Apprentissage statistique
TD2 : Inégalités de concentration

Lucie Le Briquer

5 février 2018

1 TD1

2 Inégalités de concentration

Exercice 2.1

1. X v.a. positive. Comme $t\mathbb{1}_{X \geqslant t} \leqslant X$, $\mathbb{P}(X \geqslant t) \leqslant t^{-1}\mathbb{E}[X]$.
2. (inégalité de Markov généralité) Y v.a. réelle telle que $\mathbb{P}(Y \in I) = 1$ avec I intervalle de $\mathbb{R}$. Soit $g\colon I \longrightarrow \mathbb{R}_+$ croissante.

$$\{Y \geqslant t\} \cap \{Y \in I\} \subset \{g(Y) \geqslant t\} \cap \{Y \in I\} \quad \text{car } g \text{ croissante sur } I$$

Alors,

$$\mathbb{P}(Y \geqslant t) = \mathbb{P}(Y \geqslant t, Y \in I) \leqslant \mathbb{P}(g(Y) \geqslant g(t), Y \in I) \leqslant g^{-1}(t)\mathbb{E}[g(Y)]$$

Exercice 2.2

1. Soit $s > 0$, posons $g_s\colon x \longmapsto \exp(sx)$. Alors, par l'exercice 1,

$$\mathbb{P}(X \geqslant t) \leqslant g(t)^{-1}\mathbb{E}[\exp(sX)] \leqslant \exp(-st)\exp\left(\frac{s^2b^2}{2}\right)$$

En minimisant en s, on obtient :

$$\mathbb{P}(X \geqslant t) \leqslant \exp\left(-\frac{t^2}{2b^2}\right)$$

2.
$$\mathbb{E}[\exp(tX)] = \mathbb{E}\left[\sum_{k=0}^{+\infty} \frac{t^k X^k}{k!}\right]$$

Montrons que :

$$\sum_{k=0}^{+\infty} \frac{\mathbb{E}[|X|^k]}{k!} < +\infty$$

On a $|X| \leqslant e^{|X|} \leqslant e^X + e^{-X}$. De plus,

$$\sum_{k\geqslant 0} \frac{\mathbb{E}[|X|^k]}{k!} \underset{\text{Beppo-Levy}}{=} \mathbb{E}[e^{|X|}] \leqslant \mathbb{E}[e^X] + \mathbb{E}[e^{-X}] < +\infty$$

L'interversion somme/intégrale est donc justifiée. Alors,

$$\mathbb{E}[e^{tX}] = 1 + t\mathbb{E}[X] + \frac{t^2}{2}\mathbb{E}[X^2] + t^3\varepsilon_1(t) \leqslant 1 + \frac{t^2b^2}{2} + t^3\varepsilon_2(t) \quad (*)$$

On a que $t\mathbb{E}[X] \underset{t\sim 0}{\leqslant} t^2\varepsilon_3(t) \ \Rightarrow\ \mathbb{E}[X] = 0$ avec t et $-t$.

De $(*)$, on en déduit :

$$\frac{t^2}{2}\mathbb{E}[X^2] \leqslant \frac{t^2b^2}{2} + \varepsilon_4(t)t^3$$

En divisant par t puis en prenant la limite on aura $\mathbb{E}[X^2] = \mathrm{Var}(X) \leqslant b^2$.

Exercice 2.3

1. $S_n = \sum X_i$, avec X_i b_i-sous-gaussienne. Il suffit de montrer que S_n est $\sqrt{\sum b_i^2}-$sous-gaussienne.

$$\mathbb{E}\left[\exp\left(s\sum_{i=1}^n X_i\right)\right] = \prod_{i=1}^n \mathbb{E}[\exp(sX_i)] = \exp\left(\frac{s^2}{2}\left(\sum_{i=1}^n b_i^2\right)\right)$$

Ainsi par l'exercice 2,

$$\mathbb{P}(S_n \geqslant t) \leqslant \exp\left(-\frac{t^2}{2\sum_{i=1}^n b_i^2}\right)$$

Comme $-S_n$ est aussi $\sqrt{\sum b_i^2}$, on a comme majoration :

$$\mathbb{P}(|S_n| \geqslant t) \leqslant \mathbb{P}(S_n \geqslant t) + \mathbb{P}(-S_n \geqslant t) \leqslant 2\exp\left(-\frac{t^2}{2\sum_{i=1}^n b_i^2}\right)$$

2. (X_i) v.a.i.i.d. $\mathcal{N}(0, \sigma^2)$. Alors $\mathbb{E}[e^{tX_i}] = \exp\left(\frac{t^2\sigma^2}{2}\right)$. Ainsi,

$$\mathbb{P}(|S_n| \geqslant t) \leqslant 2\exp\left(-\frac{t^2}{2n\sigma^2}\right)$$

3. (X_i) v.a.i.i.d. de loi de Rademacher. On a :

$$\mathbb{E}[e^{tX}] = \frac{e^t + e^{-t}}{2} = \mathrm{ch}(t) \leqslant \exp\left(\frac{t^2}{2}\right)$$

Exercice 2.4

Il suffit de montrer que $\forall i$, X_i est $\frac{M_i - m_i}{2}$−sous-gaussienne. Soit X v.a. telle que $m \leqslant X \leqslant M$, X est $\frac{M-m}{2}$−sous-gaussienne ? Soit $s \in \mathbb{R}$. On peut décomposer X comme :

$$X = M\frac{X-m}{M-m} + m\frac{M-X}{M-m}$$

$$\exp(sX) \leqslant \frac{(X-m)e^{sM} + (M-X)e^{sm}}{M-m} \quad \text{par convexité}$$

Ainsi,

$$\mathbb{E}[e^{sX}] \leqslant \frac{-m}{M-m}e^{sM} + \frac{M}{M-m}e^{sm} \quad \text{car } \mathbb{E}[X] = 0$$

Par passage au log, on obtient :

$$\begin{aligned} \log(\mathbb{E}[e^{sX}]) &\leqslant \log\left(e^{sm}\left(\frac{M}{M-m} - \frac{m}{M-m}e^{s(M-m)}\right)\right) \\ &\leqslant sm + \log\left(\frac{M}{M-m} - \frac{m}{M-m}e^{s(M-m)}\right) \\ &= \varphi_p(u) \end{aligned}$$

avec $u = s(M-m)$ et $p = \frac{M}{M-m}$. Or $u \mapsto \varphi_p(u)$ $\mathcal{C}^\infty(\mathbb{R}_+)$.

$$\varphi_p'(0) = 0 \qquad \varphi_p''(u) = \frac{p(1-p)\exp(u)}{(p+(1-p)\exp u)^2} \leqslant \frac{1}{4} \quad \forall u \in \mathbb{R}_+$$

Car $\frac{(a+b)^2}{4} = \frac{a^2+b^2}{4} + \frac{ab}{2} ge \frac{2ab}{4} + \frac{ab}{2} \geqslant ab$.

$$\begin{aligned} \varphi_p(u) &= \varphi_p(0) + \varphi_p'(0)u + \int_0^u (u-v)\varphi_p''(v)dv \\ |\varphi_p(u)| &\leqslant 0 + 0 + \int_0^u (u-v)\frac{1}{4}dv \\ &\leqslant \frac{u^2}{8} \qquad \forall u \geqslant 0 \end{aligned}$$

Conclusion : pour tout $s \geqslant 0$,

$$\mathbb{E}[\exp(sX)] \leqslant \exp\left(\frac{s^2(M-m)^2}{2}\right)$$

En considérant $-X$, on l'obtient pour tout $x \in \mathbb{R}$, ce qui conclut la preuve.

Exercice 2.5

(X_n) une $(\mathcal{F}_n)$−martingale. $S_n = \sum_{i=1}^n X_i$. On suppose que :

$$|X_{i+1} - X_i| \leqslant b_{i+1} < +\infty \quad \mathbb{P} - \text{p.s.}$$

1.
$$\mathbb{E}[\exp(t(Y_{n+1}-Y_n))|\mathcal{F}_n] \leqslant \exp(t^2 b_{n+1}^2/2) \quad (*)$$

$$\begin{aligned}
\mathbb{E}[\exp(t(Y_n-Y_0))] &= \mathbb{E}\left[\exp\left(t\sum_{i=0}^{n} Z_i\right)\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\exp\left(t\sum_{i=0}^{n-1} Z_i\right) \;\middle|\; \mathcal{F}_n\right]\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\exp\left(t\sum_{i=0}^{n-2} Z_i\right)\right] \mathbb{E}\left[\exp(tZ_{n-1}) \mid \mathcal{F}_n\right]\right] \\
&\leqslant e^{t^2 b_{n+1}^2/2}\mathbb{E}\left[\exp\left(t\sum_{i=0}^{n-2} Z_i\right)\right] \quad \text{car } (X_n)\ (\mathcal{F}_n)-\text{martingale}
\end{aligned}$$

où $Z_i = Y_{i+1} - Y_i$. Par récurrence on obtient alors que $S_n = Y_n - Y_0$ est $\sqrt{\sum b_i^2}$−sous-gaussienne.

2. Il suffit de montrer que (X_n) satisfait $(*)$. On sait que l'on a :

$$|w_n| \leqslant |X_{n+1} - X_n| \leqslant b_{n+1} \ \mathbb{P}-\text{p.s.}$$

Alors,
$$m = -b_{n+1} \leqslant w_n \leqslant b_{n+1} = M$$
D'après la preuve de l'inégalité d'Hoeffding, on a $\forall t \in \mathbb{R}$:

$$w_n \leqslant M\frac{w_n - m}{M-m} + m\frac{M-w_n}{M-m}$$

D'où,
$$\mathbb{E}[\exp(tw_n) \mid \mathcal{F}_n] \leqslant \frac{e^{tM}}{M-m}\mathbb{E}[w_n - m \mid \mathcal{F}_n] + \frac{e^{tm}}{M-m}\mathbb{E}[M - w_n \mid \mathcal{F}_n]$$
En utilisant $\mathbb{E}[w_n|\mathcal{F}_n] = 0$ car (X_n) est une $(\mathcal{F}_n)$−martingale, on peut copier la preuve de Hoeffding et conclure que X_n satisfait $(*)$.

Exercice 2.6

1. Soit $(\mathcal{F}_k)_{k\leqslant n} = (\sigma(X_1,\ldots,X_k))_{k\leqslant n}$ et $\mathcal{F}_i = \mathcal{F}_n$ pour $i > n$. Posons $M_i = \mathbb{E}[f(X_1,\ldots,X_n) \mid \mathcal{F}_i]$ pour $i = 0,\ldots,n$ et $M_i = f(X_1,\ldots,X_n)$ pour $i \geqslant n+1$. $(M_i)_{i\in\mathbb{N}}$ est une $(\mathcal{F}_i)_{i\in\mathbb{N}}$−martingale. Il faut vérifier les hypothèses d'Azuma-Hoeffding.

$$|M_{i+1} - M_i| = \Big|\mathbb{E}[f(X_1,\ldots,X_n) \mid \mathcal{F}_{i+1}] - \mathbb{E}[f(X_1,\ldots,X_n) \mid \mathcal{F}_i]\Big|$$

Les (X_i) sont indépendants, donc :

$$\mathbb{E}[f(X_1,\ldots,X_n)|\mathcal{F}_i] = \psi_i(X_1,\ldots,X_i)$$

où $\psi_i(x_1,\dots,x_i) = \mathbb{E}[f(x_1,\dots,x_i,X_{i+1},\dots,X_n)]$. Donc :

$$\begin{aligned}|M_{i+1}-M_i| &= |\psi_{i+1}(X_1,\dots,X_{i+1}) - \psi_i(X_1,\dots,X_i)| \\ &= \big|\mathbb{E}[f(x_1,\dots,x_{i+1},X_{i+2},\dots,X_n) - f(x_1,\dots,x_i,X_{i+1},\dots,X_n)]\big| \\ &\leqslant b_{i+1}\end{aligned}$$

Conclusion, $|M_{i+1}-M_i| \leqslant b_{i+1}$ $\mathbb{P}-$p.s.

2. $(Y_{i,k})_{i\in 1:n,\ k\in 1:m}$.

$$f(Y_{1,1},\dots,Y_{n,m}) = g(Z_1^{(m)},\dots,Z_n^{(m)})$$

avec $Z_i = \frac{1}{\sqrt{m}}\sum_{k=0}^m Y_{i,k}$. Par le TCL en dimension n on a $(Z_1,\dots,Z_n) \longrightarrow \mathcal{N}(0,\mathrm{id}_n)$. D'après le théorème du porte-manteau, et comme g est continue, il suffit de montrer que $\forall m\in\mathbb{N}^*$:

$$\mathbb{P}\Big(g(Z_1^{(m)},\dots,Z_n^{(m)}) - \mathbb{E}[g(Z_1^{(m)},\dots,Z_n^{(m)})]\ \geqslant t\Big) \leqslant \exp\left(-\frac{t^2}{2n}\right) \quad (*)$$

Pour montrer $(*)$ pour tout m, on applique Mc Diarmid à $f(Y_{1,1},\dots,Y_{n,m})$. Comme :

$$\big|f(y_{1,1},\dots,y_{i,k},\dots,y_{n,m} - f(y_{1,1},\dots,\tilde{y}_{i,k},\dots,y_{n,m})\big| \leqslant |y_{i,k}-\tilde{y}_{i,k}| \leqslant \frac{1}{\sqrt{m}}$$

On obtient en appliquant Mc Diarmid :

$$\mathbb{P}\Big(g(Z_1^{(m)},\dots,Z_n^{(m)}) - \mathbb{E}[g()]\ \geqslant t\Big) \leqslant \exp\left(-\frac{t^2}{2\sum_{(i,k)\in\{1,\dots,n\}\times\{1,\dots,m\}}\frac{1}{m}}\right) = \exp\left(-\frac{t^2}{2n}\right)$$

Remarque.

$$\begin{aligned}(X_n) \Rightarrow Z\sim\mathcal{N}(0,\mathrm{id}) &\overset{\text{def}}{\Leftrightarrow} \mathbb{E}[h(Y_n)] \longrightarrow \mathbb{E}[h(Z)]\ \forall h\in\mathcal{C}_b(\mathbb{R}^d)) \\ &\overset{\text{porte-manteau}}{\Leftrightarrow} \mathbb{P}(g(X_n)\leqslant t) \longrightarrow \mathbb{P}(g(Z)\leqslant t)\ \forall t\in\mathbb{R} \text{ et } g\colon\mathbb{R}^d\longrightarrow\mathbb{R} \text{ continue}\end{aligned}$$

3 Concepts de base de l'apprentissage statistique : classifieurs et fonctions de perte

Exercice 3.1

1. Soit f un classifieur.

$$\begin{aligned}R_{\mathbb{P}}^{c_\omega}(f) &= \mathbb{E}[c_\omega(Y,f(X))] = \mathbb{E}\big[\omega_0\mathbb{1}_{Y=1}\mathbb{1}_{f(X)=0} + \omega_1\mathbb{1}_{Y=0}\mathbb{1}_{f(X)=1}\big] \\ &= \mathbb{E}\big[\omega_0 Y(1-f(X)) + \omega_1(1-Y)f(X)\big] \\ &= \mathbb{E}\Big[\mathbb{E}\big[\omega_0 Y(1-f(X)) + \omega_1(1-Y)f(X)\mid X\big]\Big] \qquad \eta(X) = \mathbb{E}[Y|X] \\ &= \mathbb{E}\big[\omega_0\eta(X)(1-f(X)) + \omega_1(1-\eta(X))f(X)\big] \\ &\geqslant \mathbb{E}\big[\min(\omega_0\eta(X)(1-f(X)),\omega_1(1-\eta(X))f(X)\big]\end{aligned}$$

On a égalité si $f(X)=1 \Leftrightarrow \omega_1(1-\eta(X)) \leqslant \omega_0\eta(X) \Leftrightarrow \eta(X) \geqslant \frac{\omega_1}{\omega_0+\omega_1}$. Donc

$$f(x) = \mathbb{1}_{\eta(x)\geqslant\frac{\omega_1}{\omega_0+\omega_1}}$$

est un classifieur de Bayes.

2. Excès de risque :

$$\begin{aligned}\rho(f,f^*) &= \mathbb{E}\big[\omega_0\eta(X)(f^*(X)-f(X)) + \omega_1(1-\eta(X))(f(X)-f^*(X))\big] \\ &= \mathbb{E}\big[\mathbb{1}_{f(X)=0}\mathbb{1}_{f^*(X)=1}(\omega_0\eta(X)-\omega_1(1-\eta(X))) \\ &\quad - \mathbb{1}_{f(X)=1}\mathbb{1}_{f^*(X)=0}(\omega_0\eta(X)-\omega_1(1-\eta(X)))\big]\end{aligned}$$

Sur $f^*(X)=1$, $\omega_0\eta(X)-\omega_1(1-\eta(X)) \geqslant 0$, sur $f^*(X)=0$, $\omega_0\eta(X)-\omega_1(1-\eta(X)) \leqslant 0$. D'où :

$$\rho(f,f^*) = (\omega_0+\omega_1)\mathbb{E}\left[\left|\eta(X)-\frac{\omega_1}{\omega_0+\omega_1}\right|\mathbb{1}_{f(X)\neq f^*(X)}\right]$$

Si f est un classifieur de Bayes, alors $\rho(f,f^*)=0$. Donc :

$$f(X)=f^*(X) \text{ sur } \eta(X) \neq \frac{\omega_1}{\omega_0+\omega_1}$$

4 L'algorithme de perceptron

Exercice 4.1

1. Notons $A = \big\{i \in \{1,\dots,T\},\ |\ \omega_i \neq \omega_{i-1}\big\}$.

$$\omega_t = \sum_{i\in A} y_i x_i$$

$$\begin{aligned}\langle\omega_t,\omega^*\rangle &= \sum_{i\in A} y_i x_i^T\omega^* \\ &\geqslant \mathrm{Card}(A)\rho\|\omega^*\| \quad \text{par hypothèse}\end{aligned}$$

Or par Cauchy-Schwarz,

$$|\langle\omega_t,\omega^*\rangle| \leqslant \|\omega^*\|\|\omega_T\|$$

Comme,

$$\begin{aligned}\|\omega_T\|^2 &= \|\omega_{T-1}\|^2 + 2\underbrace{y_t\langle\omega_{y-1},x_t\rangle}_{\leqslant 0 \text{ si } t\in A}\mathbb{1}_{t\in A} + \|x_t\|^2\mathbb{1}_{\{t\in A\}} \\ &\leqslant \|\omega_{T-1}\|^2 + r^2\mathbb{1}_{t\in A} \\ &\leqslant r^2\mathrm{Card}(A) \qquad \text{par récurrence}\end{aligned}$$

On a donc :

$$\rho\|\omega^*\|\mathrm{Card}(A) \leqslant \|\omega^*\| r\sqrt{\mathrm{Card}(A)}$$