Inégalités de concentration
Sorbonne Université
2020-2021

Corrigé

1. Une amélioration de l'inégalité d'Azuma–Hoeffding

(1) Soit $(Y_k)_{k=0}^n$ est une sous-martingale positive et $a > 0$. On introduit le temps d'arrêt $\tau = \inf\{s \geq 0,\, Y_s \geq a\} \wedge n$. Comme τ est presque sûrement borné, le théorème d'arrêt de Doob donne $\mathbf{E}[Y_\tau] \leq \mathbf{E}[Y_n]$. D'autre part, on remarque que

$$Y_\tau \geq \mathbb{1}_{\left\{\max_{0\leq k\leq n} Y_k \geq a\right\}} a\,.$$

En prenant l'espérance, on obtient

$$\mathbf{P}\left(\max_{0\leq k\leq n} Y_k \geq a\right) \leq \frac{\mathbf{E}[Y_\tau]}{a} \leq \frac{\mathbf{E}[Y_n]}{a}\,.$$

(2) Soit $\lambda \in \mathbb{R}$ et $k \in [\![1,n]\!]$. Par l'inégalité de Jensen conditionnelle, on a

$$\mathbf{E}\left[e^{\lambda(Z_k - Z_0)} \;\middle|\; \mathcal{F}_{k-1}\right] \geq e^{\lambda \mathbf{E}\left[Z_k - Z_0 \;\middle|\; \mathcal{F}_{k-1}\right]} = e^{\lambda(Z_{k-1}-Z_0)}\,.$$

La suite $\left(e^{\lambda(Z_k-Z_0)}\right)_{k=0}^n$ est donc bien une sous-martingale.

(3) Pour tout $\lambda \in \mathbb{R}_+$, on a, par ce qui précède,

$$\begin{aligned}\mathbf{P}\left(\max_{0\leq k\leq n}(Z_k - Z_0) \geq t\right) &\leq \mathbf{P}\left(\max_{0\leq k\leq n} e^{\lambda(Z_k-Z_0)} \geq e^{\lambda t}\right)\\ &\leq e^{-\lambda t}\mathbf{E}\left[e^{\lambda(Z_n-Z_0)}\right]\,.\end{aligned}$$

La suite de la preuve est alors exactement celle de l'inégalité d'Azuma–Hoeffding : on obtient

$$\log \mathbf{E}\left[e^{\lambda(Z_n-Z_0)}\right] \leq \frac{\lambda^2}{8}\sum_{i=1}^n (b_i - a_i)^2\,,$$

et l'optimisation en $\lambda \geq 0$ donne

$$\mathbf{P}\left(\max_{0\leq k\leq n}(Z_k - Z_0) \geq t\right) \leq \exp\left(-\frac{2t^2}{\sum_{i=1}^n (b_i-a_i)^2}\right)\,.$$

2. Percolation de premier passage

(1) Par l'inégalité d'Efron–Stein, on a

$$\mathbf{Var}(Z) \leq \mathbf{E}\left[\sum_{e\in E}(Z'_e - Z)_+^2\right]\,,$$

où Z'_e correspond au poids minimal lorsque l'on remplace X_e par une copie indépendante X'_e. Soit $P_\star$ un chemin optimal pour les poids $(X_e)_{e\in E}$. On remarque que si $e \notin P_\star$ alors $Z'_e - Z \leq 0$ (le chemin $P_\star$ garde le même poids donc on peut toujours faire aussi bien que lui). Ainsi,

$$\mathbf{Var}(Z) \leq \mathbf{E}\left[\sum_{e\in P_\star}(Z'_e - Z)_+^2\right] \leq \mathbf{E}\left[\sum_{e\in P_\star}(X'_e)^2\right]\,.$$

Comme X'_e est indépendante de $\mathbb{1}_{e\in P_\star}$, on obtient bien $\mathbf{Var}(Z) \leq \sigma^2\mathbf{E}[L_\star]$.

(2) Remarquons que pour tous $x, y \in [0,1]^n$, on a

$$f(y) - f(x) = \min_{t\in S} \sum_{e\in E} t_e y_e - \min_{t\in S} \sum_{e\in E} t_e x_e \le \sum_{e\in E} t^\star_e(x)(y_e - x_e)\,,$$

où $t^\star(x)$ correspond à un vecteur de S tel que $\sum_{e\in E} t^\star_e(x) x_e = \min_{t\in S} \sum_{e\in E} t_e x_e$. Comme x_e, y_e sont entre 0 et 1, on obtient

$$f(y) - f(x) \le \sum_{e\in E} t^\star_e(x) \mathbb{1}_{y_e \neq x_e}\,.$$

On a

$$\sum_{e\in E} \mathbf{E}[t^\star_e(X)^2] = \sum_{e\in E} \mathbf{E}[\mathbb{1}_{e\in P_\star}] = \mathbf{E}[L_\star]\,,$$

et

$$\sup_{x\in[0,1]^n} \sum_{e\in E} t^\star_e(x)^2 \le d_{u,v}\,.$$

Le résultat vient alors de la proposition 5.6 des notes de cours.

3. Matrice de covariance empirique

(1) (a) On a

$$\|X\,{}^tX - \Sigma\|_{\mathsf{op}} \le \|X\,{}^tX\|_{\mathsf{op}} + \|\Sigma\|_{\mathsf{op}} \le \|X\|^2 + \|\Sigma\|_{\mathsf{op}} \le c^2 + \|\Sigma\|_{\mathsf{op}}\,.$$

Et

$$\begin{aligned}\|\Sigma\|_{\mathsf{op}} = \|\mathbf{E}[X\,{}^tX]\|_{\mathsf{op}} &= \max_{z\in\mathbb{S}^{d-1}} |\,{}^tz\mathbf{E}[X\,{}^tX]z| \\ &= \max_{z\in\mathbb{S}^{d-1}} \mathbf{E}[(\,{}^tzX)^2] \le \max_{z\in\mathbb{S}^{d-1}} \|z\|^2\mathbf{E}[\|X\|^2] \le c^2\,.\end{aligned}$$

Ainsi $\|X\,{}^tX - \Sigma\|_{\mathsf{op}} \le 2c^2$. D'autre part,

$$\mathbf{E}\big[(X\,{}^tX - \Sigma)^2\big] = \mathbf{E}\big[(X\,{}^tX)^2\big] - \Sigma^2 \prec \mathbf{E}\big[(X\,{}^tX)^2\big] = \mathbf{E}\big[\|X\|^2 X\,{}^tX\big] \prec c^2\Sigma\,.$$

Donc (comme il s'agit de matrices semi-définies positives) $\big\|\mathbf{E}\big[(X\,{}^tX - \Sigma)^2\big]\big\|_{\mathsf{op}} \le c^2\|\Sigma\|_{\mathsf{op}}$.

(b) La Proposition 8.3 des notes de cours, appliquée avec $K = 2c^2$ et

$$\sigma^2 = \Big\|\sum_{i=1}^n \mathbf{E}\big[(X_i\,{}^tX_i - \Sigma)^2\big]\Big\|_{\mathsf{op}} \le nc^2\|\Sigma\|_{\mathsf{op}}\,,$$

donne, pour tout $t \ge 0$,

$$\mathbf{P}\left(\Big\|\sum_{i=1}^n (X_i\,{}^tX_i - \Sigma)\Big\|_{\mathsf{op}} \ge t\right) \le 2d\exp\left(-\frac{t^2}{2c^2\left(n\|\Sigma\|_{\mathsf{op}} + \frac{2t}{3}\right)}\right)\,,$$

d'où l'on déduit que

$$\mathbf{P}\left(\|\widehat{\Sigma} - \Sigma\|_{\mathsf{op}} \ge t\right) \le 2d\exp\left(-\frac{nt^2}{2c^2\left(\|\Sigma\|_{\mathsf{op}} + \frac{2t}{3}\right)}\right)\,.$$

(2) (a) Soit $k \ge 1$. En utilisant que Y est centrée et sous-gaussienne, on a

$$\begin{aligned}\mathbf{E}[Y^{2k}] &= \int_0^{+\infty} \mathbf{P}\left(|Y| > t^{1/2k}\right) dt \\ &\le 2\int_0^{+\infty} e^{-\frac{t^{1/k}}{2\sigma^2}}\,dt = 2^{k+1}\sigma^{2k}k\int_0^{+\infty} e^{-u}u^{k-1}du = 2^{k+1}\sigma^{2k}k!\,,\end{aligned}$$

où l'on a effectué le changement de variable $u = \frac{t^{1/k}}{2\sigma^2}$.

Soit maintenant $\lambda \le 0$. En utilisant l'inégalité $e^u \le 1 + u + u^2/2$ pour $u \le 0$, on a

$$\log \mathbf{E} e^{\lambda(Y^2 - \mathbf{E}Y^2)} = -\lambda \mathbf{E}[Y^2] + \log\left(1 + \lambda \mathbf{E}[Y^2] + \frac{\lambda^2 \mathbf{E}[Y^4]}{2}\right) .$$

Et en utilisant $\log(1+u) \le u$ pour $u > -1$, et le fait que $\mathbf{E}[Y^4] \le 16\sigma^4$, on obtient

$$\log \mathbf{E} e^{\lambda(Y^2 - \mathbf{E}Y^2)} \le \frac{\lambda^2 \mathbf{E}[Y^4]}{2} \le 8\sigma^4\lambda^2 .$$

Pour $\lambda \in [0, 1/2\sigma^2[$, on a

$$\begin{aligned}\log \mathbf{E} e^{\lambda(Y^2 - \mathbf{E}Y^2)} &= -\lambda \mathbf{E}[Y^2] + \log\left(1 + \lambda \mathbf{E}[Y^2] + \sum_{k\ge 2} \frac{\lambda^k \mathbf{E}[Y^{2k}]}{k!}\right) \\ &\le \sum_{k\ge 2} \lambda^k 2^{k+1} \sigma^{2k} = \frac{8\sigma^4\lambda^2}{1 - 2\sigma^2\lambda} .\end{aligned}$$

(b) Soit $v \in \mathbb{S}^{d-1}$. On a

$$^t v(\widehat{\Sigma} - \Sigma)v = \frac{1}{n}\sum_{i=1}^n \left(Y_i^2 - \mathbf{E}[Y_i^2]\right) ,$$

avec $Y_i = {}^t v X_i$. Par la question précédente et l'indépendance des Y_i, on a

$$\log \mathbf{E} e^{\lambda \sum_{i=1}^n (Y_i^2 - \mathbf{E}Y_i^2)} \le \begin{cases} 8n\sigma^4\lambda^2 & \text{si } \lambda \le 0, \\ \frac{8n\sigma^4\lambda^2}{1-2\sigma^2\lambda} & \text{si } 0 \le \lambda < \frac{1}{2\sigma^2}. \end{cases}$$

Autrement dit, la variable $\sum_{i=1}^n (Y_i^2 - \mathbf{E}Y_i^2)$ est sous-gaussienne à gauche avec facteur variance $16n\sigma^4$ et sous-Gamma à droite avec facteur variance $16n\sigma^4$ et facteur d'échelle $2\sigma^2$. Par la méthode de Chernoff, on en déduit que

$$\begin{aligned}\mathbf{P}\left(\left|\sum_{i=1}^n (Y_i^2 - \mathbf{E}Y_i^2)\right| \ge t\right) &\le \exp\left(-\frac{t^2}{4\,(8\sigma^4 n + \sigma^2 t)}\right) + \exp\left(-\frac{t^2}{32\sigma^4 n}\right) \\ &\le 2\exp\left(-\frac{t^2}{4\,(8\sigma^4 n + \sigma^2 t)}\right) ,\end{aligned}$$

ce qui donne

$$\mathbf{P}\left(\left|{}^t v(\widehat{\Sigma} - \Sigma)v\right| \ge t\right) \le 2\exp\left(-\frac{nt^2}{4(8\sigma^4 + \sigma^2 t)}\right) .$$

(c) Notons $A = \widehat{\Sigma} - \Sigma$. Soit $z \in \mathbb{S}^{d-1}$ et $u \in S_\delta$ tel que $\|z - u\| \le \delta$. On a

$$\left|{}^t zAz - {}^t uAu\right| = \left|{}^t zA(z-u) - {}^t uA(u-z)\right| \le \left|{}^t zA(z-u)\right| + \left|{}^t uA(u-z)\right| .$$

Or $\left|{}^t zA(z-u)\right| \le \|z\|\|A\|_{\mathsf{op}}\|z-u\| \le \delta\|A\|_{\mathsf{op}}$. De même, $\left|{}^t uA(u-z)\right| \le \delta\|A\|_{\mathsf{op}}$. Ainsi (comme $||a| - |b|| \le |a-b|$), on a

$$\left||{}^t zAz| - |{}^t uAu|\right| \le 2\delta\|A\|_{\mathsf{op}} .$$

On a donc

$$\|A\|_{\mathsf{op}} = \max_{z\in\mathbb{S}^{d-1}} |{}^t zAz| \le \max_{u\in S_\delta} |{}^t uAu| + 2\delta\|A\|_{\mathsf{op}} .$$

(d) Pour $\delta = 1/4$, la question précédente donne $\|\widehat{\Sigma} - \Sigma\|_{\mathsf{op}} \le 2\max_{u\in S} |{}^t u(\widehat{\Sigma} - \Sigma)u|$, où S est un 1/4-recouvrement tel que $|S| \le 9^d$. Par une borne union, on obtient

$$\mathbf{P}\left(\|\widehat{\Sigma} - \Sigma\|_{\mathsf{op}} \ge t\right) \le \sum_{u\in S} \mathbf{P}\left(|{}^t u(\widehat{\Sigma} - \Sigma)u| \ge t/2\right) .$$

En utilisant le résultat de la question 2.(b), on obtient bien

$$\mathbf{P}\left(\|\widehat{\Sigma}-\Sigma\|_{\mathsf{op}} \geq t\right) \leq 2 \cdot 9^d \exp\left(-\frac{nt^2}{8(16\sigma^4+\sigma^2 t)}\right) .$$

4. Méthode entropique et inégalité de Harris

(A) On raisonne par récurrence sur $n \geq 1$. Pour $n = 1$, ce résultat est connu sous le nom d'inégalité d'association de Chebyshev : soient $f, g : \mathbb{R} \to \mathbb{R}$ deux fonctions croissantes et X une variable aléatoire réelle. Soit X' une copie indépendante de X. Comme f et g sont croissantes, on a toujours $(f(X) - f(X'))\,(g(X) - g(X')) \geq 0$. Ainsi

$$\mathbf{E}\left[\left(f(X)-f(X')\right)\left(g(X)-g(X')\right)\right] = 2\left(\mathbf{E}\left[f(X)g(X)\right] - \mathbf{E}\left[f(X)\right]\mathbf{E}\left[g(X)\right]\right) \geq 0\,.$$

Soit maintenant $n \geq 1$ et supposons le résultat vrai pour n. Soient $f, g : \mathbb{R}^{n+1} \to \mathbb{R}$ deux fonctions croissantes et $X = (X_1, \ldots, X_{n+1})$ une suite de variables aléatoires réelles indépendantes. En utilisant l'inégalité d'association de Chebyshev, on a

$$\begin{aligned}\mathbf{E}\left[f(X)g(X)\right] &= \mathbf{E}\left[\mathbf{E}\left[f(X)g(X) \;\middle|\; X_1,\ldots,X_n\right]\right] \\ &\geq \mathbf{E}\left[\mathbf{E}\left[f(X) \;\middle|\; X_1,\ldots,X_n\right]\mathbf{E}\left[g(X) \;\middle|\; X_1,\ldots,X_n\right]\right] ,\end{aligned}$$

car, conditionnellement à $X_1, \ldots, X_n$, f et g sont des fonctions de X_{n+1}. Comme les variables $X_1, \ldots, X_{n+1}$ sont indépendantes, les fonctions $f', g' : \mathbb{R}^n \to \mathbb{R}$ définies par

$$f'(x_1,\ldots,x_n) = \mathbf{E}\left[f(X) \;\middle|\; X_1 = x_1,\ldots,X_n = x_n\right]$$

et

$$g'(x_1,\ldots,x_n) = \mathbf{E}\left[g(X) \;\middle|\; X_1 = x_1,\ldots,X_n = x_n\right]$$

sont des fonctions croissantes en chacune de leurs coordonnées. Ainsi, par l'hypothèse de récurrence, on a

$$\mathbf{E}\left[f'(X_1,\ldots,X_n)g'(X_1,\ldots,X_n)\right] \geq \mathbf{E}\left[f'(X_1,\ldots,X_n)\right]\mathbf{E}\left[g'(X_1,\ldots,X_n)\right] = \mathbf{E}\left[f(X)\right]\mathbf{E}\left[g(X)\right] .$$

(B)

(1) Par la sous-additivité de l'entropie, on a, pour tout $\lambda \in \mathbb{R}$,

$$\mathbf{Ent}\left[e^{\lambda Z}\right] \leq \sum_{i=1}^n \mathbf{E}\left[\mathbf{Ent}^{(i)}[e^{\lambda Z}]\right] .$$

Or, comme X_i' est indépendante de $(X_1, \ldots, X_n)$, l'entropie de $e^{\lambda Z}$ conditionnelle à $X^{(i)}$ est égale à l'entropie de $e^{\lambda Z}$ conditionnelle à $(X^{(i)}, X_i')$. Ainsi, par la caractérisation variationnelle de l'entropie (Proposition 4.5 des notes de cours),

$$\begin{aligned}\mathbf{Ent}^{(i)}[e^{\lambda Z}] &\leq \mathbf{E}\left[e^{\lambda Z}\left(\lambda Z - \lambda Z_i'\right) - \left(e^{\lambda Z} - e^{\lambda Z_i'}\right) \;\middle|\; X^{(i)}, X_i'\right] \\ &= \mathbf{E}\left[e^{\lambda Z}\phi\left(\lambda(Z_i' - Z)\right) \;\middle|\; X^{(i)}, X_i'\right] .\end{aligned}$$

On obtient donc

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

(2) Comme $\phi(0) = 0$, on a

$$\mathbf{E}\left[e^{\lambda Z}\phi\left(\lambda(Z_i'-Z)\right)\right] = \mathbf{E}\left[e^{\lambda Z}\phi\left(\lambda(Z_i'-Z)_+\right)\right] + \mathbf{E}\left[e^{\lambda Z}\phi\left(-\lambda(Z-Z_i')_+\right)\right] .$$

Or par échangeabilité de Z et Z'_i,

$$\begin{aligned}
\mathbf{E}\left[e^{\lambda Z}\phi\left(-\lambda(Z-Z'_i)_+\right)\right] &= \mathbf{E}\left[e^{\lambda Z'_i}\phi\left(-\lambda(Z'_i-Z)_+\right)\right] \\
&= \mathbf{E}\left[e^{\lambda Z}e^{\lambda(Z'_i-Z)}\phi\left(-\lambda(Z'_i-Z)_+\right)\right] \\
&= \mathbf{E}\left[e^{\lambda Z}e^{\lambda(Z'_i-Z)_+}\phi\left(-\lambda(Z'_i-Z)_+\right)\right] ,
\end{aligned}$$

où pour la dernière égalité on a encore utilisé que $\phi(0) = 0$. Ainsi

$$\begin{aligned}
\mathbf{E}\left[e^{\lambda Z}\phi\left(\lambda(Z'_i-Z)\right)\right] &= \mathbf{E}\left[e^{\lambda Z}\left(\phi\left(\lambda(Z'_i-Z)_+\right)+e^{\lambda(Z'_i-Z)_+}\phi\left(-\lambda(Z'_i-Z)_+\right)\right)\right] \\
&= \mathbf{E}\left[e^{\lambda Z}\tau\left(\lambda(Z'_i-Z)_+\right)\right] ,
\end{aligned}$$

car $\phi(u) + e^u\phi(-u) = u(e^u - 1) = \tau(u)$.

(3) On remarque que pour $u \le 0$, on a $\tau(u) \le u^2$. Ainsi, pour $\lambda \le 0$,

$$\begin{aligned}
\mathbf{Ent}\left[e^{\lambda Z}\right] &\le \sum_{i=1}^{n}\mathbf{E}\left[e^{\lambda Z}\tau\left(\lambda(Z'_i-Z)_+\right)\right] \\
&\le \sum_{i=1}^{n}\mathbf{E}\left[\lambda^2 e^{\lambda Z}(Z'_i-Z)_+^2\right] .
\end{aligned}$$

(4) Sous l'hypothèse que $\mathbf{E}\left[\sum_{i=1}^n (Z'_i - Z)_+^2 \bigm| X\right] \le g(X)$, on a, pour tout $\lambda \le 0$,

$$\mathbf{Ent}\left[e^{\lambda Z}\right] \le \lambda^2\mathbf{E}\left[e^{\lambda f(X)}g(X)\right] .$$

Pour $\lambda \le 0$, la fonction $x \mapsto e^{\lambda f(x)}$ est décroissante en chacune de ses coordonnées. Ainsi, par l'inégalité de Harris, on obtient

$$\mathbf{Ent}\left[e^{\lambda Z}\right] \le \lambda^2\mathbf{E}\left[g(X)\right]\mathbf{E}\left[e^{\lambda f(X)}\right] .$$

L'argument de Herbst et la méthode de Chernoff permettent de conclure.