

TD 7 – Inégalités de concentration et intervalles de confiance

1. Inégalités

Exercice 1. (Inégalité de Jensen) Soit X une v.a. réelle telle que $\mathbb{E}[|X|] < \infty$. Soit $\varphi: \mathbb{R} \rightarrow \mathbb{R}_+$ une fonction convexe. Montrer que

$$\mathbb{E}[\varphi(X)] \geq \varphi(\mathbb{E}[X]).$$

Pour cela, on utilisera qu'une fonction convexe est égale au supremum de toutes les fonctions affines qui lui sont inférieures :

$$\forall x \in \mathbb{R}, \quad \varphi(x) = \sup_{(a,b) \in \mathcal{E}_\varphi} (ax + b),$$

où $\mathcal{E}_\varphi = \{(a, b) \in \mathbb{R}^2 : \forall x \in \mathbb{R}, \varphi(x) \geq ax + b\}$.

Exercice 2. Soit $X_1, \dots, X_n \in L^2(\Omega, \mathcal{A}, \mathbb{P})$ des v.a. indépendantes. Soit $\beta > 0$. Pour $k \in \llbracket 1, n \rrbracket$, supposons que $\mathbb{E}[X_k] = 0$ et $\text{Var}(X_k) = k^\beta$.

1. Montrer que, pour tout $\varepsilon > 0$,

$$\mathbb{P}\left(\left|\frac{X_1 + \dots + X_n}{n}\right| \geq \varepsilon\right) \leq \frac{n^{\beta-1}}{\varepsilon^2}.$$

2. Pour quels β peut-on dire que la moyenne empirique $(X_1 + \dots + X_n)/n$ se concentre en 0 quand $n \rightarrow \infty$?

2. Intervalles de confiance

Exercice 3. (Confiance exacte en les lois normales) On observe des v.a. $X_1, \dots, X_n$ i.i.d. dont la loi appartient à la famille $(P_\theta)_{\theta \in \mathbb{R}}$, où $P_\theta = \mathcal{N}(\theta, 1)$. Soit $\alpha \in]0, 1[$.

1. Si $X_1, \dots, X_n$ sont de loi P_θ , montrer que $X_1 + \dots + X_n$ a la même loi que $\sqrt{n}Z + n\theta$ où Z est une v.a. de loi $\mathcal{N}(0, 1)$.
2. En déduire un intervalle de confiance I pour θ de niveau *exactement* $1 - \alpha$, i.e. tel que, si $X_1, \dots, X_n$ ont loi P_θ , alors

$$\mathbb{P}(\theta \in I) = 1 - \alpha.$$

On exprimera cet intervalle en utilisant Φ^{-1} , où Φ est la fonction de répartition de la loi $\mathcal{N}(0, 1)$.

Exercice 4. (Lois uniformes) On observe des v.a. $X_1, \dots, X_n$ i.i.d. dont la loi appartient à la famille $(P_\theta)_{\theta \in [0, 2]}$, où P_θ est la loi uniforme sur $[0, \theta]$. Soit $\alpha \in]0, 1[$.

- 1.(a) Construire un intervalle de confiance pour θ de niveau $1 - \alpha$ en utilisant l'inégalité de Bienaymé–Chebychev.
(b) Construire un intervalle de confiance pour θ de niveau $1 - \alpha$ en utilisant l'inégalité de Hoeffding.
(c) Déterminer pour quelles valeurs de α le second intervalle de confiance est meilleur que le premier.
2. Soit $M_n = \max(X_1, \dots, X_n)$.
(a) Si $X_1, \dots, X_n$ sont de loi P_θ , calculer $\mathbb{P}(M_n \in [\theta - \varepsilon, \theta])$ pour $\varepsilon > 0$.
(b) En déduire un intervalle de confiance pour θ de niveau $1 - \alpha$.
(c) Le comparer aux intervalles de confiance précédents dans le régime où α est fixé et $n \rightarrow \infty$.

3. Compléments

Exercice 5. (Concentration avec un moment d'ordre 4)

1. Soit $X \in L^4(\Omega, \mathcal{A}, \mathbb{P})$. Montrer que, pour tout $a > 0$,

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq a) \leq \frac{\mathbb{E}[(X - \mathbb{E}[X])^4]}{a^4}.$$

2. Soit $Y_1, \dots, Y_n$ des v.a. réelles i.i.d. telles que $\mathbb{E}[Y_1^4] < \infty$ et $\mathbb{E}[Y_1] = 0$. Montrer que

$$\mathbb{E}[(Y_1 + \dots + Y_n)^4] = 3n(n-1)\mathbb{E}[Y_1^2]^2 + n\mathbb{E}[Y_1^4].$$

3. Soit $X_1, \dots, X_n$ des v.a. réelles i.i.d. telles que $\mathbb{E}[X_1^4] < \infty$. Montrer que, pour tout $\varepsilon > 0$,

$$\mathbb{P}\left(\left|\frac{X_1 + \dots + X_n}{n} - \mathbb{E}[X_1]\right| \geq \varepsilon\right) \leq \frac{1}{n^2\varepsilon^4} \left(3\text{Var}(X_1) + \frac{1}{n}\mathbb{E}[(X_1 - \mathbb{E}[X_1])^4]\right).$$

Exercice 6. (Lois normales de variance inconnue) On observe des v.a. $X_1, \dots, X_n$ i.i.d. dont la loi appartient à la famille $(P_\theta)_{\theta \in \Theta}$, où $P_\theta = \mathcal{N}(0, \theta^2)$. Soit $\alpha \in]0, 1[$.

1. On suppose que $\Theta =]0, \sigma_{\max}]$, où $\sigma_{\max} > 0$ est un paramètre connu.

- (a) En considérant

$$\frac{X_1^2 + \dots + X_n^2}{n}$$

et en utilisant l'inégalité de Bienaymé–Chebychev, construire un intervalle de confiance pour θ^2 de niveau $1 - \alpha$.

- (b) En déduire un intervalle de confiance pour θ de niveau $1 - \alpha$.

2. On ne suppose plus que l'on connaît a priori une borne supérieure pour le paramètre θ . On considère donc $\Theta = \mathbb{R}_+^*$.

- (a) Montrer que, pour tout $\varepsilon \in]0, 1[$,

$$\mathbb{P}\left(\left|\frac{X_1^2 + \dots + X_n^2}{n} - \theta^2\right| \geq \frac{X_1^2 + \dots + X_n^2}{n}\varepsilon\right) \leq \mathbb{P}\left(\left|\frac{X_1^2 + \dots + X_n^2}{n} - \theta^2\right| \geq \frac{\theta^2\varepsilon}{1 - \varepsilon}\right).$$

- (b) En déduire un intervalle de confiance de niveau $1 - \alpha$ pour θ^2 puis pour θ .

Exercice 7. Soit $X_1, \dots, X_n$ des v.a. i.i.d. de loi exponentielle de paramètre $\lambda > 0$.

1. Calculer $\mathbb{E}[X_1]$ et $\mathbb{E}[e^{tX_1}]$ pour tout $t \in \mathbb{R}$.

2. Soit $S = X_1 + \dots + X_n$. En appliquant l'inégalité de Markov à e^{tS} pour $t \geq 0$, puis en optimisant en t , montrer que, pour tout $\varepsilon > 0$,

$$\mathbb{P}\left(S \geq \frac{n}{\lambda} + n\varepsilon\right) \leq \exp(-n(\lambda\varepsilon - \log(1 + \lambda\varepsilon))).$$

3. En procédant de même mais avec e^{-tS} pour $t \geq 0$, montrer que, pour tout $\varepsilon \in]0, 1/\lambda[$,

$$\mathbb{P}\left(S \leq \frac{n}{\lambda} - n\varepsilon\right) \leq \exp(n(\lambda\varepsilon + \log(1 - \lambda\varepsilon))).$$

4. En déduire que pour tout $\varepsilon > 0$,

$$\mathbb{P}\left(\left|\frac{X_1 + \dots + X_n}{n} - \mathbb{E}[X_1]\right| > \varepsilon\right) \leq 2\exp(-n(\lambda\varepsilon - \log(1 + \lambda\varepsilon))).$$