UNIVERSITÉ DE VERSAILLES-ST-QUENTIN **Licence Sciences, Semestre 4**
Probabilités LSMA430 **2019/2020**

Corrigé n°6 : Convergences – Théorèmes limite

Exercice 1 : (pile ou face équitable)

Deux joueurs à pile ou face ont obtenu 48% de pile l'un en jouant 100 fois, l'autre en jouant 10 000 fois. Tous les deux pensent que la pièce qu'ils ont utilisée est équilibrée. Qu'en pensez-vous ?

Correction : Nous sommes dans le contexte d'une épreuve de Bernoulli. Le joueur 1 répète l'expérience du lancer d'une pièce 100 fois et le joueur 2 la répète 10 000 fois. On note X_i la variable aléatoire qui prend la valeur 1 si le résultat de l'expérience est "pile" et 0 sinon. On a donc $X_i \sim \mathcal{B}(p)$ où p est le paramètre de succès.

Par la loi faible des grands nombres (dont les hypothèses ici sont évidemment vérifiées), en notant $\overline{X_n}$ la moyenne empirique, on a

$$\overline{X_n} := \frac{1}{n}\sum_{i=1}^{n} X_i \longrightarrow m = \mathbb{E}(X_1)$$

en probabilité quand $n \to \infty$. Comme $X_1 \sim \mathcal{B}(p)$, on a $m = p$.

Si la pièce n'est pas truquée alors $p = 0.5$. On se demande si l'écart à la moyenne obtenu par les 2 joueurs est normal. En appliquant l'inégalité de Bienaymé-Tchebychev à la variable $\overline{X_n}$ avec $\varepsilon = 0.5 - 0.48 = 0.02$, on a

$$\begin{aligned}\mathbb{P}(|\overline{X_n} - m| > 0.02) &\le \frac{Var(\overline{X_n})}{0.02^2}\\ &= \frac{Var(X_1)}{n \times 0.02^2}\\ &= \frac{p(1-p)}{n \times 4 \times 10^{-4}}\end{aligned}$$

Pour la première égalité, on utilise l'indépendance des variables aléatoires X_i et pour la seconde le fait qu'elles soient distribuées selon une Bernoulli. On remarque que quand n augmente la borne diminue. En d'autres termes, plus n est grand, plus la probabilité de s'éloigner de la moyenne m diminue. En appliquant cette borne avec $n = 100$, et en supposant que la pièce est équilibrée, i.e. $p = 0.5$, on obtient

$$\mathbb{P}(|\overline{X}_{100} - m| > 0.02) \le \frac{0.25}{100 \times 4 \times 10^{-4}} = 6.25.$$

De même avec $n = 10000$, on a

$$\mathbb{P}(|\overline{X}_{10000} - m| > 0.02) \le \frac{0.25}{10000 \times 4 \times 10^{-4}} = 0.0625.$$

Conclusion : Au bout de 100 lancers, on ne peut pas conclure sur l'équilibrage de la pièce (la borne que l'on obtient par l'inégalité de BC est plus grande que 1). Au bout de 10 000 lancers, l'écart entre la moyenne empirique $\overline{X_n}$ et la moyenne théorique m semble trop grand. La pièce est donc sûrement truquée (on en est sûr à 94%).

Exercice 2 : (erreur de calcul)

Un micro-processeur effectue, lors d'une opération, une erreur de calcul avec la probabilité $p = 10^{-5}$. Soit P_n la proportion d'erreurs faites en n opérations. Calculer l'espérance et la variance de P_n.

Déterminer le nombre n d'opérations nécessaires pour que cette proportion d'erreurs P_n vérifie

$$\mathbb{P}\Big(\frac{|P_n - p|}{p} > 0,2\Big) \leq 0,05$$

(a) à partir de l'inégalité de Bienaymé-Tchebychev ;
(b) à partir du théorème central limite.

Indication : la table de loi normale donne $\Phi(1,96) = 0,975$ où Φ est la fonction de répartition de la loi normale centrée réduite.

Correction :

Soit la variable aléatoire X_i qui prend la valeur 1 s'il y a une erreur à l'opération i et 0 sinon. $X_i \sim \mathcal{B}(p)$ avec $p = 10^{-5}$. Soit P_n la proportion d'erreurs faites en n opérations. On a

$$P_n = \frac{\sum_{i=1}^n X_i}{n}$$

En utilisant successivement la linéarité de l'espérance, puis le fait que les X_i soient de même loi, il vient

$$\begin{aligned}\mathbb{E}(P_n) &= \mathbb{E}\left(\frac{\sum_{i=1}^n X_i}{n}\right)\\ &= \frac{1}{n}\sum_{i=1}^n \mathbb{E}(X_i)\\ &= \frac{1}{n}\ n\mathbb{E}(X_1) = \frac{1}{n}\ np = p\end{aligned}$$

On aurait pu également utiliser le fait que nP_n suit une loi binomiale de paramètres n et p.

De façon similaire et grâce à l'indépendance, il vient

$$\begin{aligned}\mathbb{V}ar(P_n) &= \frac{1}{n^2}\sum_{i=1}^n \mathbb{V}ar(X_i)\\ &= \frac{1}{n^2}\ n\mathbb{V}ar(X_1) = \frac{1}{n^2}\ np(1-p) = \frac{1}{n}p(1-p)\end{aligned}$$

a) En remarquant que

$$\mathbb{P}\Big(\frac{|P_n - p|}{p} > 0.2\Big) = \mathbb{P}\Big(|P_n - p| > 0.2\ p\Big)$$

on peut appliquer l'inégalité de Bienaymé-Tchebychev :

$$\begin{aligned}\mathbb{P}(|\overline{P_n} - p| > 0.2p) &\leq \frac{Var(X_1)}{n(0.2p)^2}\\ &= \frac{p(1-p)}{n(0.2p)^2}\end{aligned}$$

Les variables étant bien iid d'espérance et de variance finies.

On cherche n tel que

$$\frac{p(1-p)}{n(0.2p)^2} \leq 0.05$$

ce qui équivaut à

$$\frac{p(1-p)}{0.05(0.2p)^2} \leq n$$

En remplaçant p par sa valeur, on obtient : $5 * 10^7 \leq n$.

b) Les variables $(X_i)_i$ étant bien iid d'espérance et de variance finies, on peut appliquer le TCL :

$$\frac{\sqrt{n}(P_n - p)}{\sigma} \to Z \sim \mathcal{N}(0,1)$$

avec $\sigma = \sqrt{p(1-p)}$. En utilisant la définition de la valeur absolue, il vient

$$\begin{aligned} |P_n - p| > 0.2p &\Leftrightarrow P_n - p > 0.2p \text{ ou } P_n - p < -0.2p \\ &\Leftrightarrow \frac{\sqrt{n}(P_n - p)}{\sigma} > \frac{\sqrt{n}(0.2p)}{\sigma} \text{ ou } \frac{\sqrt{n}(P_n - p)}{\sigma} < -\frac{\sqrt{n}(0.2p)}{\sigma} \end{aligned}$$

En utilisant le TCL pour n assez grand, il vient

$$|P_n - p| > 0.2p \Leftrightarrow Z > \frac{\sqrt{n}(0.2p)}{\sigma} \text{ ou } Z < -\frac{\sqrt{n}(0.2p)}{\sigma}$$

En posant $a = \frac{\sqrt{n}(0.2p)}{\sigma}$, il vient pour n assez grand

$$\begin{aligned} \mathbb{P}\left(\frac{|P_n - p|}{p} > 0.2\right) &= \mathbb{P}(Z > a) + \mathbb{P}(Z < -a) \\ &= 2(1 - \Phi(a)) \end{aligned}$$

L'objectif est de trouver n tel que

$$\begin{aligned} &\mathbb{P}\Big(\frac{|P_n - p|}{p} > 0.2\Big) \leq 0.05 \\ &\Leftrightarrow 2(1 - \Phi(a)) \leq 0.05 \\ &\Leftrightarrow \Phi(a) \geq 0.975 \end{aligned}$$

En utilisant la remarque de l'énoncé, il vient

$$\begin{aligned} &a \geq 1,96 \\ &\Leftrightarrow \frac{\sqrt{n}(0.2p)}{\sqrt{p(1-p)}} \geq 1,96 \\ &\Leftrightarrow n \geq \frac{p(1-p)}{(0.2p)^2}(1.96)^2 \end{aligned}$$

En remplaçant p par sa valeur, on obtient : $9.6 * 10^6 \leq n$.

Les deux bornes donnent des résultats différents. La borne issue de l'inégalité de Bienaymé-Tchebychev est plus restrictive que celle issue du TCL.

Exercice 3 : (l'indépendance est primordiale)

Soit $(X_n)_n$ une suite de variables aléatoires toutes égales :

$$X_1 = X_2 = \dots$$

de moyenne $m = 0$ et d'écart-type $\sigma = 1$. Soit $\overline{X_n} = \dfrac{X_1 + \dots + X_n}{n}$ leur moyenne empirique. Montrer que

$$\frac{\sqrt{n}\left(\overline{X_n} - m\right)}{\sigma} = X_1\sqrt{n}.$$

Constater que cette quantité ne converge pas vers une loi gaussienne $\mathcal{N}(0,1)$.

Correction :

Les $(X_i)_{i\geq 1}$ sont identiquement distribuées mais ne sont pas indépendantes. On a $\overline{X_n} = \dfrac{X_1 + \dots + X_n}{n} = \dfrac{n\ X_1}{n}$ car $X_1 = X_2 = \dots$. D'où $\overline{X_n} = X_1$.

$$
\begin{aligned}
\frac{\sqrt{n}\left(\overline{X_n}-m\right)}{\sigma} &= \frac{\sqrt{n}\left(\frac{X_1+\cdots+X_n}{n}-m\right)}{\sigma} \\
&= \sqrt{n}\left(\frac{X_1+\cdots+X_n}{n}\right) \text{ car } m=0 \text{ et } \sigma=1 \\
&= X_1\sqrt{n} \text{ on a montré que } \overline{X_n}=X_1.
\end{aligned}
$$

$X_1\sqrt{n}$ ne converge pas vers la loi $\mathcal{N}(0,1)$ car :

$$\mathbb{V}ar(X_1\sqrt{n}) = n\mathbb{V}ar(X_1) = n \to +\infty \text{ quand } n \to +\infty \text{ et non vers 1.}$$

Remarque : l'hypothèse d'indépendance des variables est primordiale pour utiliser le théorème central limite.

Exercice 4 : (limite de Poisson(s))

Soient n variables aléatoires $X_1, X_2, \dots X_n$ indépendantes et de même loi de Poisson de paramètre $\lambda = 4$. Soit

$$\overline{X_n} = \frac{X_1+\cdots+X_n}{n}.$$

(a) Rappeler $\mathbb{E}(X_1)$ et $Var(X_1)$.

(b) Calculer $\mathbb{E}\left(\overline{X_n}\right)$ et $Var\left(\overline{X_n}\right)$.

(c) Ecrire le théorème central limite pour $\overline{X_n}$.

(d) En déduire une approximation de $\mathbb{P}(X_1+\cdots+X_n \geq 425)$ exprimée en fonction de n et de Φ, la fonction de répartition de la loi normale $\mathcal{N}(0,1)$.

Rappel : $\Phi(t) = \displaystyle\int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}\,du$

Application numérique : $n = 100$. Calculer $\mathbb{P}(X_1+\cdots+X_n \geq 425)$.

Correction :

(a) Commes les X_i sont iid de loi de Poisson de paramètre $\lambda = 4$, on a $\mathbb{E}(X_1) = Var(X_1) = \lambda = 4$.

(b) En utilisant successivement la définition $\overline{X_n}$, la linéarité de l'espérance, puis le fait que les X_i soient de même loi $\mathcal{P}(\lambda = 4)$, il vient

$$
\begin{aligned}
\mathbb{E}\left(\overline{X_n}\right) &:= \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \\
&= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(X_i) \\
&= \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}(X_1) \\
&= \frac{1}{n} n\lambda = \lambda = 4.
\end{aligned}
$$

Par des arguments similaires, en notant toutefois que l'indépendance ici est cruciale pour obtenir la deuxième égalité, on obtient

$$
\begin{aligned}
Var\left(\overline{X_n}\right) &= \frac{1}{n^2} Var\left(\sum_{i=1}^{n} X_i\right) \\
&= \frac{1}{n^2} \sum_{i=1}^{n} Var(X_i) \\
&= \frac{1}{n^2} \sum_{i=1}^{n} Var(X_1) \\
&= \frac{1}{n^2} n\lambda = \frac{\lambda}{n} = \frac{4}{n}.
\end{aligned}
$$

(c) On applique le TCL à la moyenne empirique $\overline{X_n}$ avec les X_i indépendants, de même loi d'espérance et de variance finies ($\mathbb{E}(X_1) = Var(X_1) = \lambda = 4$ d'après le (a)). On a alors quand $n \to \infty$

$$
\frac{\sqrt{n}(\overline{X_n} - m)}{\sigma} \longrightarrow \mathcal{N}(0,1).
$$

Ce qui donne, avec $\sigma = 2$ et $m = \sigma^2 = 4$

$$
Y_n = \frac{\sqrt{n}(\overline{X_n} - 4)}{2} \longrightarrow \mathcal{N}(0,1).
$$

(d) En utilisant ce qui précède, on a

$$
\begin{aligned}
\mathbb{P}\left(\sum_{i=1}^{n} X_i \geq 425\right) &= \mathbb{P}(n\overline{X_n} \geq 425) \\
&= \mathbb{P}(\overline{X_n} \geq 425/n) \\
&= \mathbb{P}\left(\frac{\sqrt{n}(\overline{X_n} - 4)}{2} \geq \frac{\sqrt{n}(425/n - 4)}{2}\right) \\
&= 1 - \Phi\left(\frac{\sqrt{n}(425/n - 4)}{2}\right).
\end{aligned}
$$

Avec $n = 100$, on a alors

$$
\begin{aligned}
\mathbb{P}\left(\sum_{i=1}^{n} X_i \geq 425\right) &= 1 - \Phi\left(\frac{10(425/100 - 4)}{2}\right) \\
&= 1 - \Phi(1.25) \\
&= 1 - 0.8944 \\
&= 0.1056.
\end{aligned}
$$

<u>Exercice 5 :</u> (une somme de Poisson est gaussienne)

Soient $X_1, X_2, \ldots, X_n, \ldots$ une suite de v.a. indépendantes de même loi de Poisson de paramètre $\lambda = 1$. Quelle est la loi de $S_n := X_1 + \cdots + X_n$? En déduire $\mathbb{P}(S_n = k)$ pour tout $k = 0, \ldots, n$.

Calculer l'espérance et la variance de S_n.

Ecrire le théorème central limite pour S_n.

Montrer la relation :

$$
\lim_{n \to +\infty} e^{-n} \sum_{k=0}^{n} \frac{n^k}{k!} = \frac{1}{2}.
$$

Correction :

(a) On va montrer que la loi de S_n est une loi de Poisson de paramètre n.

Preuve : On va faire la preuve pour $n = 2$. Soient $X \sim \mathcal{P}(\lambda)$ et $Y \sim \mathcal{P}(\mu)$ indépendantes. Alors

$$\begin{aligned}\mathbb{P}(X+Y=k) &= \sum_{j=0}^{k} \mathbb{P}(X=j)\mathbb{P}(Y=k-j) \\ &= \sum_{j=0}^{k} \frac{\lambda^j}{j!}e^{-\lambda}\frac{\mu^{k-j}}{(k-j)!}e^{-\mu} \\ &= e^{-(\lambda+\mu)}\sum_{j=0}^{k} \frac{\lambda^j}{j!}\frac{\mu^{k-j}}{(k-j)!}.\end{aligned}$$

la formule du binôme de Newton donne

$$(\lambda+\mu)^k = \sum_j \frac{k!}{j!(k-j)!}\mu^{k-j}\lambda^j$$

Il vient ainsi

$$\mathbb{P}(X+Y=k) = \frac{e^{-(\lambda+\mu)}}{k!}(\lambda+\mu)^k.$$

En d'autres termes, $X+Y \sim \mathcal{P}(\lambda+\mu)$. Ainsi, si $\lambda = \mu = 1$ alors $X+Y \sim \mathcal{P}(2)$. La preuve se généralise facilement au cas où n est arbitraire.

(b) On a montré dans (a) que $S_n \sim \mathcal{P}(n)$. On a donc $\mathbb{P}(S_n = k) = \frac{n^k}{k!}e^{-n}$ pour tout $k \in [0, n]$.

(c) $\mathbb{E}(S_n) = Var(S_n) = n$. En effet, en utilisant successivement la linéarité de l'espérance, et le fait que les X_i soient de même loi de Poisson de paramètre 1, on obtient

$$\mathbb{E}(S_n) = \mathbb{E}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n}\mathbb{E}(X_i) = \sum_{i=1}^{n}\mathbb{E}(X_1) = n$$

et de façon similaire, grâce à l'indépendance, on a

$$Var(S_n) = Var\left(\sum_{i=1}^{n} X_i\right) = Var\sum_{i=1}^{n}(X_1) = n.$$

(d) On applique le TCL à la suite de variables aléatoires indépendantes et identiquement distribuées $X_1, \dots, X_n$ d'espérance $m = 1 < \infty$ et de variance $Var(X_1) = 1 < \infty$ dont la somme est d'espérance $\mathbb{E}(S_n) = n$ et $\sigma(S_n) = \sqrt{n}$. On obtient

$$Y_n = \frac{S_n - n}{\sqrt{n}} \sim \mathcal{N}(0,1).$$

(e) Soit $Z \sim \mathcal{N}(0,1)$. D'après le TCL, quand $n \to \infty$

$$\mathbb{P}\left(\frac{X_1 + \cdots + X_n - n}{\sqrt{n}} \leq 0\right) \longrightarrow \mathbb{P}(Z \leq 0) = \frac{1}{2}.$$

On conclut donc que

$$\mathbb{P}(X_1 + \cdots + X_n \leq n) = \sum_{k=0}^{n} e^{-n}\frac{n^k}{k!} \longrightarrow \frac{1}{2}$$

quand $n \to \infty$.