Rappels

On rappelle les éléments suivants du cours de Probabilités, considérés comme acquis par la suite:

- **Loi (forte) des grands nombres**: Si $X_1, \dots, X_n$ sont des v.a. i.i.d. sur $\mathbb{R}$ satisfaisant $\mathbb{E}|X_1| < +\infty$, alors
$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow[n\to+\infty]{\text{p.s.}} \mathbb{E}(X_1).$$
- **Théorème central limite**: Si $X_1, \dots, X_n$ sont des v.a. i.i.d. sur $\mathbb{R}$, satisfaisant $\mathbb{E}|X_1|^2 < +\infty$, alors
$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} (X_i - \mathbb{E}(X_i)) \underset{n\to+\infty}{\rightsquigarrow} \mathcal{N}(0, \mathrm{Var}(X_1)).$$
  On rappelle aussi les définitions équivalentes de la convergence en loi (notée $\rightsquigarrow$): $X_n \rightsquigarrow X$ ssi, de manière équivalente
  1. $\forall f \in C_b(\mathbb{R})$ $\mathbb{E}(f(X_n)) \to \mathbb{E}(f(X))$.
  2. Si F_X est continue en t, $F_{X_n}(t) \to F_X(t)$ (fonctions de répartition).
  3. Pour tout $t \in \mathbb{R}$, $\varphi_{X_n}(t) \to \varphi_X(t)$ (fonctions caractéristiques).
- **Inégalité de Markov**: Si X est positive, alors, pour tout $t > 0$,
$$\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}(X)}{t}.$$

Exercices: outils de base en concentration

Exercice 1 (Autour de Byenaymé Tchebychev). — Soit X une variable réelle satisfaisant $\mathbb{E}(X^2) < +\infty$.

1. Montrer l'inégalité de Byenaymé Tchebychev: $\forall t > 0 \quad \mathbb{P}(|X - \mathbb{E}(X)| \geq t) \leq \frac{\mathrm{Var}(X)}{t^2}$.
2. Montrer l'inégalité de Cantelli: $\forall t > 0 \quad \mathbb{P}(X \geq \mathbb{E}(X) + t) \leq \frac{\mathrm{Var}(X)}{\mathrm{Var}(X)+t^2}$.
3. Comparer ces deux inégalités.
4. Montrer l'inégalité de Paley-Zygmund: en supposant de plus $X \geq 0$,
$$\forall t \in [0,1] \quad \mathbb{P}(X \geq t\mathbb{E}(X)) \geq (1-t)^2\frac{\mathbb{E}(X)^2}{\mathbb{E}(X^2)}.$$

Exercice 2 (Inégalité de Hoeffding). — On va commencer par prouver le Lemme de Hoeffding: si Z est une v.a. telle que $Z \in [a,b]$ p.s. et $\mathbb{E}(Z) = 0$, alors
$$\forall t \in \mathbb{R} \quad \psi_Z(t) \leq \frac{(b-a)^2 t^2}{8},$$
où $\psi_Z(t) := \log(\mathbb{E}(\exp(tZ)))$. Dans la suite on se donne une telle v.a. Z.

1. Montrer que ψ_Z est bien définie.

2. Montrer que, si Y est à valeurs dans $[a, b]$, alors
$$\mathrm{Var}(Y) \leq \frac{(b-a)^2}{4}.$$
3. En déduire que, pour tout $t \in \mathbb{R}$, $\psi_Z''(t) \leq (b-a)^2/4$.
4. Conclure.

Dans une deuxième partie on va prouver l'inégalité de Hoeffding. Soient $X_1, \ldots, X_n$ des variables aléatoires indépendantes telles que $X_i \in [a_i, b_i]$ p.s. et $\mathbb{E}(X_i) = 0$. On note $S = \sum_{i=1}^n X_i$.

5. Montrer que, pour tout $\varepsilon \geq 0$ et $t \in \mathbb{R}$,
$$\mathbb{P}\left(S \geq \varepsilon\right) \leq \exp\left(-t\varepsilon + \psi_S(t)\right).$$
6. En choisissant soigneusement t, montrer que, pour tout $\varepsilon \geq 0$,
$$\mathbb{P}\left(S \geq \varepsilon\right) \leq \exp\left(-\frac{2\varepsilon^2}{\sum_{i=1}^n (b_i - a_i)^2}\right),$$
ce qui constitue l'inégalité de Hoeffding.
7. Application: si $X_1, \ldots, X_n$ sont i.i.d. de loi $\mathcal{B}(p)$, montrer que, pour tout $\varepsilon \geq 0$,
$$\mathbb{P}\left(|\bar{X}_n - p| \geq \varepsilon\right) \leq 2\exp\left(-2n\varepsilon^2\right).$$

Exercice: un résultat utile de convergence

Exercice 3 (Convergence des quantiles). — Soit X v.a. sur $\mathbb{R}$ de fonction de répartition F. Pour $p \in [0, 1]$, le **quantile** d'ordre p de X est défini par
$$q(p) = \inf\left\{t \in \mathbb{R} \mid F(t) \geq p\right\}.$$
Soient $X_1, \ldots, X_n$ i.i.d. de même loi que X. On se donne $p \in [0, 1]$. On note q_n le quantile empirique $q_n(p)$ (associé à la fonction de répartition empirique $F_n(t) = \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{X_i \leq t}$). On suppose que F est strictement croissante en $q(p)$, que l'on abrègera en q.

1. Soit $u \in \mathbb{R}$. Montrer que
$$q \leq u \Leftrightarrow F(u) \geq p.$$
2. Soit $\varepsilon > 0$. Montrer que
$$\mathbb{P}(q_n \leq q - \varepsilon) \leq \mathbb{P}\left(\sum_{i=1}^n \mathbb{1}_{X_i \leq q-\varepsilon} \geq np\right).$$
3. En déduire l'existence de $\delta_-(\varepsilon) > 0$ tel que
$$\mathbb{P}(q_n \leq q - \varepsilon) \leq \exp(-2n\delta_-(\varepsilon)^2).$$
4. Soit $\varepsilon > 0$. Montrer que
$$\mathbb{P}(q_n > q + \varepsilon) \leq \mathbb{P}\left(\sum_{i=1}^n \mathbb{1}_{X_i \leq q+\varepsilon} < np\right).$$
5. En déduire l'existence de $\delta_+(\varepsilon) > 0$ tel que
$$\mathbb{P}(q_n > q + \varepsilon) \leq \exp(-2n\delta_+(\varepsilon)^2).$$
6. En déduire que $q_n \xrightarrow[n\to+\infty]{\text{p.s.}} q$.