TD 3

Les exercices marqués d'une étoile ne sont pas prioritaires.

3.1 Loi du 0-1

Exercice 1. Soit X une variable aléatoire de loi $\mathcal{N}(m, \sigma^2)$. Comparer les probabilités $\mathbb{P}(|X-m| \geq c\sigma)$ lues dans la table et les majorations obtenues par l'inégalité de Tchebychev pour

$$c = 0.5, \quad c = 1, \quad c = 1.5, \quad c = 2, \quad c = 2.5.$$

On a $\mathbb{P}(|X-m| \geq c\sigma) = \mathbb{P}(|Y| \geq c)$ avec $Y \sim \mathcal{N}(0,1)$, d'où

$$\mathbb{P}(|X-m| \geq c\sigma) = \mathbb{P}(Y \geq c) + \mathbb{P}(Y \leq -c) = 2\mathbb{P}(Y \geq c) = 2 - 2\mathbb{P}(Y \leq c),$$

qu'on calcule par lecture dans la table. Par ailleurs l'inégalité de Tchébychev donne

$$\mathbb{P}(|X-m| \geq c\sigma) \leq \sigma^2/(c\sigma)^2 = c^{-2}.$$

On obtient

c	0.5	1	1.5	2	2.5
$\mathbb{P}(\vert X-m\vert \geq c\sigma)$	0.6170	0.3174	0.1336	0.0456	0.0124
c^{-2}	4	1	0.4444	0.25	0.16

Exercice 2.* Soit $(X_n)_{n\in\mathbb{N}^*}$ une suite de variables aléatoires réelles indépendantes, $\mathcal{F}_n = \sigma(X_n)$, et soit $(a_n)_{n\in\mathbb{N}^*}$ une suite de réels strictement positifs tels que $\lim a_n = +\infty$. Montrer que l'ensemble des $\omega \in \Omega$ pour lesquels la suite $u_n(\omega) = \frac{1}{a_n}\sum_{k=1}^n X_k(\omega)$ converge est dans la tribu terminale de la suite $(\mathcal{F}_n)_{n\in\mathbb{N}}$. En déduire que c'est un événement certain ou impossible.

Soit $\mathcal{F}_n = \sigma(X_n)$ la tribu engendrée par X_n. La famille de tribus $(\mathcal{F}_n)$ est indépendante puisque les (X_n) le sont. Remarquons que $u_n(\omega) = \frac{1}{a_n}\sum_{k=1}^n X_k(\omega) = \frac{1}{a_n}\sum_{k=1}^{j-1} X_k(\omega) + \frac{1}{a_n}\sum_{k=j}^n X_k(\omega)$ pour tout j fixé. Comme la suite (a_n) tend vers l'infini, à j fixé on a donc u_n converge ssi $\frac{1}{a_n}\sum_{k=j}^n X_k(\omega)$ converge. L'événement $A = \{\omega \in \Omega; u_n(\omega) \text{ converge }\}$ peut donc se récrire pour tout j comme

$$A = \{\omega \in \Omega; \frac{1}{a_n}\sum_{k=j}^n X_k(\omega) \text{ converge }\}$$

A est donc un événement de $\overline{\mathcal{F}}_j = \sigma(\mathcal{F}_k, k \geq j)$ pour tout j, donc de $\cap \overline{\mathcal{F}}_j = \overline{\mathcal{F}}_\infty$ la tribu terminale de la famille $(\mathcal{F}_n)$. Par la loi du $0-1$, A est donc un événement de probabilité 0 ou 1.

Exercice 3. Soit $(\Omega, \mathcal{F}, \mathbb{P}) = ([0,1], \mathcal{B}([0,1]), \lambda)$ où λ est la mesure de Lebesgue sur $[0,1]$. Pour tout entier $n \geq 1$, on pose $A_n =]0, 1/n]$.

1. Expliciter l'événement (A_n is).

$(A_n \text{ is}) = \bigcap_{n\in\mathbb{N}}\bigcup_{m\geq n} A_m = \bigcap_{n\in\mathbb{N}}\bigcup_{m\geq n}]0, 1/m] = \bigcap_{n\in\mathbb{N}}]0, 1/n] = \emptyset$

2. Calculer $\sum_n \mathbb{P}(A_n)$.

   $\sum_n \mathbb{P}(A_n) = \sum_n 1/n = \infty$.

3. Qu'en pensez-vous ?

   On a $\sum_n \mathbb{P}(A_n) = \infty$ mais $\mathbb{P}(A_n \text{ is}) = 0$. On voit que le réciproque du lemme de Borel Cantelli est fausse sans l'hypothèse d'indépendance des (A_n).

Exercice 4.* Soit (x_n) une suite (déterministe) de réels. On définit sa limite supérieure et sa limite inférieure respectivement par

$$\begin{aligned} \limsup_{n\to\infty} x_n &= \lim_{m\to\infty} \sup_{n\geq m} x_n, \\ \liminf_{n\to\infty} x_n &= \lim_{m\to\infty} \inf_{n\geq m} x_n. \end{aligned}$$

1. Montrer que $\limsup x_n$ est bien définie dans $\mathbb{R}\cup\{\pm\infty\}$ et $\liminf x_n$ est bien définie dans $\mathbb{R}\cup\{\pm\infty\}$ pour toute suite réelle (x_n).

   La suite $(\sup_{n\geq m} x_n)$ est décroissante et à valeurs dans $\mathbb{R}\cup\{\pm\infty\}$, la suite $(\inf_{n\geq m} x_n)$ est croissante et à valeurs dans $\mathbb{R}\cup\{\pm\infty\}$, donc elles convergent dans $\mathbb{R}\cup\{\pm\infty\}$.

2. Montrer que la suite (x_n) converge dans $\mathbb{R}\cup\{\pm\infty\}$ si et seulement si on a $\limsup x_n \leq \liminf x_n$.

   Si x_n converge vers x, alors $\sup_{n\geq m} x_n$ et $\inf_{n\geq m} x_n$ convergent également vers x, donc $\limsup x_n = \liminf x_n$.
   Réciproquement, si $\limsup x_n \leq \liminf x_n$ alors nécessairement ces deux quantités sont égales et à valeurs dans $\mathbb{R}\cup\{\pm\infty\}$. Soit x leur valeur commune.
   Supposons que $x\in\mathbb{R}$. On a donc que pour tout $\epsilon>0$ il existe un rang M_1 et un rang M_2 tels que

   $$\begin{aligned} 0 \leq \quad & x - \sup_{n\geq M_1} x_n & \leq \epsilon \\ 0 \leq \quad & \inf_{n\geq M_2} x_n - x & \leq \epsilon, \end{aligned}$$

   autrement dit, pour tout $n \geq \max\{M_1, M_2\}$, on a $|x_n - x| \leq \epsilon$, donc la suite (x_n) converge vers x.
   Supposons que $x = +\infty$. Alors pour tout $A > 0$ il existe un rang M_1 et un rang M_2 tels que

   $$\begin{aligned} \sup_{n\geq M_1} x_n &\geq A \\ \inf_{n\geq M_2} x_n &\geq A, \end{aligned}$$

   autrement dit, pour tout $n \geq \max\{M_1, M_2\}$, on a $x_n \geq A$, donc la suite (x_n) converge vers $+\infty$. On fait de même avec $-\infty$.

3.2 Convergence presque sure et en probabilité

Exercice 5. Soit $(X_n)_{n\in\mathbb{N}^*}$ une suite de variables aléatoires réelles et M un réel.

1. Montrer que

$$(\limsup_{n\to\infty} X_n < M) \subset \bigcup_m \bigcap_{n\geq m} (X_n < M) \subset (\limsup_{n\to\infty} X_n \leq M).$$

Par définition, on a $(\limsup_{n\to\infty} X_n < M) = (\lim_m \sup_{n\geq m} X_n < M)$.
Or $(\sup_{n\geq m} X_n)$ est une suite décroissante, donc si sa limite est $< M$ alors la suite est $< M$ à partir d'un certain rang. On a donc
$(\limsup_{n\to\infty} X_n < M) \Rightarrow \exists m;\ \forall n \geq m\ (X_n < M)$.
Réciproquement, si $\exists m;\ \forall n \geq m\ (X_n < M)$, alors $\sup_{n\geq m} X_n \leq M$ et par décroissance de la suite des sup, on obtient $(\limsup_{n\to\infty} X_n \leq M)$.

2. Montrer que s'il existe un réel M pour lequel $\sum \mathbb{P}(X_n \geq M) < \infty$, alors on a $\limsup_{n\to\infty} X_n \leq M$ presque sûrement.

Soit $A_n = (X_n \geq M)$. Par le lemme de Borel Cantelli, $\sum \mathbb{P}(X_n \geq M) < \infty$ implique que $\mathbb{P}(A_n \text{ is}) = 0$ ce qui se traduit par passage au complémentaire par $\bigcup_{m\in\mathbb{N}} \bigcap_{n\geq m} A_n^c$ a lieu presque sûrement, autrement dit ps, $\exists m$, $\forall n \geq m$, $(X_n < M)$. Comme cet événement est inclus dans $(\limsup_{n\to\infty} X_n \leq M)$, on a $(\limsup_{n\to\infty} X_n \leq M)$ presque sûrement.

3. En déduire que si $(Y_n)_{n\geq 1}$ et Y sont des variables aléatoires à valeurs dans $\mathbb{R}^d$, et que pour tout $\epsilon > 0$ on a $\sum \mathbb{P}(|Y_n - Y| > \epsilon) < \infty$ alors (Y_n) converge presque sûrement vers Y.

On pose $X_n = |Y_n - Y|$ qui est bien une suite réelle. Comme elle est positive, il est clair que X_n tend vers 0 ps ssi $\limsup X_n \leq \epsilon$ pour tout $\epsilon > 0$. Il suffit donc d'appliquer la question 2.

4. Montrer que si $\limsup_{n\to\infty} X_n < M$ presque sûrement et que les variables aléatoires $(X_n)_{n\in\mathbb{N}^*}$ sont indépendantes alors on a $\sum \mathbb{P}(X_n \geq M) < \infty$.

Supposons que $\limsup_{n\to\infty} X_n < M$ presque sûrement et $\sum \mathbb{P}(X_n \geq M) = \infty$. Soit $A_n = (X_n \geq M)$. Par le lemme de Borel Cantelli, $\sum \mathbb{P}(X_n \geq M) = \infty$ implique que $\mathbb{P}(A_n \text{ is}) = 1$ puisqu'on a indépendance, ce qui se traduit par passage au complémentaire par $\bigcup_{m\in\mathbb{N}} \bigcap_{n\geq m} A_n^c$ est impossible, ce qui est une contradiction. Donc $\sum \mathbb{P}(X_n \geq M) < \infty$.

Exercice 6.* Soit $(X_n)_{n\in\mathbb{N}^*}$ une suite de variables aléatoires réelles indépendantes et de même loi.

1. Montrer que $\int_\Omega |X_1| d\mathbb{P} = \int_{\mathbb{R}_+} \mathbb{P}(|X_1| > t)dt$.

On a déjà vu ce résultat avec la fonction de survie d'une variable positive. Par le théorème de Fubini Tonelli, on a

$$\begin{aligned}
\int_{\mathbb{R}_+} \mathbb{P}(|X_1] > t)dt &= \int_{\mathbb{R}_+} \mathbb{E}[\mathbb{1}_{|X_1]>t}]dt \\
&= \mathbb{E}\Big[\int_{\mathbb{R}_+} \mathbb{1}_{|X_1|>t} dt\Big] \\
&= \mathbb{E}[|X_1|].
\end{aligned}$$

2. Montrer que pour tout $\epsilon > 0$ on a

$$\epsilon \sum_{n=0}^{\infty} \mathbb{P}(|X_1| > (n+1)\epsilon) \leq \sum_{n=0}^{\infty} \int_{n\epsilon}^{(n+1)\epsilon} \mathbb{P}(|X_1| > t)dt \leq \epsilon \sum_{n=0}^{\infty} \mathbb{P}(|X_1| > n\epsilon).$$

Comme la fonction de survie est décroissante, on a directement les inégalités voulues : $\int_{n\epsilon}^{(n+1)\epsilon} \mathbb{P}(|X_1| > (n+1)\epsilon)dt \leq \int_{n\epsilon}^{(n+1)\epsilon} \mathbb{P}(|X_1| > t)dt \leq \int_{n\epsilon}^{(n+1)\epsilon} \mathbb{P}(|X_1| > n\epsilon)dt$ donc $\epsilon \sum_{n=0}^{\infty} \mathbb{P}(|X_1| > (n+1)\epsilon) \leq \sum_{n=0}^{\infty} \int_{n\epsilon}^{(n+1)\epsilon} \mathbb{P}(|X_1| > t)dt \leq \epsilon \sum_{n=0}^{\infty} \mathbb{P}(|X_1| > n\epsilon)$.

3. En déduire que pour tout $\epsilon > 0$ on a $\mathbb{E}[|X_1|] < \infty \iff \sum_{n=1}^{\infty} \mathbb{P}(|X_1| > n\epsilon) < \infty$.

   On a $\mathbb{E}[|X_1|] = \int_{\mathbb{R}_+} \mathbb{P}(|X_1| > t)dt = \sum_{n=0}^{\infty} \int_{n\epsilon}^{(n+1)\epsilon} \mathbb{P}(|X_1| > t)dt$. Donc par ce qui précède $\mathbb{E}[|X_1|] < \infty \iff \sum_{n=1}^{\infty} \mathbb{P}(|X_1] > n\epsilon) < \infty$.

4. En utilisant l'exercice 4, montrer que $\mathbb{P}\left(\frac{X_n}{n} \to 0\right) = 1$ implique $\mathbb{E}[|X_1|] < \infty$.

   Si on a $\mathbb{P}\left(\frac{X_n}{n} \to 0\right) = 1$, alors $\limsup X_n/n = 0$ ps, donc $\mathbb{P}(\limsup |X_n|/n > 1) = 0$, ou bien $\mathbb{P}(\limsup |X_n|/n \leq 1) = 1$, et par l'exercice précédent (puisqu'on a indépendance) $\sum \mathbb{P}(|X_n| > n) = \mathbb{E}[|X_1|] < \infty$

5. Montrer de la même façon que $\mathbb{E}[|X_1|] < \infty$ implique $\mathbb{P}\left(\frac{X_n}{n} \to 0\right) = 1$.

   Si $\mathbb{E}[|X_1|] < \infty$, on a pour tout $\epsilon > 0$
   $\mathbb{E}[|X_1|] = \sum_{n=0}^{\infty} \int_{n\epsilon}^{(n+1)\epsilon} \mathbb{P}(|X_1| > t)dt \geq \epsilon \sum_{n=0}^{\infty} \mathbb{P}(|X_1| > (n+1)\epsilon)$.
   Par identité en loi, on a $\sum_{n=0}^{\infty} \mathbb{P}(|X_n| > (n+1)\epsilon) < \infty$ et le lemme de Borel Cantelli donne $\mathbb{P}(|X_n| > (n+1)\epsilon \text{ is}) = 0$ pour tout $\epsilon > 0$. Comme une union d'ensembles de mesure nulle est de mesure nulle, on a $\mathbb{P}(\cup_{\epsilon \in \mathbb{Q}_+}(|X_n| > (n+1)\epsilon \text{ is})) = 0$. Or $\mathbb{P}\left(\frac{X_n}{n} \to 0\right) \geq \mathbb{P}(\cap_{\epsilon \in \mathbb{Q}_+} \cup_n \cap m \geq n(|X_m|/(m+1) > \epsilon \text{ is})) = 1 - \mathbb{P}(\cup_{\epsilon \in \mathbb{Q}_+}(|X_n| > (n+1)\epsilon \text{ is})) = 1$.

Exercice 7. Soit (X_n) une suite de variables aléatoires indépendantes de loi exponentielle de même paramètre 1, et $M_n = \max_{k \leq n}\{X_k\}$.

1. Calculer la fonction de répartition de M_n.

   $\mathbb{P}(M_n \leq t) = \mathbb{P}(X_1 \leq t, \cdots X_n \leq t) = (1 - \exp(-t))^n$ par indépendance.

2. Montrer que la série $\sum_n \mathbb{P}(M_n/\log n \leq 1 - \epsilon)$ converge pour tout $0 < \epsilon < 1$.

   On a
   $\mathbb{P}(M_n \leq (1-\epsilon)\log n) = (1 - n^{-1+\epsilon})^n = \exp(n \log(1 - n^{-1+\epsilon})) = \exp(-n^{\epsilon}(1 + o(1)))$, donc la série converge.

3. En déduire que $\liminf M_n/\log n \geq 1 - \epsilon$ presque sûrement.

   Le lemme de Borel Cantelli donne $\mathbb{P}(M_n/\log n \leq 1 - \epsilon \text{ is}) = 0$ ce qui se traduit par passage au complémentaire par $\bigcup_{n \in \mathbb{N}} \bigcap_{m \geq n}(M_m/\log m \leq 1 - \epsilon)^c$ a lieu presque sûrement, autrement dit ps, $\exists n, \forall m \geq n, (M_m/\log m > 1 - \epsilon)$, ce qui est exactement la définition de $\liminf M_n/\log n \geq 1 - \epsilon$ presque sûrement.

4. Montrer que $\mathbb{P}(M_n > (1+\epsilon)\log n) = n^{-\epsilon}(1 + o(1))$.

   On a
   $\mathbb{P}(M_n > (1+\epsilon)\log n) = 1 - \mathbb{P}(M_n \leq (1+\epsilon)\log n) = 1 - \exp(-n^{\epsilon}(1 + o(1))) = n^{-\epsilon}(1 + o(1))$.

5. En déduire que pour la sous-suite (n_k) définie par $n_k = [(k+1)^{\delta}]$ pour $\delta > 1/\epsilon$, $\limsup M_{n_k}/\log n_k \leq 1 + \epsilon$ presque sûrement.

   Pour cette sous-suite, $\mathbb{P}(M_{n_k} > (1+\epsilon)\log n_k) = n_k^{-\epsilon}(1+o(1)) \leq (k+1)^{-\delta\epsilon}$, terme générique d'une suite sommable. Par le lemme de Borel Cantelli, on obtient donc $\limsup M_{n_k}/\log n_k \leq 1 + \epsilon$ presque sûrement.

6. Utiliser la croissance de la suite (M_n) pour en déduire que $\limsup M_n/\log n \leq 1+\epsilon$ presque sûrement.

   (M_n) est une suite croissante par construction. Pour tout n, il existe k tel que $n_k \leq n < n_{k+1}$, d'où $\frac{M_n}{\log n} \leq \frac{M_{n_{k+1}}}{\log n_{k+1}} \frac{\log n_{k+1}}{\log n_k}$. La limite de la dernière fraction est 1, on a donc le résultat voulu.

7. En déduire que $M_n/\log n$ converge vers 1 presque sûrement.

   On a
   $1-\epsilon \leq \liminf \frac{M_n}{\log n} \leq \limsup \frac{M_n}{\log n} \leq 1+\epsilon$ ps. En faisant tendre ϵ vers 0 le long s'une suite dénombrable (pour garder un ensemble de mesure pleine), on obtient le résultat voulu.

Exercice 8. Soit (X_n) une suite de variables aléatoires indépendantes. Montrer que X_n tend vers 0 presque sûrement si et seulement si $\sum \mathbb{P}(|X_n| \geq \epsilon) < \infty$ pour tout $\epsilon > 0$.

Supposons que X_n tend vers 0 presque sûrement. Alors, pour presque tout ω, pour tout $\epsilon > 0$ il existe n_0 tel que pour tout $n \geq n_0$, $|X_n(\omega)| < \epsilon$. Autrement dit, $\mathbb{P}(\cap_{\epsilon\in\mathbb{Q}_+^*} \cup_{n_0} \cap_{n\geq n_0} |X_n| < \epsilon) = 1$. On en déduit que pour tout $\epsilon > 0$ (rationnel, puis quelconque par croissance de la probabilité), $\mathbb{P}(\cup_{n_0} \cap_{n\geq n_0} |X_n| < \epsilon) = 1$. Par passage au complémentaire, on a $\mathbb{P}(\cap_{n_0} \cup_{n\geq n_0} |X_n| \geq \epsilon) = 0 = \mathbb{P}(|X_n| \geq \epsilon \text{ is})$. Le lemme de Borel Cantelli réciproque (puisqu'on a indépendance) donne alors $\sum \mathbb{P}(|X_n| \geq \epsilon) < \infty$.

Si on suppose $\sum \mathbb{P}(|X_n| \geq \epsilon) < \infty$, on applique le lemme de Borel Cantelli pour obtenir $\mathbb{P}(\cap_{n_0} \cup_{n\geq n_0} |X_n| < \epsilon) = 0 = \mathbb{P}(|X_n| < \epsilon \text{ is})$ et remonter à $\mathbb{P}(\cap_{\epsilon\in\mathbb{Q}_+^*} \cup_{n_0} \cap_{n\geq n_0} |X_n| < \epsilon) = 1$, qui est exactement X_n tend vers 0 presque sûrement.

Exercice 9. Soit (X_n) une suite de variables aléatoires indépendantes de loi de Bernoulli de paramètre respectif p_n.

1. Montrer que la suite (X_n) converge en probabilité vers 0 si et seulement si $\lim p_n = 0$.

   Soit $0 < \epsilon < 1$ on a $\mathbb{P}(|X_n| \geq \epsilon) = \mathbb{P}(X_n = 1) = p_n$. Donc la suite (X_n) converge en probabilité vers 0 si et seulement si $\lim p_n = 0$.

2. Montrer que la suite (X_n) converge presque sûrement vers 0 si et seulement si $\sum p_n < \infty$.

   C'est une application de l'exercice précédent : X_n tend vers 0 presque sûrement si et seulement si pour tout $\epsilon > 0$, $\sum \mathbb{P}(|X_n| \geq \epsilon) = \sum p_n < \infty$.

Exercice 10.* Il n'y a pas de lemme de Cesaro pour la convergence en probabilité Soit (X_n) une suite de variables aléatoires réelles indépendantes, avec X_n de fonction de répartition

$$F_n(t) = (1-(t+n)^{-1})\mathbb{1}_{[0,+\infty[}(t).$$

On note $S_n = \sum_{k=1}^n X_k$, $M_n = \max\{X_1, \cdots, X_n\}$ et $Y_n = S_n/n$.

1. Vérifier que F_n est bien une fonction de répartition.

   F_n est à valeurs dans $[0,1]$, croissante, continue à droite avec limite à gauche (en fait continue partout sauf en 0), nulle sur $\mathbb{R}_-$ donc avec une limite nulle en $-\infty$ et avec limite 1 en $+\infty$. C'est donc bien une fonction de répartition.

2. Montrer que la suite (X_n) converge vers 0 en probabilité.

   Pour tout $\epsilon > 0$, on a $\mathbb{P}(|X_n| > \epsilon) = 1 - F_n(\epsilon) = (\epsilon+n)^{-1}$ donc $\lim \mathbb{P}(|X_n| > \epsilon) = 0$ et la suite converge en probabilité vers 0.

3. Calculer la fonction de répartition de M_n.

   $\mathbb{P}(M_n \leq t) = \mathbb{P}(X_1 \leq t, \cdots, X_n \leq t) = \prod_{k=1}^n (1-(t+k)^{-1})$ par indépendance.

4. Montrer que $\mathbb{P}(Y_n > \epsilon) \geq \mathbb{P}(M_n > n\epsilon)$.

   $\mathbb{P}(M_n > n\epsilon) = \mathbb{P}(\max\{X_i\}/n > \epsilon) \leq \mathbb{P}(Y_n > \epsilon)$ car $\max \leq \sum$.

5. En déduire que la suite (Y_n) ne converge pas vers 0 en probabilité.

   $\liminf \mathbb{P}(Y_n > \epsilon) \geq \liminf \mathbb{P}(M_n > n\epsilon) = \liminf(1 - \mathbb{P}(M_n \leq n\epsilon))$
   $\geq \liminf 1 - ((1-(\epsilon n + n)^{-1}))^n = 1 - \exp(-1/(1+\epsilon)) > 0$

3.3 Convergence dans L^p

Exercice 11.

1. Montrer qu'une famille finie $(X_1, \cdots, X_n)$ de variables aléatoires intégrables est équi-intégrable.

   $\lim_{c\to\infty} \sup_{1\leq i\leq n} \int_{|X_i|>c} |X_i| d\mathbb{P} = \lim_{c\to\infty} \max_{1\leq i\leq n} \mathbb{E}[\mathbb{1}_{|X_i|>c}|X_i|]$ qui tend vers 0 par convergence dominée (on peut majorer le max par la somme).

2. Montrer qu'une famille de variables aléatoires $(X_i)_{i\in I}$ dominées par Y intégrable ($|X_i| \leq |Y|$ pour tout i) est équi-intégrable.

   $\lim_{c\to\infty} \sup_{i\in I} \int_{|X_i|>c} |X_i| d\mathbb{P} = \lim_{c\to\infty} \sup_{i\in I} \mathbb{E}[\mathbb{1}_{|X_i|>c}|X_i|]$
   $\leq \lim_{c\to\infty} \mathbb{E}[\mathbb{1}_{|Y|>c}|Y|]$ qui tend encore vers 0 par convergence dominée.

3. Soit $(X_i)_{i\in I}$ une famille équi-intégrable et X une variable aléatoire intégrable. Montrer que la famille $(X_i) \cup \{X\}$ est encore équi-intégrable.

   Notons $X = X_0$ et $J = I \cup \{0\}$.
   $\lim_{c\to\infty} \sup_{j\in J} \int_{|X_j|>c} |X_j| d\mathbb{P} \leq \lim_{c\to\infty} \sup_{i\in I} \int_{|X_i|>c} |X_i| d\mathbb{P} + \lim_{c\to\infty} \int_{|X|>c} |X| d\mathbb{P} = 0$ car une famille finie est équi-intégrable.

3.4 Loi des grands nombres

Exercice 12. Soit (X_n) une suite de variables aléatoires réelles indépendantes de même loi uniforme sur l'intervalle $[0,1]$. Calculer la limite de

$$\mathbb{E}\left[\frac{X_1^2 + \cdots X_n^2}{X_1 + \cdots + X_n}\right].$$

La loi uniforme est une loi intégrable, on peut donc utiliser la loi des grands nombres pour avoir $(X_1 + \cdots + X_n)/n$ converge ps vers $\mathbb{E}[X_1] = 1/2$ et $(X_1^2 + \cdots X_n^2)/n$ converge ps vers $\mathbb{E}[X_1^2] = \int_0^1 x^2 dx = 1/3$. En faisant le rapport (ce qui est possible pour la convergence ps), on en déduit que $\frac{X_1^2+\cdots X_n^2}{X_1+\cdots+X_n}$ converge ps vers $2/3$. De plus, comme les X_k sont à valeurs dans $[0,1]$, on a $X_k^2 \leq X_k$ ps, donc $\frac{X_1^2+\cdots X_n^2}{X_1+\cdots+X_n} \leq 1$. Par convergence dominée (ou en utilisant qu'une suite bornée est équi-intégrable) on en déduit que $\mathbb{E}\left[\frac{X_1^2+\cdots X_n^2}{X_1+\cdots+X_n}\right]$ converge également vers $2/3$.

3.5 Application de a loi des grands nombres à l'estimation ponctuelle

Exercice 13. Pour estimer la proportion p des foyers disposant d'une tablette dans une population de grande taille, on tire au sort deux échantillons indépendants de n_1 et n_2 foyers respectivement. A chaque foyer i on associe la variable aléatoire X_i qui vaut 1 si le foyer a une tablette, et 0 sinon. On note F_1 et F_2 les moyennes empiriques de chaque échantillon.

1. Donner la loi de X_i pour tout i.

   X_i suit la loi de Bernoulli de paramètre p.

2. Calculer l'espérance et la variance de F_1 et F_2.

   $\mathbb{E}[F_1] = \mathbb{E}[(X_1 + \cdots + X_{n_1})/n_1] = n_1\mathbb{E}[X_1]/n_1 = \mathbb{E}[X_1] = p$
   $\mathsf{Var}(F_1) = \mathsf{Var}((X_1+\cdots+X_{n_1})/n_1) = (\mathsf{Var}(X_1)+\cdots+\mathsf{Var}(X_{n_1}))/n_1^2 = n_1p(1-p)/n_1^2 = p(1-p)/n_1$ par indépendance. On a des résultats analogues pour F_2.

3. L'estimateur $F = \frac{F_1+F_2}{2}$ est-il sans biais ? Calculer sa variance.

   L'estimateur de la moyenne empirique est sans biais, donc $\mathbb{E}[F] = (\mathbb{E}[F_1]+\mathbb{E}[F_2])/2 = (p+p)/2 = p$. Donc F est encore sans biais. Comme les échantillons sont indépendants, F_1 et F_2 qui en sont les moyennes empiriques sont également indépendants. Donc $\mathsf{Var}(F) = (\mathsf{Var}(F_1) + \mathsf{Var}(F_2))/4 = p(1-p)(1/n_1 + 1/n_2)/4$.

4. Comparer les variances respectives de F, F_1 et F_2 en supposant $n_1 > n_2$. A quelles conditions la variance de F est-elle inférieure à celles de F_1 et F_2 ?

   $\mathsf{Var}(F_1) = p(1-p)/n_1$, $\mathsf{Var}(F_2) = p(1-p)/n_2$, $\mathsf{Var}(F) = p(1-p)(1/n_1 + 1/n_2)/4$. Si $n_1 > n_2$, alors $\mathsf{Var}(F_1) < \mathsf{Var}(F_2)$.
   On a $\mathsf{Var}(F) < \mathsf{Var}(F_1) \Leftrightarrow (1/n_1 + 1/n_2)/4 < 1/n_1 \Leftrightarrow n_1 < 3n_2$.

5. Déterminer les coefficients réels a et b pour que $F^* = aF_1 + bF_2$ soit un estimateur de p sans biais et de variance minimale.

   $\mathbb{E}[F^*] = (a + b)p$, donc on doit avoir $a + b = 1$.
   $\mathsf{Var}(F^*) = (a^2/n_1 + b^2/n_2)p(1-p)$, on cherche donc à minimiser $f(a) = a^2/n_1 + (1-a)^2/n_2 = a^2(1/n_1 + 1/n_2) - 2a/n_2 + 1/n^2$ sur l'intervalle $[0, 1]$. Il s'agit d'une fonction quadratique qui tend vers l'infini en $\pm\infty$, son minimum est donc atteint en l'unique point où sa dérivée s'annule $f'(a) = 0 \Leftrightarrow a = n_1/(n_1 + n_2)$. On a donc $F^* = n_1/(n_1 + n_2)F_1 + n_2/(n_1 + n_2)F_2$ qui est exactement l'estimateur de la moyenne empirique de l'échantillon complet de taille $n_1 + n_2$.

Exercice 14.* Soit $(X_1, \cdots, X_n)$ un échantillon aléatoire d'une loi normale d'espérance θ et de variance $\theta(1-\theta)$, où θ est un paramètre inconnu de $]0; 1[$.

1. Montrer que les estimateurs

$$T_1 = \frac{1}{n}\sum_{i=1}^{n} X_i, \qquad T_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2$$

   sont sans biais et convergents.

   La convergence est donnée par la loi des grands nombres.
   $\mathbb{E}[T_1] = \sum \mathbb{E}[X_i]/n = n\theta/n = \theta$ donc T_1 est sans biais.
   $\mathbb{E}[T_2] = \sum \mathbb{E}[X_i^2]/n = n(\mathsf{Var}(X) + \mathbb{E}[X]^2)/n = \theta(1-\theta) + \theta^2 = \theta$ donc T_2 est également sans biais.

2. Pour Y de loi normale normale $\mathcal{N}(m, \sigma^2)$ calculer $\mathbb{E}[Y^4]$.

   On peut calculer directement l'intégrale, mais on peut aussi faire une développement limité de la fonction caractéristique à l'ordre 4

$$\begin{aligned}
\Phi(t) &= \exp(imt - \sigma^2 t^2/2) \\
&= 1 + imt - \sigma^2 t^2/2 + (imt - \sigma^2 t^2/2)^2/2 + (imt - \sigma^2 t^2/2)^3/6 \\
&\quad +(imt - \sigma^2 t^2/2)^4/24 + o(t^4) \\
&= 1 + imt - \sigma^2 t^2/2 + (-m^2 t^2 - im\sigma^2 t^3 + \sigma^4 t^4/4)/2 \\
&\quad +(-im^3 t^3 + 3m^2\sigma^2 t^4/2)/6 + m^4 t^4/24 + o(t^4) \\
&= 1 + imt + i^2(m^2 + \sigma^2)/t^2/2 + i^3(3m\sigma^2 + m^3)t^3/6 \\
&\quad +i^4(3\sigma^4 + 6m^2\sigma^2 + m^4)t^4/24 + o(t^4)
\end{aligned}$$

   On identifie donc $\mathbb{E}[Y] = m$, $\mathbb{E}[Y^2] = m^2 + \sigma^2$, $\mathbb{E}[Y^3] = 3m\sigma^2 + m^3$ et $\mathbb{E}[Y^4] = 3\sigma^4 + 6m^2\sigma^2 + m^4$.

3. Quel estimateur choisir ?

   On choisit celui qui a la variance la plus petite. Par indépendance, on a $\mathsf{Var}(T_1) = n\,\mathsf{Var}(X_i)/n^2 = \mathsf{Var}(X_i)/n = \theta(1-\theta)/n$

$$\begin{aligned}
\mathsf{Var}(T_2) &= n\,\mathsf{Var}(X_i^2)/n^2 \\
&= \mathsf{Var}(X_i^2)/n \\
&= (\mathbb{E}[X_i^4] - \mathbb{E}[X_i^2]^2)/n \\
&= (3\theta^2(1-\theta)^2 + 6\theta^2\theta(1-\theta) + \theta^4 - (\theta^2 + \theta(1-\theta))^2)/n \\
&= 2\theta^2(1-\theta)(1+\theta)
\end{aligned}$$

   Donc $\mathsf{Var}(T_2) < \mathsf{Var}(T_1) \Leftrightarrow 2\theta^2(1-\theta)(1+\theta) < \theta(1-\theta) \Leftrightarrow 2\theta(1+\theta) < 1 \Leftrightarrow \theta < (-1+\sqrt{3})/2$. Donc on choisit T_1 si $\theta \geq (-1+\sqrt{3})/2$, T_2 sinon.