Chapitre 15

Théorèmes limites en probabilité.

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Toutes les variables aléatoires seront définies sur cet espace probabilisé.

15.1 Loi faible des grands nombres

Définition 15.1 (Moyenne empirique)

Soit $(X_n)_{n\in\mathbb{N}^*}$ une suite de variables aléatoires réelles définies sur une même espace probabilisé.
Pour tout $n \in \mathbb{N}^*$, la variable aléatoire suivante est appelée la moyenne empirique de $X_1, \dots, X_n$:

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

On la note aussi parfois $\overline{X}_n$.

Théorème 15.1 (Loi faible des grands nombres)

Soit $(X_n)_{n\in\mathbb{N}^*}$ une suite de variables aléatoires réelles définies sur une même espace probabilisé, mutuellement indépendantes, admettant une même espérance m. Pour tout $n \in \mathbb{N}^*$, on note

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

Alors

$$\forall \varepsilon > 0, \quad \lim_{n\to+\infty} \mathbb{P}\left(|M_n - m| \geq \varepsilon\right) = 0.$$

Remarque 15.1.

1. De manière plus concrète, la loi des grands nombres signifie que la moyenne de n variables aléatoires indépendantes de même espérance et variance a de grandes chances, si n est grand, d'être une valeur approchée de l'espérance.
2. C'est ce résultat qui justifie les résultats empiriques observés en TP : en simulant des variables aléatoires de même loi de manière indépendante, la moyenne observée tend, dans un certain sens, vers l'espérance lorsque le nombre de simulations tend vers l'infini.

Exemple 15.1 (Exemple important)**.** Soit A un événement et $(X_n)_{n\in\mathbb{N}^*}$ une suite de variables aléatoires de loi $\mathcal{B}(\mathbb{P}(A))$.
Alors
$$\forall \varepsilon > 0, \quad \lim_{n\to+\infty} \mathbb{P}\left(|M_n - \mathbb{P}(A)| \geq \varepsilon\right) = 0.$$

15.2 Convergence en loi

Définition 15.2 (Convergence en loi)

Soit $(X_n)_{n\in\mathbb{N}^*}$ une suite de variables aléatoires réelles et X une variable aléatoire réelle.
On dit que la suite $(X_n)_{n\in\mathbb{N}^*}$ **converge en loi vers** X si
$$\lim_{n\to+\infty} F_{X_n}(x) = F_X(x)$$
en tout réel x où F_X est continue.
On notera alors $X_n \xrightarrow{\mathcal{L}} X$.

Remarque 15.2.

1. **Rappel** : la fonction F_X est continue en $x \in \mathbb{R}$ si et seulement si $P(X = x) = 0$.
2. En particulier, si $X_n \xrightarrow{\mathcal{L}} X$ et si $a < b$ sont deux réels en lesquels F_X est continue alors
$$\lim_{n\to+\infty} P(a < X_n \leq b) = P(a < X \leq b).$$

Méthode 15.1 (Montrer qu'une suite de variables aléatoires converge en loi)

- Si la variable aléatoire X est donnée, pour vérifier si une suite $(X_n)_{n\in\mathbb{N}^*}$ de variables aléatoires converge en loi vers X :
  1. on détermine les fonctions de répartition F_{X_n} et F_X ;
  2. on fixe x et on étudie $\lim\limits_{n\to+\infty} F_{X_n}(x)$;
  3. on vérifie si pour tout x où F_X est continue, la limite ci-dessus vaut $F_X(x)$.
- Pour vérifier si une suite $(X_n)_{n\in\mathbb{N}^*}$ de variables aléatoires converge en loi :
  1. on détermine les fonctions de répartition F_{X_n} ;
  2. on fixe x et on étudie $\lim\limits_{n\to+\infty} F_{X_n}(x)$;
  3. en notant $F(x)$ la limite $\lim\limits_{n\to+\infty} F_{X_n}(x)$, on cherche à vérifier si $x \mapsto F(x)$ coïncident avec une fonction de répartition d'une variable aléatoire X en les points où F_X est continue.

Théorème 15.2 (Critère pour les variables discrètes)

Soit $(X_n)_{n\in\mathbb{N}^*}$ une suite de variables aléatoires **à valeurs dans** $\mathbb{Z}$ et X une variable aléatoire **à valeurs dans** $\mathbb{Z}$.
Alors $(X_n)_{n\in\mathbb{N}^*}$ converge en loi vers X si et seulement si
$$\forall k \in \mathbb{Z}, \quad \lim_{n\to+\infty} P(X_n = k) = P(X = k).$$

Proposition 15.1

Soit $(X_n)_{n\in\mathbb{N}^*}$ telle que pour tout $n > \lambda$, $X_n \hookrightarrow \mathcal{B}\left(n, \frac{\lambda}{n}\right)$ où $\lambda > 0$. Alors $(X_n)_{n\in\mathbb{N}^*}$ converge en loi vers une variable aléatoire suivant la loi de Poisson $\mathcal{P}(\lambda)$.

15.3 Théorème central limite

Théorème 15.3 (Théorème Central Limite (TCL))

Soit $(X_n)_{n\in\mathbb{N}^*}$ une suite de variables aléatoires mutuellement indépendantes, de même loi admettant une variance σ^2 non nulle et une espérance m.
On note, pour tout $n \in \mathbb{N}^*$:

$$M_n^* = \sqrt{n}\left(\frac{M_n - m}{\sigma}\right).$$

Alors $(M_n^*)_{n\in\mathbb{N}^*}$ converge en loi vers une variable aléatoire suivant une loi normale centrée réduite.

Remarque 15.3.

1. La variable M_n^* est la variable centrée réduite associée à la moyenne empirique M_n.
2. Intuitivement, le TCL signifie que pour n suffisamment grand M_n^* suit approximativement une loi $\mathcal{N}(0,1)$.
3. En notant $S_n = X_1 + \cdots + X_n$ on a $M_n = \frac{S_n}{n}$ et

$$S_n^* = \frac{S_n - nm}{\sigma\sqrt{n}} = M_n^*$$

donc $(S_n^*)_{n\in\mathbb{N}^*}$ converge aussi en loi vers une variable aléatoire suivant une loi normale centrée réduite.

4. Intuitivement, le TCL signifie que pour n suffisamment grand S_n suit approximativement une loi $\mathcal{N}(nm, n\sigma^2)$:
Pour tout $x \in \mathbb{R}$

$$\begin{aligned} F_{S_n}(x) = P(S_n \leq x) &= P(S_n - nm \leq x - nm) \\ &= P\left(\frac{S_n - nm}{\sigma\sqrt{n}} \leq \frac{x - nm}{\sigma\sqrt{n}}\right) \\ &= P\left(S_n^* \leq \frac{x - nm}{\sigma\sqrt{n}}\right) \\ &= F_{S_n^*}\left(\frac{x - nm}{\sigma\sqrt{n}}\right) \\ &\simeq F_Z\left(\frac{x - nm}{\sigma\sqrt{n}}\right) \end{aligned}$$

où $Z \hookrightarrow \mathcal{N}(0,1)$. Donc

$$F_{S_n}(x) \simeq F_Z\left(\frac{x - nm}{\sigma\sqrt{n}}\right) = P\left(\sigma\sqrt{n}Z + nm \leq x\right) = F_{\sigma\sqrt{n}Z+nm}(x)$$

avec $\sigma\sqrt{n}Z + nm \hookrightarrow \mathcal{N}(nm, n\sigma^2)$.

Proposition 15.2 (Moivre-Laplace)

Soit $(S_n)_{n\in\mathbb{N}^*}$ une suite de variables aléatoires tel que $S_n \hookrightarrow \mathcal{B}(n,p)$ avec $p \in]0,1[$. Alors la suite

$$\left(\frac{S_n - np}{\sqrt{np(1-p)}}\right)_{n\in\mathbb{N}^*}$$

converge en loi vers une variable aléatoire suivant la loi normale centrée réduite.
On approximera donc la loi $\mathcal{B}(n,p)$ par la loi $\mathcal{N}(np, np(1-p))$ dès que n est suffisamment grand.

Remarque 15.4.

1. L'approximation la loi $\mathcal{B}(n,p)$ par la loi $\mathcal{N}(np, np(1-p))$ est raisonnable dès que $n \geq 20$ et p proche de $\frac{1}{2}$ mais cela peut varier d'un auteur à l'autre.
2. On approxime donc une loi discrète par une loi continue. En pratique, on approxime $P(S_n = k)$ par

$$P(k - 0,5 \leq Z_n \leq k + 0,5) \quad \text{où} \quad Z_n \hookrightarrow \mathcal{N}(np, np(1-p)).$$

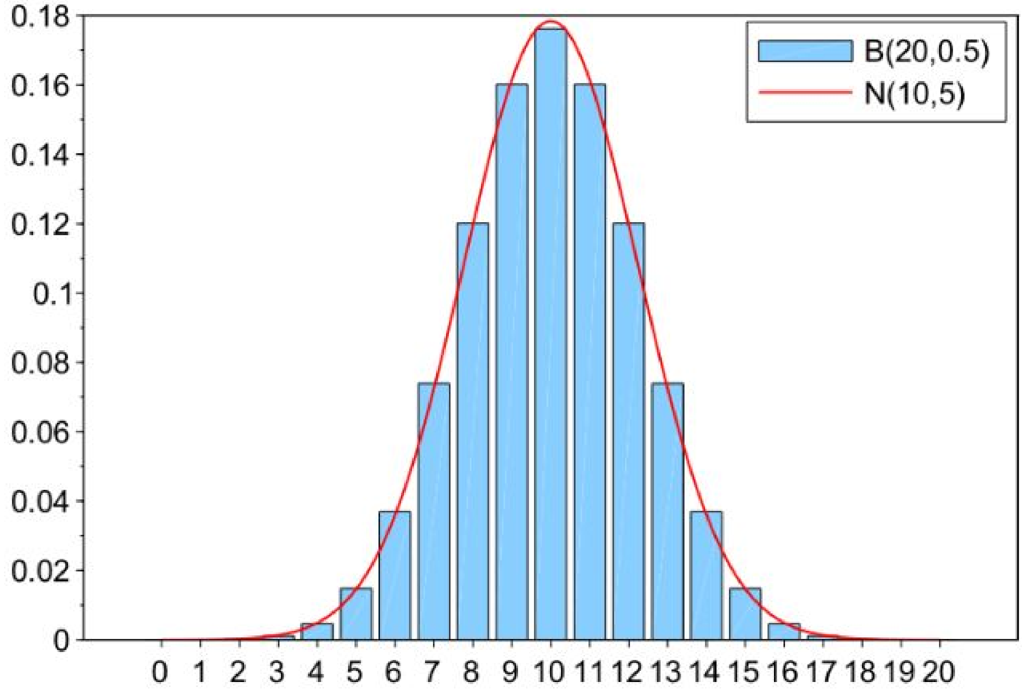

Figure 15.1 – Approximation d'une loi binomiale par une loi normale

Définition 15.3 (Variance empirique)

Soit $X_1, \dots, X_n$ des variables aléatoires.
On appelle **variance empirique** et on note S_n^2 la variable aléatoire définie par :

$$S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - M_n)^2.$$

On appelle **écart-type empirique** la racine carrée de la variance empirique.

Théorème 15.4 (Théorème Central Limite (TCL))

Soit $(X_n)_{n\in\mathbb{N}^*}$ une suite de variables aléatoires mutuellement indépendantes, de même loi admettant une variance σ^2 non nulle et une espérance m.
La suite des variables :

$$\sqrt{n}\left(\frac{M_n - m}{\sqrt{S_n^2}}\right)$$

converge en loi vers une variable aléatoire suivant une loi normale centrée réduite.

15.4 Introduction aux tests

15.4.1 Principe des tests

On considère une population Ω dont les individus possèdent un caractère X (modélisé par une variable aléatoire) dont la loi (ou un certains nombres de paramètres) sont inconnus.

On souhaite formuler une hypothèse sur le(s) paramètre(s) inconnu(s) (par exemple, une hypothèse sur sa valeur) et porter un jugement sur cette hypothèse (est-elle raisonnable ?), en se basant sur l'observation d'un échantillon prélevé.

- **Un test d'hypothèse** est une démarche qui a pour but de fournir une règle de décision permettant, sur la base de résultats d'échantillon, de faire un choix entre deux hypothèses.
- **Hypothèse nulle,** H_0 : c'est l'hypothèse que l'on cherche à tester.
- **Seuil de signification d'un test d'hypothèse** : c'est le risque α consenti à l'avance de rejeter à tort l'hypothèse H_0, alors qu'elle est vraie.

  On utilisera en général $\alpha = 5\%$.
- **Statistique de test ou variable d'échantillonnage** : c'est une variable aléatoire T (en lien avec le problème initial) dont on connait la loi sous l'hypothèse H_0.

  On sépare alors $\mathbb{R}$ en deux zones : une zone de rejet notée R_{rejet} et une zone de non-rejet notée $R_{\text{non-rejet}}$ telles que :

  $$\mathbb{P}(T \in R_{\text{rejet}}) = \alpha \quad ; \quad \mathbb{P}(T \in R_{\text{non-rejet}}) = 1 - \alpha.$$

  -**Utilisation du test** : à partir d'un échantillon, on calcule une valeur observée de T , notée t_{obs}.
  - Si $t_{obs} \notin R_{\text{non-rejet}}$ c'est-à-dire $t_{obs} \in R_{\text{rejet}}$ alors l'hypothèse H_0 est rejetée.
  - Sinon, l'hypothèse H_0 n'est pas rejetée.

Exemple 15.2. On dispose d'une pièce qu'on peut modéliser par une variable aléatoire suivant une loi de Bernoulli de paramètre inconnue p (la probabilité de faire pile par exemple).
On souhaite déterminer si notre pièce est équilibrée ou non c'est-à-dire si $p = \frac{1}{2}$.

— L'hypothèse nulle est $H_0 : p = \frac{1}{2}$.
— La stratégie intuitive est de lancer la pièce un grand nombre de fois et tester si la proportion de pile est d'environ $\frac{1}{2}$.
Chaque lancer est modéliser par une variable X_i de loi $\mathcal{B}(p)$ et la proportion de pile sur n tirages est alors la moyenne empirique M_n : c'est notre statistique de test.
Rmq : le choix de M_n comme statistique de test est justifié par la loi des grands nombres.
— Il reste à déterminer, pour α donné, une zone de non-rejet .

15.4.2 Test de conformité à la moyenne

Proposition 15.3

Soit $(X_n)_{n\in\mathbb{N}^*}$ une suite de variables aléatoires mutuellement indépendantes, de même loi admettant une variance σ^2 non nulle et une espérance m.
Soit $\alpha \in]0, 1[$ et notons Φ la fonction de répartition de la loi normale centrée réduite.
Alors

1. Il existe un réel t_α tel que $\Phi(t_\alpha) = 1 - \frac{\alpha}{2}$.
2. $\lim\limits_{n\to+\infty} \mathbb{P}\left(M_n \in \left[m - t_\alpha\sqrt{\frac{S_n^2}{n}}; m + t_\alpha\sqrt{\frac{S_n^2}{n}}\right]\right) = 1 - \alpha$.

Remarque 15.5. Pour $\alpha = 0.05$, $t_\alpha \simeq 1.96$.

Application : test de conformité à la moyenne On considère une population Ω dont les individus possèdent un caractère X (modélisé par une variable aléatoire) dont la loi (ou un certains nombres de paramètres) sont inconnus.

On souhaite formuler une hypothèse sur la valeur de la moyenne m de X, en se basant sur l'observation d'un échantillon prélevé.

— L'hypothèse nulle H_0 est de la forme $m = m_0$ où m_0 est la valeur hypothétique de la moyenne.
— L'échantillon prélevé est modélisé par une suite de variables $(X_n)_{n\in\mathbb{N}^*}$ indépendantes et de même loi que X.
— La statistique de test est alors la moyenne empirique M_n (moyenne des valeurs mesurées sur l'échantillon).
— Si n est suffisamment grand alors :

$$\mathbb{P}\left(M_n \in \left[m_0 - t_\alpha\sqrt{\frac{S_n^2}{n}}; m_0 + t_\alpha\sqrt{\frac{S_n^2}{n}}\right]\right) \simeq 1 - \alpha.$$

L'intervalle $\left[m_0 - t_\alpha\sqrt{\frac{S_n^2}{n}}; m_0 + t_\alpha\sqrt{\frac{S_n^2}{n}}\right]$ est donc une zone de non-rejet au seuil (environ) α.

— Décision :
  — si $M_n \in \left[m_0 - t_\alpha\sqrt{\frac{S_n^2}{n}}; m_0 + t_\alpha\sqrt{\frac{S_n^2}{n}}\right]$ on ne rejette pas l'hypothèse $H_0 : m = m_0$;
  — si $M_n \notin \left[m_0 - t_\alpha\sqrt{\frac{S_n^2}{n}}; m_0 + t_\alpha\sqrt{\frac{S_n^2}{n}}\right]$ on rejette l'hypothèse $H_0 : m = m_0$.