LA LOI DES GRANDS NOMBRES ET LE THÉORÈME DE LA LIMITE CENTRALE

MATTHIEU KOWALSKI

1. INTRODUCTION

La démarche statistique consiste à observer une expérience aléatoire dans le but de mieux connaître ses caractéristiques.

Définition 1 (n-échantillon)**.** *On appelle n-échantillon de loi P l'ensemble $X_1, \dots, X_n$ de n variables aléatoires indépendantes et identiquement distribuées selon P.*

Autrement dit, on a

$$\forall i \in \{1, \dots, n\}\ X_i \sim P \quad \text{et} \quad \forall i \neq j X_i \text{ et } X_j \text{ indépendants .}$$

Le n-échantillon représente l'information dont on dispose.

Définition 2. *On appelle modèle statistique la famille de loi possible pour notre n-échantillon. On le note $\{P_\theta, \theta \in \Theta\}$, où θ est un paramètre qui caractérise la loi.*

1.1. Estimateur de la moyenne.

Le but de l'estimation est, à partir d'un n-échantillon de loi P_θ, d'estimer le paramètre θ de la loi P.

Définition 3. *Soit $X_1, \dots, X_n$ un n-échantillon de loi P tel que $\forall i\ \mathbb{E}\{X_i\} = \mu$ et $\mathrm{Var}\{X_i\} = \sigma^2 < +\infty$. On définit la quantité*

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$$

appelée moyenne empirique.

La moyenne empirique est un estimateur de la moyenne μ. On a la propriété suivante :

Proposition 1.

$$\mathbb{E}\{\bar{X}_n\} = \mu \quad \textit{et} \quad \mathrm{Var}\left\{\bar{X}_n\right\} = \frac{\sigma^2}{n} .$$

On verra un exemple pratique avec la modélisation du jeu de pile ou face.

2. LA LOI DES GRANDS NOMBRES

La loi des grands nombres a été formalisée au XVIIe XVIIe siècle lors de la découverte de nouveaux langages mathématiques. Essentiellement, la loi des grands nombres indique que lorsque l'on fait un tirage aléatoire dans une série de grande taille, plus on augmente la taille de l'échantillon, plus les caractéristiques statistiques du tirage (l'échantillon) se rapprochent des caractéristiques statistiques de la population. Mais il est intéressant de noter que la taille de l'échantillon à prendre

pour approcher les caractéristiques de la population initiale ne dépend que faiblement voire pas du tout de la taille de la série initiale : pour un sondage au Luxembourg ou aux États-Unis, il suffit, pour obtenir une précision égale de prendre un échantillon de même taille. C'est sur cette loi que reposent la plupart des sondages Ils interrogent un nombre suffisamment important de personnes pour connaître l'opinion (probable) de la population entière. De même, sans la formalisation de la loi des grands nombres, l'assurance n'aurait jamais pu se développer avec un tel essor. En effet, cette loi permet aux assureurs de déterminer les probabilités que les sinistres dont ils sont garants se réaliseront ou non. La loi des grands nombres sert aussi en statistique inférentielle, pour déterminer une loi de probabilité à partir d'une série d'expériences.

Théoreme 1 (Loi faible des grands nombres)**.** *Soit $X_1, X_2, \ldots, X_n$ n variables aléatoires indépendantes et identiquement distribuées d'espérance commune μ et de variance commune $\sigma^2 < \infty$. On pose $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ la moyenne empirique. Alors*

$$\forall \varepsilon > 0, \quad \lim_{n \to +\infty} \mathbb{P}\left\{|\bar{X}_n - m| > \varepsilon\right\} = 0$$

Démonstration. Le théorème se montre en utilisant l'inégalité de Tchebychev :

$$\mathbb{P}\left\{|\bar{X}_n - \mathbb{E}\{\bar{X}\}| > \varepsilon\right\} \leq \frac{\mathrm{Var}\left\{\bar{X}_n\right\}}{\varepsilon^2}$$

avec

$$\mathbb{E}\{\bar{X}\} = \mathbb{E}\left\{\frac{1}{n}\sum_{i=1}^{n} X_i\right\} = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\left\{X_i\right\} = \mu$$

et, parce que les X_i sont indépendantes,

$$\mathrm{Var}\left\{\bar{X}\right\} = \mathrm{Var}\left\{\frac{1}{n}\sum_{i=1}^{n} X_i\right\} = \frac{1}{n^2}\sum_{i=1}^{n} \mathrm{Var}\left\{X_i\right\} = \frac{\sigma^2}{n}$$

Par conséquent $\mathbb{P}\left\{|\bar{X}_n - \mathbb{E}\{\bar{X}\}| > \varepsilon\right\} \leq \frac{\mathrm{Var}\left\{\bar{X}_n\right\}\sigma^2}{n\varepsilon^2}$ et donc

$$\lim_{n \to +\infty} \mathbb{P}\left\{|\bar{X}_n - m| > \varepsilon\right\} = 0$$

∎

Il ne faut toutefois pas confondre la moyenne des gains et le gain absolu. Si deux joueurs jouent très longtemps à pile ou face, celui qui perd donnant un euro à celui qui gagne, la moyenne des gains de chaque joueur tendra effectivement vers 0 (la moyenne étant définie comme le gain divisé par le nombre de parties jouées), mais le gain de chaque joueur passera alternativement par des hauts et des bas.

3. Modélisation du jeu de pile ou face

On peut modéliser un jeu de pile ou face par une suite de variables aléatoires indépendantes identiquement distribuées $(X_k)_{k \in \mathbb{N}^*}$ à valeurs dans $\{0, 1\}$ telles que

$$\mathbb{P}\{\text{le } k\text{-ième lancer donne 'pile'}\} = \mathbb{P}\{X_k = 1\} = p$$
$$\mathbb{P}\{\text{le } k\text{-ième lancer donne 'face'}\} = \mathbb{P}\{X_k = 0\} = 1 - p$$

où p est un paramètre fixé dans $[0,1]$. La loi des $(X_k)_{k\in\mathbb{N}^*}$ est appelée *loi de Bernoulli* de paramètre p et est notée $\mathcal{B}(p)$.

Proposition 2. *On rappelle que l'espérance et la variance d'une variable aléatoire X suivant une loi de Bernoulli de paramètre p sont données par*

$$\mathbb{E}\{X\} = p \quad et \quad \mathrm{Var}\{X\} = p(1-p)\ .$$

La première grandeur à laquelle on s'intéresse est le nombre de 'pile' obtenu lors de n tirages successifs (le nombre de 'face' s'en déduit...). On introduit donc la variable aléatoire :

$$N_n = \sum_{i=1}^{n} X_i\ .$$

La loi de N_n est appelée *loi binômiale* de paramètre n (le nombre de tirages) et p (la proabilité de tomber sur 'pile'), et est noté $\mathcal{B}(n,p)$.

Remarque 1. *On peut remarquer que la loi binômiale de paramètres $n = 1$ et $p \in [0,1]$ quelconque,* **est** *la loi de Bernoulli de paramètre p.*

Proposition 3. *L'espérance et la variance d'une variable aléatoire X suivant une loi Binômilae de paramètres $n \in \mathbb{N}^*$ et $p \in [0,1]$ sont données par*

$$\mathbb{E}\{X\} = np \quad et \quad \mathrm{Var}\{X\} = np(1-p)\ .$$

Expérience 1. *Prendre une pièce de monnaie. Le paramètre p de cette pièce est* a priori *inconnu. Répeter plusieurs fois l'expérience suivante :*

«Lancer 10 fois la pièce de monnaie, et compter le nombre de 'pile'.»

(1) *Quelle est la loi qui permet de modéliser cette expérience ?*

(2) *Théoriquement, quelles valeurs peut prendre le nombre de 'pile' obtenu au cours d'une expérience ?*

(3) *Faire un diagramme en baton qui représente les résultats obtenus par cette expérience.*

On va maintenant étudier le comportement asymtotique de $\frac{N_n}{n}$, c'est à dire l'évolution du nombre moyen de 'pile' dans n tirages pour n grand.

Expérience 2. *Lancer plusieurs fois la pièce de monnaie. À chaque lancé, recalculer le nombre de 'pile' obtenu en moyenne.*

(1) *Vers quelle valeur la moyenne semble-t-elle converger ?*

(2) *Dessinez sur un graphique l'évolution de la moyenne en fonction du nombre de lancé. Que remarquez-vous ?*

(3) *En quoi est-ce que cette expérience illustre la loi des grands nombres ?*

Le résultat de cet exercice est illustré sur les figures 1 (a) et (b). On simule le jet d'une pièce de monaie mal équilibrée : on a 7 chance sur 10 de tomber sur pile. Autrement dit, la pièce de monaie suit une loi de Bernoulli $\mathcal{B}(0.7)$. On fait une première expérience (figure 1 (a)), où l'on jette 500 fois la pièce. À chaque lancé, on calcule la moyenne empirique. On s'intéresse alors à l'évolution de cette moyenne empirique au cours des lancés.

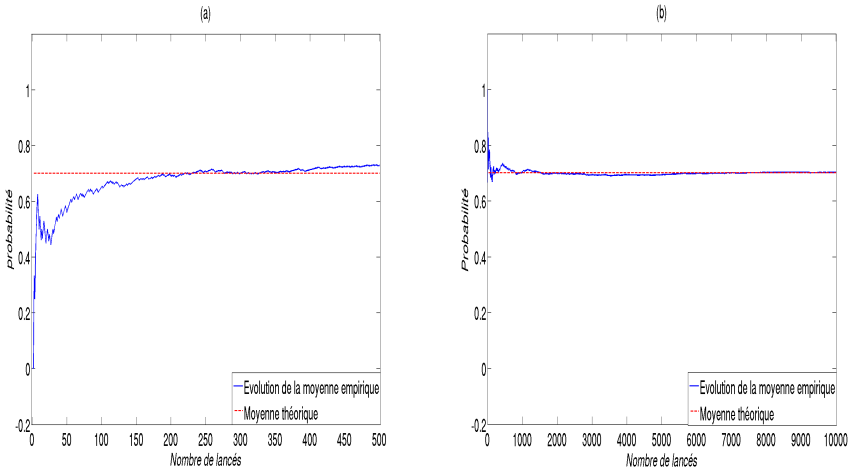

FIG. 1. Illustration de la convergence en probabilité de la loi des grands nombres

On refait ensuite la même expérience, mais cette fois ci en jettant 10000 fois la pièce (figure 1 (b)). Les figures 1 (a) et (b) montre l'évolution de ces moyennes empiriques au cours des lancés, par rapport à la moyenne théorique.

On remarque sur ces figures que plus on fait de lancés, plus la courbe de la moyenne empirique se rapproche de la courbe de moyenne théorique. De plus, la moyenne empirique peut «fluctuer» autour de la moyenne théorique, sans rester «collée». Ce comportement illustre la convergence en probabilité, et sera précisé par le théorème de la limite centrale en section 5.

4. PARADOXE OU «BON SENS» ?

On vous propose de jouer au jeu suivant. Trois boites identiques se présente à vous. L'une d'entre elle contient un lot (par exemple, une grosse somme d'argent) et les deux autres sont vides. Le but du jeu est simple : si vous choisissez la boite qui contient le lot, alors vous le gagnez. Le jeu se déroule en trois étapes :

(1) Vous choisissez une boite, au hasard.

(2) Le maitre du jeu vous montre une boite qui est vide.

(3) Vous avez alors deux possibilités :

   (a) Rester sur votre 1er choix

   (b) Changer d'avis, et choisir l'autre boite restée fermée.

D'après-vous, quelle est la bonne stratégie : rester sur son choix, ou bien changer d'avis ?

On va répondre à cette question de deux manières différentes. L'une est un simple calcul théorique. L'autre viendra pour se convaincre de la justesse de ce calcul, et consistera à appliquer la loi des grands nombres en jouant plusieurs fois à ce jeu.

4.1. **Pourquoi il faut changer.** Au début du jeu, le joueur choisit une boite, qu'on notera la boite B_C (comme «boite choisie». Comme il y a trois boites, la boite B_C a une chance sur trois de contenir le lot. Ensuite, le maitre du jeu montre une boite **vide**. Cette boite, qu'on notera B_{MJ} à donc **zéro** chance de contenir le lot ! Il ne reste plus qu'à calculer la probabilité pour la boite restante (qu'on notera B_R) de contenir le lot. Pour cela, on utilise la loi fondamentale des probabilités qui dit : «la somme des probabilités est égale à 1».

Si on résume :

$$\mathbb{P}\{B_C \text{ contient le lot}\} = \frac{1}{3}$$
$$\mathbb{P}\{B_{MJ} \text{ contient le lot}\} = 0$$
$$\mathbb{P}\{B_C \text{ contient le lot}\} + \mathbb{P}\{B_{MJ} \text{ contient le lot}\} + \mathbb{P}\{B_R \text{ contient le lot}\} = 1\ .$$

On en déduit donc que :

$$\mathbb{P}\{B_R \text{ contient le lot}\} = 1 - 0 - \frac{1}{3} = \frac{2}{3}\ !$$

C'est-à-dire que cette boite à deux fois plus de chance de contenir le lot que c'elle qu'on avait choisie au départ ! Autrement-dit, il serait stupide de ne pas choisir l'autre boite...

4.2. **Jouons !** Pour s'en convaincre, il suffit de jouer un grand nombre de fois et d'appliquer la loi des grands nombres.

Expérience 3.

(1) *Modéliser le jeu afin de pouvoir appliquer la loi des grands nombres, et retrouver $\mathbb{P}\{B_C$ contient le lot$\}$ et $\mathbb{P}\{B_R$ contient le lot$\}$.*

(2) *Jouer un certain nombre de fois, et noter à chaque fois quelle boite contient le lot (B_C ou B_R).*

(3) *Appliquer la loi des grands nombres et conclure.*

On a simulé la stratégie suivante : on choisi une boite, puis on change d'avis à chaque fois. On joue 1000 fois, et l'on compte le nombre de fois où l'on a gagné et perdu sur ces 1000 fois. D'après la loi des grands nombre, cela nous permet d'estimer la probabilité théorique de gagner et de perdre avec cette stratégie. Le résultat de cette simulation est illustré sur la figure 2, où l'on voit clairement qu'on se rapproche de la probabilité $\frac{2}{3}$ de gagner avec cette stratégie, ce qui correspond au calcul théorique !

5. Le théorème de la limite centrale

5.1. **Théorème et illustration statistique.** Les lois de probabilité gaussiennes apparaissent comme des lois limite dans des situations où on additionne des v.a.r. indépendantes de carrés intégrables et de variance petite devant celle de la somme. Précisons :

Théoreme 2 (De Moivre–Laplace)**.** *Si (X_n) est une suite de v.a.r. i.i.d. de Bernoulli ($X_i \sim \mathcal{B}(p)$), alors pour tout $-\infty \leq a < b \leq +\infty$*

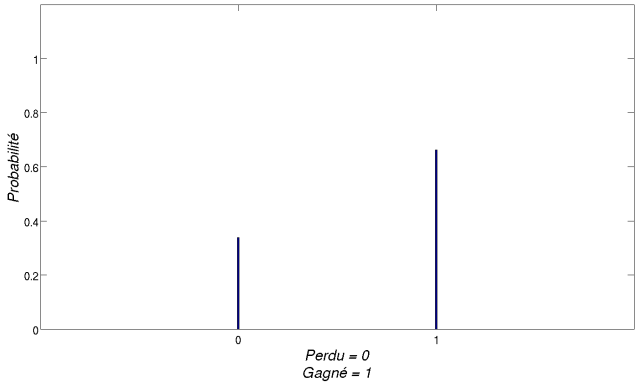

FIG. 2. Utilisation de la loi des grands nombres pour tester la stratégie théoriquement gagnante au jeu des boites.

$$\lim_{n\to+\infty} \mathbb{P}\left\{ a < \frac{S_n - np}{\sqrt{np(1-p)}} < b \right\} = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-x^2/2}\, dx$$

Autrement dit, la variable aléatoire $Z = \frac{S_n - np}{\sqrt{np(1-p)}}$ suit une loi normale centrée réduite.

Ce théorème se généralise de la manière suivante

Théoreme 3 (limite centrale)**.** *Soit (X_n) une suite de v.a.r. i.i.d. de carrés int'égrables, d'espérance commune μ et de variance commune σ^2. Alors*

$$\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \to \mathcal{N}(0,1)$$

ou de manière équivalente

$$\frac{\bar{X}_n - \mu}{\sqrt{\frac{\sigma^2}{n}}} \to \mathcal{N}(0,1)$$

Pour vérifier ce comportement, on propose d'abord la simulation suivante, qui est une illustration du théorème de De Moivre–Laplace. On fixe n grand (par exemple $n = 10000$), et l'on simule un grand nombre de fois la variable aléatoire $\bar{X}_n$ (par exemple 1000 fois), pour une variable aléatoire $X \sim \mathcal{B}(p)$. Pour simuler la variable aléatoire $\bar{X}_n$, il suffit de lancer n fois la pièce de monaie de paramètre p, et de faire la moyenne empirique de la probabilité d'apparition de pile. On répète alors cette expérience 1000 fois.

Pratiquement, ce théorème nous dit comment se comporte la variable aléatoire $\bar{X}_n$, à n fixé et grand :

$$\bar{X}_n \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$$

On illustre se comportement, sur la figure 4 cette fois ci avec une variable aléatoire suivant une loi bibomiale : $X \sim \mathcal{B}(100, 0.5)$. Pour cette simulation, on doit donc répèter l'expérience «jeter 100 fois la pièce de monaie» n fois (on a pris ici

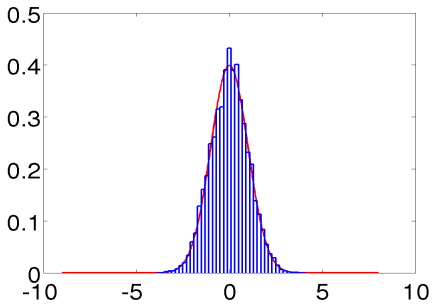

FIG. 3. Illustration du théorème de De Moivre – Laplace, avec $X\mathcal{B}(0.5)$

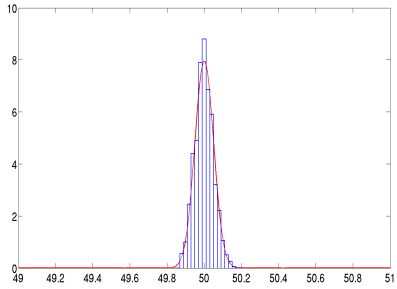

FIG. 4. Illustration du comportement de la variable aléatoire $\bar{X}_n$ où $X\mathcal{B}(100, 0.5)$.

$n = 10000$) pour calculer $\bar{X}_n$. Et on simule 1000 fois la variable aléatoire $\bar{X}_n$ (ce qui fait en tout 1 milliard de jet de pièce de monaie !). On rappelle que $\mathbb{E}\{X\} = np = 50$ et $\text{Var}\{X\} = np(1-p) = 25$.

5.2. **Utilisation du théorème en pratique : intervalles de confiance.** Le théorème de la limite centrale sert surtout en pratique pour simplifier les calculs. En particulier, on pourra l'utiliser dans le calcul d'intervalles de confiance.

Définition 4 (Intervalle de confiance)**.** *I est un intervalle de confiance au risque α pour un paramètre θ si et ssi*

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

Le théorème de la limite centrale va nous permettre de calculer un intervalle de confiance pour la moyenne μ d'une loi quelconque. Intuitivement, l'estimateur donné par la moyenne empirique doit être relativement proche de la moyenne théorique (d'après la loi des grands nombres). On va donc chercher I tel que

$$\mathbb{P}\{\mu \in I\} = \mathbb{P}\{\mu \in [\bar{X}_n - t; \bar{X}_n + t]\} = 1 - \alpha \, , \text{ où } t \in \mathbb{R}_+ \, .$$

Trouver l'intervalle de confiance I revient donc à trouver t.

$$\begin{aligned}\mathbb{P}\{\mu \in [\bar{X}_n - t; \bar{X}_n + t]\} &= \mathbb{P}\{|\bar{X}_n - \mu| < t\} \\ &= \mathbb{P}\left\{\frac{|\bar{X}_n - \mu|}{\sqrt{\sigma^2/n}} < \frac{t}{\sqrt{\sigma^2/n}}\right\} .\end{aligned}$$

Or d'après le théorème de la limite centrale, si n est grand, la loi de la variable aléatoire $Z = \frac{|\bar{X}_n - \mu|}{\sqrt{\sigma^2/n}}$ est approximativement une loi normale centrée réduite. Soit alors $Z \sim \mathcal{N}(0,1)$, et soit z l'unique valeur telle que

$$\mathbb{P}\{|Z| < z\} = 1 - \alpha .$$

La valeur de z nous est donnée par les tables de la loi normale. Par exemple, si $\alpha = 0.95$, alors $z = 1.96$. Si $\alpha = 0.99$, alors $z = 2.57$. z est donc maintenant connu, et on peut poursuivre le calcul. On a donc :

$$\mathbb{P}\left\{\frac{|\hat{X}_n - \mu|}{\sqrt{\sigma^2/n}} < \frac{t}{\sqrt{\sigma^2/n}}\right\} \geq 1 - \alpha ,$$

avec

$$\frac{t}{\sqrt{\sigma^2/n}} = z .$$

et donc

$$t = z\sqrt{\sigma^2/n} .$$

Au final, on peut parier (avec un risque α d'avoir tort) que

$$\mu \in [\hat{X}_n - z\sqrt{\sigma^2/n}; \hat{X}_n + z\sqrt{\sigma^2/n}] .$$

Si σ^2 est connu, alors le calcul est fini. Malheureusement, on ne connait pas toujours σ^2. On doit donc estimer sa valeur pour trouver l'intervalle de confiance (voir l'application à une loi de Bernoulli).