Inégalités de concentration
Sorbonne Université
2021-2022

Corrigé

Exercice 1

Pour $x = (x_1, \dots, x_n) \in \mathcal{X}^n$, notons $f(x) = \sup_{A\in\mathcal{A}} \left|\frac{1}{n}\sum_{i=1}^n \mathbb{1}_{x_i\in A} - P(A)\right|$ et $A(x) \in \mathcal{A}$ le sous-ensemble tel que

$$f(x) = \left|\frac{1}{n}\sum_{i=1}^n \mathbb{1}_{x_i\in A(x)} - P(A(x))\right| .$$

Montrons d'abord que f est une fonction à différences bornées avec $c_i = \frac{1}{n}$. Soit Pour tout $x = (x_1, \dots, x_n) \in \mathcal{X}^n$ et $x_i' \in \mathcal{X}$, on a

$$\begin{aligned} f(x_1, &\dots, x_i, \dots, x_n) - f(x_1, \dots, x_i', \dots, x_n) \\ &\le \left|\frac{1}{n}\sum_{j=1}^n \mathbb{1}_{x_j\in A(x)} - P(A(x))\right| - \left|\frac{1}{n}\sum_{j\neq i} \mathbb{1}_{x_j\in A(x)} + \frac{1}{n}\mathbb{1}_{x_i'\in A(x)} - P(A(x))\right| \\ &\le \frac{1}{n}\mathbb{1}_{x_i'\in A(x)} \le \frac{1}{n} , \end{aligned}$$

par l'inégalité triangulaire. Par le même raisonnement avec $A(x_1, \dots, x_i', \dots, x_n)$, on obtient

$$f(x_1, \dots, x_i, \dots, x_n) - f(x_1, \dots, x_i', \dots, x_n) \ge -\frac{1}{n} .$$

La fonction f est donc bien à différences bornées avec $c_i = \frac{1}{n}$.

Soit maintenant $Z = f(X)$ avec $X = (X_1, \dots, X_n)$ i.i.d. de loi P, et définissons la variable Z_i par

$$Z_i = \frac{1}{2}\left(\inf_{x_i} f(X_1, \dots, x_i, \dots, X_n) + \sup_{x_i} f(X_1, \dots, x_i, \dots, X_n)\right) .$$

Par ce qui précède, on a $|Z - Z_i| \le \frac{1}{2n}$. L'inégalité d'Efron-Stein donne alors

$$\mathbf{Var}(Z) \le \sum_{i=1}^n \mathbf{E}\left[(Z - Z_i)^2\right] \le \frac{1}{4n} \cdot$$

Exercice 2

(1) La fonction f est à différences bornées avec $c_i = 1$. Par l'inégalité de McDiarmid, on obtient

$$\mathbf{P}\left(|Z - \mathbf{E}Z| \ge t\right) \le 2\exp\left(-\frac{2t^2}{n}\right) .$$

(2) Les x_i tels que $x_i = y_i$ peuvent être rangés dans $f(y)$ segments. Et pour les x_i tels que $x_i \neq y_i$, on peut les ranger dans des nouveaux segments de telle sorte qu'au plus un segment soit moins qu'à moitié rempli (s'il y en a deux, on peut les regrouper). Pour un tel rangement, le nombre de boîtes est inférieur à deux fois la somme, plus 1. On obtient bien

$$f(x) \le f(y) + 2\sum_{i=1}^n x_i \mathbb{1}_{y_i\neq x_i} + 1 .$$

(3) Soit Y une variable aléatoire de Q, dont la loi jointe avec X sera explicitée plus bas. Par la question précédente, on a

$$\begin{aligned}\mathbf{E}f - \mathbf{E}_Q f &= \mathbf{E}f(X) - \mathbf{E}f(Y) \\ &\leq 2\sum_{i=1}^n \mathbf{E}\left[X_i \mathbb{1}_{Y_i \neq X_i}\right] + 1 \\ &= 2\sum_{i=1}^n \mathbf{E}\left[X_i \mathbf{P}(Y_i \neq X_i \mid X)\right] + 1 \\ &\leq 2\sum_{i=1}^n \mathbf{E}\left[X_i^2\right]^{1/2} \mathbf{E}\left[\mathbf{P}(Y_i \neq X_i \mid X)^2\right]^{1/2} + 1 \\ &\leq 2\sqrt{\sum_{i=1}^n \mathbf{E}\left[X_i^2\right]}\sqrt{\sum_{i=1}^n \mathbf{E}\left[\mathbf{P}(Y_i \neq X_i \mid X)^2\right]} + 1\end{aligned}$$

Par l'inégalité de transport conditionnelle de Marton, il existe un couplage de (X, Y) tel que

$$\sum_{i=1}^n \mathbf{E}\left[\mathbf{P}(Y_i \neq X_i \mid X)^2\right] \leq 2\mathrm{D}(Q \mid P)\,.$$

Ainsi

$$\mathbf{E}f - \mathbf{E}_Q f \leq 2\sqrt{2v\mathrm{D}(Q \mid P)} + 1\,.$$

(4) Par le lemme de transport (appliqué à la variable $-Z$), on a, pour tout $\lambda \geq 0$,

$$\log \mathbf{E} e^{\lambda(\mathbf{E}Z - Z)} = \sup_{Q \ll P}\left\{\lambda(\mathbf{E}f - \mathbf{E}_Q f) - \mathrm{D}(Q \mid P)\right\}\,.$$

Ainsi, par la majoration de la question précédente,

$$\log \mathbf{E} e^{\lambda(\mathbf{E}Z - Z)} \leq \sup_{Q \ll P}\left\{\left(2\sqrt{2v\mathrm{D}(Q \mid P)} + 1\right)\lambda - \mathrm{D}(Q \mid P)\right\}\,.$$

En maximisant la fonction $u \mapsto \left(2\sqrt{2vu} + 1\right)\lambda - u$ sur $\mathbb{R}_+$, on obtient que le supremum est atteint pour Q telle que $\mathrm{D}(Q \mid P) = 2v\lambda^2$, ce qui donne

$$\log \mathbf{E} e^{\lambda(\mathbf{E}Z - Z)} \leq 2v\lambda^2 + \lambda\,.$$

Puis, par la méthode de Chernoff, on a, pour tout $t \geq 1$,

$$\begin{aligned}\mathbf{P}\left(Z - \mathbf{E}Z \leq -t\right) &\leq e^{-\sup_{\lambda \geq 0}\left\{\lambda t - \log \mathbf{E} e^{\lambda(\mathbf{E}Z - Z)}\right\}} \\ &\leq e^{-\sup_{\lambda \geq 0}\left\{\lambda(t-1) - 2v\lambda^2\right\}}\,.\end{aligned}$$

En optimisant, on trouve bien

$$\mathbf{P}\left(Z - \mathbf{E}Z \leq -t\right) \leq e^{-\frac{(t-1)^2}{8v}}\,.$$

Exercice 3

(1) Par convexité de $u \mapsto e^u$ et comme $X \in [0, 1]$, on a, pour tout $\lambda \in \mathbb{R}$,

$$e^{\lambda X} \leq X e^{\lambda} + 1 - X\,.$$

En prenant l'espérance, on a

$$\mathbf{E} e^{\lambda X} \leq p e^{\lambda} + 1 - p\,,$$

qui correspond à la transformée de Laplace d'une variable de Bernoulli de paramètre p.

(2) Pour tout $q \in]p,1[$, on a

$$\begin{aligned} g_p'(q) &= \frac{1}{(q-p)^2}\left(\log q + 1 - \log p - \log(1-q) - 1 + \log(1-p) - \frac{2}{q-p}\left(q\log\frac{q}{p} + (1-q)\log\frac{1-q}{1-p}\right)\right) \\ &= \frac{1}{(q-p)^2}\left(\frac{2-q-p}{q-p}\log\frac{1-p}{1-q} - \frac{q+p}{q-p}\log\frac{q}{p}\right) \\ &= \frac{1}{(q-p)^2}\left(H\left(\frac{q-p}{1-p}\right) - H\left(\frac{q-p}{q}\right)\right). \end{aligned}$$

(3) Pour tout $u \in]0,1[$, on a

$$\begin{aligned} H(u) &= \left(1-\frac{2}{u}\right)\left(-u - \frac{u^2}{2} - \frac{u^3}{3} - \dots\right) \\ &= 2 + u^2\left(\frac{2}{3}-\frac{1}{2}\right) + u^3\left(\frac{1}{2}-\frac{1}{3}\right) + \dots + u^k\left(\frac{2}{k+1}-\frac{1}{k}\right) + \dots \end{aligned}$$

Tous les termes de la somme sont strictement croissants, donc H est strictement croissante.

(4) Pour tout $q \in]p,1[$, on a

$$\begin{aligned} g_p'(q) > 0 &\Leftrightarrow H\left(\frac{q-p}{1-p}\right) > H\left(\frac{q-p}{q}\right) \\ &\Leftrightarrow \frac{q-p}{1-p} > \frac{q-p}{q} \\ &\Leftrightarrow q > 1-p\,. \end{aligned}$$

On en déduit que si $0 < p < 1/2$, alors le minimum sur $[p,1]$ de g_p est atteint en $q = 1-p$ et vaut

$$g_p(1-p) = \frac{\log\left(\frac{1-p}{p}\right)}{1-2p},$$

et que si $1/2 \le p < 1$, alors g_p est strictement croissante sur $[p,1]$. Son minimum sur $[p,1]$ est donc atteint en p et vaut

$$g_p(p) = \frac{1}{2p(1-p)}.$$

(5) Soit $P = \mathcal{B}(p)$, et $Q \ll P$. On a $Q = \mathcal{B}(q)$ pour un certain $q \in [0,1]$. Par le lemme de transport (version sous-gaussienne), et comme $\mathbf{E}_Q X - \mathbf{E}_P X = q - p$, il suffit de montrer que

$$q - p \le \sqrt{2v_p \mathrm{D}(Q \mid P)}\,.$$

Si $q \le p$, c'est immédiat. Et si $q \ge p$, cela équivaut à montrer que

$$(q-p)^2 \le 2v_p\left(q\log\frac{q}{p} + (1-q)\log\frac{1-q}{1-p}\right),$$

c'est-à-dire que $g_p(q) \ge \frac{1}{2v_p}$. C'est ce que l'on a montré à la question précédente.

Exercice 4

Pour tout $i \in [\![1,n]\!]$, on définit

$$Z_i = \inf_{x_i \in \mathbb{R}} f(X_1, \dots, X_{i-1}, x_i, X_{i+1}, \dots, X_n)\,.$$

Par un résultat du chapitre sur la méthode entropique, on sait que pour tout $\lambda \ge 0$,

$$\mathbf{Ent}\left[e^{\lambda Z}\right] \le \frac{\lambda^2}{2}\sum_{i=1}^n \mathbf{E}\left[e^{\lambda Z}(Z - Z_i)^2\right].$$

En utilisant les hypothèses de l'énoncé, on sait que

$$\mathbb{1}_{X\in A}\sum_{i=1}^{n}(Z-Z_i)^2 \leq v\,,$$

et

$$\mathbb{1}_{X\notin A}\sum_{i=1}^{n}(Z-Z_i)^2 \leq C\,.$$

Ainsi,

$$\begin{aligned}\mathbf{Ent}\left[e^{\lambda Z}\right] &\leq \frac{\lambda^2}{2}\mathbf{E}\left[e^{\lambda Z}\left(v\mathbb{1}_{X\in A}+C\mathbb{1}_{X\notin A}\right)\right]\\ &\leq \frac{\lambda^2}{2}\left(v\mathbf{E}\left[e^{\lambda Z}\right]+C\mathbf{E}\left[e^{\lambda Z}\mathbb{1}_{X\notin A}\right]\right)\,.\end{aligned}$$

Or, comme $\lambda \geq 0$, comme f est croissante, et comme A est croissant, les fonctions $x \mapsto e^{\lambda f(x)}$ et $x \mapsto \mathbb{1}_{x\in A}$ sont croissantes. Par l'inégalité de Harris, on a donc

$$\mathbf{E}\left[e^{\lambda f(X)}\mathbb{1}_{X\notin A}\right] \leq \mathbf{E}\left[e^{\lambda f(X)}\right]\mathbf{P}(X\notin A)\,.$$

Finalement, on a

$$\mathbf{Ent}\left[e^{\lambda Z}\right] \leq \frac{\lambda^2}{2}\left(v+C\mathbf{P}(X\notin A)\right)\mathbf{E}\left[e^{\lambda Z}\right]\,.$$

On conclut alors par l'argument de Herbst et la méthode de Chernoff.

Exercice 5

(1) On a

$$\sum_{i=1}^{n}\mu_i = \mathsf{tr}(A) = 0\,,$$

puisque les coefficients diagonaux sont nuls. De plus, comme A n'est pas nulle, on a $\mu_1 > 0$.

(2) On écrit $A = {}^tQ\Lambda Q$, où Q est une matrice orthogonale, et Λ la matrice diagonale des valeurs propres. Ainsi $Z = {}^t(QX)\Lambda QX$. Or, par invariance par rotation, le vecteur QX a la même loi que X. Ainsi Z a la même loi que

$${}^tX\Lambda X = \sum_{i=1}^{n}\mu_i X_i^2 = \sum_{i=1}^{n}\mu_i(X_i^2-1)\,,$$

où la dernière égalité vient du fait que $\sum_{i=1}^{n}\mu_i = 0$.

(3) Soit $Y \sim \mathcal{N}(0,1)$. On a $Y^2 \sim \Gamma\left(\frac{1}{2},\frac{1}{2}\right)$. En particulier, pour tout $\lambda < \frac{1}{2}$, on a

$$\log \mathbf{E}e^{\lambda(Y^2-1)} = -\frac{1}{2}\log(1-2\lambda)-\lambda = \frac{1}{2}\left(-\log(1-2\lambda)-2\lambda\right)\,.$$

En utilisant que $-\log(1-u)-u \leq \frac{u^2}{2}$ pour $u \leq 0$, et que $-\log(1-u)-u \leq \frac{u^2}{2(1-u)}$ pour $u \in [0,1[$, on obtient bien les inégalités voulues.

(4) Pour tout $\lambda \in \left[0, \frac{1}{2\mu_1}\right[$, on a

$$
\begin{aligned}
\log \mathbf{E} e^{\lambda Z} &= \sum_{i=1}^{n} \log \mathbf{E} e^{\lambda \mu_i (X_i^2 - 1)} \\
&\leq \sum_{i=1}^{n} \left\{ \mathbb{1}_{\mu_i \leq 0} \mu_i^2 \lambda^2 + \mathbb{1}_{\mu_i > 0} \frac{\mu_i^2 \lambda^2}{1 - 2\mu_i \lambda} \right) \\
&\leq \frac{\lambda^2 \sum_{i=1}^{n} \mu_i^2}{1 - 2\mu_1 \lambda} \cdot
\end{aligned}
$$