Oral presentation - 23 November 2023

Jérémy Bettinger & Simon Viel

Concentration inequalities for martingales

Under the direction of Bernard Delyon & Adrien Saumard
ENSAI, Rennes.

Summary

Hoeffding's inequality

Freedman's inequality

An application to matrices

Hoeffding's Inequality

Theorem (Hoeffding)

Let $(X_k)_{k \in [\![1;n]\!]}$ *be independent random variables such that :*

$$\forall k \in [\![1;n]\!], \exists a_k, b_k \in \mathbb{R} \quad a_k \leqslant X_k \leqslant b_k \ \ a.s.$$

If $S_n = \sum_{k=1}^{n} X_k$ *and* $x \geqslant 0$, *then we have :*

$$\mathbb{P}(|S_n - \mathbb{E}(S_n)| \geqslant x) \leqslant 2 \exp\left(\frac{-2x^2}{\sum_{k=1}^{n}(b_k - a_k)^2}\right).$$

Hoeffding's Inequality

Theorem (Hoeffing)

Let $(X_k)_{k\in[\![1;n]\!]}$ *be independent random variables such that :*

$$\forall k \in [\![1;n]\!], \exists a_k, b_k \in \mathbb{R} \quad a_k \leqslant X_k \leqslant b_k \ \ a.s.$$

If $S_n = \sum_{k=1}^n X_k$ *and* $x \geqslant 0$, *then we have :*

$$\mathbb{P}(|S_n - \mathbb{E}(S_n)| \geqslant x) \leqslant 2\exp\left(\frac{-2x^2}{\sum_{k=1}^n (b_k - a_k)^2}\right).$$

- Our objective here is to generalise Hoeffding's inequality by replacing S_n by any square-integrable martingale without assuming that the increments are independent.

Azuma-Hoeffding Inequality

Theorem (Azuma-Hoeffding)

Let $(M_k)_{k\in[\![0;n]\!]}$ *be a martingale with finite variance such that* $M_0 = 0$ *and :*

$$\forall k \in [\![1;n]\!], \exists a_k, b_k \in \mathbb{R} \quad a_k \leqslant \Delta M_k := M_k - M_{k-1} \leqslant b_k \ \text{ a.s.}$$

If $x \geqslant 0$, *then we have :*

$$\mathbb{P}(|M_n| \geqslant x) \leqslant 2\exp\left(\frac{-2x^2}{\sum_{k=1}^{n}(b_k - a_k)^2}\right).$$

Demonstration - Azuma Hoeffding

By definition, $M_n = M_{n-1} + \Delta M_n$.

Demonstration - Azuma Hoeffding

By definition, $M_n = M_{n-1} + \Delta M_n$.
Let $t > 0$,

$$\begin{aligned}\mathbb{E}(\exp(tM_n)) &= \mathbb{E}[\mathbb{E}(\exp(tM_n)|\mathcal{F}_{n-1})] \\ &= \mathbb{E}[\exp(tM_{n-1})\mathbb{E}(\exp(t\Delta M_n)|\mathcal{F}_{n-1})].\end{aligned}$$

Demonstration - Azuma Hoeffding

By definition, $M_n = M_{n-1} + \Delta M_n$.
Let $t > 0$,

$$\begin{aligned}\mathbb{E}(\exp(tM_n)) &= \mathbb{E}[\mathbb{E}(\exp(tM_n)|\mathcal{F}_{n-1})] \\ &= \mathbb{E}[\exp(tM_{n-1})\mathbb{E}(\exp(t\Delta M_n)|\mathcal{F}_{n-1})].\end{aligned}$$

Technical Lemma

Let X a be real-valued centered variable such that $a \leqslant X \leqslant b$ a.s.
Then for all $t \geqslant 0$,

$$\mathbb{E}(\exp(tX)) \leqslant exp\left(\frac{t^2(b-a)^2}{8}\right).$$

Technical Lemma

Let X a be real-valued centered variable such that $a \leqslant X \leqslant b$ a.s. Then for all $t \geqslant 0$,

$$\mathbb{E}(\exp(tX)) \leqslant exp\left(\frac{t^2(b-a)^2}{8}\right).$$

Technical Lemma

Let X a be real-valued centered variable such that $a \leqslant X \leqslant b$ a.s. Then for all $t \geqslant 0$,

$$\mathbb{E}(\exp(tX)) \leqslant exp\left(\frac{t^2(b-a)^2}{8}\right).$$

We apply this lemma to the conditional expectation of $X = \Delta M_n$ which is bounded by hypothesis and centered by the martingale property : $\mathbb{E}(\Delta M_n | \mathcal{F}_{n-1}) = 0$.

Therefore, for all $t \geqslant 0$:

$$\mathbb{E}(\exp(t\Delta M_n)|\mathcal{F}_{n-1}) \leqslant \exp\left(\frac{t^2(b_n - a_n)^2}{8}\right).$$

Therefore, for all $t \geqslant 0$:

$$\mathbb{E}(\exp(t\Delta M_n)|\mathcal{F}_{n-1}) \leqslant \exp\left(\frac{t^2(b_n - a_n)^2}{8}\right).$$

The previous equality

$$\mathbb{E}(\exp(tM_n)) = \mathbb{E}[\exp(tM_{n-1})\mathbb{E}(\exp(t\Delta M_n)|\mathcal{F}_{n-1})]$$

thus becomes

$$\mathbb{E}(\exp(tM_n)) \leqslant \mathbb{E}[\exp(tM_{n-1})] \exp\left(\frac{t^2(b_n - a_n)^2}{8}\right).$$

By iterating the relation

$$\mathbb{E}(\exp(tM_n)) \leqslant \mathbb{E}[\exp(tM_{n-1})] \exp\left(\frac{t^2(b_n - a_n)^2}{8}\right),$$

By iterating the relation

$$\mathbb{E}(\exp(tM_n)) \leqslant \mathbb{E}[\exp(tM_{n-1})] \exp\left(\frac{t^2(b_n - a_n)^2}{8}\right),$$

we obtain :

$$\mathbb{E}(\exp(tM_n)) \leqslant \mathbb{E}[\exp(tM_0)] \exp\left(\frac{t^2 v_n}{8}\right)$$

where $v_n = \sum_{k=1}^{n}(b_k - a_k)^2$.

By iterating the relation

$$\mathbb{E}(\exp(tM_n)) \leqslant \mathbb{E}[\exp(tM_{n-1})] \exp\left(\frac{t^2(b_n - a_n)^2}{8}\right),$$

we obtain :

$$\mathbb{E}(\exp(tM_n)) \leqslant \mathbb{E}[\exp(tM_0)] \exp\left(\frac{t^2 v_n}{8}\right)$$

where $v_n = \sum_{k=1}^{n}(b_k - a_k)^2$.
Since $M_0 = 0$, we have :

$$\forall t \geqslant 0, \quad \mathbb{E}(\exp(tM_n)) \leqslant \exp\left(\frac{t^2 v_n}{8}\right).$$

$$\forall t \geqslant 0, \quad \mathbb{E}(\exp(tM_n)) \leqslant \exp\left(\frac{t^2 v_n}{8}\right).$$

$$\forall t \geqslant 0, \quad \mathbb{E}(\exp(tM_n)) \leqslant \exp\left(\frac{t^2 v_n}{8}\right).$$

Using Markov inequality, we obtain for all $t, x \geqslant 0$:

$$\mathbb{P}(M_n \geqslant x) \leqslant \exp(-tx)\mathbb{E}(\exp(tM_n))$$

$$\forall t \geqslant 0, \quad \mathbb{E}(\exp(tM_n)) \leqslant \exp\left(\frac{t^2 v_n}{8}\right).$$

Using Markov inequality, we obtain for all $t, x \geqslant 0$:

$$\mathbb{P}(M_n \geqslant x) \leqslant \exp(-tx)\mathbb{E}(\exp(tM_n))$$

which gives

$$\mathbb{P}(M_n \geqslant x) \leqslant \exp\left(-tx + \frac{t^2 v_n}{8}\right).$$

$$\forall t \geqslant 0, \quad \mathbb{E}(\exp(tM_n)) \leqslant \exp\left(\frac{t^2 v_n}{8}\right).$$

Using Markov inequality, we obtain for all $t, x \geqslant 0$:

$$\mathbb{P}(M_n \geqslant x) \leqslant \exp(-tx)\mathbb{E}(\exp(tM_n))$$

which gives

$$\mathbb{P}(M_n \geqslant x) \leqslant \exp\left(-tx + \frac{t^2 v_n}{8}\right).$$

We optimize in t the right bound with $t = 4x/v_n$.
So

$$\mathbb{P}(M_n \geqslant x) \leqslant \exp\left(\frac{-2x^2}{v_n}\right).$$

We have shown that

$$\mathbb{P}(M_n \geqslant x) \leqslant \exp\left(\frac{-2x^2}{v_n}\right).$$

We can obtain the following inequality in the same way by replacing M_n by $-M_n$:

$$\mathbb{P}(M_n \leqslant -x) \leqslant \exp\left(\frac{-2x^2}{v_n}\right).$$

We have shown that

$$\mathbb{P}(M_n \geqslant x) \leqslant \exp\left(\frac{-2x^2}{v_n}\right).$$

We can obtain the following inequality in the same way by replacing M_n by $-M_n$:

$$\mathbb{P}(M_n \leqslant -x) \leqslant \exp\left(\frac{-2x^2}{v_n}\right).$$

We can conclude for all $x \geqslant 0$:

$$\mathbb{P}(|M_n| \geqslant x) = \mathbb{P}(M_n \geqslant x) + \mathbb{P}(M_n \leqslant -x) \leqslant 2\exp\left(\frac{-2x^2}{v_n}\right)$$

We have shown that

$$\mathbb{P}(M_n \geqslant x) \leqslant \exp\left(\frac{-2x^2}{v_n}\right).$$

We can obtain the following inequality in the same way by replacing M_n by $-M_n$:

$$\mathbb{P}(M_n \leqslant -x) \leqslant \exp\left(\frac{-2x^2}{v_n}\right).$$

We can conclude for all $x \geqslant 0$:

$$\mathbb{P}(|M_n| \geqslant x) = \mathbb{P}(M_n \geqslant x) + \mathbb{P}(M_n \leqslant -x) \leqslant 2\exp\left(\frac{-2x^2}{v_n}\right)$$

hence the desired inequality

$$\forall x \geqslant 0 \qquad \mathbb{P}(|M_n| \geqslant x) \leqslant 2\exp\left(\frac{-2x^2}{\sum_{k=1}^{n}(b_k - a_k)^2}\right).$$

$\square$

Bennett's Inequality

Theorem (Bennett)

Let $(X_k)_{k\in[\![1;n]\!]}$ *be independent random variables centered and square integrable such that :*

$$\forall k \in [\![1; n]\!], \exists b > 0 \quad X_k \leqslant b \;\; a.s.$$

Then for all $n \in \mathbb{N}^*, x \in [0; b]$,

$$\mathbb{P}(S_n \geqslant x) \leqslant \exp\left(- \frac{x^2}{2(Var(S_n) + bx/3)} \right).$$

Bennett's Inequality

Theorem (Bennett)

Let $(X_k)_{k\in[\![1;n]\!]}$ be independent random variables centered and square integrable such that :

$$\forall k \in [\![1;n]\!], \exists b > 0 \quad X_k \leqslant b \ \ a.s.$$

Then for all $n \in \mathbb{N}^*, x \in [0;b]$,

$$\mathbb{P}(S_n \geqslant x) \leqslant \exp\left(-\frac{x^2}{2(Var(S_n) + bx/3)}\right).$$

- We obtain Poissonian tails that have the advantage of taking the variance into account.

Bennett's Inequality

Theorem (Bennett)

Let $(X_k)_{k\in[\![1;n]\!]}$ be independent random variables centered and square integrable such that :

$$\forall k \in [\![1;n]\!], \exists b > 0 \quad X_k \leqslant b \ \ a.s.$$

Then for all $n \in \mathbb{N}^*, x \in [0;b]$,

$$\mathbb{P}(S_n \geqslant x) \leqslant \exp\left(-\frac{x^2}{2(Var(S_n) + bx/3)} \right).$$

- We obtain Poissonian tails that have the advantage of taking the variance into account.
- Our objective here is to generalise Bennett's inequality by replacing S_n by any square-integrable martingale without assuming that the increments are independent.

Freedman's Inequality

Theorem (Freedman(1975))

Let $(M_n)_{n\in\mathbb{N}}$ be a square integrable martingale starting from 0 such that the increments $\Delta M_k = M_k - M_{k-1}$, $k \in \mathbb{N}^*$ are as bounded from above by $b \in \mathbb{R}_+^*$, then for all $n \in \mathbb{N}^*, x \in [0; b]$ and $y \in \mathbb{R}_+^*$,

$$\mathbb{P}(M_n \geqslant nx, \langle M\rangle_n \leqslant ny) \leqslant \exp\left(-\frac{nx^2}{2(y + bx/3)}\right).$$

Freedman's Inequality

Theorem (Freedman(1975))

Let $(M_n)_{n\in\mathbb{N}}$ be a square integrable martingale starting from 0 such that the increments $\Delta M_k = M_k - M_{k-1}$, $k \in \mathbb{N}^*$ are as bounded from above by $b \in \mathbb{R}_+^*$, then for all $n \in \mathbb{N}^*, x \in [0; b]$ and $y \in \mathbb{R}_+^*$,

$$\mathbb{P}(M_n \geqslant nx, \langle M \rangle_n \leqslant ny) \leqslant \exp\left(-\frac{nx^2}{2(y + bx/3)}\right).$$

Since the upper bound is homogeneous in b, in the demonstration we will assume that $b = 1$, that is $\Delta M_k \leqslant 1$ for all $k \in \mathbb{N}^*$, almost surely.

Demonstration : the case of S_n

In order to guide our demonstration for the martingale case, we first have to understand what happens in the case of the sum of independent variables.

Demonstration : the case of S_n

In order to guide our demonstration for the martingale case, we first have to understand what happens in the case of the sum of independent variables.

Let $X_1, \ldots, X_n$ be independent centered variables with finite variance and let $S_n = \sum_{k=1}^{n} X_k$, we compute the Cramér transform of S_n. By independence of $X_1, \ldots, X_n$, we have

$$\mathcal{L}_{S_n}(t) = \ln(\mathbb{E}[\exp(tS_n)]) = \sum_{k=1}^{n} \ln(\mathbb{E}[\exp(tX_k)]).$$

Demonstration : the case of S_n

In order to guide our demonstration for the martingale case, we first have to understand what happens in the case of the sum of independent variables.

Let $X_1, \ldots, X_n$ be independent centered variables with finite variance and let $S_n = \sum_{k=1}^{n} X_k$, we compute the Cramér transform of S_n. By independence of $X_1, \ldots, X_n$, we have

$$\mathcal{L}_{S_n}(t) = \ln(\mathbb{E}[\exp(tS_n)]) = \textstyle\sum_{k=1}^{n} \ln(\mathbb{E}[\exp(tX_k)]).$$

Thus, using the concavity of the logarithm yields

$$\frac{1}{n}\mathcal{L}_{S_n}(t) \leqslant \ln\left(\frac{1}{n}\sum_{k=1}^{n}\mathbb{E}[\exp(tX_k)]\right).$$

$$\frac{1}{n}\mathcal{L}_{S_n}(t) \leqslant \ln\left(\frac{1}{n}\sum_{k=1}^{n}\mathbb{E}[\exp(tX_k)]\right).$$

Let X be a random variable with distribution $\frac{1}{n}\sum_{k=1}^{n} P_{X_k}$.
By the construction of X, $\frac{1}{n}\sum_{k=1}^{n}\mathbb{E}[\exp(tX_k)] = \mathbb{E}[\exp(tX)]$.

$$\frac{1}{n}\mathcal{L}_{S_n}(t) \leqslant \ln\left(\frac{1}{n}\sum_{k=1}^{n}\mathbb{E}[\exp(tX_k)]\right).$$

Let X be a random variable with distribution $\frac{1}{n}\sum_{k=1}^{n} P_{X_k}$.
By the construction of X, $\frac{1}{n}\sum_{k=1}^{n}\mathbb{E}[\exp(tX_k)] = \mathbb{E}[\exp(tX)]$.

Lemma

Suppose that $X \leq 1$ *as*, $\mathbb{E}[X] = 0$ and $\mathbb{E}[X^2] = v = Var(S_n)/n$.
Let $\xi \sim \frac{v}{1+v}\delta_1 + \frac{1}{1+v}\delta_{-v}$ be a centered Bernoulli random variable with variance v, then almost surely

$$\mathbb{E}[\exp(tX)] \leqslant \mathbb{E}[\exp(t\xi)] = \frac{v}{1+v}e^{t} + \frac{1}{1+v}e^{-vt}.$$

We obtain the upper bound for the Cramér transform $\mathcal{L}_{S_n}(t) \leqslant nL_v(t)$ where $L_v(t) = \ln(ve^t + e^{-vt}) - \ln(1+v)$.

We obtain the upper bound for the Cramér transform $\mathcal{L}_{S_n}(t) \leqslant nL_v(t)$ where $L_v(t) = \ln(ve^t + e^{-vt}) - \ln(1+v)$.

We now compute the Fenchel-Legendre derivative of L_v and we eventually find $L_v^*(x) = \frac{x+v}{1+v}\ln(1+\frac{x}{v}) + \frac{1-x}{1+v}\ln(1-x)$ that can be lower-bounded by $\frac{x^2}{2(v+x/3)}$.

We obtain the upper bound for the Cramér transform $\mathcal{L}_{S_n}(t) \leqslant nL_v(t)$ where $L_v(t) = \ln(ve^t + e^{-vt}) - \ln(1+v)$.

We now compute the Fenchel-Legendre derivative of L_v and we eventually find $L_v^*(x) = \frac{x+v}{1+v}\ln(1+\frac{x}{v}) + \frac{1-x}{1+v}\ln(1-x)$ that can be lower-bounded by $\frac{x^2}{2(v+x/3)}$.

Using Cramér's inequality finally leads to Bennett's inequality

$$\mathbb{P}(S_n \geqslant nx) \leqslant \exp\left(-\frac{nx^2}{2(v+bx/3)}\right).$$

□

Demonstration : adaptation to the martingale case

Similarly to what we have seen for the sum of independent variables, we introduce $L_v : t \mapsto \ln(ve^t + e^{-vt}) - \ln(1+v)$

Demonstration : adaptation to the martingale case

Similarly to what we have seen for the sum of independent variables, we introduce $L_v : t \mapsto \ln(ve^t + e^{-vt}) - \ln(1+v)$ and the random variable $W_n(t) = \exp(tM_n - \sum_{k=1}^{n} L_{V_k}(t))$ where

$$V_k = \langle V \rangle_k - \langle V \rangle_{k-1} = \mathbb{E}[(M_k - M_{k-1})^2 \mid \mathcal{F}_{k-1}].$$

Demonstration : adaptation to the martingale case

Similarly to what we have seen for the sum of independent variables, we introduce $L_v : t \mapsto \ln(ve^t + e^{-vt}) - \ln(1+v)$ and the random variable $W_n(t) = \exp(tM_n - \sum_{k=1}^n L_{V_k}(t))$ where

$$V_k = \langle V \rangle_k - \langle V \rangle_{k-1} = \mathbb{E}[(M_k - M_{k-1})^2 \mid \mathcal{F}_{k-1}].$$

The same technical lemma than previously ensures that $\mathbb{E}[\exp(t\Delta M_k) \mid \mathcal{F}_{k-1}] \leqslant \mathbb{E}[\exp(t\xi)]$ with ξ a centered Bernoulli variable with variance V_k conditionnally to $\mathcal{F}_{k-1}$.

We obtain $\mathbb{E}[\exp(t\Delta M_k) \mid \mathcal{F}_{k-1}] \leqslant \exp(L_{V_k}(t))$ almost surely.

We obtain $\mathbb{E}[\exp(t\Delta M_k) \mid \mathcal{F}_{k-1}] \leqslant \exp(L_{V_k}(t))$ almost surely.

Recall that $W_n(t) = \exp(tM_n - \sum_{k=1}^{n} L_{V_k}(t))$. Then we get

$$\begin{aligned} \mathbb{E}[W_n(t) \mid \mathcal{F}_{n-1}] &= W_{n-1}(t)\mathbb{E}[\exp(t\Delta M_n) \mid \mathcal{F}_{n-1}] \exp(-L_{V_n}(t)) \\ &\leqslant W_{n-1}(t). \end{aligned}$$

We obtain $\mathbb{E}[\exp(t\Delta M_k) \mid \mathcal{F}_{k-1}] \leqslant \exp(L_{V_k}(t))$ almost surely.

Recall that $W_n(t) = \exp(tM_n - \sum_{k=1}^n L_{V_k}(t))$. Then we get

$$\begin{aligned}\mathbb{E}[W_n(t) \mid \mathcal{F}_{n-1}] &= W_{n-1}(t)\mathbb{E}[\exp(t\Delta M_n) \mid \mathcal{F}_{n-1}]\exp(-L_{V_n}(t)) \\ &\leqslant W_{n-1}(t).\end{aligned}$$

It means that $(W_n(t))_{n\in\mathbb{N}}$ is a supermartingale starting from 1 and for all $n \in \mathbb{N}$, $\mathbb{E}[W_n(t)] \leqslant 1$.

End of the demonstration

Let A_n denote the event $(M_n \geqslant nx, \langle M \rangle_n \leqslant ny)$.
For $t \in \mathbb{R}_+^*$, using Markov's inequality yields

$$\mathbb{P}(A_n) \leqslant \mathbb{E}\big[\exp(tM_n - tnx)\mathbb{1}_{\langle M \rangle_n \leqslant ny}\big].$$

End of the demonstration

Let A_n denote the event $(M_n \geqslant nx, \langle M \rangle_n \leqslant ny)$.
For $t \in \mathbb{R}_+^*$, using Markov's inequality yields

$$\mathbb{P}(A_n) \leqslant \mathbb{E}[\exp(tM_n - tnx)\mathbb{1}_{\langle M \rangle_n \leqslant ny}].$$

So,

$$\mathbb{P}(A_n) \leqslant \mathbb{E}\left[\exp\left(tM_n - \sum_{k=1}^{n} L_{V_k}(t)\right)\exp\left(\sum_{k=1}^{n} L_{V_k}(t) - tnx\right)\mathbb{1}_{\langle M \rangle_n \leqslant ny}\right]$$

End of the demonstration

Let A_n denote the event $(M_n \geqslant nx, \langle M \rangle_n \leqslant ny)$.
For $t \in \mathbb{R}_+^*$, using Markov's inequality yields

$$\mathbb{P}(A_n) \leqslant \mathbb{E}[\exp(tM_n - tnx)\mathbb{1}_{\langle M \rangle_n \leqslant ny}].$$

So,

$$\mathbb{P}(A_n) \leqslant \mathbb{E}\left[\exp\left(tM_n - \sum_{k=1}^{n} L_{V_k}(t)\right)\exp\left(\sum_{k=1}^{n} L_{V_k}(t) - tnx\right)\mathbb{1}_{\langle M \rangle_n \leqslant ny}\right]$$

Since the function $v \mapsto L_v(t)$ is concave and non-decreasing, we have $\frac{1}{n}\sum_{k=1}^{n} L_{V_k}(t) \leqslant L_{\langle M \rangle_n / n}(t) \leqslant L_y(t)$ on the event A_n.

Therefore

$$\mathbb{P}(A_n) \leqslant \mathbb{E}[\exp\left(tM_n - \sum_{k=1}^{n} L_{V_k}(t)\right)\exp\left(nL_y(t) - tnx\right)\mathbb{1}_{\langle M\rangle_n \leqslant ny}]$$

$$\leqslant \mathbb{E}[W_n(t)]\exp(nL_y(t) - tnx).$$

Therefore

$$\begin{aligned}\mathbb{P}(A_n) &\leqslant \mathbb{E}[\exp\Big(tM_n - \sum_{k=1}^{n} L_{V_k}(t)\Big)\exp\Big(nL_y(t) - tnx\Big)\mathbb{1}_{\langle M\rangle_n \leqslant ny}] \\ &\leqslant \mathbb{E}[W_n(t)]\exp(nL_y(t) - tnx).\end{aligned}$$

In the end, we find that

$$\mathbb{P}(A_n) \leqslant \exp(-nL_y^*(x)).$$

Therefore

$$\mathbb{P}(A_n) \leqslant \mathbb{E}[\exp\left(tM_n - \sum_{k=1}^{n} L_{V_k}(t)\right)\exp\left(nL_y(t) - tnx\right)\mathbb{1}_{\langle M\rangle_n \leqslant ny}]$$
$$\leqslant \mathbb{E}[W_n(t)]\exp(nL_y(t) - tnx).$$

In the end, we find that

$$\mathbb{P}(A_n) \leqslant \exp(-nL_y^*(x)).$$

Since we have already computed the Legendre-Fenchel derivative of L_y, the proof is complete. □

An application to matrices

Theorem

Let $p < n \in \mathbb{N}^*$, $A \in M_{p,n}(\mathbb{R})$ *with coefficients bounded by* $b > 0$. *Let* π *be a uniformy distributed random variable on the set of one-to-one maps from* $[\![1; p]\!]$ *to* $[\![1; n]\!]$ *and let* $S = \sum_{i=1}^{p} A[i, \pi(i)]$.

Then for all $x \in \mathbb{R}_+^*$,

$$\mathbb{P}(|S - \mathbb{E}[S]| \geqslant x) \leqslant 2 \exp\left(- \frac{x^2}{4(\theta v + 2xb/3)} \right)$$

where

- $v = \frac{1}{n} \sum_{i=1}^{p} \sum_{j=1}^{n} A[i, j]^2$
- $\theta = -\alpha - \frac{1+\alpha}{\alpha} \ln(1 - \alpha)$
- $\alpha = p/n$.

We set, for all $i \in [\![1; p]\!]$, $\mathcal{F}_i = \sigma(\pi(1), \dots, \pi(i))$ and $M_i = \mathbb{E}[S \mid \mathcal{F}_i] - \mathbb{E}[S]$. Then $(M_i)_{i \in [\![1;p]\!]}$ is a martingale.

We set, for all $i \in [\![1;p]\!]$, $\mathcal{F}_i = \sigma(\pi(1), \ldots, \pi(i))$ and $M_i = \mathbb{E}[S \mid \mathcal{F}_i] - \mathbb{E}[S]$. Then $(M_i)_{i \in [\![1;p]\!]}$ is a martingale.

But given $\mathcal{F}_i$, $\pi_{|[\![i+1;p]\!]}$ is uniformly distributed on the set of one-to-one maps from $[\![i+1;p]\!]$ to $[\![1;n]\!] \backslash \{\pi(1), \ldots, \pi(i)\}$.

We set, for all $i \in [\![1; p]\!]$, $\mathcal{F}_i = \sigma(\pi(1), \ldots, \pi(i))$ and $M_i = \mathbb{E}[S \mid \mathcal{F}_i] - \mathbb{E}[S]$. Then $(M_i)_{i \in [\![1;p]\!]}$ is a martingale.

But given $\mathcal{F}_i$, $\pi_{|[\![i+1;p]\!]}$ is uniformly distributed on the set of one-to-one maps from $[\![i+1; p]\!]$ to $[\![1; n]\!] \backslash \{\pi(1), \ldots, \pi(i)\}$.

Thus by defining $m_i(k, :) = \mathbb{E}[A[k, \pi(i+1)] \mid \mathcal{F}_i]$, we obtain $m_i(k, :) = \frac{1}{n-i} \sum_{j \notin \{\pi(1), \ldots, \pi(i)\}} A[k, j]$

We set, for all $i \in [\![1;p]\!]$, $\mathcal{F}_i = \sigma(\pi(1), \ldots, \pi(i))$ and $M_i = \mathbb{E}[S \mid \mathcal{F}_i] - \mathbb{E}[S]$. Then $(M_i)_{i \in [\![1;p]\!]}$ is a martingale.

But given $\mathcal{F}_i$, $\pi_{|[\![i+1;p]\!]}$ is uniformly distributed on the set of one-to-one maps from $[\![i+1;p]\!]$ to $[\![1;n]\!] \setminus \{\pi(1), \ldots, \pi(i)\}$.

Thus by defining $m_i(k,:) = \mathbb{E}[A[k, \pi(i+1)] \mid \mathcal{F}_i]$, we obtain $m_i(k,:) = \frac{1}{n-i} \sum_{j \notin \{\pi(1),\ldots,\pi(i)\}} A[k,j]$ which in turn gives

$$\Delta M_i = A[i, \pi(i)] - m_{i-1}(i,:) + \sum_{k=i+1}^{p} [m_i(k,:) - m_{i-1}(k,:)].$$

We set, for all $i \in [\![1; p]\!]$, $\mathcal{F}_i = \sigma(\pi(1), \ldots, \pi(i))$ and $M_i = \mathbb{E}[S \mid \mathcal{F}_i] - \mathbb{E}[S]$. Then $(M_i)_{i \in [\![1;p]\!]}$ is a martingale.

But given $\mathcal{F}_i$, $\pi_{|[\![i+1;p]\!]}$ is uniformly distributed on the set of one-to-one maps from $[\![i+1; p]\!]$ to $[\![1; n]\!] \setminus \{\pi(1), \ldots, \pi(i)\}$.

Thus by defining $m_i(k, :) = \mathbb{E}[A[k, \pi(i+1)] \mid \mathcal{F}_i]$, we obtain $m_i(k, :) = \frac{1}{n-i} \sum_{j \notin \{\pi(1), \ldots, \pi(i)\}} A[k, j]$ which in turn gives

$$\Delta M_i = A[i, \pi(i)] - m_{i-1}(i, :) + \sum_{k=i+1}^{p} [m_i(k, :) - m_{i-1}(k, :)].$$

We eventually find that $|\Delta M_i| \leqslant 4b$ as since

$$m_i(k, :) - m_{i-1}(k, :) = \frac{1}{n-i}(m_{i-1}(k, :) - A[k, \pi(i)]).$$

Applying Freedman's inequality to both M and $-M$ yields

$$\mathbb{P}(|M_p| \geqslant x) \leqslant 2\exp\Big(-\frac{x^2}{2(\|\langle M\rangle_p\|_\infty + 4bx/3)}\Big).$$

Applying Freedman's inequality to both M and $-M$ yields

$$\mathbb{P}(|M_p| \geqslant x) \leqslant 2 \exp\Big(- \frac{x^2}{2(\|\langle M \rangle_p\|_\infty + 4bx/3)} \Big).$$

We admit that $\langle M \rangle_p \leqslant 2\theta v$ almost surely where

$$v = \frac{1}{n} \sum_{i=1}^{p} \sum_{j=1}^{n} A[i,j]^2 \;\; ; \quad \theta = -\alpha - \frac{1+\alpha}{\alpha} \ln(1-\alpha) \;\; ; \quad \alpha = \frac{p}{n}$$

giving the result. □

Applying Freedman's inequality to both M and $-M$ yields

$$\mathbb{P}(|M_p| \geqslant x) \leqslant 2\exp\left(-\frac{x^2}{2(\|\langle M\rangle_p\|_\infty + 4bx/3)}\right).$$

We admit that $\langle M\rangle_p \leqslant 2\theta v$ almost surely where

$$v = \frac{1}{n}\sum_{i=1}^{p}\sum_{j=1}^{n} A[i,j]^2 \quad ; \quad \theta = -\alpha - \frac{1+\alpha}{\alpha}\ln(1-\alpha) \quad ; \quad \alpha = \frac{p}{n}$$

giving the result. □

- We would like to have a similar result for square matrices.

Applying Freedman's inequality to both M and $-M$ yields

$$\mathbb{P}(|M_p| \geqslant x) \leqslant 2\exp\left(-\frac{x^2}{2(\|\langle M\rangle_p\|_\infty + 4bx/3)}\right).$$

We admit that $\langle M\rangle_p \leqslant 2\theta v$ almost surely where

$$v = \frac{1}{n}\sum_{i=1}^{p}\sum_{j=1}^{n} A[i,j]^2 \quad ; \quad \theta = -\alpha - \frac{1+\alpha}{\alpha}\ln(1-\alpha) \quad ; \quad \alpha = \frac{p}{n}$$

giving the result. □

- We would like to have a similar result for square matrices.
- Because the constant θ is not defined for $\alpha = 1$, that is $p = n$, we will decompose the matrix into two rectangular matrices and apply the previous result twice.

The square matrix version

Let $n \geqslant 2$, $A \in M_n(\mathbb{R})$, $\pi \sim \mathcal{U}(\mathfrak{S}_n)$ and $S = \sum_{i=1}^n A[i, \pi(i)]$. Set $v = \frac{1}{n} \sum_{i,j=1}^n A[i,j]^2$ and suppose that $|A[i,j]| \leqslant b$ for all $i, j \in [\![1; n]\!]$.

The square matrix version

Let $n \geqslant 2$, $A \in M_n(\mathbb{R})$, $\pi \sim \mathcal{U}(\mathfrak{S}_n)$ and $S = \sum_{i=1}^{n} A[i, \pi(i)]$. Set $v = \frac{1}{n}\sum_{i,j=1}^{n} A[i,j]^2$ and suppose that $|A[i,j]| \leqslant b$ for all $i, j \in [\![1; n]\!]$.

Now write a decomposition $S = P + Q$ with $P = \sum_{i=1}^{p} A[i, \pi(i)]$ and $Q = \sum_{i=p+1}^{n} A[i, \pi(i)]$ where $p = \lfloor \frac{n}{2} \rfloor$.

The square matrix version

Let $n \geqslant 2$, $A \in M_n(\mathbb{R})$, $\pi \sim \mathcal{U}(\mathfrak{S}_n)$ and $S = \sum_{i=1}^n A[i, \pi(i)]$. Set $v = \frac{1}{n} \sum_{i,j=1}^n A[i,j]^2$ and suppose that $|A[i,j]| \leqslant b$ for all $i, j \in [\![1; n]\!]$.

Now write a decomposition $S = P + Q$ with $P = \sum_{i=1}^p A[i, \pi(i)]$ and $Q = \sum_{i=p+1}^n A[i, \pi(i)]$ where $p = \lfloor \frac{n}{2} \rfloor$.

Let $x \in \mathbb{R}_+^*$, we have :

$$\mathbb{P}(|S - \mathbb{E}[S]| \geqslant x) \leqslant \mathbb{P}(|P - \mathbb{E}[P]| \geqslant x/2) + \mathbb{P}(|Q - \mathbb{E}[Q]| \geqslant x/2).$$

The square matrix version

Let $n \geqslant 2$, $A \in M_n(\mathbb{R})$, $\pi \sim \mathcal{U}(\mathfrak{S}_n)$ and $S = \sum_{i=1}^{n} A[i, \pi(i)]$.
Set $v = \frac{1}{n} \sum_{i,j=1}^{n} A[i,j]^2$ and suppose that $|A[i,j]| \leqslant b$ for all $i, j \in [\![1; n]\!]$.

Now write a decomposition $S = P + Q$ with $P = \sum_{i=1}^{p} A[i, \pi(i)]$ and $Q = \sum_{i=p+1}^{n} A[i, \pi(i)]$ where $p = \lfloor \frac{n}{2} \rfloor$.

Let $x \in \mathbb{R}_+^*$, we have :

$$\mathbb{P}(|S - \mathbb{E}[S]| \geqslant x) \leqslant \mathbb{P}(|P - \mathbb{E}[P]| \geqslant x/2) + \mathbb{P}(|Q - \mathbb{E}[Q]| \geqslant x/2).$$

In the end, we get $\mathbb{P}(|S - \mathbb{E}[S]| \geqslant x) \leqslant 4 \exp\left(-\frac{x^2}{16(\theta v + xb/3)}\right)$ with the constant $\theta = \frac{5}{2}\ln(3) - \frac{2}{3}$. □

Remark

Defining $S = \sum_{i=1}^{n} A[i, \pi(i)]$, we have shown that :

$$\mathbb{P}(|S - \mathbb{E}[S]| \geqslant x) \leqslant 4 \exp\left(-\frac{x^2}{16(\theta v + xb/3)}\right)$$

with the constant $\theta = \frac{5}{2}\ln(3) - \frac{2}{3}$.

Remark

When $(x_1, y_1), \ldots, (x_n, y_n)$ are iid observations of a distribution P and $A[i, j] = f(x_i, y_j)$, this inequality may be used for a test of independence of the marginals of P.

Do you have any questions ?

References :

[1] Bercu, Delyon and Rio, Concentration inequalities for sums and martingales (2015).
[2] Chung and Lu, Concentration inequalities and martingale inequalities : a survey (2006).

Happy birthday Nolwenn !