Inégalités de concentration et applications : synthèse d'articles

Nicolas Goix, M2 Probabilités et Modèles aléatoires

Encadrant : Rama Cont

Table des matières

Ce document présente des résultats généraux sur les inégalités de concentration de mesure et quelques applications, notamment aux graphes aléatoires. Il est basé sur les articles de Colin McDiarmid "Concentration", de Terence Tao "Talagrand's concentration inequality", et de Svante Janson "On concentration of probability". Dans ce travail, j'ai choisi de mettre en relation les trois articles de la facon suivante : J'utilise l'article de McDiarmid pour illustrer et démontrer dans un cadre plus général les résultats fondamentaux exposés dans l'article de Janson, dont le but est de donner un survol des différentes méthodes pour obtenir des inégalités de concentration. J'ai choisi quelques applications présentées par McDiarmid, ainsi que dans l'article de Tao. Nous verrons deux méthodes pour obtenir des inégalités de concentration : la méthode des différences bornées, qui est une méthode de martingales, puis la méthode assez récente due à Talagrand, qui donne souvent des résultats plus forts que la précédente.

1 Méthodes des différences bornées

1.1 Deux résultats fondamentaux

Les deux résultats suivants permettent de déduire un grand nombre d'inégalités de concentration, comme celles d'Hoeffding, Azuma, Bernstein, McDiarmid.

1.1.1 Définitions préliminaires

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité. Soit X une variable aléatoire sur cet espace et $\mathcal{G}$ une sous-tribu de $\mathcal{F}$.

Définition 1.1. *On suppose X réelle. On définit alors $\sup(X|\mathcal{G})$ comme étant l'unique variable aléatoire réelle $f : \Omega \to \mathbb{R}$ qui vérifie :*
(i) f est $\mathcal{G}$-mesurable
(ii) $X \leq f$
(iii) Si $g : \Omega \to \mathbb{R}$ vérifie (i) et (ii) alors $f \leq g$

Remarque : On a clairement $\sup(X|\mathcal{G}) \geq \mathbb{E}(X|\mathcal{G})$ et $\sup(X|\mathcal{G}_1) \geq \sup(X|\mathcal{G}_2)$ si $\mathcal{G}_1 \subset \mathcal{G}_2$.

Définition 1.2. *On suppose X bornée. Soit $(\mathcal{F}_k)_{0\leq k\leq n}$ filtration de $\mathcal{F}$ telle que X $\mathcal{F}_n$ -mesurable. On note $X_1, ..., X_n$ la martingale $X_k = \mathbb{E}(X|\mathcal{F}_k)$ et $Y_k = X_k - X_{k-1}$ la différence de martingale associée. On définit alors :*
- ⋆ $\mathbf{ran(X|}\mathcal{G}\mathbf{)} = sup(X|\mathcal{G}) + sup(-X|\mathcal{G})$
- ⋆ $\mathbf{ran_k} = ran(Y_k|\mathcal{F}_{k-1}) = ran(X_k|\mathcal{F}_{k-1})$
- ⋆ $\mathbf{R^2} = \sum_1^n ran_k^2$ *et* $\mathbf{\hat{r}^2} = supess(R^2)$

Définition 1.3. *On se place dans le même cadre que ci-dessus, sans supposer X bornée. On définit alors :*
- $\mathbf{var(X|}\mathcal{G}\mathbf{)} = \mathbb{E}((X - \mathbb{E}(X|\mathcal{G}))^2|\mathcal{G})$
- $\mathbf{var_k} = var(Y_k|\mathcal{F}_{k-1}) = var(X_k|\mathcal{F}_{k-1})$
- $\mathbf{V} = \sum var_k$ *et* $\hat{\nu} = supess(V)$ *la somme maximale des variances conditionnelles.*
- ∗ $\mathbf{dev_k^+} = sup(Y_k|\mathcal{F}_{k-1})$
- ∗ $\mathbf{maxdev^+} = supess(\max_{0\leq k\leq n} dev_k^+)$ *la déviation conditionnelle maximale.*

1.1.2 Enoncés

Théorème 1.1. *Soit X une variable aléatoire bornée avec $\mathbb{E}(X) = \mu$. Soit $(\mathcal{F}_k)_{0\leq k\leq n}$ une filtration de $\mathcal{F}$ avec $\mathcal{F}_0 = (\emptyset, \Omega)$ et telle que X est $\mathcal{F}_n$-mesurable. Alors $\forall t \geq 0, \mathbb{P}(X - \mu \geq t) \leq e^{-2t^2/\hat{r}^2}$,*
et plus généralement $\forall t \geq 0, \forall r^2 \geq 0, \mathbb{P}((X - \mu \geq t) \cap (R^2 \leq r^2)) \leq e^{-2t^2/r^2}$.

On a besoin des deux lemmes suivants, qui sont faciles à prouver (le premier par récurrence, le second en utilisant la convexité de $x \to e^{hx}$:

Lemme utile 1.1. *Soit $(\mathcal{F}_k)_{0\leq k\leq n}$ une filtration de $\mathcal{F}$ avec $\mathcal{F}_0 = (\emptyset, \Omega)$. Soit $(Y_k)_{1\leq k\leq n}$ une différence de martingale pour cette filtration telle que chaque Y_k soit borné. Soit Z une fonction indicatrice. Alors :*

$$\mathbb{E}(Ze^{h\sum Y_k}) \leq sup(Z\prod_k \mathbb{E}(e^{hY_k}|\mathcal{F}_{k-1}))$$

Remarque : Ce lemme joue en quelque sorte le même role que l'indépendance des termes d'une somme de variables, si on oublie le "sup".

Lemme utile 1.2. *Soit X une variable aléatoire telle que $\mathbb{E}(X) = 0$ et $a \leq X \leq b$, alors $\forall h > 0, \mathbb{E}(e^{hX}) \leq e^{\frac{1}{8}h^2(b-a)^2}$. Ce résultat reste vrai avec des espérances conditionnelles (la condition est alors $\mathbb{E}(X|\mathcal{G}) = 0$) .*

Démonstration. (théorème 1.1) :

La preuve du théorème (1.1) suit un schéma "traditionnel", comme on le verra dans la preuve directe de l'inégalité d'Hoeffding, basée en trois étapes : L'inégalité de Markov exponentielle, la majoration du terme exponentielle par indépendance (ou dans le cas présent par le lemme utile (1.1), qui joue le même rôle), puis par le lemme utile (1.2). Enfin l'optimisation en h :

Soit $X_k = \mathbb{E}(X|\mathcal{F}_{k-1})$ et $Y_k = X_k - X_{k-1}$ la différence de martingales associée. On pose $Z = \mathbb{1}_{R^2 \leq r^2}$.
L'inégalité de Markov exponentielle donne,

$$\begin{aligned}
\forall h,\ \mathbb{P}((X-\mu \geq t) \cap (R^2 \leq r^2)) &= \mathbb{P}(Ze^{h(X-\mu)} \geq e^{ht}) \\
&\leq e^{-ht}\mathbb{E}(Ze^{h(X-\mu)}) \\
&\leq e^{-ht}\mathbb{E}(Ze^{h(\sum Y_k)})
\end{aligned}$$

Par le lemme utile (1.2), $\mathbb{E}(e^{hY_k}|\mathcal{F}_{k-1}) \leq e^{\frac{1}{8}h^2r_k^2}$
D'où par le lemme utile (1.1) :

$$\begin{aligned}
\mathbb{E}(Ze^{h\sum Y_k}) &\leq sup(Z\prod \mathbb{E}(e^{hY_k}|\mathcal{F}_{k-1})) \\
&\leq sup(Z\prod e^{\frac{1}{8}h^2r_k^2}) \\
&= sup(Ze^{\frac{1}{8}h^2R^2}) \\
&\leq e^{\frac{1}{8}sup(ZR^2)} \\
&\leq e^{\frac{1}{8}h^2r^2}
\end{aligned}$$

Finalement, on obtient :

$$\mathbb{P}((X-\mu \geq t) \cap (R^2 \leq r^2)) \leq e^{-ht+\frac{1}{8}h^2r^2} \leq e^{-2t^2/r^2}$$

en posant $h = \frac{4t}{r^2}$.

□

Théorème 1.2. *Soit X une variable aléatoire avec $\mathbb{E}(X) = \mu$ et $(\mathcal{F}_k)_{0\leq k\leq n}$ une filtration de $\mathcal{F}$ avec $\mathcal{F}_0 = (\emptyset, \Omega)$ et telle que X est $\mathcal{F}_n$-mesurable. Soit $b = maxdev^+$ la déviation conditionnelle maximale supposée finie, et $\hat{\nu} = supessV$ la somme maximale des variances conditionnelles supposée finie également.*
Alors $\forall t \geq 0,\ \mathbb{P}(X-\mu \geq t) \leq e^{-\frac{t^2}{2(\hat{\nu}+bt/3)}}$
et plus généralement $\forall t \geq 0, \forall v \geq 0,\ \mathbb{P}((X-\mu \geq t) \cap (V \leq v)) \leq e^{-\frac{t^2}{2(v+bt/3)}}$.

Nous avons besoin de deux lemmes : le lemme utile (1.1) ci-dessus exploitant la décomposition en somme de différences de martingales et remplacant l'indépendance des termes de cette somme dans la preuve, et le lemme suivant qui joue le même rôle que le lemme utile (1.2) pour des variables non necessairement bornées mais à variance bornée :

Lemme utile 1.3. *Soit g la fonction croissante définie par $\forall x \neq 0,\ g(x) = \frac{e^x-1-x}{x^2}$, et X une variable aléatoire satisfaisant $\mathbb{E}(X) = 0$ et $X \leq b$.*
Alors $\mathbb{E}(X) \leq e^{g(b)var(X)}$. Ce résultat reste vrai avec des espérances et des variances conditionnelles en remplacant b par $sup(X|\mathcal{G})$.

La preuve de ce lemme est immédiate en remarquant que $e^x \leq 1+x+x^2g(b)$ pour $x \leq b$.

Démonstration. (théorème (1.2)) : Elle suit le même schéma classique que le théorème (1.1). Soient $Y_1, ..., Y_n$ les différences de martingales associées à X et à $(\mathcal{F}_k)$.

On pose $Z = \mathbb{1}_{V \leq v}$.
L'inégalité de Markov exponentielle donne,

$$\begin{aligned} \forall h,\ \mathbb{P}((X-\mu \geq t) \cap (V \leq v)) &= \mathbb{P}(Ze^{h(X-\mu)} \geq e^{ht}) \\ &\leq e^{-ht}\mathbb{E}(Ze^{h(X-\mu)}) \\ &\leq e^{-ht}\mathbb{E}(Ze^{h(\sum Y_k)}) \end{aligned}$$

Par le lemme utile (1.3), $\mathbb{E}(e^{hY_k}|\mathcal{F}_{k-1}) \leq e^{h^2g(hdev_k^+)var_k} \leq e^{h^2g(hb)var_k}$
D'où par le lemme utile (1.1) :

$$\begin{aligned} \mathbb{E}(Ze^{h\sum Y_k}) &\leq sup(Z\prod \mathbb{E}(e^{hY_k}|\mathcal{F}_{k-1})) \\ &\leq sup(Z\prod e^{h^2g(hb)var_k}) \\ &= sup(Ze^{h^2g(hb)V}) \\ &\leq e^{h^2g(hb)sup(ZV)} \\ &\leq e^{h^2g(hb)v} \end{aligned}$$

Finalement, on obtient :

$$\begin{aligned} \mathbb{P}((X-\mu \geq t) \cap (R^2 \leq r^2)) &\leq e^{-ht+h^2g(hb)v} \\ &\leq e^{-\frac{t^2}{2(v+bt/3)}} \end{aligned}$$

en posant $h = \frac{1}{b}ln(1+\frac{bt}{v})$ puis en utilisant que $\forall x \geq 0,\ (1+x)ln(1+x)-x \geq 3x^2/(6+2x)$.

□

1.2 Inégalités célèbres

1.2.1 Sans hypothèse de variance

Théorème 1.3. *(Inégalité de Azuma-Hoeffding) Soit $(\mathcal{F}_k)_{0\leq k\leq n}$ une filtration de $\mathcal{F}$ avec $\mathcal{F}_0 = (\emptyset, \Omega)$, soit Z une martingale et Y la différence de martingale associée. Si $\forall k,\ |Y_k| \leq c_k$, alors :*

$\mathbb{P}(|\sum Y_k| \geq t) \leq 2e^{-\frac{t^2}{2\sum c_k^2}}$

Démonstration. On veut appliquer le théorème (1.1) en posant $X = \sum_1^n Y_k$, $\mathcal{F}_k = \sigma(Y_1, ..., Y_k)$ et $X_k = \mathbb{E}(X|\mathcal{F}_k)$: On a alors $\mu = 0$, $X_k = \mathbb{E}(X|\mathcal{F}_k) = \sum_1^k Y_i$ car Z martingale, et $Y_i = X_i - X_{i-1}$.
Donc $ran_k = ran(Y_k|\mathcal{F}_k) \leq 2c_k$ d'où $R^2 \leq 4\sum c_k^2$ et $\hat{r}^2 \leq 4\sum c_k^2$.
Par le théorème 1.1 $\quad \mathbb{P}(X - \mu \geq t) \leq e^{\frac{-2t^2}{\hat{r}^2}} \leq e^{-\frac{t^2}{2\sum c_k^2}}$
En appliquant cette inégalité à -X, on obtient $\mathbb{P}(|\sum Y_k| \geq t) \leq 2e^{-\frac{t^2}{2\sum c_k^2}}$. □

Théorème 1.4. *(Inégalité de McDiarmid) Soit $X = (X_1, ..., X_n)$ où les X_i sont indépendants à valeurs dans A_i. Soit $f : \prod A_k \to \mathbb{R}$ vérifiant la condition de Lipschitz :*

$$\forall x, x' \in \prod_1^n A_k, \ |f(x) - f(x')| \leq c_k \quad \textit{si x et x' ne diffèrent seulement de la kième coordonnée.} \quad (1)$$

On note $\mu = \mathbb{E}(f(X))$. Alors $\forall t \geq 0, \ \mathbb{P}(f(X) - \mu \geq t) \leq e^{-2t^2/\sum c_k^2}$
et donc $\forall t \geq 0, \ \mathbb{P}(|f(X) - \mu| \geq t) \leq 2e^{-2t^2/\sum c_k^2}$

Démonstration. La condition de Lipschitz (1) implique que f est bornée, donc par le thèorème (1.1) on a :
$\mathbb{P}(f(X) - \mu \geq t) \leq e^{-2t^2/\hat{r}^2}$ où $\hat{r}^2$ est défini en posant $\mathcal{F}_k = \sigma(X_1, ..., X_k)$ et $X = f(X_1, ..., X_n)$
On remarque que cette inégalité est vraie sans hypothèse d'indépendance ni de la condition de Lipschitz, seulement avec l'hypothèse f bornée. L'hypothèse de Lipschitz et l'indépendance vont permettent de majorer $\hat{r}^2$:
$ran_k = ran(\ \mathbb{E}(f(X)|\mathcal{F}_k) - \mathbb{E}(f(X)|\mathcal{F}_{k-1}) \ \ |\mathcal{F}_{k-1}) \leq c_k$
□

Théorème 1.5. *(Inégalité de Hoeffding) Soient $X_1, ..., X_n$ variables indépendantes telles que $\forall i, \ a_i \leq X_i \leq b_i$. Soit $S_n = \sum X_k$ et $\mu = \mathbb{E}(S_n)$.*
Alors $\mathbb{P}(|S_n - \mu| \geq t) \leq 2e^{-2t^2/\sum(b_k - a_k)^2}$.

Démonstration. Il s'agit d'une conséquence de l'inégalité de McDiarmid avec $A_k = [a_k, b_k]$, $f(x) = \sum x_k$ et $c_k = b_k - a_k$. On a alors $\hat{r}^2 \leq b_k - a_k$. □

Remarque : La preuve directe de ce résultat suit le même schéma que celle du théorème (1.1) : Inégalité de Markov exponentielle, puis hypothèse de somme de variables indépendantes (ou de différences de martingales), et enfin utilisation du lemme utile (1.2) avant l'optimisation en h :

$$\begin{array}{rcl} \mathbb{P}(S_n - \mu \geq t) & \leq & \mathbb{E}(e^{h(S_n - \mu)})e^{-ht} \\ \mathbb{E}(\prod e^{h(X_k - \mathbb{E}X_k)}) & = & \prod \mathbb{E}(e^{h(X_k - \mathbb{E}X_k)}) \text{ par indépendance} \\ & \leq & e^{\frac{1}{8}h^2\sum(b_k - a_k)^2} \text{ par le lemme utile (1.2)} \end{array}$$

puis on pose $h = \frac{4t}{\sum(b_k - a_k)^2}$.

Remarque : On se rend compte en comparant ces deux dernières inégalités au théorème (1.1) que la décomposition en somme de différences de martingales permet une généralisation du cas d'une somme de variables indépendantes. Sous réserve d'introduire des outils de contrôle plus précis comme $\hat{r}^2$, on n'a plus besoin d'indépendance ni de condition de Lipschitz.
Ces deux hypothèses supplémentaires peuvent être vues comme une manière de majorer $\hat{r}^2$ pour obtenir une forme plus simple.

1.2.2 Variance bornée

Théorème 1.6. *(Bernstein) Soient $X_1, ..., X_n$ des variables aléatoires indépendantes avec $X_k - \mathbb{E}(X_k) \leq b$. On pose $S_n = \sum X_k$, $V = var(S_n)$ et $\mathbb{E}(S_n) = \mu$.*

$$\begin{array}{rl} \textit{Alors :} & \forall t \geq 0,\ \mathbb{P}(S_n - \mu \geq t) \leq e^{-\frac{t^2}{2(V+bt/3)}} \\ \textit{et plus généralement :} & \mathbb{P}((S_n - \mu \geq t) \cap (V \leq v)) \leq e^{-\frac{t^2}{2(v+bt/3)}} \end{array}$$

Démonstration. On pose $F_k = \sigma(X_1, ..., X_n)$, $X = \sum (X_k - \mathbb{E}X_k) = S_n - \mu$, $\tilde{X}_k = \mathbb{E}(X|\mathcal{F}_k) = \sum_1^k (X_i - \mathbb{E}X_i)$ et $Y_k = \tilde{X}_k - \tilde{X}_{k-1}$.
Alors $Y_k = X_k - \mathbb{E}X_k$ d'où $dev_k^+ \leq b$ puis $maxdev^+ \leq b$ et $var_k = var(Y_k|\mathcal{F}_{k-1}) = \mathbb{E}((Y_k - \mathbb{E}(Y_k|\mathcal{F}_{k-1}))^2|\mathcal{F}_{k-1}) = \mathbb{E}((Y_k - \mathbb{E}Y_k)^2) = var(Y_k)$ donc $\hat{\nu} = supess(\sum var_k) = supess(V) = V$.
Le théorème (1.2) s'applique donc et donne :

$$\mathbb{P}(S_n - \mu \geq t) \leq e^{-\frac{t^2}{2(V+bt/3)}}$$
$$\mathbb{P}((S_n - \mu \geq t) \cap (V \leq v)) \leq e^{-\frac{t^2}{2(v+bt/3)}}$$

□

Théorème 1.7. *Soit $(Y_k)_{1 \leq k \leq n}$ une différence de martingale telle que $-a_k \leq Y_k \leq 1 - a_k$, $a = \frac{1}{n}\sum a_k$*

(i) $\forall t \geq 0,\ \mathbb{P}(|\sum Y_k| \geq t) \leq 2e^{-2t^2/n}$

(ii) $\forall \epsilon > 0,\ \mathbb{P}(\sum Y_k \geq \epsilon an) \leq e^{-\frac{\epsilon^2 an}{2(1+\epsilon/3)}}$

(iii) $\forall \epsilon > 0,\ \mathbb{P}(\sum Y_k \leq -\epsilon an) \leq e^{-\frac{1}{2}\epsilon^2 an}$

Démonstration. (i) est une conséquence immédiate du théorème (1.1), puisque : $-a_k \leq Y_k \leq 1 - a_k \quad \Rightarrow \quad ran_k \leq 1 \quad \Rightarrow \quad \hat{r}^2 \leq n$

(ii) On pose $X = Z_n$, $\mathcal{F}_k = \sigma(Z_1, ..., Z_k)$ et $\forall 1 \leq k \leq n, X_k = \mathbb{E}(X|\mathcal{F}_k)$.
On a $X_k = Z_k$ car Z_k martingale, donc $Y_k = X_k - X_{k-1}$.
Comme $-a_k \leq Y_k \leq 1 - a_k$, on a $dev_k^+ = sup(Y_k|\mathcal{F}_{k-1}) \leq 1$ et donc $b = maxdev^+ \leq 1$.
Soit $W_k := Y_k + a_k$. On a $0 \leq W_k \leq 1$, $\mathbb{E}(W_k|\mathcal{F}_{k-1}) = a_k$ et $var(Y_k|\mathcal{F}_{k-1}) = var(W_k|\mathcal{F}_{k-1})$. Or :

$$\begin{aligned} var(W_k|\mathcal{F}_{k-1}) &= \mathbb{E}(W_k^2|\mathcal{F}_{k-1}) - \mathbb{E}(W_k|\mathcal{F}_{k-1})^2 \\ &\leq \mathbb{E}(W_k|\mathcal{F}_{k-1}) - \mathbb{E}(W_k|\mathcal{F}_{k-1})^2 \\ &= a_k - a_k^2 \\ &= a_k(1 - a_k) \\ &\leq a_k \end{aligned}$$

Donc $V = \sum var_k \leq \sum a_k = na$ d'où $\hat{\nu} \leq na$
Le théorème (1.2) donne donc :

$$\mathbb{P}(Z_n - Z_0 \geq t) \leq e^{-\frac{t^2}{2(\hat{\nu}+bt/3)}} \leq e^{-\frac{t^2}{2(an+t/3)}}$$

$$\text{d'où il suit que :} \quad \mathbb{P}(\sum Y_k \geq \epsilon an) \leq e^{-\frac{\epsilon^2 a^2 n^2}{2(an+\epsilon an/3)}} = e^{-\frac{\epsilon^2 an}{2(1+\epsilon/3)}}$$

(iii) On a $-a_k \leq Y_k \leq 1-a_k \quad \Rightarrow \quad -\bar{a}_k \leq -Y_k \leq 1-\bar{a}_k$ où $\bar{a}_k = 1-a_k$. En posant $\bar{a} = 1-a$, (ii) appliqué à $(-Y_k, \bar{a}_k)$ donne :

$$\forall \epsilon > 0,\ \mathbb{P}(\sum(-Y_k) \geq \epsilon \bar{a} n) \leq e^{-\frac{\epsilon^2 \bar{a} n}{2(1+\epsilon/3)}}.$$

Avec $\epsilon = ta/\bar{a}$ on obtient :

$$\forall t > 0,\ \mathbb{P}(\sum(-Y_k) \geq tan) \leq e^{-\frac{t^2 an}{2(a\bar{a}+a^2 t/3)}}$$

Comme $-Y_k \leq a_k$ on a $\sum(-Y_k) \leq an\ p.s.$ donc :

$$\forall t > 1,\ \mathbb{P}(\sum(-Y_k) \geq tan) = 0 \tag{2}$$

De plus : $\quad \forall t \in]0,1],\ 2(a\bar{a} + a^2 t/3) \leq 2(1/4+1/3) \leq 2$

Donc :

$$\forall t \in]0,1],\ \mathbb{P}(\sum(-Y_k) \geq tan) \leq e^{-\frac{1}{2}t^2 an}$$

ce qui est aussi valable pour $t > 1$ par (2). Ainsi,

$$\begin{aligned} \forall t > 0,\ \mathbb{P}(-\sum Y_k \geq tan) &\leq e^{-\frac{1}{2}t^2 an} \\ \forall t > 0,\ \mathbb{P}(\sum Y_k \leq -tan) &\leq e^{-\frac{1}{2}t^2 an} \end{aligned}$$

□

On déduit immédiatement de ce théorème (avec $a_k = \mathbb{E}(X_k)$, $Y_k = X_k - a_k$, ce qui donne $-a_k \leq Y_k \leq 1-a_k$ et $\sum Y_k = S_n - \mu$) le théorème bien connu suivant :

Théorème 1.8. *Soient $X_1, ..., X_n$ variables aléatoires indépendantes avec $0 \leq X_k \leq 1$, $S_n = \sum X_k$, $\mu = \mathbb{E}S_n$ et $p = \mu/n$. Alors :*
(i) $\forall t > 0,\ \mathbb{P}(S_n \geq \mu + t) \leq e^{-\frac{t^2}{2(\mu+t/3)}}$
(ii) $\forall t > 0,\ \mathbb{P}(S_n \leq \mu - t) \leq e^{-\frac{t^2}{2\mu}}$

2 Inégalité de Talagrand

2.1 L'inégalité

Définition 2.1. *Soit $x = (x_1, ..., x_n) \in \prod A_i$ et $y = (y_1, ..., y_n) \in \prod A_i$. Soit $\alpha = (\alpha_1, ..., \alpha_n) \in \mathbb{R}^n$, $\alpha_i \geq 0$. On définit :*

$$\begin{aligned} d_\alpha(x,y) &= \sum_1^n \alpha_i \mathbb{1}_{x_i \neq y_i} \\ d_\alpha(x,A) &= \inf_{y \in A} d_\alpha(x,y) \\ d_T(x,y) &= \sup_{\|\alpha\|=1} d_\alpha(x,A) \end{aligned}$$

Théorème 2.1. *(Inégalité de Talagrand) Soit $X = (X_1, ..., X_n)$ où les X_i sont des variables aléatoires indépendantes, $X_i : \Omega \to A_i$. Soit $B \subset A = \prod A_i$. Alors :*

$$\mathbb{P}(X \in B)\mathbb{P}(d_T(X, B) \geq t) \leq e^{-t^2/4}$$

Remarque : Avec la méthode des différences bornées (Inégalité de McDiarmid 1.4), on peut montrer que pour tout vecteur α unitaire positif :

$$\mathbb{P}(X \in B)\mathbb{P}(d_\alpha(X, B) \geq t) \leq e^{-t^2/2}.$$

Cependant, l'inégalité de Talagrand est beaucoup plus puissante car elle se réfère simultanément à tous les vecteurs α.

Démonstration. (Résumé)

On montre que $\mathbb{P}(X \in B)\mathbb{E}(e^{\frac{1}{4}d_T(X,B)^2}) \leq 1$ ce qui donne le résultat par Markov exponentielle.

1ère étape : Définition équivalente de d_T :

Soit $x \in \mathbb{R}^n, B \subset \mathbb{R}^n, U = U(x, B)$ l'ensemble des vecteurs binaires u tels que partant de x, on peut atteindre un vecteur $y \in B$ en changeant seulement des coordonnées x_i telles que $u_i = 1$, i.e : $U(x, B) = \{u \in \{0,1\}^n, \exists y \in B, \forall i \;\; t.q \;\; u_i = 0, y_i = x_i\}$. Soit $V = V(x, B)$ l'enveloppe convexe de U. Alors $d_T(x, B) = \min_{v \in V} \|v\|$.

2ème étape : Deux lemmes techniques pour l'étape suivante.

3ème étape : Preuve principale par récurrence sur n :

(i) Lien entre $d_T(x, B)$ en dimension $n + 1$ avec des distances ne prenant en compte que les n premières coordonnées, grâce à des "ω-section de $B \subset \prod_1^{n+1} A_k$" : $B_\omega = \{x \in \prod_1^n A_k, (x, \omega) \in B\}$ et des projections de B : $\hat{B} = \{x \in \prod_1^n, (x, \omega) \in B$ pour un $\omega \in A_{n+1}\}$. Ceci fait apparaître un paramètre λ :

Soit $z = (x, \omega) \in (\prod_1^n A_k) \times A_{n+1}$ et soit $0 \leq \lambda \leq 1$. Alors :

$$d_T(z, B)^2 \leq \lambda d_T(x, B_\omega)^2 + (1 - \lambda)d_T(x, \hat{B})^2 + (1 - \lambda)^2. \qquad (3)$$

(ii) (3) nous permet d'appliquer l'hypothèse de récurrence. Tout d'abord :

$$\mathbb{E}(e^{\frac{1}{4}d_T((X,X_{n+1}),B)^2}) \;\; = \;\; \mathbb{E}(\mathbb{E}(E(X_{n+1})|\mathcal{F}_{n+1}))$$

où :

$$\begin{aligned} E(\omega) &= \mathbb{E}(e^{\frac{1}{4}d_T((X,\omega),B)^2}) \\ &\leq e^{\frac{1}{4}(1-\lambda)^2}\, \mathbb{E}(e^{(\frac{1}{4}d_T(X,B_\omega))^2})^\lambda \, \mathbb{E}(e^{\frac{1}{4}d_T(X,\hat{B})^2})^{1-\lambda} \text{ par (3) et l'inégalité de Hölder.} \\ &\leq e^{\frac{1}{4}(1-\lambda)^2}(\nu_n(B_\omega))^{-\lambda}(\nu_n(\hat{B}))^{-(1-\lambda)} \\ &\qquad \text{par hypothèse de récurrence, où } \nu_n(A) = \mathbb{P}(X \in A) \end{aligned}$$

(iii) L'optimisation en λ donne :

$$E(\omega) \quad \leq \quad (\nu_n(\hat{B}))^{-1}(2 - \frac{\nu_n(B_\omega)}{\nu_n(\hat{B})})$$

Comme $\nu_n(B_{X_{n+1}}) = \nu_{n+1}(B)$, on obtient :

$$\nu_{n+1}(B)\ \mathbb{E}(e^{\frac{1}{4}d_T((X,X_{n+1}),B)^2}) \leq x(2-x) \leq 1$$

où : $x = \frac{\nu_{n+1}(B)}{\nu_n(\hat{B})}$, ce qui termine la preuve.

□

2.2 Un lemme utile

L'inégalité de Talagrand permet souvent d'obtenir des inégalités de concentration autour de la médiane. Le lemme suivant permet de retrouver des concentrations autour de l'espérance :

Lemme utile 2.1. *Soit X une variable aléatoire de moyenne μ et de médiane m. Soient $a, b > 0$*

(i) *Si* $\forall t > 0,\ \mathbb{P}(X - m \geq t) \leq ae^{-t^2/b}$ *alors* $\mu - m \leq (\sqrt{\pi}/2)a\sqrt{b}$

Si $\forall t > 0,\ \mathbb{P}(X - m \leq -t) \leq ae^{-t^2/b}$ *alors* $\mu - m \geq -(\sqrt{\pi}/2)a\sqrt{b}$

(ii) *Si* $\forall t > 0,\ \mathbb{P}(X - m \geq t) \leq ae^{-t^2/b(m+t)}$ *alors* $\mu - m \leq \sqrt{\pi/2}a\sqrt{bm} + 2abe^{-m/2b}$

Démonstration. Facile en remarquant que :
$\mu - m = \mathbb{E}(Y - m) \leq \mathbb{E}((Y-m)^+) = \int_0^\infty \mathbb{P}(Y - m > t)dt.$ □

3 Applications

3.1 Applications de la méthode des différences bornées au nombre chromatique

Soit $G_{n,p}$ un graphe aléatoire à n sommets, tel que pour chaque pairs de ces sommets, il existe une arête les reliant avec probabilité p. On note $E(G_{n,p})$ l'ensemble des arêtes de $G_{n,p}$. Soit $(A_1, ..., A_m)$ une partition de l'ensemble des arêtes possibles (i.e de l'ensemble des arêtes du graphe complet $\mathbb{K}_n$). Alors $E(G_{n,p})$ peut s'écrire $(E_1, ..., E_m)$ avec $E_i \in \mathcal{P}(A_i)$ et on identifie par la suite un graphe sur les n sommets avec l'ensemble de ses arêtes qui s'écrit de manière unique dans $\prod_1^m \mathcal{P}(A_i)$.

Théorème 3.1. *Soit f une fonction sur l'ensemble des graphes sur les n sommets telle que*

$$\exists k,\ E(G)\Delta E(G') \subset A_k \quad \Rightarrow \quad |f(G) - f(G')| \leq 1 \tag{4}$$

Alors la variable $Y := f(G_{n,p})$ vérifie :

$$\forall t > 0,\ \mathbb{P}(Y - \mathbb{E}Y \geq t) \leq e^{-2t^2/m}$$

Démonstration. Il s'agit de l'inégalité de McDiarmid (1.4) appliquée à $X = E(G_{n,p}) = (E_1, ..., E_m) \in \prod_1^m \mathcal{P}(A_i)$ avec $\forall 0 \leq k \leq m,\ c_k = 1$.
En effet, la condition (4) signifie simplement que si $E(G)$ et $E(G')$ ne diffèrent que de la kième coordonnée, alors $|f(G) - f(G')| \leq c_k = 1$. $\square$

Corollaire 3.2. *Supposons que $|f(G) - f(G')| \leq 1$ dès que G et G' ne diffèrent que d'une arête. Alors :*

$$\forall t > 0, \mathbb{P}(f(G_{n,p}) - \mathbb{E}f(G_{n,p}) \geq t) \leq e^{-4t^2/n^2}$$

Démonstration. Il s'agit du théorème (3.1) appliqué avec $m = \begin{pmatrix} n \\ 2 \end{pmatrix}$ et $\forall i, |A_i| = 1$:

$$\forall t > 0, \mathbb{P}(f(G_{n,p}) - \mathbb{E}G_{n,p} \geq t) \leq e^{-\frac{2t^2}{n(n-1)/2}} \leq e^{-4t^2/n^2}$$

$\square$

On peut appliquer ce Corollaire (3.2) à un résultat de Bollobas :

$$\mathbb{P}((1-\epsilon)\frac{n}{2log_b n} \leq \chi(G_{n,p}) \leq (1+\epsilon)\frac{n}{2log_b n})) \xrightarrow[n\infty]{} 1 \tag{5}$$

où $b = \frac{1}{q} = \frac{1}{1-p}$.

Nous allons prouver la borne supérieure de (5). Soit $\tilde{p}(n)$ la probabilité que $G_{n,p}$ ne contienne aucun stable de $s(n) = \lceil (2-\epsilon) log_b n \rceil$ sommets.

Montrons tout d'abord que $\tilde{p}(n) = \Theta(e^{-n^{4/3}})$. Etant donné un graphe G à n sommets, on définit $f(G)$ comme étant le nombre maximum de stables de taille $s(n)$ qui n'ont deux à deux qu'un sommet commun. f vérifie $|f(G) - f(G')| \leq 1$ dès que G et G' ne diffèrent que d'une arête, donc $\tilde{p}(n) = \mathbb{P}(f(G_{n,p}) = 0) = \mathbb{P}(f(G_{n,p}) - \mu_n \leq -\mu_n) \leq e^{-4\mu_n^2/n^2}$ où $\mu_n = \mathbb{E}f(G_{n,p})$. On peut montrer que $\mu_n \geq n^{5/3}$ pour n assez grand donc $\tilde{p}(n) \leq e^{-4n^{10/3-2}} = e^{-4n^{4/3}}$.

Soit $\tilde{n} = \lceil \frac{n}{log^2 n} \rceil$. Nous dirons qu'un ensemble W d'au moins $\tilde{n}$ sommets de $G_{n,p}$ est "mauvais" si il ne contient aucun stable de taille supérieure à $s(\tilde{n})$. La probabilité qu'il y ait un ensemble mauvais est au plus $2^n \tilde{p}(\tilde{n}) = o(1)$. Mais s'il n'y en a pas, on peut colorier un stable de taille $s(\tilde{n})$ et le supprimer, et recommencer jusqu'à ce qu'il reste moins de $\tilde{n}$ sommets, qu'on colorie tous d'une couleur différente. Le nombre total de couleurs utilisées est inférieur à :

$$\begin{aligned}
\frac{n}{s(\tilde{n})}+\tilde{n} &= \frac{n}{\lceil(2-\epsilon)log_b(\tilde{n})\rceil}+\tilde{n}\\
&\leq \frac{n}{(2-\epsilon)log_b(\frac{n}{log^2n}-1)}+\frac{n}{log^2n}\\
&\leq \frac{n}{log_bn}(\frac{1}{2-\epsilon}+o(1))
\end{aligned}$$

D'où $\chi(G_{n,p})\frac{2log_bn}{n} \leq \frac{1}{1-\frac{\epsilon}{2}}+o(1)$ et comme $\frac{1}{1-\frac{\epsilon}{2}} < 1+\epsilon$ on obtient bien :

$$\mathbb{P}(\chi(G_{n,p})\frac{2log_bn}{n} \leq (1+\epsilon)) \underset{n\infty}{\longrightarrow} 1.$$

3.2 Applications de l'inégalité de Talagrand

Théorème 3.3. *Soient $X_1,...,X_n$ des variables aléatoires indépendantes, $X_k \in A_k$, $A=\prod_1^n A_k$. On pose $X=(X_1,...,X_n)$ et $f : A \to \mathbb{R}$ telle que*

$$\forall x \in A, \exists \alpha \in \mathbb{R}_+^n, \|\alpha\|=1, \forall y \in A, f(x) \leq f(y)+cd_\alpha(x,y) \tag{6}$$

Alors $\mathbb{P}(|f(X)-m| \geq t) \leq 4e^{-t^2/4c^2}$ où m médiane de $f(X)$. La même conclusion est valable si on remplace (6) par $f(y) \leq f(x)+cd_\alpha(x,y)$.

Remarque : Il existe un théorème analogue pour f c-configuration, i.e. telle que $\forall x,\ \exists\alpha,\ \forall y,\ f(x) \leq f(y)+\sqrt{cf(x)}d_\alpha(x,y)$. On a alors, en notant $m =$ médiane$(f(X))$:

$$\mathbb{P}(f(X) \geq m+t) \leq 2e^{-\frac{t^2}{4c(m+t)}}$$
$$\mathbb{P}(f(X) \leq m-t) \leq 2e^{-\frac{t^2}{4cm}}$$

Démonstration. (théorème (3.3)) Pour tout $a \in \mathbb{R}$, on pose $A_a=\{y \in \mathbb{R}, f(y) \leq a\}$. Soit $x \in A$. Il existe $\alpha \in (\mathbb{R}_+^n), \|\alpha\|=1$, tel que $\forall y \in A, f(x) \leq f(y)+cd_\alpha(x,y)$. D'où :

$$\begin{aligned}
\forall y \in A_a, f(x) \leq a+cd_\alpha(x,y) &\Rightarrow f(x) \leq a+cd_\alpha(x,A_a)\\
&\Rightarrow f(x) \leq a+cd_T(x,A_a)
\end{aligned}$$

donc $f(x) \geq a+t \Rightarrow d_T(x,A_a) \geq t/c$

$$\begin{aligned}
\text{d'où } \mathbb{P}(f(x)\leq a)\mathbb{P}(f(x)\geq a+t) &\leq \mathbb{P}(X \in A_a)\mathbb{P}(d_T(X,A_a)\geq t/c)\\
&\leq e^{-\frac{t^2}{4c^2}}
\end{aligned}$$

$$\text{Pour } a=m,\ \mathbb{P}(f(X)\geq m+t) \leq 2e^{-\frac{t^2}{4c^2}}$$
$$\text{Pour } a=m-t,\ \mathbb{P}(f(X)\leq m-t) \leq 2e^{-\frac{t^2}{4c^2}}$$

Pour la deuxième condition, il suffit de considérer $g := -f$. $\square$

Les trois points suivants sont des applications de ce théorème 3.3.

3.2.1 Problème du voyageur de commerce

Soit $A = ([0,1]^2)^n$. Pour $x \in A$, on note $tsp(x)$ la longueur minimale d'un cycle passant par tous ces points.

On utilise le résultat déterministe suivant :

(7)

> Il existe $c \geq 0$ tel que :
> $\forall n \geq 0, \forall x \in A, \exists\ T(x)$ cycle recouvrant x, tel que la somme des carrés des longueurs des arêtes de ce tour est au plus c.

Soit $x \in A$. On note β_k la somme des longueurs des deux arêtes partant de x_k dans $T(x)$. Ainsi $\sum_1^n \beta_k^2 \leq 4c$ (en utilisant que $(a+b)^2 \leq 2a^2 + 2b^2$.)

Propriété 1. *$\forall y \in A$, $tsp(x) \leq tsp(y) + d_\beta(x,y)$. Ainsi, en notant $\alpha = \frac{\beta}{\|\beta\|}$, on a $tsp(x) \leq tsp(y) + (2\sqrt{c})d_\alpha(x,y)$ d'où par le théorème précédent (3.3), $\mathbb{P}(|tsp(X) - m| \geq t) \leq 4e^{-\frac{t^2}{16c}}$*

Démonstration. On note de la même façon le vecteur $x \in A$ et l'ensemble $\{x_1, ..., x_n\}$.
Si $x \cap y = \emptyset$, $d_\beta(x,y) = 2|T(x)|$ donc la propriété est vraie.
Sinon, $x \cap y \neq \emptyset$, on construit l'ensemble d'arête F de la façon suivante : Pour tout segment de $T(x)$ de la forme $a, v_1, ..., v_j, b$ où $a, b \in y$ et $v_i \notin y$, on met dans F les arêtes $av_1, v_1v_2, ..., v_{j-1}v_j, v_jv_{j-1}, ..., v_1a$ si $av_1 < v_jb$ et la même chose en remplaçant av_1 et v_1a par bv_j et v_jb sinon.
Ainsi, on met dans F un cycle contenant exactement un point appartenant à y et dont la longueur n'excède pas la somme des coordonnées de β correspondant aux v_i. (C'est pour cela qu'on a mis $av_1 \wedge bv_j$ dans F)

Tous ces cycles contenus dans F contiennent tous les points de $x \backslash y$ et leur longueur totale est inférieure ou égale à $d_\beta(x,y)$. De plus, ils contiennent chacun exactement un point de y. Il suffit maintenant d'ajouter ces cycles à $T^*(y)$, un cycle optimal pour y, pour former un graphe contenant tous les sommets $x \cup y$ et de longueur plus petite que $|T^*(y)| + d_\beta(x,y)$.
Ce graphe formé de $T^*(y)$, sur lequel on a ajouté des cycles partant de certains points de $T^*(y)$, donne un cycle Eulérien et en supprimant les points déjà visités du parcours (par exemple si b a déjà été visité et qu'il est visité de nouveau avec les arêtes ab et bc, on remplace ces deux arêtes par une seule : l'arête ac), on obtient un parcours du voyageur de commerce d'une longueur inférieure (par l'inégalité triangulaire), et recouvrant ainsi tous les points $x \cup y$, en particulier ceux de x, d'où :

$$tsp(x) \leq |T^*(y)| + d_\beta(x,y) \leq tsp(y) + d_\beta(x,y)$$

$\square$

3.2.2 Arbres de Steiner

Un arbre de Steiner pour $x \in A = ([0,1]^2)^n$ est un arbre dont les sommets contiennent $\{x_1, ..., x_n\} =: x$. On note $st(x)$ la longueur minimale d'un arbre de Steiner pour x. Nous allons utiliser le résultat déterministe (7) admis dans l'application précédente, notamment le cycle $T(x)$ correspond à la constante c, et le β correspondant. Soit $y \in A$ et $S^*(y)$ un arbre de Steiner optimal. On rajoute à $S^*(y)$ les arêtes de $T(x)$ ayant au moins une extrémité hors de y. On augmente ainsi la longueur de $S^*(y)$ d'au plus $d_\beta(x,y)$ et on forme ainsi un graphe recouvrant $x \cup y$, connexe, de longueur inférieure ou égale à $st(y) + d_\beta(x,y)$. Pour obtenir un arbre, il suffit d'enlever certaines arêtes pour éliminer les cycles, d'où $st(x) \leq st(y) + d_\beta(x,y)$, et par le théorème (3.3),

$$\mathbb{P}(|st(X) - m| \geq t) \leq 4e^{-\frac{t^2}{16c}}$$

3.2.3 Arbres couvrants aléatoires minimaux

Soit K_n le graphe complet à n sommets avec des arêtes de longueur $(X_e)_{1 \leq e \leq N}$ où $N = \begin{pmatrix} n \\ 2 \end{pmatrix}$ et les X_e sont uniforméments distribués sur $[0,1]$. Soit $L_n :=$ $mst(X)$ la longueur minimale d'un arbre couvrant. On admet que : $\mathbb{E}(L_n) \underset{n\infty}{\longrightarrow}$ $\sum_1^\infty j^{-3} = \zeta(3)$. Soit $L_n^{(b)}$ la longueur minimal d'un arbre couvrant quand X_e est remplacé par $Y_e = min(X_e, b)$ i.e. $L_n^{(b)} = mst(Y)$. On admet le lemme suivant, qui montre que les grandes arêtes ne jouent pas un rôle déterminant :

Lemme 1.

$$\forall t > 0, \quad \exists c_1 > 0, \quad \exists \nu > 0, \quad \forall n \geq 0, \quad \mathbb{P}(L_n - L_n^{(b)} \geq t) \;\leq\; e^{-\nu n} \quad \textit{où} \quad b = c_1/n$$

Le théorème suivant montre que L_n est fortement concentré autour de $\zeta(3)$:

Théorème 3.4. $\forall t > 0, \exists \delta > 0, \forall n \geq 0, \mathbb{P}(|L_n - \xi(3)| \geq t) \leq e^{-\delta n}$

Démonstration. Soit $N = \begin{pmatrix} n \\ 2 \end{pmatrix}$ le nombre d'arêtes. Soit $0 < b \leq 1$ et $B =$ $[0,b]^N$. Soit $x \in B$, et $T = T(x)$ un arbre couvrant de longueur minimale, où plutôt l'ensemble de ses arêtes. Ainsi, $mst(x) = |T(x)| = \sum_{i \in T} x_i$.
Soit $\beta \in \mathbb{R}^N$ défini par $\beta_i = b$ si $i \in T$ et 0 sinon, et $\alpha = \beta/\|\beta\| = \beta/(b\sqrt{n-1})$.
Alors :

$$\begin{aligned} \forall y \in B, \quad mst(y) &\leq \sum_{i \in T} y_i \\ &\leq \sum_{i \in T} x_i + \sum_{i \in T} (y_i - x_i) \\ &\leq mst(x) + \sum_{i \in T} b \mathbb{1}_{x_i \neq y_i} \\ &\leq mst(x) + d_\beta(x,y) \end{aligned}$$

$$\text{d'où} \quad \forall y \in B, \qquad mst(y) \leq mst(x) + b\sqrt{n}\; d_\alpha(x,y)$$

Donc par le théorème 3.3, $\mathbb{P}(|mst(X) - m| \geq t) \leq 4e^{-\frac{t^2}{4b^2 n}}$ où m médiane de $mst(X)$. Par le Lemme 1 ($b = c_1/n$) :

$$\begin{aligned}\mathbb{P}(|mst(Y) - m| \geq 2t) &\leq \mathbb{P}(mst(Y) - mst(X) \geq t) + \mathbb{P}(|mst(X) - m| \geq t)\\ &\leq e^{-\nu n} + 4e^{-\frac{t^2 n}{4c_1^2}}\end{aligned}$$

Donc :

$$\forall t > 0,\ \exists\, \delta_1 = \delta_1(t) > 0,\ \forall n \geq 0,\ \mathbb{P}(|L_n - m| \geq t) \leq e^{-\delta_1 n} \tag{8}$$

Il reste à remplacer dans l'inégalité précédente m par $\zeta(3)$:

$$\begin{aligned}|\mathbb{E}(L_n) - m| &\leq \mathbb{E}|L_n - m|\\ &\leq \mathbb{E}\left[\frac{t}{4}\mathbb{1}_{|L_n - m| \leq \frac{t}{4}}\right] + \mathbb{E}\left[|l_n - m|\mathbb{1}_{|L_n - m| > \frac{t}{4}}\right]\\ &\leq t/4 + n\mathbb{P}(|L_n - m| > t/4) \ \text{ car } \ L_n \leq n\\ &\leq t/3 \quad \text{pour n assez grand}\end{aligned}$$

par (8).
De plus, on a admis que $\mathbb{E}(L_n) \xrightarrow{n\infty} \zeta(3)$ donc pour n assez grand, $|\mathbb{E}(L_n) - \zeta(3)| \leq t/3$ et donc $|m - \zeta(3)| \leq 2t/3$. D'où, pour n assez grand,

$$\mathbb{P}(|L_n - \zeta(3)| \geq t) \leq \mathbb{P}(|L_n - m| \geq t/3) \leq e^{-\delta n}$$

où $\delta = \delta_1(t/3)$ dépend de t.

□

3.2.4 Concentration sur le n-cube

On considère $X = (X_1, ..., X_n) \in \{-1, 1\}^n$ un vecteur aléatoire distribué uniformément sur le cube $\{-1, 1\}^n$, et un sous espace vectoriel V de $\mathbb{R}^n$ de dimension $1 \leq d \leq n$.

Propriété 2. $\mathbb{E}(dist(X, V)^2) = n - d$

Démonstration. Soit $P = (p_{ij})_{1 \leq i,j \leq n}$ la matrice de projection orthogonale sur $V^\perp$. On a : $dist(X, V)^2 = X.PX = \sum_{j=1}^n \sum_{j=1}^n X_i X_j p_{ij}$ d'où $\mathbb{E}(dist(X, V)^2) = \sum_1^n p_{ii} = tr(P) = n - d$.

□

Théorème 3.5. *(Inégalité de Talagrand pour la distance euclidienne) Soit A un convexe de $\mathbb{R}^n$. Alors $\mathbb{P}(X \in A)\mathbb{P}(X \notin A_t) \leq e^{-ct^2}$ où $A_t = \{x \in \mathbb{R}^n, dist(x, A) \leq t\}$.*

Démonstration.

Lemme 2. *Si* $A \subset [-1,1]^n,\ \forall x \in [-1,1]^n,\ d_T(x,A) \le t \quad \Rightarrow \quad dist(x,A) \le 2t$

Démonstration. (du lemme) Supposons que $d_T(x,A) \le t$.
On utilise la définition équivalente de d_T : Soit $U_A(x) = \{v \in \{0,1\}^n, \exists z \in A,\ v_i = 0 \Rightarrow z_i = x_i\}$ Soit $V = V_A(x)$ l'enveloppe convexe de U. Alors $d_T(x,A) = \min_{v \in V} \|v\|$.
Ainsi $\exists w \in V_A(x), \|w\| \le t$. On peut écrire $w = \sum_1^m \lambda_i w_i$ où $W_i \in U_A(x)$. Par définition de $U_A(x)$, il existe $y_i \in A - x$ tel que $y_{ik} = 0\ \forall k\ t.q\ w_{ik} = 0$.
On pose $y = \sum_1^m \lambda_i y_i \in A - x$ car A est convexe. Comme $\|y_i\| \le 1$ et $\forall k,\ w_{ik} = 0 \Rightarrow y_{ik} = 0$, on a $\|y_i\| \le 2w_i$ donc $\|y\| \le \sum_1^m \lambda_i 2w_i \le 2t$ donc $dist(x,A) \le 2t$.

□

Ainsi,

$$\begin{aligned}
\mathbb{P}(X \in A)\ \mathbb{P}(dist(X,A) \ge t) &\le \mathbb{P}(X \in A \cap [-1,1]^n)\ \mathbb{P}(dist(X, A \cap [-1,1]^n) \ge t) \\
&\le \mathbb{P}(X \in A \cap [-1,1]^n)\ \mathbb{P}(d_T(X, A \cap [-1,1]^n) \ge t/2) \\
&\le e^{-t^2/8}
\end{aligned}$$

□

Propriété 3. $\mathbb{P}(|dist(X,V) - \sqrt{n-d}| \ge t) \le Ce^{-ct^2}$

Remarque : Avec la méthode des différences bornées, on obtiendrait une majoration en $Ce^{-ct^2/n}$.

Démonstration. (propriété 3) On pose $A = \{x \in \mathbb{R}^n, dist(x,V) \le r\}$, on a alors $A_t = \{x \in \mathbb{R}^n, dist(x,V) \le r + t\}$ d'où par le théorème précédent

$$\mathbb{P}(dist(X,V) \le r)\mathbb{P}(dist(X,V) > r + t) \le e^{-ct^2} \tag{9}$$

On applique (9) avec $r = M$ puis $r = M - t$, où M médiane de $d(X,V)$:

$$\begin{aligned}
\frac{1}{2}\mathbb{P}(dist(X,V) > M + t) &\le e^{-ct^2} \\
\frac{1}{2}\mathbb{P}(dist(X,V) \le M - t) &\le e^{-ct^2}
\end{aligned}$$

d'où $\mathbb{P}(|dist(X,V) - M| > t) \le 4e^{-ct^2}$.

Montrons maintenant que $M = \sqrt{n-d}$:

$$\begin{aligned}
med(dist(X,V))^2 &= med(dist(X,V)^2) \qquad \text{(puisque } dist(X,V) \geq 0\text{)} \\
\text{De plus :} \quad d(X,V)^2 &= \sum_{i \neq j} X_i X_j p_{ij} + \sum X_i^2 p_{ii} \\
&= \sum_{i \neq j} X_i X_j p_{ij} + tr(P) \\
&= \sum_{i \neq j} X_i X_j p_{ij} + n - d
\end{aligned}$$

et $\mathbb{P}(\sum_{i \neq j} X_i X_j p_{ij} \geq 0) = \mathbb{P}(\sum_{i \neq j} X_i X_j p_{ij} \leq 0)$, donc $\mathbb{P}(d(X,V)^2 \geq n-d) = \mathbb{P}(d(X,V)^2 \leq n-d)$ puis $M^2 = n-d$.
Finalement on a bien $\mathbb{P}(|dist(X,V) - \sqrt{n-d}| \geq t) \leq Ce^{-ct^2}$

□

Remarque : Si on prend pour coordonnées de X des variables i.i.d Gaussiennes $X_i \equiv \mathcal{N}(0,1)$ plutôt que des signes aléatoires, la propriété 3 reste vraie, et on a pas besoin de l'inégalité de Talagrand pour la montrer.

Remarque : On a vu que dans le cas où A est convexe et borné, on est capable de majorer la distance euclidienne $d(x,A)$ par celle de Talagrand $d_T(x,A)$ (cf lemme 2). Ainsi, on peut donner la version suivante de l'inégalité de Talagrand :

Théorème 3.6. *Soit $K > 0$, $X = (X_1, ..., X_n)$ où les X_i sont indépendants, $|X_i| \leq K$, et F une fonction 1-Lipschitz.*
Alors : $\quad \mathbb{P}(|F(X) - M| \geq t) \leq Ce^{-ct^2/K^2}$.

Démonstration.

$$\begin{aligned}
\mathbb{P}(X \in A)\ \mathbb{P}(dist(X,A) \geq t) &\leq \mathbb{P}(X \in A)\ \mathbb{P}(d_T(X,A) \geq ct) \\
&\leq Ce^{-c't^2}
\end{aligned}$$

Puis on pose $A_a = \{x, F(x) \leq a\}$ qui est bien un ensemble convexe, et on prend $a = M$ puis $a = M - t$.

□

Donnons pour terminer une autre conséquence pratique de l'inégalité de Talagrand, pour des fonctions lipschitz (séparement pour chaque coordonnée) et certifiables :

Théorème 3.7. *Soient $X = (X_1, ..., X_n)$ où les X_k sont indépendants à valeurs dans A_k. Soit $f : \prod A_k \to \mathbb{R}$ vérifiant les deux conditions suivantes :*

(L) $\forall x, x' \in \prod_1^n A_k,\ |f(x) - f(x')| \leq c_k$ *si x et x' ne diffèrent seulement de la kième coordonnée.*

(C) Si $f(z) \geq r$*, alors il existe* $J \subset \{1, ..., n\}$ *certificat pour z, i.e :* $\sum_{k \in J} c_k^2 \leq \psi(r)$*, et* $\forall y \in \prod_1^n A_k,\ (y_k = z_k\ \forall k \in J) \Rightarrow (f(y) \geq r)$*.*

Alors : $\forall r \in \mathbb{R}, \forall t \geq 0, \mathbb{P}(f(X) \leq r - t)\ \mathbb{P}(X \geq r) \leq e^{-\frac{t^2}{4\psi(r)}}$

Démonstration. On pose $A = \{z, f(z) \leq r - t\}$ et $B = \{z, f(z) \geq r\}$. Il suffit de montrer que : $\forall w \in B,\ d_T(A, w) \geq \frac{t}{\sqrt{\psi(r)}}$.

Soit $w \in B$, et $J \subset \{1, ..., n\}$ un certificat pour w.

On pose $\alpha_i = \begin{cases} \frac{c_i}{\sqrt{\sum_J c_i^2}} & \text{si } i \in J \\ 0 & \text{sinon} \end{cases}$

Soit $w' \in A$, on pose $w''_i = \begin{cases} w_i & si\ i \in J \\ w'_i & sinon \end{cases}$

On a $f(w'') \geq r$ par définition de J, et $f(w') \leq r - t$, d'où $|f(w') - f(w'')| \geq t$

$\Rightarrow\ d_c(w', w'') := \sum_1^n c_i \mathbb{1}_{w'_i \neq w''_i} \geq t$ et les indices qui diffèrent sont dans J.

$\Rightarrow\ d_c(w', w) \geq t$ et les indices qui diffèrent sont dans J.

d'où :

$$d_c(w', w) = \sum_{i \in J} c_i\ \mathbb{1}_{w'_i \neq w''_i} = \sqrt{\sum_{i \in J} c_i^2}\ \sum_{i \in J} \alpha_i \mathbb{1}_{w'_i \neq w''_i} = \sqrt{\sum_{i \in J} c_i^2}\ d_\alpha(w, w') \geq t$$

Donc :

$$d_\alpha(w, w') \geq \frac{t}{\sqrt{\sum_{i \in J} c_i^2}} \geq \frac{t}{\sqrt{\psi(r)}}$$

□