Introduction aux lois du Zéro-Un

R. Rossignol et B. Ycart

MAP5, FRE CNRS 2428,
Université René Descartes, Paris
{rost/ycart}@math-info.univ-paris5.fr

Table des matières

1 Introduction

A l'échelle microscopique, la matière est essentiellement aléatoire. Or le monde tel que nous le percevons est déterministe, prédictible. Le mouvement désordonné des molécules de gaz ne nous empêche pas de mesurer des températures ou des pressions. Malgré l'agitation moléculaire, nous sommes capables de prévoir la quantité de combustible nécessaire pour chauffer une pièce connaissant son volume, ou la puissance transmise par un piston à de l'air comprimé. Quel miracle fait que, en passant de l'échelle microscopique à l'échelle macroscopique, ce qui était aléatoire devient déterministe ? Nous avons tous une réponse en tête : la "loi des grands nombres".

La loi des grands nombres des probabilistes dit qu'une moyenne de variables indépendantes est "essentiellement constante". On peut définir rigoureusement cette notion, en décidant qu'une suite de variables aléatoires (X_n) est essentiellement constante si elle est équivalente à la suite des espérances $(\mathbb{E}[X_n])$, à savoir si $Prob[\, X_n \leq c\mathbb{E}[X_n]\,]$ converge vers 0 pour $c < 1$, vers 1 pour $c \geq 1$. L'inégalité de Bienaymé-Chebyshev fournit une condition pour qu'il en soit ainsi : il suffit que l'écart-type de X_n soit petit devant son espérance.

$$\lim_{n\to\infty} \frac{\sqrt{Var[X_n]}}{\mathbb{E}[X_n]} = 0 \quad \Longrightarrow \quad \lim_{n\to\infty} Prob\left[\left|\frac{X_n}{\mathbb{E}[X_n]} - 1\right| > \varepsilon\right] = 0\,,\ \forall \varepsilon > 0\,.$$

Pour une somme de n variables indépendantes, l'espérance est proportionnelle à n, l'écart-type à $\sqrt{n}$, et ce critère s'applique. Mais ceci n'est qu'un cas particulier d'une démarche très générale, qui est commune non seulement à tous les probabilistes, mais aussi à bon nombre de physiciens, chimistes, biologistes, et bien sûr informaticiens. Devant un phénomène aléatoire (variable ou processus), on commence par se placer à son *échelle de localisation* (son espérance). Il y a "loi des grands nombres" si à l'échelle de localisation (macroscopique) le phénomène est essentiellement déterministe. On se place ensuite autour de l'espérance, à l'*échelle des fluctuations*, donnée par l'écart-type. A cette échelle (microscopique), le phénomène est bien aléatoire, et on cherche à démontrer une convergence en loi. Dans le cas des sommes de n variables indépendantes, la localisation est en $O(n)$, les fluctuations en $O(\sqrt{n})$ et la limite en loi est gaussienne : c'est le théorème central limite ([37] p. 258). Mais bien d'autres comportements sont possibles. Par exemple pour la connexité des graphes aléatoires la localisation est en $O(\log(n)/n)$, les fluctuations en $O(1/n)$ et la loi limite est la loi de Gumbel ([101] p. 303 et théorème 2.17 ci-après). Il peut se faire qu'il n'y ait pas convergence en loi à l'échelle des fluctuations, comme pour les runs de 1 consécutifs ([69] et théorème 2.15). L'informatique fournit un catalogue particulièrement riche et varié de ces comportements asymptotiques (voir le livre de Flajolet et Sedgewick [97]). Notre propos ici n'est pas de les recenser. L'idée des lois du zéro-un est de prédire pour une classe de phénomènes la plus riche possible dans un modèle donné, que leurs fluctuations seront effectivement négligeables.

L'intuition que chacun a de la "loi des grands nombres" dépasse les sommes de variables indépendantes. Ce serait plutôt l'énoncé vague suivant.

Dans un gros espace,
tout événement raisonnable
a une probabilité proche de zéro ou un.

On appelle *loi du zéro-un* un théorème qui, dans un certain contexte, rend rigoureuse l'intuition ci-dessus. Pour ce faire, il convient de donner un sens précis aux trois membres de phrase, qui donnent le plan de ce cours :

"Dans un gros espace" :
Nous présenterons dans la première partie les modèles probabilistes sur lesquels on peut démontrer une loi du zéro-un. Ce sont en majorité des espaces produits. Nous insisterons surtout sur $\{0,1\}^n$, muni d'un produit de lois de Bernoulli (2.1). Ses variantes sont multiples. En premier lieu viennent les graphes aléatoires (2.2), étudiés depuis longtemps. Nous présenterons également le modèle le plus simple d'image aléatoire (2.3). Le problème de la k-satisfiabilité (2.4) relève aussi de $\{0,1\}^n$, et il a suscité une intense activité ces dernières années. Nous discuterons en 2.5 de la possibilité d'étendre les lois du zéro-un à d'autres modèles que le modèle produit (structures combinatoires, faible dépendance...).

"tout événement raisonnable" :
Dans n'importe quel modèle, il est facile d'exhiber des foules d'événements dont la probabilité reste loin de 0 et 1. Le problème pour un modèle donné est de définir une classe d'événements, la plus large possible, pour laquelle on sache démontrer que leur probabilité est proche de 0 ou 1. Les réponses classiques (Borel-Cantelli, Kolmogorov, Hewitt-Savage) traitent d'événements dépendant d'une infinité de variables aléatoires, et ont donc une portée pratique assez faible (3.1). Le théorème de Glebskii-Fagin est beaucoup plus intéressant, puisqu'il traite tous les événements exprimables dans la logique du premier ordre (3.2). L'approche plus récente de Friedgut et Kalai, avec la notion de seuil, donne un résultat plus précis qu'une simple convergence, pour des événements croissants, dépendant de façon symétrique des coordonnées (3.3). Mais ces deux résultats, pour spectaculaires qu'ils soient, traitent de valeurs fixes du paramètre p de la loi de Bernoulli. Or il est bien connu, pour les graphes aléatoires ou la k-satisfiabilité par exemple, que les phénomènes intéressants se produisent pour des valeurs de p tendant vers 0 avec n. Ceci a conduit à introduire les notions de seuil grossier ou fin : coarse ou sharp (3.4).

"a une probabilité proche de 0 *ou* 1*" :*
Encore faut-il être capable de quantifier des probabilités faibles, ou d'en donner des majorants suffisamment précis. En théorie des probabilités, il existe des livres entiers d'inégalités [108]. Nous commencerons par en rappeler un petit nombre, parmi les plus utilisées (4.1). Le but de ces inégalités est de quantifier la concentration de la mesure dans un espace produit. Récemment, Talagrand a développé divers aspects de la concentration (4.2). Ses démonstrations étant très délicates, beaucoup d'efforts ont été faits pour proposer une autre approche, à partir des inégalités de Sobolev logarithmiques (4.3). Enfin, dans la dernière partie (4.4), nous reviendrons sur la relation entre les inégalités de concentration au sens propre, et les inégalités sur les largeurs de seuil du type Friedgut et Kalai.

2 Dans un gros espace ...

2.1 Un peu de mousse

Le but de cette section n'est pas de démontrer de "vraies" lois du zéro-un, ni même d'énoncer des résultats originaux. Nous allons tourner un peu autour de la version la plus simple de la loi des grands nombres, pour en faire ressortir quelques aspects qui seront généralisés ou étendus par la suite.

Dans tout ce qui suit, n est un entier, et E_n désigne l'ensemble fini $\{0,1\}^n$. Les éléments de E_n seront appelés "configurations". Selon les interprétations, nous les verrons comme :
- des vecteurs $x = (x(i))_{i=1,\dots,n}$, dont les coordonnées valent 0 ou 1,
- des applications de $\{1,\dots,n\}$ dans $\{0,1\}$,
- des sous-ensembles de $\{1,\dots,n\}$,
- des vecteurs de booléens,
- des tables de vérité d'un ensemble de n variables logiques.

L'interprétation logique nous conduira à nommer "propriétés" les sous-ensembles de E_n. On munit E_n de la loi de probabilité pour laquelle les coordonnées d'une configuration sont indépendantes, chacune valant 1 avec probabilité p.

Définition 2.1 *La loi produit de n lois de Bernoulli de paramètre p, notée $\mu_{n,p}$ est définie par :*

$$\forall x = (x(i)) \in E_n\ ,\ \mu_{n,p}(x) = \prod_{i=1}^{n} p^{x(i)}(1-p)^{1-x(i)} = p^{\sum x(i)}(1-p)^{n-\sum x(i)}\ . \tag{2.1}$$

Dans le cas particulier $p = 1/2$, la probabilité de toute configuration est $1/2^n$: la loi $\mu_{n,1/2}$ est donc l'équiprobabilité sur E_n.

Soit X_n une variable aléatoire à valeurs dans E_n, de loi $\mu_{n,p}$ et S_n la variable aléatoire réelle égale à la somme des coordonnées de X_n. La loi de S_n est la loi binomiale $\mathcal{B}(n,p)$. Le résultat suivant, considéré de tout temps comme évident par les joueurs, est un des plus anciens succès de la théorie des probabilités. Il est dû à Jacques Bernoulli (1654-1705).

Théorème 2.2 *Lorsque n tend vers l'infini, S_n/n converge en probabilité vers p.*

Il s'agit bien d'une loi du zéro-un, mais sa portée est restreinte aux propriétés du type "$S_n/n \in I$", où I est un intervalle ouvert de $I\!R$.

Corollaire 2.3 *Soit I un intervalle ouvert de $I\!R$, et A_I la propriété :*

$$A_I = \{ x = (x(i)) \in E_n \, ; \, \frac{1}{n} \sum x(i) \in I \} \, . \tag{2.2}$$

Alors $\mu_{n,p}(A_I)$ converge vers 1 si I contient p, vers 0 sinon.

Quand une loi du zéro-un est vraie pour une famille de propriétés, elle reste vraie pour sa *fermeture booléenne.*

Définition 2.4 *Soit $\mathcal{A}$ une famille de propriétés. La* fermeture booléenne *de $\mathcal{A}$ est la famille $\widetilde{\mathcal{A}}$ de propriétés définie par :*

- $\mathcal{A} \subset \widetilde{\mathcal{A}}$,
- *Pour tout* $A \in \widetilde{\mathcal{A}}$, $\neg A \in \widetilde{\mathcal{A}}$,
- *Pour tout* $A, B \in \widetilde{\mathcal{A}}$, $A \wedge B \in \widetilde{\mathcal{A}}$.

La fermeture booléenne de $\mathcal{A}$ contient toutes les combinaisons logiques d'un nombre fini de propriétés de $\mathcal{A}$ par "non" ($\neg$), "et" ($\wedge$), "ou" ($\vee$).

Lemme 2.5 *Soit $\mathcal{A}$ une famille de propriétés et $\widetilde{\mathcal{A}}$ sa fermeture booléenne. Si la probabilité de toute propriété dans $\mathcal{A}$ tend vers 0 ou 1, il en est de même pour $\widetilde{\mathcal{A}}$.*

Démonstration : Cela se vérifie immédiatement en utilisant les propriétés élémentaires suivantes des probabilités :

- $Prob[\neg A] = 1 - Prob[A]$,
- $Prob[A] + Prob[B] - 1 \le Prob[A \wedge B] \le \min\{Prob[A], Prob[B]\}$.

□

Le lemme 2.5 permet de faire mousser immédiatement un résultat tel que le corollaire 2.3. Nous ne nous attarderons pas à caractériser la fermeture booléenne de l'ensemble des propositions A_I, car son pouvoir d'expression est plutôt faible : nous obtiendrons en 3.2 des lois du zéro-un beaucoup plus générales en logique du premier ordre.

Nous considérons maintenant la fonction de survie de S_n, à savoir les probabilités d'événements du type $S_n > k$. Notons A_k la proposition correspondante :

$$A_k = \{ x = (x(i)) \in E_n \, ; \, \sum x(i) > k \} \, . \tag{2.3}$$

Vue comme fonction de p, la probabilité $\mu_{n,p}(A_k)$ est croissante, proche de 0 tant que p est inférieur à k/n, proche de 1 après : la figure 1 illustre le cas $n = 100, k = 50$. Le fait que $\mu_{n,p}(A_k)$ soit une fonction croissante de p mérite justification. Le lemme 2.7 ci-dessous le montre pour toutes les *propriétés croissantes.*

Définition 2.6 *Soit $A \subset E_n$ une propriété. On dit que A est* croissante *si sa fonction indicatrice est une fonction croissante pour l'ordre de E_n défini coordonnée par coordonnée.*

$$(\forall i, x(i) \le y(i) \text{ et } x \in A) \Longrightarrow (y \in A) \, .$$

En d'autres termes, si une propriété croissante est vraie pour une configuration, elle reste vraie quand on remplace des 0 par des 1. Pour tout k, la propriété A_k définie par (2.3) est croissante, et nous rencontrerons bien d'autres exemples dans les paragraphes suivants.

Lemme 2.7 *Soit A une propriété croissante, alors $\mu_{n,p}(A)$ est une fonction croissante de p.*

Nous donnons la démonstration par couplage pour deux raisons. D'une part elle fournit une interprétation dynamique (sous forme de processus) de $\mu_{n,p}$, interprétation utile pour les graphes aléatoires aussi bien que pour les images, et très proche de la simulation. D'autre part, elle ramène les propriétés

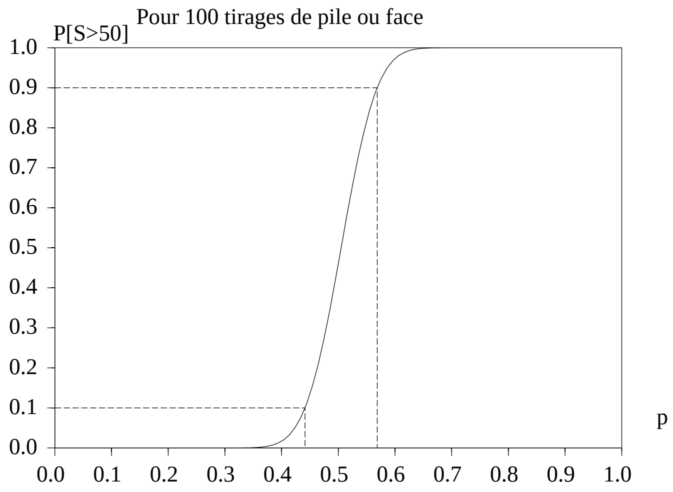

FIG. 1 – Probabilité d'obtenir plus de la moitié de piles sur 100 tirages à pile ou face, en fonction de la probabilité de pile.

de $\mu_{n,p}$ à celles de la mesure de Lebesgue sur l'hypercube $[0,1]^n$, indiquant ainsi que les phénomènes que nous étudions ne sont pas limités à E_n.

Démonstration : Soit $U = (U(i))_{i=1,\dots,n}$ un échantillon de la loi uniforme sur $[0,1]$. Les coordonnées $U(i)$ sont des variables aléatoires indépendantes et identiquement distribuées (n appels de Random successifs). La loi du vecteur U est la loi uniforme sur $[0,1]^n$. Pour tout $i = 1,\dots,n$ et pour tout $p \in [0,1]$, posons $X_p = (X_p(i))$, où $X_p(i) = \mathbb{1}_{[0,p]}(U(i))$ (1 si $U(i) \leq p$, 0 sinon). Les trajectoires du processus $\{X_p\,,\ 0 \leq p \leq 1\}$ sont croissantes au sens de l'ordre coordonnées par coordonnées : si A est une propriété croissante, vraie pour X_p avec un certain p, alors elle est aussi vraie pour $X_{p'}$, pour tout $p' \geq p$. Mais bien sûr, pour tout $p \in [0,1]$, la loi de X_p est $\mu_{n,p}$. D'où le résultat. □

La probabilité $\mu_{n,p}(A_k)$ est donc bien une fonction croissante de p, et si k est strictement compris entre 0 et n, elle vaut 0 pour $p = 0$ et 1 pour $p = 1$. L'intervalle de valeurs de p sur lequel s'effectue la transition de 0 à 1 s'appelle le *seuil*. Prenons d'abord $k = 0$ (ce qui en sera dit vaut pour toute valeur fixée de k). La probabilité de A_0 vaut $1-(1-p)^n$, qui tend vers 1 si p est constante. Pour $p = p(n)$ tendant vers 0, $(1-\mu_{n,p(n)}(A_0))$ est équivalent à e^{-np}. Donc $\mu_{n,p(n)}(A_0)$ tendra vers 0 pour $p(n)$ petit devant $1/n$, vers 1 pour $p(n)$ grand devant $1/n$. On dit que $1/n$ est une *fonction seuil* pour la propriété A_0.

Définition 2.8 *Soit A une propriété croissante. On dit que $s(n)$ est une* fonction seuil *pour la propriété A si :*

$$\lim_{n\to\infty} \frac{p(n)}{s(n)} = 0 \Longrightarrow \lim_{n\to\infty} \mu_{n,p(n)}(A) = 0\ ,$$

et :

$$\lim_{n\to\infty} \frac{p(n)}{s(n)} = +\infty \Longrightarrow \lim_{n\to\infty} \mu_{n,p(n)}(A) = 1\ .$$

Notons qu'une fonction seuil n'est qu'un ordre de grandeur : si $s(n)$ est une fonction seuil, alors $k(n)s(n)$ est encore fonction seuil de la même propriété, pourvu que $k(n)$ reste entre deux constantes strictement positives. La fonction seuil donne l'échelle de localisation du seuil. L'échelle des fluctuations est la *largeur du seuil.*

Définition 2.9 *Soit A une propriété croissante. Pour tout $\alpha \in [0,1]$, notons p_α la valeur de p telle que $\mu_{n,p_\alpha}(A) = \alpha$. On dit que la propriété A a un* seuil de largeur *$\ell(n)$ si pour tout $\varepsilon \in]0,1/2[$, on a :*

$$p_{1-\varepsilon} - p_\varepsilon = O(\ell(n))\ .$$

La figure 1 illustre le seuil de A_k pour $n = 100$, $k = 50$ et $\varepsilon = 0.1$. Comme la fonction seuil, $\ell(n)$ n'est qu'un ordre de grandeur. Nous commettrons désormais l'abus de langage consistant à parler de *la* fonction seuil ou de *la* largeur du seuil.

Nous avons vu que pour $p = p(n)$ tendant vers 0, $\mu_{n,p(n)}(A_0)$ est équivalent à e^{-np}. Avec la notation de la définition ci-dessus, p_α est équivalent à $\log(1/\alpha)/n$. La largeur du seuil est donc $1/n$.

Nous cherchons maintenant à borner la largeur du seuil pour la propriété A_k, indépendamment de k. La loi de S_n a pour espérance np et pour écart-type $\sqrt{np(1-p)}$. Pour préciser le comportement de S_n à l'échelle des fluctuations, nous utiliserons le résultat suivant qui se déduit des inégalités de Chernov ou de Hoeffding (cf. 4.1).

Théorème 2.10 *Pour tout n, p et pour tout $c > 0$ on a :*

$$\mu_{n,p}(S_n - np > c\sqrt{n}) \leq \exp(-2c^2) \ , \tag{2.4}$$

$$\mu_{n,p}(S_n - np < -c\sqrt{n}) \leq \exp(-2c^2) \ , \tag{2.5}$$

et donc :

$$\mu_{n,p}(|S_n - np| > c\sqrt{n}) \leq 2\exp(-2c^2) \ . \tag{2.6}$$

Le principal avantage de ces inégalités est que le majorant ne dépend ni de n ni de p. En revanche, il ne faut pas en attendre une évaluation précise. Par exemple, pour $n = 10000$ et $p = 0.5$, S_n est compris entre 4900 et 5100 avec une probabilité de 0.9545. L'inégalité (2.6) dit que cette probabilité est supérieure à 0.7293. Pour p constant, (2.6) donne le bon ordre de grandeur, $O(\sqrt{n})$, pour les fluctuations de S_n. Mais nous avons vu que pour $p = p(n) = O(1/n)$, l'ordre de grandeur des fluctuations de S_n est $O(1/n)$. L'obtention d'inégalités précises, plus générales que (2.6) est un domaine très important de la théorie des probabilités (voir par exemple [91, 92]). Nous y reviendrons dans la partie 4.

Le théorème 2.10 va nous permettre d'obtenir très simplement un contrôle indépendant de k, pour les largeurs de seuil de toutes les propriétés A_k. Dans la proposition 2.11, n et k sont quelconques, et p_α désigne la valeur de p pour laquelle $\mu_{n,p}(A_k)$ vaut α (cf. définition 2.9).

Proposition 2.11 *Pour tout $\varepsilon \in]0, 1/2[$, on a :*

$$p_{1-\varepsilon} - p_\varepsilon \leq \frac{\sqrt{2\log(1/\varepsilon)}}{\sqrt{n}} \ .$$

Démonstration : Fixons ε entre 0 et 1/2, et posons $c = \sqrt{\frac{\log(1/\varepsilon)}{2}}$, de sorte que $\exp(-2c^2) = \varepsilon$. Si p_ε est tel que $\mu_{n,p_\varepsilon}(A_k) = \varepsilon$, alors k ne peut pas être trop loin de np_ε. L'inégalité (2.4) permet d'affirmer que k est inférieur à $np_\varepsilon + \sqrt{n\log(1/\varepsilon)/2}$. Posons alors $q = p_\varepsilon + 2\sqrt{\log(1/\varepsilon)/(2n)}$. Par l'inégalité (2.5) appliquée à nq, on obtient $\mu_{n,q}(A_k) \geq 1 - \varepsilon$, et donc $q \geq p_{1-\varepsilon}$, d'où le résultat. □

Pour faire mousser la proposition 2.11, nous avons besoin d'une définition qui ne prendra véritablement sa force qu'en 3.3.

Définition 2.12 *Soit $A \subset E_n$ une propriété. On dit que A est* symétrique *si A est invariante par l'action d'un groupe G de permutations des coordonnées agissant transitivement sur $\{1, \ldots, n\}$.*

$$\forall i, j = 1, \ldots, n \ , \quad \exists \sigma \in G \ , \ \sigma(i) = j \ ,$$

et :

$$\forall \sigma \in G \ , \quad (x(i)) \in A \Longrightarrow (x(\sigma(i))) \in A \ .$$

On dit que A est totalement symétrique *si elle est invariante par toute permutation des coordonnées ($G = \mathcal{S}_n$).*

Corollaire 2.13 *Soit A une propriété croissante et totalement symétrique. Pour tout $\alpha \in [0,1]$, notons p_α la valeur de p telle que $\mu_{n,p_\alpha}(A) = \alpha$. Pour tout $\varepsilon \in]0,1/2[$, on a :*

$$p_{1-\varepsilon} - p_\varepsilon \le \frac{\sqrt{2\log(1/\varepsilon)}}{\sqrt{n}} .$$

Contrairement aux apparences, ce résultat ne contient rien de plus que la proposition 2.11. En effet si A est une propriété croissante et totalement symétrique, alors il existe un entier k tel que $A = A_k$. Son seul intérêt est d'introduire le théorème 3.16 (loi du zéro-un de Friedgut et Kalai), qui porte sur les propriétés croissantes et symétriques.

Nous terminons cette section par quelques exemples.

Exemple 1 *Stabilité*
Considérons d'abord la propriété S definie comme suit.

Définition 2.14 *On munit $\{1,\dots,n\}$ de la structure de graphe cyclique non orienté dont l'ensemble d'arêtes est :*

$$C_n = \Big\{ \{1,n\}, \{i,i+1\}, i = 1,\dots,n-1 \Big\} .$$

Une configuration, identifiée à un sous-ensemble de $\{1,\dots,n\}$ est dite stable *si elle est un sous-ensemble stable pour le graphe cyclique :*

$$x = (x(i)) \in S \iff (\forall \{i,j\} \in C_n\,,\ x(i) = 1 \Longrightarrow x(j) = 0) . \tag{2.7}$$

En d'autres termes, une configuration est stable si elle ne comporte pas deux 1 voisins. Nous notons S la propriété de stabilité. On peut calculer la fonction génératrice des ensembles stables pour de nombreux types de graphes (voir [42]). Dans le cas du graphe cyclique à n sommets, elle est particulièrement simple. Notons $s_{k,n}$ le nombre de stables à k sommets dans le graphe cyclique à n sommets. On a :

$$g_n(z) = \sum_{k=0}^{n} s_{k,n} z^k = \left(\frac{1-\sqrt{1+4z}}{2}\right)^n + \left(\frac{1+\sqrt{1+4z}}{2}\right)^n .$$

Il est facile d'en déduire l'expression explicite de $\mu_{n,p}(S)$.

$$\mu_{n,p}(S) = (1-p)^n g_n\left(\frac{p}{1-p}\right) .$$

La propriété S est invariante par les permutations cycliques de $\{1,\dots,n\}$. Sa négation est croissante, et donc $\mu_{n,p}(S)$ est fonction décroissante de p. Pour toute valeur strictement positive de p, $\mu_{n,p}(S)$ tend vers 0 quand n tend vers l'infini. Pour $p = p(n)$ tendant vers 0, on vérifie facilement que $\mu_{n,p}(S)$ est équivalent à $\exp(-np^2)$. La fonction seuil de S est $1/\sqrt{n}$. La largeur de seuil est aussi $1/\sqrt{n}$, bien que S ne soit pas totalement symétrique.

Exemple 2 *Longueur de runs*
Afin d'étendre l'exemple 1, notons M_n la variable aléatoire égale au nombre maximal de 1 consécutifs dans la configuration aléatoire X_n (toujours au sens de la structure de graphe cyclique C_n). Par rapport à la propriété de stabilité S de l'exemple précédent, on a :

$$X_n \in S \iff M_n \le 1 .$$

La distribution des "runs" de 1 consécutifs dans une configuration aléatoire cyclique a souvent été étudiée (voir par exemple [49]). On démontre que la variable aléatoire M_n est localisée en $O(\log(n))$, avec des fluctuations d'ordre $O(1)$, sans qu'il y ait convergence en loi à l'échelle des fluctuations. Le résultat suivant, dû à Erdös et Révész [34], est démontré dans [76] (voir aussi [52, 69] et [59] pour une généralisation).

Théorème 2.15 *Posons :*

$$a(n) = \left\lfloor \frac{\log(n)}{\log(1/p)} \right\rfloor \quad \textit{et} \quad b(n) = \frac{\log(n)}{\log(1/p)} - \left\lfloor \frac{\log(n)}{\log(1/p)} \right\rfloor \ ,$$

où $\lfloor \cdot \rfloor$ *désigne la partie entière. Notons* R_k *la propriété :*

$$R_k = \text{“}M_n - a(n) \leq k\text{”}.$$

Quand n *tend vers l'infini :*

$$\mu_{n,p}(R_k) = \exp(-p^{k-b(n)}) + o(1) \ .$$

Exemple 3 *Parité*
Nous passons maintenant à des exemples de propriétés dont la probabilité ne converge pas vers 0 ou 1. La plus souvent citée est “n est pair”, dont la probabilité vaut alternativement 0 ou 1, indépendamment de p. Considérons plutôt :

$$A = \{ \, x = (x(i)) \, , \, \sum_{i=1}^{n} x(i) \text{ est pair } \} \ .$$

On obtient aisément sa probabilité :

$$\mu_{n,p}(A) = \frac{1}{2}(1 + (1-2p)^n) \ .$$

Elle tend vers $1/2$ quand n tend vers l'infini, pour tout $p \in]0,1[$, bien que A soit totalement symétrique. On comprend aisément comment fabriquer sur le même patron des propriétés totalement symétriques dont la probabilité converge vers un rationnel h/k : il suffit de demander que la somme des coordonnées modulo k soit comprise entre 1 et h.

Exemple 4 *Tirages binaires biaisés*
Si une propriété ne dépend que d'un nombre fixe de coordonnées parmi les n, elle ne vérifiera pas non plus de loi du zéro-un. Par exemple $A = \{x = (x(i)) \, , \, x(1) = 1\}$ a pour probabilité p, quel que soit n. Nous allons fabriquer une propriété dépendant de toutes les variables, bien que de manière non symétrique, dont la probabilité pour $\mu_{n,1/2}$ converge vers une valeur $p \in [0,1]$, arbitraire. Considérons le développement binaire de p, supposé tel qu'il ne se termine pas par une infinité de 1.

$$p = \sum_{i=1}^{\infty} \frac{a(i)}{2^i} \ .$$

Soit $a = (a(i))$ la configuration égale aux n premiers termes du développement de p. Etant donnée une configuration $x \in E_n$, on décide qu'elle appartient à A si en la première coordonnée où x et a diffèrent, a vaut 1 et x vaut 0. Par convention, $a \notin A$. On vérifie aisément que :

$$\mu_{n,1/2}(A) = \sum_{i=1}^{n} \frac{a(i)}{2^i} \ .$$

Ceci montre qu'on peut utiliser une pièce non truquée pour fabriquer des tirages biaisés, pour lesquels la probabilité de pile est p.

2.2 Graphes aléatoires

La théorie des graphes aléatoires a débuté par l'article magistral de Erdös et Rényi [33], qui en décrivait déjà les principaux résultats. Elle est traitée dans plusieurs manuels, parmi lesquels [6, 89, 101]. Le modèle que nous considérons ici est celui du graphe non orienté à n sommets pour lequel les arêtes existent indépendamment avec probabilité p.

Définition 2.16 *On appelle* graphe aléatoire, *et on note* $\mathcal{G}$ *ou* $\mathcal{G}(n,p)$, *la variable aléatoire à valeurs dans l'ensemble des graphes non orientés d'ensemble de sommets* $\{1, \ldots, n\}$, *telle que l'arête* $\{i,j\}$ *existe avec probabilité* p, *ces événements étant indépendants.*

Posons $\alpha(n) = \binom{n}{2} = n(n-1)/2$. Quitte à numéroter toutes les arêtes possibles de 1 à $\alpha(n)$, l'ensemble des arêtes $\mathcal{A}$ du graphe peut être identifié à une configuration de $\{0,1\}^{\alpha(n)}$. On note habituellement $i \sim j$ la propriété $\{i,j\} \in \mathcal{A}$ (relation de voisinage). Le modèle que nous avons défini revient à mettre la probabilité $\mu_{\alpha(n),p}$ sur $\{0,1\}^{\alpha(n)}$ (définition 2.1). Notons que ce modèle est différent de celui initialement proposé par Erdös et Rényi [33], qui considéraient l'équiprobabilité sur l'ensemble des graphes ayant un nombre fixe d'arêtes. En fait les deux modèles ont les mêmes propriétés asymptotiques : nous y reviendrons en 2.5.

Notre propos ici n'est pas de donner un aperçu, même bref, de la théorie des graphes aléatoires. Nous nous contenterons d'illustrer par trois exemples le phénomène de concentration de la mesure $\mu_{\alpha(n),p}$. Nous ne l'avons abordé en 2.1 que par la variable aléatoire S_n, qui correspond ici au nombre total d'arêtes. Deux graphes "tirés au hasard" (réalisations de $\mathcal{G}$) n'ont pas seulement des nombres d'arêtes proches : ils ont le même aspect, les mêmes propriétés.

Exemple 5 *Connexité*
Un des résultats les plus frappants de la théorie des graphes aléatoires concerne la connexité. Sa fonction seuil est $\log(n)/n$, avec une largeur de seuil égale à $1/n$. L'énoncé précis est le suivant ([101] p. 303).

Théorème 2.17 *Soit :*

$$p = p(n) = \frac{\log(n)}{n} + \frac{c}{n} + o(\frac{1}{n}) \; .$$

Alors :

$$\lim_{n\to\infty} \mu_{\alpha(n),p(n)}(\text{ Connexité }) = e^{-e^{-c}} \; .$$

Exemple 6 *Diamètre*
Une fois la connexité acquise, se pose la question du diamètre du graphe, à savoir la distance en nombre d'arêtes entre deux sommets quelconques. Considérons la propriété D_k, réalisée si tout couple de sommets est relié par un chemin contenant k arêtes :

$$\forall x,y \in \{1,\ldots,n\} \, , \; \exists z_1,\ldots,z_{k-1} \, , \; x \sim z_1, z_1 \sim z_2, \ldots, z_{k-1} \sim y \; .$$

La propriété D_k admet pour fonction seuil $n^{-\frac{k-1}{k}}(\log(n))^{\frac{1}{k}}$. En particulier pour p constant, la propriété D_2 a une probabilité proche de 1 : deux sommets quelconques non voisins dans le graphe $\mathcal{G}(n,1/2)$ ont au moins un voisin en commun.

Le monde est petit ! Le folklore scientifique dit que deux personnes quelconques sont toujours reliées par une chaîne de 5 connaissances au plus. On trouve même sur le réseau des sites dont le but est de vérifier "expérimentalement" cette loi sur les utilisateurs du web. Des asymptotiques précises sur la probabilité des propriétés D_k vont nous permettre de quantifier ce phénomène. On démontre en effet que pour $p = p(n)$ de l'ordre de $n^{-\frac{k-1}{k}}(\log(n))^{\frac{1}{k}}$, la probabilité $\mu_{\alpha(n),p}(D_k)$ est proche de $e^{-\lambda}$, avec :

$$\lambda = \frac{n^2}{2}\exp(-n^{k-1}p^k) \; .$$

(Voir [101] p. 315 pour l'énoncé précis). Comme application numérique, prenons $n = 6\,10^9$ (6 milliards d'individus sur terre), et $p = 1/(6\,10^{-7})$, de sorte que $np = 100$ (chaque individu connaît en moyenne 100 personnes). Le calcul donne :

$$\mu_{\alpha(n),p}(D_5) \simeq 10^{-10^{18}} \quad \text{et} \quad \mu_{\alpha(n),p}(D_6) \simeq 1 - 10^{-56} \; .$$

Il n'y a donc aucune chance que le diamètre soit 5, il est presque sûrement égal à 6. Voici les valeurs de np pour lesquelles $\mu_{\alpha(n),p}(D_5)$ vaut respectivement 0.01, 0.5 et 0.99.

np	191.4	193.1	196.6
$\mu_{\alpha(n),p}(D_5)$	0.01	0.50	0.99

Les trois valeurs sont étonnamment proches : la propriété D_5 a un seuil très étroit.

Exemple 7 *Nombre clique*
Voici une autre illustration spectaculaire de la concentration de la mesure dans les graphes aléatoires. Le nombre clique d'un graphe G, noté $\omega(G)$, est le nombre de sommets de la plus grande clique (sous-graphe complet) de G. Le théorème de concentration, démontré indépendamment par Bollobás, Erdös et Matula en 76 (voir [6], p. 251), affirme que pour p fixé, le nombre clique du graphe aléatoire $\mathcal{G}(n,p)$ ne peut prendre que deux valeurs au plus. Nous donnons l'énoncé pour $p = 1/2$.

Théorème 2.18 *Il existe une fonction k, de $\mathbb{N}$ dans $\mathbb{N}$, telle que*

$$k(n) \sim 2\log(n)/\log(2) \ ,$$

et :

$$\lim_{n\to\infty} \mu_{\alpha(n),1/2}(\,\omega(\mathcal{G}) = k(n) \ \textit{ou}\ k(n)+1\,) = 1 \ .$$

En fait, pour la plupart des valeurs de n, le nombre clique est concentré non pas sur deux, mais sur une valeur seulement !

2.3 Images aléatoires

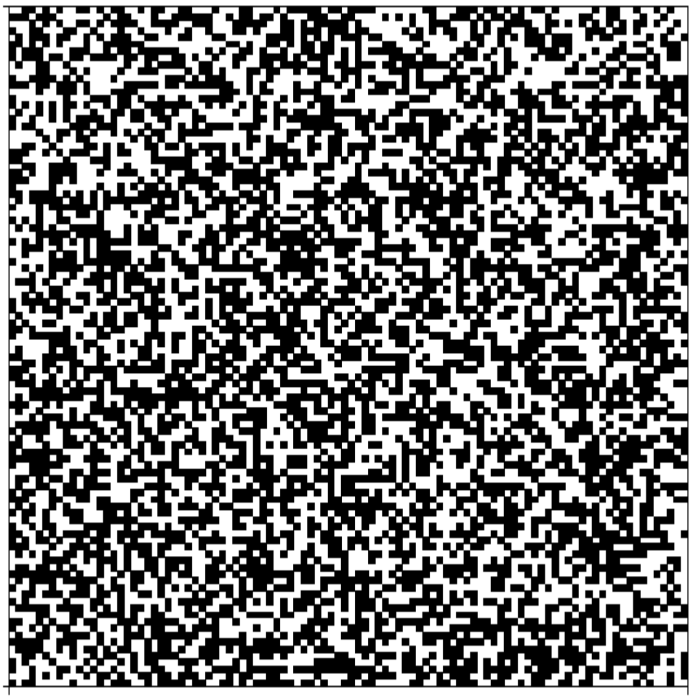

Fig. 2 – Image aléatoire de 100×100 pixels indépendants, noirs ou blancs avec probabilité 0.5.

Nous utiliserons ici les images aléatoires comme un autre moyen de visualiser la concentration de la mesure dans un espace produit. Nous ne considérons que des images carrées $n \times n$, binaires, pour lesquelles les pixels sont noirs avec probabilité p ou blancs avec probabilité $1-p$, indépendamment les uns des autres.

Définition 2.19 *On appelle* image aléatoire*, et on note $\mathcal{I}$, ou $\mathcal{I}(n,p)$, la matrice aléatoire de taille $n \times n$, dont les coefficients, indépendants, valent 1 avec probabilité p et 0 avec probabilité $1-p$.*

C'est une nouvelle variante du modèle de Bernoulli : l'espace est identifiable à $\{0,1\}^{n^2}$ et la loi d'une image aléatoire est $\mu_{n^2,p}$ (définition 2.1).

L'idée des images aléatoires est de fournir une base quantitative à l'intuition qui consiste à dire que l'œil voit dans une image ce qui est inhabituel, qui tranche sur un bruit de fond, bref, qui est improbable. Les algorithmes de détection cherchent donc les groupes de pixels qui seraient de probabilité faible dans une image aléatoire (voir [22]).

La figure 2 représente une image aléatoire pour $n = 100$ et $p = 1/2$. Comme déjà observé en 2.1, le nombre total de pixels noirs est concentré autour de 5000. Mais cette concentration est aussi vraie pour des sous-images. Divisons l'image observée en 100 sous-images de taille 10×10. La probabilité que *chacune* des 100 sous-images ait un nombre de pixels noirs compris entre 30 et 70 est supérieure à 0.99.

Exemple 8 *Percolation*
Une autre manière d'aborder le même modèle est la *percolation de sites* (voir [54] p. 24). En percolation on considère plutôt un réseau infini, $\mathbb{Z}^2$ dans notre cas, dans lequel les sommets (pixels) sont noirs avec probabilité p. On s'intéresse aux composantes connexes de pixels noirs. Un résultat fondamental est l'existence d'une probabilité critique p_c, dont la valeur dans le cas de $\mathbb{Z}^2$ n'est toujours pas connue exactement. Pour $p < p_c$, les composantes connexes sont toutes finies et la distribution de leurs tailles admet des moments de tous ordres. Pour $p > p_c$, il existe presque sûrement une composante connexe infinie unique. Ce phénomène est à rapprocher de la composante connexe géante des graphes aléatoires ([6] p. 131). Pour établir le lien avec les images aléatoires, introduisons la propriété C : “il existe un chemin formé uniquement de pixels noirs voisins, joignant le bord gauche au bord droit de l'image.”. On peut montrer ([110]) que $\mu_{n^2,p}(C)$ tend vers 0 si $p < p_c$, vers 1 si $p > p_c$. Une conjecture généralement admise est que $\mu_{n^2,p_c}(C)$ tend vers 1/2.

Exemple 9 *Sous-images*
Nous nous intéresserons ensuite à l'apparition de sous-images. Soit V une image *fixée* de taille $m \times m$, que nous appellerons “vignette” pour la distinguer de l'image complète (par exemple la croix de la figure 3).

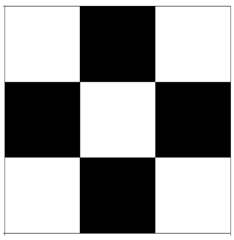

FIG. 3 – Vignette de 3×3 pixels. L'image aléatoire de la figure 2 en contient plusieurs copies, incluses et exactes.

Pour $n \geq m$, nous nous intéressons aux copies de V contenues dans l'image $\mathcal{I}$. Afin de simplifier les notations en évitant les problèmes de bord, nous supposerons que le réseau carré portant les pixels a une condition de bord *périodique*. De sorte que les indices apparaissant dans la définition suivante sont ajoutés entre eux *modulo n*.

Définition 2.20 *Soit $I = (x(i,j))\,,\, i,j = 1,\ldots,n$ une image $n \times n$. Pour tout $i,j = 1,\ldots,n$, notons :*

$$I_{i,j} = (x(h,k))\,,\; h = i,\ldots,i+m-1\,,\; k = j,\ldots,j+m-1\,,$$

la sous-image extraite de I, de taille $m \times m$, dont le coin supérieur gauche est le pixel de coordonnées (i,j).

1. *On appelle* vignette incluse *la propriété :*

$$VI = \{I \,;\, \exists (i,j)\,,\, I_{i,j} \geq V\,\}\ .$$

2. *On appelle* vignette exacte *la propriété :*

$$VE = \{I \,;\, \exists (i,j)\,,\, I_{i,j} = V\,\}\ .$$

Dans la définition de VI, l'ordre est l'ordre induit sur E_{n^2} composante par composante : une image vérifie VI si elle contient une sous-image qui a au moins les mêmes pixels noirs que V. Elle vérifie VE si elle contient une copie exacte de V. Les deux propriétés sont invariantes par le groupe des transformations du réseau des pixels (tore). La propriété VI a évidemment beaucoup moins d'intérêt que VE pour les applications en image. Elle présente l'avantage d'être croissante. Pour une valeur fixée de p, strictement comprise entre 0 et 1, $\mu_{n^2,p}(VI)$ et $\mu_{n^2,p}(VE)$ tendent vers 1. Toute vignette finit par apparaître dans une image aléatoire suffisamment grande. C'est la version planaire du paradoxe "du singe et de la machine à écrire". Une étude précise des probabilités de VI et VE pour p proche de 0 et 1 a été réalisée par Coupier [17]. Il démontre en particulier le résultat d'approximation poissonnienne suivant :

Proposition 2.21 *Soit b le nombre de pixels noirs de la vignette V. Posons $p = p(n) = cn^{-2/b}$. Alors :*

$$\lim_{n\to\infty} \mu_{n^2,p}(VI) = \lim_{n\to\infty} \mu_{n^2,p}(VE) = 1 - \exp\left(-\frac{8c^b}{a}\right)\ ,$$

où a est le nombre de transformations géométriques (symétries ou rotations) qui laissent V invariante.

Ce type d'approximation est bien connu dans le cadre des graphes aléatoires, où l'analogue de la propriété VI est l'apparition d'un sous-graphe (voir [101], p. 309). On déduit immédiatement de la proposition 2.21 que VI admet pour fonction seuil et largeur de seuil la même fonction $n^{-2/b}$.

2.4 La k-satisfiabilité

Le problème de la k-satisfiabilité, archétype des problèmes NP-complets (pour $k \geq 3$: voir [18]), a suscité une immense littérature, qu'il est hors de propos de résumer ici. Ce qui suit est basé sur les présentations de Monasson et al. [87] et Friedgut [43] (voir aussi [13, 51, 61, 66, 85, 86]).

On se donne un ensemble de n variables logiques, $X_1, \ldots, X_n$. Une *clause* de longueur k est une disjonction de k de ces variables, affirmées ou niées. Par exemple, $X_{i_1} \vee \neg X_{i_2} \vee X_{i_3}$ est une clause de longueur 3. Il y a $\binom{n}{k}$ ensembles de k variables possibles, et une fois choisies les k variables, chacune peut apparaître affirmée ou niée dans la clause. Il est donc possible de former $\alpha(n,k) = 2^k\binom{n}{k}$ clauses de longueur k sur n variables. Notons $C_{n,k}$ l'ensemble des clauses. Choisissons maintenant un sous-ensemble $\{c_1, \ldots, c_m\}$ de $C_{n,k}$, que nous appellerons *formule*. Cette formule est dite *satisfiable* s'il existe une affectation booléenne de $X_1, \ldots, X_n$ telles que $c_1, \ldots, c_m$ soient simultanément vraies. La satisfiabilité est une propriété, que nous identifions à un sous-ensemble de l'ensemble des formules (elles-mêmes ensembles de clauses !). Une formule peut être vue comme une configuration dans $E_{n,k} = \{0,1\}^{\alpha(n,k)}$. La non-satisfiabilité est clairement une propriété croissante, au sens de la définition 2.6 : plus on rajoute de clauses, plus il devient difficile de les satisfaire simultanément.

Le problème est de définir une loi de probabilité sur $E_{n,k}$. Comme pour les graphes aléatoires, avec le modèle $\mathcal{G}(n,p)$ et celui d'Erdös-Rényi, l'alternative consiste à fixer le nombre total M de clauses dans la configuration, ou bien à les affecter chacune d'une probabilité p. Le premier modèle est le plus classique (voir [87]). Une fois fixé le nombre total de clauses M, chaque configuration sera choisie avec la même probabilité $\binom{\alpha(n,k)}{M}^{-1}$. Nous préfèrerons le second modèle, introduit par Friedgut [43], qui est plus proche de la présentation que nous avons suivie jusqu'ici, tout en étant asymptotiquement équivalent au modèle classique (voir 2.5).

Définition 2.22 *On appelle formule aléatoire, et on note $\mathcal{F}(n,p)$, une variable aléatoire à valeurs dans $E_{n,k}$, telle que chaque clause apparaît dans la formule avec probabilité p, indépendamment des autres.*

La loi de $\mathcal{F}(n,p)$ est donc $\mu_{\alpha(n,k),p}$ (définition 2.1). De très nombreux articles ont été consacrés à l'étude du comportement asymptotique de $\mu_{\alpha(n,k),p}(S)$ en fonction de k, et relativement peu de conjectures ont pu être démontrées à ce jour.

Le cas de la 1-satisfiabilité ne pose pas problème. Si toutes les clauses sont de longueur 1, chacune est constituée d'une seule variable, affirmée ou niée. Une formule est satisfiable si et seulement si elle ne contient pas à la fois une variable et sa négation. On a donc :

$$\mu_{\alpha(n,1),p}(S) = (1-p^2)^n \ .$$

Cette probabilité tend vers 0, pour toute valeur de p fixée strictement positive. Pour $p = p(n)$ tendant vers 0, on a :

$$\mu_{\alpha(n,1),p}(S) \sim e^{-np^2} \ .$$

La propriété S admet donc $1/\sqrt{n}$ pour fonction seuil et pour largeur de seuil, comme la stabilité de l'exemple 1.

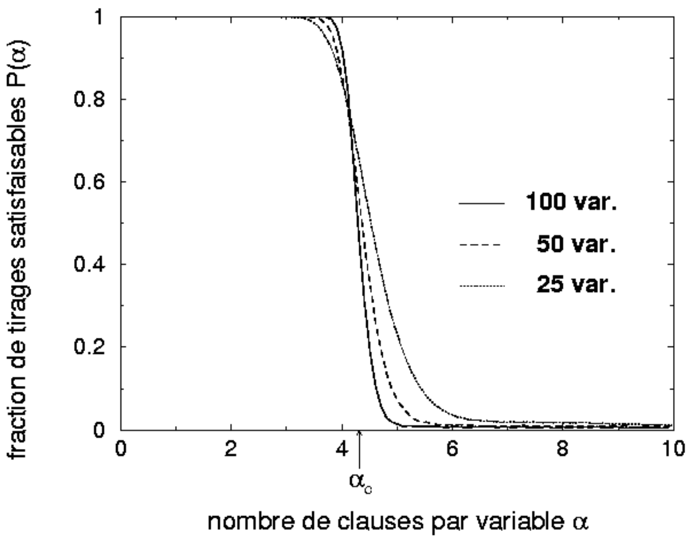

Fig. 4 – Résultats expérimentaux montrant le phénomène de seuil pour la 3-satisfiabilité : figure extraite de [13].

A partir de $k = 2$, la k-satisfiabilité est beaucoup plus difficile à étudier. Un raisonnement intuitif montre que la probabilité de S doit être forte si le nombre de clauses est faible devant le nombre de variables. En effet dans ce cas, il y a de fortes chances pour que les clauses fassent intervenir des variables toutes différentes, ce qui entraîne trivialement la satisfiabilité. Inversement, si le rapport du nombre de clauses au nombre de variables tend vers l'infini, alors le problème sera largement surcontraint, et la probabilité de satisfiabilité sera faible. Or pour $p = p(n)$ donné, le nombre de clauses dans la formule aléatoire $\mathcal{F}(n,p)$ a pour espérance $\alpha(n,k)p(n)$, et pour k donné, $\alpha(n,k)$ est d'ordre $O(n^k)$. Il en découle que la fonction seuil pour la propriété S est n^{1-k}. Le résultat suivant, dû à Friedgut [43], montre que la largeur du seuil est strictement plus petite que la fonction seuil.

Théorème 2.23 *Pour tout $k \geq 2$, il existe une fonction $c_k(n)$, telle que pour tout $\varepsilon > 0$:*

$$\lim_{n\to\infty} \mu_{\alpha(n,k),p(n)}(S) = \begin{cases} 1 & si\ p(n) = (c_k(n) - \varepsilon)n^{1-k} \ , \\ 0 & si\ p(n) = (c_k(n) + \varepsilon)n^{1-k} \ . \end{cases}$$

Pour $k = 2$, il a été démontré ([50, 109]) qu'on peut prendre $c_2(n) \equiv 1$ dans le théorème 2.23. Les spécialistes (voir [27, 28, 87]) s'accordent à penser que $c_k(n)$ peut être considérée comme constante pour toute valeur de k, ce qui est confirmé par de nombreux résultats de simulation (voir [87, 66] et [13], d'où est extraite la figure 4, fournie par R. Monasson).

Signalons pour finir, deux variantes du problème de la k-satisfiabilité. La première, développée par Monasson et al. dans [87], consiste à mélanger dans une formule une certaine proportion de clauses de longueur 3 avec des clauses de longueur 2. L'autre, étudiée par Creignou et Daudé [19], consiste à étudier la satisfiabilité pour des clauses contruites à partir du "ou exclusif" (XOR).

2.5 Sortir de $\{0,1\}^n$

Dans les lois du zéro-un que l'on trouve dans les livres d'informatique ou de logique (par exemple [3, 30, 71]), il n'est pas question d'indépendance. Elles n'ont de rapport avec les probabilités que par la combinatoire, car elles portent sur des structures finies munies de l'équiprobabilité. Dans ce cas, il s'agit de savoir si le nombre d'objets vérifiant une propriété est équivalent au nombre total (probabilité proche de 1), ou bien négligeable (probabilité proche de 0).

A propos de la loi $\mu_{n,p}$ (définition 2.1), nous avions déjà observé que l'équiprobabilité sur $\{0,1\}^n$ correspondait au cas particulier $p = 1/2$. Ceci s'étend à d'autres espaces produits. Par exemple, le produit de n copies de la loi uniforme sur $[0,1]$ est la loi uniforme sur l'hypercube $[0,1]^n$. Mais inversement, quels modèles d'équiprobabilité pourraient remplacer $\mu_{n,p}$ pour $p \neq 1/2$? Nous en avons déjà rencontré deux à propos des graphes aléatoires (graphes à nombres de sommets et d'arêtes fixés) ainsi que de la k-statisfiabilité (formules à nombres de variables et de clauses fixés). Dans les deux cas, nous avions annoncé sans justification que l'équiprobabilité était proche d'une loi produit. Il n'est pas facile de donner un énoncé rigoureux totalement convaincant. Commençons par définir le modèle d'équiprobabilité dans le cadre de $\{0,1\}^n$.

Définition 2.24 *Soit m un entier fixé ($0 \leq m \leq n$). Notons $\nu_{n,m}$ l'équiprobabilité sur l'ensemble des configurations de taille n, dont la somme des coordonnées vaut m. Elle est définie par :*

$$\nu_{n,m}(x) = \begin{cases} \binom{n}{m}^{-1} & si\ \sum x(i) = m \ , \\ 0 & sinon. \end{cases}$$

Sous la loi $\mu_{n,p}$, la somme est concentrée autour de np (théorème 2.2). Il est donc raisonnable d'approcher $\mu_{n,p}$ par la loi $\nu_{n,m(n)}$, avec $m(n) \sim np$. Le résultat suivant est une simple extension du théorème classique de convergence de la loi hypergéométrique vers la loi binomiale. Il justifie l'approximation ci-dessus, pour une valeur fixe de p.

Proposition 2.25 *Soit $z \in E_k$ un mot binaire fixé de longueur k. Notons $h = \sum z(i)$ son nombre de 1 et $A_z \subset E_n$ la propriété :*

$$A_z = \{\, x \in E_n \ , \ (x(1), \dots, x(k)) = z \,\} \ .$$

Les probabilités de A_z sous $\nu_{n,m}$ et $\mu_{n,p}$ sont :

$$\nu_{n,m}(A_z) = \frac{\binom{n-k}{m-h}}{\binom{n}{m}} \quad et \quad \mu_{n,p}(A_z) = \mu_{k,p}(A_z) \ .$$

Si $m = m(n)$ est tel que $m(n) \sim np$, alors :

$$\lim_{n\to\infty} \nu_{n,m(n)}(A_z) = \mu_{k,p}(A_z) \ .$$

Mais le cas où p est fixe n'est pas le cas le plus intéressant. On trouve dans Bollobás [6] (théorème 3 p. 36) le résultat suivant qui fournit un encadrement pour les probabilités de propriétés croissantes.

Proposition 2.26 *Soient $p_1(n)$, $p_2(n)$ et $m(n)$ trois fonctions telles que :*

1. $0 < p_1(n) < p_2(n) < 1$,
2. $m(n) \in \mathbb{N}$,
3. $$\lim_{n\to\infty} \frac{m(n) - np_1(n)}{\sqrt{n\,p_1(n)(1-p_1(n))}} = \lim_{n\to\infty} \frac{np_2(n) - m(n)}{\sqrt{n\,p_2(n)(1-p_2(n))}} = \infty \ .$$

Soit A une propriété croissante (définition 2.6). Alors :

$$\mu_{n,p_1(n)}(A) + o(1) \leq \nu_{n,m(n)}(A) \leq \mu_{n,p_2(n)}(A) + o(1) \ .$$

On peut déduire ce résultat du lemme 2.7 et du théorème 2.10. Notons qu'il ne couvre toujours pas les cas où $p_1(n)$ et $p_2(n)$ sont très proches de 0. La raison pour laquelle il est plus facile de travailler avec $\mu_{n,p}$ qu'avec $\nu_{n,m}$ est l'indépendance des coordonnées, et les calculs explicites qu'elle permet (formule (2.1)).

Au delà des graphes, des images et de la k-satisfiabilité, de nombreuses structures combinatoires peuvent se ramener à des variantes ou des approximations de $\{0,1\}^n$ muni de la loi $\mu_{n,p}$: chemins de Dyck, arbres, modèles d'urnes [26, 47, 63], modèles de bases de données [2, 3] etc... Mais l'intuition sous-jacente aux lois du zéro-un ne leur est pas limitée : on souhaiterait pouvoir dire que deux objets combinatoires de grande taille tirés au hasard se ressemblent toujours, qu'il s'agisse d'arbres, de permutations, de tries ou de forêts (cf. [53, 111]). Nous ne sommes pas en mesure de faire mieux qu'indiquer une voie de recherche pour le justifier. Ce qui suit est basé sur la combinatoire analytique de Flajolet [39] que nous remercions pour son aide.

Récemment, Duchon et al. [29] ont introduit la notion de "loi de Boltzmann" pour une structure combinatoire. Soit $\mathcal{C}$ une structure combinatoire ordinaire (on pourrait étendre le raisonnement aux structures labellisées). Nous noterons $|\cdot|$ sa fonction de taille. Pour tout n entier, l'ensemble $\mathcal{C}_n$ des objets de taille n de la structure est fini, de cardinal C_n. Nous noterons C la fonction génératrice associée :

$$C(z) = \sum_{n=0}^{\infty} C_n\, z^n \ .$$

Définition 2.27 *Soit x un réel positif tel que $C(x) < \infty$. On appelle* loi de Boltzmann *de paramètre x, la loi de probabilité qui à un objet $\gamma \in \mathcal{C}$ associe :*

$$\mu_x(\gamma) = \frac{x^{|\gamma|}}{C(x)} \ .$$

On aurait pu aussi définir cette loi en disant que la taille d'un objet tiré selon μ_x est une variable aléatoire à valeurs entières, de fonction génératrice $C(zx)/C(x)$, et la loi conditionnelle d'un objet sachant que sa taille vaut n est l'équiprobabilité sur $\mathcal{C}_n$.

Considérons maintenant une propriété $\mathcal{A}$ de la structure, à savoir un sous-ensemble de $\mathcal{C}$. Notons A_n le nombre d'objets de taille n qui possèdent la propriété $\mathcal{A}$. Nous souhaiterions pouvoir dire que le rapport A_n/C_n tend vers 0 ou 1. Ceci revient à comparer le comportement de deux fonctions génératrices au voisinage d'une singularité (voir [38, 40, 41]). Notons $A(z)$ la fonction génératrice des A_n :

$$A(z) = \sum_{n=0}^{\infty} A_n\, z^n \ .$$

La probabilité de $\mathcal{A}$ pour la loi de Boltzmann μ_x est le rapport des deux fonctions génératrices :

$$\mu_x(A) = \frac{A(x)}{C(x)} \ .$$

La proposition 2.28 donne un début de contenu à l'intuition selon laquelle la probabilité d'une propriété tend généralement vers 0 ou 1. En tant que séries entières à coefficients entiers positifs les fonctions $A(z)$ et $C(z)$ ont un rayon de convergence inférieur ou égal à 1. Nous notons r le rayon de convergence de $C(z)$, que nous supposerons non nul. Le réel r est une singularité de $C(z)$. Les équivalents au voisinage de r sont :

$$A(z) \sim \alpha_A \, (r-z)^{-\beta_A} \quad \text{et} \quad C(z) \sim \alpha_C \, (r-z)^{-\beta_C} \; .$$

Par convention, β_A peut être nul (si $A(z)$ est holomorphe en r), et β_C peut être infini (si la singularité de $C(z)$ est essentielle).

Proposition 2.28

$$\lim_{n\to\infty} \frac{A_n}{C_n} = \left\{ \begin{array}{lll} 0 & si & \beta_A < \beta_C \; , \\ \frac{\alpha_A}{\alpha_C} & si & \beta_A = \beta_C \; . \end{array} \right.$$

Prenons par exemple la structure élémentaire des mots binaires, de fonction génératrice $C(z) = (1-2z)^{-1}$. La taille est la longueur du mot. Pour la loi de Boltzmann de paramètre $x < \frac{1}{2}$, la loi de la taille est la loi géométrique de paramètre $1-2x$. Considérons la propriété de stabilité (exemple 1). La fonction génératrice $A(z)$ vaut :

$$A(z) = \left(1 - z\left(\frac{1-\sqrt{5}}{2}\right)\right)^{-1} + \left(1 - z\left(\frac{1+\sqrt{5}}{2}\right)\right)^{-1} \; .$$

Dans ce cas $r = 1/2$ et $\beta_A = 0$ (le rayon de convergence de A est strictement supérieur à r). Le rapport A_n/C_n tend vers 0 à vitesse exponentielle, ce qui n'est pas un scoop. Considérons ensuite la propriété "la somme des coordonnées du mot est paire" (exemple 3). Sa fonction génératrice est $A(z) = \frac{1}{2}((1-2z)^{-1}+1)$. Donc $\beta_A = \beta_C = 1$ et $\alpha_A/\alpha_C = 1/2$. Le rapport A_n/C_n tend vers $1/2$, comme prévu.

Il est plus intéressant de biaiser les lois de Boltzmann pour obtenir une définition paramétrée, proche des modèles que nous avons étudiés jusqu'ici. Considérons une seconde fonction à valeurs entières définie sur la structure $\mathcal{C}$. Nous la nommerons arbitrairement le *poids*, et nous la noterons w. Notons $C_{n,k}$ le nombre d'objets de la structure, de taille n et de poids k. On considère la fonction génératrice double :

$$C(z,t) = \sum_{n,k} C_{n,k} \, z^n t^k \; .$$

Soient x et y deux réels positifs tels que $C(x,y) < \infty$. Il est naturel de définir sur $\mathcal{C}$ la loi de probabilité $\mu_{x,y}$ qui à un objet $\gamma \in \mathcal{C}$ associe :

$$\mu_{x,y}(\gamma) = \frac{x^{|\gamma|} y^{w(\gamma)}}{C(x,y)} \; .$$

Sous $\mu_{x,y}$, la loi de la taille a pour fonction génératrice $C(xz,y)/C(x,y)$. La probabilité conditionnelle d'un objet sachant que sa taille vaut n n'est plus uniforme sur $\mathcal{C}_n$: elle est fonction du poids de cet objet. C'est une autre loi de Boltzmann sur $\mathcal{C}_n$, paramétrée par y. Etant donnée une propriété A, nous notons maintenant $A_{n,k}$ le nombre d'objets de taille n et de poids k qui la possèdent, et $A(z,t)$ la fonction génératrice double correspondante. La probabilité $\mu_{x,y}(A)$ est le rapport $A(x,y)/C(x,y)$. Dans l'optique des lois du zéro-un, on souhaite montrer que la probabilité conditionnelle de $\mathcal{A}$ pour les objets de taille n tend vers 0 ou 1. La loi conditionnelle d'un objet sachant que sa taille vaut n sera notée $\mu_{n,y}$. La probabilité $\mu_{n,y}(A)$ est le rapport des deux fonctions de y, coefficients de x^n dans les fonctions $A(x,y)$ et $C(x,y)$:

$$\mu_{n,y}(\mathcal{A}) = \frac{[A(x,y)]_{x,n}}{[C(x,y)]_{x,n}} \; .$$

Pour savoir si $\mu_{n,y}(\mathcal{A})$ converge, il faut étudier le comportement de $A(x,y)$ au voisinage de la singularité (en x) de $C(x,y)$.

Pour les mots binaires, le poids sera le nombre total de 1 (la somme des coordonnées). La fonction génératrice devient $C(z,t) = (1-z(1+t))^{-1}$. La loi de la taille reste une loi géométrique (de paramètre $x(1+y)$), et la loi conditionnelle sur $\mathcal{C}_n$, sachant que la taille vaut n, est la loi $\mu_{n,p}$ (définition 2.1), au changement de paramètre $p = \frac{y}{1+y}$ près. Pour la stabilité des mots binaires, la fonction génératrice est :

$$A(z,t) = \left(1 - z\left(\frac{1-\sqrt{1+4t}}{2}\right)\right)^{-1} + \left(1 - z\left(\frac{1+\sqrt{1+4t}}{2}\right)\right)^{-1} .$$

Pour la parité, la fonction génératrice est $A(z,t) = \frac{1}{2}(C(z,t) + C(z,-t))$. Dans les deux cas, la conclusion du cas équiprobable ($y = 1$) s'étend à une valeur de y quelconque.

Pour transformer cette approche analytique en une loi du zéro-un véritablement intéressante, il resterait à caractériser les propriétés $\mathcal{A}$ pour lesquelles on peut décider a priori que le comportement de $A(z)$ au voisinage de la singularité de $C(z)$ est tel que la probabilité converge, de manière à cerner leur pouvoir d'expression logique. Signalons dans cette direction les travaux de Compton [15], qui relient les propriétés analytiques des fonctions génératrices au langage de la logique du premier ordre (voir aussi le livre récent de Burris [10]).

On peut aussi penser à sortir du cadre binaire, qui n'est qu'un cas très particulier pour les inégalités de concentration que nous étudierons en 4. Rossignol [94] a démontré l'analogue de la proposition 2.11 pour un produit d'espaces finis quelconques. La difficulté majeure dans ce cas est qu'on n'y dispose plus aussi naturellement de l'outil puissant qu'est la monotonie. On peut résoudre cette difficulté en paramétrant la loi d'une composante par un paramètre unidimensionnel, comme par exemple dans [90], où une application de la loi du zéro-un de Friedgut et Kalai est proposée dans le contexte des systèmes cohérents en fiabilité.

Une autre voie d'extension des lois du zéro-un est la *dépendance faible*. Nous nous contenterons d'indiquer les éléments sur lesquels on peut fonder quelque espoir. Comme nous l'avons vu en 2.1, il est possible d'énoncer des lois du zéro-un (certes rudimentaires) en se basant uniquement sur la loi des grands nombres (théorème 2.2) ou des inégalités exponentielles (théorème 2.10). Or ces résultats ont été depuis longtemps étendus sous différentes hypothèses de dépendance faible. Le modèle le plus ancien est celui des chaînes de Markov, pour lequel la loi des grands nombres et le théorème central limite sont des résultats de base (cf. Chung [11] p. 94). Plus récemment, plusieurs inégalités exponentielles de type Chernov sont aussi apparues (cf. [79]). Des résultats analogues sont également disponibles pour d'autres hypothèses de dépendance faible, moins contraignantes que le cas markovien (voir Doukhan [25], Rio [93], ou le recueil d'articles [20]). La loi du zéro-un de Kolmogorov que nous rappellerons en 3.1 a également été généralisée aux chaînes et processus de Markov [14, 62]. Le comportement asymptotique des chaînes de Markov sur de gros espaces d'états donne lieu à des transitions abruptes dans le temps, que l'on doit rapprocher des phénomènes de seuil que nous étudions ici. Nous ne les aborderons pas : voir [82, 112, 113, 114, 115], et aussi Lynch [81] qui introduit la notion de fonction seuil pour les chaînes de Markov avec un point de vue de théorie des graphes.

3 ... tout événement raisonnable ...

3.1 Les grands ancêtres

L'étude de la convergence presque sûre d'une suite de variables aléatoires passe le plus souvent par le théorème (ou lemme) de Borel-Cantelli ([36] p. 200).

Théorème 3.1 *Soit $(B_n)_{n\in\mathbb{N}}$ une suite d'événements (quelconques).*

$$\sum_{n\in\mathbb{N}} Prob[B_n] < \infty \Longrightarrow Prob\left[\bigcap_{n_0\geq 0}\bigcup_{n\geq n_0} B_n\right] = 0 \; .$$

Ce résultat doit se lire ainsi : si les probabilités $Prob[B_n]$ tendent vers 0 assez rapidement pour que la série $\sum Prob[B_n]$ converge, alors presque sûrement (avec probabilité 1) seul un nombre fini des B_n sont réalisés simultanément.

Démonstration : Posons :

$$C_{n_0} = \bigcup_{n\geq n_0} B_n \; .$$

La suite d'événements (C_{n_0}) est décroissante ($C_{n_0} \supset C_{n_0+1}$) et donc :

$$Prob\left[\bigcap_{n_0\geq 0} C_{n_0}\right] = \lim_{n_0\to\infty} Prob\left[C_{n_0}\right] \; .$$

Or :

$$Prob[C_{n_0}] = Prob\left[\bigcup_{n\geq n_0} B_n\right] \leq \sum_{n=n_0}^{\infty} Prob[B_n] \; .$$

La probabilité de C_{n_0} est majorée par le reste de la série convergente $\sum Prob[B_n]$, elle tend donc vers 0. □

Ce théorème admet une sorte de réciproque dans le cas d'événements indépendants.

Théorème 3.2 *Si les événements $(B_n)_{n\in\mathbb{N}}$ sont indépendants alors :*

$$\sum Prob[B_n] = \infty \Longleftrightarrow P\left[\bigcap_{n_0\geq 0}\bigcup_{n\geq n_0} B_n\right] = 1 \; .$$

Si les B_n sont indépendants et si leurs probabilités sont suffisamment fortes pour que la série diverge alors une infinité de B_n seront réalisés simultanément.

Démonstration : Avec les mêmes notations que dans la démonstration précédente, la probabilité de C_{n_0} se calcule en utilisant l'hypothèse d'indépendance :

$$\begin{aligned} Prob[C_{n_0}] &= Prob\left[\bigcup_{n\geq n_0} B_n\right] = 1 - Prob\left[\bigcap_{n\geq n_0} \overline{B_n}\right] \\ &= 1 - \prod_{n\geq n_0} (1 - Prob[B_n]) \; . \end{aligned}$$

Dire que la série $\sum Prob[B_n]$ diverge équivaut à dire que le produit $\prod(1 - Prob[B_n])$ tend vers 0, soit que la probabilité de C_{n_0} vaut 1 pour tout n_0. Donc la probabilité de leur intersection vaut 1. □

Voici quel est le rapport avec la convergence presque sûre d'une suite de variables aléatoires. La suite $(X_n)_{n\in\mathbb{N}}$ et sa limite éventuelle X étant données, définissons l'événement :

$$B_n^\varepsilon = \text{``}|X_n - X| \geq \varepsilon\text{''} .$$

Dire que X_n converge vers X c'est dire que pour tout ε, le complémentaire $\overline{B_n^\varepsilon}$ est toujours réalisé à partir d'un certain rang n_0. La suite (X_n) converge vers X *presque sûrement* (c'est à dire avec probabilité 1) si et seulement si pour tout $\varepsilon > 0$:

$$Prob\left[\bigcap_{n_0\geq 0}\bigcup_{n\geq n_0} B_n^\varepsilon\right] = 0 .$$

Corollaire 3.3 *Si la série de terme général $Prob[|X_n - X| \geq \varepsilon]$ converge, alors (X_n) tend vers X presque sûrement.*

Considérons par exemple une suite (X_n) de variables aléatoires de Bernoulli indépendantes :

$$Prob[X_n = 1] = p(n) \qquad Prob[X_n = 0] = 1 - p(n) .$$

Si $\sum p(n) < \infty$ alors X_n converge p.s. vers 0 : la suite ne prendra qu'un nombre fini de fois la valeur 1. Si $\sum p(n) = \infty$, la suite prendra une infinité de fois la valeur 1.

Il est légitime de se demander si une suite de variables aléatoires peut converger avec une probabilité différente de 0 ou 1. La loi du zéro-un de Kolmogorov ([37] p. 124) montre que ce n'est pas possible si les variables sont indépendantes.

Théorème 3.4 *Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires indépendantes. Soit A un événement exprimable en fonction de $(X_n)_{n\in\mathbb{N}}$, mais indépendant de $(X_1, \ldots, X_n)$ pour tout n.*
Alors $Prob(A) = 0$ ou 1.

Démonstration : Nous admettrons que l'on peut approcher l'événement A à ε près au sens de la probabilité sur l'espace produit, par une suite d'événements (A_n), tels que pour tout n A_n soit exprimable en fonction de $(X_1, \ldots, X_n)$.

$$Prob[A \setminus A \cap A_n] < \varepsilon \quad \text{et} \quad Prob[A_n \setminus A \cap A_n] < \varepsilon . \tag{3.1}$$

Or par hypothèse, A et A_n sont indépendants. On en déduit :

$$Prob[A] - Prob[A]Prob[A_n] < \varepsilon ,$$

et en passant à la limite,

$$Prob[A] = (Prob[A])^2 .$$

L'événement A est donc indépendant de lui-même. □

Parmi les événements A auxquels la loi du zéro-un de Kolmogorov s'applique, aucun ne peut dépendre que d'un nombre fini de variables. Ils concernent tous des propriétés asymptotiques de la suite : convergence de (X_n) ou de $\sum X_n$, propriétés de la limite supérieure, etc...

La loi du zéro-un de Hewitt-Savage ([37] p. 125) part d'une hypothèse plus restrictive, mais sa portée est un peu plus large. Elle concerne des événements invariants par permutation finie des variables.

Définition 3.5 *Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires.*
Soit $A = A(X_1, X_2, \ldots)$ un événement exprimable en fonction de $(X_n)_{n\in\mathbb{N}}$. On dit que A est invariant par permutation finie des variables si pour tout $i < j \in \mathbb{N}$:

$$A(X_1, \ldots, X_i, \ldots, X_j, \ldots) = A(X_1, \ldots, X_j, \ldots, X_i, \ldots) .$$

Le fait que A ne soit pas modifié si on échange deux variables de la suite entraîne que A reste invariant si on en permute un nombre fini quelconque (comparer avec la définition 2.12).

Théorème 3.6 *Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires indépendantes et de même loi. Soit A un événement exprimable en fonction de $(X_n)_{n\in\mathbb{N}}$, invariant par permutation finie des coordonnées. Alors $Prob(A) = 0$ ou 1.*

Démonstration : Reprenons la suite (A_n) d'événements approchant A de la démonstration précédente. Notons B_n l'événement obtenu en inversant l'ordre des $2n$ premières variables de la suite, les autres restant fixes. Par hypothèse, les inégalités (3.1) vraies pour A_n, le sont aussi pour B_n. On en déduit que la différence symétrique entre A et $A_n \cap B_n$ a une probabilité inférieure à 4ε. Mais comme A_n dépend des n premières variables, B_n dépend des n suivantes, et ils sont donc indépendants. Donc :

$$Prob[A] - Prob[A_n]Prob[B_n] < 4\varepsilon ,$$

et en passant à la limite,

$$Prob[A] = (Prob[A])^2 .$$

□

Parmi les événements A auxquels s'applique la loi du zéro-un de Hewitt-Savage on retrouve les événements asymptotiques du théorème 3.4. Mais il y en a d'autres. Posons par exemple $S_n = X_1 + \cdots + X_n$ et $B_n =$ "$\frac{1}{n}S_n \in I$", où I est un intervalle de $\mathbb{R}$. L'événement :

$$A = \bigcap_{n_0 \geq 0} \bigcup_{n \geq n_0} B_n ,$$

"les moyennes de Cesaro de la suite (X_n) visitent I une infinité de fois" a pour probabilité 0 ou 1 : comparer avec le corollaire 2.3.

3.2 Glebskii-Fagin

Pour un logicien, une loi du zéro-un n'est pas nécessairement une bonne nouvelle : si dans un langage donné les probabilités sont proches de 0 ou 1, cela signifie que toute proposition ou sa négation est presque une tautologie, et donc que le langage considéré n'a qu'un faible pouvoir d'expression (voir [1, 2, 3, 15, 60]). Néanmoins les lois du zéro-un en logique ont quelque chose de fascinant, par la généralité et la simplicité de leur énoncé. Celle que nous allons démontrer (théorème 3.13) est due indépendamment à Glebskii et al. [48] et Fagin [35], et concerne la logique du premier ordre. N'étant pas en mesure de développer un cours de logique (voir par exemple [16, 31, 70, 71]), nous nous contenterons d'en reprendre quelques notions élémentaires en les reliant aux exemples de modèles qui ont été introduits dans la partie précédente.

Nous partons d'un ensemble fini $\mathcal{R}$ de *relations*, (appelées aussi *prédicats*) concernant chacune un nombre fixé d'objets. Ce nombre est l'*arité* de la relation. L'ensemble des objets auxquels s'appliquent les relations est fini, mais sa taille est destinée à tendre vers l'infini. C'est le *domaine*, que nous noterons X_n. Le couple $(\mathcal{R}, X_n)$ constitue un *modèle*.

Dans le premier modèle que nous avons introduit, le domaine est $\{1, \ldots, n\}$, auquel s'applique une relation unaire R : pour $i \in \{1, \ldots, n\}$, Ri signifie que la coordonnée i vaut 1 et $\neg Ri$ qu'elle vaut 0. Pour une image binaire, le domaine est l'ensemble des pixels, identifié à $\{1, \ldots, n\}^2$. La couleur C des pixels est aussi une relation unaire : Ci pour un pixel noir, $\neg Ci$ pour un blanc.

Tous les modèles que nous considérons contiennent implicitement l'égalité, qui est une relation binaire. A partir du moment où nous introduisons une structure de graphe, comme dans les configurations cycliques (exemples 1 et 2), une autre relation binaire apparaît. Le domaine est alors vu comme un ensemble de sommets et si i et j sont deux sommets, Rij signifie qu'ils sont reliés par une arête. Pour parler d'une image en distinguant la verticale de l'horizontale, nous aurons besoin de deux relations binaires, V et H. Tous les graphes que nous avons considérés jusqu'ici sont non orientés et sans boucle. Mais pour une relation binaire quelconque, Rii est possible et Rij n'implique pas nécessairement Rji.

Un modèle $(\mathcal{R}, X_n)$ étant fixé, on définit une *structure* en donnant une liste de faits relatifs aux relations de $\mathcal{R}$ appliquées aux éléments de X_n. Chaque fait possible $Ri_1, \dots i_k$ doit apparaître dans la liste, soit affirmé, soit nié (mais pas les deux). La figure 5 donne un exemple de structure pour une seule relation binaire et le domaine $\{1, 2, 3\}$ (graphe orienté).

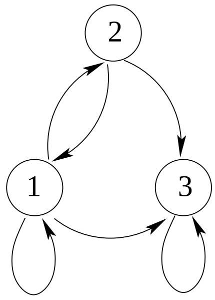

FIG. 5 – Exemple de structure pour une relation binaire (graphe orienté à 3 sommets) : $\{R11, \neg R22, R33, R12, R21, R23, \neg R32, R13, \neg R31\}$.

Nous notons E_n l'ensemble de toutes les structures relatives au modèle $(\mathcal{R}, X_n)$. Supposons que $\mathcal{R}$ ne contienne qu'une relation k-aire R. On peut voir une structure S comme une application de $(X_n)^k$ dans $\{0, 1\}$, qui à tout k-uplet $(i_1, \dots, i_k)$ d'éléments de X_n, associe 1 si S contient le fait $Ri_1 \dots i_k$ et 0 si S contient $\neg Ri_1, \dots, i_k$. Nous avons utilisé ce genre d'identification dans la partie précédente pour les graphes, les images et la k-satisfiabilité. Pour une seule relation k-aire sur $X_n = \{1, \dots, n\}$, il y a 2^{n^k} structures et nous noterons $E_{n,R}$ leur ensemble. Dans le cas général, on peut par le même raisonnement identifier E_n à un produit cartésien, dont chacune des composantes correspond à une relation de la structure :

$$E_n = \underset{R \in \mathcal{R}}{\times} E_{n,R} \ .$$

Nous munissons E_n d'une loi de probabilité paramétrée qui généralise celles de la partie précédente.

Définition 3.7 *Soit E_n l'ensemble des structures relatives à un modèle $(\mathcal{R}, X_n)$. A chaque relation $R \in \mathcal{R}$, associons un réel p_R compris entre 0 et 1, et notons $\mathbf{p}$ l'ensemble des paramètres p_R, pour $R \in \mathcal{R}$. On définit la probabilité $\mu_{n,\mathbf{p}}$ sur E_n comme suit.*

1. *Pour tout $S \in E_n$, la probabilité que S contienne le fait $Ri_1 \dots i_k$ est p_R.*
2. *Tous ces événements sont indépendants dans leur ensemble.*

Dans le cas d'une seule relation unaire sur $\{1, \dots, n\}$, on retrouve la probabilité $\mu_{n,p}$ de 2.1. Si $\mathcal{R}$ contient une seule relation binaire (non symétrique), une structure est un graphe orienté, avec boucles. Dans ce cas la probabilité $\mu_{n,\mathbf{p}}$ est l'analogue pour les graphes orientés de la loi des graphes aléatoires de la section 2.2 (voir par exemple [65, 80]).

La logique du premier ordre est une formalisation construite à partir des symboles logiques $\forall, \exists, \neg, \wedge, \vee, =$, de relations d'arité finie $R_1, R_2, \dots$, et de variables $x_1, x_2, \dots$ Si $R \in \mathcal{R}$ est une relation k-aire et $(x_1, \dots, x_k)$ un k-uplet de variables, la formule élémentaire $Rx_1 \dots x_k$ est un *atome*. Nous utiliserons l'abréviation classique $\mathbf{x}$ pour le k-uplet $(x_1, \dots, x_k)$ et $R\mathbf{x}$ pour l'atome $Rx_1, \dots x_k$. Par définition, la logique du premier ordre est constituée de l'ensemble des formules que l'on peut construire de manière récursive à partir des atomes, en utilisant les connecteurs logiques habituels.

Définition 3.8 *La logique du premier ordre relative à l'ensemble de relations $\mathcal{R}$ est l'ensemble de formules $\mathcal{L}_1$ défini de la façon suivante.*

1. *Pour tout k, toute relation k-aire $R \in \mathcal{R}$ et tout k-uplet $\mathbf{x}$ de variables, $R\mathbf{x} \in \mathcal{L}_1$.*
2. *Si A et B sont des formules de $\mathcal{L}_1$, alors $(\neg A)$, $(\forall \mathbf{x}\, A\mathbf{x})$ et $(A \wedge B)$ appartiennent aussi à $\mathcal{L}_1$.*

Comme conséquence, si A et B sont deux formules de $\mathcal{L}_1$, $(\exists \mathbf{x}\, A\mathbf{x})$, $(A \vee B)$, $(A \to B)$ et $(A \leftrightarrow B)$ sont aussi dans $\mathcal{L}_1$: les logiciens notent $A \to B$ l'implication $(\neg A \vee B)$ et $A \leftrightarrow B$ l'équivalence $(A \to B) \wedge (B \to A)$.

A partir du moment où un univers X_n a été fixé (disons $\{1, \ldots, n\}$), les relations s'appliquent aux éléments de X_n. Considérons par exemple la stabilité d'un sous-graphe (exemple 1). Notons R_1 la relation unaire et R_2 la relation binaire de voisinage. La formule de définition (2.7) s'écrira :

$$\forall x, y \ (R_1 x \wedge R_2 xy) \to \neg R_1 y \ . \tag{3.2}$$

Parmi toutes les formules de $\mathcal{L}_1$, seules nous intéressent celles dont on peut décider si elles sont vraies ou fausses pour une structure donnée. Nous conviendrons donc d'appeler *proposition* une formule de $\mathcal{L}_1$ dont toutes les variables sont quantifiées (formule close ou phrase en logique). Soit S une structure et A une proposition de $\mathcal{L}_1$. Comme A est construite à partir des relations de $\mathcal{R}$ (définition 3.8), elle est ou non compatible avec S. Si elle l'est, on dit que S "satisfait" A, et on note $S \models A$. La donnée de A partitionne l'ensemble E_n en deux sous-ensembles : l'ensemble des structures qui satisfont A, que nous noterons A_n, et son complémentaire. Nous dirons aussi que A_n est l'ensemble des structures pour lesquelles A *est vraie*. Les logiciens munissent habituellement l'ensemble E_n de l'équiprobabilité, auquel cas la probabilité que A soit vraie est le rapport du cardinal de A_n au cardinal de E_n. C'est un cas particulier de la définition 3.7 (si tous les p_R valent $\frac{1}{2}$). Dans le cas général, la probabilité que A soit vraie sera :

$$\mu_{n,\mathbf{p}}(A_n) = \sum_{S \models A} \mu_{n,\mathbf{p}}(S) \ .$$

Le résultat principal de cette section (théorème 3.13) dit que pour tout $\mathcal{R}$, toute famille de paramètres $\mathbf{p}$, et toute proposition du premier ordre A, la probabilité $\mu_{n,\mathbf{p}}(A_n)$ que A soit vraie tend vers 0 ou 1 quand la taille n du domaine tend vers l'infini. La clé de la démonstration est le théorème 3.11 dû à Gaifman [45]. Nous ne donnerons pas sa démonstration, qui est basée sur le jeu d'Ehrenfeucht [32] (voir [30] p. 44 et [101] p. 318 pour le cas des graphes). Il ramène l'étude de $\mathcal{L}_1$ à des propositions d'un type particulier, les extensions.

Définition 3.9 *Soit $\mathcal{X}_n = \{x_1, \ldots, x_n\}$ un ensemble de n variables distinctes. On appelle* description complète *de $\mathcal{X}_n$ une conjonction D d'atomes telle que pour tout k, toute relation k-aire $R \in \mathcal{R}$ et tout k-uplet $\mathbf{x}$ de variables de $\mathcal{X}_n$, D contient soit $R\mathbf{x}$, soit $\neg R\mathbf{x}$ (mais pas les deux).*

Appliquée à un domaine X_n, Une description complète répond donc à toutes les questions possibles concernant X_n. Il existe une structure et une seule qui la satisfait. Dans la structure des graphes à n sommets, une description complète donne la liste de tous les couples de sommets reliés et de tous les couples qui ne le sont pas.

Définition 3.10 *Soit $\mathcal{X}_m = \{x_1, \ldots, x_m\}$ un m-uplet de variables, et $D = D(\mathcal{X}_m)$ une description complète de $\mathcal{X}_m$. Soit y une nouvelle variable, et $D' = D'(\mathcal{X}_m, y)$ une description complète de $\mathcal{X}_m \cup \{y\}$. On dit que D'* étend *D si $D' \to D$.*

On appelle extension *toute proposition du type :*

$$\forall x_1, \ldots, x_m \ , \left(\bigwedge_{1<i<j<m} x_i \neq x_j \right) \wedge D(\mathcal{X}_m) \ \to \ \exists y \ , \left(\bigwedge_{1<i<m} y \neq x_i \right) \wedge D'(\mathcal{X}_m, y) \ , \tag{3.3}$$

où :

- *$\mathcal{X}_m = \{x_1, \ldots, x_m\}$ désigne un sous-ensemble de cardinal m de variables distinctes,*
- *$D(\mathcal{X}_m)$ est une description complète de $\mathcal{X}_m$,*
- *y est une variable distincte des x_i,*
- *$D'(\mathcal{X}_m, y)$ est une description complète de $\mathcal{X}_m \cup \{y\}$ qui étend $D(\mathcal{X}_m)$.*

Voici deux exemples d'extensions dans le langage des graphes non orientés (la relation de voisinage R est symétrique).

$$\forall x \exists y \ , \ y \neq x \wedge Rxy \ .$$

(Aucun sommet n'est isolé).

$$\forall x_1, x_2 \ , \ (x_1 \neq x_2 \wedge Rx_1x_2) \ \rightarrow \ \exists y \ (y \neq x_1 \wedge y \neq x_2 \wedge Rx_1y \wedge Rx_2y) \ .$$

(Toute arête est dans un triangle).
La figure 6 illustre une extension dans le modèle des graphes orientés (pour une relation binaire).

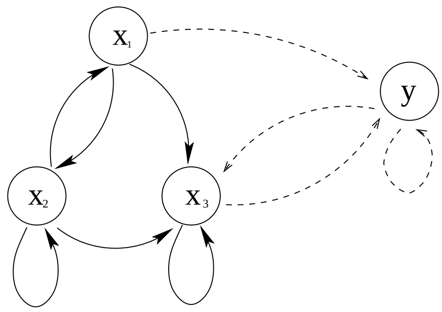

FIG. 6 – Exemple d'extension dans le modèle des graphes orientés.

Théorème 3.11 *L'ensemble $\mathcal{T} \subset \mathcal{L}_1$ de toutes les extensions est consistant et complet.*

En clair, toute proposition de la logique du premier ordre ou son contraire est impliquée par une conjonction finie d'extensions. On en déduit immédiatement le corollaire suivant.

Corollaire 3.12 *Si la probabilité de toute extension tend vers* 1*, alors la probabilité de toute proposition du premier ordre tend vers* 0 *ou* 1*.*

Théorème 3.13 *Soit :*

- *$\mathcal{R}$ un ensemble fini de relations,*
- *n un entier et X_n un domaine de cardinal n,*
- *E_n l'espace des structures du modèle $(\mathcal{R}, X_n)$,*
- *$\mathbf{p} = \{p_R \ , \ R \in \mathcal{R}\}$ une famille de paramètres strictement compris entre* 0 *et* 1*,*
- *$\mu_{n,\mathbf{p}}$ la loi de probabilité de paramètre $\mathbf{p}$ sur E_n (définition 3.7).*

Soit $A \in \mathcal{L}_1$ une proposition du premier ordre quelconque et $A_n \subset E_n$ l'ensemble des structures qui satisfont A. Alors :

$$\lim_{n\to\infty} \mu_{n,\mathbf{p}}(A_n) = 0 \ \textit{ou} \ 1 \ .$$

Démonstration : D'après le corollaire 3.12, il nous suffit d'examiner les probabilités des extensions. Soit T l'extension définie par la formule (3.3), et T_n l'ensemble des structures de E_n qui satisfont T. Nous allons montrer que $\mu_{n,\mathbf{p}}(E_n \setminus T_n)$ tend vers 0. La proposition $\neg T$ s'écrit :

$$\exists x_1, \ldots, x_m \ , \ \left(\bigwedge_{1<i<j<m} x_i \neq x_j \right) \wedge D(\mathcal{X}_m) \wedge \forall y \ \left(\bigwedge_{1<i<m} y \neq x_i \right) \rightarrow \neg D'(X_m, y) \ .$$

Choisissons m éléments distincts de X_n, notons X_m leur ensemble et $\overline{T}_{n,X_m}$ l'ensemble des structures de E_n qui satisfont :

$$D(X_m) \wedge \forall y \, \neg D'(X_m, y) \ .$$

L'ensemble $E_n \setminus T_n$ est la réunion des $\overline{T}_{n,X_m}$: il y en a $\binom{n}{m}$. Par symétrie de la loi $\mu_{n,\mathbf{p}}$ leurs probabilités sont égales. On peut donc fixer un sous-ensemble X_m et écrire :

$$\mu_{n,\mathbf{p}}(E_n \setminus T_n) \leq \binom{n}{m} \mu_{n,\mathbf{p}}(\overline{T}_{n,X_m}) \ .$$

Par rapport à $D(X_m)$, la description $D'(X_m, y)$ contient en plus tous les atomes relatifs à y. Pour un y fixé, le nombre de ces atomes est borné indépendamment de n. Donc la probabilité de $\neg D'(X_m, y)$

est un certain réel $\tilde{p}$, indépendant de n dès que $n > m$. Comme tous les p_R sont différents de 0 et 1, il en est de même pour $\tilde{p}$. A cause de l'indépendance des faits, la probabilité de $\overline{T}_{n,X_m}$ est majorée par $\tilde{p}^{n-m}$ (car $n-m$ est le cardinal de $X_n \setminus X_m$). On obtient donc :

$$\mu_{n,\mathbf{p}}(E_n \setminus T_n) \leq \binom{n}{m} \tilde{p}^{n-m} ,$$

qui tend vers 0 quand n tend vers l'infini. □

Le théorème 3.13 est certes important mais il ne couvre pas les modèles intéressants en pratique. Par exemple, il s'applique aux graphes orientés avec boucle (si $\mathcal{R}$ est constitué d'une seule relation binaire) mais pas aux graphes non orientés de la définition 2.16. Ceux-ci sont contraints de vérifier la propriété suivante :

$$(\forall x \, \neg Rxx) \wedge (\forall x \neq y \; Rxy \rightarrow Ryx) . \tag{3.4}$$

La probabilité $\mu_{\alpha(n),p}$ sur les graphes non orientés est en fait une probabilité *conditionnelle* sachant (3.4). On démontre que le théorème 3.13 s'étend à tous les modèles dits *paramétriques*, pour lesquels le conditionnement s'exprime par une conjonction de formules du type $\forall \mathbf{x} \; \psi(\mathbf{x})$, comme (3.4). Ce résultat est dû à Oberschelp [88] (voir aussi [30] p. 74 et [101] p. 317 pour les graphes non orientés).

Mais de nombreux modèles ne relèvent ni du théorème 3.13, ni de son extension aux conditionnements de type paramétrique. C'est le cas en particulier des images aléatoires (définition 2.19). Le langage des images comporte une relation unaire (la couleur C) et deux relations binaires de voisinage (horizontal H et vertical V). Le domaine est $X_n = \{1, \ldots, n\}^2$. Parmi toutes les structures, seules nous intéressent celles pour lesquelles le graphe décrit par les faits Hxy et Vxy est exactement le tore à deux dimensions. Associons aux trois relations une famille de paramètres $\mathbf{p} = \{p_C, p_H, p_V\}$. Pour la loi $\mu_{n^2,\mathbf{p}}$, la probabilité de l'ensemble I_n des images de taille $n \times n$ décroît exponentiellement avec n. La loi de probabilité des images aléatoires (définition 2.19), est la loi *conditionnelle* $\mu_{n^2,\mathbf{p}}(\,\cdot \,|\, I_n)$. Dans le modèle des configurations cycliques (exemples 1 et 2), la relation binaire est contrainte de manière à décrire le graphe cyclique C_n (définition 2.14). Différents types de contraintes pour des ensembles de relations unaires et binaires dans le cadre des logiques de description sont proposés dans [95]. Dans [35] p. 57, Fagin montre que la loi du zéro-un pour $\mu_{n,\mathbf{p}}$ n'entraîne pas nécessairement un résultat analogue pour des probabilités conditionnelles (voir aussi le chapitre 3 de [30] ainsi que [57, 58]). Dans la section 8 de [15], Compton liste quelques problèmes ouverts dans ce domaine. Nous nous contenterons de démontrer la loi du zéro-un pour les configurations cycliques aléatoires. Ce qui suit s'adapte sans difficulté aux images aléatoires.

Proposition 3.14 *On considère le modèle $(\mathcal{R}, X_n)$ défini par :*

- *$\mathcal{R} = \{R_1, R_2\}$, où R_1 est unaire et R_2 binaire symétrique.*
- *$X_n = \{1, \ldots, n\}$.*

Soit $C_n \subset E_n$ l'ensemble des structures dont les faits relatifs à R_2 sont les suivants (cf. definition 2.14).

$$\begin{cases} R_2 ij & si\ j = i \pm 1 \ modulo\ n , \\ \neg R_2 ij & sinon. \end{cases}$$

Soit $p = p_{R_1} \in [0,1]$ un paramètre réel. Notons $\mu_{n,p}$ la probabilité conditionnelle $\mu_{n,\mathbf{p}}(\,\cdot \,|\, C_n)$ (la probabilité paramétrant R_2 n'intervient pas dans la définition de $\mu_{n,p}$).

Soit $A \in \mathcal{L}_1$ une proposition du premier ordre quelconque et $A_n \subset C_n$ l'ensemble des structures qui satisfont A. Alors :

$$\lim_{n \to \infty} \mu_{n,p}(A_n) = 0 \ \ ou \ 1 .$$

Démonstration : Contrairement au théorème 3.13, certaines extensions ont une probabilité qui tend vers 0, d'autres vers 1 et le corollaire 3.12 ne s'applique pas. Nous allons utiliser un autre théorème de Gaifman, qui exprime différemment le caractère local de la logique du premier ordre. On note d la distance usuelle entre sommets du graphe cyclique, et $S(x, r)$ (sphère de centre x et de rayon r)

l'ensemble des sommets à distance au plus r de x. On appelle *proposition locale basique* une proposition du type suivant.

$$\exists x_1 \ldots \exists x_m \left(\bigwedge_{1\leq i<j\leq m} d(x_i,x_j) > 2r \right) \wedge \left(\bigwedge_{1\leq i\leq m} \psi_i(x_i) \right) , \tag{3.5}$$

où :

- m et r sont des entiers positifs fixés,
- pour tout $i = 1,\ldots,m$, $\psi_i(x)$ désigne une formule de $\mathcal{L}_1$ pour laquelle seule la variable x est libre (non quantifiée), et les autres variables appartiennent toutes à la sphère $S(x,r)$.

Le théorème de Gaifman (voir [46] ou [30] p. 30) affirme que toute proposition du premier ordre est équivalente à une combinaison booléenne finie de propositions locales basiques. Pour démontrer la proposition 3.14, il suffit donc de montrer que la probabilité de toute proposition locale basique tend vers 0 ou 1 (cf. lemme 2.5).

Examinons tout d'abord les formules $\psi_i(x)$. Soit l'une d'entre elles est insatisfiable, auquel cas la proposition locale basique (3.5) est de probabilité nulle, soit toutes sont satisfiables, et nous allons montrer que la probabilité de (3.5) tend vers 1. Nous supposons désormais que n est strictement plus grand que $m(2r+1)$. Dire que $\psi_i(x)$ est satisfiable équivaut à dire qu'il existe une description complète de la sphère $S(x,r)$ qui implique $\psi_i(x)$. Pour tout $i = 1,\ldots,m$, nous fixons une telle description complète, que nous notons $D_i(x)$. Nous allons ensuite limiter les centres de sphères possibles. Nous dirons que $x_1,\ldots,x_m$ sont *r-consécutifs* s'ils sont distincts et si pour tout $i = 2,\ldots,m$:

$$d(x_{i-1},x_i) = 2r+1 \; .$$

Il est clair que si $x_1,\ldots,x_m$ sont r-consécutifs, alors deux quelconques d'entre eux sont à distance strictement supérieure à $2r$. La proposition locale basique (3.5) est alors impliquée par la proposition suivante.

$$\exists x_1,\ldots,x_m \; r\text{-consécutifs} \bigwedge_{1\leq i\leq m} D_i(x_i) \; . \tag{3.6}$$

Notons D cette proposition et D_n l'ensemble des structures de C_n qui satisfont D. Nous voulons montrer que $\mu_{n,p}(D_n)$ tend vers 1. La sphère $S(x,r)$ a $2r+1$ éléments, donc la probabilité de chaque $D_i(x)$ est minorée par $\tilde{p} = \min\{p, 1-p\}^{2r+1}$. Si $x_1,\ldots,x_m$ sont r-consécutifs, $D_1(x_1)\wedge\ldots\wedge D_m(x_m)$ est une description complète de la réunion des sphères $S(x_i,r)$, qui est un segment du graphe cyclique, contenant exactement $m(2r+1)$ points. Sa probabilité est minorée par $\tilde{p}^m$. La probabilité qu'une telle description complète apparaisse dans une configuration cyclique aléatoire de taille n tend vers 1 quand n tend vers l'infini. □

Les logiciens ont évidemment cherché très tôt à étendre le théorème 3.13 à la logique du second ordre. La loi du zéro-un est vraie pour certains fragments, fausse pour d'autres : voir les articles de synthèse de Kolaitis et Vardi [67, 68] et Le Bars [73]. Ce dernier est l'auteur de plusieurs exemples de propriétés dont la probabilité ne tend ni vers 0 ni vers 1 (voir [72, 74, 75]).

Dans la section précédente, nous avons plusieurs fois observé que les propriétés intéressantes sont souvent localisées autour de valeurs de la probabilité tendant vers 0 avec n. Il semble donc souhaitable de faire varier avec n le paramètre $\mathbf{p}$ de la définition 3.7. Pour les images aléatoires, la proposition 2.21 montre que la fonction seuil pour l'existence d'une sous-image fixée est une puissance rationnelle de n. C'est le cas aussi pour l'apparition d'un sous-graphe dans un graphe aléatoire ([101] p. 300). Si on choisit pour $p(n)$ une puissance *non rationnelle* de n, alors la probabilité d'apparition d'une sous-image ou d'un sous-graphe tendra vers 0 ou 1. Shelah et Spencer ont montré que c'est aussi le cas pour les extensions (voir [98], [101] p. 315 ainsi que [99, 100, 102]).

Théorème 3.15 *Soit α un réel* non rationnel*, strictement positif. Considérons le modèle de graphe aléatoire $\mathcal{G}(n, n^{-\alpha})$ (cf. définition 2.16). La probabilité de toute proposition du premier ordre tend vers* 0 *ou* 1*.*

3.3 Friedgut-Kalai

Cette section est consacrée au théorème de Friedgut et Kalai [44], que nous avons déjà évoqué en 2.1. Nous reprenons donc le modèle de base de 2.1, à savoir l'espace $E_n = \{0,1\}^n$, muni de la loi produit $\mu_{n,p}$ (définition 2.1). Afin d'alléger un peu les notations, nous remplacerons $\mu_{n,p}$ par μ_p dans toute cette section. Le théorème 3.16 donne un majorant de la largeur du seuil pour toute propriété *croissante* (définition 2.6) et *symétrique* (définition 2.12).

Théorème 3.16 *Soit A une propriété croissante et symétrique. Pour tout $\alpha \in [0,1]$, notons p_α la valeur de p telle que $\mu_{p_\alpha}(A) = \alpha$. Il existe une constante universelle $C \leq 7.03$ telle que pour tout $\varepsilon \in]0,1/2]$, on a :*

$$p_{1-\varepsilon} - p_\varepsilon \leq C \frac{\log \frac{1-\varepsilon}{\varepsilon}}{\log n} .$$

La démonstration de ce théorème sera l'occasion de découvrir une nouvelle manière d'approcher les lois du zéro-un en étudiant la largeur du seuil. La première remarque importante correspond à l'intuition selon laquelle la probabilité de A passera d'autant plus rapidement de 0 à 1 que le graphe de la fonction $p \mapsto \mu_p(A)$ aura une pente plus raide près du seuil. On cherchera donc à montrer que $\frac{d\mu_p(A)}{dp}$ est grand devant $\mu_p(A)$ lorsque $\mu_p(A) \leq \frac{1}{2}$, et grand devant $1-\mu_p(A)$ lorsque $\mu_p(A) \geq \frac{1}{2}$. Cela revient au même de montrer que $\frac{d\mu_p(A)}{dp}$ est grand devant $\mu_p(A)(1-\mu_p(A))$. La formulation précise est donnée par le lemme suivant.

Lemme 3.17 *Soit a un réel positif tel que pour tout $p \in [0,1]$:*

$$\frac{d\mu_p(A)}{dp} \geq a\, \mu_p(A)(1-\mu_p(A)) .$$

Alors, pour tout ε dans $]0,1/2]$,

$$p_{1-\varepsilon} - p_\varepsilon \leq \frac{2}{a} \log \frac{1-\varepsilon}{\varepsilon} .$$

Démonstration : On a :

$$\frac{\frac{d\mu_p(A)}{dp}}{\mu_p(A)(1-\mu_p(A))} \geq a ,$$

soit :

$$\frac{d \log \frac{\mu_p(A)}{1-\mu_p(A)}}{dp} \geq a .$$

D'où, en intégrant entre p_ε et $p_{1-\varepsilon}$,

$$2 \log \frac{1-\varepsilon}{\varepsilon} \geq a\,(p_{1-\varepsilon} - p_\varepsilon) .$$

□

La deuxième remarque fondamentale est que la dérivée $\frac{d\mu_p(A)}{dp}$ s'exprime à l'aide de la somme des *influences* des coordonnées, que l'on peut voir comme une mesure de la *frontière* de A. Ce fait est bien connu en théorie de la percolation sous le nom de "lemme de Russo". Pour l'énoncer précisément, nous avons besoin de quelques définitions. Nous devons notamment définir, pour tout $k = 1, \ldots, n$ l'ensemble des points de A tels que le changement de la k-ième coordonnée les fait sortir de A :

$$A_k = \{x \in A \text{ t.q. } T_k(x) \in E_n \setminus A\} ,$$

où $T_k(x)$ désigne la configuration y telle que $y(i) = x(i)$ pour $i \neq k$ et $y(k) = 1 - x(k)$. L'ensemble A_k apparaît comme la frontière de A relative à la k-ième coordonnée.

L'*influence* de la k-ième coordonnée est une mesure de la frontière relative à la k-ième coordonnée, et donc apparaît logiquement comme une mesure en dimension $n-1$.

Définition 3.18 *Pour tout* $u \in E_{n-1}$, *on note :*

$$l_k(u) = \{(u_1, \ldots, u_{k-1}, 0, u_k, \ldots, u_{n-1}), (u_1, \ldots, u_{k-1}, 1, u_k, \ldots, u_{n-1})\} .$$

On appelle influence de la k-ième coordonnée sur A *et on note* $I_A(k)$ *le nombre :*

$$I_A(k) = \mu_{n-1,p}(\{u \in E_{n-1} \ t.q.\ \mathbb{1}_A \text{ n'est pas constante sur } l_k(u)\}) .$$

Dans la mesure où A est un ensemble croissant, il est facile de voir que $pI_A(k) = \mu_p(A_k)$. Le lemme de Russo montre que la dérivée de $\mu_p(A)$ est égale à la somme des influences (voir [54], p. 35 théorème 2.25).

Lemme 3.19 *Si* A *est croissante,*

$$\frac{d\mu_p(A)}{dp} = \sum_{k=1}^{n} I_A(k) .$$

Démonstration : Fixons k dans $\{1, \ldots, n\}$, et donnons-nous deux n-uplets $\mathbf{p} = (p_1, \ldots, p_n)$ et $\mathbf{p}' = (p'_1, \ldots, p'_n)$ de $[0,1]^n$ tels que :

$$\forall i \neq k \ , \ p_i = p'_i \quad \text{et} \quad p_k \leq p'_k .$$

Comme dans la démonstration du lemme 2.7, considérons un échantillon $(U(i))_{i=1,\ldots,n}$ de n variables aléatoires indépendantes de loi uniforme sur $[0,1]$, et posons :

$$\forall i \in \{1, \ldots, n\}, \ X(i) = \mathbb{1}_{[0,p_i]}(U(i)) \quad \text{et} \quad Y(i) = \mathbb{1}_{[0,p'_i]}(U(i)) ,$$

de sorte que

$$\forall i \in \{1, \ldots, n\}, \ X(i) \leq Y(i) .$$

Les vecteurs aléatoires $X = (X(i))$ et $Y = (Y(i))$ ont pour lois respectives $\mu_{\mathbf{p}}$ et $\mu_{\mathbf{p}'}$, où $\mu_{\mathbf{p}}$ désigne la probabilité sur E_n telle que :

$$\mu_{\mathbf{p}}(x) = \prod_{i=1}^{n} p_i^{x_i}(1-p_i)^{1-x_i} .$$

Comme A est croissante, $X \in A$ entraîne $Y \in A$, et :

$$\begin{aligned}
\mu_{\mathbf{p}'}(A) - \mu_{\mathbf{p}}(A) &= Prob[Y \in A \text{ et } X \notin A] , \\
&= Prob[(X \notin A \text{ et } T_k(X) \in A \text{ et } p_k \leq U(k) < p'_k] , \\
&= Prob\left[T_k(X) \in A_k \text{ et } p_k \leq U(k) < p'_k\right] , \\
&= I_A(k)\,(p'_k - p_k) ,
\end{aligned}$$

en adaptant la définition de $I_A(k)$ à la mesure $\mu_{\mathbf{p}}$. On en déduit que :

$$\frac{\partial \mu_{\mathbf{p}}(A)}{\partial p_k} = I_A(k) .$$

Donc, si $\mathbf{p} = (p, p, \ldots, p)$,

$$\frac{d\mu_p(A)}{dp} = \sum_{i=1}^{n} I_A(k) .$$

□

Une inégalité de la forme $\frac{d\mu_p(A)}{dp} \geq a\mu_p(A)(1-\mu_p(A))$ peut être interprétée comme une inégalité isopérimétrique. En effet, comme l'indique le lemme de Russo, $\frac{d\mu_p(A)}{dp}$ est une mesure de la frontière (surface) de A, tandis que $\mu_p(A)(1-\mu_p(A))$ est une mesure de son volume. On verra dans la section 4.2 les liens étroits entre isopérimétrie et concentration de la mesure.

Nous sommes maintenant ramenés à l'étude de la somme des influences. Or si A est symétrique, les influences $I_A(k)$ sont toutes égales. Il suffit alors de montrer que pour A croissante (pas nécessairement symétrique) au moins une des influences dépasse $\frac{2}{C}\frac{\log n}{n}\mu_p(A)(1-\mu_p(A))$, pour en déduire le théorème 3.16. Cela a été prouvé par Kahn, Kalaï et Linial [64] dans le cas $p=\frac{1}{2}$, puis par Talagrand [105] pour le cas p quelconque.

La démonstration utilise des techniques d'analyse harmonique. On peut munir E_n d'une structure algébrique isomorphe à $(\mathbb{Z}/2\mathbb{Z})^n$. La transformation de Fourier sur les groupes finis (voir par exemple [96]) établit une dualité entre les configurations de E et les sous-ensembles de $\{1,\dots,n\}$. L'addition correspond alors à la différence symétrique, et la multiplication à l'intersection. En ce sens, E_n est son propre dual. L'analyse de Fourier sur le groupe $(\mathbb{Z}/2\mathbb{Z})^n$ est un fil directeur de ce qui suit, même si elle n'est pas indispensable pour comprendre le raisonnement.

Pour tout $i=1,\dots,n$, on définit l'application r_i, de E_n dans $\mathbb{R}$ par :

$$r_i(x)=\begin{cases} -\sqrt{\frac{1-p}{p}} & \text{si} \quad x_i=1 \\ \sqrt{\frac{p}{1-p}} & \text{si} \quad x_i=0 \end{cases} .$$

Pour $S\subset\{1,\dots,n\}$, on définit l'application r_S par :

$$r_S(x)=\prod_{i\in S} r_i(x) .$$

L'ensemble des applications $(r_S)_{S\subset\{0,\dots,n\}}$ forme une base orthonormée de $L^2(E_n,\mu_p)$. Dans le cas $p=\frac{1}{2}$, c'est un résultat d'algèbre bien connu, car les r_S sont alors les caractères du groupe $((\mathbb{Z}/2\mathbb{Z})^n,+)$.

On définit aussi n endomorphismes de $L^2(E_n,\mu_p)$ par :

$$\Delta_k f(x)=\begin{cases} (1-p)(f(x)-f(T_k(x))) & \text{si } x_k=1 , \\ p(f(x)-f(T_k(x))) & \text{si } x_k=0 . \end{cases}$$

L'opérateur Δ_k est un analogue discret de la dérivée partielle suivant la k-ième coordonnée, et il est caractérisé sur notre base de $L^2(E_n,\mu_p)$ par :

$$\Delta_k r_S=\begin{cases} r_S & \text{si } k\in S , \\ 0 & \text{si } k\notin S . \end{cases}$$

Si $f=\mathbb{1}_A$, on a, pour tout $q\geq 1$:

$$\|\Delta_k f\|_q^q = I_A(k)(p(1-p)^q+(1-p)p^q) . \tag{3.7}$$

Notamment, en posant $\alpha_S=\langle f,r_S\rangle=\int_{E_n} f(x)r_S(x)\; d\mu_p(x)$, on peut écrire :

$$f=\sum_{S\subset\{0,\dots,n\}} \alpha_S r_S ,$$

et :

$$\|\Delta_k f\|_2^2=\|\sum_{S\ni k}\alpha_S r_S\|_2^2=\sum_{S\ni k}\alpha_S^2 .$$

On en déduit donc, $(r_S)_{S\subset\{0,\dots,n\}}$ formant une base orthonormée de $L^2(E_n,\mu_p)$, que :

$$\begin{aligned}\sum_{k=1}^n I_A(k) &= \frac{1}{p(1-p)}\sum_{k=1}^n \|\Delta_k f\|_2^2 \\ &= \frac{1}{p(1-p)}\sum_{k=1}^n \|\sum_{S\ni k}\alpha_S r_S\|_2^2 \\ &= \frac{1}{p(1-p)}\sum_{k=1}^n\sum_{S\ni k}\alpha_S^2 \\ &= \frac{1}{p(1-p)}\sum_{S}|S|\,\alpha_S^2 .\end{aligned}$$

D'un autre côté, on a :

$$\mu_p(A) = \|f\|_2^2 = \sum_S \alpha_S^2 ,$$

et comme $\alpha_\emptyset = \mu_p(A)$,

$$\sum_{S\neq\emptyset} \alpha_S^2 = \mu_p(A)(1-\mu_p(A)) .$$

Si cette somme n'est pas portée principalement par les α_S pour $|S|$ petit, alors $\sum_S |S|\alpha_S^2$ sera significativement plus grande. Cela peut être montré pour les $\Delta_k f$, qui ont un support petit, mais non vide : elles sont trop irrégulières pour que $\alpha_S 1\!\!1_{S\ni k}$ soit toujours petit quand $|S|$ est grand.

Le noyau dur de la démonstration du théorème 3.16 est un lemme d'hypercontractivité démontré par Beckner [5] pour $p=\frac{1}{2}$ et modifié par Talagrand [105] pour traiter le cas p quelconque.

Lemme 3.20 *Pour tout nombre $q \geq 2$, pour tout $p \in [0,1]$ et toute famille $(a_S)_{|S|\leq k}$, on a :*

$$\left\| \sum_{|S|\leq k} a_S r_S \right\|_{q,\mu_p} \leq \left(\sqrt{\frac{q-1}{p(1-p)}}\right)^k \left\| \sum_{|S|\leq k} a_S r_S \right\|_{2,\mu_p} .$$

Démonstration : *(Fin de la démonstration du théorème 3.16.)*
La démonstration qui suit est tirée de l'article de Bourgain et Kalai ([9]). Dans la suite, on notera :

$$\rho = \sum_{S\neq\emptyset} \alpha_S^2 = \mu_p(A)(1-\mu_p(A)) .$$

Soit $\beta \in [0,1]$, et K, dépendant de β, défini comme suit :

$$K = \inf\left\{ j \in \{1,\ldots,n\} \text{ t.q. } \sum_{0<|S|\leq j} \alpha_S^2 \geq \rho\beta \right\} .$$

On définit :

$$g = \sum_{0<|S|\leq K} \alpha_S r_S .$$

On a alors :

$$\begin{aligned} \rho\beta \leq \sum_{0<|S|\leq K} |S|\alpha_S^2 &= \sum_{S\in E^n} |S| \langle f, r_S\rangle \langle g, r_S\rangle \\ &= \sum_{k=1}^n \sum_{S\ni k} \langle f, r_S\rangle\langle g, r_S\rangle \\ &= \sum_{k=1}^n \sum_S \langle \Delta_k f, r_S\rangle \langle \Delta_k g, r_S\rangle \\ &= \sum_{k=1}^n \langle \Delta_k f, \Delta_k g\rangle . \end{aligned}$$

En utilisant l'inégalité de Hölder avec $\frac{4}{3}$ et 4 comme exposants conjugués, on obtient :

$$\rho\beta \leq \sum_{k=1}^n \|\Delta_k f\|_{\frac{4}{3}} \|\Delta_k g\|_4 .$$

De l'égalité 3.7, on déduit :

$$\|\Delta_k f\|_{\frac{4}{3}} = ((1-p)^{\frac{1}{3}} + p^{\frac{1}{3}})^{\frac{3}{4}} \|\Delta_k f\|_2^{\frac{3}{2}} \leq \sqrt{2}\|\Delta_k f\|_2^{\frac{3}{2}} .$$

D'autre part, le lemme 3.20 implique :

$$\|\Delta_k g\|_4 \leq C_1^K \|\Delta_k g\|_2 ,$$

où :

$$C_1 = \sqrt{\frac{3}{p(1-p)}} .$$

On a alors :

$$\begin{aligned}
\rho\beta &\leq \sqrt{2}\, C_1^K \sum_{k=1}^{n} \|\Delta_k f\|_2^{\frac{3}{2}} \|\Delta_k g\|_2 \\
&\leq \sqrt{2}\, C_1^K \max_i \|\Delta_k g\|_2^{\frac{1}{2}} \sum_{k=1}^{n} \|\Delta_k f\|_2^{\frac{3}{2}} \|\Delta_k g\|_2^{\frac{1}{2}} .
\end{aligned}$$

L'inégalité de Cauchy-Schwarz entraîne alors :

$$\begin{aligned}
\rho\beta &\leq \sqrt{2}\, C_1^K \max_i \|\Delta_k g\|_2^{\frac{1}{2}} \left(\sum_{k=1}^{n} \|\Delta_k f\|_2^2\right)^{\frac{3}{4}} \left(\sum_{k=1}^{n} \|\Delta_k g\|_2^2\right)^{\frac{1}{4}} \\
&\leq \sqrt{2}\, C_1^K \max_k \|\Delta_k g\|_2^{\frac{1}{2}} \left(\sum_{S} |S|\alpha_S^2\right)^{\frac{3}{4}} \left(\sum_{|S|\leq K} |S|\alpha_S^2\right)^{\frac{1}{4}} \\
&\leq \sqrt{2}\, C_1^K \max_k \|\Delta_k g\|_2^{\frac{1}{2}} \left(\sum_{S} |S|\alpha_S^2\right) \\
&\leq \sqrt{2}\, C_1^K \max_k \|\Delta_k g\|_2^{\frac{1}{2}} \sum_{k=1}^{n} I_k(A)p(1-p) .
\end{aligned}$$

Or, A étant symétrique, les normes $\|\Delta_k g\|_2$ sont toutes égales :

$$\begin{aligned}
\|\Delta_k g\|_2^2 &= \frac{1}{n} \sum_{j=1}^{n} \sum_{\substack{S \ni j \\ |S|\leq K}} \alpha_S^2 \\
&= \frac{1}{n} \sum_{|S|\leq K} |S|\alpha_S^2 \\
&\leq \frac{K}{n}\rho .
\end{aligned}$$

On en déduit :

$$\rho\beta \leq \sqrt{2}\, C_1^K \left(\frac{K}{n}\rho\right)^{\frac{1}{4}} I(A)p(1-p) ,$$

où on a posé $I(A) = \sum_{k=1}^{n} I_k(A)$. Donc :

$$I(A)p(1-p) \geq \frac{\beta\rho^{\frac{3}{4}} n^{\frac{1}{4}}}{\sqrt{2}\, C_1^K K^{\frac{1}{4}}} . \tag{3.8}$$

D'un autre côté, on a :

$$\begin{aligned}
I(A)p(1-p) &\geq K \sum_{|S|\geq K} \alpha_S^2 \\
&\geq K \left(\rho - \sum_{0<|S|\leq K-1} \alpha_S^2\right) .
\end{aligned}$$

Compte tenu de la définition de K :

$$I(A)p(1-p) \geq K\rho(1-\beta) \ . \tag{3.9}$$

Fixons maintenant $\tau \in]0,1[$, et choisissons :

$$\beta = \frac{\left(\frac{(1-\tau)\log n}{2\log\frac{3}{p(1-p)}}\right)^{\frac{5}{4}}}{n^{\frac{\tau}{4}}} \ .$$

On a alors deux cas possibles :

- Si $K \geq \frac{(1-\tau)\log n}{2\log\frac{3}{p(1-p)}}$, alors d'après (3.9),

$$I(A)p(1-p) \geq \rho\frac{(1-\tau)\log n}{2\log\frac{3}{p(1-p)}}(1-\beta) \ .$$

- Si $K \leq \frac{(1-\tau)\log n}{2\log\frac{3}{p(1-p)}}$, alors on remarque que l'application $x \mapsto x^{\frac{1}{4}}C_1^x$ est croissante pour $x > 0$ (car $C_1 \geq \sqrt{12}$), et donc en utilisant (3.8),

$$\begin{aligned} I(A)p(1-p) &\geq \frac{(1-\tau)\log n}{2\sqrt{2}\log\frac{3}{p(1-p)}}\rho^{\frac{3}{4}} \\ &\geq \frac{(1-\tau)\log n}{2\sqrt{2}\log\frac{3}{p(1-p)}}\rho \ . \end{aligned}$$

Par exemple, si l'on prend $\tau = \frac{1}{2}$, on obtient que $\beta \leq 0.289$ (le maximum est atteint pour $n = 22026$), et on a alors :

$$I(A) \geq C_2\frac{\log n}{p(1-p)\log\frac{3}{p(1-p)}}\rho \ ,$$

avec $C_2 = \frac{1-0.289}{4}$. Or pour tout $p \in [0,1]$:

$$p(1-p)\log\frac{3}{p(1-p)} \leq \frac{\log(12)}{4} \ ,$$

Donc, dans le premier cas, on a :

$$I(A) \geq 0.286\log n\,\mu_p(A)(1-\mu_p(A)) \ .$$

Dans le second cas, il suffit de majorer $p(1-p)\log\frac{3}{p(1-p)}$ pour obtenir :

$$I(A) \geq \frac{1}{\sqrt{2}\log 12}\log n\,\mu_p(A)(1-\mu_p(A)) \ .$$

Or $\frac{1}{\sqrt{2}\log 12} \simeq 0.2846 < 0.286$. On obtient donc, dans les deux cas possibles :

$$I(A) \geq \frac{\log n}{\sqrt{2}\log 12}\mu_p(A)(1-\mu_p(A)) \ ,$$

En utilisant le lemme 3.17, on conclut donc :

$$\forall \varepsilon \in]0,\frac{1}{2}] \ , \quad p_{1-\varepsilon} - p_\varepsilon \leq C\,\frac{\log\frac{1-\varepsilon}{\varepsilon}}{\log n} \ ,$$

avec $C = 2\sqrt{2}\log 12 \leq 7.03$. □

La propriété de symétrie (cf. définition 2.12) requise dans le théorème 3.3 peut paraître surprenante. En fait, de telles conditions sont assez naturelles. Donnons quelques exemples.

Exemple 10 *Symétrie totale*
Si une propriété $A \subset E_n$, monotone, est invariante sous l'action du groupe des permutations $\mathcal{S}_n$ tout entier (propriété totalement symétrique), le théorème 3.3 s'applique, mais la largeur de seuil obtenue, $\frac{1}{\log n}$ est bien plus grande que celle donnée par la proposition 2.11, $\frac{1}{\sqrt{n}}$.

Exemple 11 *Stabilité*
Considérons la propriété $\neg S$ "ne pas être stable" (définition 2.14) concernant les configurations du graphe cyclique à n sommets. Cette propriété est croissante, et invariante par le groupe des permutations cycliques, qui est bien transitif sur $\{1, \ldots, n\}$. Le théorème 3.3 s'applique donc, mais une fois encore, la largeur de seuil obtenue, $\frac{1}{\log n}$ est bien plus grande que celle donnée par le calcul de la fonction génératrice, $\frac{1}{\sqrt{n}}$.

Exemple 12 *Runs de longueur $u \log n$*
Considérons à nouveau le graphe cyclique à n sommets. Soit M_n la propriété d'avoir au moins un run de longueur $u \log n$ ($u > 0$ est fixé : voir l'exemple 2 de la section 2.1). La propriété M_n s'identifie également à un sous-ensemble croissant de E_n, invariant sous l'action du groupe cyclique. Le théorème 3.3 s'applique encore, et donne cette fois le bon ordre de largeur de seuil $\frac{1}{\log n}$.

Exemple 13 *Graphes aléatoires*
Nous reprenons les notations et les définitions de la section 2.2. Notamment, on pose $\alpha(n) = \binom{n}{2} = n(n-1)/2$. Il est alors naturel de dire qu'un sous-ensemble de $E_{\alpha(n)}$ est une propriété de graphe s'il est invariant par tout automorphisme de graphe. Ceci revient à dire que $A \subset E_{\alpha(n)}$ est une propriété de graphe si et seulement si A est invariante sous l'action de $\mathcal{S}_n$ (permutations de sommets) sur les arêtes :

$$\forall \sigma \in \mathcal{S}_n,\ \forall i \neq j,\ \sigma.(i,j) = (\sigma(i), \sigma(j)) \ .$$

Les graphes considérés ne comportant pas de boucle, l'action de $\mathcal{S}_n$ sur $\{1, \ldots, \alpha(n)\}$ est transitive. On peut donc appliquer le théorème 3.3, et en déduire que toute propriété de graphe, au sens énoncé plus haut, a une largeur de seuil de l'ordre de $\frac{1}{\log n}$. Mais évidemment, de nombreuses propriétés auront un seuil plus étroit, comme par exemple la connexité (théorème 2.17).

Exemple 14 *Images aléatoires*
Pour les images aléatoires de 2.3, on peut convenir qu'une propriété d'image est invariante par l'action sur les pixels du groupe engendré par les deux translations élémentaires :

$$\tau_v : (i,j) \mapsto (i+1, j) \ ,$$

$$\tau_h : (i,j) \mapsto (i, j+1) \ ,$$

où on identifie l'ensemble des pixels à $(\mathbb{Z}/n\mathbb{Z})^2$ (conditions de bord toriques). Le groupe de permutations ainsi obtenu agit bien transitivement sur l'ensemble des pixels. On peut donc appliquer le théorème 3.16, et en déduire que toute propriété d'image a une largeur de seuil de l'ordre de $\frac{1}{\log n}$ au plus. Là encore, nous avons donné plusieurs exemples de propriétés dont le seuil est plus étroit.

Exemple 15 *La k-satisfiabilité*
Considérons deux types d'action sur les clauses (pour les définitions, voir le paragraphe 2.4). Tout d'abord, l'action de chaque élément σ de $\mathcal{S}_n$, qui consiste à remplacer chaque variable X_i par $X_{\sigma(i)}$ dans les clause. Ensuite, les transformations τ_k, qui consistent à échanger X_k et $\neg X_k$ dans les clauses. Voici deux exemples pour des clauses de longuer 3 :

$$(145).(\neg X_4 \vee X_6 \vee X_5) = (\neg X_5 \vee X_6 \vee X_4) \ ,$$

$$\tau_4.(\neg X_5 \vee X_6 \vee X_4) = (\neg X_5 \vee X_6 \vee \neg X_4) \ .$$

Le groupe engendré par ces éléments agit transitivement sur l'ensemble des clauses de longueur k, et on le fait agir par l'intermédiaire des clauses sur $E_{n,k}$. On peut donc appliquer le théorème 3.16, et en déduire que toute propriété invariante sous l'action du groupe que l'on vient de décrire a une largeur de seuil en $\frac{1}{\log n}$.

Si on compare le théorème 3.16 à la proposition 2.11, on voit qu'en passant du groupe $\mathcal{S}_n$ tout entier à un sous-groupe de $\mathcal{S}_n$ dont on requiert seulement qu'il soit transitif sur $\{1,\dots,n\}$, la largeur de seuil s'en trouve grandement affectée : elle est au plus de $\frac{1}{\sqrt{n}}$ dans le premier cas, et au plus de $\frac{1}{\log n}$ dans le second. Deux questions se posent alors naturellement :

- L'ordre de grandeur $O(\frac{1}{\log n})$ est-il optimal ? L'exemple 12 ci-dessus montre que pour la propriété M_n de contenir un run de longueur $\log n$, la largeur du seuil est effectivement en $O(\frac{1}{\log n})$. Mais nous avons donné de nombreux exemples de seuils beaucoup plus étroits.
- Peut-on espérer obtenir une largeur de seuil entre $\frac{1}{\log n}$ et $\frac{1}{\sqrt{n}}$ en regardant une propriété invariante sous l'action d'un groupe qui est "grand", mais pas $\mathcal{S}_n$ tout entier ?

Un début de réponse est apporté par le théorème 3.21, dû à Bourgain et Kalai [9], qui relie les largeurs de seuil à une caractéristique du groupe, définie comme suit. Soit G un groupe de permutations agissant transitivement sur $\{1,\dots,n\}$. Pour tout $k \in \{1,\dots,n\}$ définissons $\phi(k)$ par :

$$\phi(k) = \phi_G(k) = \inf_{\substack{S\subset\{1,\dots,n\}\\ |S|=k}} \log\left(|\{\sigma(S) \text{ t.q. } \sigma \in G\}|\right) .$$

Pour tout $\tau > 0$, posons alors :

$$a_\tau(G) = \sup\left\{\phi(k) \text{ t.q. } \phi(k) > k^{1+\tau}\right\} .$$

Théorème 3.21 *Soit G un groupe de permutations agissant transitivement sur $\{1,\dots,n\}$. Pour tout $\tau > 0$, il existe une constante $c_\tau > 0$ telle que, pour toute propriété monotone A de $\{0,1\}^n$ invariante sous G :*

$$\frac{d\mu_p(A)}{dp} > c_\tau a_\tau(G)\mu_p(A)(1-\mu_p(A)) ,$$

pourvu que $p(1-p)$ reste éloigné de 0 au sens où :

$$\log\left(\frac{1}{p(1-p)}\right) \le C^{te} \log\log n .$$

Par conséquent, la largeur du seuil pour une telle propriété A est au plus :

$$p_{1-\varepsilon} - p_\varepsilon \le \frac{2\log\frac{1-\varepsilon}{\varepsilon}}{c_\tau a_\tau(G)} .$$

Ce résultat leur permet, dans le même article [9], de garantir des largeurs de seuil un peu plus fines que $\frac{1}{\log n}$ à partir de considérations algébriques sur le groupe G. Par exemple, on obtient, dans le cas des graphes, que pour tout $\tau > 0$, il existe une constante $C_\tau > 0$ telle que pour toute propriété de graphe croissante, la largeur du seuil soit majorée par $\frac{\log\frac{1-\varepsilon}{\varepsilon}}{C_\tau(\log n)^{2-\tau}}$. Ceci est presque fin pour ce qui concerne les graphes, puisqu'on peut montrer que la propriété de contenir une clique (graphe complet) de taille $\lfloor\log n\rfloor$ admet un seuil de largeur $\frac{1}{(\log n)^2}$.

3.4 Coarse ou sharp

Revenons un instant sur les exemples des sections 2.1, 2.2, 2.3 et 2.4. Dans certains cas, les résultats 2.11, 3.16 et 3.21 réunis sont précis, et dans d'autres cas pas du tout. Citons quelques cas où ces résultats sont (relativement) précis :

- Si A est la propriété de E_n : "$\sum_{i=1}^n x_i \ge \frac{n}{2}$", alors le seuil de A se situe en $\frac{1}{2}$, et est en effet de largeur $\frac{1}{\sqrt{n}}$, comme on peut le voir en appliquant le théorème central limite à $\sum_{i=1}^n x_i$. La proposition 2.11 donne le bon ordre de grandeur.
- Si A est la propriété "contenir un run de longueur $\log n$", le seuil de A se situe en e^{-1}, et est de largeur $\frac{1}{\log n}$. C'est l'ordre de grandeur prévu par le théorème 3.16.
- Si A est la propriété "contenir une clique de taille $\lfloor\log n\rfloor$", elle admet un seuil de largeur $\frac{1}{(\log n)^2}$. Le théorème 3.21 donne "presque" le bon ordre de grandeur.

Voici des cas où ces théorèmes ne sont pas précis :

- Si A est la propriété de E_n : "$\sum_{i=1}^n x_i \geq 1$", alors $\mu_p(A) = 1 - (1-p)^n$, la fonction seuil de A tend vers 0 :
$$p_\varepsilon(n) \sim \frac{\log \frac{1}{1-\varepsilon}}{n} .$$
Elle est du même ordre que la largeur du seuil :
$$p_{1-\varepsilon} - p_\varepsilon \sim \frac{\log \frac{1-\varepsilon}{\varepsilon}}{n} .$$
- Si A est la propriété de E_n : "$\sum_{i=1}^n x_i \geq n$", alors $\mu_p(A) = p^n$, le seuil de A est proche de 1 :
$$p_\varepsilon(n) \sim 1 - \frac{\log \frac{1}{\varepsilon}}{n} ,$$
Dans ce cas, $1 - p_\varepsilon$ est du même ordre que la largeur du seuil :
$$p_{1-\varepsilon} - p_\varepsilon \sim \frac{\log \frac{1-\varepsilon}{\varepsilon}}{n} .$$
- Si A est la propriété "ne pas être stable", pour une configuration du graphe cyclique à n sommets,
$$p_\varepsilon(n) \sim \sqrt{\frac{\log \frac{1}{1-\varepsilon}}{n}} ,$$
du même ordre que la largeur de seuil.
- Si A est la propriété "contenir un triangle", pour les graphes aléatoires, alors (cf. [101], p. 296) :
$$p_\varepsilon(n) \sim \frac{\left(6 \log \frac{1}{1-\varepsilon}\right)^{\frac{1}{3}}}{n} ,$$
du même ordre que la largeur de seuil.
- Si A est la propriété de connexité pour les graphes aléatoires, alors (théorème 2.17) :
$$p_\varepsilon(n) \sim \frac{\log n}{n} + \frac{\log(1/\log(1/\varepsilon))}{n} .$$
Dans ce cas, la largeur du seuil est petite devant la fonction seuil.
- Si A est la propriété d'image "contenir une vignette incluse de taille $m \times m$", la proposition 2.21 nous donne :
$$p_\varepsilon(n) \sim \left(\frac{a \log \frac{1}{1-\varepsilon}}{8n^2}\right)^{\frac{1}{b}} ,$$
et la fonction seuil est encore du même ordre que la largeur de seuil.
- Enfin, pour la k-sat, on peut montrer que le seuil est également proche de 0 (la fonction seuil est n^{1-k}). Par contre, on verra que la largeur du seuil est plus petite que la fonction seuil.

On observe donc que les théorèmes cités jusqu'ici ont tendance à être précis lorsque la localisation du phénomène est loin de 0 et 1, et qu'ils sont souvent mauvais lorsqu'elle tend vers 0 ou 1. Tellement qu'il suffit en général de regarder la vitesse à laquelle $p_\varepsilon(n)$ tend vers 0 pour trouver une majoration de la largeur du seuil bien meilleure que celle donnée par la proposition 2.11 et les théorèmes 3.16 et 3.21. Que peut-on dire de la largeur du seuil d'une propriété dont la fonction seuil tend vers 0 ? Comme le montrent les exemples ci-dessus, il peut se faire que la largeur du seuil soit strictement inférieure à la fonction seuil, comme pour la connexité des graphes aléatoires. Mais dans de nombreux cas (apparition d'un sous-graphe ou d'une vignette par exemple), les deux fonctions sont du même ordre. Or pour reprendre l'argumentaire développé en 2.1, la fonction seuil est l'échelle de localisation du phénomène, et on s'attend à ce que les fluctuations aient lieu à une échelle inférieure. Pour distinguer les deux types de comportement, on introduit la définition suivante.

Définition 3.22 *Soit $(A_n)_{n\in\mathbb{N}} \subset E_n$ une suite de propriétés croissantes dont la fonction seuil tend vers 0. On dit que $(A_n)_{n\in\mathbb{N}}$ admet un seuil* fin *(sharp) si et seulement si pour tout $\varepsilon \in]0, \frac{1}{2}]$:*

$$\lim_{n\to\infty} \frac{p_{1-\varepsilon} - p_\varepsilon}{p_{1/2}} = 0 \ .$$

On dit qu'elle admet un seuil grossier *(coarse) si pour tout $\varepsilon \in]0, \frac{1}{2}]$:*

$$\liminf_{n\to\infty} \frac{p_{1-\varepsilon} - p_\varepsilon}{p_{1/2}} > 0 \ .$$

Si on dispose d'un développement asymptotique de p_ε en fonction de n, il est facile de "voir" sur ce développement si le seuil est fin ou non. En effet, si le terme dominant dépend de ε, alors le seuil est grossier. Dans le cas contraire, il est fin. Prenons l'exemple de A_T, la propriété de graphe "contenir un triangle". On peut montrer un résultat d'approximation poissonnienne pour la variable comptant le nombre de triangles inclus dans le graphe (cf. [101], p. 296), et on a notamment :

$$\text{Soit } p(n) \text{ tel que } \binom{n}{3} p^3(n) = \lambda \ ,$$

$$\lim_{n\to\infty} \mu_{p(n)}(A_T) = 1 - e^{-\lambda} \ .$$

Ceci permet d'obtenir le développement asymptotique :

$$p_\varepsilon(n) = \frac{\left(6 \log \frac{1}{1-\varepsilon}\right)^{\frac{1}{3}}}{n} + o(\frac{1}{n}) \ .$$

On en déduit :

$$\frac{p_{1-\varepsilon} - p_\varepsilon}{p_{1/2}} = \frac{(\log \frac{1}{\varepsilon})^{\frac{1}{3}} - \left(\log \frac{1}{1-\varepsilon}\right)^{\frac{1}{3}}}{(\log 2)^{\frac{1}{3}}} + o(1) \ ,$$

et le seuil est grossier. On pourrait voir de la même manière que le seuil de la propriété "contenir une copie de H" est grossier, dès que H est un graphe de taille fixée. On peut montrer que l'union d'un nombre fixé de propriétés admettant des seuils grossiers admet également un seuil grossier. Par conséquent, si on fixe un nombre entier k, et une liste $\mathcal{L}$ de graphes ayant un nombre d'arêtes plus petit que k, la propriété "contenir une copie d'un graphe de $\mathcal{L}$" aura un seuil grossier. On peut alors se demander si toute propriété de graphe ayant un seuil grossier est de ce type. La réponse est à peu près affirmative comme le montre le théorème qui suit, dû à Friedgut (théorème 1.1 dans [43]). Dans cet énoncé, on appelle graphe minimal d'une propriété B_n un graphe qui la satisfait et dont aucun sous-graphe ne la vérifie. On notera $\|B_n\|$ la plus grande taille (i.e. le plus grand nombre d'arêtes) des graphes minimaux de B_n. Remarquons que pour une propriété B_n croissante, B_n et l'ensemble des graphes G contenant un graphe minimal de B_n sont identiques.

Théorème 3.23 *Il existe une fonction $k(\varepsilon, c)$, telle que pour tout $c > 0$, pour tout entier n, toute propriété de graphe A_n telle que $p\frac{d\mu_p(A_N)}{dp} \leq c$, pour tout $\varepsilon > 0$, il existe une propriété de graphe B_n telle que $\|B_n\| \leq k(\varepsilon, c)$ et $\mu_p(A_n \Delta B_n) \leq \varepsilon$ (Δ désigne la différence symétrique).*

Pour relier ce résultat à la définition 3.22, il suffit d'observer que la dérivée de $\mu_p(A_N)$ est l'inverse de la dérivée de la fonction réciproque, qui à α associe p_α. Majorer $p\frac{d\mu_p(A_N)}{dp}$ revient donc à minorer $\frac{1}{p_\alpha}\frac{dp_\alpha}{d\alpha}$, ce qui entraîne un seuil grossier.

Un résultat en tout point analogue (le théorème 5.1 de l'article [43]) peut être démontré pour les propriétés concernant les formules de la k-sat (cf. section 2.4 et l'exemple 15 de la section 3.3). Il permet de prouver que le seuil de la k-sat est fin, en montrant que la propriété de ne pas être satisfiable ne peut pas être approchée de la manière décrite dans le théorème 3.23.

4 ... a une probabilité proche de 0 ou 1.

4.1 Inégalités classiques

Comme le montre le corollaire 3.3, la clé de la démonstration d'une convergence p.s. est l'étude d'une série de probabilités :

$$\sum_{n \in \mathbb{N}} Prob[|X_n - X| > t] \,.$$

Les probabilités du type $Prob[Y > t]$ ou $Prob[Y < t]$ se rencontrent très souvent. On les appelle probabilités de déviations, de queues, ou d'ailes comme disent plus joliment les canadiens. Les ordres de grandeur exponentiels des probabilités de déviations que nous allons étudier dans cette partie sont des cas particuliers d'une théorie générale, celle des *grandes déviations* (voir [21, 23, 104]).

Les probabilités de déviations ne sont en général pas faciles à calculer et on cherche à les majorer par des quantités plus accessibles. De nombreuses inégalités plus ou moins sophistiquées ont été mises au point pour cela, et beaucoup de livres présentent ces outils de base (voir par exemple les chapitres 2 et 3 de [24] sur les inégalités de concentration utilisées en statistique non-paramétrique).

La plus simple est l'inégalité de Markov :

Proposition 4.1 *Soit Y une variable aléatoire presque sûrement positive, admettant une espérance. Pour tout $t > 0$:*

$$Prob[Y > t] \leq \frac{1}{t} \mathbb{E}[Y] \,. \tag{4.1}$$

Démonstration : Elle consiste à minorer $\mathbb{E}[Y]$, assez brutalement.

$$\mathbb{E}[Y] = \int_{\mathbb{R}} y \, P_Y(dy) \geq \int_t^{+\infty} t P_Y(dy) = t \, Prob[Y > t] \,.$$

□

Pour $Y = (X - \mathbb{E}[X])^2$, (4.1) est l'inégalité de Bienaymé-Chebyshev. L'inégalité de Chernov consiste à appliquer (4.1) à la *fonction génératrice des moments*. Si X est une variable aléatoire, on appelle fonction génératrice des moments de X (ou plutôt de sa loi), la fonction qui à s associe $\mathbb{E}[\exp(sX)]$, si cette espérance existe. On utilise aussi la *transformée de Laplace* $\mathbb{E}[\exp(-sX)]$. Pour éviter la confusion avec les fonctions génératrices en combinatoire, nous parlerons systématiquement de transformée de Laplace, même pour désigner $\mathbb{E}[\exp(sX)]$.

Proposition 4.2 *Soit X une variable aléatoire. Pour tout $t > 0$:*

$$Prob[X > t] \leq \inf_{s > 0} \mathbb{E}[\exp(s(X - t))] \,. \tag{4.2}$$

Démonstration : Pour tout $s > 0$:

$$Prob[X > t] = Prob[X - t > 0] = Prob[\exp(s(X - t)) > 1] \,.$$

Il suffit alors d'appliquer l'inégalité de Markov à $Y = \exp(s(X - t))$. □

Il peut paraître étonnant qu'une inégalité aussi rudimentaire donne des résultats précis. C'est pourtant le cas. Nous illustrons d'abord l'inégalité de Chernov sur deux exemples élémentaires, la loi binomiale et la loi de Poisson.

Corollaire 4.3 *Soit X une variable aléatoire, suivant la loi binomiale $\mathcal{B}(n,p)$. Pour tout $b \in]0,1[$, posons :*

$$h(p,b) = \left(\frac{1-p}{1-b}\right)^{1-b} \left(\frac{p}{b}\right)^{b} \,.$$

Alors :

$$Prob[X > nb] \leq h^n(p,b) \quad si \quad b > p \;, \tag{4.3}$$

et :

$$Prob[X < nb] \leq h^n(p,b) \quad si \quad b < p \;. \tag{4.4}$$

Démonstration : La transformée de Laplace de la loi binomiale $\mathcal{B}(n,p)$ s'écrit :

$$\mathbb{E}[\exp(sX)] = ((1-p) + pe^s)^n \;.$$

Le minimum de $\mathbb{E}[\exp(s(X-nb))]$ est atteint pour $e^s = \frac{b(1-p)}{p(1-b)}$, ce qui conduit à la première inégalité. On obtient la seconde en changeant le signe. □

Evidemment les inégalités du corollaire 4.3 ne sont pas très lisibles. On peut les améliorer en posant $b = p \pm \frac{c}{\sqrt{n}}$. Quelques manipulations fastidieuses mais élémentaires conduisent au théorème 2.10.

Voici l'analogue pour la loi de Poisson.

Corollaire 4.4 *Soit X une variable aléatoire, suivant la loi de Poisson $\mathcal{P}(\lambda)$. Alors :*

$$Prob[X > b] \leq \left(\frac{\lambda}{b}\right)^b e^{b-\lambda} \quad si \quad b > \lambda \;, \tag{4.5}$$

et :

$$Prob[X < b] \leq \left(\frac{\lambda}{b}\right)^b e^{b-\lambda} \quad si \quad b < \lambda \;. \tag{4.6}$$

Démonstration : La transformée de Laplace de la loi de Poisson $\mathcal{P}(\lambda)$ est la suivante.

$$\mathbb{E}[\exp(sX)] = \exp(\lambda(e^s - 1)) \;.$$

Le minimum de $\mathbb{E}[\exp(s(X-b))]$ est atteint pour $e^s = \frac{b}{\lambda}$, ce qui conduit à la première inégalité. On obtient la seconde en changeant le signe. □

Ici encore, (4.5) et (4.6) ne sont pas forcément les plus faciles à utiliser. On est en général amené à transformer ces inégalités, quitte à les affaiblir un peu. Pour $b > \lambda$, on obtient par exemple :

$$Prob[X > b] \leq \exp\left(-\frac{b}{2}\left(1-\frac{\lambda}{b}\right)^2\right) \;. \tag{4.7}$$

Pour passer de (4.5) à (4.7), il suffit d'observer que pour $x > 1$:

$$\log(x) - 1 + \frac{1}{x} > \frac{1}{2}\left(1 - \frac{1}{x}\right)^2 \;.$$

La transformée de Laplace d'une somme de variables aléatoires indépendantes est le produit des transformées de Laplace des composantes. Il n'est donc pas surprenant que la technique de changement de variable exponentiel donne de bons résultats pour les inégalités portant sur des sommes de variables. Il existe de très nombreuses variantes de cette technique (voir par exemple [91, 92]). Nous présentons ci-dessous une des plus utilisées en probabilités discrètes, l'inégalité de Hoeffding.

Théorème 4.5 *Soient $X_1, X_2, \ldots, X_n$ des variables aléatoires indépendantes à valeurs dans $[a_i, b_i]$. Alors, pour tout $t > 0$:*

$$Prob\left[\sum_{i=1}^{n}(X_i - \mathbb{E}[X_i]) \geq t\right] \leq \exp\left(-\frac{2t^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right) \;, \tag{4.8}$$

et :

$$Prob\left[\sum_{i=1}^{n}(X_i - \mathbb{E}[X_i]) \leq -t\right] \leq \exp\left(-\frac{2t^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right) . \qquad (4.9)$$

Dans le cas particulier où les X_i suivent toutes la loi de Bernoulli de paramètre p, leur somme suit la loi binomiale $\mathcal{B}(n,p)$. Si on applique (4.8) et (4.9) avec $t = c\sqrt{n}$, on retrouve les inégalités du théorème 2.10.

Démonstration : Remarquons tout d'abord que pour tout $i = 1, \ldots, n$,

$$-\frac{b_i - a_i}{2} \leq X_i - \left(a_i + \frac{b_i - a_i}{2}\right) \leq \frac{b_i - a_i}{2} .$$

Par conséquent, la variance de X_i est majorée par $\frac{1}{4}(b_i - a_i)^2$, et ce quelle que soit la loi de probabilité de X_i. L'égalité ne peut avoir lieu que si X_i prend les valeurs a_i et b_i avec probabilité $\frac{1}{2}$.

Nous examinons maintenant la transformée de Laplace, ou plutôt son logarithme : pour une variable aléatoire Y quelconque, nous notons ψ_Y la fonction définie par :

$$\psi_Y(s) = \log \mathbb{E}\left[e^{sY}\right] .$$

Cette fonction, que nous appellerons "transformée de log-Laplace", apparaîtra plusieurs fois dans les sections suivantes. Les deux premières dérivées de ψ_Y se calculent aisément (les variables que nous considérons étant bornées, les problèmes de dérivation des intégrales ne se posent pas). On trouve :

$$\psi'_Y(s) = \mathbb{E}\left[Y\frac{\exp(sY)}{\mathbb{E}[\exp(sY)]}\right] ,$$

et :

$$\psi''_Y(s) = \mathbb{E}\left[Y^2\frac{\exp(sY)}{\mathbb{E}[\exp(sY)]}\right] - \left(\mathbb{E}\left[Y\frac{\exp(sY)}{\mathbb{E}[\exp(sY)]}\right]\right)^2 .$$

Les dérivées $\psi'_Y(s)$ et $\psi''_Y(s)$ apparaissent donc respectivement comme l'espérance et la variance de Y relatives à une nouvelle loi de probabilité, obtenue en multipliant l'ancienne par $\exp(sY)/\mathbb{E}[\exp(sY)]$.

En considérant la transformée de log-Laplace de $Y_i = X_i - \mathbb{E}[X_i]$, dont la variance est bornée, on obtient :

$$\psi''_{Y_i}(s) \leq \frac{(b_i - a_i)^2}{4} . \qquad (4.10)$$

Comme $\psi'_{Y_i}(0) = \psi_{Y_i}(0) = 0$, en intégrant deux fois (4.10) entre 0 et λ, on trouve :

$$\psi_{Y_i}(s) \leq \frac{s^2}{8}(b_i - a_i)^2 .$$

Considérons maintenant $S_n = \sum_{i=1}^{n}(X_i - \mathbb{E}[X_i])$. Grâce à l'hypothèse d'indépendance, sa transformée de log-Laplace est la somme de celles des Y_i. Donc :

$$\psi_{S_n}(s) \leq \frac{s^2}{8}\sum_{i=1}^{n}(b_i - a_i)^2 .$$

Nous pouvons maintenant appliquer l'inégalité de Chernov à S_n :

$$Prob[S_n > t] \leq \mathbb{E}[\exp(s(S_n - t))] \leq \exp\left(\frac{s^2}{8}\sum_{i=1}^{n}(b_i - a_i)^2 - st\right) .$$

Le minimum en s est atteint pour $s = (4t)/\sum_{i=1}^{n}(b_i - a_i)^2$ et (4.8) en découle. L'inégalité (4.9) s'obtient en remplaçant X_i par $-X_i$. □

L'inégalite de Hoeffding est fine lorsque la loi de chaque X_i est concentrée sur les valeurs extrêmes a_i et b_i. Elle n'est plus fine lorsque la variance des X_i est faible devant la largeur des intervalles. L'inégalité de Bernstein ci-dessous corrige ce défaut.

Théorème 4.6 *Soient $X_1, X_2, \ldots, X_n$ des variables aléatoires indépendantes d'espérance nulle et c une constante positive telle que pour tout $i = 1, \ldots, n$, $|X_i| \leq c$. Notons σ^2 la variance moyenne des X_i :*

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^{n} Var[X_i] \, .$$

Alors, pour tout $t > 0$:

$$Prob\left[\sum_{i=1}^{n} X_i \geq t\right] \leq \exp\left(-\frac{t^2}{2n\sigma^2 + ct}\right) \, , \tag{4.11}$$

et :

$$Prob\left[\sum_{i=1}^{n} X_i \leq -t\right] \leq \exp\left(-\frac{t^2}{2n\sigma^2 + ct}\right) \, . \tag{4.12}$$

Démonstration : Il s'agit à nouveau d'appliquer l'inégalité de Chernov, et donc de majorer la transformée de Laplace. Pour une variable aléatoire X d'espérance nulle, bornée en valeur absolue par c, et de variance σ^2, on a :

$$\begin{aligned}
\mathbb{E}[e^{sX}] &= \mathbb{E}\left[1 + sX + \sum_{k=2}^{\infty} \frac{s^k X^k}{k!}\right] \\
&= 1 + \mathbb{E}\left[\sum_{k=2}^{\infty} \frac{s^k X^k}{k!}\right] \\
&\leq 1 + \frac{\sigma^2}{c^2} \sum_{k=2}^{\infty} \frac{(sc)^k}{k!} \\
&= 1 + \frac{\sigma^2}{c^2}(e^{sc} - 1 - sc) \\
&\leq \exp\left(\frac{\sigma^2}{c^2}(e^{sc} - 1 - sc)\right) \, .
\end{aligned}$$

En utilisant l'indépendance des X_i, on obtient la majoration suivante pour la transformée de Laplace de leur somme.

$$\mathbb{E}\left[\exp(sS_n)\right] \leq \exp\left(\frac{n\sigma^2}{c^2}(e^{sc} - 1 - sc)\right) \, .$$

On applique alors l'inégalité de Chernov.

$$Prob[S_n > t] \leq \mathbb{E}[\exp(s(S_n - t))] \leq \exp\left(\frac{n\sigma^2}{c^2}(e^{sc} - 1 - sc) - st\right) \, .$$

La minimisation en s est un peu plus délicate que précédemment. Elle conduit à (4.11) après quelques manipulations élémentaires. Comme toujours, (4.12) s'obtient en remplaçant X_i par $-X_i$. □

Les deux théorèmes précédents portaient sur des sommes de variables indépendantes. Or nous avons déjà eu l'occasion de constater que la concentration de la mesure dans les espaces produits va bien au delà des sommes de coordonnées. Le principe du changement de variable exponentiel qui fonde l'inégalité de Chernov s'applique non seulement à des sommes de variables indépendantes, mais plus généralement à des martingales. L'inégalité de McDiarmid (théorème 4.7) en est une illustration. Elle porte sur une fonction quelconque d'un ensemble de variables indépendantes, sous une hypothèse de régularité relativement peu contraignante en ce qui concerne les applications aux probabilités discrètes. On trouve dans le chapitre 8 de [101] un cas particulier, appliqué sous le nom d'inégalité d'Azuma à la concentration du nombre chromatique des graphes aléatoires.

Théorème 4.7 *Soient $X_1, \ldots, X_n$ des variables aléatoires indépendantes à valeurs dans E, et f une application de E dans $I\!R$ vérifiant la condition suivante.*

$$\forall i = 1, \ldots, n\,,\ \forall x_1, \ldots, x_n, x_i' \in E\,,$$
$$|f(x_1, \ldots, x_n) - f(x_1, \ldots, x_{i-1}, x_i', x_{i+1}, \ldots, x_n)| \leq c_i\ .$$

Notons Y la variable aléatoire $Y = f(X_1, \ldots, X_n)$. Alors pour tout $t > 0$:

$$Prob[Y - I\!E[Y] \geq t] \leq \exp\left(-\frac{2t^2}{\sum_{i=1}^n c_i^2}\right)\ , \tag{4.13}$$

et :

$$Prob[Y - I\!E[Y] \leq -t] \leq \exp\left(-\frac{2t^2}{\sum_{i=1}^n c_i^2}\right)\ , \tag{4.14}$$

Démonstration : Notons $M_0 = I\!E[Y]$, et pour $i = 1, \ldots, n$:

$$M_i = I\!E[Y \,|\, X_1, \ldots, X_i] \text{ et } V_i = M_i - M_{i-1}\ .$$

On vérifie immédiatement que $(M_i)_{i=0,\ldots,n}$ est une martingale relativement à la filtration naturelle de $(X_1, \ldots, X_n)$:

$$I\!E[M_i \,|\, X_1, \ldots, X_{i-1}] = I\!E[Y \,|\, X_1, \ldots, X_{i-1}] = M_{i-1}\ .$$

Nous souhaitons appliquer l'inégalité de Chernov à :

$$Y - I\!E[Y] = \sum_{i=1}^n V_i\ .$$

Nous devons pour cela majorer la transformée de Laplace de la somme des V_i.

$$\begin{aligned}
I\!E\left[\exp\left(s\sum_{i=1}^n V_i\right)\right] &= I\!E\left[\prod_{i=1}^n e^{sV_i}\right] \\
&= I\!E\left[I\!E\left[\prod_{i=1}^n e^{sV_i} \,|\, X_1, \ldots, X_{n-1}\right]\right] \\
&= I\!E\left[I\!E\left[\left(\prod_{i=1}^{n-1} e^{sV_i}\right) I\!E\left[e^{sV_n} \,|\, X_1, \ldots, X_{n-1}\right]\right]\right] \\
&\ \vdots \\
&= I\!E\left[\prod_{i=1}^n I\!E\left[e^{sV_i} \,|\, X_1, \ldots, X_{i-1}\right]\right]\ .
\end{aligned}$$

Il reste donc à majorer les espérances conditionnelles $I\!E\left[e^{sV_i} \,|\, X_1, \ldots, X_{i-1}\right]$. Or on peut déduire de la condition de Lipschitz sur f que chacune des variables V_i est bornée en valeur absolue par c_i. Le raisonnement que nous avons effectué dans la démonstration de l'inégalité de Hoeffding s'applique donc ici, en remplaçant a_i par $-c_i$ et b_i par c_i. On en déduit donc :

$$I\!E\left[e^{sV_i} \,|\, X_1, \ldots, X_{i-1}\right] \leq \exp\left(\frac{s^2 c_i^2}{2}\right)\ .$$

On termine alors la démonstration en appliquant l'inégalité de Chernov, comme pour le théorème 4.5.
□

4.2 Inégalités de Talagrand

Talagrand, notamment à la suite des travaux de Milman et Gromov, a largement contribué à développer et diffuser dans la communauté mathématique la notion de "concentration de la mesure", si bien que son nom lui est immédiatement associé. Ses travaux se sont situés dans la lignée du fondateur de cette notion (Milman), et nous allons dans cette section en présenter les caractéristiques majeures (pour une présentation plus précise et très agréable, voir [107]). L'idée de la concentration de la mesure sera d'abord liée aux élargissements d'ensembles. De manière approximative, on pensera que la concentration de la mesure a lieu dans un espace de probabilité si, dès qu'un ensemble A a pour mesure $\frac{1}{2}$, la quasi-totalité de l'espace est située à une distance "faible" de A.

Exprimons cela de manière plus rigoureuse. Soit $\mathcal{X}$ un espace muni d'une distance d, et d'une probabilité P. On désigne par A_t le "t-élargissement" de A, c'est à dire l'ensemble :

$$A_t = \{x \in \mathcal{X}\ ,\ d(x, A) \leq t\}\ .$$

La *fonction de concentration* $\alpha(P, t)$ est définie pour $t \geq 0$ par :

$$\alpha(P,t) = \inf \left\{ \alpha > 0 \text{ t.q. } P(A) \geq \frac{1}{2} \Longrightarrow 1 - P(A_t) \leq \alpha \right\}\ .$$

Dans de nombreux cas importants, $\alpha(P, t)$ décroît très rapidement avec t. Plus cette décroissance est rapide, plus on considèrera la mesure comme "concentrée". Pour voir le rapport avec les fonctions de $\mathcal{X}$ dans $\mathbb{R}$, considérons une telle fonction f, et notons M_f une médiane de f. Si on pose $A = \{f \leq M_f\}$, la probabilité de A est supérieure ou égale à $\frac{1}{2}$, par définition de M_f. Supposons maintenant que f soit 1-lipschitzienne :

$$\forall x, y \in \mathcal{X}\ ,\ |f(x) - f(y)| \leq d(x, y)\ ,$$

c'est le côté "raisonnable" de la fonction. Alors,

$$x \in A_t \Longrightarrow f(x) \leq t + M_f\ .$$

Et ainsi,

$$P(f > M_f + t) \leq 1 - P(A_t) \leq \alpha(P, t)\ .$$

On peut faire le même raisonnement avec $-f$, et donc :

$$P(|f - M_f| > t) \leq 2\alpha(P, t)\ .$$

On obtient ainsi une "inégalité de concentration de f autour de sa médiane". Si on avait supposé f L-lipschitzienne, on aurait obtenu :

$$P(|f - M_f| > t) \leq 2\alpha(P, \frac{t}{L})\ .$$

Nous n'avons pas encore parlé d'espace produit, ni d'indépendance. Pour l'instant, il se trouve que la quasi-totalité des inégalités de concentration démontrées l'ont été pour des espaces produits, munis d'une probabilité produit (sauf pour les résultats de Marton, [83], qui concernent une dépendance markovienne), mais les définitions nécessaires au cadre de la concentration sont susceptibles d'intéresser tout espace disposant d'une distance et d'une probabilité. Pour ce qui concerne les distances, c'est un aspect qui a été très approfondi par Talagrand : il a développé des résultats de concentration avec de nombreux exemples de distances différentes (on en verra un exemple plus loin : celui de la "distance convexe"). L'intérêt est bien sûr que pour prouver ensuite une inégalité de concentration pour une fonction, il faut qu'elle se comporte bien vis-à-vis de la distance. Multiplier les types de distance revient à multiplier les types de fonctions, ou plutôt les types de dépendance de ces fonctions à leurs variables. Pour bien cerner cela, donnons plusieurs exemples de concentration.

Le premier concerne la sphère S_n de $\mathbb{R}^{n+1}$, munie de la distance géodésique, et de la mesure de Haar normalisée Q_n. On a dans ce cas :

$$\alpha(Q_n, t) \leq \left(\frac{\pi}{8}\right)^{\frac{1}{2}} e^{-\frac{n-1}{2}t^2}\ . \tag{4.15}$$

Le second concerne $\mathbb{R}^n$, muni de la mesure gaussienne standard γ_n et de la distance euclidienne. On obtient :

$$\alpha(\gamma_n, t) \leq \frac{1}{2} e^{-\frac{t^2}{2}} . \tag{4.16}$$

Enfin, le dernier concerne un produit quelconque d'espaces de probabilité $(\mathcal{X}_i, P_i)$, muni de la probabilité produit $\mathbb{P}_n = \bigotimes_{i=1}^n P_i$ et de la distance de Hamming $d_H(x,y) = \sum_{i=1}^n \mathbb{1}_{x_i \neq y_i}$. On a :

$$\alpha(\mathbb{P}_n, t) \leq 2 e^{-\frac{t^2}{n}} . \tag{4.17}$$

On voit dans ces trois exemples que la concentration de la mesure est un terme très vague, et qu'il faut toujours se demander "Pour quel espace ? Pour quelle distance ? Pour quelle probabilité ?". Par exemple, on peut appliquer (4.17) au cadre du deuxième exemple pour donner une deuxième fonction de concentration, sur le même espace muni de la même mesure : $(\mathbb{R}^n, \gamma_n)$. Mais les distances sont différentes et une fonction lipschitzienne pour la distance euclidienne ne l'est pas forcément pour la distance de Hamming.

Nous n'avons pas encore expliqué comment trouver des inégalités telles que (4.15), (4.16), et (4.17). C'est bien sûr là que réside la difficulté. Historiquement, ces inégalités ont été démontrées dans cet ordre, et (4.15) est réellement un résultat fondateur de Borel. L'idée de la démonstration est de nature isopérimétrique : un résultat de Paul Lévy affirme que sur la sphère, ce sont les "calottes" (intersections de la sphère avec un demi-espace affine) qui ont le plus petit élargissement possible pour un volume donné. C'est à dire que, si Q_n est la mesure de Haar sur la sphère unité S_n de $\mathbb{R}^{n+1}$, on a :

$$\forall C \text{ calotte}, \forall A \subset S_n \, , \; Q_n(A) = Q_n(C) \Longrightarrow \forall t > 0, Q_n(A_t) \geq Q_n(C_t) \, . \tag{4.18}$$

Grâce à un tel résultat, il suffit de calculer $Q_n(C_t)$ pour une demi-sphère C, pour en conclure l'inégalité (4.15). Pour se convaincre que l'équation (4.18) est bien de nature isopérimétrique, on peut définir une mesure de la surface de $A \subset S_n$ par :

$$Vol_n(\partial A) = \lim_{t \to 0} \frac{Q_n(A_t) - Q_n(A)}{t} \, .$$

Ainsi, l'équation (4.18) est équivalente à l'affirmation :

$$\forall A \subset S_n \, , \; Q_n(A) = Q_n(C) \Longrightarrow Vol_n(\partial A) \leq Vol_n(\partial C) \, .$$

qui est bien une inégalité isopérimétrique.

Remarquons aussi que (4.16) a été démontrée a partir de (4.15), en obtenant la mesure gaussienne sur $\mathbb{R}^n$ comme la limite de la projection de la mesure de Haar Q_{n+r} pour la sphère S_{n+r} sur un sous-espace de dimension n passant par l'origine, et en transposant le rôle joué par les calottes dans l'isopérimétrie sur la sphère en un rôle similaire, joué cette fois par les demi-espaces affines.

Un des inconvénients de cette méthode "isopérimétrique" est qu'elle fait appel à un résultat d'isopérimétrie (celui de Lévy), que ce type de résultat est déjà difficile à montrer dans des cas où la distance considérée est "classique", et qu'on ne sait absolument pas comment s'y prendre dès que l'on complique cette distance. C'est pourquoi la première preuve de (4.17), due à Talagrand, est de nature très différente. Nous n'expliquerons pas la méthode de Talagrand, car depuis, une autre méthode inventée sans doute par Cirel'son, Ibragimov et Sudakov [12], puis approfondie par Ledoux puis Massart (cf. [4, 7, 8, 77, 78, 84]), s'est finalement avérée capable de prouver beaucoup de résultats de Talagrand (et surtout ceux les plus utilisés) de manière plus facile à concevoir. Les outils de cette méthode sont les inégalités de Sobolev logarithmiques, nous y reviendrons dans la section suivante.

Pour finir cette section, nous allons donner un autre résultat de concentration, dû à Talagrand [106], très important pour l'utilisation qui en a été faite en combinatoire (voir entre autres, [103] et [106]). Définissons tout d'abord une sorte de distance à un ensemble :

Définition 4.8 *Soit $(\Omega, \mathcal{A}, \mu)$ un espace de probabilité, x un point de Ω^n et A un sous-ensemble de Ω^n. On définit la "distance convexe de x à A" par :*

$$d_T(x, A) = \sup_{\substack{(\alpha_i(x))_{i=1\ldots n}\geq 0 \\ \sum_{i=1}^n \alpha_i(x)^2 \leq 1}} \left\{ \inf_{y\in A} \sum_{i=1}^n \alpha_i(x) 1\!\!1_{x_i\neq y_i} \right\} .$$

L'avantage d'une telle distance est la possibilité de faire dépendre les poids de x. Le résultat de concentration pour cette distance est alors le suivant :

Théorème 4.9 *Pour tout $A \subset \Omega^n$, en notant P la mesure produit $\mu^{\otimes n}$, on a :*

$$\int_{\Omega^n} e^{\frac{1}{4}d_T^2(x,A)}\, dP(x) \leq \frac{1}{P(A)} , \tag{4.19}$$

et par conséquent,

$$P(d_T(x, A) \geq t) \leq \frac{e^{-\frac{t^2}{4}}}{P(A)} , \tag{4.20}$$

ce qui signifie :

$$P(A_t) \geq 1 - \frac{e^{-\frac{t^2}{4}}}{P(A)} .$$

On passe de (4.19) à (4.20) en effectuant une manipulation de type Chernov. N'ayant pas exactement affaire à une distance, l'argument permettant de passer à la concentration d'une fonction, même s'il est simple, mérite d'être détaillé. On pourra comparer la proposition 4.10 ci-dessous à l'inégalité de McDiarmid (théorème 4.7).

Proposition 4.10 *Soit F, de Ω^n dans $I\!R$, 1-lipschitzienne au sens suivant :*

$$\forall x \in \Omega^n, \exists (b_i(x))_{i=1,\ldots,n} \geq 0 \ \textit{tels que} \ \sum_{i=1}^n b_i(x)^2 \leq c^2 \ \textit{et} \ \forall y \in \Omega^n ,$$

$$F(x) \leq F(y) + \sum_{i=1}^n b_i(x) 1\!\!1_{x_i\neq y_i} , \tag{4.21}$$

alors, si M_F est une médiane de F pour P,

$$P(|F - M_F| \geq t) \leq 4e^{-\frac{t^2}{4c^2}} .$$

Démonstration : Notons :

$$A = \{x \in \Omega^n \text{ t.q. } F(x) \leq M_F\} ,$$
$$B = \{x \in \Omega^n \text{ t.q. } F(x) \geq M_F + t\} ,$$
$$C = \{x \in \Omega^n \text{ t.q. } d_T(x, A) \geq \frac{t}{c}\} .$$

Si on arrive à montrer que $B \subset C$, comme $P(A) \geq \frac{1}{2}$, on aura d'après le théorème 4.9 :

$$P(B) \leq P(C) \leq 2e^{-\frac{t^2}{4c^2}} .$$

Soit $x \in B$. D'après l'hypothèse sur F, il existe $(b_i(x))_{i=1,\dots,n} \geq 0$ tels que $\sum_{i=1}^n b_i(x)^2 \leq c^2$, et, pour tout $y \in A$,

$$\begin{aligned}
F(x) &\leq F(y) + \sum_{i=1}^n b_i(x) \mathbb{1}_{x_i \neq y_i} \\
&\leq M_F + c \sum_{i=1}^n \frac{b_i(x)}{\sqrt{\sum_{i=1}^n b_i^2(x)}} \mathbb{1}_{x_i \neq y_i} \\
&\leq M_F + c \inf_{y \in A} \sum_{i=1}^n \frac{b_i(x)}{\sqrt{\sum_{i=1}^n b_i^2(x)}} \mathbb{1}_{x_i \neq y_i} \\
&\leq M_F + c\, d_T(x, A) .
\end{aligned}$$

Donc :

$$d_T(x, A) \geq \frac{F(x) - M_F}{c} \geq \frac{t}{c} ,$$

et $x \in C$. On peut faire la même chose avec :

$$A' = \{x \in \Omega^n \text{ t.q. } F(x) \geq M_F\} ,$$

$$B' = \{x \in \Omega^n \text{ t.q. } F(x) \leq M_F - t\} ,$$

$$C' = \{x \in \Omega^n \text{ t.q. } d_T(x, A) \geq \frac{t}{c}\} ,$$

et on obtient que $P(B') \leq 2e^{-\frac{t^2}{4c^2}}$, ce qui finit de prouver la proposition. □

Appliquons maintenant ce résultat à un problème célèbre de combinatoire : le problème du voyageur de commerce, ou “Travelling Salesman Problem” (TSP). La présentation qui suit est celle de Steele [103] p. 124-125.
Soient $X_1, X_2, \dots, X_n$, n points tirés au hasard, indépendamment les uns des autres, dans le carré $[0,1]^2$. Le problème du voyageur de commerce est de relier ces n points et de rallier son point de départ (choisi parmi les n points) en parcourant la distance la plus courte possible. Nous noterons $L_n(x_1, \dots, x_n)$ la fonction désignant cette plus courte distance possible pour $x_1, \dots, x_n$ appartenant à $[0,1]^2$. La question qui nous préoccupe ici est celle de la concentration de la variable aléatoire $L_n(X_1, \dots, X_n)$. Le point crucial est de décider de poids $b_i(x)$ pour $x = (x_1, \dots, x_n) \subset [0,1]^2$. L'objectif est évidemment de satisfaire l'inégalité (4.21), tout en conservant $\sum_{i=1}^n b_i(x)^2 \leq c^2$. Pour notre problème, il existe une heuristique qui nous indique comment les choisir : celle dite de la “space-filling curve”. Imaginons en effet une courbe continue remplissant $[0,1]^2$, c'est à dire une fonction continue et surjective $\psi : [0,1] \to [0,1]^2$. On peut imaginer que plus cette courbe est “lisse”, plus elle aura tendance à passer par les points $(x_1, \dots, x_n)$ “sans perdre de temps”. On pourra alors définir un trajet parmi ces points dans l'ordre où la fonction ψ les visite. Un fait important est alors qu'il existe des fonctions telles que ψ qui sont $\frac{1}{2}$-Hölderiennes, ce qui signifie qu'il existe une constante c_ψ telle que pour tous s, et t de $[0,1]$, $|\psi(s) - \psi(t)| \leq c_\psi |s-t|^{\frac{1}{2}}$. On peut alors en conclure que pour tout $\{x_1, \dots, x_n\} \subset [0,1]^2$, il existe une permutation σ de $\{1, \dots, n\}$ telle que :

$$\sum_{i=1}^{n-1} |x_{\sigma(i)} - x_{\sigma(i+1)}|^2 < c_\psi^2 .$$

En effet, soit $\psi : [0,1] \to [0,1]^2$, continue, surjective, et $\frac{1}{2}$-Hölderienne, et notons t_i des points tels que $\psi(t_i) = x_i$. Puis, ordonnons les t_i, par une permutation σ de $\{1, \dots, n\}$ telle que :

$$t_{\sigma(1)} \leq t_{\sigma(2)} \leq \dots \leq t_{\sigma(n)} .$$

On a alors :

$$\begin{aligned}
\sum_{i=1}^{n-1} |x_{\sigma(i)} - x_{\sigma(i+1)}|^2 &= \sum_{i=1}^{n-1} |\psi(t_{\sigma(i)}) - \psi(t_{\sigma(i+1)})|^2 \\
&\leq c_\psi^2 \sum_{i=1}^{n-1} |t_{\sigma(i)} - t_{\sigma(i+1)}| \\
&\leq c_\psi^2 \sum_{i=1}^{n-1} (t_{\sigma(i+1)} - t_{\sigma(i)}) \\
&= c_\psi^2 (t_{n-1} - t_1) \\
&\leq c_\psi^2 \ .
\end{aligned}$$

Décidons de prendre pour $b_i(x)$ deux fois la longueur des deux arêtes incidentes à x_i, dans le trajet donné par une telle fonction ψ que l'on fixe une fois pour toutes. De cette manière, on sait que :

$$\forall x \in \left([0,1]^2\right)^n , \ \sum_{i=1}^{n} b_i(x)^2 \leq c^2 \ ,$$

où c vaut $2c_\psi$. Cherchons maintenant à vérifier (4.21) :

$$L_n(x) \leq L_n(y) + \sum_{i=1}^{n} b_i(x) \mathbb{1}_{x_i \neq y_i} \ .$$

Pour voir ceci, considérons la figure 7.

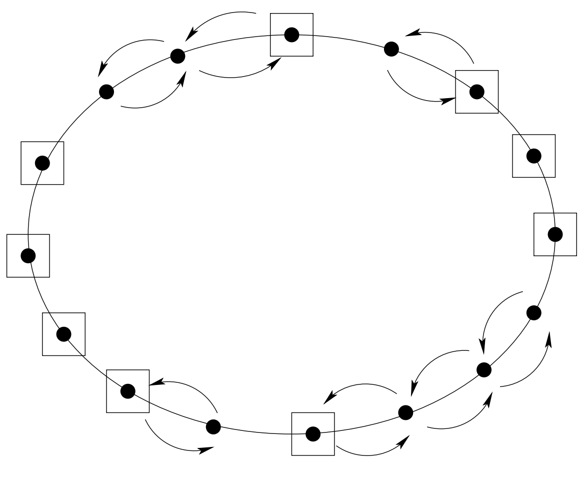

FIG. 7 – Illustration du mode de dépendance de L_n en ses variables

L'ellipse indique le trajet suivi par ψ passant par chacun des x_i (et retour au point de départ), et chaque x_i est désigné par un point noir. Les carrés entourent les éléments communs à x et y. Considérons un trajet optimal par les points de y, auquel on ajoute les "excroissances" désignées par les flèches sur la figure 7. Ces excroissances sont des cycles qui passent par un seul point de y. Le trajet auquel on a ajouté les excroissances est un trajet qui passe notamment par tous les points de x. Sa longueur est donc supérieure à $L_n(x)$. D'un autre côté, sa longueur est égale à la longueur du tour optimal par les points de y augmentée de la somme des longueurs de chaque "excroissance". Avec notre choix des $b_i(x)$, elle est donc inférieure à $L_n(y) + \sum_{i=1}^n b_i(x) \mathbb{1}_{x_i \neq y_i}$. On peut donc appliquer la proposition 4.10 pour obtenir une inégalité de concentration sur L_n : il existe une constante $c > 0$ telle que :

$$P(|L_n - m_n| \geq t) \leq 4e^{-\frac{t^2}{4c^2}} ,$$

où m_n désigne une médiane de L_n.

4.3 Inégalités de Sobolev logarithmiques

L'objectif de cette section est de montrer comment les inégalités de Sobolev logarithmiques peuvent impliquer des résultats de concentration. Il existe sur ce sujet un livre remarquable (et en français) [4], dont un chapitre est consacré aux connections avec la concentration de la mesure, mais qui ignore certains développements récents, notamment ceux de Boucheron Lugosi et Massart (cf. [8]), qu'il nous paraît indispensable de mentionner ici.

Commençons avec l'exemple de l'inégalité de Sobolev logarithmique gaussienne, démontrée par Gross en 1976 (cf. [55]) :

Théorème 4.11 *Soit γ_n la mesure gaussienne standard sur $\mathbb{R}^n$, et f une fonction continûment différentiable sur $\mathbb{R}^n$, et à valeurs dans $\mathbb{R}$. Alors,*

$$\int f^2 \log f^2 \, d\gamma_n - \left(\int f^2 \, d\gamma_n\right) \log \int f^2 \, d\gamma_n \leq 2 \int \|\nabla f\|_2^2 \, d\gamma_n .$$

Ce modèle illustre une forme générale des inégalités de Sobolev logarithmiques : il s'agit de majorer l'entropie d'une fonction f par une forme de son énergie. La forme de l'énergie peut être très variée, et elle est en fait associée à la mesure sous laquelle on prend l'entropie (dans le théorème 4.11, il s'agit de γ_n). L'objet théorique essentiel dont la définition d'inégalité de Sobolev logarithmique découle est alors un générateur infinitésimal et le semi-groupe de diffusion associé. Nous en dirons un peu plus dans la section 4.4.

Nous allons montrer que ce théorème permet d'obtenir un résultat de concentration. Le procédé décrit ci-après est couramment nommé "argument de Herbst". Il consiste à appliquer une inégalité de Sobolev logarithmique comme celle du théorème 4.11 à une certaine fonction $e^{\frac{sg}{2}}$ pour obtenir une inéquation différentielle sur la transformée de log-Laplace de g, puis à intégrer cette inéquation différentielle.

Soit g une fonction continûment différentiable sur $\mathbb{R}^n$, et à valeurs dans $\mathbb{R}$, que l'on supposera de plus L-lipschitzienne. On a donc, pour tout x dans $\mathbb{R}^n$, $\|\nabla g\|_2 \leq L$. On peut appliquer le théorème 4.11 à la fonction $f = e^{\frac{sg}{2}}$, où $s > 0$. Observons que :

$$\forall x \in \mathbb{R}^n,\ \|\nabla f(x)\|_2^2 = \frac{s^2}{4}\|\nabla g(x)\|_2^2 e^{sg(x)} \leq \frac{s^2L^2}{4} e^{sg(x)} .$$

On en déduit :

$$s\int g e^{sg} \, d\gamma_n - \int e^{sg} \, d\gamma_n \log \int e^{sg} \, d\gamma_n \leq \frac{s^2L^2}{2} \int e^{sg} \, d\gamma_n .$$

En posant $F(s) = \int e^{s(g - \int g \, d\gamma_n)} \, d\gamma_n$, l'équation précédente est équivalente à :

$$\frac{F'(s)}{sF(s)} - \frac{1}{s^2} \log F(s) \leq \frac{L^2}{2} ,$$

ce qui se simplifie si l'on pose $H(s) = \frac{\log F(s)}{s}$:

$$H'(s) \leq \frac{L^2}{2} .$$

En constatant que $H(s)$ tend vers 0 lorsque s tend vers 0, on obtient que $H(s) \leq \frac{sL^2}{2}$. C'est à dire :

$$\int e^{s(g - \int g \, d\gamma_n)} \, d\gamma_n \leq e^{\frac{s^2}{2}L^2} .$$

Par un argument de régularisation, on peut obtenir la même équation en supposant seulement que g est L-lipschitzienne. Finalement, l'inégalité de Chernov permet de conclure. Pour tout $t > 0$:

$$\gamma_n\left(g(x) \geq \int g\,d\gamma_n + t\right) \leq e^{-\frac{t^2}{2L^2}}\ . \tag{4.22}$$

Et en appliquant ce qui précède à $-g$, on obtient :

$$\gamma_n\left(g(x) \leq \int g\,d\gamma_n - t\right) \leq e^{-\frac{t^2}{2L^2}}\ . \tag{4.23}$$

On retrouve ainsi l'ordre de grandeur de la concentration donnée par (4.16).

Remarquons que ce procédé, au lieu de situer la concentration autour de la médiane, la situe autour de l'espérance. Cela dit, en appliquant (4.22) à $t = M_g - \int g\,d\gamma_n$, si $M_g \geq \int g\,d\gamma_n$, et (4.23) à $t = \int g\,d\gamma_n - M_g$, si $M_g \leq \int g\,d\gamma_n$, on obtient :

$$\left|M_g - \int g\,d\gamma_n\right| \leq L\sqrt{2\log 2}\ .$$

En particulier, quand l'espérance est grande devant le facteur de Lipschitz L, la médiane l'est aussi. On peut écrire, pour tout $t > 0$:

$$\gamma_n\left(g(x) \geq M_g + L\sqrt{2\log 2} + t\right) \leq e^{-\frac{t^2}{2L^2}}\ ,$$

et

$$\gamma_n\left(g(x) \leq M_g - L\sqrt{2\log 2} - t\right) \leq e^{-\frac{t^2}{2L^2}}\ .$$

Mentionnons à présent les développements récents dûs à Boucheron Lugosi et Massart (cf. [8]) qui partent tous d'une inégalité de type Sobolev logarithmique très générale (pour sa démonstration, voir [84] ou [7]) :

Théorème 4.12 *Soient $X_1, \ldots, X_n$ des variables aléatoires indépendantes, à valeurs dans un epace mesurable $\mathcal{X}$, et f une fonction mesurable de $\mathcal{X}$ dans $\mathbb{R}$. Considérons également $X_1^{'}, \ldots, X_n^{'}$, n copies indépendantes de $X_1, \ldots, X_n$.*
On note $Z = f(X_1, \ldots, X_n)$, $Z^{(i)} = f(X_1, \ldots, X_{i-1}, X_i^{'}, X_{i+1}, \ldots, X_n)$ et $\psi(x) = x(e^x - 1)$.
Alors, pour tout nombre s réel :

$$s\mathbb{E}\left(Ze^{sZ}\right) - \mathbb{E}\left(e^{sZ}\right)\log\mathbb{E}\left(e^{sZ}\right) \leq \sum_{i=1}^{n}\mathbb{E}\left(e^{sZ}\psi(-s(Z - Z^{(i)}))\mathbb{1}_{Z>Z^{(i)}}\right)\ ,$$

et :

$$s\mathbb{E}\left(Ze^{sZ}\right) - \mathbb{E}\left(e^{sZ}\right)\log\mathbb{E}\left(e^{sZ}\right) \leq \sum_{i=1}^{n}\mathbb{E}\left(e^{sZ}\psi(s(Z^{(i)} - Z))\mathbb{1}_{Z<Z^{(i)}}\right)\ .$$

A partir de ces inégalités, Boucheron, Lugosi et Massart démontrent plusieurs inégalités de concentration tout à fait remarquables. Ils retrouvent, entre autres, comme conséquence de deux de leurs résultats, le théorème 4.9 (bien qu'avec une moins bonne constante dans l'exposant). Ils trouvent également les théorèmes généraux suivants, qui s'appliquent aux fonctions qui comptent le nombre de sous-graphes du graphe aléatoire $\mathcal{G}(n,p)$ isomorphes à un graphe donné de petite taille. Définissons :

$$V_+ = \mathbb{E}\left(\sum_{i=1}^{n}(Z - Z^{(i)})^2\mathbb{1}_{Z>Z^{(i)}}\,\middle|\,X_1, \ldots, X_n\right)\ .$$

Théorème 4.13 *Supposons qu'il existe des constantes a et b telles que :*

$$V_+ \leq aZ + b\ .$$

Alors, pour tout $s \in]0, \frac{1}{a}[$,

$$\log \mathbb{E}\left[e^{s(Z-\mathbb{E}[Z])}\right] \leq \frac{s^2}{1-as}(a\mathbb{E}[Z]+b)\ ,$$

et, pour tout $t > 0$,

$$Prob[Z > \mathbb{E}[Z] + t) \leq \exp\left(-\frac{t^2}{4a\mathbb{E}[Z]+4b+2at}\right)\ .$$

Théorème 4.14 *Supposons que f soit positive, et qu'il existe une variable aléatoire W telle que :*

$$V_+ \leq WZ\ .$$

Alors, pour tout $\theta > 0$ et tout $s \in]0, \frac{1}{\theta}[$,

$$\log \mathbb{E}\left[e^{s\left(\sqrt{Z}-\mathbb{E}[\sqrt{Z}]\right)}\right] \leq \frac{s\theta}{1-s\theta} \log \mathbb{E}\left[e^{\frac{sW}{\theta}}\right]\ .$$

Appliquons ce résultat à la déviation supérieure de la variable aléatoire Z comptant le nombre de triangles dans le graphe $\mathcal{G}(n,p)$. Il est facile de voir que :

$$\mathbb{E}[Z] = \binom{n}{3} p^3 \approx \frac{n^3p^3}{6}\ ,$$

et on peut montrer également :

$$Var[Z] = \binom{n}{3}(p^3 - p^6) + \binom{n}{4}\binom{4}{2}(p^5 - p^6)\ .$$

Pour tout i on définit la variable aléatoire B_i comme suit : si u et v désignent les extrémités de l'arête i, B_i est le nombre de sommets w tels que les arêtes (u,i) et (i,v) soient présentes dans le graphe aléatoire. On a alors, en notant X_i l'état de l'arête i :

$$\begin{aligned} V_+ &= \sum_{i=1}^{m} X_i(1-p)B_i^2 \\ &\leq \sum_{i=1}^{m} X_i B_i^2\ . \end{aligned}$$

Remarquons de plus que $\sum_{i=1}^{m} X_i B_i = 3Z$. Donc :

$$\begin{aligned} V_+ &\leq \left(\max_{j=1,\dots,m} B_j\right) \sum_{i=1}^{m} X_i B_i \\ &\leq 3\left(\max_{j=1,\dots,m} B_j\right) Z\ . \end{aligned}$$

Notons donc $W = 3\max_{j=1,\dots,m} B_j$, et afin d'appliquer le théorème 4.14, essayons de majorer la fonction génératrice de W. Remarquons déjà que W est bornée par $3n$, et donc, le théorème 4.13 nous donne :

$$Prob[Z > \mathbb{E}[Z] + t) \leq \exp\left(-\frac{t^2}{12n\mathbb{E}[Z]+6nt}\right)\ . \tag{4.24}$$

Puis remarquons que :

$$\sum_{i=1}^{m}(W - W^{(i)})^2 \mathbb{1}_{W > W^{(i)}} \leq 18W \ .$$

Ainsi, le théorème 4.13 nous donne :

$$\log \mathbb{E}\left[e^{s(W-\mathbb{E}[W])}\right] \leq \frac{s^2}{1-18s}(18\mathbb{E}[W]) \ .$$

On peut alors appliquer le théorème 4.14, avec par exemple $\theta = 1$. Notons $Y = \sqrt{Z}$:

$$\begin{aligned} \log \mathbb{E}\left[e^{s(Y-\mathbb{E}[Y])}\right] &\leq \frac{s}{1-s}\left(\frac{s^2}{1-18s}(18\mathbb{E}[W]) + s\mathbb{E}[W]\right) \\ &\leq \frac{s^2\mathbb{E}[W]}{1-19s} \ . \end{aligned}$$

On obtient, après Chernov, minimisation en s et quelques menus calculs :

$$Prob[Y > \mathbb{E}[Y] + t] \leq \exp\left(-\frac{t^2}{4\mathbb{E}[W] + 13t}\right) \ .$$

Il ne nous reste plus qu'à disposer d'une borne supérieure raisonnable pour $\mathbb{E}[W]$. Sachant que $\frac{W}{3}$ est le sup de $\alpha(n)$ binomiales de paramètres (n, p^2), on peut effectuer la manipulation qui suit. Grâce à l'inégalité de Jensen :

$$\begin{aligned} \frac{\mathbb{E}[W]}{3} &\leq \log\left(\mathbb{E}\left[\max_{i=1,\dots,m} e^{B_i}\right]\right) \\ &\leq \log\left(\mathbb{E}\left[\sum_{i=1}^{m} e^{B_i}\right]\right) \\ &= \log\alpha(n) + \log\left(\mathbb{E}\left[e^{B_1}\right]\right) \ . \end{aligned}$$

On est donc ramené à obtenir une borne supérieure de $\mathbb{E}\left[e^{B_1}\right]$. Or, B_1 est la somme de n variables indépendantes de Bernoulli I_i de paramètres p^2. On a donc :

$$\begin{aligned} \log\left(\mathbb{E}\left[e^{B_1}\right]\right) &= n\log\left(\mathbb{E}\left[e^{I_1}\right]\right) \\ &= n\log(1 + p^2(e-1)) \\ &\leq (e-1)np^2 \ . \end{aligned}$$

D'où :

$$Prob[Y \geq \mathbb{E}[Y] + t] \leq \exp\left(-\frac{t^2}{12((e-1)np^2 + 2\log n) + 13t}\right) \ .$$

On peut maintenant en déduire une borne sur la queue de Z, en utilisant, grâce à l'inégalité de Jensen, que $\mathbb{E}[Y] \leq \sqrt{\mathbb{E}[Y^2]}$:

$$\begin{aligned} Prob[Z > \mathbb{E}[Z] + t] &= Prob\left[Y \geq \sqrt{\mathbb{E}[Y^2] + t}\right] \\ &\leq Prob\left[Y \geq \mathbb{E}[Y] + x\right] \\ &\leq \exp\left[-\frac{x^2}{12((e-1)np^2 + 2\log n) + 13x}\right] \ , \end{aligned}$$

où $x = \sqrt{\mathbb{E}[Z] + t} - \sqrt{\mathbb{E}[Z]}$. Donc, t valant $x^2 + 2x\sqrt{\mathbb{E}[Z]}$,

$$Prob[Z > \mathbb{E}[Z] + 2x\sqrt{\mathbb{E}[Z]} + x^2] \leq \exp\left(-\frac{x^2}{12((e-1)np^2 + 2\log n) + 13x}\right) \ .$$

D'où, en posant $u = x\sqrt{\mathbb{E}[Z]}$,

$$Prob\left[Z > \mathbb{E}[Z] + 2u + \frac{u^2}{\mathbb{E}[Z]}\right] \leq \exp\left(-\frac{u^2/\mathbb{E}[Z]}{12((e-1)np^2 + 2\log n) + \frac{13u}{\sqrt{\mathbb{E}[Z]}}}\right) . \tag{4.25}$$

Remarquons que pour $\alpha > 0$, et $u > 0$, on a :

$$\alpha u \geq \frac{u^2}{\mathbb{E}[Z]} \Leftrightarrow u \leq \alpha \mathbb{E}[Z] .$$

Avec (4.24) et (4.25), on peut alors décrire différents régimes de concentration selon l'échelle à laquelle se situe t par rapport à p, à $\mathbb{E}[Z]$ et à $(e-1)np^2 + 2\log n$ (voir l'article original [8] pour plus de détails).

4.4 Lemme de Beckner et hypercontractivité

Nous avons vu dans la section 3.3 que l'ingrédient essentiel à la démonstration du théorème 3.16 était un lemme de Beckner modifié par Talagrand (cf. [105]). Nous allons tout d'abord remettre la version originale de ce résultat dans le cadre général de l'hypercontractivité. Cette section est principalement fondée sur le livre de Ané et al. [4].

Rappelons la définition générale d'un semi-groupe. Pour cela, on se donne un espace topologique mesuré $(\Omega, \mathcal{F}, \mu)$, et un espace vectoriel normé complet de fonctions continues bornées de Ω dans $\mathbb{R}$, que l'on notera $(\mathcal{B}, ||.||)$. Par exemple, pour $\Omega = \{0,1\}^n$, $\mathcal{B}$ pourra être l'ensemble des fonctions de Ω dans $\mathbb{R}$, muni de la norme $||.||_\infty$. On pourra prendre aussi $\Omega = \mathbb{R}^n$, et $\mathcal{B} = \mathcal{C}_b(\mathbb{R}^n, \mathbb{R})$, muni encore de la norme $||.||_\infty$. On suppposera toujours que la fonction $1\!\!1_\Omega$, constante égale à 1, appartient à $\mathcal{B}$.

Définition 4.15 *On dit qu'une famille $(S_t)_{t\leq 0}$ d'opérateurs linéaires sur $\mathcal{B}$ est un semi-groupe de Markov si et seulement si :*
- (i) $S_0 = Id$,
- (ii) *pour toute fonction* $f \in \mathcal{B}$, $t \mapsto S_t f$ *est continue sur* $\mathbb{R}^+$,
- (iii) *pour tous* $s,\ t \geq 0, S_{t+s} = S_t \circ S_s$,
- (iv) $S_t 1\!\!1_\Omega = 1\!\!1_\Omega$ *et* $S_t f \geq 0$ *pour* $f \geq 0$,
- (v) *pour toute fonction* $f \in \mathcal{B}$, $||S_t f|| \leq ||f||$.

On peut alors définir l'hypercontractivité de la manière suivante :

Définition 4.16 *Etant donnée une fonction strictement croissante q, de $\mathbb{R}^+$ dans $[q(0), +\infty[$, on dit qu'un semi-groupe $(S_t)_{t\leq 0}$ est hypercontractif, de fonction de contraction q, si et seulement si pour toute fonction f de $\mathcal{B}$, et tout $t > 0$,*

$$||S_t f||_{q(t)} \leq ||f||_{q(0)} .$$

Cela revient donc à dire que S_t est une contraction de $L^{q(0)} \bigcap \mathcal{B}$ dans $L^{q(t)} \bigcap \mathcal{B}$. Evidemment, une telle définition n'est intéressante que si $L^{q(0)} \bigcap \mathcal{B}$ n'est pas vide. Elle traduit alors une certaine capacité du semi-groupe à "lisser" les fonctions.

Nous allons nous intéresser à un semi-groupe de Markov particulier sur $\Omega = \{0,1\}^n$ muni de l'equiprobabilité $\mu_{n,1/2}$. Pour tout $t > 0$, notons K_t le noyau de transition sur $\{0,1\}$ défini par :

$$K_t(x,x) = \frac{1+e^{-t}}{2} ,$$

$$K_t(x,y) = \frac{1-e^{-t}}{2} \text{ si } x \neq y .$$

Puis, notons S_t l'opérateur défini à l'aide de ce noyau :

$$S_t f(x) = \int_{\{0,1\}} f(y) K_t(y,x)\, d\mu(y) ,$$

où μ est l'équiprobabilité sur $\{0,1\}$. Remarquons que :

$$S_t f(x) = e^{-t} f(x) + (1-e^{-t}) \int_{\{0,1\}} f(y)\, d\mu(y) \ .$$

De sorte que pour toute fonction f de $\{0,1\}$ dans $\mathbb{R}$,

$$S_t f \xrightarrow[n\to+\infty]{} \int_{\{0,1\}} f(y)\ d\mu(y) \ .$$

On peut voir facilement que la famille $(S_t)_{t\geq 0}$ forme un semi-groupe, en choisissant par exemple comme norme sur l'ensemble des fonctions, la norme de $L^2(\{0,1\})$. On peut définir le semi-groupe produit sur $\{0,1\}^n$ avec pour famille de noyaux :

$$K_{n,t}((x_1,\dots,x_n),(y_1,\dots,y_n)) = \prod_{i=1}^n K_t(x_i,y_i) \ ,$$

en posant, pour toute fonction f de $\{0,1\}^n$ dans $\mathbb{R}$:

$$S_{n,t} f(x) = \int_{\{0,1\}^n} f(y) K_{n,t}(y,x)\, d\mu^{\otimes n}(y) \ .$$

Sous cette forme, on montre par récurrence que :

$$S_{n,t} f(x) = \sum_{k=0}^n \sum_{\substack{I\subset\{1,\dots,n\}\\ |I|=k}} e^{-kt}(1-e^{-t})^{n-k} \mathbb{E}\left[f(X_1,\dots,X_n)\,|(X_i)_{i\in I} = (x_i)_{i\in I}\right] \ ,$$

ce qui permet de démontrer facilement que les fonctions propres de $S_{n,t}$ sont les fonctions r_S déjà définies dans la section 3.3. Ici, p valant $\frac{1}{2}$, leur expression est très simple :

$$r_S(x) = (-1)^{\sum_{i\in S} x_i} \ .$$

On a alors :

$$S_{n,t} r_S = e^{-t|S|} r_S \ .$$

On peut maintenant énoncer le lemme de Beckner sous sa forme originale (cf. [5]) :

Lemme 4.17 *Le semi-groupe $(S_{n,t})_{t\geq 0}$ est hypercontractif, de fonction de contraction $q(t) = 1 + e^{2t}$. Dit de manière équivalente, pour toute famille de réels $(a_S)_{S\subset\{1\dots,n\}}$,*

$$\left\| \sum_S e^{-t|S|} a_S r_S \right\|_{q(t)} \leq \left\| \sum_S a_S r_S \right\|_2 \ .$$

Nous avons donc une notion théorique fondamentale pour avoir des théorèmes de "raideur de seuil" : l'hypercontractivité, et une autre pour obtenir des résultats de concentration : les inégalités de Sobolev logarithmiques. Or, il se trouve que ces deux notions sont liées, et nous allons voir comment.

Il existe un être mathématique très important pour les semi-groupes de Markov, c'est le générateur infinitésimal (voir par exemple les chapitres 3 et 5 de [115]). Le générateur infinitésimal d'un semi-groupe de Markov $(S_t)_{t\geq 0}$ est un opérateur linéaire $\mathbf{L}$ défini comme suit, lorsque la limite existe :

$$\mathbf{L} f = \lim_{t\to 0} \frac{S_t f - f}{t} \ .$$

L'ensemble des fonctions f telles que cette limite existe est noté $\mathcal{D}(\mathbf{L})$ et est appelé le domaine de $\mathbf{L}$. Grâce au caractère markovien du semi-groupe, le générateur en enferme toutes les caractéristiques. La donnée du générateur est équivalente à celle du semi-groupe. On peut alors donner une nouvelle définition des inégalités de Sobolev logarithmique à partir du générateur, et pour la mesure invariante par le semi-groupe associé.

Définition 4.18 *Soit $\boldsymbol{L}$ un générateur infinitésimal de mesure invariante μ (c'est à dire que μ est invariante par le semi-groupe de Markov associé à $\boldsymbol{L}$). On dira que μ satisfait à une inégalité de Sobolev logarithmique de constante c si, pour toute fonction $f \in \mathcal{D}(\boldsymbol{L})$,*

$$\int f^2 \log f^2 \, d\mu - \left(\int f^2 \, d\mu\right) \log \int f^2 \, d\mu \leq c \int -f\boldsymbol{L}f \, d\mu \, .$$

Cette définition coïncide notamment avec celle sous-entendue dans le théorème 4.11, qui présente en fait une inégalité de Sobolev logarithmique de constante 2 pour le semi-groupe d'Ornstein-Uhlenbeck. Par contre, elle ne couvre pas les inégalités de Boucheron-Lugosi-Massart (théorème 4.12).

Nous allons donc finir ce chapitre (et ce cours) en laissant à votre réflexion le théorème suivant, dû à Gross [56].

Théorème 4.19 *Soit $(S_t)_{t\geq 0}$ un semi-groupe de Markov admettant μ pour mesure* réversible *(cf. [115] p. 121). Alors, les deux propositions suivantes sont vérifiées :*

(i) S'il existe une constante $c > 0$ telle que le semi-groupe $(S_t)_{t\geq 0}$ soit hypercontractif de fonction de contraction $q(t) = 1+e^{\frac{4t}{c}}$, alors, la mesure μ satisfait à une inégalité de Sobolev logarithmique de constante c.

(ii) Si μ satisfait à une inégalité de Sobolev logarithmique de constante $c > 0$, alors pour tout $q(0) > 1$, le semi-groupe $(S_t)_{t\geq 0}$ est hypercontractif, de fonction de contraction $q(t) = 1 + (q(0)-1)e^{\frac{4t}{c}}$.

Références

[1] S. Abiteboul, K. Compton, and V. Vianu. Expressive power of query languages. In J.D. Ullman, editor, *Theoretical Studies in Computer Science*, 207–251. Academic Press, New York, 1991.

[2] S. Abiteboul, K. Compton, and V. Vianu. Queries are easier than you thought (probably). In *Proc. 11th ACM Symp. on Principles of Database Systems*, 23–32, San Diego, CA, 1992.

[3] S. Abiteboul, R. Hull, and V. Vianu. *Foundations of databases.* Addison-Wesley, Reading, 1995.

[4] C. Ané, S. Blachère, D. Chafaï, P. Fougères, I. Gentil, F. Malrieu, C. Roberto, and G. Scheffer. *Sur les inégalités de Sobolev logarithmiques.* Société Mathématique de France, Paris, 2000.

[5] W. Beckner. Inequalities in Fourier analysis. *Ann. Math.*, 102:159–182, 1975.

[6] B. Bollobás. *Random Graphs.* Academic Press, London, 1985.

[7] S. Boucheron, G. Lugosi, and P. Massart. A sharp concentration inequality with applications. *Rand. Struct. Algo.*, 16:277–292, 2000.

[8] S. Boucheron, G. Lugosi, and P. Massart. Concentration inequalities using the entropy method. *Ann. Probab.*, to appear.

[9] J. Bourgain and G. Kalai. Influences of variables and threshold intervals under group symmetries. *Geom. Funct. Analysis*, 7:438–461, 1997.

[10] S.N. Burris. *Number theoretic density and logical limit laws.* American Mathematical Society, Providence, 2001.

[11] K.L. Chung. *Markov chains with stationary transition probabilities.* Springer-Verlag, New York, 1960.

[12] B.S. Cirel'son, I.A. Ibragimov, and V.N. Sudakov. Norms of gaussian sample functions. In *Proc. 3rd Japan-USSR Symp. Probab. Theory, Taschkent 1975*, *L.N. in Mathematics* **550**, 20–41. Springer-Verlag, New York, 1976.

[13] S. Cocco, O. Dubois, J. Mandler, and R. Monasson. A la rescousse de la complexité calculatoire. *Pour la Science*, 295:2–11, may 2002.

[14] H. Cohn. On the tail σ-field of the countable Markov chains. *Revue Roumaine de Mathématiques Pures et Appliquées*, 21(6):667–675, 1976.

[15] K.J. Compton. 0-1 laws in logic and combinatorics. In I. Rival, editor, *Algorithms and order*, 353–383. Kluwer, Dordrecht, 1989.

[16] R. Cori and D. Lascar. *Logique Mathématique, Tomes I et II.* Masson, Paris, 1993.

[17] D. Coupier. Techniques de graphes aléatoires appliquées au traitement d'images. Mémoire de DEA, Orsay, 2002.

[18] N. Creignou. The class of problems that are linearly equivalent to satisfiability or a uniform method for proving NP-completeness. *Theor. Comp. Sci.*, 145:111–145, 1995.

[19] N. Creignou and H. Daudé. Satisfiability threshold for random XOR-CNF formulas. *Discrete Applied Math.*, 41–53, 1999.

[20] H. Dehling, T. Mikosch, and M. Sørensen, editors. *Empirical process techniques for dependent data.* Birkhäuser, Basel, 2002.

[21] A. Dembo and O. Zeitouni. *Large deviations techniques and applications.* Springer-Verlag, New York, 2nd edition, 1998.

[22] A. Desolneux, M. Moisan, and J.M. Morel. Meaningful alignments. *Int. J. Computer Vision*, 40(1):7–23, 2000.

[23] J.D. Deuschel and D.W. Stroock. *Large deviations.* Academic Press, New York, 1989.

[24] L. Devroye and G. Lugosi. *Combinatorial methods in density estimation.* Springer-Verlag, New York, 2001.

[25] P. Doukhan. *Mixing: Properties and examples*, *L.N. in Statistics* **85**. Springer-Verlag, Berlin, 1994.

[26] M. Drmota, D. Gardy, and B. Gittenberger. A unified presentation of some urn models. *Algorithmica*, 29(1-2):120–147, 2001.

[27] O. Dubois, P. André, Y. Boufkhad, and J. Carlier. SAT versus UNSAT. *DIMACS*, 26:415–436, 1996.

[28] O. Dubois and Y. Boufkhad. A general upperbound of the satisfiability threshold for random r-SAT formulae. *J. Algorithms*, 24:395–420, 1997.

[29] P. Duchon, P. Flajolet, G. Louchard, and G. Schaeffer. Random sampling from Boltzmann principles. In *Proc. 29th Int. Conf. on Automata, Languages, and Programming, L.N. in Computer Science* **2380**. Springer-Verlag, New York, 2002.

[30] H.D. Ebbinghaus and J. Flum. *Finite model theory.* Springer-Verlag, Berlin, 1995.

[31] H.D. Ebbinghaus, J. Flum, and W. Thomas. *Mathematical logic.* Springer-Verlag, Berlin, 2nd edition, 1984.

[32] A. Ehrenfeucht. An application of games to the completeness problem for formalized theories. *Fund. Math.*, 49:129–141, 1961.

[33] P. Erdös and A. Rényi. On the evolution of random graphs. *Mat. Kuttató. Int. Közl.*, 5:17–60, 1960.

[34] P. Erdös and P. Révész. On the length of the longest head run. In *Topics in information theory*, volume 16, 219–228. Keszthely, Hungary, 1975.

[35] R. Fagin. Probabilities on finite models. *J. of Symbolic Logic*, 41:50–58, 1976.

[36] W. Feller. *An introduction to probability theory and its applications*, volume I. Wiley, London, 3rd edition, 1968.

[37] W. Feller. *An introduction to probability theory and its applications*, volume II. Wiley, London, 2nd edition, 1971.

[38] P. Flajolet and A.M. Odlyzko. Singularity analysis of generating functions. *SIAM J. on Algebraic and Discrete Methods*, 3(2):216–240, 1990.

[39] P. Flajolet and R. Sedgewick. Analytic combinatorics. To appear, `http ://algo.inria.fr/flajolet/Publications/books.html`, 200X.

[40] P. Flajolet and M. Soria. Gaussian limiting distributions for the number of components in combinatorial structures. *J. Combinatorial Theory*, 53:165–182, 1990.

[41] P. Flajolet and M. Soria. General combinatorial schemas: Gaussian limit distributions and exponential tails. *Discrete Mathematics*, 114:159–180, 1993.

[42] F. Forbes and B. Ycart. Counting stable sets on Cartesian products of graphs. *Discrete Mathematics*, 186:105–116, 1998.

[43] E. Friedgut. Sharp threshold of graph properties and the k-sat problem (with an appendix by Jean Bourgain). *J. Amer. Math. Soc.*, 12(4):1017–1054, 1999.

[44] E. Friedgut and G. Kalai. Every monotone graph property has a sharp threshold. *Proc. Amer. Math. Soc.*, 124:2993–3002, 1996.

[45] H. Gaifman. Concerning measures in first-order calculi. *Israel J. of Mathematics*, 2:1–18, 1964.

[46] H. Gaifman. On local and non local properties. In J. Stern, editor, *Logic Colloquium '81*, 105–135. North Holland, Amsterdam, 1982.

[47] D. Gardy. Occupancy urn models in the analysis of algorithms. *J. Statistical Planning and Inference*, 101(1-2):95–105, 2002.

[48] Y.V. Glebskii, D.I. Kogan, M.I. Liogonkii, and V.A. Talanov. Range and degree of realizability of formulas in the restricted predicate calculus. *Cybernetics*, 5:142–154, 1969.

[49] A.P. Godbole and S.G. Papastavridis, editors. *Runs and patterns in probability: selected papers.* Kluwer, Dordrecht, 1994.

[50] A. Goerdt. A threshold for unsatisfiability. *J. Comput. Syst. Sci.*, 53:460–486, 1996.

[51] C.P. Gomes and B. Selman. Satisfied with physics. *Science*, 297:784–785, August 2002.

[52] L. Gordon, M.F. Schilling, and M.S. Waterman. An extreme value theory for long head runs. *Probab. Th. Rel. Fields*, 72:279–287, 1986.

[53] E. Grädel and A. Malmström. 0-1 Laws for recursive structures. *Archive for Mathematical Logic*, 38:205–215, 1999.

[54] G. Grimmett. *Percolation.* Springer-Verlag, New York, 1989.

[55] L. Gross. Logarithmic Sobolev inequalities. *Amer. J. of Math.*, 97(4):1061–1083, 1976.

[56] L. Gross. Logarithmic sobolev inequalities and contractivity properties of semigroups. In *Dirichlet forms (Varenna 1992)*, 54–88. Springer-Verlag, Berlin, 1993.

[57] A.J. Grove, J.Y. Halpern, and D. Koller. Asymptotic conditional probabilities: The non unary case. *J. Symbolic Logic*, 61(1):250–275, 1996.

[58] A.J. Grove, J.Y. Halpern, and D. Koller. Asymptotic conditional probabilities: The unary case. *SIAM J. on Computing*, 25(1):1–51, 1996.

[59] L.J. Guibas and A.M. Odlyzko. Long repetitive patterns in random sequences. *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, 53:241–262, 1980.

[60] Y. Gurevich. Zero-one laws. *Bull. EATCS*, 46:90–106, 1992.

[61] B. Hayes. Can't get no satisfaction. *American Scientist*, 85:108–112, march-april 1997.

[62] R. Isaac. The tail σ-fields of recurrent Markov processes. *Aplikace Matematiky*, 22(6):397–408, 1977.

[63] N. Johnson and S. Kotz. *Urn models and their applications.* Wiley, New York, 1977.

[64] J. Kahn, G. Kalai, and N. Linial. The influence of variables on Boolean functions. In *Proc. 29-th Ann. Symp. on Foundations of Comp. Sci.*, pages 68–80. IEEE, Washington, 1988.

[65] R.M. Karp. The transitive closure of a random digraph. *Rand. Struct. Algo.*, 1(1):73–94, 1990.

[66] S. Kirkpatrick and B. Selman. Critical behavior in the satisfiability of random Boolean expressions. *Science*, 264(5163):1297–1301, 27 May 1994.

[67] P.G. Kolaitis and M.Y. Vardi. 0-1 laws and decision problems for fragments of second-order logic. *Information and Computation*, 87(1-2):302–338, 1990.

[68] P.G. Kolaitis and M.Y. Vardi. 0-1 laws for fragments of existential second-order logic: a survey. In M. Nielsen and B. Rovan, editors, *Math. Found. of Computer Science*, *L.N. in Computer Science* **1893**, 84–98. Springer-Verlag, Berlin, 2000.

[69] B. Kopociński. On the distribution of the longest success-run in Bernoulli trials. *Rosczniki Polskiego Towarzystwa Mat.*, 24:5–13, 1991.

[70] R. Lassaigne and M. de Rougemont. *Logique et fondements de l'informatique.* Hermès, Paris, 1993.

[71] R. Lassaigne and M. de Rougemont. *Logique et complexité.* Hermès, Paris, 1996.

[72] J.M. Le Bars. Fragments of existential second-order logic without 0-1 laws. In *Proc. 13th annual IEEE symposium on logic in computer science*, 527–537, 1998.

[73] J.M. Le Bars. Counterexamples of the 0-1 law for fragments of existential second order logic: an overview. *Bull. Symbolic Logic*, 9:67–82,, 2000.

[74] J.M. Le Bars. The 0-1 law fails for the monadic existential second-order logic on undirected graphs. *Information and Processing Letters*, 77:43–48, 2001.

[75] J.M. Le Bars. The 0-1 law fails for frame satisfiability of propositional modal logic. In *Proc. 17th IEEE Symposium on Logic in Computer Science,*, 225–234, 2002.

[76] M.R. Leadbetter, G. Lingren, and H. Rootzén. *Extremes and related properties of random sequences and processes.* Springer-Verlag, New York, 1983.

[77] M. Ledoux. On Talagrand's deviation inequalities for product measures. *ESAIM Probab. Stat.*, 1:63–87, 1996.

[78] M. Ledoux. Concentration of measures and logarithmic Sobolev inequalities. In J. Azéma and M. Yor, editors, *Séminaire de Probabilités XXXIII, L.N. in Mathematics* **1709**, pages 120–216. Springer-Verlag, New York, 1999.

[79] P. Lézaud. Chernoff-type bounds for finite Markov chains. *Ann. Applied Probab.*, 8(3):849–867, 1998.

[80] T. Łuczac. The phase transition in the evolution of random digraphs. *J. Graph Theory*, 14(2):217–223, 1990.

[81] J. Lynch. Threshold functions for Markov chains: A graph-theoretic approach. *Combin., Probab, and Comput.*, 2:351–362, 1993.

[82] S. Martínez and B. Ycart. Decay rates and cutoff for convergence and hitting times of Markov chains with countably infinite state space. *Adv. Applied Probab.*, 33(1):188–205, 2001.

[83] K. Marton. A measure concentration inequality for contracting Markov chains. *Ann. Probab.*, 24(2):857–866, 1996.

[84] P. Massart. About the constants in Talagrand's concentration inequalities for empirical processes. *Ann. Probab.*, 28:863–885, 2000.

[85] R. Monasson and R. Zecchina. Entropy of the k-satisfiability problem. *Phys. Rev. Lett.*, 76:3881, 1996.

[86] R. Monasson and R. Zecchina. Statistical mechanics of the random k-sat model. *Phys. Rev. E*, 56:1357, 1997.

[87] R. Monasson, R. Zecchina, S. Kirkpatrick, B. Selman, and L. Troyansky. Determining complexity from characteristic 'phase transitions'. *Nature*, 400:133–137, july 1999.

[88] W. Oberschelp. Asymptotic 0-1 laws in combinatorics. In D. Jungnickel, editor, *Combinatorial theory, L.N. in Mathematics* **969**, 276–292. Springer-Verlag, Berlin, 1982.

[89] E.M. Palmer. *Graphical evolution: An introduction to the theory of random graphs.* Wiley, New York, 1985.

[90] C. Paroissin and B. Ycart. Zero-one law for the non-availability of multi-states repairable systems. Submitted, 2002.

[91] V.V. Petrov. *Sums of independent Random Variables.* Springer-Verlag, New York, 1975.

[92] V.V. Petrov. *Limit theorems of probability theory.* Oxford University Press, Oxford, 1995.

[93] E. Rio. *Théorie asymptotique des processus aléatoires faiblement dépendants*, *Mathématiques et applications* **31**. Springer-Verlag, Berlin, 1999.

[94] R. Rossignol. Largeur du seuil pour une propriété croissante et symétrique. Mémoire de DEA, Orsay, 2000.

[95] M.C. Rousset and B. Ycart. A zero-one law for random sentences in description logics. In D. Gardy and A. Mokkadem, editors, *Proc. Colloquium on Mathematics and computer science*, 329–340. Birkäuser, Basel, 2000.

[96] W. Rudin. *Fourier analysis on groups.* Wiley, New York, 1990.

[97] R. Sedgewick and P. Flajolet. *Introduction à l'analyse des algorithmes.* Int. Thomson Publishing, France, 1996.

[98] S. Shelah and J. Spencer. Zero-one laws for sparse random graphs. *J. Amer. Math. Soc.*, 1:97–115, 1988.

[99] J. Spencer. Counting extensions. *J. Combinatorial Theory*, 55:247–255, 1990.

[100] J. Spencer. Threshold functions for extension statements. *J. Combinatorial Theory*, 53:286–305, 1990.

[101] J. Spencer. Nine lectures on Random Graphs. In P. Bernard, editor, *Ecole d'été de probabilité de Saint-Flour XXI, L.N. in Mathematics* **1541**, 293–343. Springer-Verlag, New York, 1991.

[102] J. Spencer. Zero-one laws with variable probability. *J. of Symbolic Logic*, 58:1–14, 1993.

[103] J.M. Steele. Probability theory and combinatorial optimization. In *CBMS-NSF Regional Conference Series in Applied Mathematics*, volume 69. SIAM, Philadelphia, 1997.

[104] D.W. Stroock. *An introduction to the theory of large deviations.* Springer-Verlag, New York, 1984.

[105] M. Talagrand. On Russo's approximate zero-one law. *Ann. Probab.*, 22:1576–1587, 1994.

[106] M. Talagrand. Concentration of measure and isoperimetric inequalities in product spaces. *Public. Mathématiques IHES*, 81:73–205, 1995.

[107] M. Talagrand. A new look at independence. *Ann. Probab.*, 24:1–34, 1996.

[108] Y.L. Tong, editor. *Inequalities in Statistics and Probability, IMS Lecture Notes* **5**, 1984.

[109] Y. Verhoeven. Random 2-sat and unsatisfiability. *Information Processing Letters*, 72:119–123, 1999.

[110] W. Werner. Communication privée, Novembre 2002.

[111] P. Winkler. Random structures and zero-one laws. In N.W. et al. Sauer, editor, *Finite and infinite combinatorics in sets and logic*, 399–420. Kluwer, Dordrecht, 1993.

[112] B. Ycart. Cutoff for samples of Markov chains. *ESAIM Probab. Stat.*, 3:89–107, 1999.

[113] B. Ycart. Stopping tests for Monte-Carlo Markov chain methods. *Meth. and Comp. in Applied Probab.*, 2(1):23–36, 2000.

[114] B. Ycart. Cutoff for Markov chains: some examples and applications. In E. Goles and S. Martínez, editors, *Complex Systems*, 261–300. Kluwer, Dordrecht, 2001.

[115] B. Ycart. *Modèles et algorithmes markoviens*, *Mathématiques et applications* **39**. Springer-Verlag, Berlin, 2002.