *si et seulement si pour tous x et y, $\mathbf{P}(X = x, Y = y) = \mathbf{P}(X = x)\mathbf{P}(Y = y)$;*
- *Les variables aléatoires $X_1, \ldots, X_k$ sont* mutuellement indépendantes *si et seulement si $\forall I \subseteq \{1, \ldots, k\}$, $\forall (x_i)_{i \in I}$,*

$$\mathbf{P}(\bigcap_{i \in I} \{X_i = x_i\}) = \prod_{i \in I} \mathbf{P}(\{X_i = x_i\}) \ ;$$

- *les variables aléatoires $X_1, \ldots, X_k$ sont* indépendantes et identiquement distribuées (i.i.d.) *si et seulement si elles sont mutuellement indépendantes et de même distribution.*

Proposition 3. *Soit $(\Omega, \mathcal{F}, \mathbf{P})$ un espace de probabilité. Soient $X_1, \ldots, X_n$ des variables aléatoires à valeurs dans respectivement $E_1, \ldots, E_n$ des ensembles au plus dénombrables et $f : E_1 \times \cdots \times E_n \to F$ une fonction. Alors $f(X_1, \ldots, X_n)$ est une variable aléatoire.*

Démonstration. X_i est une variable aléatoire, donc $\{\omega \mid X_i(\omega) = y_i\} \in \mathcal{F}$ pour tout $y_i \in E_i$. Or, pour tout $x \in F$,

$$\{\omega \mid f(X_1(\omega), \ldots, X_n(\omega)) = x\} = \bigcup_{\{y_i, \ldots, y_n \mid f(y_1, \ldots, y_n) = x\}} \bigcap_{i=1}^{n} \{\omega \mid X_i(\omega) = y_i\}$$

Donc $\{f(X_1, \ldots, X_n) = x\} \in \mathcal{F}$ par stabilité par intersection et union dénombrable. □

En particulier, si X et Y sont des variables aléatoires et f une fonction, $X + Y$, XY, $f(X)$ en sont aussi.

Proposition 4. *Soit $(\Omega, \mathcal{F}, \mathbf{P})$ un espace de probabilité. Soient $X_1, \ldots, X_n$ des variables aléatoires mutuellement indépendantes à valeurs dans respectivement $E_1, \ldots, E_n$ des ensembles au plus dénombrables et $f_i : E_i \to F_i$ une fonction. Alors $f_1(X_1), \ldots, f_n(X_n)$ sont des variables aléatoires mutuellement indépendantes.*

Démonstration. Nous montrons le résultat pour $n = 2$ seulement, le cas général n'est qu'une généralisation d'écriture.

Pour tout $(j_1, j_2) \in F_1 \times F_2$,

$$\begin{aligned} p(f_1(X_1) = j_1, f_2(X_2) = j_2) &= \sum_{i_1 \in E_1, f_1(i_1) = j_1} \sum_{i_2 \in E_2, f_2(i_2) = j_2} \mathbf{P}(X_1 = i_1, X_2 = i_2) \\ &= \sum_{i_1 \in E_1, f_1(i_1) = j_1} \mathbf{P}(X_1 = i_1) \sum_{i_2 \in E_2, f_2(i_2) = j_2} \mathbf{P}(X_2 = i_2) \\ &= \mathbf{P}(f_1(X_1) = j_1)\mathbf{P}(f_2(X_2) = j_2). \end{aligned}$$

□

Exemples de distributions

1. **Loi constante égale à a :** $\mathbf{P}(X = a) = 1$.
2. **Loi de Bernoulli :**

$$X \sim \mathcal{B}\text{er}(p) \text{ si } \mathbf{P}(X = 1) = p \text{ et } \mathbf{P}(X = 0) = 1 - p.$$

   Interprétation : pile ou face biaisé. Un exemple fondamental de v.a. de ce type de distribution est la fonction caractéristique d'un événement : pour $A \in \mathcal{F}$, $\mathbf{1}_A$ définie telle que $\mathbf{1}_A(\omega) = 1$ si $\omega \in A$ et $\mathbf{1}_A(\omega) = 0$ sinon est une v.a. de loi Bernoulli de paramètre $\mathbf{P}(A)$.

3. **Loi binomiale de paramètres n et p :**

$$X \sim \mathcal{B}\text{in}(n, p) \text{ si } \mathbf{P}(X = j) = \binom{n}{j} p^j (1 - p)^{n-j}.$$

   Interprétation : n lancers d'une pièce qui donne face avec probabilité p. $p(j)$ est la probabilité d'obtenir j fois face exactement parmi n lancers indépendants. $X = \sum_{i=1}^{n} X_i$ où X_i est l'issue du i-ème lancer. Les X_i sont mutuellement indépendants et de loi $\mathcal{B}\text{er}(p)$.

4. **Loi géométrique de paramètre p :**

$$x \sim \mathcal{G}\text{eo}(p) \text{ si } \mathbf{P}(X = n) = (1 - p)^{n-1} p \text{ pour } n \geq 1.$$

   Interprétation : On lance une pièce jusqu'à obtenir face. Quelle est la probabilité de lancer la pièce n fois exactement ?

Proposition 5. *La loi géométrique est* sans mémoire *: soit X une v.a. de loi géométrique. Alors pour tous $k \geq 0$ et $n \geq 1$, $\mathbf{P}(X = n + k \mid X > k) = \mathbf{P}(X = n)$.*

Démonstration.

$$\begin{aligned}\mathbf{P}(X = n+k \mid X > k) &= \frac{\mathbf{P}(X = n+k,\ X > k)}{\mathbf{P}(X > k)} = \frac{\mathbf{P}(X = n+k)}{\mathbf{P}(X > k)} = \frac{(1-p)^{n+k-1}p}{\sum_{i=k}^{\infty}(1-p)^i p} \\ &= \frac{(1-p)^{n+k-1}p}{(1-p)^k} = (1-p)^{n-1}p = \mathbf{P}(X = n).\end{aligned}$$

□

5. **Loi uniforme sur [0,1]** (exception : v.a. non discrète). $\forall x \leq 1$, $\mathbf{P}(X \leq x) = x$.

1.2.2 Espérance

Définition 8. *Soit X une v.a. à valeurs dans E de distribution $(p(x),\ x \in E)$. Soit $f : E \to \overline{\mathbb{R}}$, où $\overline{\mathbb{R}} = \mathbb{R} \cup \{+\infty, -\infty\}$ une fonction. L'espérance de $f(X)$, notée $\mathbf{E}[f(X)]$ est :*

(a) si $\forall x \in E,\ f(x) \geq 0$, alors $\mathbf{E}[f(X)] = \sum_{x \in E} f(x)p(x)$

(b) sinon, soit $f^+(x) = \max(f(x), 0)$ et $f^-(x) = \max(-f(x), 0)$

1. *si $\mathbf{E}[|f(X)|] < \infty$ (f est intégrable), alors $\mathbf{E}[f(X)] = \mathbf{E}[f(X)^+] - \mathbf{E}[f^-(X)]$.*
2. *si $\mathbf{E}[|f(X)|] = \infty$ et $\mathbf{E}[f(X)^+]$ ou $\mathbf{E}[f^-(X)] < \infty$ (f est sommable), alors $\mathbf{E}[f(X)] = \mathbf{E}[f(X)^+] - \mathbf{E}[f^-(X)] = \pm\infty$.*
3. *sinon $\mathbf{E}[f(X)]$ n'existe pas / n'est pas définie.*

Si X est à valeurs dans $E \subseteq \overline{\mathbb{R}}$, alors $\mathbf{E}[X]$ est sa moyenne.

Propriétés importantes

Théorème 5 (Linéarité)**.** *Soient X et Y des variables aléatoires réelles d'espérance finie et a et b des réels. Alors*

$$\mathbf{E}[aX + bY] = a\mathbf{E}[X] + b\mathbf{E}[Y].$$

Démonstration. On montre ceci en deux étapes pour plus de clarté. Tout d'abord,

$$\mathbf{E}[aX] = \sum_{x \in E} axp(x) = a \sum_{x \in E} xp(x) = a\mathbf{E}[X].$$

Puis,

$$\begin{aligned}\mathbf{E}[X+Y] &= \sum_{x,y \in E} (x+y)\mathbf{P}(X = x, Y = y) \\ &= \sum_{x \in E}\sum_{y \in E} x\mathbf{P}(X = x, Y = y) + \sum_{y \in E}\sum_{x \in E} y\mathbf{P}(X = x, Y = y) \\ &= \sum_{x \in E} x \sum_{y \in E} \mathbf{P}(X = x, Y = y) + \sum_{y \in E} y \sum_{x \in E} \mathbf{P}(X = x, Y = y) \\ &= \sum_{x \in E} x\mathbf{P}(X = x) + \sum_{y \in E} y\mathbf{P}(Y = y) \\ &= \mathbf{E}[X] + \mathbf{E}[Y].\end{aligned}$$

□

Remarque 2. *Ce théorème est aussi valide, par le même raisonnement si X et Y sont positives ainsi que a et b dans le cas où l'une de ces v.a. au moins est d'espérance infinie.*

Soient X une v.a. et $\mathcal{P}$ une propriété. Si $\mathbf{P}(X \text{ vérifie } \mathcal{P}) = 1$, on dit que $\mathcal{P}$ est vérifiée presque sûrement (p.s.).

Proposition 6 (Monotonie)**.** *Soient X une v.a. et $f, g : E \to \overline{\mathbb{R}}$ deux fonctions telles que les espérances de $f(X)$ et $g(X)$ existent. Alors, si $f(X) \leq g(X)$ p.s., alors $\mathbf{E}[f(X)] \leq \mathbf{E}[g(X)]$.*

Démonstration. $\mathbf{E}[f(X)] = \sum_{x \in E} f(x)p(x) \leq \sum_{x \in E} g(x)p(x) = \mathbf{E}[g(X)]$. □

Inégalité de Jensen pour les fonctions convexes

Théorème 6 (Inégalité de Jensen)**.** *Si ϕ est une fonction convexe sur un intervalle $I \subseteq \mathbb{R}$ et X une v.a. sur I, alors si X et $\phi(X)$ sont intégrables, alors*

$$\mathbf{E}[\phi(X)] \geq \phi(\mathbf{E}[X]).$$

Démonstration. Si ϕ est convexe sur I, alors pour tout $x_0 \in \overset{\circ}{I}$, il existe α tel que pour tout $x \in I$, $\phi(x) \geq \phi(x_0) + \alpha(x - x_0)$. Deux cas se présentent : si X est constante presque sûrement, on a $\phi(\mathbf{E}[X]) = \mathbf{E}[\phi(X)]$. Sinon, $\mathbf{E}[X] \in \overset{\circ}{I}$ et on peut poser $x_0 = \mathbf{E}[X]$. Alors, par monotonie, on a

$$\phi(X) \geq \phi(\mathbf{E}[X]) + \alpha(X - \mathbf{E}[X])$$

et en prenant l'espérance, par linéarité,

$$\mathbf{E}[\phi(X)] \geq \mathbf{E}[\phi(\mathbf{E}[X])] + \alpha\mathbf{E}[X] - \alpha\mathbf{E}[\mathbf{E}[X]] = \phi(\mathbf{E}[X]).$$

□

Espérance conditionnelle par rapport à un événement Soit X une v.a. et A un événement. L'espérance conditionnelle de X par rapport à A est

$$\mathbf{E}[X \mid A] = \sum_{x \in E} xp(X = x \mid A).$$

Lemme 1. *Soient X et Y des variables aléatoires réelles telle que $\mathbf{E}[X]$ existe. Alors*

$$\mathbf{E}[X] = \sum_{y \in E} \mathbf{P}(Y = y)\mathbf{E}[X \mid Y = y].$$

Démonstration.

$$\begin{aligned}\sum_{y \in E} \mathbf{P}(Y = y)\mathbf{E}[X \mid Y = y] &= \sum_{y \in E} \mathbf{P}(Y = y) \sum_{x \in E} x\mathbf{P}(X = x \mid Y = y) \\ &= \sum_{x \in E} x \sum_{y \in E} \mathbf{P}(X = x, Y = y) = \sum_{x \in E} x\mathbf{P}(X = x) = \mathbf{E}[X].\end{aligned}$$

□

Exemples

1. Si X est constante égale à a, alors $\mathbf{E}[X] = a$.
2. Si $X \sim \mathcal{B}\text{er}(p)$, alors
$$\mathbf{E}[X] = 0 \times (1-p) + 1 \times p = p.$$
Si A est un événement, on utilise souvent la réécriture $\mathbf{E}[\mathbf{1}_A] = \mathbf{P}(A)$.
3. Si $X \sim \mathcal{B}\text{in}(n,p)$, alors
$$\mathbf{E}[X] = \mathbf{E}[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} \mathbf{E}[X_i] = np,$$
avec $X_i \sim \mathcal{B}\text{er}(p)$.
4. Si $X \sim \mathcal{G}\text{eo}(p)$, alors
$$\mathbf{E}[X] = \sum_{i=1}^{\infty} ip(1-p)^{i-1} = 1/p.$$
On peut faire le calcul direct ou utiliser la propriété suivante :

Proposition 7. *Soit X une v.a. à valeurs dans $\mathbb{N}$. Alors $\mathbf{E}[X] = \sum_{i=1}^{\infty} \mathbf{P}(X \geq i)$.*

Démonstration.
$$\sum_{i=1}^{\infty} \mathbf{P}(X \geq i) = \sum_{i=1}^{\infty} \sum_{j=i}^{\infty} \mathbf{P}(X = j) = \sum_{j=1}^{\infty} \sum_{i=1}^{j} \mathbf{P}(X = j) = \sum_{j=1}^{\infty} j\mathbf{P}(X = j) = \mathbf{E}[X].$$
□

Si $X \sim \mathcal{G}\text{eo}(p)$, alors $\mathbf{P}(X \geq i) = \sum_{j=i}^{\infty} \mathbf{P}(X = j) = (1-p)^{i-1}$ et donc $\mathbf{E}[X] = \sum_{i=1}^{\infty} (1-p)^{i-1} = 1/p$.

Application au collectionneur de coupons

Un individu collectionne les images dans les boîtes de céréales. Chaque boîte en contient une et une seule, et il y a n images différentes au total.

L'image contenue dans une boîte est choisie uniformément parmi les n images, et indépendamment des autres boîtes.

Combien de boîtes faut-il acheter en moyenne pour obtenir toutes les images ?

Soit X le nombre de boîtes ouvertes pour obtenir les n images. On cherche $\mathbf{E}[X]$. Soit X_i le nombre de boîtes ouvertes depuis que le collectionneur a obtenu $i-1$ images jusqu'à l'obtention de la i-ème image. On a $X = \sum_{i=1}^{n} X_i$.

Distribution des X_i : pour chaque boîte, le collectionneur a un nouveau coupon avec probabilité $(n-(i-1))/n$. Donc $X_i \sim \mathcal{G}\text{eo}(1 - \frac{i-1}{n})$ et son espérance est $\mathbf{E}[X_i] = \frac{n}{n-i+1}$. Ainsi
$$\mathbf{E}[X] = \mathbf{E}[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} \mathbf{E}[X_i] = \sum_{i=1}^{n} \frac{n}{n-i+1} = n \sum_{i=1}^{n} \frac{1}{i} = nH(n)$$
avec $H(n) = \sum_{i=1}^{n} \frac{1}{i} = \ln n + \mathcal{O}(1)$. En effet,
$$H(n) - 1 = \sum_{i=2}^{n} \frac{1}{i} \leq \ln n = \int_1^n \frac{1}{x} dx \leq \sum_{i=1}^{n} \frac{1}{i} = H(n).$$

1.2.3 Variance

Définition 9. *Soit X une variable aléatoire réelle.*
- *Le* k-ème moment *d'une v.a.r. X est* $\mathbf{E}[X^k]$.
- *La* variance *d'une v.a.r. X est* $\mathbf{V}\text{ar}(X) = \mathbf{E}[(X - \mathbf{E}[X])^2] = \mathbf{E}[X^2] - \mathbf{E}[X]^2$
- *la* covariance *de deux v.a.r. X et Y est* $\mathbf{C}\text{ov}(X, Y) = \mathbf{E}[(X - \mathbf{E}[X])(Y - \mathbf{E}[Y])]$.
- *l'*écart-type *de X est* $\sigma(X) = \sqrt{\mathbf{V}\text{ar}(X)}$.

Justification des deux expressions de la variance : $\mathbf{E}[(X - \mathbf{E}[X])^2] = \mathbf{E}[X^2 - 2\mathbf{E}[X]X + \mathbf{E}[X]^2] = \mathbf{E}[X^2] - 2\mathbf{E}[X]\mathbf{E}[X] + \mathbf{E}[X]^2 = \mathbf{E}[X^2] - \mathbf{E}[X]^2$

La variance et l'écart-type donnent une mesure de la différence entre X et $\mathbf{E}[X]$. Si la variance est grande, alors avec une forte probabilité, X est très différent de $\mathbf{E}[X]$. Au contraire, si la variance est petite, alors la probabilité que X soit proche de $\mathbf{E}[X]$ est élevée.

Exemple (variance). Soit X une v.a. telle que $\mathbf{P}(X = 0) = 1 - 1/k$ et $\mathbf{P}(X = k) = 1/k$. Alors $\mathbf{E}[X] = 1$ et $\mathbf{V}\text{ar}(X) = k - 1$: $\mathbf{E}[X^2] = k^2/k = k$.

Proposition 8. *Soient X et Y deux v.a.r. Alors*
- $\mathbf{V}\text{ar}(X + Y) = \mathbf{V}\text{ar}(X) + \mathbf{V}\text{ar}(Y) + 2\mathbf{C}\text{ov}(X, Y)$
- *Si X et Y sont indépendantes, alors* $\mathbf{E}[XY] = \mathbf{E}[X]\mathbf{E}[Y]$.
- *Si X et Y sont indépendantes, alors* $\mathbf{C}\text{ov}(X, Y) = 0$ *et* $\mathbf{V}\text{ar}(X+Y) = \mathbf{V}\text{ar}(X)+\mathbf{V}\text{ar}(Y)$.

Démonstration.

$$\begin{aligned}\mathbf{E}[(X + Y - \mathbf{E}[X + Y])^2] &= \mathbf{E}[(X + Y - \mathbf{E}[X] - \mathbf{E}[Y])^2] \\ &= \mathbf{E}[(X - \mathbf{E}[X])^2 + (Y - \mathbf{E}[Y])^2 + 2(X - \mathbf{E}[X])(Y - \mathbf{E}[Y])] \\ &= \mathbf{E}[(X - \mathbf{E}[X])^2] + \mathbf{E}[(Y - \mathbf{E}[Y])^2] + 2\mathbf{E}[(X - \mathbf{E}[X])(Y - \mathbf{E}[Y])] \\ &= \mathbf{V}\text{ar}(X) + \mathbf{V}\text{ar}(Y) + 2\mathbf{C}\text{ov}(X, Y).\end{aligned}$$

Si X et Y sont indépendantes, alors

$$\begin{aligned}\mathbf{E}[XY] &= \sum_i \sum_j (i \cdot j)\mathbf{P}(X = i, Y = j) = \sum_i \sum_j (i \cdot j)\mathbf{P}(X = i)\mathbf{P}(Y = j) \\ &= \sum_i i\mathbf{P}(X = i) \sum_j j\mathbf{P}(Y = j) = \mathbf{E}[X]\mathbf{E}[Y].\end{aligned}$$

Mais alors, $\mathbf{C}\text{ov}(X, Y) = \mathbf{E}[(X - \mathbf{E}[X])(Y - \mathbf{E}[Y])] = \mathbf{E}[(X - \mathbf{E}[X])]\mathbf{E}[(Y - \mathbf{E}[Y])] = 0$ et $\mathbf{V}\text{ar}(X + Y) = \mathbf{V}\text{ar}(X) + \mathbf{V}\text{ar}(Y)$. □

Exemples

- Si X est constante égale à a, alors $\mathbf{V}\text{ar}(X) = 0$;
- Si $X \sim \mathcal{B}\text{er}(p)$, alors $X^2 = X$ et $\mathbf{E}[X^2] = p$. Alors $\mathbf{V}\text{ar}(X) = p - p^2 = p(1 - p)$.
- Si $X \sim \mathcal{B}\text{in}(n, p)$, alors $X = \sum_{i=1}^{n} X_i$ avec $X_i \sim \mathcal{B}\text{er}(p)$ i.i.d et donc $\mathbf{V}\text{ar}(X) = np(1-p)$.
- Si $X \sim \mathcal{G}\text{eo}(p)$, alors $\mathbf{V}\text{ar}(X) = \frac{1-p}{p^2}$. En effet, soit $X_0 \sim \mathcal{B}\text{er}(p)$ et $Y \sim \mathcal{G}\text{eo}(p)$ deux v.a. indépendantes. La v.a. X peut s'interpréter comme $X = 1$ si $X_0 = 1$ et $X = Y+1$ sinon.

$\mathbf{E}[X^2] = \mathbf{E}[X^2 \mid X_0 = 0]\mathbf{P}(X_0 = 0) + \mathbf{E}[X^2 \mid X_0 = 1]\mathbf{P}(X_0 = 1) = (1-p)\mathbf{E}[(Y+1)^2] + p$. Or X et Y ont même distribution, donc

$$\mathbf{E}[X^2] = (1-p)(\mathbf{E}[Y^2] + 2\mathbf{E}[Y] + 1) + p = (1-p)\mathbf{E}[X^2] + 2\frac{1-p}{p} + 1.$$

Finalement on trouve $\mathbf{E}[X^2] = \frac{2-p}{p^2}$, d'où $\mathbf{V}\mathrm{ar}[X] = \frac{2-p}{p^2} - \frac{1}{p^2} = \frac{1-p}{p^2}$.

1.2.4 Déviations : premières inégalités

Le but des inégalités que nous allons voir dans ce paragraphe est de borner la probabilité que l'écart à la moyenne soit plus grand qu'une certaine valeur. Par exemple, cela pourra être pour un algorithme probabiliste la probabilité que le temps d'exécution soit trop long, pour le collecteur de coupons que le nombre de boîtes à acheter soit trop grand...). On borne plus précisément la quantité $\mathbf{P}(X \geq a)$.

Théorème 7 (Inégalité de Markov)**.** *Soit X une v.a. réelle positive ou nulle. Alors pour tout $a > 0$,*

$$\mathbf{P}(X \geq a) \leq \frac{\mathbf{E}[X]}{a}.$$

Démonstration. Soit $a > 0$. On pose $I = 1$ si $X \geq a$ et $I = 0$ sinon. Alors $I \leq \frac{X}{a}$, puisque $a \geq 0$. Alors, $\mathbf{E}[I] = \mathbf{P}(I = 1) = \mathbf{P}(X \geq a)$ et $\mathbf{E}[I] \leq \mathbf{E}[X/a] = \frac{\mathbf{E}[X]}{a}$ (monotonie). □

Exemple (lancer de pièces). On lance une pièce non faussée n fois. Borner la probabilité d'obtenir au moins $3n/4$ fois face. On pose $X = \sum_{i=1}^{n} X_i$ où $X_i = 1$ si le i-ème lancer est face et $X_i = 0$ s'il est pile. Alors, $\mathbf{E}[X_i] = \mathbf{P}(X_i = 1) = 1/2$. Donc $\mathbf{E}[X] = n/2$. En appliquant l'inégalité de Markov, on obtient $\mathbf{P}(X \geq 3n/4) \leq \frac{n/2}{3n/4} = \frac{2}{3}$.

C'est la borne la moins précise, mais qui est à la base de toutes les autres. On a la même borne pour toutes le v.a. ayant la même espérance et donc cette borne ne dépend pas ici de n. Intuitivement, $\mathbf{P}(X \geq 3n/4)$ est décroissante en n.

Théorème 8 (Inégalité de Tchebychev)**.** *Soient X une v.a.r et $a > 0$. Alors*

$$\mathbf{P}(|X - \mathbf{E}[X]| \geq a) \leq \frac{\mathbf{V}\mathrm{ar}(X)}{a^2}.$$

Démonstration. On a $\mathbf{P}(|X - \mathbf{E}[X]| \geq a) = \mathbf{P}((X - \mathbf{E}[X])^2 \geq a^2)$. Comme $(X - \mathbf{E}[X])^2 \geq 0$, on peut lui appliquer l'inégalité de Markov :

$$\mathbf{P}((X - \mathbf{E}[X])^2 \geq a^2) \leq \frac{\mathbf{E}[(X - \mathbf{E}[X])^2]}{a^2} = \frac{\mathbf{V}\mathrm{ar}(X)}{a^2}.$$

□

Exemple (lancer de pièces). Reprenons le même exemple du lancer de pièces. On obtient cette fois

$$\mathbf{P}(X \geq 3n/4) \leq \mathbf{P}(|X - \mathbf{E}[X]| \geq n/4) = \frac{\mathbf{V}\mathrm{ar}(X)}{(n/4)^2} = \frac{4}{n}.$$

Collectionneur de coupons - suite On a vu que $\mathbf{E}[X] = nH(n)$.

Quelle est la probabilité que le temps pour collectionner les n coupons soit le double au moins de cette espérance ? Borner $\mathbf{P}(X \geq 2nH(n))$ à l'aide des inégalités de Markov et de Tchebychev.

— *Inégalité de Markov :* $\mathbf{P}(X \geq 2nH(n)) \leq \frac{nH(n)}{2nH(n)} = \frac{1}{2}$

— *Inégalité de Tchebychev :* $\mathbf{P}(X \geq 2nH(n)) \leq \mathbf{P}(|X - nH(n)| \geq nH(n)) \leq \frac{\mathbf{V}\mathrm{ar}(X)}{(nH(n))^2}$. Il faut maintenant calculer la variance de X. On pose $X = \sum_{i=1}^{n} X_i$, où X_i est défini comme précédemment. Les X_i sont indépendants, donc

$$\begin{aligned} \mathbf{V}\mathrm{ar}(X) &= \sum_{i=1}^{n} \mathbf{V}\mathrm{ar}(X_i) \\ &= \sum_{i=1}^{n} \frac{n(i-1)}{(n-i+1)^2} \\ &\leq \sum_{i=1}^{n} \frac{n^2}{(n-i+1)^2} \leq n^2 \sum_{i=1}^{n} \frac{1}{i^2} \leq n^2 \frac{\pi^2}{6}. \end{aligned}$$

On a donc $\mathbf{P}(X \geq 2nH(n)) \leq \frac{\pi^2}{6H(n)^2} = \mathcal{O}(\frac{1}{\ln^2 n})$

— *Union-bound* Il existe une autre méthode, meilleure et plus simple. Soit E_i l'événement « le coupon i n'a pas été obtenu après ouverture des $\lceil n \ln + cn \rceil$ premières boîtes ». On a $\mathbf{P}(E_i) = (1 - \frac{1}{n})^{\lceil n(\ln n + c) \rceil} \leq e^{-(\ln n + c)} = \frac{1}{e^c n}$.
Ainsi, $\mathbf{P}(X \geq \lceil n \ln + cn \rceil) = \mathbf{P}(\cup_{i=1}^{n} E_i) \leq \sum_{i=1}^{n} \mathbf{P}(E_i) \leq \frac{1}{e^c}$. Avec $c = \ln n$, on obtient $\mathbf{P}(X \geq \lceil 2n \ln n \rceil) \leq \frac{1}{n}$.

1.2.5 Classification des algorithmes

Algorithmes de Las Vegas [Exemple : algorithme de tri rapide]

— Résout un problème exactement

— avec une complexité moyenne finie (que l'on cherche à minimiser)

Algorithmes de Monte Carlo [Exemple : recherche de la médiane]

— Résout un problème de manière approchée (avec une erreur contrôlée)

— avec une complexité qui est une fonction déterministe de la donnée.

Si la probabilité d'erreur d'un algorithme Monte Carlo est λ, cela signifie que sur chaque entrée, la probabilité d'erreur est au plus λ, et pas qu'une proportion λ de données fournit une réponse erronée.

Passage d'un algorithme de Monte Carlo à un algorithme Las Vegas Considérons un algorithme Monte Carlo de complexité C_{MC} qui renvoie soit la réponse correcte, soit la réponse `erreur` (avec probabilité au plus λ). On peut transformer cet algorithme en un algorithme Las Vegas en le répétant tant que la réponse renvoyée est `erreur`. La complexité de l'algorithme Las Vegas est alors

$$\mathbf{E}(C_{LV}) \leq \frac{C_{MC}}{\lambda}.$$

Erreur unilatérale Les algorithmes Monte Carlo qui renvoient `vrai` ou `faux` sont à erreur unilatérale s'ils se trompent seulement sur une réponse, comme c'était le cas pour la vérification de multiplication matricielle ou de polynôme. Pour minimiser l'erreur, il suffit de répéter l'algorithme plusieurs fois.

Erreur bilatérale Si un algorithme Monte Carlo qui renvoie les réponses `vrai` ou `faux` peut se tromper sur les deux réponses, il est à erreur bilatérale. Si la probabilité d'erreur est inférieure strictement à $1/2$, alors on peut diminuer cette probabilité en exécutant plusieurs fois l'algorithme et en renvoyant la réponse majoritaire.

Terminaison

— Un algorithme termine avec probabilité λ si pour toute instance l'algorithme termine avec probabilité au moins λ.
— Un algorithme termine presque sûrement si pour toute instance il termine avec probabilité 1.
— Un algorithme termine (sûrement) s'il termine sur toutes les entrées, quel que soit le tirage aléatoire.

1.3 La méthode probabiliste

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

1.3.1 Argument de comptage

Idée : On dispose d'une collection au plus dénombrable d'objets a_i, $i \in I$. On veut prouver que l'un d'eux au moins satisfait une propriété $\mathcal{P}$. Pour ce faire, on peut choisir un objet a_i aléatoirement en introduisant une variable aléatoire X sur $\{a_i\}_{i\in I}$ et si l'on peut prouver que $\mathbf{P}(X \text{ satisfait } \mathcal{P}) > 0$, alors il existe bien a_i qui satisfait $\mathcal{P}$.

Application : nombre de Ramsey : coloriage des arêtes d'un graphe complet K_n en deux couleurs, rouge et bleu de telle manière qu'il n'y ait pas de grande clique monochrome. $R(k)$ est la taille minimale du graphe (n) telle qu'il est impossible de trouver un coloriage des arêtes tel qu'il n'y a pas clique de taille k monochrome.

Théorème 9. *Si $\binom{n}{k}2^{1-\binom{k}{2}} < 1$, alors $R(k) > n$, c'est-à-dire qu'il est possible de colorier les arêtes de K_n de telle sorte qu'il n'y a pas de clique de taille k monochrome.*

Démonstration. Il y a $2^{\binom{n}{2}}$ coloriages possibles des arêtes de K_n avec deux couleurs. On choisit un coloriage uniformément parmi tous ces coloriages possibles. Cela correspond à colorier chaque arête aléatoirement en rouge ou en bleu (chaque couleur avec probabilité 1/2) et indépendamment de la couleur des autres arêtes.

Soit $i = 1, \ldots, \binom{n}{k}$ une énumération des cliques de taille k. Soit A_i l'événement « i est une clique monochrome ». Alors

$$\mathbf{P}(A_i) = 2^{-\binom{k}{2}+1} \quad \text{(deux choix parmi } 2^{\binom{k}{2}}\text{)}.$$

Donc

$$\mathbf{P}(\bigcup_{i=1}^{\binom{n}{k}} A_i) \leq \sum_{i=1}^{\binom{n}{k}} \mathbf{P}(A_i) = \binom{n}{k} 2^{-\binom{k}{2}+1} < 1.$$

et $\mathbf{P}(\bigcap_{i=1}^{\binom{n}{k}} A_i^c) = 1 - \mathbf{P}(\bigcup_{i=1}^{\binom{n}{k}} A_i) > 0$ et il existe bien un coloriage avec la propriété voulue. □

Construction effective : (k est une constante) on utilise un algorithme de type Monte-Carlo.

- Colorier chaque arête uniformément et indépendamment.
- vérifier qu'il n'y a pas de clique de taille k dans un graphe : $\mathcal{O}(n^k)$
- probabilité d'échec : $p = \mathbf{P}(\bigcup_{i=1}^{\binom{n}{k}} A_i)$.
- transformation en un algorithme Las-Vegas : répéter tant qu'on trouve une clique monotone. $\mathbf{E}(\text{temps d'exécution}) = \mathcal{O}\left(\frac{n^k}{1-\binom{n}{k}2^{-\binom{k}{2}+1}}\right)$.

1.3.2 Méthode du premier moment (argument de l'espérance)

Idée : Si l'espérance d'une v.a. X est μ, alors avec probabilité strictement positive, X prend des valeurs inférieures et supérieures à μ.

Théorème 10. *Soit X une v.a.r. Alors $\mathbf{P}(X \geq \mathbf{E}[X]) > 0$ et $\mathbf{P}(X \leq \mathbf{E}[X]) > 0$.*

Démonstration. $\mathbf{E}[X] = \sum_x x\mathbf{P}(X = x)$. Si $\mathbf{P}(X \geq \mathbf{E}[X]) = 0$, alors

$$\mathbf{E}[X] = \sum_{x<\mathbf{E}[X]} x\mathbf{P}(X = x) < \sum_{x<\mathbf{E}[X]} \mathbf{E}[X]\mathbf{P}(X = x) = \mathbf{E}[X],$$

ce qui est absurde.

Le même raisonnement est valide si $\mathbf{P}(X \leq \mathbf{E}[X]) = 0$. □

Application 1 : MAXSAT

Problème : F formule en forme normale conjonctive (CNF) (par exemple, $F = (x_1 \vee x_2 \vee x_3) \wedge (x_2 \vee \neg x_3 \vee x_4) \wedge (x_1 \vee \neg x_2 \vee \neg x_4) \ldots$).
Question : quel est le nombre maximal de clauses satisfiables ?

Problème de décision associé (NP-complet) :

Données : F, k.

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

Théorème 11. *Soient F une formule à m clauses, k_i le nombre de littéraux de la i-ème clause et $k = \min_{i=1}^m k_i$. Il existe une affectation des variables qui satisfait au moins $\sum_{i=1}^m (1-2^{-k_i}) \geq m(1-2^{-k})$ clauses.*

Démonstration. On note $x_1, \ldots, x_n$ les variables de la formule. On affecte à chaque variable `vrai` ou `faux` de manière indépendante et uniforme (si $X_i = \mathbf{1}_{x_i \texttt{ vrai}}$ alors $X_i \sim \mathcal{B}\text{er}(1/2)$).

Soit E_j l'événement « la j-ème clause est satisfaite » et $Y_j = \mathbf{1}_{E_j}$. On a $\mathbf{E}[Y_j] = 1 - 2^{-k_j}$. Soit $Y = \sum_{j=1}^m Y_j$. On a

$$\mathbf{E}[Y] = \sum_j 1 - 2^{-k_j} \geq \sum_j (1 - 2^{-k}) = m(1 - 2^{-k}).$$

Comme $\mathbf{P}(Y \geq \mathbf{E}[Y]) > 0$, il existe une affectation qui satisfait ce nombre de clauses au moins. □

Algorithme effectif de construction : utiliser les espérances conditionnelles par rapport à une affectation déjà en partie construite.

$$\begin{aligned}
\mathbf{E}[Y] &= \mathbf{E}[Y \mid X_1 = 1]\mathbf{P}(X_1 = 1) + \mathbf{E}[Y \mid X_1 = 0]\mathbf{P}(X_1 = 0) \\
&= \frac{1}{2}\left(\mathbf{E}[Y \mid X_1 = 1] + \mathbf{E}[Y \mid X_1 = 0]\right) \\
&\leq \max(\mathbf{E}[Y \mid X_1 = 1], \mathbf{E}[Y \mid X_1 = 0]).
\end{aligned}$$

Soit $\mathbf{E}[Y \mid X_1 = 1]$ ou $\mathbf{E}[Y \mid X_1 = 0]$ est supérieur à $\mathbf{E}[Y]$. On choisit l'affectation qui maximise l'espérance : x_i est `vrai` si $\mathbf{E}[Y \mid X_1 = 1] \geq \mathbf{E}[Y \mid X_1 = 0]$, et on recommence pour l'affectation de x_2 et des variables suivantes : si on a fixé les valeurs de $X_1, \ldots, X_k$ à $x_1, \ldots, x_k$ respectivement, on fixe celle de x_{k+1}. On a

$$\begin{aligned}\mathbf{E}[Y \mid X_1 = x_1, \ldots, X_k = x_k] &= \mathbf{E}[Y \mid X_1 = x_1, \ldots, X_k = x_k, X_{k+1} = 1]\mathbf{P}(X_{k+1} = 1) \\ &\quad + \mathbf{E}[Y \mid X_1 = x_1, \ldots, X_k = x_k, X_{k+1} = 0]\mathbf{P}(X_{k+1} = 0) \\ &\leq \max(\mathbf{E}[Y \mid X_1 = x_1, \ldots, X_k = x_k, X_{k+1} = 1], \\ &\quad \mathbf{E}[Y \mid X_1 = x_1, \ldots, X_k = x_k, X_{k+1} = 0])\end{aligned}$$

et on fixe X_{k+1} à 0 ou 1 selon l'espérance.

Application 2 : ensembles indépendants

Soit $G = (V, E)$ un graphe. On note $|V| = n$ et $|E| = m$. Un sous-ensemble de sommets $I \subseteq V$ est indépendant si $\forall u, v \in I$, $(u, v) \notin E$ (I est une clique dans $G' = (V, \overline{E})$).

Théorème 12. *Soit $G = (V, E)$ un graphe connexe de n sommets et m arêtes. Si $\frac{2m}{n} \geq 1$, alors G possède un ensemble indépendant de taille au moins $n^2/4m$.*

Démonstration. Si le graphe contient plus de sommets que d'arêtes, alors on sait qu'il existe un indépendant de taille $n - m$. En effet, on peut enlever un sommet extrémité de chaque arête, et l'on obtient un graphe sans arête d'au moins $n - m$ sommets. Une première phase d'un algorithme de construction consiste donc à enlever des sommets (et les arêtes qui y sont adjacentes) afin de construire un tel graphe.

Soit $p \in [0, 1]$.

Algorithme 3 : Trouver un indépendant

début

Effacer chaque sommet de G avec probabilité $1 - p$ indépendamment ;
Pour chaque arête restante, la retirer, ainsi qu'un des sommets adjacents.

On peut calculer l'espérance du nombre de sommets et d'arêtes obtenus à la fin de la seconde phase.

1. Soit X le nombre de sommets restant après la première étape : $\mathbf{E}[X] = np$.
2. Soit Y le nombre d'arêtes à la fin de la première étape. Une arête survit si aucun des sommets qu'elle relie n'est supprimé, ce qui arrive avec probabilité p^2. Ainsi, $\mathbf{E}[Y] = mp^2$.

À la deuxième étape, on retire un sommet par arête au plus. Le nombre de sommets restants est alors au moins $X - Y$ et $\mathbf{E}[X - Y] = np - mp^2$. Cette quantité est maximisée pour $p = \frac{n}{2m} \leq 1$ par hypothèse. On obtient alors $\mathbf{E}[X - Y] = \frac{n^2}{2m} - \frac{n^2}{4m} = \frac{n^2}{4m}$. Notons que $d = \frac{1}{p} = \frac{2m}{n}$ est le degré moyen du graphe. □

Une autre inégalité, qui découle directement de l'inégalité de Markov, va être très utile dans le prochain paragraphe :

Théorème 13. *Soit X une variable aléatoire sur $\mathbb{N}$. Alors*

$$\mathbf{P}(X \neq 0) \leq \mathbf{E}[X].$$

1.3.3 Méthode du second moment

Théorème 14. *Si X est une v.a. sur $\mathbb{N}$, alors $\mathbf{P}(X = 0) \leq \frac{\mathbf{V}\mathrm{ar}(X)}{\mathbf{E}[X]^2}$.*

Démonstration. $\mathbf{P}(X = 0) \leq \mathbf{P}(|X - \mathbf{E}[X]| \geq \mathbf{E}[X]) \leq \frac{\mathbf{V}\mathrm{ar}(X)}{\mathbf{E}[X]^2}$ en appliquant l'inégalité de Tchebychev. □

Application aux graphes aléatoires (Erdös-Rényi)

Soient $n \in \mathbb{N}$ et $p \in [0, 1]$. L'espace $\mathcal{G}(n, p)$ est l'espace des graphes non orientés avec n sommets et où chaque arête a une probabilité p d'exister, indépendamment des autres. Plus précisément, $\mathcal{G}(n, p) = (\Omega_n, \mathcal{P}(\Omega_n), \mathbf{P})$, où

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

Il y a au plus $N = \binom{n}{2}$ arêtes dans un graphe à n sommets et il y a 2^N graphes dans $\mathcal{G}(n, p)$. Dans la suite, $G_{n,p}$ est un élément aléatoire de $\mathcal{G}(n, p)$.

Exemple (Graphes aléatoires). Dans $\mathcal{G}(n, p)$,
— le graphe complet a probabilité p^N ;
— le graphe vide a probabilité $(1 - p)^N$;
— la probabilité que $G_{n,p}$ ait m arêtes est $\binom{N}{m} p^m (1 - p)^{N-m}$.

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

1. p est fixé. (cf TD : les propriétés du premier ordre sont asymptotiquement vérifiées avec probabilité 0 ou 1)
2. $p = p(n)$ varie avec n.

Dans le dernier cas, on veut alors trouver une *fonction seuil*. Pour la propriété A, une telle fonction seuil est une fonction $g(n)$ telle que

(i) si $\lim_{n\to\infty} p(n)/g(n) = 0$ (ou $p \ll g$), alors $\lim_{n\to\infty} \mathbf{P}(G_{n,p(n)}$ satisfait $A) = 0$.
(ii) si $\lim_{n\to\infty} g(n)/p(n) = 0$ (or $p \gg g$), alors $\lim_{n\to\infty} \mathbf{P}(G_{n,p(n)}$ satisfait $A) = 1$.

Une fonction seuil peut s'interpréter de la manière suivante : à chaque paire de sommets $\{u, v\}$, associer un nombre aléatoire $p_{u,v}$ choisi uniformément dans $[0, 1]$ et indépendamment des autres nombres. Pour $p \in [0, 1]$, le graphe est constitué des arêtes telles que $p_{u,v} \leq p$. Maintenant, lorsque p croît de 0 jusqu'à 1, le graphe $G_{n,p}$ croît aussi. Si $g(n) \gg p$, alors $\mathbf{P}(G_{n,p}$ satisfait $A) = 0$; et si $g(n) \ll p$, alors $\mathbf{P}(G_{n,p}$ satisfait $A) = 1$. La table 1.1 donne quelques fonctions seuil.

Théorème 15. *Si $A =$ « contenir une clique de taille 4 », alors la fonction seuil est $g(n) = n^{-2/3}$. Plus précisément,*
— *si $p(n) \ll n^{-2/3}$, alors $\lim_{n\to\infty} \mathbf{P}(G_{n,p}$ satisfait $A) = 0$;*
— *si $p(n) \gg n^{-2/3}$, alors $\lim_{n\to\infty} \mathbf{P}(G_{n,p}$ satisfait $A) = 1$.*

propriété	fonction seuil $g(n)$
contenir un chemin de longueur k	$n^{-\frac{k+1}{k}}$
ne pas être planaire	$\frac{1}{n}$
contenir un chemin hamiltonien	$\frac{\ln n}{n}$
être connecté	$\frac{\ln n}{n}$
contenir une clique de taille k	$n^{-\frac{2}{k-1}}$

TABLE 1.1 – Exemples de fonctions seuil.

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

Soit $C_1, \ldots, C_{\binom{n}{4}}$ une énumération des ensembles de 4 sommets et définissons les variables aléatoires $X_i \in \{0,1\}$, $i \in \{1, \ldots, \binom{n}{4}\}$

$$X_i = 1 \Leftrightarrow C_i \text{ est une clique de taille 4.}$$

On pose $X = \sum_i X_i$.

— $\mathbf{E}[X] = \sum \mathbf{E}[X_i] = \binom{n}{4} p(n)^6 = (\frac{1}{24} n^4 + o(n^4)) p(n)^6$;

— $\mathbf{E}[X^2] = \sum \mathbf{E}[X_i] + \sum_{i \neq j} \mathbf{E}[X_i X_j]$. On doit étudier séparément plusieurs cas, selon le nombre de sommets communs à C_i et C_j. Ces cas sont décrits dans la table 1.2.

$\lvert C_i \cap C_j \rvert$	$\mathbf{E}[X_i X_j]$	nombre
≤ 1	$p(n)^{12}$	$\binom{n}{4}(\binom{n-4}{4} + 4\binom{n-4}{3})$
2	$p(n)^{11}$	$\binom{n}{4} 6 \binom{n-4}{2}$
3	$p(n)^9$	$\binom{n}{4} 4(n-4)$

TABLE 1.2 – Cliques de taille 4 : disjonction de cas pour le calcul de $\mathbf{V}\text{ar}(X)$.

Ainsi,

$$\mathbf{E}[X^2] = (\frac{1}{24} n^4 + o(n^4)) p(n)^6 + (\frac{1}{24^2} n^8 + o(n^8)) p(n)^{12} + (\frac{6}{48} n^6 + o(n^6)) p(n)^{11} + (\frac{4}{24} n^5 + o(n^5)) p(n)^9$$

et

$$\mathbf{V}\text{ar}[X] = (\frac{1}{24} n^4 + o(n^4)) p(n)^6 + (o(n^8)) p(n)^{12} + (\frac{6}{48} n^6 + o(n^6)) p(n)^{11} + (\frac{4}{24} n^5 + o(n^5)) p(n)^9.$$

Maintenant,

— si $p(n) = o(n^{-2/3})$, alors par l'inégalité de Markov,

$$\mathbf{P}(X \neq 0) \leq \mathbf{E}[X] = (\frac{1}{24} n^4 + o(n^4)) p(n)^6 = o(1).$$

— si $n^{-2/3} = o(p(n))$, $n^4 p(n)^6 \overset{n \to \infty}{\longrightarrow} \infty$ alors par la méthode du second moment,

$$\mathbf{P}(X = 0) \leq \frac{\mathbf{V}\text{ar}(X)}{\mathbf{E}[X]^2} = O(n^{-4} p(n)^{-6}) + o(1) + O(n^{-2} p(n)^{-1}) + O(n^{-3} p(n)^{-3}) = o(1).$$

□

Application aux flots de données massives

On veut analyser des données avec les contraintes suivantes.
— La mémoire dont on dispose est limitées (sous-linéaire) ;
— On ne peut accéder aux données que séquentiellement et une seule fois ;
— On doit analyser les données très rapidement

Numéro manquant Un joueur dit à un deuxième une suite des nombres de 1 à n sauf un (et dans le désordre).

Comment le deuxième joueur peut-il trouver le nombre manquant sans mémoriser les n nombres ? Si n n'est pas connu ?

Si n est connu, alors il suffit à chaque fois qu'un numéro est donné de le retrancher de la valeur initiale $n(n+1)/2$. Si n n'est pas connu, on fait alors la somme des nombres, et on compte aussi le nombre de boules

Schéma de Flajolet-Martin

But : compter les éléments distincts dans un multi-ensemble donné par un flot, c'est-à-dire qu'on ne voit les éléments de l'ensemble qu'une fois, et que l'espace dont on dispose n'est pas suffisamment grande pour garder tous les nombres en mémoire.

On considère donc S un multi-ensemble de N entiers dans l'ensemble $[0, D]$ où D est polynomial en N. Ainsi, le codage d'un entier se fait en $\mathcal{O}(\ln N)$.

On note F le nombre d'éléments distincts dans S et on cherche à calculer $\tilde{F}$ une approximation de F.

Le but exact est de calculer $\tilde{F}$ en une seule passe sur les entiers en utilisant $\mathcal{O}(\ln N)$ espace mémoire.

L'idée principale est d'utiliser une fonction de hachage parfaite h : pour tout $k \in S$, $h(k) \sim \mathcal{U}\text{nif}[0, 2^w - 1]$, où $w = \lceil \log N \rceil$. En particulier, pour $k_1 \neq k_2$, $h(k_1)$ et $h(k_2)$ sont indépendants.

On adopte le schéma suivant :
— Soit z_k le nombre de 0 en tête de la représentation binaire de $h(k)$.
— $Z = \max_{k \in S} z_k$
— $\tilde{F} = 2^Z$.

Par exemple, si $w = 5$ et $h(k) = 6 = (00110)_2$, $z_k = 2$.

Théorème 16.

$$\forall c \geq 3,\ \mathbf{P}(\frac{1}{c} \leq \frac{\tilde{F}}{F} \leq c) \geq 1 - \frac{3}{c}.$$

Démonstration. Il y a 2^w valeurs de clé possible et 2^{w-r} clés qui commencent pas r zéros, donc

$$\forall r \in [0, w],\ P(z_k \geq r) = \frac{2^{w-r}}{2^w} = 2^{-r}.$$

Pour $r \in [0, w]$ et $k \in S$, on définit la v.a. $X_k(r) = \mathbf{1}_{z_k \geq r}$ et $X(r) = \sum_{k \text{ distincts de } S} X_k(r)$. $X_k(r)$ est une v.a. Bernoulli de paramètre 2^{-r}. Donc $\mathbf{E}(X_k(r)) = 2^{-r}$ et $\mathbf{V}\text{ar}(X_k(r)) = 2^{-r}(1 - 2^{-r})$.

Les $X_k(r)$ sont indépendants, et il y a F valeurs différentes de k, donc

$$\mathbf{E}(X(r)) = F2^{-r} \qquad \text{et} \qquad \mathbf{V}\text{ar}(X(r)) = F2^{-r}(1-2^{-r}) \leq F2^{-r}.$$

Soient r_1 le plus petit entier r tel que $2^r > cF$ et r_2 le plus petit entier r tel que $2^r \geq F/c$. On dit que le schéma est correct si $\frac{1}{c} \leq \frac{\tilde{F}}{F} \leq c$. Or, on a l'équivalence suivante :

$$\begin{aligned} \frac{1}{c} \leq \frac{\tilde{F}}{F} \leq c \quad &\Leftrightarrow \quad \frac{1}{c}F \leq 2^Z \leq cF \\ &\Leftrightarrow \quad r_2 \leq Z < r_1. \end{aligned}$$

De plus, si $X(r_1) = 0$, alors il n'y a aucun $k \in S$ tel que $X_k(r_1) = 1$ donc $Z < r_1$ et si $X(r_2) \neq 0$, alors il existe $k \in S$ tel que $X_k(r_2) = 1$ donc $Z \geq r_2$. Donc le schéma est correct si $X(r_1) = 0$ et $X(r_2) \neq 0$.

Il ne reste maintenant plus qu'à borner ces quantités :

— D'après l'inégalité de Markov,

$$\mathbf{P}(X(r_1) \geq 1) \leq \mathbf{E}[X(r_1)] = F2^{-r_1} < \frac{1}{c}.$$

— D'après l'inégalité de Tchebychev,

$$\mathbf{P}(X(r_2) = 0) \leq \frac{\mathbf{V}\text{ar}(X(r_2))}{\mathbf{E}[X(r_2)]^2} \leq \frac{F2^{-r_2}}{F^2 2^{-2r_2}} = \frac{2^{r_2}}{F} \leq \frac{2}{c}.$$

On déduit donc le théorème :

$$\mathbf{P}(X(r_1) = 0 \text{ et } X(r_2) \neq 0) \geq 1 - \mathbf{P}(X(r_1) \neq 0) - \mathbf{P}(X(r_2) = 0) \geq 1 - \frac{3}{c}.$$

□

Amélioration : la technique de la médiane. On voudrait maintenant avoir une probabilité de succès de $1-\epsilon$. Pour ce faire, on utilise plusieurs fonctions de hachage mutuellement indépendantes, et pour chacune d'elles, on exécute l'algorithme. Pour s fonctions de hachage, on obtient $\tilde{F}_1, \ldots, \tilde{F}_s$ et on choisit $\tilde{F}$ comme la médiane de ces valeurs.

Théorème 17. *$\forall c > 6$, $\exists s = \mathcal{O}(\ln \frac{1}{\epsilon})$ tel que $\mathbf{P}(\frac{F}{c} \leq \tilde{F} \leq cF) \geq 1 - \epsilon$.*

Démonstration. Soit $i \in [1, s]$ et $X_i = 0$ si $\tilde{F}_i \in [\frac{F}{c}, cF]$ et $= 1$ sinon. On a $\rho = \mathbf{P}(X_i = 1) \leq \frac{3}{c} \leq 1/2$.

Soit $X = \sum_{i=1}^{s} X_i$. On a donc $\mathbf{E}[X] = s\rho$.

Si $X < \frac{s}{2}$, alors $\frac{F}{c} \leq \tilde{F} \leq cF$ et le schéma est correct. Appliquons les bornes de Chernoff :

$$\mathbf{P}(X \geq \frac{s}{2}) = \mathbf{P}(X \geq s\rho[1 + (\frac{1}{2\rho} - 1)]) \leq \exp(-\frac{1}{3}(\frac{1}{2\rho} - 1)^2 \rho s).$$

Si $s \geq \frac{3}{\rho}(\frac{1-2\rho}{2\rho})^2 \ln \frac{1}{\epsilon}$, alors $\mathbf{P}(X \geq \frac{s}{2}) \leq \epsilon$. □

1.4 Le lemme local de Lovász

Les méthodes étudiées jusqu'à présent utilisent pour la plupart le fait que les événements mis en jeu sont indépendants ou alors utilisent l'union bound.

1.4.1 Lemme symétrique de Lovász

Si $E_1, \ldots, E_n$ sont des *mauvais* événements et qu'on veut montrer qu'il existe des situations où aucun de ces mauvais événements n'a lieu, il suffit de montrer que

$$\mathbf{P}(\overline{\cup E_i}) > 0.$$

Deux solutions sont possibles :

1. $\mathbf{P}(\overline{\cup E_i}) = 1 - \mathbf{P}(\cup E_i) = 1 - \sum_{i=1}^{n} \mathbf{P}(E_i)$ (Union bound)
2. $\mathbf{P}(\overline{\cup E_i}) = \mathbf{P}(\cap \overline{E_i}) = \prod_{i=1}^{n} \mathbf{P}(\overline{E_i})$ si les E_i sont mutuellement indépendants.

Comment faire si la première méthode ne suffit pas ou si les événements ne sont pas mutuellement indépendants ? Le lemme local de Lovász permet de donner une réponse à cette question lorsque les événements ne sont pas *trop* dépendants.

On dit que qu'un événement E est *mutuellement indépendant* de $E_1, \ldots, E_n$ si pour tout $I \subseteq \{1, \ldots, n\}$, $\mathbf{P}(E \mid \cap_{i \in I} E_i) = \mathbf{P}(E)$.

Définition 10 (Graphe de dépendance)**.** *Un graphe de dépendance de $E_1, \ldots, E_n$ est un graphe $G = (V, E)$ tel que $V = \{1, \ldots, n\}$ et E_i est mutuellement indépendant des $\{E_j \mid (i, j) \notin E\}$.*

Par exemple, prenons l'exemple d'une formule en CNF :

$$F = (x_1 \vee x_2 \vee \neg x_3) \wedge (x_1 \vee \neg x_2 \vee x_3) \wedge (x_2 \vee x_3 \vee x_4) \wedge (x_4 \vee x_5 \vee \neg x_6).$$

Les mauvais événements sont E_j : « la j-ème clause n'est pas satisfaite ». Supposons que les valeurs des variables sont affectées les unes indépendamment des autres. Alors, par exemple, E_4 est mutuellement indépendant de E_1 et E_2 car la 4ème clause ne partage pas de littéraux avec les deux premières.

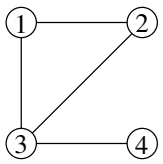


FIGURE 1.1 – Graphe de dépendance correspondant à F.

Théorème 18 (Lemme local de Lovász symétrique)**.** *Soient $E_1, \ldots, E_n$ des événements tels que*

1. *$\forall i \in \{1, \ldots n\}$, $\mathbf{P}(E_i) \leq p$.*
2. *Le degré des sommets du graphe de dépendance de $E_1, \ldots, E_n$ est borné par d.*
3. *$4pd \leq 1$.*

Alors

$$\mathbf{P}(\cap_{i=1}^{n} \bar{E_i}) > 0.$$

En utilisant l'Union bound, il faudrait que $pn < 1$. Le lemme de Lovász donne de meilleurs résultats dès que $d \leq n/4$.

Exemple ($k - SAT$). Données : Une formule F en CNF avec k variables distinctes exactement par clause.
Question : F est-elle satisfiable ?

Théorème 19. *Soit F une formule en CNF avec k littéraux exactement par clause. Si aucune variable n'apparaît plus de $2^k/4k$ fois ou si aucune clause ne partage de variable en commun avec 2^{k-2} autres clauses, alors F est satisfiable*

Démonstration. Soit E_j l'événement « la clause j n'est pas satisfaite ». On affecte les variables indépendamment selon une loi de Bernoulli de paramètre 1/2. Alors $\mathbf{P}(E_j) \leq 2^{-k}$.

E_j est mutuellement indépendant des événements concernant des clauses qui n'ont pas de variable en commun avec la j-ème.

Chaque variable apparaît au plus $2^k/4k$ fois, donc $d \leq k \times 2^k/4k = 2^{k-2}$. (On a le même degré avec l'autre énoncé). Alors, $4dp \leq 1$. □

Avec $k = 3$, une clause peut partager d'autres variables avec deux clauses.

1.4.2 Preuve constructive

Il existe une preuve plus classique qui se fait par récurrence, mais qui ne donne aucune intuition sur le résultat. Nous présentons ici une preuve constructive, un peu plus compliquée, mais basée sur un algorithme (Las Vegas) de construction.

On se place dans un cadre légèrement plus restrictif, mais qui est adapté à ce qui se passe en pratique.

On suppose l'espace de probabilité structuré de la manière suivante :

$$\Omega = C_1 \times \cdots \times C_m$$

où C_j est un ensemble fini et les dimensions sont indépendantes les unes des autres : il existe $\mathbf{P}_j$ une probabilité sur C_j tel que pour tout $(a_1, \ldots, a_m) \in C_1 \times \cdots \times C_m$,

$$\mathbf{P}(\{a_1, \ldots, a_m\}) = \prod_{j=1}^{m} \mathbf{P}_j(\{a_j\}).$$

On note A_j une v.a sur C_j.

On suppose en outre que les événements E_i s'expriment en fonction d'événements du type $A_j \in$ .... Pour tout $i \in \{1, \ldots, n\}$, on note $e(i)$ est l'ensemble minimal des j tels que E_i s'exprime en fonction des A_j.

Le graphe de dépendance peut alors se construire comme $(i, j) \in E \Leftrightarrow e(i) \cap e(j) \neq \emptyset$.

Par exemple, pour le problème k-SAT, $\Omega = \{0, 1\}^n$, où n est le nombre de variables de la formule F, et les affectations des variables sont indépendantes. Pour l'exemple précédent, E_1 est «$\{x_1 = \texttt{faux}\} \cap \{x_2 = \texttt{faux}\} \cap \{x_3 = \texttt{vrai}\}$» et $e(1) = \{1, 2, 3\}$. De la même manière, $e(3) = \{2, 3, 4\}$...

Théorème 20. *L'algorithme suivant permet d'obtenir une évaluation des v.a. A_j qui satisfait $\cap\overline{E_i}$ si $\mathbf{P}(A_j) \leq p$ et $p\frac{(d+1)^{d+1}}{d^d} \leq 1$. De plus, pour d et n fixés, cet algorithme a une complexité moyenne en $\mathcal{O}(m)$.*

C'est une version plus forte que la précédents car $\frac{(d+1)^{d+1}}{d^d} \leq 4d$.

Algorithme 4 : Algorithme de Moser et Tardos (2009)

début
  Affecter les variables (A_j) selon leur distribution;
  tant que *il existe* $i \in \{1, \dots, n\}$ *tel que* E_i *a lieu* **faire**
    choisir i tel que E_i a lieu;
    Réaffecter les variables A_j pour $j \in e(i)$ indépendamment selon leur distribution [`RESET`]

Preuve de l'algorithme

On peut tout d'abord remarquer que si l'algorithme s'arrête, alors on a bien une affectation des variables qui est dans $\cap\overline{E}_i$.

On va maintenant analyser le déroulement de l'algorithme. On note

— $\texttt{LOG} = i_1, i_2, \dots, i_t, \dots$ la suite des indices des événements choisis à chaque étape.

— pour tout i, $\texttt{COUNT}(i)$ le nombre de fois où i apparaît dans `LOG`

Si on montre que $\mathbf{E}[\texttt{COUNT}(i)] \leq \frac{1}{d}$, alors $\mathbf{E}[|\texttt{LOG}|] \leq \frac{n}{d} \leq \frac{m(d+1)}{d} \leq 2m$. On a en effet $n \leq (d+1)m$, car chaque variable apparaît dans au plus $d+1$ événements (sinon le degré de dépendance serait $> d$).

On construit un *arbre de dépendance* $\texttt{TREE}(t)$ pour chaque étape de l'algorithme, qui tente de donner une explication à la question « pourquoi réinitialise-t-on $(A_j)_{j\in e(i_t)}$ à l'étape t ? »

Plus précisément, « quelles sont les étapes de l'algorithme qui font que E_i a lieu ? »

Construction de $\texttt{TREE}(t)$ Cet arbre est étiqueté par les $i \in \{1, \dots, n\}$ qui ont déjà été réinitialisés de la manière suivante.

1. La racine est étiquetée par i_t, correspondant à l'événement choisi pour le `RESET`.
2. Si $e(i_t) \cap e(i_{t-1}) = \emptyset$, alors i_{t-1} n'apparaît pas dans l'arbre (quelle que soit la nouvelle affectation des variables de $e(i_{t-1})$ à l'étape $t-1$, l'événement choisi à l'étape t a lieu).
3. Si $e(i_t) \cap e(i_{t-1}) \neq \emptyset$, alors i_{t-1} est fils de i_t.
4. Une fois les nœuds potentiels $i_{u+1}, \dots, i_t$ traités, on s'intéresse à i_u.
   — Si pour tout i apparaissant dans l'arbre, $e(i_u) \cap e(i) = \emptyset$, alors i_u n'apparaît pas dans l'arbre.
   — Sinon soit s le nœud le plus bas (le plus loin de la racine) parmi les nœuds v tels que $e(i_v) \cap e(i_u) \neq \emptyset$. Alors i_u est fils de i_s (s'il y a plusieurs choix possibles, on choisit s minimal parmi les possibilités).

Exemple ($\texttt{TREE}(3)$).

— $[i_3]$ si $e(i_3) \cap e(i_2) = \emptyset$ et $e(i_3) \cap e(i_1) = \emptyset$

— $[i_3 \to i_1]$ si $e(i_3) \cap e(i_2) = \emptyset$ et $e(i_3) \cap e(i_1) \neq \emptyset$
— $[i_3 \to i_2]$ si $e(i_3) \cap e(i_2) \neq \emptyset$ et $(e(i_3) \cup e(i_2)) \cap e(i_1) = \emptyset$
— $[i_3 \to i_1; i_3 \to i_2]$ si $e(i_3) \cap e(i) \neq \emptyset$ et $e(i_2) \cap e(i_1) = \emptyset$
— $[i_3 \to i_2 \to i_1]$ si $e(i_2) \cap e(i) = \emptyset$ et la relation entre $e(i_3)$ et $e(i_1)$ n'intervient pas.

Propriétés de l'arbre On se place dans l'exécution de l'algorithme, on construit les arbres $\texttt{TREE}(1), \ldots, \texttt{TREE}(t), \ldots$.

1. Tous les arbres $\texttt{TREE}(t)$, $t \geq 1$ sont différents : deux arbres, $\texttt{TREE}(s)$ et $\texttt{TREE}(t)$, soit ils n'ont pas la même racine, soit ils ont la même racine i (même événement `RESET`), mais alors, chaque occurrence de i dans `LOG` jusqu'à l'itération t apparaît dans $\texttt{TREE}(t)$, et donc le nombre de nœuds étiquetés par cet i est différent. En fait, deux arbres ayant la même étiquette à la racine, ont une structure différente : ils n'ont pas le même nombe de sommets.
2. Si $e(i_u) \cap e(i_v) \neq \emptyset$ et $u < v$, alors $e(i_u)$ est strictement plus bas que $e(i_v)$.
3. Le nombre maximal de fils pour chaque nœud est $d + 1$.

L'étape suivante consiste à borner la probabilités d'obtenir un arbre d'une certaine forme et de compter les arbres possibles.

Probabilité d'un arbre Soit T un arbre étiqueté par les $i \in \{1, \ldots, n\}$ satisfaisant les propriétés ci-dessus. On pose $\texttt{OCCUR}(T)$ l'événement « $\exists t,\ \texttt{TREE}(t) = T$ ».

Lemme 2. $\mathbf{P}(\texttt{OCCUR}(T)) \leq p^{|T|}$.

Démonstration. On procède par récurrence sur la taille de l'arbre.

Regardons tout d'abord le cas où T est composé uniquement de sa racine : $T = [i]$, alors c'est que dans l'algorithme on a choisi pour la première fois l'événement E_i et que tous les événements choisis avant sont indépendants de lui. La probabilité d'avoir cet arbre est alors au plus la probabilité que dans l'affectation initiale E_i a lieu. Cette probabilité est par hypothèse au plus p.

Dans le cas général, on analyse l'arbre du bas vers le haut.

— **La feuille (étiquette i) la plus basse a une probabilité au plus** p : c'est aussi la probabilité que l'évènement qu'elle représente soit satisfait dès l'affectation initiale et les variables dont cet événement dépend ne sont pas modifiées jusqu'à la sélection de cet événement.
— **L'arbre privé de cette feuille (T') a probabilité au plus** $p^{|T|-1}$ (par hypothèse de récurrence) car l'affectation des variables après le `RESET` a la même distribution qu'initialement.
— **L'arbre T' est indépendant de la feuille i supprimée**. L'événement E_i ne dépend que de variables $(A_j)_{j \in e(i)}$. Soit les événements appairassent dans l'arbre en sont indépendants, soit ils en dépendent. Mais comme les variables sont réinitialisées, le fait que ces événements aient lieu est indépendant du fait qu'ils aient lieu dans l'affectation initiale.

Donc la probabilité d'obtenir T est inférieure à $p^{|T|}$. □

Comme chaque arbre n'apparaît qu'une fois au plus dans le déroulement de l'algorithme,

$$\mathbf{E}(\texttt{COUNT}(i)) = \sum_{T \text{ possible de racine } i} \mathbf{P}(\texttt{OCCUR}(T)) \leq \sum_{T \text{ possible de racine } i} p^{|T|} \leq \sum_{T \in T_{d+1}} p^{|T|}$$

où T_{d+1} est l'ensemble des sous-arbres finis où le nombre maximal de fils (ordonnés) de chaque nœud est au plus $d+1$.

Compter les arbres Soit g_s le nombre d'arbres finis de degré au plus $d+1$ de hauteur au plus s et donc chaque nœud est pondéré par p. On a :

$$\begin{aligned} g_0 &= p \\ g_{s+1} &= p(1 + (d+1)g_s + \binom{d+1}{2} g_s^2 + \cdots + \binom{d+1}{k} g_s^k + \cdots + g_s^{d+1} \\ &= p(1+g_s)^{d+1} = f(g_s) \end{aligned}$$

(g_s) est une suite croissante, donc tend soit vers l'infini, soit vers son plus petit point fixe. Or, par hypothèse,

$$p(1+\frac{1}{d})^{d+1} = p(\frac{d}{d+1})^{d+1} = \frac{1}{d}[p\frac{(d+1)^{d+1}}{d^d}] \leq \frac{1}{d}.$$

Donc il existe un point fixe et ainsi

$$\mathbf{E}(\texttt{COUNT}(i)) = \sum_{T \in T_{d+1}} p^{|T|} \leq \frac{1}{d}.$$

Application : cycles dans un graphe orienté

Théorème 21. *Soit $G = (V, E)$ un graphe orienté, avec un degré sortant minimal δ et un degré entrant maximal Δ. Si $4\Delta\delta(1-\frac{1}{k})^\delta \leq 1$, alors G contient un cycle orienté de longueur $0 \mod k$.*

Démonstration. Quitte à supprimer des arêtes en trop, on peut supposer que le degré sortant de chaque nœud est exactement δ. Ce faisant, on n'augmente pas le degré entrant des nœuds.

Soit $f : V \to \{0, \ldots, k-1\}$ un coloriage aléatoire uniforme des sommets : chaque sommet est colorié selon une distribution uniforme, indépendamment des autres sommets selon k couleurs possibles.

Pour chaque $v \in V$, soit A_u l'événement « il n'y a pas de sommet v tel que $(u,v) \in E$ et $f(v) \equiv f(u)+1 \mod k$ ». On a :

— $\mathbf{P}(A_u) = (1-\frac{1}{k})^\delta$ (il y a δ sommets tels que $(u,v) \in E$)

— A_u est mutuellement indépendant des A_v qui ne satisfont pas

$$N^+(v) \cap N^+(u) \neq \emptyset,$$

avec $N(u)^+ = \{v \mid (u,v) \in E\}$. Il y a au plus $\delta\Delta$ sommets qui satisfont cette propriété. On peut donc appliquer le lemme de Lovász : il existe un coloriage tel que pour tout u, il existe v tel que $f(v) \equiv f(u)+1 \mod k$. Une fois le coloriage obtenu, il suffit de suivre les arêtes telles que $f(v) \equiv f(u)+1 \mod k$. On obtient nécessairement deux fois sur le même sommet, ce qui forme un cycle, dont la longueur ne peut être que $0 \mod k$. □

1.4.3 Version générale du lemme de Lovász

Théorème 22. *Soient $E_1, \ldots, E_n$ des événements tels qu'il existe des nombres $x_i \in [0,1)$ tels que pour tout $i \in \{1, \ldots, n\}$*

$$\mathbf{P}(E_i) \leq x_i \prod_{(i,j)\in E} (1 - x_j)$$

Alors $\mathbf{P}(\cap_{i=1}^n \overline{E_i}) \geq \prod_{i=1}^n (1 - x_i)$.

Démonstration. Le principe est le même que pour le cas symétrique. On l'admet. □

Corollaire 2 (Version asymétrique du Lemme de Lovász)**.** *Si pour tout i, $\sum_{\{j \mid (i,j)\in E\}} \mathbf{P}(E_j) \leq 1/4$, alors $\mathbf{P}(\cap_{i=1}^n \overline{E_i}) \geq \prod_{i=1}^n (1 - 2\mathbf{P}(E_i)) > 0$.*

Démonstration. On pose $x_i = 2\mathbf{P}(E_i)$. Alors,

$$\begin{aligned} \mathbf{P}(E_i) &\leq 2\mathbf{P}(E_i)(1 - \sum_j 2\mathbf{P}(E_j)) = x_i(1 - \sum_j x_j) \\ &\leq x_i \prod_j (1 - x_j). \end{aligned}$$

□

1.5 Fonctions génératrices des moments et applications

1.5.1 Définition

Soient X une variable aléatoire sur $\mathbb{N}$ et $\phi : \mathbb{N} \to \mathbb{C}$ une fonction complexe. On note ϕ_R et Φ_I ses parties réelle et imaginaire. Pourvu que les espérances existent et sont finies, on a

$$\mathbf{E}[\phi(X)] = \mathbf{E}[\phi_R(X)] + i\mathbf{E}[\phi_I(X)].$$

Définition 11. *Soit X une v.a. sur $\mathbb{N}$. Sa* fonction génératrice (des moments) *est*

$$\begin{array}{rcl} g_X : \{s \in \mathbb{C} \mid |s| \leq 1\} & \to & \mathbb{C} \\ s & \mapsto & \mathbf{E}[s^X] = \sum_{k=0}^{\infty} s^k \mathbf{P}(X = k). \end{array}$$

On utilisera principalement cette fonction génératrice comme une fonction définie sur $[0, 1]$.

g_X est $\mathcal{C}^\infty$ sur $]-1, 1[$ on a $g_X(0) = \mathbf{P}(X = 0)$, $g_X(1) = 1$, $\mathbf{P}(X = n) = g_X^{(n)}(0)/n!$.

Proposition 9. *La fonction génératrice g_X est croissante et convexe sur $[0, 1]$. De plus, si $\mathbf{P}(X = 0) < 1$, alors g est strictement croissante et si $\mathbf{P}(X \leq 1) < 1$, alors g est strictement convexe.*

Démonstration.

$g_X(s) = \sum_{n\in\mathbb{N}} \mathbf{P}(X = n)s^n$ est croissante et même strictement croissante si $\mathbf{P}(X \geq 1) > 0$. $g'_X(s) = \sum_{n\in\mathbb{N}} \mathbf{P}(X = n + 1)s^{n+1}$ est croissante et même strictement croissante si $\mathbf{P}(X \geq 2) > 0$, donc g_X est convexe et strictement convexe si $\mathbf{P}(X \leq 1) < 1$. □

Proposition 10. *Soient X et Y deux variables aléatoires indépendantes, de fonctions génératrices respectives g_X et g_Y. Alors la fonction génératrice de $X + Y$ est $g_{X+Y} = g_X g_Y$.*

Démonstration. Pour tout s, $g_{X+Y}(s) = \mathbf{E}[s^{X+Y}] = \mathbf{E}[s^X]\mathbf{E}[s^Y]$. □

Exemple (distributions classiques).

- $X \sim \mathcal{B}\text{er}(p) : g_X(s) = 1 - p + ps$;
- $X \sim \mathcal{B}\text{in}(n, p) : g_X(s) = (1 - p + ps)^n$;
- $X \sim \mathcal{G}\text{eo}(p) : g_X(s) = \sum_{n\geq 1} s^n(1 - p)^{n-1}p = \frac{ps}{1-(1-p)s}$.

Lien entre les dérivées successives et les moments Soit X une variable aléatoire sur $\mathbb{N}$ et g_X sa fonction génératrice.

- $g'(s) = \sum_{n=1}^{\infty} n\mathbf{P}(X = n)s^{n-1}$ donc $g'_X(1) = \mathbf{E}[X]$.
- $g''_X(s) = \sum_{n=2}^{\infty} n(n-1)\mathbf{P}(X = n)s^{n-2} = \sum_{n=2}^{\infty} n^2\mathbf{P}(X = n)s^{n-2} - \sum_{n=2}^{\infty} n\mathbf{P}(X = n)s^{n-2}$, donc ($1^2 = 1$, on peut sommer à partir de $n = 1$) $g''_X(1) = \mathbf{E}[X^2] - \mathbf{E}[X]$ et $\mathbf{V}\text{ar}(X) = g''_X(1) + g'_X(1) - g'_X(1)^2$.
- Plus généralement, $g_X^{(k)}(s) = \sum_{n=k}^{\infty} n(n-1)\cdots(n-k+1)\mathbf{P}(X = n)s^{n-k}$ et donc si $g_X^{(k)}(1)$ existe, elle peut s'écrire sous la forme $g_X^{(k)}(1) = \mathbf{E}[X^k] + \alpha_{k-1}\mathbf{E}[X^{k-1}] + \cdots + \alpha_1\mathbf{E}[X]$.

Lien entre fonction génératrice et distribution

Proposition 11. *Soient X et Y deux variables aléatoires de fonctions génératrices respectives g_X et g_Y. Si $\forall s \in [0, \delta]$, $g_X(s) = g_Y(s)$, alors X et Y ont la même distribution.*

Cette proposition se déduit de la théorie des séries entières : si deux séries sont égales sur un intervalle $[0, \rho[$, avec $\rho > 0$, alors leurs coefficients sont égaux.

Somme aléatoire, égalité de Wald

Théorème 23. *Soient T une variable aléatoire à valeurs dans $\mathbb{N}$ et $(Z_i)_{i\in\mathbb{N}\setminus\{0\}}$ une suite de variables aléatoires i.i.d et indépendantes de T. Soit $X = \sum_{i=1}^{T} Z_i$. Soient g_Z, g_T et g_X les fonctions génératrices respectives de Z_1, T et X. Alors*

$$g_X = g_T \circ g_Z.$$

Démonstration.

$$s^{Z_1+\cdots+Z_T} = \sum_{n=0}^{\infty} \mathbf{1}_{\{T=n\}} s^{Z_1+\cdots+Z_n},$$

donc

$$\begin{aligned}
\mathbf{E}(s^{Z_1+\cdots+Z_T}) &= \sum_{n=0}^{\infty} \mathbf{E}[\mathbf{1}_{\{T=n\}} s^{Z_1+\cdots+Z_n}] \text{ (linéarité)} \\
&= \sum_{n=0}^{\infty} \mathbf{E}[\mathbf{1}_{\{T=n\}}]\mathbf{E}[s^{Z_1+\cdots+Z_n}] \quad \text{(indépendance de } T \text{ et } Z_i) \\
&= \sum_{n=0}^{\infty} \mathbf{P}(T=n)[g_Z(s)]^n \quad \text{(indépendance des } Z_i) \\
&= \mathbf{E}[g_Z(s)^T] = g_T(g_Z(s)).
\end{aligned}$$

□

Corollaire 3 (Égalité de Wald)**.** *Soient T une variable aléatoire à valeurs dans $\mathbb{N}$ et $(Z_i)_{i\in\mathbb{N}\setminus\{0\}}$ une suite de variables aléatoires i.i.d et indépendantes de T. Soient $X = \sum_{i=1}^{T} Z_i$. Let g_Z, g_T et g_X les fonctions génératrices respectives de Z_1, T et X. Alors*

$$\mathbf{E}[X] = \mathbf{E}[Z]\mathbf{E}[T].$$

Démonstration.

$$\mathbf{E}[X] = g'_X(1) = g'_Z(1)g'_T(g_Z(1)) = g'_Z(1)g'_T(1) = \mathbf{E}[Z]\mathbf{E}[T].$$

□

1.5.2 Processus de branchement de Galton-Watson

Le processus de Galton-Watson a initialement été introduit pour étudier l'extinction ou la survie des nom de famille des familles aristocratiques dans l'Angleterre victorienne. La construction est la suivante :

— $X_0 = 1$ (la racine, niveau 0) ;

— X_n est le nombre de nœuds au niveau n (ou à la n-ième génération).

On note $Z_i^{(n)}$ le nombre d'enfants du i-ème nœud de la n-ème génération ; les $(Z_i^{(n)})_{i,n}$ sont i.i.d. avec la même loi qu'une v.a. notée Z.

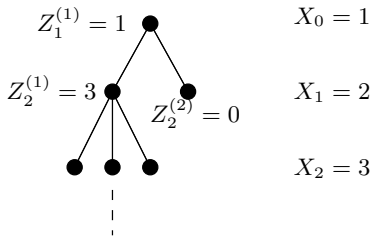


FIGURE 1.2 – Processus de branchement de Galton-Watson.

On a donc

$$X_{n+1} = \sum_{i=1}^{X_n} Z_i^{(n)}.$$

Le moyen le plus simple d'étudier ce processus est d'utiliser les fonctions génératrices des moments. Notons $g(s) = \mathbf{E}[s^Z]$ la fonction génératrice de Z, et $\phi_n = \mathbf{E}[s^{X_n}]$ celle de X_n.

Lemme 3. *$\phi_{n+1} = g_Z(\phi_n)$.*

Démonstration. D'après le théorème 23, on a $\phi_{n+1} = \phi_n \circ g_Z$. Alors,

$$\phi_{n+1} = \phi_0 \circ g_Z \circ \cdots \circ g_Z = \phi_0 \circ g_Z^{n+1}.$$

Mais $\mathbf{P}(X_0 = 1) = 1$, donc $\phi_0(s) = s$ et $\phi_{n+1} = g_Z^{n+1}$. □

Soit $p_e = \mathbf{P}(\exists n \in \mathbb{N},\ X_n = 0) = \mathbf{P}(\cup_{n\in\mathbb{N}}\{X_n = 0\})$ la probabilité d'extinction du processus. Comme $\{X_n = 0\} \subseteq \{X_{n+1} = 0\}$, par continuité séquentielle, on a $p_e = \lim_{n\to\infty} \mathbf{P}(X_n = 0)$.

Lemme 4. *$p_e = g_Z(p_e)$.*

Démonstration. On sait que $\phi_{n+1}(0) = g_Z(\phi_n(0))$. Mais $\phi_{n+1}(0) = \mathbf{P}(X_{n+1} = 0)$ et $\phi_n(0) = \mathbf{P}(X_n = 0)$. Alors, par continuité (g_Z est continue en 0), $p_e = g_Z(p_e)$. □

Théorème 24 (Point fixe)**.** *Soit l'équation $p = g(p)$ où g est la fonction génératrice d'une variables aléatoire X. 5 est strictement convexe. Si $\mathbf{P}(X = 1) < 1$ et si $\mathbf{E}[X] \leq 1$, alors l'équation $x = g(x)$ a une unique solution dans $[0, 1]$, et c'est $x = 1$. Si $\mathbf{E}[X] > 1$, alors l'équation $x = g(x)$ a deux solutions dans $[0, 1]$, $x = 1$ et $\beta \in [0, 1[$.*

Démonstration. $x = 1$ est une solution triviale de l'équation. Maintenant, en utilisant la convexité de g_Z, si $\mathbf{P}(X = 1) < 1$ et $\mathbf{E}[X] \leq 1$, alors $g'_z(1) \leq 1$ et, comme la fonction g est convexe, $\forall x < 1$, $g'_Z(x) \leq 1$ et $g_Z(x) > x$.

Si $\mathbf{E}[X] > 1$, sur un intervalle $[1 - \epsilon, 1[$, $g_Z(x) < x$. Mais $g_Z(0) \geq 0$, donc il existe β tel que $\beta = g_Z(\beta)$. □

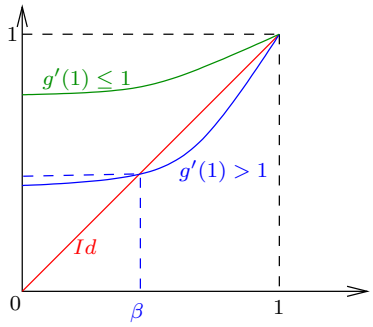


FIGURE 1.3 – Point(s) fixe(s) d'une fonction génératrice g.

Théorème 25. *Soit p_e la probabilité d'extinction du processus de Galton-Watson.*

1. *Si $\mathbf{P}(Z \neq 1) > 0$ et $\mathbf{E}[Z] \leq 1$ alors $p_e = 1$;*
2. *Si $\mathbf{P}(Z > 1) = 0$ et $\mathbf{E}[Z] = 1$, alors $p_e = 0$;*
3. *Si $\mathbf{E}[Z] > 1$, alors $p_e = \beta < 1$.*

Démonstration. Soit $x_n = \mathbf{P}(X_n = 0)$. Nous savons que $x_0 = 0$, donc $x_0 \leq \beta$. Or, si $x_n \leq \beta$, alors, comme g_Z est croissante, $x_{n+1} = g_Z(x_n) \leq g_Z(\beta) = \beta$. Donc $p_e \leq \beta$ et finalement $p_e = \beta$. □

1.5.3 Bornes de Chernoff

Inégalités

Principe : appliquer l'inégalité de Markov aux fonctions génératrices.

Théorème 26. *Soit X une v.a. sur $\mathbb{N}$. Pour tout $a \geq 0$,*

— $\mathbf{P}(X \geq a) = \inf_{s>1} \frac{\mathbf{E}[s^X]}{s^a}$;

— $\mathbf{P}(X \leq a) = \inf_{s<1} \frac{\mathbf{E}[s^X]}{s^a}$.

Démonstration. $\forall s > 1$, $\mathbf{P}(X \geq a) = \mathbf{P}(s^X \geq s^a) \leq \frac{\mathbf{E}[s^X]}{s^a}$

$\forall s < 1$, $\mathbf{P}(X \leq a) = \mathbf{P}(s^X \geq s^a) \leq \frac{\mathbf{E}[s^X]}{s^a}$ □

Quelque formes particulières :

Théorème 27. *Soient $X_1, \ldots, X_n$ n v.a. mutuellement indépendantes, $X_i \sim \mathcal{B}\text{er}(p_i)$. Soit $X = \sum_{i=1}^n X_i$ et notons $\mu = \mathbf{E}[X]$. Alors*

1. *$\forall \delta > 0$, $\mathbf{P}(X \geq (1 + \delta)\mu) \leq \left(\frac{e^\delta}{(1+\delta)^{1+\delta}}\right)^\mu$.*

2. $\forall \delta \in]0,1]$, $\mathbf{P}(X \geq (1+\delta)\mu) \leq e^{-\mu\frac{\delta^2}{3}}$.
3. $\forall R \geq 6\mu$, $\mathbf{P}(X \geq R) \leq 2^{-R}$.

Démonstration. Soit g_i la fonction génératrice de X_i. On a

$$g_i(s) = 1 - p_i + p_i s = 1 + p_i(s-1) \leq e^{p_i(s-1)}.$$

Alors

$$g_X(s) = \prod_{i=1}^{n} g_i(s) \leq \prod_{i=1}^{n} e^{p_i(s-1)} = e^{\mu(s-1)}.$$

Or, $\forall s > 1$, $\mathbf{P}(X \geq (1+\delta)\mu) \leq \frac{\mathbf{E}[s^X]}{s^{(1+\delta)\mu}} \leq \frac{e^{\mu(s-1)}}{s^{(1+\delta)\mu}}$. En prenant $s = 1+\delta$ (ce qui minimise la borne supérieure), on obtient

$$\mathbf{P}(X \geq (1+\delta)\mu) \leq \left(\frac{e^\delta}{(1+\delta)^{1+\delta}}\right)^\mu.$$

Pour montrer les autres inégalités, il suffit de remarquer que

$$\forall \delta \in]0,1], \ \frac{e^\delta}{(1+\delta)^{1+\delta}} = e^{\delta-(1+\delta)\ln(1+\delta)} \leq e^{-\frac{\delta^2}{3}}$$

et que

$$\forall \delta \geq 5, \ \left(\frac{e^\delta}{(1+\delta)^{1+\delta}}\right)^\mu \leq \left(\frac{e}{\delta+1}\right)^{(\delta+1)\mu} \leq \left(\frac{1}{2}\right)^R,$$

avec $R = (1+\delta)\mu$. □

On a le théorème similaire suivant.

Théorème 28. *Soient $X_1, \ldots, X_n$ n v.a. mutuellement indépendantes, $X_i \sim \mathcal{B}\mathrm{er}(p_i)$. Soit $X = \sum_{i=1}^n X_i$ et notons $\mu = \mathbf{E}[X]$. Alors pour tout $\delta \in]0,1[$,*

1. $\mathbf{P}(X \leq (1-\delta)\mu) \leq \left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^\mu$.
2. $\mathbf{P}(X \leq (1-\delta)\mu) \leq e^{-\mu\frac{\delta^2}{2}}$.

Démonstration. On procède exactement de la même manière, pour $s < 1$:

$$\mathbf{P}(X \leq (1-\delta)\mu) \leq \frac{\mathbf{E}[s^X]}{s^{(1-\delta)\mu}} \leq \frac{e^{\mu(s-1)}}{s^{(1-\delta)\mu}}.$$

On choisit $s = 1-\delta$ et

$$\mathbf{P}(X \leq (1-\delta)\mu) \leq \left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^\mu.$$

□

Exemple (Lancer de pièces). En utilisant les bornes de Chernoff, la probabilité d'obtenir plus de $3n/4$ fois face en n lancers est

$$\mathbf{P}(X \geq (1+1/2)\frac{n}{2}) \leq e^{-1/3\times n/2\times 1/4} = e^{-n/24}.$$

Application au tri rapide

On veut évaluer la probabilité que le tri s'effectue avec une forte probabilité ($\geq 1 - 1/n$) en au plus $\beta n \ln n$ comparaisons, β à définir. Plus précisément, on montre que

$$\exists \beta,\ \forall n \in \mathbb{N}, \mathbf{P}(QS(n) \geq \beta n \ln n) \leq 1/n,$$

où $QS(n)$ est le nombre de comparaisons à effectuer par l'algorihtme de tri rapide dans un tableau de n éléments distincts.

Pour ce faire, on considère l'arbre binaire des pivots : les nœuds sont les sous-tableaux à trier (un nœud a donc deux fils qui sont les sous-tableaux à séparer selon un pivot choisi, la taille du tableau père est donc la somme des tailles des sous-tableaux des fils plus 1). On identifie les nœuds et les pivots.

On s'intéresse à la hauteur de cet arbre.

On appelle *bon nœud* un nœud qui correspond à un sous-tableau de taille au plus $s/2$ si la taille du père est s. Soit b une branche de l'arbre. On montre les points suivants.

1. Il existe $\alpha >$ tel qu'il y a au plus $\alpha \ln n$ bons nœuds sur b.
2. Il existe β tel que $\mathbf{P}(|b| \geq \beta \ln n) \leq 1/n^2$.
3. $\mathbf{P}(\max_b |b| \geq \beta \ln n) \leq 1/n$, où b parcourt l'ensemble des branches de l'arbre.
4. $\mathbf{P}(QS(T) \geq \beta n \ln n) \leq 1/n$.

Nombre de bons nœuds sur une branche S'il y a k bons nœuds sur une branche, alors la taille finale du tableau feuille est $\ell \leq (1/2)^k n$. On doit donc avoir $(1/2)^k n \geq 1 \Leftrightarrow k \leq \frac{1}{\ln 2} \ln n$. On peut donc poser $\alpha = 1.5 (> \frac{1}{\ln 2} = 1.44)$.

Borner la longueur d'une branche Soit i un pivot sur cette branche correspondant au tri d'un tableau de taille s. Soit X_i la v.a. telle que $X_i = 1$ si et seulement si i est un bon nœud. Alors $X_i \sim \mathcal{B}\text{er}(p_i)$, avec $p_i \geq 1/2$, et les v.a. X_i sur une branche ne sont pas tout à fait indépendantes : les p_i dépendent de la parité des tableaux construits. Cependant, on peut définir des v.a. Y_i mutuellement indépendantes, $Y_i \sim \mathcal{B}\text{er}(1/2)$ et $Y_i \leq X_i$ p.s.

Essayons maintenant de borner la probabilité que b soit de longueur au moins m. Ceci peut arriver uniquement s'il y a moins de $\alpha \ln n$ bons nœuds sur e. On a donc

$$\begin{aligned}
\mathbf{P}(|b| \geq m) &= \mathbf{P}(\sum_{i \in b} X_i \leq \alpha \ln n \text{ et } |b| \geq m) \\
&= \mathbf{P}(\sum_{i \in b} Y_i \leq \alpha \ln n \text{ et } |b| \geq m) \\
&\leq \mathbf{P}(\sum_{i=1}^{m} Y_i \leq \alpha \ln n),
\end{aligned}$$

où X_i est la variable aléatoire égale à 1 si le i-ème nœud est un bon nœud et 0 sinon : il ne peut y avoir de branche avec plus de $\alpha \ln n$ bons nœuds. On cherche m de la forme $\beta \ln n$. On cherche donc à borner

$$\mathbf{P}(\sum_{i=1}^{\beta \ln n} Y_i \leq \alpha \ln n),$$

où pour tout i, $Y_i \sim \mathcal{B}\text{er}(1/2)$, Y_i sont mutuellement indépendantes.

On applique donc la deuxième version de la borne de Chernoff, deuxième forme (plus facile et suffisante pour les calculs).

$$\mathbf{P}(\sum_{i=1}^{\beta \ln n} Y_i \leq \alpha \ln n) \leq e^{-\mu\delta^2/2},$$

où
— $\mu = \sum_{i=1}^{\beta \ln n} 1/2 = \frac{\beta \ln n}{2}$;
— $(1-\delta)\mu = \alpha \ln n$, donc $\delta = 1 - \frac{\alpha \ln n}{\mu} = \frac{\beta - 2\alpha}{\beta}$.
On a donc

$$\mathbf{P}(\sum_{i=1}^{\beta \ln n} X_i \leq \alpha \ln n) \leq e^{-\frac{(\beta-2\alpha)^2}{4\beta} \ln n} = n^{-\frac{(\beta-2\alpha)^2}{4\beta}}.$$

Il suffit donc de choisir β tel que $\beta^2 - 4\alpha\beta + 4\alpha^2 \geq 8\beta$. On a pris $\alpha = 1.5$, donc on peut prendre $\beta = 14$.

On a donc bien $\mathbf{P}(|b| \geq 14 \ln n) \leq \frac{1}{n^2}$.

Borner la hauteur de l'arbre On veut borner $\mathbf{P}(\max_b |b| \geq \ln n)$. En utilisant l'*union-bound*, $\mathbf{P}(\max_b |b| \geq \ln n) \leq \sum_b P(|b| \geq 14 \ln n)$. Or, il y a n nœuds, donc au plus n branches, et

$$\mathbf{P}(\max_b |b| \geq \ln n) \leq n \sum_b P(|b| \geq 14 \ln n) \leq \frac{1}{n}.$$

Conclusion Le nombre de comparaisons effectuées est égal à la somme, pour chaque pivot, du nombre de comparaisons effectuées pour ce pivot. Chaque nœud induit donc un nombre de comparaisons égal au nombre de ses descendants. Si maintenant on compte le coût d'une comparaison non pas sur le pivot, mais sur l'autre nœud, le coût d'un nœud est le nombre de ses ancêtres, et donc c'est la profondeur de ce nœud. La probabilité que le tri s'effectue en plus que $14n \ln n$ est au plus la probabilité que l'arbre ait une hauteur plus que $14 \ln n$. Cette probabilité est au plus $1/n$.

Estimation d'un paramètre

On veut par exemple évaluer l'intensité d'une mutation génétique. On veut obtenir cette estimation à partir d'un nombre restreint d'échantillons (indépendants).

On note p la valeur à trouver (probabilité qu'il y ait une mutation) et n le nombre d'échantillons.

En ce qui concerne les observations, on note $\tilde{p}$ le taux de mutation observé : soit X la variable aléatoire du nombre de mutations, après avoir analysé n échantillons, on obtient $X = \tilde{p}n$.

Définition 12. *Un intervalle de confiance de $1-\gamma$ pour un paramètre p est un intervalle $[\tilde{p} - \delta, \tilde{p} + \delta]$ tel que $\mathbf{P}(p \in [\tilde{p} - \delta, \tilde{p} + \delta]) \geq 1 - \gamma$.*

On veut à la fois que δ et γ soient petits.

Les échantillons sont indépendants, donc $X \sim \mathcal{B}\text{in}(n, p)$, donc $\mathbf{E}[X] = np$ et $X = n\tilde{p}$. Donc si $p \notin [\tilde{p} - \delta, \tilde{p} + \delta]$, l'un des deux événements suivants a lieu :

1. $p < \tilde{p} - \delta$ et $X = n\tilde{p} > n(p + \delta) = \mathbf{E}[X](1 + \delta/p)$
2. $p > \tilde{p} - \delta$ et $X = n\tilde{p} < n(p + \delta) = \mathbf{E}[X](1 - \delta/p)$

On peut appliquer les bornes de Chernoff dans les deux cas :

$$\begin{aligned}
\mathbf{P}(p \notin [\tilde{p} - \delta, \tilde{p} + \delta]) &= \mathbf{P}(X < np(1 - \delta/p)) + \mathbf{P}(X > np(1 + \delta/p)) \\
&\leq e^{-np(\frac{\delta}{p})^2/2} + e^{-np(\frac{\delta}{p})^2/3} \\
&= e^{-n\delta^2/2} + e^{-n\delta^2/3} = \gamma.
\end{aligned}$$

1.6 Balles et Urnes - approximation poissonnienne

1.6.1 Le paradoxe de l'anniversaire

Il y a k personnes dans une salle. Quelle est la probabilité qu'au moins deux personnes aient leur anniversaire le même jour ? On suppose les dates de naissance indépendantes et uniformément distribuées. La probabilité que la $i+1$-ième personne ait son anniversaire un jour différent des i premières personnes sachant qu'elles ont un anniversaire différent est $1 - \frac{i}{365}$. Donc d'après la formule de Bayes, la probabilité que toutes les dates de naissance soient différentes est

$$p = \left(1 - \frac{1}{365}\right) \cdot \left(1 - \frac{2}{365}\right) \cdots \left(1 - \frac{k-1}{365}\right).$$

Par exemple, si $k = 30$ et $n = 365$, $p = 0,2937$ et il y a 70% de chances d'avoir un anniversaire commun.

Informellement, si k est suffisamment petit,

$$\begin{aligned} p &= \prod_{j=1}^{k-1} \left(1 - \frac{j}{n}\right) = \exp\left(\sum_{j=1}^{k-1} \ln\left(1 - \frac{j}{n}\right)\right) \\ &\simeq \exp\left(-\sum_{j=1}^{k-1} \frac{j}{n}\right) = \exp\left(-\frac{-k(k-1)}{2n}\right) \simeq e^{-\frac{k^2}{2n}}. \end{aligned}$$

Alors $p = 1/2 \overset{\simeq}{\Leftrightarrow} \frac{k^2}{2n} = \ln 2$ et $k = \sqrt{2n \ln 2}$, ce qui donne $k \simeq 23$.

1.6.2 Balles et Urnes

On lance m balles dans n urnes. Chaque balle atterrit dans une urne de manière uniforme et indépendamment des autres balles.

Si on note $X_{i,j}$ la variable égale à 1 si la balle i est lancée dans l'urne j et 0 sinon, et X_j le nombre de balles dans l'urne j après lancer des m balles (on dit aussi la charge de l'urne j), on a $X_{i,j} \sim \mathcal{B}\text{er}(1/n)$ et $X_j \sim \mathcal{B}\text{in}(m, 1/n)$.

Application : tri par paquet On suppose $n = 2^m$ éléments à trier, choisis uniformément et indépendamment dans $[0, 2^k[$, $k \geq m$.

1. On place chaque nombre dans l'urne selon les m premiers chiffres (en $\mathcal{O}(n)$) ;
2. on prend les paquets dans l'ordre et on les trie. On peut même prendre pour cela un tri naïf, en temps quadratique.

Pour chaque paquet i, de taille X_i, la complexité est donc de cX_i^2, et $X_i \sim \mathcal{B}\text{in}(n, \frac{1}{n})$. Alors, on a $\mathbf{E}[X_i^2] = n\frac{1}{n}(1 - \frac{1}{n}) + n^2(\frac{1}{n})^2 = 2 - \frac{1}{n}$. Ainsi, la complexité totale est

$$\mathbf{E}\Big[\sum_i cX_i^2\Big] = c\sum_i \mathbf{E}\left[X_i^2\right] < 2cn.$$

On s'intéresse maintenant aux deux questions suivantes :

1. Quelles est la charge maximale d'une urne ?
2. Quelle est la distribution de la charge d'une urne quand m et n tendent vers $+\infty$?

Théorème 29. *Si n balles sont jetées indépendamment et uniformément dans n urnes, alors pour n suffisamment grand, la probabilité que la charge maximale d'une urne dépasse $3\frac{\ln n}{\ln\ln n}$ est au plus $1/n$. Autrement dit,*

$$\exists N \text{ tel que } \forall n \geq N,\ \mathbf{P}\left(\max_i X_i \geq 3\frac{\ln n}{\ln\ln n}\right) \leq \frac{1}{n}.$$

Démonstration. Soit M un entier. Alors $\mathbf{P}(X_i \geq M) \leq \binom{n}{M}(\frac{1}{n})^M$. En effet, pour chaque choix de M boules parmi les m, la probabilité qu'elles soient toutes placées dans l'urne i est $1/n^M$, et il y a $\binom{n}{M}$ choix possibles. Par union bound, on obtient donc

$$\mathbf{P}(\max X_i \geq M) = \mathbf{P}\Big(\bigcup_i \{X_i \geq M\}\Big) \leq \sum_i \mathbf{P}(X_i \geq M) \leq n\binom{n}{M}\left(\frac{1}{n}\right)^M.$$

On sait que $e^M = \sum_i \frac{M^i}{i!} \geq \frac{M^M}{M!}$ et ainsi que $\binom{n}{M} \leq \frac{n^M}{M!} \leq (\frac{ne}{M})^M$. Donc,

$$\binom{n}{M}\left(\frac{1}{n}\right)^M \leq \left(\frac{e}{M}\right)^M.$$

En prenant $M = \lceil 3\frac{\ln n}{\ln\ln n}\rceil$, on obtient

$$\begin{aligned}
\mathbf{P}\left(\max X_i \geq 3\frac{\ln n}{\ln\ln n}\right) &\leq n\left(\frac{e}{M}\right)^M \\
&\leq n\left(\frac{e\ln\ln n}{3\ln n}\right)^{3\frac{\ln n}{\ln\ln n}} \\
(\tfrac{e}{3} \leq 1) \qquad &\leq \exp\left(\ln n + \frac{3\ln n}{\ln\ln n}(\ln\ln\ln n - \ln\ln n)\right) \\
&\leq \exp\left(-2\ln n + 3\ln n\frac{\ln\ln\ln n}{\ln\ln n}\right) \\
&\leq \frac{1}{n} \text{ pour } n \text{ grand.}
\end{aligned}$$

En effet, $\frac{\ln\ln\ln n}{\ln\ln n} \overset{n\to\infty}{\longrightarrow} 0$. □

1.6.3 Limite de la loi binomiale

Reprenons le modèle à m balles et n urnes. On note $U_{i,r}$ l'événement « il y a r balles dans l'urne i ». On a alors $\mathbf{P}(U_{1,0}) = \left(1-\frac{1}{n}\right)^m$ et la proportion d'urnes vides est aussi

$$\left(1-\frac{1}{n}\right)^m \simeq e^{-m/n}.$$

Plus généralement, on a

$$\mathbf{P}(U_{i,r}) = \binom{m}{r}\left(\frac{1}{n}\right)^r\left(1-\frac{1}{n}\right)^{m-r} \simeq \frac{1}{r!}e^{-m/n}\left(\frac{m}{n}\right)^r.$$

Théorème 30. *Soit* $X_n \sim \mathcal{B}\text{in}(n, p(n))$ *avec* $np(n) \overset{n\to\infty}{\longrightarrow} \lambda$. *Alors* $\lim_{n\to\infty} \mathbf{P}(X_n = k) = e^{-\lambda}\frac{\lambda^k}{k!}$.

En particulier, pour un système de m balles et n urnes où $\frac{n}{m} \to \lambda$,

$$\lim_{n\to\infty} \mathbf{P}(X_i = r) = \frac{e^{-\lambda}\lambda^r}{r!}.$$

Cette loi limite est la loi de Poisson ($X \sim \mathcal{P}\text{oi}(\lambda)$) : $x \sim \mathcal{P}\text{oi}(\lambda)$ si pour tout $k \geq 0$, $\mathbf{P}(X = k) = \frac{e^{-\lambda}\lambda^k}{k!}$.

Démonstration. On écrit p à la place de $p(n)$ pour ne pas alourdir l'écriture. Soit $X_n \sim \mathcal{B}\text{in}(n, p)$. Donc $g_n(s) = g_{X_n}(s) = (ps + (1-p))^n = (1 - p(1-s))^n$. On admet que si (g_n) converge simplement vers g une fonction génératrice sur $[0, 1]$, alors la distribution de X_n tend vers la distribution caractérisée par la fonction génératrice g. Posons $\lambda(n) = pn$. On a pour tout $s \in [0, 1]$,

$$\begin{aligned}
g_n(s) &= \exp(n \ln[1 - p(1-s)]) \\
&= \exp\left(n \ln\left[1 - \frac{\lambda(n)}{n}(1-s)\right]\right) \\
&= \exp\left(n \ln\left[1 - \frac{\lambda}{n}(1-s) + o(1/n)\right]\right) \\
&= \exp\left(n\left[-\frac{\lambda}{n}(1-s) + o(1/n)\right]\right) \\
&= e^{-\lambda(1-s)}(1 + o(1)).
\end{aligned}$$

Donc $g_n(s) \overset{n\to+\infty}{\longrightarrow} e^{-\lambda(1-s)} = \sum_{k\geq 0} e^{-\lambda}\frac{\lambda^k}{k!}s^k$, qui est une fonction génératrice, et

$$\mathbf{P}(X_n = k) \overset{n\to\infty}{\longrightarrow} e^{-\lambda}\frac{\lambda^k}{k!}.$$

□

Quelques propriétés de la loi de Poisson

— Sa fonction génératrice est : $g(s) = \sum_{k\geq 0} s^k \frac{e^{-\lambda}\lambda^k}{k!} = e^{-\lambda}e^{s\lambda} = e^{\lambda(s-1)}$;
— son espérance est donc $\mathbf{E}[X] = g'(1) = \lambda$;
— et sa variance est $\mathbf{V}\text{ar}(X) = g''(1) + g'(1) - g'(1)^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$.

Lemme 5. *Si deux variables indépendantes* X_1 *et* X_2 *vérifient* $X_1 \sim \mathcal{P}\text{oi}(\lambda_1)$ *et* $X_2 \sim \mathcal{P}\text{oi}(\lambda_2)$, *alors*

$$X_1 + X_2 \sim \mathcal{P}\text{oi}(\lambda_1 + \lambda_2)$$

Démonstration. La fonction génératrice de $X_1 + X_2$ est $g_{X_1+X_2} = \mathbf{E}[s^{X_1+X_2}]$ et donc par indépendance de X_1 et X_2, $g_{X_1+X_2} = g_{X_1}g_{X_2} = e^{(\lambda_1+\lambda_2)(s-1)}$. □

1.6.4 Approximation poissonnienne

On veut gérer la dépendance des urnes : si l'urne 1 est vide, les m balles sont distribuées sur les $m-1$ autres urnes. Ainsi les urnes ne sont pas indépendantes. On aimerait toutefois les traiter comme si elles l'étaient, afin, par exemple, d'utiliser les bornes de Chernoff (où l'indépendance permet d'obtenir des bornes plus simples à calculer).

On pose

— $Y_1^{(m)}, \dots, Y_n^{(m)} \sim \mathcal{P}\text{oi}(m/n)$ des variables mutuellement indépendantes (modèle poissonnien) ;

— $X_1^{(m)}, \dots, X_n^{(m)}$ le nombre respectif de balles dans les urnes numérotées de 1 à n. Ces variables aléatoires ne sont pas indépendantes (modèle d'urnes). En particulier, $\sum_{i=1}^n X_i^{(m)} = m$.

Théorème 31. *La distribution de $(Y_1^{(m)}, \dots, Y_n^{(m)})$ conditionnée à $\sum_{i=1}^n Y_i^{(m)} = k$ est la même que celle de $(X_1^{(k)}, \dots, X_n^{(k)})$ (distribution des urnes obtenue après qu'on a jeté k balles).*

Démonstration. On veut montrer que pour tout k, $k_1, \dots, k_n$ tels que $\sum_i k_i = k$,

$$\mathbf{P}(Y_1^{(m)} = k_1, \dots, Y_n^{(m)} = k_n \mid \sum_{i=1}^n Y_i^{(m)} = k) = \mathbf{P}(X_1^{(k)} = k_1, \dots, X_n^{(k)} = k_n).$$

D'une part, on a

$$A = \mathbf{P}(X_1^{(k)} = k_1, \dots, X_n^{(k)} = k_n) = \frac{k!}{k_1!k_2!\cdots k_n!}\frac{1}{n^k}.$$

D'autre part,

$$B = \mathbf{P}(Y_1^{(m)} = k_1, \dots, Y_n^{(m)} = k_n \mid \sum_{i=1}^n Y_i^{(m)} = k) = \frac{\mathbf{P}(Y_1^{(m)} = k_1, \dots, Y_n^{(m)} = k_n)}{\mathbf{P}(\sum_{i=1}^n Y_i^{(m)} = k)}.$$

Or les variables $Y_i^{(m)} \sim \mathcal{P}\text{oi}(m/n)$ sont mutuellement indépendantes. Donc $\sum_{i=1}^n Y_i^{(m)} \sim \mathcal{P}\text{oi}(m)$. Donc

$$B = \frac{\prod_{i=1}^n e^{-m/n}(m/n)^{k_i}/k_i!}{e^{-m}m^k/k!} = \frac{1}{n^k}\frac{k!}{k_1!\cdots k_n!} = A.$$

□

Théorème 32. *Soit $f : \mathbb{N}^n \to \mathbb{R}_+$. Alors,*

$$\mathbf{E}[f(X_1^{(m)}, \dots, X_n^{(m)})] \leq e\sqrt{m}\mathbf{E}[f(Y_1^{(m)}, \dots, Y_n^{(m)})].$$

Démonstration.

$$\begin{aligned}
\mathbf{E}[f(Y_1^{(m)},\dots,Y_n^{(m)})] &= \sum_{k=0}^{\infty}\mathbf{E}[f(Y_1^{(m)},\dots,Y_n^{(m)}) \mid \sum_{i=1}^{n}Y_i^{(m)}=k]\mathbf{P}(\sum_{i=1}^{n}Y_i^{(m)}=k)\\
&\geq \mathbf{E}[f(Y_1^{(m)},\dots,Y_n^{(m)}) \mid \sum_{i=1}^{n}Y_i^{(m)}=m]\mathbf{P}(\sum_{i=1}^{n}Y_i^{(m)}=m)\\
&= \mathbf{E}[f(X_1^{(m)},\dots,X_n^{(m)})]\frac{e^{-m}m^m}{m!}\\
&\geq \frac{1}{e\sqrt{m}}\mathbf{E}[f(X_1^{(m)},\dots,X_n^{(m)})]
\end{aligned}$$

car $m! \leq e\sqrt{m}\,(\frac{m}{e})^m$. En effet, $\ln m! = \sum_{i=1}^{m}\ln i$. Or $\int_{i-1}^{i}\ln x dx \geq \frac{\ln i+\ln(i-1)}{2}$ car ln est concave.

Alors, $\int_1^m \ln x dx \geq \frac{\ln 1}{2}+2\sum_{i=2}^{m-1}\frac{\ln i}{2}+\frac{\ln m}{2} = \sum_{i=1}^{m}\ln i - \frac{\ln m}{2}$. D'autre part, une intégration par parties donne $\int_1^m \ln x dx = [x\ln x - x]_1^m = m\ln m - m + 1 \geq \ln m! - \frac{\ln m}{2}$. Donc en prenant l'exponentielle des deux expressions, on obtient $\frac{m^m}{e^m}e \geq \frac{m!}{\sqrt{m}}$.

□

Corollaire 4. *La probabilité qu'un événement ait lieu dans le cas Poisson avec probabilité p a lieu dans le cas des urves avec probabilité inférieure à $e\sqrt{m}p$.*

Application : borne inférieure sur la charge maximum des urnes

Théorème 33. *Quand n balles sont lancées uniformément et indépendamment dans n urnes, la charge maximale est au moins $\frac{\ln n}{\ln\ln n}$ avec probabilité au moins $1-1/n$ pour n suffisamment grand :*

$$\exists N \text{ tel que } \forall n \geq N,\ \mathbf{P}\left(\max_i X_i \leq \frac{\ln n}{\ln\ln n}\right) \leq \frac{1}{n}.$$

Démonstration. Plaçons-nous dans le modèle poissonnien. La probabilité que l'urne 1 ait une charge au moins M est $\mathbf{P}(Y_1^{(n)} \geq M) \geq \mathbf{P}(Y_1^{(n)} = M) = \frac{e^{-1}}{M!}$. (dans le cas où $m = n$, $(Y_1^{(n)} \sim \mathcal{P}\text{oi}(1))$

Ainsi, la probabilité qu'aucune boîte n'ait de charge au moins M est

$$\begin{aligned}
\mathbf{P}(\max_i Y_i^{(n)} < M) &\leq \left(1-\frac{1}{eM!}\right)^n \quad \text{(indépendance)}\\
&\leq e^{-\frac{n}{eM!}}.
\end{aligned}$$

Dans le modèle d'urnes, on aura donc

$$\mathbf{P}(\max_i X_i^{(n)} \leq M) \leq e\sqrt{n}e^{-\frac{n}{eM!}}.$$

Si on choisit M tel que $e^{-\frac{n}{eM!}} \leq n^{-2}$, alors pour n assez grand, on aura bien le résultat souhaité.

Il reste à vérifier que c'est bien le cas avec $M = \lfloor\frac{\ln n}{\ln\ln n}\rfloor$.

$$\begin{aligned}
-\frac{n}{eM!} \leq -2\ln n &\Leftrightarrow M! \leq \frac{n}{2e\ln n}\\
&\Leftrightarrow \ln M! \leq \ln n - \ln 2 - 1 - \ln\ln n.
\end{aligned}$$

Or $M! \leq e\sqrt{M}\left(\frac{M}{e}\right)^M \leq M\left(\frac{M^M}{e^M}\right)$ et

$$\begin{aligned}
\ln M! &\leq \ln M + M\ln M - M \\
&\leq \ln\left(\frac{\ln n}{\ln\ln n}\right) + \frac{\ln n}{\ln\ln n}\ln\left(\frac{\ln n}{\ln\ln n}\right) - \frac{\ln n}{\ln\ln n} \\
&\leq \frac{\ln n}{\ln\ln n}(\ln\ln n - \ln\ln\ln n) + \ln\ln n - \ln\ln\ln n - \frac{\ln n}{\ln\ln n} \\
&\leq \ln n\left(1 - \frac{\ln\ln\ln n}{\ln\ln n} + \frac{\ln\ln n}{\ln n} - \frac{\ln\ln\ln n}{\ln\ln n}\right) - \frac{\ln n}{\ln\ln n} \\
&\leq \ln n - \frac{\ln n}{\ln\ln n} \qquad \text{pour } n \text{ assez grand} \\
&\leq \ln n - \ln 2 - 1 - \ln\ln n \qquad \text{car } \ln\ln n = o\left(\frac{\ln n}{\ln\ln n}\right)
\end{aligned}$$

□

1.6.5 Puissance de deux choix

On considère le problème d'urne suivant, avec n urnes et n balles. Pour chaque balle, on tire $d \geq 2$ urnes au hasard, indépendamment, et on place la balle dans l'urne la moins pleine (si égalité, on choisit l'urne au hasard parmi les moins pleines).

Théorème 34. *Après affectation des n balles, la charge maximale d'une urne est au plus $\ln\ln n/\ln d + O(1)$ avec probabilité $1 - o(1/n)$.*

Dans la preuve, pour $m \in \mathbb{N}$ et $q \in [0,1]$, on note $\mathcal{B}\text{in}(m,q)$ une v.a. de loi binomiale de paramètres m et q, indépendante des autres variables introduites.

Lemme 6. $\mathbf{P}(\mathcal{B}\text{in}(m,q) \geq 2mq) \leq e^{-mq/3}$.

Démonstration. On applique la borne de Chernoff avec $\mu = mq$ et $\delta = 1$: $\mathbf{P}(\mathcal{B}\text{in}(m,q) \geq 2mq) \leq e^{-\mu\delta^2/3} = e^{-mq/3}$. □

On différencie deux types d'urnes :

1. celles dont la charge est très petite ;
2. celles dont la charge est moins petite, c'est-à-dire en $\ln\ln n/\ln d + O(1)$, donc c'est la charge maximale que l'on vise.

On note B_i le nombre d'urnes contenant au moins i balles et on définit $\beta_4 = n/4$, $\beta_{i+1} = 2\beta_i^d/n^{d-1}$ et $p_i = \beta_i^d/n^d$. Soit $i^* = \min\{i \in \mathbb{N} \mid p_i < 6\ln n/n\}$.

Lemme 7. $i^* = \ln\ln n/\ln d + O(1)$.

Démonstration. On commence par montrer par récurrence que $\beta_{4+i} = \frac{n}{2^{2d^i - \sum_{j=0}^{i-1} d^j}} \leq \frac{n}{2^{d^i}}$.

En effet, c'est vrai pour $i = 0$, et

$$\beta_{4+i+1} = \frac{2\beta_i^d}{n^{d-1}} = \frac{2n^d}{n^{d-1}2^{d(2d^i - \sum_{j=0}^{i-1} d^j)}}.$$

Donc $\beta_{i+4} \leq \frac{n}{2^{d^i}}$ et $\frac{n^d}{2^{d^{i+1}}}/n^d < 6\ln n/n \Rightarrow \beta_{i+4}^d/n^d < 6\ln n/n$. Or

$$\begin{aligned}\frac{1}{2^{d^{i+1}}} < 6\ln n/n &\Leftrightarrow 2^{d^{i+1}} > n(6\ln n)^{-1} \\ &\Leftrightarrow \ln\ln 2 + \ln d^{i+1} > \ln(\ln n - \ln 6 - \ln\ln n) \\ &\Leftrightarrow (i+1)\ln d > -\ln\ln 2 + \ln(\ln n - \ln 6 - \ln\ln n),\end{aligned}$$

d'où le résultat. □

Cas $i \leq i^*$ (ou $np_i \geq 6\ln n$) : Pour $i \leq i^*$, soit $\mathcal{E}_i$ l'événement $\{B_i \leq \beta_i\}$.

Lemme 8. $\mathbf{P}(\neg\mathcal{E}_{i^*}) \leq \frac{i^*}{n^2}$.

Démonstration. On montre ce lemme grâce à une relation entre les $\mathcal{E}_i$.

Tout d'abord, on remarque que $\mathbf{P}(\mathcal{E}_4) = 1$. En effet, s'il y avait strictement plus de $n/4$ urnes avec plus de 4 balles, alors il y aurait strictement plus de n balles.

Ensuite, montrons que si $p_i n \geq 6\ln n$, $\mathbf{P}(\neg\mathcal{E}_{i+1}) \leq \mathbf{P}(\neg\mathcal{E}_i) + \frac{1}{n^2}$.
Tout d'abord,

$$\mathbf{P}(\neg\mathcal{E}_{i+1}) = \mathbf{P}(\neg\mathcal{E}_{i+1}, \mathcal{E}_i) + \mathbf{P}(\neg\mathcal{E}_{i+1} \mid \neg\mathcal{E}_i)\mathbf{P}(\neg\mathcal{E}_i) \leq \mathbf{P}(\neg\mathcal{E}_{i+1}, \mathcal{E}_i) + \mathbf{P}(\neg\mathcal{E}_i).$$

On remarque que $\beta_{i+1} = 2p_i n$. De plus, une balle tombe dans une urne de charge au moins $i+1$ si elle tire d urnes de charge au moins i. Cela arrive avec probabilité au plus $(\beta_i/n)^d = p_i$ si $\mathcal{E}_i$ a lieu. On a donc

$$\mathbf{P}(\neg\mathcal{E}_{i+1}, \mathcal{E}_i) = \mathbf{P}(B_{i+1} > \beta_{i+1}, \mathcal{E}_i) \leq \mathbf{P}(\mathcal{B}\mathrm{in}(n, p_i) > 2p_i n) \leq e^{-p_i n/3}.$$

Le passage de la première inégalité n'est pas trivial. En effet, les lancers de boules ne sont pas indépendants dans le sens où une boule va dans une urne dont le nombre de boules dépend des lancers précédents, et le nombre d'urnes de charge au moins $i+1$ depend de la distribution du nombre de balles dans les urnes, et pas seulement du nombre de balles de charge au moins i.

On effectue la transformation suivante sur le modèle. À chaque boule, on associe une hauteur, qui est le nombre de boules dans l'urne où elle est placée, à l'instant où elle y est placée. On note Y_j la variable aléatoire caractéristique de l'événement « la boule j a une hauteur au moins $i+1$ et $B_i(j) \leq \beta_i$ », où $B_i(j-1)$ est le nombre d'urne de charge au moins i quand $j-1$ balles ont été lancées. Le nombre de boules de hauteur au moins $i+1$ est supérieur au nombre d'urnes au final de charge au moins $i+1$, donc l'événement « $\neg\mathcal{E}_{i+1} \cap \mathcal{E}_i$ » est inclus dans l'événement « $\sum_{j=1}^n Y_j > \beta_{i+1} = 2np_i$ ».

D'une part,

$$\mathbf{P}(Y_j = 1) = \mathbf{P}(Y_j = 1 \mid \mathcal{E}_i)\mathbf{P}(\mathcal{E}_i) + \mathbf{P}(Y_j = 1 \cap \neg\mathcal{E}_i) \leq \mathbf{P}(Y_j = 1 \mid \mathcal{E}_i) + 0 \leq \left(\frac{\beta_i}{n}\right)^d = p_i.$$

Les Y_j ne sont pas indépendants, mais on peut construire une famille i.i.d $(Z_j)_{j=1}^n$ de variables Bernoulli de paramètre p_i telles que $Z_j \geq Y_j$ p.s. Les Y_j sont construites par rapport aux Z_j selon le procédé suivant :

— $Z_1 \sim \mathcal{B}\text{er}(p_i)$, $Y_1 = 0$ (on rappelle que $i > 4$, et la pemière balle lancée est toujours de hauteur 1). On a bien $Z_1 \geq Y_1$ p.s ;
— Supposons $Z_1, \ldots, Z_j$ et $Y_1 \ldots, Y_j$ définis. Soit $Z_{j+1} \sim \mathcal{B}\text{er}(p_i)$ indépendant de ces variables. Pour chaque réalisation possible des j premiers tirs, on s'intéresse à la probabilité que $Y_{j+1} = 1$. On note ω cette réalisation.
  — Soit $B_i(j)(\omega) > \beta_i$, auquel cas $Y_{j+1}(\omega) = 0$;
  — soit $B_i(j)(\omega) \leq \beta_i$, et $\mathbf{P}(Y_{j+1}(\omega)) \leq p_i$.
  On peut donc toujours faire dépendre Y_{j+1} de Z_{j+1} de sorte que $Y_{j+1} = 0$ si $Z_{j+1} = 0$ et $Z_{j+1} = 1$ si $Y_{j+1} = 1 : Z_{j+1} \geq Y_{j+1}$ p.s.

Si $p_i n \geq 6 \ln n$, alors $\mathbf{P}(\neg\mathcal{E}_{i+1},\ \mathcal{E}_i) \leq e^{-p_i n/3} \leq \frac{1}{n^2}$. On en déduit que

$$\mathbf{P}(\neg\mathcal{E}_{i+1}) \leq \frac{1}{n^2} + \mathbf{P}(\neg\mathcal{E}_i).$$

Enfin, $\mathbf{P}(\neg\mathcal{E}_{i^*}) \leq \frac{1}{n^2} + \mathbf{P}(\neg\mathcal{E}_{i^*-1}) \leq \cdots \leq \frac{i^*-4}{n^2} + \mathbf{P}(\neg\mathcal{E}_4) \leq \frac{i^*}{n^2}$ □

Cas $i > i^*$: On suppose maintenant que $p_i n < 6 \ln n$ (donc les calculs précédents ne sont plus valides). On définit $\mathcal{E}_{i^*+1} = \{B_{i^*+1} \leq 12 \ln n\}$ et $\mathcal{E}_{i^*+2} = \{B_{i^*+2} \leq 1\}$ et $\mathcal{E}_{i^*+3} = \{B_{i^*+3} = 0\}$.

Lemme 9. $\mathbf{P}(\neg\mathcal{E}_{i^*+3}) = o(1/n)$.

Démonstration. On borne tout d'abord $\mathbf{P}(\neg\mathcal{E}_{i^*+2})$. Tout d'abord, $\mathbf{P}(\neg\mathcal{E}_{i^*+1},\ \mathcal{E}_{i^*}) \leq \frac{1}{n^2}$, et $\mathbf{P}(\neg\mathcal{E}_{i^*+1}) \leq \frac{i^*+1}{n^2}$: de la même manière que précédemment,

$$\mathbf{P}(\neg\mathcal{E}_{i^*+1},\ \mathcal{E}_{i^*}) \leq \mathbf{P}(\mathcal{B}\text{in}(n, 6 \ln n/n) \geq 12 \ln n) \leq e^{-6 \ln n/3} \leq n^{-2}.$$

Donc, $\mathbf{P}(\neg\mathcal{E}_{i^*+1}) \leq \mathbf{P}(\neg\mathcal{E}_{i^*}) + n^{-2} \leq \frac{i^*+1}{n^2}$.

Maintenant, $\mathbf{P}(\neg\mathcal{E}_{i^*+3}) = \mathbf{P}(B_{i*+3} \geq 0) \leq \mathbf{P}(B_{i*+2} \geq 2) = \mathbf{P}(\neg\mathcal{E}_{i^*+2})$. Or, toujours sur le même principe, $\mathbf{P}(\neg\mathcal{E}_{i^*+2}) \leq \mathbf{P}(\neg\mathcal{E}_{i^*+2},\ \mathcal{E}_{i^*+1}) + \mathbf{P}(\neg\mathcal{E}_{i^*+1})$. On a

$$\mathbf{P}(\neg\mathcal{E}_{i^*+2},\ \mathcal{E}_{i^*+1}) \leq \mathbf{P}(\mathcal{B}\text{in}(n, (12 \ln n/n)^d) \geq 2)) \leq n^2 \left(\frac{12 \ln n}{n}\right)^{2d} = o(n^{2d-3}) = o(n^{-1}),$$

d'où le résultat. □

En conclusion on remarque que $P(\mathcal{E}_{i^*+2}) = 1 - o(1/n)$ est la probabilité qu'il n'y ait pas d'urne de charge au moins $i^* + 2$ et que $i^* + 2 = \ln \ln n / \ln d + O(1)$.