Probabilités

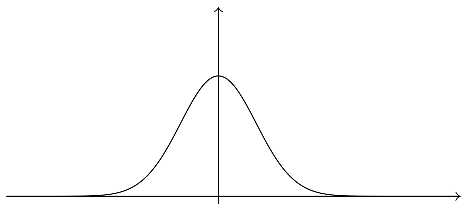
Chapitre 1 : Inégalités de concentration classiques et intervalles de confiance

Lucie Le Briquer

Sommaire

1 Introduction

Pour la loi normale $\mathcal{N}(0,1)$, en traçant le graphe de la densité on s'aperçoit que la masse se trouve concentrée autour de la moyenne, dès que l'on s'en éloigne la masse décroît très vite. En fait si $X \sim \mathcal{N}(0,1)$ alors $\mathbb{P}(|g| > t) \sim e^{\frac{-t^2}{2}}$.



En revanche si on regarde une Bernoulli $\xi \sim \mathcal{B}(1/2)$. $\mathbb{P}(\xi = 0) = \mathbb{P}(\xi = 1) = \frac{1}{2}$, il n'y a pas de concentration (idem pour la loi uniforme).
Mais lorsque qu'on tensorise la loi (i.e. on prend des copies indépendantes de la même V.A.) $\xi_1, ..., \xi_n$ indépendantes, $\frac{1}{n}\sum_{i=1}^{n} \xi_i \underset{n\to\infty}{\to} \mathbb{E}(\xi) = \frac{1}{2}$ d'après LFGN.
Quand n devient grand, on s'attend à revoir le phénomène de concentration i.e. la proportion de piles se concentre autour de la moyenne qui est $\frac{1}{2}$. Le TCL va aussi dans ce sens.

But : quantifier cette convergence et savoir à partir de combien de lancers on peut presque garantir d'avoir le même nombre de piles et de faces.

Théorème 1 (Bienaymé-Chebychev)

Soit X une v.a. de moyenne m et de variance σ^2. Alors :

$$\forall t > 0, \ \mathbb{P}(|X - m| \geq t) \leq \frac{\sigma^2}{t^2}$$

Remarque.
Reprenons notre exemple. $\xi_1, ..., \xi_n \sim \mathcal{B}(\frac{1}{2})$ et posons $X = \frac{1}{n}\sum_{i=1}^{n} \xi_i$.
$\mathbb{E}(X) = \frac{1}{2}$ et $\mathbb{V}(X) = \frac{1}{4n}$ alors d'après B-C $\mathbb{P}(|X - \frac{1}{2}| \geq t) \leq \frac{1}{4nt^2}$.
$n \geq \frac{1}{2\xi^2}$, on a *Proportion de piles* $\in [\frac{1}{2} - \xi, \frac{1}{2} + \xi]$ avec une probabilité $\geq \frac{1}{2}$.
On a juste utilisé que notre v.a. a un moment d'ordre 2.

Propriété 2

Si $X_1, ..., X_n$ v.a. indépendantes tq $\mathbb{E}(X_i) = m \ \forall i$, $M_4 = \sup\limits_{1\leq i\leq n} \mathbb{E}((X_i - m)^4)$. Alors :

$$\forall t > 0, \ \mathbb{P}\left(\left|\frac{1}{n}\sum_{i=0}^{n} X_i - m\right| \geq t\right) \leq \frac{3M_4}{n^2t^4}$$

Preuve.

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=0}^{n}X_i - m\right| \geq t\right) \leq \frac{1}{n^4t^4}\mathbb{E}\left(\left|\sum_{i=0}^{n}(X_i-m)\right|^4\right)$$

$$\mathbb{E}\left(\left|\sum_{i=0}^{n}(X_i-m)\right|^4\right) = \mathbb{E}\left(\sum_{i,j,k,l=1}^{n}(X_i-m)(X_j-m)(X_k-m)(X_l-m)\right)$$

Toutes les expressions contenant un terme d'ordre 1 s'annulent. Il reste dans la somme les expressions de la forme une puissance de 4 ou un produit de deux carrés.

Notons que :

$$\mathbb{E}((X_i-m)^2(X_j-m)^2) \underset{\text{C.S.}}{\leq} [\mathbb{E}((X_i-m)^4)]^{\frac{1}{2}}[\mathbb{E}((X_j-m)^4)]^{\frac{1}{2}} \leq M_4$$

et :

$$\mathbb{E}((X_i-m)^4) \leq M_4$$

Mais on a n termes puissances de 4 et $3n(n-1)$ termes de la forme un produit de carrés. Ainsi :

$$\mathbb{E}\left(\left|\sum_{i=1}^{n}X_i - m\right|^4\right) \leq M_4(n+3n(n-1)) = nM_4(3n-2) \leq 3M_4n^2$$

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n}X_i - m\right| \geq t\right) \leq \frac{1}{n^4t^4}3M_4n^2 = \frac{3M_4}{n^2}t^4$$

□

2 Concentration à travers les moments

Sur l'exemple précédent on a vu que si on était plus régulier on pouvait obtenir une meilleure inégalité de concentration. On va essayer de généraliser cette observation.

Définition 1 (fonction génératrice des moments)

Soit X une v.a., sa fonction génératrice des moments est donnée par $M_X(\lambda) := \mathbb{E}(e^{\lambda X})$

Remarque.

M_X permet de calculer les moments :

$$M'_X(0) = \mathbb{E}(X)$$

$$M''_X(0) = \mathbb{E}(X^2)$$

Exemple.
Fonction génératrice des moments de la loi normale $X \sim \mathcal{N}(0,1)$

$$M_X(\lambda) = \mathbb{E}(e^{\lambda X}) = \int_{-\infty}^{+\infty} e^{\lambda t} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}(t^2 - 2\lambda t + \lambda^2)} e^{-\frac{\lambda^2}{2}} dt = e^{\frac{\lambda^2}{2}}$$

Fonction génératrice de la *loi de Rademacher* : $X \sim \mathcal{R}(1/2)$

$$M_X(\lambda) = e^{\lambda}\mathbb{P}(X=1) + e^{-\lambda}\mathbb{P}(X=-1) = \frac{e^{\lambda} + e^{-\lambda}}{2} = ch(\lambda) \leq e^{\lambda^2/2}$$

Lemme 3 (Borne de Chernoff)

Soit S une v.a réelle. Alors :

$$\forall t \in \mathbb{R},\ \mathbb{P}(S \geq t) \leq \inf_{\lambda \geq 0} \{e^{-\lambda t} M_S(\lambda)\}$$

Preuve.
Soit $\lambda \geq 0$.

$$\mathbb{P}(S \geq t) = \mathbb{P}(\lambda S \geq \lambda t) = \mathbb{P}\left(e^{\lambda S} \geq e^{\lambda t}\right) \leq e^{-\lambda t}\mathbb{E}\left(e^{\lambda S}\right) = e^{-\lambda t} M_S(\lambda)$$

D'où le résultat en passant à l'infimum.

□

Exemple.
Si on regarde $X \sim \mathcal{N}(0,1)$. Soit $t \in \mathbb{R}^+$. Alors :

$$\mathbb{P}(X \geq t) \leq \inf_{\lambda \geq 0} \{e^{-\lambda t} e^{\lambda^2/2}\} = e^{-t^2/2}$$

Définition 2 (K-sous-Gaussienne)

On dit qu'une v.a centrée X est K-sous-Gaussienne si

$$M_X(\lambda) \leq e^{\lambda^2 K^2/2}$$

Exemple.

Ainsi une variable gaussienne est 1-sous-gaussienne.

Remarque.
On peut enlever *centrée* de la définition car si $M_X(\lambda) \leq \exp(\lambda^2 K^2/2)$ alors forcément X est centrée et $\mathbb{E}(X^2) \leq K^2$:

$$M_X(\lambda) = \sum_{n=0}^{+\infty} \frac{\lambda^n}{n!}\mathbb{E}(X^n) \leq \sum_{n=0}^{+\infty} \frac{\lambda^{2n}}{2^n n!}K^{2n}$$

d'où :

$$\lambda\mathbb{E}(X) + \frac{\lambda^2}{2}\mathbb{E}(X^2) + o(\lambda^2) \leq \frac{K^2\lambda^2}{2} + o(\lambda^2)$$

On fait tendre λ vers 0^+ et 0^- pour obtenir les résultats.

Théorème 4 (Bienaymé-Tchebychev)

Soit X une va centrée. Alors les assertions suivantes sont équivalentes :

1. $\exists K$ tq X est K-sous-gaussienne.
2. $\exists c > 0,\ \forall t \geq 0,\ \mathbb{P}(|X| \geq t) \leq 2e^{-ct^2}$
3. $\exists a > 0,\ E(e^{aX^2}) \leq 2$ (cette inégalité est connue comme la condition Ψ_2)

Preuve.

$(1) \implies (2)$:
On sait que

$$M_X(\lambda) \leq e^{(\lambda^2 K^2)/2} \quad \forall\lambda$$

Le lemme de Chernoff donne alors :

$$\forall t \geq 0,\ \mathbb{P}(X \geq t) \leq e^{-t^2/2K^2}$$

On fait le même travail pour $-X$:

$$\mathbb{P}(-X \geq t) \leq e^{-t^2/(2K^2)}$$

On en déduit :

$$\mathbb{P}(|X| \geq t) \leq \mathbb{P}(X \geq t) + \mathbb{P}(-X \geq t) \leq 2e^{-t^2/2K^2}$$

Et on obtient (2) avec $c := 1/2K^2$.

$(2) \implies (3)$:
On sait que

$$\forall t \geq 0,\ \mathbb{P}(|X| \geq t) \leq 2e^{-ct^2}$$

On aimerait trouver a tel que $\mathbb{E}(e^{aX^2}) \leq 2$. Or

$$\mathbb{E}(e^{aX^2}) = \int_0^\infty \mathbb{P}(e^{aX^2} \geq u)du = \int_0^1 \mathbb{P}(e^{aX^2} \geq u)du + \int_1^\infty \mathbb{P}(e^{aX^2} \geq u)du$$

Soit $a < c$:

$$\mathbb{E}(e^{aX^2}) \leq 1 + \int_0^\infty 2ate^{at^2}\mathbb{P}(|X| \geq t)dt \leq 1 + 2a\int_0^\infty 2te^{-(c-a)t^2}dt = 1 + \frac{2a}{c-a} = 2 \text{ (pour a=c/3)}$$

On obtient (3) avec $a = c/3$.

(3) $\Longrightarrow$ (1) :
On sait que $\mathbb{E}(e^{aX^2}) \leq 2 \quad (\star)$.

$$M_X(\lambda) = \mathbb{E}(e^{\lambda X}) = 1 + \int_0^1 (1-y)\mathbb{E}[(\lambda X)^2 e^{y\lambda X} dy] \leq 1 + \frac{\lambda^2}{2}\mathbb{E}(X^2 e^{|\lambda X|})$$

Or $|\lambda X| \leq \frac{\lambda^2}{2a} + \frac{aX^2}{2}$ (IAG). On obtient alors :

$$M_X(\lambda) \leq 1 + \frac{\lambda^2}{2} e^{\lambda^2/2a}\mathbb{E}\left(X^2 e^{aX^2/2}\right) \underset{CS}{\leq} 1 + \frac{\lambda^2}{2} e^{\lambda^2/2a}\left[\mathbb{E}\left(X^4\right)\right]^{1/2}\left[\mathbb{E}\left(e^{aX^2}\right)\right]^{1/2}$$

On déduit du développement de $(\star)$ que $\mathbb{E}\left(X^4\right) \leq 2/a^2$. Ainsi :

$$M_X(\lambda) \leq 1 + \frac{\lambda^2}{2} e^{\lambda^2/2a}\sqrt{\frac{2}{a^2}}\sqrt{2} = 1 + \frac{\lambda^2}{a}\ e^{\lambda^2/2a} \leq \left(1 + \frac{\lambda^2}{a}\right)\ e^{\lambda^2/2a} \leq e^{\lambda^2/a} e^{\lambda^2/2a} = e^{\frac{\lambda^2}{2}\frac{3}{a}}$$

Donc X est $\sqrt{\frac{3}{a}}$-Sous-Gaussienne.

□

Remarque.
Les dépendances par rapport à K sont (2) $c = \frac{1}{2K^2}$ et (3) $a = \frac{1}{6K^2}$

Corollaire 5

Soit ξ sous gaussienne de paramètre κ. Si on note $\eta = \xi^2$. Alors

$$\forall l \in \mathbb{N}, \quad \mathbb{E}(\eta^l) \leq 2l!(6\kappa^2)^l = \frac{1}{2} l!(6K^2)^{l-2}(12K^2)^2$$

Remarque.
Seconde forme utile pour appliquer Bernstein.

Preuve.
D'après le théorème 4 (2), on a que

$$\begin{aligned} & \mathbb{E}(e^{\frac{\xi^2}{6\kappa^2}}) \leq 2 \\ \text{Donc} \quad & \mathbb{E}(e^{\frac{\eta}{6\kappa^2}}) \leq 2 \\ \text{Alors} \quad & \mathbb{E}(\frac{\eta}{6\kappa^2})^l \leq 2l! \quad \forall l \quad \text{en développant l'exponentielle, (interversion } \textstyle\sum - \mathbb{E}) \end{aligned}$$

□

Propriété 6

Si $(\xi_1, \ldots, \xi_n)$ v.a sous-gaussiennes indépendantes, de paramètres κ_i chacune. Alors $\sum \xi_i$ sous gaussienne de paramètre $\sqrt{\sum(\kappa_i)^2}$.
En particulier,

$$\forall t \leq 0, \quad \mathbb{P}\left(\sum_{i=1}^{n} \xi_i \geq t\right) \leq \exp\left(-\frac{t^2}{2\sum_{i=1}^{n} \kappa_i^2}\right)$$

Preuve.
Soit $\lambda \in \mathbb{R}$, on a par définition puis par indépendance.

$$M_{\sum \xi_i}(\lambda) = \mathbb{E}(e^{\lambda \sum \xi_i}) = \prod_{i=1}^{n} \mathbb{E}(e^{\lambda \xi_i}) \leq \prod_{i=1}^{n} e^{\frac{\lambda^2 \kappa_i}{2}}$$

Ainsi

$$M_{\sum \xi_i}(\lambda) \leq e^{\frac{\lambda^2}{2} \sum_{i=1}^{n} \kappa_i^2}$$

Et donc en utilisant le théorème 4, on obtient que $\sum \xi_i$ est sous-gaussienne de paramètre $\sqrt{\sum_{i=1}^{n} \kappa_i}$.
De plus, si ξ sous-gaussienne de paramètre κ alors

$$\mathbb{P}(\xi \geq t) \leq \exp\left(\frac{-t}{2\kappa^2}\right)$$

Ce qui achève la démonstration.

□

3 Exemple concrets, Inégalité d'Hoeffding, Bennett, Bernstein

On va déjà donner un exemple de v.a sous gaussienne.

Lemme 7 (Hoeffding)

Si X v.a à valeur dans $[a, b]$. Alors $X - \mathbb{E}(X)$ est sous gaussienne de paramètre $\frac{b-a}{2}$.

Preuve.
On va supposer que X est centrée. On doit calculer $M_X(\lambda) = \mathbb{E}(e^{\lambda X})$.
Par convexité,

$$\forall x \in [a, b], \quad e^{\lambda x} \leq \frac{b-x}{b-a} e^{\lambda a} + \frac{x-a}{b-a} e^{\lambda b}$$

Comme $x = \frac{b-x}{b-a}a + \frac{x-a}{b-a}b$, on a alors

$$M_X(\lambda) \leq \frac{b}{b-a}e^{\lambda a} - \frac{a}{a-b}e^{\lambda b} \quad \text{car X centrée}$$

On pose alors $u = (b-a)\lambda$, $\Phi(u) = \ln(f(\lambda))$ et $p = \dfrac{-a}{b-a}$. Alors, $1 - p = \dfrac{b}{b-a}$ et

$$\Phi(u) = -pu + \ln(1 - p + pe^u)$$

On a $\Phi(0) = \Phi'(0) = 0$ et

$$\Phi''(u) = \frac{(1-p)pe^u}{(1-p+pe^u)^2} = \frac{\alpha\beta}{(\alpha+\beta)^2} \leq \frac{1}{4}$$

D'après la formule de Taylor-Lagrange :

$$\Phi(u) \leq \frac{1}{4}\frac{u^2}{2}$$

ce qui termine la preuve.

□

Théorème 8 (Hoeffding)

Soit $X_1, \dots, X_n$ v.a indépendantes telles que $X_i \in [a_i, b_i]$ presque sûrement. Alors

$$\forall t \geq 0, \quad \mathbb{P}\left(\sum_{i=1}^{n}(X_i - \mathbb{E}(X_i)) \geq t\right) \leq \exp\left(\frac{-2t^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right) \tag{1}$$

$$\mathbb{P}\left(\sum_{i=1}^{n}(X_i - \mathbb{E}(X_i)) \leq -t\right) \leq \exp\left(\frac{-2t^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right) \tag{2}$$

$$\mathbb{P}\left(\left|\sum_{i=1}^{n}(X_i - \mathbb{E}(X_i))\right| \geq t\right) \leq 2\exp\left(\frac{-2t^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right) \tag{3}$$

Preuve.
On a grâce au lemme 7 que, pour tout i, $X_i - \mathbb{E}(X_i)$ est sous-gaussienne de paramètre $\frac{(b_i-a_i)}{2}$. Et d'après la propriété 6 on a que $\sum_{i=1}^{n}(X_i - \mathbb{E}(X_i))$ sous-gaussienne de paramètre $\sqrt{\sum(\frac{b_i-a_i}{2})^2}$. Et donc en utilisant la théorème 4, les majorations en découlent.

□

Remarque.
Le théorème précédent ne tient pas compte des lois des variables aléatoires en question. Par exemple il traite de la même manière des Bernoulli standards et des Bernoulli de paramètres très petits.

- En particulier, si l'on prend ξ v.a telle que $\mathbb{P}(\xi = 0) = 0,99$ et $\mathbb{P}(\xi = 10^{10}) = 0,01$, on constate que le 10^{10} fait exploser la majoration dans Hoeffding la rendant alors inutile.

- Si $\xi_1, ..., \xi_n$ indépendantes $\sim \mathcal{B}(p)$, on sait que $S = \sum_{i=1}^n \xi_i \sim \mathcal{B}(n,p)$. Or si p fixé et n devient grand, on a S approché par $\mathcal{B}(np, np(1-p))$. On doit donc s'attendre à avoir

$$\mathbb{P}\left(\left|\sum_{i=1}^n (\xi_i - np)\right| \geq t\right) \leq 2\exp\left(\frac{-t^2}{2np(1-p)}\right)$$

  Alors que Hoeffding (théorème 8) nous donne uniquement

$$\mathbb{P}\left(\left|\sum_{i=1}^n (\xi_i - np)\right| \geq t\right) \leq 2\exp\left(\frac{-t^2}{n}\right)$$

  Qui est moins bien car $p(1-p) \leq 1/4$.

- En prenant cette fois-ci p qui décroit avec n, ie $np \to \lambda$ quand $n \to \infty$, alors S est bien approché par $\mathcal{P}(\lambda)$ et on a (cf TD 2 exo 2)

$$\mathbb{P}(\xi - \lambda \geq t) \leq \exp\left(-\lambda H(\frac{t}{\lambda})\right)$$

  Où $H(x) = (1+x)\ln(1+x) - x$ et $\xi \sim \mathcal{P}(\lambda)$.

Lemme 9

Soit $b \geq 0$ et ξ une v.a centrée avec $\mathbb{E}(\xi^2) \leq \sigma^2$ ($\sigma \geq 0$). Alors

1. Si $|\xi| \leq b$ p.s, alors
$$M_\xi(\lambda) \leq \exp\left(\frac{\sigma^2}{b^2}(e^{\lambda b} - \lambda b - 1)\right)$$

2. Si $\forall k \geq 3,\ \mathbb{E}(\xi^k) \leq \frac{1}{2}k!\sigma^2 b^{k-2}$ alors
$$M_\xi(\lambda) \leq \exp\left(\frac{\lambda^2\sigma^2}{2(1-\lambda b)}\right), \quad \forall \lambda < b$$

Remarque.

Si ξ vérifie Lemme 9(1), alors elle vérifie (2) car $|\xi| \leq b$, alors

$$\mathbb{E}(\xi^k) \leq \mathbb{E}(\xi^2\xi^{k-2}) \leq b^{k-2}\underbrace{\mathbb{E}(\xi^2)}_{\leq \sigma^2}$$

Ainsi (2) est beaucoup plus souple. On dit que c'est une *condition de croissance exponentielle des moments.*
Si ξ sous gaussienne de paramètre κ, alors ξ^2 vérifie (2) avec $\sigma = 2\sqrt{6}\kappa$ et $b = 6\kappa^2$.

Preuve.

1. On a

$$\begin{aligned}
M_\xi(\lambda) = \mathbb{E}(e^{\lambda\xi}) = \mathbb{E}\left(\sum_{k=0}^{\infty} \frac{\lambda^k}{k!}\xi^k\right) \\
&\leq 1 + \overbrace{0}^{\text{(car variable centrée)}} + \frac{\lambda^2\sigma^2}{2} + \sum_{k=3}^{\infty} \frac{\lambda^k}{k!}\mathbb{E}(\xi^k) \\
&\leq 1 + \frac{\lambda^2\sigma^2}{2} + \sum_{k=3}^{\infty} \frac{\lambda^k}{k!} b^{k-2}\sigma^2 \\
&\leq 1 + \frac{\lambda^2\sigma^2}{2} + \frac{\sigma^2}{b^2}\sum_{k=3}^{\infty} \frac{\lambda^k b^k}{k!} \\
&\leq 1 + \frac{\sigma^2}{b^2}(e^{\lambda b} - \lambda b - 1)
\end{aligned}$$

   Puis on utilise le fait que $e^x - (x+1) \geq 0$.

2. On a

$$\begin{aligned}
M_\xi(\lambda) = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!}\mathbb{E}(\xi^k) &\leq 1 + \frac{\lambda^2\sigma^2}{2} + \frac{1}{2}\sum_{k\geq 3} \lambda^k \sigma^2 b^{k-2} \\
&= 1 + \frac{\lambda^2\sigma^2}{2} + \frac{\lambda^2\sigma^2}{2}\sum_{k=3}^{\infty} (\lambda b)^{k-2} \\
&= 1 + \frac{\lambda^2\sigma^2}{2} + \frac{\lambda^2\sigma^2}{2}\frac{\lambda b}{1-\lambda b} \quad \text{si } |\lambda b| < 1 \\
&= 1 + \frac{\lambda^2\sigma^2}{2(1-\lambda b)}
\end{aligned}$$

□

Théorème 10 (inégalité de Bernstein)

Soient $\xi_1, \ldots, \xi_n$ v.a centrées indépendantes telles que pour tout i, $\mathbb{E}(\xi_i^2) \leq \sigma_i^2$, et

$$\exists b \geq 0, \forall k \geq 3, \quad \mathbb{E}(|\xi_i|^k) \leq \frac{1}{2}k!\sigma_i^2 b^{k-2}$$

Alors,

$$\forall t \geq 0, \quad \mathbb{P}\left(\left|\sum_{i=1}^{n} \xi_i\right| \geq t\right) \leq 2\exp\left(-\frac{t^2}{2(\sigma^2 + tb)}\right)$$

Où $\sigma^2 = \sum_{i=1}^n \sigma_i^2$.

Remarque.

- Si $t \leq \frac{\sigma^2}{b}$ alors $\mathbb{P}(|\sum_{i=1}^n \xi_i| \geq t) \leq 2\exp(-\frac{t^2}{4\sigma^2})$ et on voit le comportement sous-gaussien.
- Si $t > \frac{\sigma^2}{b}$ alors $\mathbb{P}(|\sum_{i=1}^n \xi_i| \geq t) \leq 2\exp(-\frac{t}{4b})$ et on voit le comportement exponentiel.

Preuve.
Soit $S = \sum_{i=1}^n \xi_i$ et soit $\lambda \geq 0$. Par indépendance, on a $M_S(\lambda) = \prod_{i=1}^n M_{\xi_i}(\lambda)$.
Mais on sait que

$$M_{\xi_i}(\lambda) \leq \exp\left(\frac{\lambda^2\sigma_i^2}{2(1-\lambda b)}\right), \quad \forall \lambda \leq \frac{1}{b}$$

On a alors

$$M_S(\lambda) \leq \exp\left(\frac{\lambda^2\sigma^2}{2(1-\lambda b)}\right)$$

La borne de Chernoff (lemme 3) nous donne que

$$\mathbb{P}(S \geq t) \leq \inf_{\lambda \geq 0}\left\{\exp\left(-\lambda t + \frac{\lambda^2\sigma^2}{2(1-\lambda b)}\right)\right\}$$

On prend $\lambda = \frac{t}{\sigma^2+tb}$, et donc

$$\mathbb{P}(S \geq t) \leq \exp\left(\frac{-t^2}{\sigma^2+tb} + \left(\frac{t}{\sigma^2+tb}\right)^2 \frac{\sigma^2}{2(1-\frac{tb}{\sigma^2+tb})}\right)$$

Puis tout se simplifie dans l'exponentielle. On obtient la majoration voulue en prenant la valeur absolue de S, nous donnant le facteur 2.

$\square$

Théorème 11 (inégalité de Bennett)

Soient $\xi_1, \ldots, \xi_n$ v.a centrées indépendantes telles que pour tout i, $\mathbb{E}(\xi_i^2) \leq \sigma_i^2$, et $|\xi_i| \leq b$.

$$\forall t \geq 0, \quad \mathbb{P}\left(\left|\sum_{i=1}^n \xi_i\right| \geq t\right) \leq 2\exp\left(-\frac{\sigma^2}{b}\ H\left(\frac{tb}{\sigma^2}\right)\right)$$

Où $H(x) = (1+x)\ln(1+x) - x$.

Preuve.
On sait que

$$\forall i \leq n, \quad M_{\xi_i}(\lambda) \leq \exp\left(\frac{\sigma_i^2}{b^2}(e^{\lambda b} - \lambda b - 1)\right)$$

Donc,

$$M_{\xi}(\lambda) \leq \exp\left(\frac{\sigma^2}{b^2}(e^{\lambda b} - \lambda b - 1)\right)$$

Chernoff (lemme 3) nous donne

$$\mathbb{P}(S \geq t) \leq \inf_{\lambda \geq 0} \left\{ \lambda t + \frac{\sigma^2}{b^2}(e^{\lambda b} - \lambda b - 1) \right\}$$

Et on prend $\lambda = \frac{1}{b \ln\left(1+\frac{tb}{\sigma^2}\right)}$

$\square$

Remarque.

- On a $H(x) \geq \frac{x^2}{2(1+\frac{x}{3})}$, ainsi Bennett est une meilleure majoration par Bernstein.
- On a aussi deux régimes : quand x petit alors $H(x) \sim x \Longrightarrow$ régime sous-gaussien. Et quand x grand alors $H(x) \sim x \ln(x) \Longrightarrow$ régime un peu mieux que l'exponentielle.
- On sait que $|\sum \xi_i| \leq nb$ p.s. Ainsi, l'inégalité de concentration a un sens quand $t \leq nb$.

4 Lemme d'applatissement de Johnson-Lindenstrauss

On a un ensemble T de n points dans un espace d'Hilbert, on peut naturellement les plonger dans $\mathbb{R}^n$

But : on aimerait compresser l'espace i.e. placer ces points dans un espace de dimension beaucoup plus petite tout en gardant presque la disposition des points (c'est-à-dire les distances qui les séparent).

On aimerait trouver :

$A : \mathbb{R}^n \longrightarrow \mathbb{R}^N$ avec $N << n$ tel que $\forall\ x, y \in T,\ \|A(x) - A(y)\|_2 \sim \|x - y\|_2$
où $\|.\|_2$ est la norme Euclidienne.

On va trouver un A linéaire.

On ne construira pas un A "déterministe" mais on prendra plutôt A aléatoire et on montrera qu'il va bien, avec une grande probabilité, effectuer la compression. A est une matrice de taille $N \times n$. On va remplir la matrice par des v.a. indépendantes sous-Gaussienne (par exemple des Bernouilli).

Théorème 12 (Johnson-Lindenstrauss)

Soit $n \in \mathbb{N}$ et $\varepsilon \in [0, 1]$ et T un ensemble à n points dans $\mathbb{R}^n$
Alors $\exists N = N(n, \varepsilon) \sim \frac{\log n}{\varepsilon^2}$ et $\exists\ A : R^n \longrightarrow \mathbb{R}^N$ tel que

$$\forall\ x, y \in T,\ (1 - \varepsilon)\|x - y\|_2 \leq \|A(x) - A(y)\|_2 \leq (1 + \varepsilon)\|x - y\|_2$$

Preuve.

Soit N que l'on trouvera à la fin de la preuve.

Soit $B = (\xi_{i,j})_{1 \leq i,j \leq n}$ une matrice $N \times n$ à entrées indépendantes sous-Gaussiennes de paramètre K (à spécifier après). Soit $L = \mathbb{E}(\xi^2) \leq K^2$

Soit $u \in S^{n-1}$, où S^{n-1} est la sphère unité de $\mathbb{R}^n$.

$$\|Bu\|_2^2 = \sum_{i=1}^{N} < ligne_i(B), u >^2 = \sum_{i=1}^{N} \eta_i$$

où $\eta_i =< ligne_i(B), u >^2$

On a que les η_i sont indépendantes

$$\mathbb{E}(\eta_i) = \mathbb{E}(< ligne_i(B), u >^2) = \mathbb{E}\left(\left(\sum_{j=1}^{n} \xi_{i,j} u_j \right)^2 \right) = \mathbb{E}\left(\sum_{j,k=1}^{n} \xi_{i,j}\xi_{i,k} u_j u_k \right) = \mathbb{E}\left(\sum_{j=1}^{n} \xi_{i,j}^2 u_j^2 \right) = L$$

$< ligne_i(B), n >= \sum_{j=1}^{n} \xi_{i,j} u_j$ est sous-Gaussienne de paramètre K (car c'est la somme de v.a. sous-Gaussiennes indépendantes)

Donc η_i est le carré d'une sous-Gaussienne alors par le corollaire (juste après les propriétés de ss-Gauss)

$$\forall L, \ \mathbb{E}(\eta_i^l) \leq \frac{1}{2} l!(6K^2)^{l-2}(12K^2)^2$$

Bernstein avec $\sigma = 12K^2$ et $b = 6K^2$:

$$\forall t > 0, \ \mathbb{P}\left(\left| \sum_{i=1}^{N} (\eta_i - \mathbb{E}(\eta_i)) \right| > t \right) \leq 2\exp\left(-\frac{t^2}{2(144K^4 + 6K^2 t)} \right)$$

mais $\sum_{i=1}^{N} \eta_i = \|Bu\|_2^2$, $\sum_{i=1}^{N} \mathbb{E}(\eta_i) = NL$

On vient de montrer que $\forall u \in S^{n-1}$ et $\forall \varepsilon > 0$ on a :

$$\mathbb{P}(| \, \|Bu\|_2^2 - NL| > NL\varepsilon) \leq 2\exp\left(-\frac{N^2L^2\varepsilon^2}{2(144K^4N + 6K^2NL\varepsilon)} \right) \leq 2\exp\left(-\frac{N^2L^2\varepsilon^2}{4 \times 144K^4N} \right)$$

On aurait du prendre $K = 1$ dès le début. Ainsi on a montré que $\forall u \in S^{n-1}, \ \forall \varepsilon > 0$:

$$\mathbb{P}\left(\frac{1}{NL} \|Bu\|_2^2 \notin [1-\varepsilon, 1+\varepsilon] \right) \leq 2\exp\left(-\frac{N\varepsilon^2}{4 \times 144} \right)$$

$$\forall \ x, y \in T \ \ \mathcal{E}_{x,y} = \left\{ \omega \mid \frac{1}{NL} \frac{\|B(x-y)\|_2^2}{\|x-y\|_2} \notin (1-\varepsilon, 1+\varepsilon) \right\}$$

$$\mathbb{P}(\mathcal{E}_{x,y}) \leq 2\exp\left(-\frac{N\varepsilon^2}{4 \times 144} \right)$$

On veut que $\mathbb{P}\left(\bigcap\limits_{x,y\in T}\mathcal{E}_{x,y}^C\right)$ soit grande.

$$\mathbb{P}\left(\bigcup_{x,y\in T}\mathcal{E}_{x,y}\right)\leq\sum_{x,y\in T}\mathbb{P}(\mathcal{E}_{x,y})\leq 2\underbrace{|T|^2}_{n^2}\exp\left(-\frac{N\varepsilon^2}{4\times 144}\right)=\exp\left(\ln(2n^2)-\frac{N\varepsilon^2}{4\times 144}\right)$$

On choisit $N=\frac{\ln(2n^2)}{\varepsilon^2}8\times 144 \;\Leftarrow\; \mathbb{P}\left(\bigcup\limits_{x,y\in T}\mathcal{E}_{x,y}\right)$ est très petite.

Il suffit de prendre une réalisation $A=\frac{1}{\sqrt{N}}B(w)$ pour $w\in\bigcap\limits_{x,y\in T}\mathcal{E}_{x,y}^C$

$\square$

5 Intervalles de confiance

5.1 Introduction

On se trouve avec loi inconnue que l'on observe sur des échantillons. Le but est de pouvoir estimer certains paramètres de cette loi à travers l'étude d'échantillons. Par exemple, estimer l'espérance et la variance.

On considère un caractère X : "avoir un accident"

$X_i=1$ si l'individu i a un accident, 0 sinon.

Sur un échantillon de n personnes, on a $\frac{1}{n}\sum_{i=1}^n X_i$ proportion d'accidents. Il est naturel d'estimer l'espérance par la moyenne empirique $\frac{1}{n}\sum_{i=1}^n X_i$. On cherche un intervalle de confiance dont on peut s'assurer (avec une petite erreur) que l'espérance est dedans).

5.2 Estimateurs

Définition 3 (n-échantillon aléatoire)

Un n-échantillon aléatoire issu d'une v.a. réelle X est un ensemble $(X_1,...,X_n)$ de n copies indépendantes de X.

On a un paramètre θ associé à X et on aimerait l'estimer.

Définition 4 (estimateur)

Un estimateur $\widehat{\theta}_n$ de θ est une fonction qui dépend du n-échantillon $(X_1, ..., X_n)$.

- On dit que l'estimateur $\widehat{\theta}_n$ est *convergent* ou *consistant* s'il est proche de θ au sens de la convergence en probabilité i.e. :

$$\forall \varepsilon > 0, \mathbb{P}(|\widehat{\theta}_n - \theta| > \varepsilon) \underset{n\to\infty}{\longrightarrow} 0$$

- *fortement consistant* si $\widehat{\theta_n}$ converge presque sûrement vers θ i.e.

$$\mathbb{P}(\lim_n |\widehat{\theta_n} - \theta| = 0) = 1$$

Définition 5 (biais)

On appelle le biais la quantité $\mathbb{E}(\widehat{\theta}_n - \theta)$

On dit que l'estimateur est *sans biais* si le biais est nul, sinon qu'il est *biaisé*.

Exemple.

La moyenne empirique est un estimateur fortement consistant sans biais de l'espérance.

On peut aussi regarder l'écart quadratique $\mathbb{E}((\widehat{\theta}_n - \theta)^2) = Var(\widehat{\theta}_n) + (biais)^2$

Pour réduire l'écart quadratique on a intérêt à trouver un estimateur sans biais et à faible variance.

5.3 Intervalles de confiance

Définition 6 (intervalle de confiance $1 - \alpha$)

Soit $\alpha \in [0, 1]$ Un intervalle de confiance $1 - \alpha$ est un couple d'estimateur $(\underline{\theta_n}, \overline{\theta_n})$ tel que :

$$\mathbb{P}\left(\theta \in [\underline{\theta_n} \ ; \ \overline{\theta_n}]\right) \geq 1 - \alpha$$

Définition 7 (intervalle de confiance asymptotique $1 - \alpha$)

Soit $\alpha \in [0, 1]$ Un intervalle de confiance asymptotique $1 - \alpha$ est un couple d'estimateur $(\underline{\theta_n}, \overline{\theta_n})$ tel que :

$$\underset{n\to\infty}{lim} \mathbb{P}\left(\theta \in [\underline{\theta_n} \ ; \ \overline{\theta_n}]\right) = 1 - \alpha$$

Exemple.

Prenons l'exemple de la mesure Gaussienne (avec variance connue). $(X_1, ..., X_n)$ un n-échantillon de noi $\mathcal{N}(\mu, \sigma^2)$

$\overline{X_n} = \frac{1}{n}\sum_{i=1}^n X_i$ est un estimateur sans biais de μ

$$\sqrt{n}\left(\frac{\overline{X_n}-\mu}{\sigma}\right) \sim \mathcal{N}(0,1)$$

$$\Rightarrow \mathbb{P}\left(-z_{1-\frac{\alpha}{2}} \leq \sqrt{n}\left(\frac{\overline{X_n}-\mu}{\sigma}\right) \leq z_{1-\frac{\alpha}{2}}\right) = 1-\alpha$$

où z_p est le quantile tel que $F(z_p) = p = \mathbb{P}(\underbrace{G}_{\sim\mathcal{N}(0,1)} \leq z_p)$

Ainsi, on a:

$$\mathbb{P}(\overline{X_n} - \frac{\sigma}{\sqrt{n}}z_{1-\frac{\alpha}{2}} \leq \mu \leq \overline{X_n} + \frac{\sigma}{\sqrt{n}}z_{1-\frac{\alpha}{2}}) = 1-\alpha$$

On pose $(\underline{\theta_n}, \overline{\theta_n}) = (\overline{X_n} - \frac{\sigma}{\sqrt{n}}z_{1-\frac{\alpha}{2}}, \overline{X_n} + \frac{\sigma}{\sqrt{n}}z_{1-\frac{\alpha}{2}})$

$(\underline{\theta_n}, \overline{\theta_n})$ est ainsi un intervalle de confiance $1-\alpha$.

Si la *variance est inconnue* :

Définition 8 (variance empirique)

La *variance empirique* d'un n-échantilon est

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X}_n)^2$$

Définition 9 (loi du chi-deux)

Soit $Y_1, ..., Y_n \sim N(0,1)$.
La *loi du chi-deux à n degrès de liberté* est la loi de la v.a. $\sum_{i=1}^n Y_i^2$. On la note $\chi^2(n)$.

Faits.

S_n^2 est un estimateur convergent de σ^2 sans biais. $\frac{n-1}{\sigma^2}S_n^2$ suit la loi du $\chi^2(n-1)$.

On ne peut plus estimer μ en considérant $\sqrt{n}\frac{\overline{X_n-\mu}}{\sigma}$. Au lieu ça, on considère $\sqrt{n}\frac{\overline{X_n-\mu}}{S_n}$. Mais celle ci ne suit pas une loi normale mais une loi dite de student à $n-1$ degrés de libertés que l'on note τ_{n-1} (sa densité est une fonction paire et on a aussi une table comme pour la loi normale). Ainsi, comme précédement :

$$\mathbb{P}\left(-t_{1-\frac{\alpha}{2}} \leqslant \sqrt{n}\left(\frac{\overline{X_n}-\mu}{\sigma}\right) \leqslant t_{1-\frac{\alpha}{2}}\right) = 1-\alpha$$

On pose :

$$\underline{\theta_n} = \overline{X_n} - \frac{S_n}{\sqrt{n}}t_{1-\frac{\sigma}{2}}, \quad \overline{\theta_n} = \overline{X_n} + \frac{S_n}{\sqrt{n}}t_{1-\frac{\sigma}{2}}$$

$(\underline{\theta_n}, \overline{\theta_n})$ est un intervalle de confiance $1-\alpha$.

Cours du 24 février

Regardons si X_i sont des Bernouilli indépendantes de paramètre θ

TCL $\Rightarrow$ $\frac{\sqrt{n}}{\sqrt{\theta(1-\theta)}}(\underbrace{\widehat{\theta_n}}_{\text{moyenne empirique}} - \theta) \longrightarrow \mathcal{N}(0,1)$

Ceci se traduit par la convergence des fonctions de répartition.

$$\lim_{n\to+\infty} \mathbb{P}\left(\frac{\sqrt{n}}{\sqrt{\theta(1-\theta)}}(\widehat{\theta_n} - \theta) \leq z_\alpha\right) = 1 - \alpha$$

Cela nous donne un intervalle de confiance asymptotique $1-\alpha$ qui est :

$$\left[\widehat{\theta_n} - z_{\alpha/2}\sqrt{\frac{\widehat{\theta_n}(1-\widehat{\theta_n})}{n}}\ ;\ \widehat{\theta_n} + z_{\alpha/2}\sqrt{\frac{\widehat{\theta_n}(1-\widehat{\theta_n})}{n}}\right]$$

Si on veut un intervalle de confiance non-asymptotique on utilisera les inégalités de concentration.
Si X_i Bernouilli de paramètre θ, $S_n = \sum_{i=1}^n X_i$ alors d'après Hoeffding :

$$\Rightarrow\ \mathbb{P}(|S_n - n\theta| \geq t) \leq 2\exp\left(\frac{-2t^2}{n}\right)$$

On veut que cette probabilité soit $\leq \alpha$. Prenons $t = \sqrt{\frac{n}{2}\ln(\frac{2}{\alpha})}$, on aurait :

$$\mathbb{P}\left(|S_n - n\theta| \geq \sqrt{\frac{n}{2}\ln\left(\frac{2}{\alpha}\right)}\right) \leq \alpha$$

Ainsi on a un intervalle de confiance $1-\alpha$ qui est :

$$\left[\widehat{\theta_n} - \sqrt{\frac{1}{2n}\ln\frac{2}{\alpha}}\ ;\ \widehat{\theta_n} + \sqrt{\frac{1}{2n}\ln\frac{2}{\alpha}}\right]$$