BASIC CONCENTRATION INEQUALITIES

THIERRY KLEIN

Table des matières

1. Introduction

The starting point of our discussion is the STRONG LAW OF LARGE NUMBER

Theoreme 1.1. *Assume $(X_n)_{n\geq 1}$ is a sequence of i.i.d random variables such that $\mathbb{E}\left(|X_n|\right) < +\infty$ then*

$$\frac{X_1 + \ldots + X_n}{n} \xrightarrow[n\to\infty]{p.s.} \mathbb{E}(X_1).$$

For the statistician $\mathbb{E}(X_1)$ represents an unknown quantity to be estimated and $\frac{X_1+\ldots+X_n}{n}$ is an natural estimator. In the real life n never goes to infinity, we only have a finite number of observations ($n = 100$, $n = 1000$). It is then natural to wonder for a fixed n if $\frac{X_1+\ldots+X_n}{n}$ is close or far from $\mathbb{E}(X_1)$. The speed of convergence is also unnatural question we can be interested in.
The first answer concerning the rate of convergence is given by the central limit theorem

Theoreme 1.2. *Let $(X_n)_{n\geq 1}$ be a sequence of i.i.d random variables such that the variance σ^2 exists (i.e.$\mathbb{E}\left(X_n^2\right) < +\infty$) then*

$$\sqrt{n}\Big(\frac{X_1 + \ldots + X_n}{n} - \mathbb{E}(X_1)\Big) \xrightarrow[n\to\infty]{Loi} \mathcal{N}(0, \sigma^2).$$

Roughly speaking this theorem tells us that $\frac{X_1+\ldots+X_n}{n}$ goes at rate $\sqrt{n}$ to $\mathbb{E}(X_1)$. Nevertheless, this is an asymptotic result and gives us nothing when n is fixed (in particular if n is small).
The ami of this course is to présent elementary results allowing to quantify the error $\frac{X_1+\ldots+X_n}{n} - \mathbb{E}(X_1)$ for a fixed n. We will présents the first results in the area of the theory of mesure concentration and of

concentration inequalities. For more detailed and more advanced mathematic the reader should take a look of the books of Michel Ledoux [?] and of Boucheron, Lugosi et Massart [?].

2. Gaussian world is too perfect

2.1. Some notations.

We consider iid gaussian random variables (we can assume without lost of generality that they are centered with variance 1). The density fundtion is

$$\phi(x) = \frac{\exp\left(-x^2/2\right)}{\sqrt{2\pi}}.$$

The cumulative distribution function is

$$\Phi(x) = \int_{-\infty}^{x} \phi(t)dt.$$

2.2. A Perfect concentration inequality ?

Take your favorite n (n is now a fixer number), we aim to know if $\frac{X_1+\ldots+X_n}{n}$ is far from $0 = \mathbb{E}(X_1$. Mathematicaly speaking this can be put in the following way. Take $x > 0$ (the distance between $\frac{X_1+\ldots+X_n}{n}$ and 0) and let us try to estimate the probability that de distance between $\frac{X_1+\ldots+X_n}{n}$ and 0 is greater than x

$$C(x) = \mathbb{P}\left(\left|\frac{X_1+\ldots+X_n}{n} - \mathbb{E}(X_1)\right| \geq x\right)$$

Here $\mathbb{E}(X_1) = 0$ et $\frac{X_1+\ldots+X_n}{n} \sim \mathcal{N}(0, 1/n)$, so

$$C(x) = \mathbb{P}\left(|X_1| \geq x\sqrt{n}\right) = 2\mathbb{P}\left(X_1 \geq x\sqrt{n}\right) = \frac{2}{\sqrt{2\pi}}\int_x^{+\infty} e^{-t^2/2}dt.$$

It is now sufficient to control sharply $\int_x^{+\infty} e^{-t^2/2}dt$.

Lemma 2.1. $\forall x > 0$

$$max\left(0, e^{-x^2/2}\left(\frac{1}{x} - \frac{1}{x^3}\right)\right) \leq \int_x^{+\infty} e^{-t^2/2}dt \leq e^{-x^2/2}\frac{1}{x} \tag{1}$$

Proof

Majoration

$$\int_x^{+\infty} e^{-t^2/2}dt = \int_x^{+\infty} \frac{t}{t}e^{-t^2/2}dt \leq \frac{1}{x}\int_x^{+\infty} te^{-t^2/2}dt = \frac{1}{x}e^{-x^2/2}.$$

Minoration Integrating by parts we get

$$\begin{aligned}\int_x^{+\infty} e^{-t^2/2}dt &= \int_x^{+\infty} \frac{t}{t}e^{-t^2/2}dt = \Big[-\frac{e^{-t^2/2}}{t}\Big]_x^{+\infty} - \int_x^{+\infty} \frac{1}{t^2}e^{-t^2/2}dt \\ &= \frac{e^{-x^2/2}}{x} - \int_x^{+\infty} \frac{t}{t^3}e^{-t^2/2}dt \geq \frac{e^{-x^2/2}}{x} - \frac{e^{-x^2/2}}{x^3}.\end{aligned}$$

■

Remark 2.2. *The bounds in the previous Lemma can be as sharp as one wants by integrating many times by parts. For example the upper bound can be replaced by*

$$e^{-x^2/2}\Big(\frac{1}{x} - \frac{1}{x^3} + \frac{3}{x^5}\Big).$$

Theoreme 2.1 (Gaussian Concentration). *For any $n \geq 1$ and any $x > 0$*

$$2max\left(0, e^{-nx^2/2}\left(\frac{1}{\sqrt{n}x} - \frac{1}{n\sqrt{n}x^3}\right)\right) \leq \mathbb{P}\left(\left|\frac{X_1+\ldots+X_n}{n} - \mathbb{E}(X_1)\right| \geq x\right) \leq 2e^{-nx^2/2}\frac{1}{\sqrt{n}x} \tag{2}$$

Remark 2.3. *Since the majoration and the minoration are of the same order of magnitude, we see that we have an perfect control of $C(x)$.*

Unfortunately Gaussian world is too perfect and it would not be possible to get equivalent results in general.

2.3. **From Markov inequality to an other concentration inequality.** Markov inequality is a powerful tool allowing to derive concentration inequalities.

Proposition 2.4. *Let Y be an non negative integrable random variable then for any $t > 0$*

$$\mathbb{P}(Y \geq t) \leq \frac{1}{t}\mathbb{E}(Y).$$

If g is increasing non negative then

$$\mathbb{P}(Y \geq t) = \mathbb{P}(g(Y) \geq g(t)) \leq \frac{\mathbb{E}(g(Y))}{g(t)}.$$

Choosing $g(x) = x^2$ we get the well known Bienaymé-Tchebichev inequality

$$\mathbb{P}(|Y - \mathbb{E}(Y)| \geq t) \leq \frac{Var(Y)}{t^2}.$$

Let us go back to the perfect gassian world.

$$C(x) = \mathbb{P}\left(\left|\frac{X_1 + \ldots + X_n}{n} - \mathbb{E}(X_1)\right| \geq x\right) = 2\mathbb{P}\left(X_1 \geq x\sqrt{n}\right)$$

For $t > 0$ consider the application $g_t(y) = e^{ty}$ (which is obviously increasing non negative) so from Markov inequality we get

$$C(x) \leq 2e^{-tx\sqrt{n}}\mathbb{E}\left(e^{tX_1}\right). \tag{3}$$

Lemma 2.5. *If $X \sim \mathcal{N}(0,1)$ then*

$$\mathbb{E}\left(e^{tX_1}\right) = e^{t^2/2}$$

Proof

$$\mathbb{E}\left(e^{tX_1}\right) = \int_{\mathbb{R}} e^{tx}e^{-x^2/2}\frac{dx}{\sqrt{2\pi}} = e^{t^2/2}\int_{\mathbb{R}} e^{-(x-t)^2/2}\frac{dx}{\sqrt{2\pi}}.$$

■

We will now optimized in $t > 0$ the right hand side of (3)

$$C(x) \leq 2e^{-tx\sqrt{n}}\mathbb{E}\left(e^{tX_1}\right) = 2e^{-tx\sqrt{n}}e^{t^2/2}$$

The optimum is reached for $t = x\sqrt{n}$, one gets

Theoreme 2.2 (Concentration Gaussienne 2)**.** *For all $n \geq 1$ and all $x > 0$*

$$\mathbb{P}\left(\left|\frac{X_1 + \ldots + X_n}{n} - \mathbb{E}(X_1)\right| \geq x\right) \leq 2e^{-nx^2/2} \tag{4}$$

Remark 2.6. *Let us compare inequality (2) and (4). The first one provide a bound of order $2e^{-nx^2/2}\frac{1}{\sqrt{nx}}$ and the second one of $2e^{-nx^2/2}$. Using Markov inequality we loose a factor of order $\frac{1}{\sqrt{nx}}$ but we obtained the dominated factor $e^{-nx^2/2}$. The advantage of the second method is that it can be easily extended outside the gaussian world.*

2.4. **A third Gaussian inequality ?** Replacing $g_t(y) = e^{ty}$ by $m_k(y) = |y|^k$ ici $k \in \mathbb{N}^*$, Markov inequality leads to

$$C(x) \leq \frac{1}{\sqrt{n}^k x^k}\mathbb{E}\left(|X_1|^k\right). \tag{5}$$

Lemma 2.7 (Computing $M_k = \mathbb{E}\left(|X_1|^k\right)$)**.**

(1) *if $k = 2p$ then*

$$M_{2p} = \frac{(2p)!}{2^p p!}$$

(2) *If* $k = 2p+1$ *then*

$$M_{2p+1} = 2^p p! \frac{2}{\sqrt{2\pi}}.$$

Preuve Set $M_k = \mathbb{E}\left(|X_1|^k\right)$ then (integrating by parts)

$$M_k = 2\int_0^\infty x^k e^{-x^2/2} dx/\sqrt{2/pi} = 2(k-1)\int_0^\infty x^{k-2} e^{-x^2/2} dx/\sqrt{2/pi} = (k-1)M_{k-2}.$$

(1) If $k = 2p$ then by induction(recall that $M_2 = 1$) one can see that

$$M_{2p} = \frac{(2p)!}{2^p p!}$$

(2) If $k = 2p+1$ again by induction (recall that $M_1 = 2/\sqrt{2\pi}$) one can see that

$$M_{2p+1} = 2^p p! \frac{2}{\sqrt{2\pi}}.$$

■

Now we have to optimized (5) in k. Doesn't seem toto be an est task!!!!!

3. Main Dish Cramér's Method

3.1. **The method.** This method is baded on Markov inequality. Take any random variable Y and $t > 0$ then by Markov inequality we have

$$\mathbb{P}(Y > x) \leq e^{-tx}\mathbb{E}\left(e^{tY}\right). \tag{6}$$

Set

$$\Psi_Y(t) = \log\left(\mathbb{E}\left(e^{tY}\right)\right).$$

Then (6) becomes

$$\mathbb{P}(Y > x) \leq e^{-tx+\Psi_Y(t)}.$$

Set $\Psi_Y^*(x)$ lthe Legendre transform of Ψ_Y

$$\Psi_Y^*(x) = \sup_{t \geq 0}\{tx - \Psi_Y(t)\}.$$

We then get

$$\mathbb{P}(Y > x) \leq e^{-\Psi_Y^*(x)}.$$

We will apply this method to empirical means of iid random variables. Take $X_1, X_2, \ldots X_n$ i.i.d such that there exists $t > 0$ for which $\mathbb{E}\left(e^{tX_1}\right) < +\infty$, we apply the previous method to $Y = \frac{X_1+\ldots+X_n}{n} - \mathbb{E}(X_1)$. Since the variable are iid it is obvious that

$$\mathbb{P}\left(\frac{X_1+\ldots+X_n}{n} - \mathbb{E}(X_1) \geq x\right) \leq \exp\left(-t(x + \mathbb{E}(X_1) + n\Psi_{X_1}(\frac{t}{n})\right)$$

At this point we have two possibilities

(1) The function $\Psi_Y^*(x)$ is well known (see for example 1) and in that case we obtain nice inequalities.

(2) The function $\Psi_Y^*(x)$ is unknown (no closed formula or even worse the law of Y is unknown) and in that case we will try to find clever upper bounds for $e^{-\Psi_Y^*(x)}$.

Example 1.

(1) *Gaussian Random variables. Let us consider now the case where the law of the random variables is* $\mathcal{N}(\mu, \sigma^2)$
Let $(X_i)_i$ *be i.i.d.* $\mathcal{N}(\mu, \sigma^2)$ *random variables then*

$$\mathbb{E}\left(e^{tX_1}\right) = e^{\frac{t^2\sigma^2}{2} + t\mu}$$

we then get

$$\mathbb{P}\left(\frac{X_1+\ldots+X_n}{n}-\mathbb{E}(X_1)\geq x\right)\leq \exp\left(-tx+\frac{t^2\sigma^2}{2n}\right)$$

The right hand side is minimized for pour $t=\frac{nx}{\sigma^2}$ *Which lead to the following bound*

$$\mathbb{P}\left(\frac{X_1+\ldots+X_n}{n}-\mathbb{E}(X_1)\geq x\right)\leq \exp\left(-\frac{nx^2}{2\sigma^2}\right)$$

We can do the same thing for left deviations and get

$$\mathbb{P}\left(\left|\frac{X_1+\ldots+X_n}{n}-\mathbb{E}(X_1)\right|\geq x\right)\leq 2\exp\left(-\frac{nx^2}{2\sigma^2}\right)$$

One should note that the smaller σ^2 *is the better the bound is.*

(2) *Case of Poisson random variable n* ($\mathcal{P}(\lambda)$).
Let $X_1,\ldots,X_n$ *be iid* $\mathcal{P}(\lambda)$ *distributed random variables.*
We begin by finding the Laplace transform of $X\sim\mathcal{P}(\lambda)$. *Let* $t\in\mathbb{R}$ *we have*

$$\psi_X(t)=\log\left(\mathbb{E}\left(e^{tX}\right)\right)=\log\left(e^{-\lambda}\sum_{k=0}^{+\infty}e^{tk}\frac{\lambda^k}{k!}\right)=\log\left(\exp\left(-\lambda+\lambda e^t\right)\right)=-\lambda+\lambda e^t.$$

We will usus again Cramér's method. For $x>0$ *and* $t>0$ *we have*

$$\mathbb{P}\left(\frac{X_1+\ldots+X_n}{n}-\lambda\geq x\right)\leq \exp\left(-t(x+\lambda)+n\Psi_X(\frac{t}{n})\right)=\exp\left(-t(x+\lambda)+n\left(-\lambda+\lambda e^{t/n}\right)\right)$$

The right hand side term is minimal fort $=n\log\left(\frac{x+\lambda}{\lambda}\right)$, *which leads to*

$$\begin{aligned}\mathbb{P}\left(\frac{X_1+\ldots+X_n}{n}-\lambda\geq x\right)&\leq \exp\left(-n\log\left(\frac{x+\lambda}{\lambda}\right)(x+\lambda)+n\left(-\lambda+\lambda\frac{x+\lambda}{\lambda}\right)\right)\\&\leq \exp\left(-n\log\left(\frac{x+\lambda}{\lambda}\right)(x+\lambda)+nx\right)=\exp\left(-n\lambda\left[(1+\frac{x}{\lambda})\log\left(1+\frac{x}{\lambda}\right)-\frac{x}{\lambda}\right]\right)\end{aligned}$$

Definition 3.1. *Let us define for* $x\geq -1$ *the function*

$$h(x)=(1+x)\log(1+x)-x.$$

We have

$$\mathbb{P}\left(\frac{X_1+\ldots+X_n}{n}-\lambda\geq x\right)\leq \exp\left(-n\lambda h\left(\frac{x}{\lambda}\right)\right). \tag{7}$$

Let us take a look a left déviations. Since Poisson variables are non negative positive $\frac{X_1+\ldots+X_n}{n}-\lambda$ *is always larger or equal than* $-\lambda$, *we shall then take* $0<x<\lambda$ *and* $t>0$

$$\begin{aligned}\mathbb{P}\left(\frac{X_1+\ldots+X_n}{n}-\lambda\leq -x\right)&=\mathbb{P}\left(-t\left(\frac{X_1+\ldots+X_n}{n}\right)\geq -t\left(\lambda-x\right)\right)=\mathbb{P}\left(e^{-t\left(\frac{X_1+\ldots+X_n}{n}\right)}\geq e^{-t(\lambda-x)}\right)\\&\leq \exp\left(t(\lambda-x)+n\Psi_X(-\frac{t}{n})\right)=\exp\left(t(\lambda-x)+n\left(-\lambda+\lambda e^{-t/n}\right)\right)\end{aligned}$$

The right hand side term is minimal for $t=-n\log\left(\frac{\lambda-x}{\lambda}\right)$, *which leads to*

$$\begin{aligned}\mathbb{P}\left(\frac{X_1+\ldots+X_n}{n}-\lambda\leq -x\right)&\leq \exp\left(-n\log\left(\frac{\lambda-x}{\lambda}\right)(\lambda-x)+n\left(-\lambda+\lambda\frac{\lambda-x}{\lambda}\right)\right)\\&\leq \exp\left(-n\log\left(\frac{\lambda-x}{\lambda}\right)(\lambda-x)-nx\right).\end{aligned}$$

So we have for $0<x<\lambda$

$$\mathbb{P}\left(\frac{X_1+\ldots+X_n}{n}-\lambda\leq -x\right)\leq \exp\left(-n\lambda h\left(-\frac{x}{\lambda}\right)\right) \tag{8}$$

Remark 3.1. *On can show that if $x \geq 0$ then*

$$h(x) \geq \frac{x^2}{2+2x/3}.$$

Hence the bound in Equation (7) is bounded by

$$\mathbb{P}\left(\frac{X_1+\ldots+X_n}{n}-\lambda \geq x\right) \leq \exp\left(-n\lambda h\left(\frac{x}{\lambda}\right)\right) \leq \exp\left(-n\frac{x^2}{2\lambda+2x/3}\right) \leq \exp\left(-\frac{nx^2}{2\lambda}\frac{1}{1+2x/(3\lambda)}\right).$$

One shall compared this bound to the gaussian bound $\exp\left(-\frac{nx^2}{2\lambda}\right)$.

(3) *Case of Bernoulli random variables ($\mathcal{B}(p)$).*
Let $X_1, \ldots, X_n$ be iid $\mathcal{B}(p)$ distributed random variables.
Proceeding as previously we compute for $X \sim \mathcal{B}(p)$ and $t \in \mathbb{R}$ the Laplace transfer of X

$$\psi_X(t) = \log\left(\mathbb{E}\left(e^{tX}\right)\right) = \log\left(pe^t+(1-p)\right).$$

Following t Cramér's method for $0 < x < 1-p$and $t > 0$ (the restriction on x is valid since Bernoulli random variables are smaller than 1) we have

$$\mathbb{P}\left(\frac{X_1+\ldots+X_n}{n}-p \geq x\right) \leq \exp\left(-t(x+p)+n\Psi_X(\frac{t}{n})\right) = \exp\left[-t(x+p)+n\log\left(pe^t+(1-p)\right)\right]$$

Set $u = pe^t$ hence the right hand side term is optimized for $u = \frac{(1-p)(x+p)}{n-(x+p)}$ (i.e. $t = \log(u/p)$).

$$\begin{aligned}
\mathbb{P}\left(\frac{X_1+\ldots+X_n}{n}-p \geq x\right) &\leq \exp\left[-\log(u/p)(x+p)+n\log\left(\frac{(1-p)(x+p)}{n-(x+p)}+(1-p)\right)\right] \\
&\leq \exp\left[-(x+p)\log\left(\frac{(1-p)(x+p)}{p(n-(x+p))}\right)+n\log\left(\frac{n(1-p)}{n-(x+p)}\right)\right] \\
&\leq \exp\left[-(n-(x+p))\log\left(\frac{(n-(x+p))}{(1-p)}\right)-(x+p)\log\left(\frac{x+p}{p}\right)+n\log(n)\right]
\end{aligned}$$

For a general Random variable these computations are not feasible We shall then proceed by comparison.

4. Classical inequalities

4.1. Sub-gaussian random variables.

Definition 4.1. *A centered random variable X is said to be sub-gaussian with variance term v if*

$$\psi_X(t) := \log\left(\mathbb{E}\left(e^{tX}\right)\right) \leq \frac{vt^2}{2}$$

We denote those variable by $X \in \mathcal{G}(v)$.

Theoreme 4.1 (Sub Gaussian concetration)**.** *Let $v > 0$, if $X_1, \ldots, X_n$ are i.i.d random variable belonging to $\mathcal{G}(v)$. Then for all $x > 0$,*

$$\mathbb{P}\left(\frac{X_1+\ldots+X_n}{n} \geq x\right) \leq \exp\left(-\frac{nx^2}{2v}\right).$$

Proof
We use (again) Cramé's method, then we bound $\psi_X$by its counterpart gaussian. ■

But what happent for non sub-gaussian random variables? For example assume you have in hand exponential random variables. Take n iid random variable with exponential law with parameter $\lambda > 0$. Then if $t < \lambda$ and if $X \sim \mathcal{E}(\lambda)$

$$\psi_X(t) = \log\left(\mathbb{E}\left(e^{tX}\right)\right) = \log\left(\int_0^{+\infty} e^{ty}\lambda e^{-\lambda y}dy\right) = \log\left(\frac{\lambda}{\lambda-t}\right).$$

Hence if $t < n\lambda$ we get

$$\mathbb{P}\left(\frac{X_1+\ldots+X_n}{n}-\frac{1}{\lambda} \geq x\right) \leq \exp\left(-t(x+\frac{1}{\lambda})+n\log\left(\frac{\lambda}{\lambda-t/n}\right)\right).$$

The right hand side is optimized for $t = \frac{nx\lambda^2}{1+x\lambda}$ (note that $t \leq n\lambda$). Which leads to

$$\begin{aligned}\mathbb{P}\left(\frac{X_1+\ldots+X_n}{n} - \frac{1}{\lambda} \geq x\right) &\leq \exp\left(-\frac{nx\lambda^2}{1+x\lambda}(x+\frac{1}{\lambda}) + n\log\left(\frac{\lambda}{\lambda - \frac{x\lambda^2}{1+x\lambda}}\right)\right), \\ &\leq \exp\left(-nx\lambda + n\log\left(1+x\lambda\right)\right), \\ &\leq \exp\left(-n\left[x\lambda - \log\left(1+x\lambda\right)\right]\right).\end{aligned}$$

Obviously this bound is greater than the gaussian bound!!!

4.2. **Hoeffding's inequality.**

4.2.1. *Bounded random variables are sub-gaussian.*

Lemma 4.1. *Let Z be a centered random variable ($\mathbb{E}(Z) = 0$) taking its values in $[a, b]$ ($a \leq Z \leq b$ a.s.). Then $Z \in \mathcal{G}\left(\frac{(b-a)^2}{4}\right)$*

Proof
Step 1 : We shall note that (the distance from Z to the middle point of $[a, b]$ is smaller than $b - a$)

$$\left|Z - \frac{b+a}{2}\right| \leq \frac{b-a}{2}$$

Hence

$$\mathrm{Var}(Z) \leq \frac{(b-a)^2}{4}$$

Step 2 : Let us compute

$$\begin{aligned}\psi_Z(t) &= \log\left(\mathbb{E}\left(e^{tZ}\right)\right) \\ \psi'_Z(t) &= \frac{\mathbb{E}\left(Ze^{tZ}\right)}{\mathbb{E}\left(e^{tZ}\right)} \\ \psi''_Z(t) &= \frac{\mathbb{E}\left(Z^2e^{tZ}\right)}{\mathbb{E}\left(e^{tZ}\right)} - \frac{\left(\mathbb{E}\left(Ze^{tZ}\right)\right)^2}{\left(\mathbb{E}\left(e^{tZ}\right)\right)^2} = \mathbb{E}\left(Z^2e^{tZ}e^{-\psi_Z(t)}\right) - \left(\mathbb{E}\left(Ze^{tZ}e^{-\psi_Z(t)}\right)\right)^2\end{aligned}$$

The last term looks like a variance. Noticing that $\mathbb{E}\left(e^{tZ}e^{-\psi_Z(t)}\right) = 1$ the fonction $u \mapsto e^{tu}e^{-\psi_Z(t)}$ is a probability density with respect to the law of Z. Hence if U is a random variable admitting this density we have

$$\mathrm{Var}(U) = \mathbb{E}\left(Z^2e^{tZ}e^{-\psi_Z(t)}\right) - \left(\mathbb{E}\left(Ze^{tZ}e^{-\psi_Z(t)}\right)\right)^2$$

Since $a \leq Z \leq b$we also have $a \leq U \leq b$ and step 1 is valid for U we then have

$$\psi''_Z(t) \leq \frac{(b-a)^2}{4}.$$

It is easy to see that $\psi_Z(0) = 0$ ($\log(1) = 0$) and that $\psi'_Z(0) = 0$ ($\mathbb{E}(Z) = 0$), so using taylor expansion of order two we have

$$\psi_Z(t) = \frac{t^2}{2}\psi''_Z(\theta) \leq \frac{t^2(b-a)^2}{8}.$$

∎

4.2.2. *Hoeffding's inequality.* Applying the two previous results we get

Theoreme 4.2. *Let $X_1, \ldots, X_n$ be n indépendant random variables. Assume that for all index i, $X_i \in [a_i, b_i]$. Set*

$$S = \sum_{i=1}^{n}\left(X_i - \mathbb{E}(X_i)\right).$$

Then if $t > 0$, we have

$$\mathbb{P}\left(S \geq t\right)\exp\left(-\frac{2t^2}{\sum_{i=1}^{n}\left(b_i - a_i\right)^2}\right)$$

4.3. **Bennett's inequality.**

Theoreme 4.3 (Bennett's inequality)**.** *Let $X_1, \dots, X_n$ be n independent random variables with finite variance. Assume that for all index i, $X_i \leq b$. Set*

$$S = \sum_{i=1}^{n} (X_i - \mathbb{E}(X_i))$$

and

$$v = \sum_{i=1}^{n} \mathbb{E}(X_i^2).$$

For $u \in \mathbb{R}$, set $\phi(u) = e^u - u - 1$ and for $u \geq -1$, $h(u) = (1+u)\log(1+u) - u$ Then

(1) *For $t > 0$*

$$\psi_S(t) := \log\left(\mathbb{E}\left(e^{tS}\right)\right) \leq n \log\left(1 + \frac{v}{nb^2\phi(bt)}\right) \leq \frac{v}{b^2}\phi(bt)$$

(2) *For $x > 0$,*

$$\mathbb{P}(S \geq x) \leq \exp\left(-\frac{v}{b^2} h\left(\frac{bx}{v}\right)\right)$$

Proof

(1)
Step 1 : One can assume (without loss of generality) that $b = 1$.
Step 2 : Note first that $u \mapsto \frac{\phi(u)}{u^2}$ is increasing. Hence since $X_i \leq 1$ it is obvious that

$$\phi(tX_i) \leq t^2 X_i^2 \phi(t) = X_i^2\left(e^t - t - 1\right)$$
$$e^{tX_i} \leq tX_i + 1 + X_i^2\left(e^t - t - 1\right)$$

Step 3 : We computs $\psi_S(t)$ and use step 2

$$\begin{aligned}\psi_S(t) &= \sum_{i=1}^{n} \log\left(\mathbb{E}\left[e^{t(X_i - \mathbb{E}(X_i))}\right]\right) = \sum_{i=1}^{n}\left(\log\left(\mathbb{E}\left[e^{tX_i}\right]\right) - t\mathbb{E}(X_i)\right) \\ &\leq \sum_{i=1}^{n}\left(\log\left(1 + t\mathbb{E}(X_i) + \mathbb{E}(X_i^2)\left(e^t - t - 1\right)\right) - t\mathbb{E}(X_i)\right)\end{aligned}$$

using the concavity of $u \mapsto \log(1+u)$ we have

$$\begin{aligned}\psi_S(t) &\leq n\left(\log\left(1 + t\frac{\sum_{i=1}^{n}\mathbb{E}(X_i)}{n} + \frac{v}{n}\left(e^t - t - 1\right)\right) - t\frac{\sum_{i=1}^{n}\mathbb{E}(X_i)}{n}\right) \\ &\leq v\left(e^t - t - 1\right).\end{aligned}$$

Which proves the first point.

(2) Cramér's method again!!

$$\mathbb{P}(S \geq x) \leq e^{-tx + \psi_S(t)} \leq e^{-tx + v\left(e^t - t - 1\right)}$$

The right hand side is optimized for $t = \log\left(1 + \frac{x}{v}\right)$, we then get

$$\begin{aligned}\mathbb{P}(S \geq x) &\leq e^{-\log\left(1+\frac{x}{v}\right)x + v\left(\frac{x}{v} - \log\left(1+\frac{x}{v}\right)\right)} \\ &\leq e^{-v\left[\left(1+\frac{x}{v}\right)\log\left(1+\frac{x}{v}\right) - \frac{x}{v}\right]}.\end{aligned}$$

■

Remark 4.2. *One can see that*

$$h(u) \geq \frac{u^2}{2(1+u/3)}.$$

which provides

$$\mathbb{P}(S \geq x) \leq e^{-\frac{x^2}{2(v+bx/3)}}.$$

4.4. **Bernstein's inequality.** Or how one can take into account information on the variances.

Theoreme 4.4 (Bernstein's inequality). *Let $X_1, \dots, X_n$ be n independent random variables such that there exists iv > 0 and > 0 satisfying $\sum_{i=1}^n \mathbb{E}(X_i^2) \le v$ and*

$$\sum_{i=1}^n \mathbb{E}\left[(X_i)_+^q\right] \le \frac{q!}{2} v c^{q-2}, \qquad \forall q \ge 3,$$

where $x_+ = \max(x, 0)$. Set

$$S = \sum_{i=1}^n (X_i - \mathbb{E}(X_i))$$

Set for $u > 0$, $h_1(u) = 1 + u - \sqrt{1+2u}$ Then

(1) *Pour $0 < t < 1/c$*

$$\psi_S(t) := \log\left(\mathbb{E}\left(e^{tS}\right)\right) \le \frac{vt^2}{2(1-ct)}$$

(2) *For $x > 0$,*

$$\mathbb{P}\left(S \ge \sqrt{2vx} + cx\right) \le \exp(-x)$$

Proof

(1) Like in the previous proof we consider the function $\phi(u) = e^u - u - 1$. This function is bounded by $u^2/2$ as soon as $u \le 0$

$$\phi(u) \le \frac{u^2}{2} \qquad \forall u \le 0.$$

Let $t > 0$ then

$$\phi(tX_i) = \sum_{q=2}^{+\infty} \frac{t^q X_i^q}{q!} \le \frac{t^2 X_i^2}{2} + \sum_{q=3}^{+\infty} \frac{t^q (X_i)_+^q}{q!}$$

$$\mathbb{E}(\phi(tX_i)) \le \frac{t^2 \mathbb{E}\left(X_i^2\right)}{2} + \sum_{q=3}^{+\infty} \frac{t^q \mathbb{E}\left[(X_i)_+^q\right]}{q!}$$

$$\sum_{i=1}^n \mathbb{E}(\phi(tX_i)) \le \frac{t^2 v}{2} + \sum_{q=3}^{+\infty} \frac{t^q v c^{q-2}}{2}$$

$$\sum_{i=1}^n \mathbb{E}(\phi(tX_i)) \le \frac{v}{2} \sum_{q=2}^{+\infty} t^q c^{q-2}.$$

The série is convergent if and only if $tc < 1$ that is $t \le 1/c$. We control now $\psi_S(t)$ (we shall use again$\log(u) \le u - 1$ for $u > 0$).

$$\begin{aligned}
\psi_S(t) = \log\left[\mathbb{E}\left(e^{tS}\right)\right] &= \sum_{i=1}^n \left(\log\left[\mathbb{E}\left(e^{t\sum_{i=1}^n X_i}\right)\right]\right) - t\sum_{i=1}^n \mathbb{E}(X_i) \\
&\le \sum_{i=1}^n \left(\mathbb{E}\left(e^{t\sum_{i=1}^n X_i}\right) - 1 - t\mathbb{E}(X_i)\right) = \sum_{i=1}^n \mathbb{E}(\phi(tX_i)) \\
&\le \frac{v}{2} \sum_{q=2}^{+\infty} t^q c^{q-2} \\
&\le \frac{vt^2}{2} \frac{1}{1-tc}.
\end{aligned}$$

Which proves the first point.

(2) Let y>0 and $0 < t < 1/c$

$$\mathbb{P}(S \ge y) \le e^{-ty + \psi_S(y)} \le e^{-ty + \frac{vt^2}{2}\frac{1}{1-tc}}. \tag{9}$$

Set $h(t) = ty - \frac{vt^2}{2}\frac{1}{1-tc}$ we aim to optimized this expression in $t \in [0, 1/c]$. Set $u = \frac{cy}{v}$ (ie $y = \frac{uv}{c}$, then

$$h(t) = \frac{uvt}{c} - \frac{vt^2}{2(1-ct)}.$$

Dividing by v the critical points are given by

$$\frac{u}{c} - \frac{t}{1-ct} - \frac{ct^2}{2(1-ct)^2} = 0.$$

Multiplying by $2(1-ct)^2$ we have

$$t^2 - \frac{2t}{c} + \frac{2u}{c^2(1+2u)} = 0.$$

the square root of the discriminant is

$$\sqrt{\Delta} = \frac{2}{c\sqrt{1+2u}}.$$

The only critical point less than$1/c$ is

$$t^* = \frac{1}{c}\left(1 - \frac{1}{\sqrt{1+2u}}\right).$$

which leads to

$$\begin{aligned} h(t^*) &= v\left[\frac{u}{c^2}\left(1 - \frac{1}{\sqrt{1+2u}}\right) - \frac{\sqrt{1+2u}\left(1-\frac{1}{\sqrt{1+2u}}\right)^2}{2c^2}\right] \\ &= \frac{v}{c^2}h_1(u) \end{aligned}$$

h_1 is defined for $u > 0$ by

$$h_1(u) = 1 + u - \sqrt{1+2u}.$$

Rewritting Equation (9) we get

$$\mathbb{P}(S \geq y) \leq \exp\left(-\frac{v}{c^2}h_1(\frac{cy}{v})\right)$$

The inequality

$$\mathbb{P}\left(S \geq \sqrt{2vx} + cx\right) \leq \exp(-x)$$

is obtained using the inverse of l $h_1(u)$. This inverse is $h_1^{-1}(t) = t + \sqrt{t}$.

■

Remark 4.3. (1) *An equivalent for of this inequality is given by*

$$\mathbb{P}(S \geq y) \leq \exp\left(-\frac{v}{c^2}h_1(\frac{cy}{v})\right)$$

(2) *One can see that for $u > 0$*

$$h_1(u) \leq \frac{u^2}{2(1+u)}.$$

Then under the same assumptions we have

$$\mathbb{P}(S \geq y) \leq \exp\left(-\frac{t^2}{2(v+ct)}\right).$$

■

5. The case of Sobol indices

The results of this Section comme from the work of Gamboa, Janon Klein, Lagnoux et Prieur [?]

5.1. **Black box models.** We consider regression models

$$Y = f(X) := f(X_1, \dots, X_p). \tag{10}$$

Here $Y \in \mathbb{R}$ and $X = (X_1, \dots X_p)$ with for $i = 1, \dots p$, the X_i's are any indépendant random objects. $\mathcal{X}_i$ will dénote the space in which X_i leaves. f is assumed to be a deterministic function and Y such that $0 < \mathrm{Var}Y < +\|infty)$.
One goal of sensitivity analysis is to determined influent variables. Sobol indices allow to quantifie this influence.
Let us recall their definitions. Take $\mathbf{u}$ a subset of $I_p := \{1, \dots, p\}$: . The closed Sobol index is (see [?])

$$S^{\mathbf{u}}_{\mathrm{Cl}} := \frac{\mathrm{Var}(\mathbb{E}(Y|X_i, i \in \mathbf{u}))}{\mathrm{Var}(Y)}.$$

5.2. **Pick and freeze method for estimation.** In real life E f is unknown and so are the cobol indices. We shall then provide some estimation procedure for them. For X and a subset $\mathbf{v}$ of I_p we set $X^{\mathbf{v}}$ lthe vector such that $X^{\mathbf{v}}_i = X_i$ if $i \in v$ and $X^{\mathbf{v}}_i = X'_i$ if $i \notin v$ where X'_i is an independent copy of X_i. We then set

$$Y^{\mathbf{v}} := f(X^{\mathbf{v}}).$$

The next lemma's gives a nice covariance représentation of the numerator of Sobol index (see [? , Lemma 1.2])

Lemma 5.1. *For any* $\mathbf{u} \subset \mathbf{I_p}$, *we have*

$$\mathrm{Var}(\mathbb{E}(Y|X_i, i \in \mathbf{u})) = \mathrm{Cov}\left(\mathbf{Y}, \mathbf{Y^u}\right). \tag{11}$$

Thanks to this Lemma, we can derive a natural estimator

$$S^{\mathbf{u}}_{N,\mathrm{Cl}} = \frac{\frac{1}{N}\sum Y_i Y^{\mathbf{u}}_i - \left(\frac{1}{N}\sum Y_i\right)\left(\frac{1}{N}\sum Y^{\mathbf{u}}_i\right)}{\frac{1}{N}\sum Y_i^2 - \left(\frac{1}{N}\sum Y_i\right)^2}. \tag{12}$$

This estimator is consistent and goes almost surely to the true cobol index. As usual, in applications, n is finite and one would want to quantify the distance between the estimator and the true value.
The main differences with respect to the previous sections are the following

(1) The estimator is not a sum of independent random variables.
(2) The estimator has a bias.

Notation
V will denote $\mathrm{Var}(Y)$ and as previously h is defined for $x > -1$ by

$$h(x) = (1 + x)\ln(1 + x) - x.$$

5.3. **Concentration inequalities for $S^{\mathbf{u}}_{N,\mathrm{Cl}}$.** Let us introduce the following random variables

$$U^{\pm}_i = Y_i Y^{\mathbf{u}}_i - (S^{\mathbf{u}}_{\mathrm{Cl}} \pm y)(Y_i)^2 \text{ et } J^{\pm}_i = (S^{\mathbf{u}}_{\mathrm{Cl}} \pm y)Y_i - Y^{\mathbf{u}}_i$$

Set V^+_U (resp. V^-_U, V^+_J and V^-_J) the moment of order 2 of the variables U^+_i (resp. U^-_i, J^+_i and J^-_i).

Theoreme 5.1. *Soit* $b > 0$ *et* $y > 0$. *We assume that* Y_i *and* $Y^{\mathbf{u}}_i$ *belongs to* $[-b, b]$. *Then*

$$\mathbb{P}\left(S^{\mathbf{u}}_{N,\mathrm{Cl}} \geq S^{\mathbf{u}}_{\mathrm{Cl}} + y\right) \leq M_1 + 2M_2 + 2M_3, \tag{13}$$

$$\mathbb{P}\left(S^{\mathbf{u}}_{N,\mathrm{Cl}} \leq S^{\mathbf{u}}_{\mathrm{Cl}} - y\right) \leq M_4 + 2M_2 + 2M_5, \tag{14}$$

where

$$M_1 = \exp\left\{-\frac{NV^+_U}{b_U^2} h\left(\frac{b_U}{V^+_U}\frac{yV}{2}\right)\right\} \qquad M_3 = \exp\left\{-\frac{NV^+_J b^2}{b_U^2} h\left(\frac{b_U}{bV^+_J}\sqrt{\frac{yV}{2}}\right)\right\}$$

$$M_2 = \exp\left\{-\frac{NV}{b^2} h\left(\frac{b}{V}\sqrt{\frac{yV}{2}}\right)\right\} \qquad M_4 = \exp\left\{-\frac{NV^-_U}{b_U^2} h\left(\frac{b_U}{V^-_U}\frac{yV}{2}\right)\right\}$$

$$M_5 = \exp\left\{-\frac{NV^-_J b^2}{b_U^2} h\left(\frac{b_U}{bV^-_J}\sqrt{\frac{yV}{2}}\right)\right\}$$

and $b_U = b^2(1 + S^{\mathbf{u}}_{\mathrm{Cl}} + y)$.

Proof
Since $S^{\mathbf{u}}_{\mathrm{Cl}}$ and $S^{\mathbf{u}}_{N,\mathrm{Cl}}$ are invariant when one translate the variables Y and $Y^{\mathbf{u}}$ we can assume that $\mathbb{E}(Y)=0$.

(1) U_i^+ et $U_i^-$are bounded by b_U, J_i^+ and J_i^- by b_U/b, moreover

$$\begin{array}{ll}\mathbb{E}(U_i^+) = -yV & \mathbb{E}(J_i^+)=0\\ \mathbb{E}(U_i^-) = yV & \mathbb{E}(J_i^-)=0\end{array}$$

and

$$V_U^{\pm} = \mathrm{Var}(YY^{\mathbf{u}}) + (S^{\mathbf{u}}_{\mathrm{Cl}}+y)^2\mathrm{Var}(Y^2) - 2(S^{\mathbf{u}}_{\mathrm{Cl}}\pm y)\mathrm{Cov}(YY^{\mathbf{u}},Y^2) + y^2V^2$$

$$V_J^{\pm} = ((S^{\mathbf{u}}_{\mathrm{Cl}}\pm y)^2+1)V - 2(S^{\mathbf{u}}_{\mathrm{Cl}}\pm y)C_u.$$

(2) Proof of (13). As

$$\{a+b\geq c\}\subset\{a\geq c/2\}\cup\{b\geq c/2\} \quad \text{et} \quad \{ab\geq c\}\subset\{|a|\geq\sqrt{c}\}\cup\{|b|\geq\sqrt{c}\}$$

we have

$$\begin{aligned}
\mathbb{P}\left(S^{\mathbf{u}}_{N,\mathrm{Cl}}\geq S^{\mathbf{u}}_{\mathrm{Cl}}+y\right) &= \mathbb{P}\left(\frac{\frac{1}{N}\sum_{i=1}^N Y_iY_i^{\mathbf{u}} - \overline{Y}_N\overline{Y}^{\mathbf{u}}_N}{\frac{1}{N}\sum_{i=1}^N (Y_i)^2 - \left(\overline{Y}_N\right)^2} \geq S^{\mathbf{u}}_{\mathrm{Cl}}+y\right)\\
&= \mathbb{P}\left(\frac{1}{N}\sum_{i=1}^N\left(U_i^+ - \mathbb{E}(U^+)\right) + \overline{Y}_N\overline{J}_N^+ \geq yV\right)\\
&\leq \mathbb{P}\left(\sum_{i=1}^N\left(U_i^+ - \mathbb{E}(U^+)\right)] \geq N\frac{yV}{2}\right) + \mathbb{P}\left(\sum_{i=1}^N Y_i \geq N\sqrt{\frac{yV}{2}}\right)\\
&\quad + \mathbb{P}\left(\sum_{i=1}^N(-Y_i)\geq N\sqrt{\frac{yV}{2}}\right) + \mathbb{P}\left(\sum_{i=1}^N J_i^+ \geq N\sqrt{\frac{yV}{2}}\right)\\
&\quad + \mathbb{P}\left(\sum_{i=1}^N(-J_i^+)\geq N\sqrt{\frac{yV}{2}}\right).
\end{aligned}$$

Inequality (13) comes from the application of Bennett's inequality (apply Bennett's result five time).

(3) Proof (14). Similarly we have

$$\begin{aligned}
\mathbb{P}\left(S^{\mathbf{u}}_{N,\mathrm{Cl}}\leq S^{\mathbf{u}}_{\mathrm{Cl}}-y\right) &= \mathbb{P}\left(\frac{1}{N}\sum_{i=1}^N\left(-U_i^- + \mathbb{E}(U^-)\right) + (-\overline{Y}_N)\overline{J}_N^- \geq yV\right)\\
&\leq \mathbb{P}\left(\sum_{i=1}^N\left(-U_i^- + \mathbb{E}(U^-)\right) \geq N\frac{yV}{2}\right) + \mathbb{P}\left(\sum_{i=1}^N Y_i \geq N\sqrt{\frac{yV}{2}}\right)\\
&\quad + \mathbb{P}\left(\sum_{i=1}^N(-Y_i)\geq N\sqrt{\frac{yV}{2}}\right) + \mathbb{P}\left(\sum_{i=1}^N J_i^- \geq N\sqrt{\frac{yV}{2}}\right)\\
&\quad + \mathbb{P}\left(\sum_{i=1}^N(-J_i^-)\geq N\sqrt{\frac{yV}{2}}\right).
\end{aligned}$$

Inequality (14) comes from the application of Bennett's inequality (apply Bennett's result five time).

■

E-mail address: thierry.klein@math.univ-toulouse.fr