Chapitre 12 : Estimation, convergence et approximation

I Estimation ponctuelle

Le but de l'estimation est de déterminer une **valeur approchée** d'une caractéristique **inconnue** d'une population, que l'on notera $g(\theta)$ dans le cours (par exemple : $g(\theta)$ est la taille moyenne de la population mondiale, ou $g(\theta)$ est la proportion de la population français qui votera pour le candidat A...).
Il est souvent impossible de déterminer la valeur exacte de $g(\theta)$, car on ne peut pas étudier tous les individus par manque de temps ou d'argent...etc... donc on se contente de faire un sondage ou un recensement c'est-à-dire d'étudier un échantillon de n individus extraits de façon aléatoire de la population, avec remise, et de trouver un réel appelé estimateur qui donne une bonne approximation de la valeur exacte de $g(\theta)$.
Le but de ce chapitre n'est pas de mettre en pratique l'estimation mais de faire une étude préalable d'un estimateur choisi pour décider si c'est un bon ou un mauvais estimateur.

1 La modélisation mathématique

On considère une variable aléatoire X dont on ne connait pas tout à fait la loi .
Plus précisément, on sait que la loi P de X appartient à une famille de lois $(P_\theta)_{\theta\in\Theta}$ dépendant d'un paramètre θ.

Le but est de déterminer, par l'étude d'expériences répétées, une valeur approchée de θ, ou plus généralement d'une valeur $g(\theta)$.

Exemple
Le sondage : On cherche à estimer la proportion θ de la population française qui votera pour le candidat A aux prochaines élections.

On pose l'expérience aléatoire suivante : on choisit une personne au hasard dans la population, et on lui demande si elle va voter pour A.
On pose alors $X = \begin{cases} 1 \text{ si elle votera pour } A. \\ 0 \text{ si elle ne votera pas pour } A. \end{cases}$.
Alors $X \hookrightarrow \mathcal{B}(\theta)$; le sondage est une estimation du paramètre θ d'une loi de Bernoulli.

Le problème de l'estimation consiste à évaluer (estimer) la valeur du paramètre $g(\theta)$ à partir d'un échantillon de données $x_1, x_2, ..., x_n$ obtenues en observant n fois le même phénomène.
On fait l'hypothèse que cet échantillon de données constitue une réalisation d'une suite $(X_1, X_2, ..., X_n)$, où les X_i sont des variables aléatoires réelles, indépendantes, de même loi que X.

Définition 1 : n-échantillon

On se place dans les conditions précédemment énoncées.
*Un n-**échantillon de la loi P de X** est une liste $(X_1, X_2, ...X_n)$ de n variables aléatoires indépendantes et de même loi que X (identiquement distribuées).*
On appelle échantillon observé la réalisation observée x_1, x_2, ...x_n de ces n variables aléatoires.

Définition 2 : estimateur

*Un **estimateur** de $g(\theta)$ est une variable aléatoire de la forme $T_n = \Phi(X_1, X_2, ..., X_n)$ où $(X_1, X_2, ..., X_n)$ est un n-échantillon de la loi de X.*
La réalisation $\Phi(x_1, x_2, ...x_n)$ de l'estimateur T_n est appelée l'estimation de $g(\theta)$. Elle ne dépend que de l'échantillon $(x_1, x_2, ..., x_n)$ observé.

Le but de cette partie est de déterminer si un estimateur choisi T_n de θ est un bon estimateur ou un mauvais estimateur de θ. Pour le moment, nous allons le décider uniquement à l'aide des notions de biais et de risque quadratique de l'estimateur.
Ces notions permettent en réalité plutôt de comparer deux estimateurs et de décider lequel est le meilleur

plutôt que de faire l'étude intrinsèque d'un estimateur qui sera plutôt faites en partie III (intervalle de confiance).

2 Etude qualitative d'un estimateur

Soit T_n un estimateur de $g(\theta)$.

a) Biais de T_n

Définition 3 : biais

*Si T_n a une espérance, on appelle **biais** de T_n la quantité :*

$$\boxed{b_\theta(T_n) = E_\theta(T_n) - g(\theta)}$$

Cela correspond à l'écart entre l'espérance de T_n et la valeur $g(\theta)$ à estimer.
*Lorsque, pour tout θ de Θ, $E_\theta(T_n) = g(\theta) \Leftrightarrow b_\theta(T_n) = 0$, on dit que T_n est un **estimateur sans biais** de $g(\theta)$.*

Remarque

T_n est un estimateur bien choisi censé donner une valeur approchée de $g(\theta)$. On espère donc qu'au moins en moyenne T_n vale $g(\theta)$ c'est-à-dire que $E_\theta(T_n) = g(\theta)$: d'où la notion d'estimateur sans biais.
Si l'estimateur n'est pas sans biais, on espère au moins que son biais soit très faible.

Définition 4 : estimateur asymptotiquement sans biais

*Une suite $(T_n)_{n\geq 1}$ d'estimateurs de $g(\theta)$ est dite **asymptotiquement sans biais** si pour tout θ de Θ, on a :*

$$\lim_{n\to+\infty} E_\theta(T_n) = g(\theta) \Leftrightarrow \lim_{n\to+\infty} b_\theta(T_n) = 0$$

b) Risque quadratique de T_n

Définition 5

*Si, pour tout θ de Θ, T_n admet un moment d'ordre 2, on appelle **risque quadratique** de T_n la quantité :*

$$\boxed{r_\theta(T_n) = E_\theta\left[(T_n - g(\theta))^2\right]}$$

C'est la moyenne des carrés de la distance de T_n à $g(\theta)$.
Si le risque quadratique est très faible, c'est qu'en moyenne la distance de T_n à $g(\theta)$ est très faible donc que T_n est un bon estimateur de $g(\theta)$.

Remarques

1. Il suffit que T_n ait un moment d'ordre 2 car $(T_n - g(\theta))^2 = T_n^2 - 2g(\theta)T_n + g(\theta)^2$.
   Donc si T_n a un moment d'ordre 2 c'est-à-dire $E_\theta(T_n^2)$ existe, alors $E_\theta(T_n)$ existe et donc par linéarité de l'espérance, $E_\theta(T_n^2 - 2g(\theta)T_n + g(\theta)^2)$ existe.
2. Le risque quadratique permet surtout de comparer deux estimateurs de $g(\theta)$: celui ayant le plus faible risque quadratique sera plus précis, donc meilleur.

Théorème 1

Soit T_n un estimateur admettant un risque quadratique, alors on a l'égalité :

$$\boxed{r_\theta(T_n) = V_\theta(T_n) + \left[b_\theta(T_n)\right]^2}$$

Preuve

Remarque : par soucis de lisibilité, la notation "indice θ" est omise dans cette preuve.

$$\begin{aligned}
r(T_n) &= E\left[\left(T_n - g(\theta)\right)^2\right] && \textit{par définition du risque quadratique} \\
&= E\left[\left(T_n - E(T_n) + E(T_n) - g(\theta)\right)^2\right] \\
&= E\left[\left(T_n - E(T_n) + b(T_n)\right)^2\right]
\end{aligned}$$

$$\begin{aligned}
&= E\Big[\big(T_n - E(T_n)\big)^2 + \big(b(T_n)\big)^2 + 2\big(T_n - E(T_n)\big)b(T_n)\Big] && \textit{par identité remarquable}\\
&= E\Big[\big(T_n - E(T_n)\big)^2\Big] + E\Big[\big(b(T_n)\big)^2\Big] + 2E\Big[\big(T_n - E(T_n)\big)b(T_n)\Big] && \textit{par linéarité de l'espérance}\\
&= E\Big[\big(T_n - E(T_n)\big)^2\Big] + \big(b(T_n)\big)^2 + 2b(T_n)E\Big[T_n - E(T_n)\Big] && \textit{car } b(T_n) \textit{ est un scalaire}\\
&= V(T_n) + (b(T_n))^2 + 2b(T_n)(E(T_n) - E(E(T_n))) && \textit{par linéarité de l'espérance}\\
&= V(T_n) + (b(T_n))^2 + 2b(T_n)(E(T_n) - E(T_n)) && \textit{car } E(T_n) \textit{ est un scalaire}\\
&= V(T_n) + (b(T_n))^2
\end{aligned}$$

c) Estimateur convergeant

Définition 6 : estimateur convergeant

Une suite $(T_n)_{n\geq 1}$ d'estimateurs de $g(\theta)$ est dite ***convergente*** *si la suite (T_n) converge en probabilité vers $g(\theta)$.*

Propriété 2 : condition suffisante de convergence d'un estimateur

Si pour tout θ de Θ, on a : $\lim\limits_{n\to+\infty} r_\theta(T_n) = 0$, alors la suite d'estimateurs $(T_n)_{n\geq 1}$ de $g(\theta)$ est convergente.

Preuve

On sait d'après l'inégalité de Markov que pour tout variable aléatoire Y admettant une espérance, on a : $\forall a > 0$, $P(Y \geq a) \leq \frac{E(Y)}{a}$.
Choisissons la variable aléatoire $Y = (T_n - g(\theta))^2$ et pour tout $\varepsilon > 0$, choisissons $a = \varepsilon^2$. On a alors : $\forall \varepsilon > 0$, $P((T_n - g(\theta))^2 \geq \varepsilon^2) \leq \frac{E\left[(T_n - g(\theta))^2\right]}{\varepsilon^2} = \frac{r(T_n)}{\varepsilon^2}$.
Or $P((T_n - g(\theta))^2 \geq \varepsilon^2) = P(|T_n - g(\theta)| \geq \varepsilon)$ en composant par la racine carrée donc :

$$\forall \varepsilon > 0, \quad P(|T_n - g(\theta)| \geq \varepsilon) \leq \frac{r(T_n)}{\varepsilon^2}$$

Supposons maintenant que pour tout θ de Θ, on a : $\lim\limits_{n\to+\infty} r(T_n) = 0$, alors , par théorème d'encadrement, l'inégalité précédente donne :

$$\forall \varepsilon > 0, \quad \lim_{n\to+\infty} P(|T_n - g(\theta)| \geq \varepsilon) = 0$$

c'est-à-dire que la suite d'estimateurs (T_n) est convergente car converge en probabilité vers $g(\theta)$.

Remarque

Si le risque quadratique d'un estimateur est convergeant, on peut affirmer que c'est un estimateur correct de $g(\theta)$.

Exercice 1 (: estimateur de Monte-Carlo pour une proportion)**.**
On souhaite estimer la proportion p d'électeurs votant pour le candidat A.
Pour un individu choisi au hasard, on note $X = \begin{cases} 1 \text{ s'il votera pour } A. \\ 0 \text{ s'il ne votera pas pour } A. \end{cases}$.

Alors $X \hookrightarrow \mathcal{B}(p)$; où p est l'inconnue à estimer.
On considère alors un n-échantillon $(X_1, X_2, ...X_n)$ de la loi de X .

1. Que représente $S_n = \sum\limits_{k=1}^{n} X_k$? Donner la loi de S_n.
2. Un estimateur naturel de p est la proportion T_n des personnes de l'échantillon votant pour A. Exprimer T_n en fonction de S_n et de n.
3. T_n est-il un estimateur sans biais de p ?
4. Déterminer le risque quadratique de T_n . T_n est-il un estimateur convergeant ?

Exercice 2 (: Etude générale de l'estimateur de Monte-Carlo)**.**
On cherche à estimer la taille moyenne θ d'une population. La taille d'un individu choisi au hasard dans la population est donc une variable aléatoire X d'espérance θ inconnue, que l'on suppose admettre une variance σ^2 connue.
On considère un n-échantillon $(X_1, X_2, ...X_n)$ de la loi de X (c'est-à-dire que l'on a fait un recensement de n personnes choisies au hasard et avec remise dans la population et que l'on a noté X_1,X_2, ...X_n leur taille respective).

1. Quel est l'estimateur le plus naturel de θ ? On notera $\overline{X_n}$ cet estimateur.
   Nous allons estimer la qualité de l'estimateur $\overline{X_n}$:

2. Donner la valeur de $E(\overline{X_n})$ en fonction de θ. $\overline{X_n}$ est-il un estimateur sans biais ?
3. Déterminer le risque quadratique de $\overline{X_n}$. $\overline{X_n}$ est-il un estimateur convergeant ?

Exercice 3 (: Comparaison de plusieurs estimateurs)**.**

Soit f définie par $f(t) = \begin{cases} 0, & \forall t \notin [0, R] \\ \frac{2t}{R^2}, & \forall t \in [0, R] \end{cases}$. On cherche à déterminer l'inconnu $R > 0$.

1. Montrer que f est une densité de probabilité. Dans la suite, X est une variable aléatoire de densité f. Déterminer sa fonction de répartition F.
2. Calculer l'espérance et la variance de X.

   Dans la suite, on considère un n-échantillon $(X_1, X_2, ..., X_n)$ de la loi de X.
3. On note $T_n = \frac{3}{2n} \sum_{i=1}^{n} X_i$. Montrer que T_n est un estimateur sans biais de R et calculer son risque quadratique $r(T_n)$.
4. Soit $M_n = Max(X_1, X_2, ..., X_n)$.
   (a) Montrer que pour tout x réel, $P(M_n \leq x) = (F(x))^n$. En déduire que M_n est une variable aléatoire à densité.
   (b) déterminer une densité g_n de M_n.
   (c) Montrer que $E(M_n) = \frac{2n}{2n+1} R$ et que $V(M_n) = \frac{n}{(n+1)(2n+1)^2} R^2$.
   (d) On cherche à estimer R avec M_n. L'estimateur M_n est-il sans biais ? Asymptotiquement sans biais ?
   (e) Déterminer un équivalent simple de $r(M_n)$ quand $n \to +\infty$.
   (f) Quels sont les avantages et les inconvénients respectifs des estimateurs T_n et M_n ?
   (g) Donner un estimateur sans biais W_n de R, proportionnel à M_n (c'est-à-dire de la forme $\lambda_n M_n$ où λ_n est un réel non aléatoire). Calculer son risque quadratique. Conclure.

Exercice 4 (: estimateur du maximum de vraisemblance)**.** Dans cet exercice, θ désigne un réel strictement positif et n un entier naturel supérieur ou égal à 2.

Pour tout k de $\mathbb{N}$, on pose : $u_k = \frac{1}{1+\theta} \left(\frac{\theta}{1+\theta}\right)^k$.

1. Montrer que la suite $(u_k)_{k \in \mathbb{N}}$ définit une loi de probabilité.

On considère maintenant une variable aléatoire X prenant ses valeurs dans $\mathbb{N}$ et dont la loi est donnée par :

$$\forall k \in \mathbb{N},\ \mathrm{P}(X = k) = u_k.$$

2. On pose $Y = X + 1$. Reconnaître la loi de Y, puis en déduire l'espérance et la variance de X.
3. Dans cette question, on souhaite estimer le paramètre θ par la méthode du maximum de vraisemblance. Pour ce faire, on considère un échantillon $(X_1, X_2, \ldots, X_n)$ composé de variables aléatoires indépendantes ayant toutes la même loi que X et on introduit L, de $\mathbb{R}_+^*$ dans $\mathbb{R}$, définie par :

$$\forall \theta \in \mathbb{R}_+^*,\ L(\theta) = \prod_{k=1}^{n} \mathrm{P}(X_k = x_k)$$

   où $x_1, x_2, \ldots, x_n$ désignent des entiers naturels éléments de $X(\Omega)$.

   L'objectif est de choisir la valeur de θ qui rend $L(\theta)$ maximale.

   (a) Écrire $\ln\big(L(\theta)\big)$ en fonction de θ et de $S_n = \sum_{k=1}^{n} x_k$.
   (b) On considère la fonction φ définie par :

$$\forall \theta \in]0; +\infty[,\ \varphi(\theta) = S_n \ln \theta - (S_n + n) \ln(1 + \theta)$$

   Montrer que la fonction φ admet un maximum, atteint en un seul réel que l'on notera $\widehat{\theta_n}$ et que l'on exprimera en fonction de S_n. Que représente $\widehat{\theta_n}$ pour la fonction L ?

4. On pose dorénavant : $T_n = \frac{1}{n} \sum_{i=1}^{n} X_i$. La variable T_n est appelée estimateur du maximum de vraisemblance pour θ.
   (a) Vérifier que T_n est un estimateur sans biais de θ.
   (b) Calculer le risque quadratique $r_{T_n}(\theta)$ de T_n et vérifier que $\lim_{n \to +\infty} r_{T_n}(\theta) = 0$.

Nous souhaiterions estimer un paramètre θ plus précisément, c'est-à-dire calculer la probabilité qu'il soit bien compris entre 2 estimateurs donnés par exemple (notion d'intervalle de confiance). Mais les probabilités sont trop difficiles à calculer car il est compliqué de calculer la loi d'une fonction T_n de n variables aléatoires $(X_1, ...X_n)$. C'est pourquoi, dans une seconde partie, nous allons apprendre à approximer des probabilités compliquées par des probabilités plus simples à calculer.

II Convergences et calculs approchés de probabilités

1 Majoration d'une probabilité par l'inégalité de Bienaymé-Tchebychev

a) L'inégalité de Markov

Propriété 3 : Inégalité de Markov

Soit Y une variable aléatoire réelle positive admettant une espérance. On a :

$$\boxed{\forall a > 0,\ P(Y \geq a) \leq \frac{E(Y)}{a}}$$

Preuve

Montrons que l'on a :

$$\forall a > 0, \quad E(Y) \geq aP(Y \geq a)$$

pour toute variable aléatoire Y positive, à densité ou discrète, admettant une espérance :

• 1) Cas d'une variable à densité :

Soit Y une variable aléatoire à densité, positive, admettant une espérance. Notons f une densité de Y.
On rappelle pour la suite que $P(Y \geq a) = P(Y \in [a; +\infty[) = \int_a^{+\infty} f(t)dt$. On a :

$$\begin{aligned}
E(Y) &= \int_{-\infty}^{+\infty} tf(t)dt && \textit{par définition de l'espérance} \\
&= \int_0^{+\infty} tf(t)dt && \textit{car } Y(\Omega) \subset \mathbb{R}_+ \\
&= \int_0^{a} tf(t)dt + \int_a^{+\infty} tf(t)dt && \textit{par relation de Chasles} \\
&\geq \int_a^{+\infty} tf(t)dt && \textit{car si } t \in [0; a], \textit{ on a } t \geq 0 \textit{ et } f(t) \geq 0 \textit{ (densité)} \\
&\geq \int_a^{+\infty} af(t)dt && \textit{car si } t \in [a; +\infty[\textit{ alors } t \geq a \textit{ puis par multiplication par } f(t) \geq 0 \textit{ et par croissance des bornes} \\
&= a\int_a^{+\infty} f(t)dt && \textit{par linéarité de l'intégration} \\
&= aP(Y \geq a)
\end{aligned}$$

• 2) Cas d'une variable discrète lorsque $a \in Y(\Omega)$:

Soit Y une variable aléatoire discrète positive admettant une espérance. On admet que $Y(\Omega) \subset \mathbb{N}$ pour simplifier la preuve.
On rappelle pour la suite que si $a \in \mathbb{N}$, $P(Y \geq a) = \sum_{k=a}^{+\infty} P(Y = k)$. On a :

$$\begin{aligned}
E(Y) &= \sum_{k=0}^{+\infty} kP(Y = k) && \textit{par définition de l'espérance} \\
&= \sum_{k=0}^{a-1} kP(Y = k) + \sum_{k=a}^{+\infty} kP(Y = k) && \textit{en séparant la somme} \\
&\geq \sum_{k=a}^{+\infty} kP(Y = k) && \textit{car } k \in \mathbb{N} \textit{ et } P(Y = k) \geq 0 \textit{ (proba)} \\
&\geq \sum_{k=a}^{+\infty} aP(Y = k) && \textit{car si } k \in [|a; +\infty|[\textit{ alors } k \geq a \textit{ puis par multiplication par } P(Y = k) \geq 0 \textit{ et par sommation des inégalités} \\
&= a\sum_{k=a}^{+\infty} P(Y = k) && \textit{par linéarité de la somme} \\
&= aP(Y \geq a)
\end{aligned}$$

b) L'inégalité de Bienaymé-Tchebychev

L'inégalité de Tchebychev majore la probabilité de dispersion d'une variable aléatoire autour de sa moyenne :

Théorème 4 : Inégalité de Bienaymé-Tchebychev

Soit X une variable aléatoire admettant un moment d'ordre 2. On a :

$$\boxed{\forall \varepsilon > 0,\ P(|X - E(X)| \geq \varepsilon) \leq \frac{V(X)}{\varepsilon^2}}$$

$P(|X - E(X)| \geq \varepsilon)$ est la probabilité pour que X prenne des valeurs éloignées de $E(X)$ d'au moins ε. Cette probabilité est d'autant plus faible que $V(X)$ est petit et que ε est grand.

Preuve

Pour tout $\varepsilon > 0$, on utilise l'inégalité de Markov en choisissant $a = \varepsilon^2$ et $Y = (X - E(X))^2$ (on remarque qu'on peut utiliser Markov car $Y \geq 0$ et comme X a une variance alors Y a une espérance) :

$$\forall \varepsilon > 0,\ P\big[(X - E(X))^2 \geq \varepsilon^2\big] \leq \frac{E\big[(X - E(X))^2\big]}{\varepsilon^2} = \frac{V(X)}{\varepsilon^2}$$

Or, en composant l'inégalité par la fonction racine carrée, croissante, on a : $P\big[(X - E(X))^2 \geq \varepsilon^2\big] = P\big[|X - E(X)| \geq \varepsilon\big]$. Ainsi,

$$\forall \varepsilon > 0,\ P\big[|X - E(X)| \geq \varepsilon\big] \leq \frac{V(X)}{\varepsilon^2}$$

Exercice 5. Soit X une variable aléatoire suivant une loi normale centrée réduite.

1. A l'aide de l'inégalité de Bienaymé-Tchebychev, démontrer que pour tout $x > 0$, $P(|X| \leq x) \geq 1 - \frac{1}{x^2}$.
2. Démontrer que pour tout $x > 0$, $P(|X| \leq x) = 2\Phi(x) - 1$.
3. En déduire l'inégalité $\int_{-\infty}^{x} e^{-\frac{t^2}{2}}\,dt \geq \sqrt{2\pi}\left(1 - \frac{1}{2x^2}\right)$.

2 Convergence en probabilités

Pour éclairer la signification de la loi faible des grands nombres, nous donnons la définition de la convergence en probabilité d'une suite de variables aléatoires.

Définition 7 : convergence en probabilités

Une suite (X_n) de variables aléatoires réelles ***converge en probabilités*** *vers un réel m si :*

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

Notation : $X_n \xrightarrow{\mathbb{P}} m$

En pratique, cela signifie que pour tout $\omega \in \Omega$ (presque sûrement) , la suite $(X_n(\omega))$ converge vers m au sens de la convergence usuelle d'une suite de réels.

3 Loi faible des grands nombres

Théorème 5 : loi faible des grands nombres

Soit (X_n) une suite de variables aléatoires indépendantes et identiquement distribuées (c'est-à-dire de même loi), admettant une espérance commune m et une variance commune σ^2.

On pose $\overline{X_n} = \frac{1}{n}\sum\limits_{k=1}^{n} X_k = \frac{X_1 + \cdots + X_n}{n}$ la moyenne des n premières variables X_k.

La loi faible des grands nombres affirme que la suite $(\overline{X_n})$ converge en probabilité vers m.

Preuve

Nous allons appliquer l'inégalité de Bienaymé-Tcheychev à $\overline{X_n}$ en remarquant qu'elle a forcément une espérance et une variance :

- $\overline{X_n}$ a une espérance par linéarité de l'espérance, donnée par :

$$E(\overline{X_n}) = E(\frac{1}{n}\sum_{k=1}^{n} X_k) = \frac{1}{n}\sum_{k=1}^{n} E(X_k) = \frac{1}{n}\sum_{k=1}^{n} m = \frac{1}{n} nm = m$$

- $\overline{X_n}$ a une variance grâce à l'indépendance des variables X_1, X_2, ..., X_n donnée par :

$$V(\overline{X_n}) = \frac{1}{n^2} V(\sum_{k=1}^{n} X_k) = \frac{1}{n^2}\sum_{k=1}^{n} V(X_k) = \frac{1}{n^2}\sum_{k=1}^{n} \sigma^2 = \frac{1}{n^2} n\sigma^2 = \frac{\sigma^2}{n}$$

L'inégalité de Bienaymé-Tchebychev donne donc :

$$\forall \varepsilon > 0,\ P(|\overline{X_n} - E(\overline{X_n})| \geq \varepsilon) \leq \frac{V(\overline{X_n})}{\varepsilon^2}$$

i.e.

$$\forall \varepsilon > 0,\ P(|\overline{X_n} - m)| \geq \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2} \underset{n\to+\infty}{\to} 0$$

Par définition de la convergence en probabilités, on obtient donc que $\overline{X_n}$ converge en probabilité vers m.

Exemple

La loi faible des grands nombres nous indique donc que l'estimateur de Monte-Carlo $\overline{X_n} = \frac{1}{n}\sum_{k=1}^{n} X_k$ obtenu à partir d'un n-échantillon de la loi de X converge en probabilité vers $m = E(X)$. C'est donc un "bon" estimateur de m.

4 Convergence en loi

a) Cas général : Fonctions de répartitions

Définition 8

Soit (X_n) une suite de variables aléatoires réelles et X une variable aléatoire, définies sur ***le même espace probabilisé****.*
On dit que la suite (X_n) converge en loi vers la variable X si et seulement si, en tout point $x \in \mathbb{R}$ tel que F_X est continue en x on a :

$$\boxed{\lim_{n\to+\infty} F_{X_n}(x) = F_X(x)}$$

Notation : $X_n \xrightarrow{\mathcal{L}} X$

Exercice 6. Soit $(X_n)_{n\geq 1}$ une suite de v.a.r. définies sur un même espace probabilisé, indépendantes, suivant une loi uniforme sur $[0,1]$. Pour $n \geq 1$, on définit :

$$M_n = max(X_1, X_2, ..., X_n) \text{ et } Y_n = n(1 - M_n)$$

1. Soit x un réel fixé. Calculer $\lim_{n\to+\infty} \left(1 - \frac{x}{n}\right)^n$.
2. Déterminer la fonction de répartition de M_n puis celle de Y_n.
3. En distinguant les cas $x < 0$ et $x \geq 0$, Montrer que la suite (Y_n) converge en loi vers une variable remarquable.

b) Cas des variables discrètes

Définition 9

Soit (X_n) une suite de variables aléatoires discrètes, X une variable aléatoire discrète, et $\mathcal{K}$ (en pratique $\mathcal{K} = \mathbb{N}$ ou $\mathcal{K} = \mathbb{Z}$) un ensemble discret tels que pour tout $n \in \mathbb{N}$, $X_n(\Omega) \subset \mathcal{K}$ et $X(\Omega) \subset \mathcal{K}$.
Alors $(X_n) \xrightarrow{\mathcal{L}} X$ si et seulement si :

$$\boxed{\forall k \in \mathcal{K},\ \lim_{n\to+\infty} P(X_n = k) = P(X = k)}$$

Exercice 7. Soit $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires telles que X_n suit la loi binomiale de paramètres $(n, \frac{\lambda}{n})$.
Soit $k \in \mathbb{N}$ fixé et $n \geq k$.

1. Donner l'expression de $P(X_n = k)$.
2. Ecrire $\frac{n!}{(n-k)!}$ sous forme d'un produit de k facteurs puis en déduire un équivalent simple de $\frac{n!}{(n-k)!}$ lorsque n tend vers $+\infty$.
3. En déduire un équivalent de $P(X_n = k)$ lorsque n tend vers $+\infty$ qui ne dépend plus de n.
4. Montrer que la suite (X_n) converge en loi vers une variable remarquable.

Exercice 8. On considère une suite de variables aléatoires $(X_n)_{n\in\mathbb{N}^*}$ et une variable Y toutes définies sur un même espace probabilisé.
Pour tout $n \in \mathbb{N}^*$, X_n suit la loi uniforme sur $[|1, n|]$.
Y suit la loi uniforme sur $[0,1]$.
On pose $Y_n = \frac{X_n}{n}$. Montrons que (Y_n) converge en loi vers Y.

1. Déterminer la loi de Y_n.

2. Montrer que $\lim\limits_{n\to+\infty} F_{Y_n}(x) = F_Y(x)$ pour tout $x < 0$ puis pour tout $x > 1$.
3. (a) Soit $n \in \mathbb{N}^*$ et soit $k \in [|0, n|]$, rappeler $P(X_n \leq k)$.
   (b) Montrer que pour tout nombre réel x, $\lim\limits_{n\to+\infty} \frac{\lfloor nx \rfloor}{n} = x$.
   (c) Soit $x \in [0, 1]$. Montrer que $\lim\limits_{n\to+\infty} F_{Y_n}(x) = F_Y(x)$.
4. Conclure.

5 Théorème limite central (ou théorème de la limite centrée)

Soit (X_n) une suite de variables aléatoires indépendantes et de même loi, admettant une espérance commune m et une variance commune σ^2.
Soit alors $\overline{X_n} = \frac{1}{n}\sum\limits_{k=1}^{n} X_k$ la moyenne empirique des n premières variables.
On remarque alors que :
• $E(\overline{X_n}) = \frac{1}{n}\sum\limits_{k=1}^{n} E(X_k) = \frac{1}{n}\sum\limits_{k=1}^{n} m = \frac{1}{n}nm = m$ par linéarité de l'espérance et ,
• $V(\overline{X_n}) = \frac{1}{n^2}V(\sum\limits_{k=1}^{n} X_k) = \frac{1}{n^2}\sum\limits_{k=1}^{n} V(X_k) = \frac{1}{n^2}n\sigma^2 = \frac{\sigma^2}{n}$ par indépendance des variables aléatoires donc $\sigma(\overline{X_n}) = \frac{\sigma}{\sqrt{n}}$.

Théorème 6 : T L C (admis)

Soit (X_n) une suite de variables aléatoires indépendantes et de même loi, admettant une espérance commune m et une variance commune σ^2.
Soit alors $\overline{X_n} = \frac{1}{n}\sum\limits_{k=1}^{n} X_k$ la moyenne empirique des n premières variables.
On pose ensuite $\overline{X_n}^* = \frac{\overline{X_n} - E(\overline{X_n})}{\sigma(\overline{X_n})} = \sqrt{n}\frac{\overline{X_n} - m}{\sigma}$ la variable centrée réduite associée à $\overline{X_n}$.

Alors la suite $(\overline{X_n}^*)$ converge en loi vers une variable X suivant **la loi normale centrée réduite $\mathcal{N}_{0,1}$**.

On a donc pour tous $a < b \in \mathbb{R}$: $P(a \leq \overline{X_n}^* \leq b) \xrightarrow[n\to+\infty]{} \frac{1}{\sqrt{2\pi}}\int_a^b e^{-\frac{t^2}{2}}\, dt = \Phi(b) - \Phi(a)$.

Remarques

1. Cette convergence est très rapide : on verra qu'à partir de $n = 30$, on pourra approximer avec une excellente précision la loi de $\overline{X_n}^*$ par $\mathcal{N}_{0,1}$ et celle de $\overline{X_n}$ par $\mathcal{N}_{m,\sigma^2/n}$.
2. T.L.C bis : La somme centrée réduite des variables X_i converge également en loi vers une normale centrée réduite : en effet, si on pose $S_n = \sum\limits_{k=1}^{n} X_k$ alors :

$$S_n^* = \frac{S_n - E(S_n)}{\sigma(S_n)} = \frac{S_n - nm}{\sqrt{n}\sigma} = \frac{\frac{S_n - nm}{n}}{\frac{\sqrt{n}\sigma}{n}} = \frac{\overline{X_n} - m}{\frac{\sigma}{\sqrt{n}}} = \overline{X_n}^*$$

Exercice 9. Ecrire un programme informatique qui simule une variable X suivant la loi normale centrée réduite à l'aide de la simulation de 30 variables uniformes sur $[0, 1]$ et indépendantes.

Exercice 10. Le but de cet exercice est d'étudier la suite $(u_n)_{n\geq 1}$ définie par : $\forall n \geq 1$, $u_n = e^{-n}\sum\limits_{i=0}^{n} \frac{n^i}{i!}$.
Soit (X_n) une suite de v.a.r. indépendantes suivant la même loi de Poisson de paramètre 1.

1. On pose : $S_n = \sum\limits_{i=1}^{n} X_i$. Donner la loi de S_n, son espérance et sa variance.
2. Exprimer u_n à l'aide de S_n et vérifier que : $\forall n \geq 1$, $u_n = P(S_n^* \leq 0)$, où S_n est la variable centrée réduite associée à S_n.
3. En appliquant le théorème de la limite centrée, déterminer $\lim\limits_{n\to+\infty} u_n$ ainsi qu'un équivalent de $\sum\limits_{i=0}^{n} \frac{n^i}{i!}$ au voisinage de $+\infty$.

6 Approximations

De façon générale, lorsqu'une suite (X_n) converge en loi vers une variable X, on peut approximer la loi de X_n par celle de X dès que n est suffisamment grand.
Le seul principe à retenir pour la suite est que si l'on vous demande d'approcher une loi par une autre, il faut choisir les paramètres de façon à ce que les 2 lois aient même espérance et même variance s'il y a besoin d'un autre paramètre.

a) Loi binomiale

i) Par la loi de Poisson

Propriété 7

Si X suit une loi binomiale $\mathcal{B}(n,p)$ où n est grand, on peut approximer la loi de X par la loi de Poisson $\mathcal{P}(\lambda)$ où $\lambda = E(X) = np$.

En pratique pour que l'approximation soit valable il faut vérifier : $\begin{cases} p \leq 0,1 \\ n \geq 30 \end{cases}$.

Preuve

On s'appuie sur le fait que $\mathcal{B}\left(n, \frac{\lambda}{n}\right) \xrightarrow{loi} \mathcal{P}(\lambda)$.

Dès que n est grand, en écrivant que $\mathcal{B}\left(n, \frac{\lambda}{n}\right) \approx \mathcal{P}(\lambda)$, on pose alors $\lambda = np$ pour obtenir : $\mathcal{B}(n,p) \approx \mathcal{P}(np)$.

ii) Par la loi normale

Propriété 8

Si X suit une loi binomiale $\mathcal{B}(n,p)$ où n est grand, on peut approximer la loi de X par la loi normale $\mathcal{N}_{m,\sigma^2}$, avec $m = E(X) = np$ et $\sigma^2 = V(X) = np(1-p)$.

En pratique pour que l'approximation soit valable il faut vérifier : $\begin{cases} n \geq 30 \\ 0,2 \leq p \leq 0,8 \end{cases}$.

Preuve

Si on note X_i des variables de Bernouilli identiques, indépendantes et de paramètre p on a :

$S_n = \sum\limits_{i=1}^{n} X_i \hookrightarrow \mathcal{B}(n,p)$.

Or d'après le théorème central limite, $S_n^* \xrightarrow{loi} \mathcal{N}_{0,1}$, donc en considérant que $S_n^* \underset{\sim}{\hookrightarrow} \mathcal{N}_{0,1}$ on obtient $S_n = E(S_n) + \sigma(S_n) S_n^* \underset{\sim}{\hookrightarrow} \mathcal{N}_{np,np(1-p)}$ par transformation affine d'une loi normale.

Ainsi, $\mathcal{B}(n,p) \approx \mathcal{N}_{np,np(1-p)}$.

Remarque

On retiendra que l'on approche une binomiale par une poisson si le paramètre p est très petit ($p \leq 0,1$) et par une loi normale si le paramètre p est plus grand ($0,2 \leq p \leq 0,8$).

b) Loi de Poisson

Propriété 9

Si X suit une loi de Poisson $\mathcal{P}(\lambda)$ où λ est grand, on peut approximer la loi de X par la loi normale $\mathcal{N}_{m,\sigma^2}$, avec $m = E(X) = \lambda$ et $\sigma^2 = V(X) = \lambda$.

Preuve

Si on note X_i des variables de Poisson identiques, indépendantes et de paramètre $\frac{\lambda}{n}$ on a :

$S_n = \sum\limits_{i=1}^{n} X_i \hookrightarrow \mathcal{P}(\lambda)$.

Or d'après le théorème central limite, $S_n^* \xrightarrow{loi} \mathcal{N}_{0,1}$, donc en considérant que $S_n^* \underset{\sim}{\hookrightarrow} \mathcal{N}_{0,1}$ on obtient $S_n = E(S_n) + \sigma(S_n) S_n^* \underset{\sim}{\hookrightarrow} \mathcal{N}_{\lambda,\lambda}$ par transformation affine d'une loi normale.

Ainsi, $\mathcal{P}(\lambda) \approx \mathcal{N}_{\lambda,\lambda}$.

Exercice 11. Un étudiant fait en moyenne une faute d'orthographe tous les 500 mots. Quelle est la probabilité qu'il ne fasse pas plus de 5 fautes dans une dictée contenant 2000 mots ?
On utilisera l'approximation d'une loi binomiale par une loi de Poisson et on pourra se servir de l'annexe de la dernière page.

Exercice 12. *On admet l'approximation* $\sqrt{2} \simeq 1,41$.
Un dé régulier est lancé 9000 fois.

1. Déterminer la probabilité d'obtenir le résultat 6 entre 1450 et 1550 fois.
   On utilisera l'approximation d'une loi binomiale par une loi normale et on pourra se servir de l'annexe en dernière page.
2. Minorer également cette probabilité par l'inégalité de Bienaymé-Tchebychev .

III Estimation par intervalle de confiance

1 Principe

Pour estimer une valeur, on effectue une réalisation effective de l'expérience, et on prend les valeurs obtenues pour calculer la réalisation t_n de T_n, qui est notre estimation.

Cependant, sur deux estimations, on n'obtiendra pas la même valeur ; laquelle est alors meilleure ? Et à quelle précision notre estimation est-elle valable ?

C'est pour répondre à ces questions qu'on utilise l'estimation par intervalles de confiance.

2 Vocabulaire

Dans tout ce paragraphe , $(U_n)_{n\geq 1}$ et $(V_n)_{n\geq 1}$ désigneront deux suites d'estimateurs de $g(\theta)$ tels que $U_n \leq V_n$.

Définition 10 : intervalle de confiance

On dit que l'intervalle $[U_n; V_n]$ est un ***intervalle de confiance*** *de $g(\theta)$ au* ***niveau de confiance*** $1-\alpha$ *($\alpha \in [0;1]$) si pour tout θ de Θ, on a :*

$$\boxed{P(U_n \leq g(\theta) \leq V_n) \geq 1-\alpha}$$

Remarque

Les variables U_n et V_n sont en fait obtenues à partir d'un estimateur T_n, en posant $U_n = T_n - a$ et $V_n = T_n + a$. Ainsi il faudra, à l'aide d'une approximation ou d'un calcul exact, déterminer la valeur de a à poser pour obtenir le niveau de confiance ou de risque souhaité.

Définition 11 : intervalle de confiance asymptotique

On appelle intervalle de confiance asymptotique au niveau de confiance $1-\alpha$ une suite d'intervalle $([U_n; V_n])_{n\geq 1}$ vérifiant : pour tout θ de Θ, il existe une suite de réels $(\alpha_n)_{n\geq 1}$, à valeurs dans $[0;1]$, de limite α, telle que pour tout $n \geq 1$:

$$\boxed{P(U_n \leq g(\theta) \leq V_n) \geq 1-\alpha_n}$$

Exercice 13 (: intervalle de confiance par l'inégalité de Bienaymé-Tchebychev)**.** Soit un dé cubique truqué dont la probabilité d'obtenir la face 1 est p à déterminer.
On lance n fois le dé et on note X_i la variable aléatoire valant 1 si on obtient la face n° 1 au ième lancer, 0 sinon.
On décide d'estimer le paramètre p par la fréquence F_n d'apparition de la face n° 1 au cours des n lancers.

1. Donner l'expression de F_n en fonction des variables X_i.
2. Montrer que F_n a une espérance et une variance et les calculer.
3. Montrer que pour tous $p \in [0,1]$, $p(1-p) \leq \frac{1}{4}$.
4. Enoncer l'inégalité de Bienaymé-Tchebychev pour la variable F_n puis en déduire que
$$P(F_n - \varepsilon \leq p \leq F_n + \varepsilon) \geq 1 - \frac{1}{4n\varepsilon^2}$$
5. Donner une valeur de ε telle que l'intervalle $[F_n - \varepsilon; F_n + \varepsilon]$ est un intervalle de confiance de p au niveau de confiance $1-\alpha$.

6. Deux cas :
   (a) On effectue 1000 lancers. Donner alors un intervalle de confiance de p au niveau de confiance de 90%
   (b) Déterminer le nombre minimal n de lancers à effectuer pour que $[F_n - 0,01, F_n + 0,01]$ soit un intervalle de confiance de p au niveau de confiance de 99%.

Exercice 14 (: intervalle de confiance asymptotique par le théorème central limite)**.** *Dans cet exercice, on admet que la table de la fonction de répartition Φ de la loi normale $\mathcal{N}(0,1)$ donne $\Phi(2,575) = 0,995$.*

10 000 000 électeurs sont inscrits sur les listes électorales d'un pays. Une élection oppose deux candidats A et B.
On cherche à estimer la proportion p d'électeurs (inscrits) du pays favorables au candidat A.
On considère un échantillon quelconque de n électeurs choisis au hasard (avec remise). Pour tout $i \in [|1;n|]$, on note X_i la variable aléatoire égale à 1 si le i-ème électeur interrogé vote pour le candidat A, et à 0 sinon.

1. On note N_n le nombre d'électeurs favorables à A dans l'échantillon et F_n la proportion d'électeurs favorables à A dans l'échantillon et F_n^* la v.a.r centrée réduite associée à F_n. On choisit F_n comme estimateur de p.
   (a) Exprimer F_n en fonction de N_n puis en fonction des variables X_1, X_2, ...X_n.
   (b) Déterminer un réel t tel que :
   $$\lim_{n\to+\infty} P(-t \leq F_n^* \leq t) = 0,99$$
   (c) Montrer que :
   $$\lim_{n\to+\infty} P\left(p - t\frac{\sqrt{p(1-p)}}{\sqrt{n}} \leq F_n \leq p + t\frac{\sqrt{p(1-p)}}{\sqrt{n}}\right) = 0,99$$
   (d) Démontrer que pour tout $p \in [0,1]$, $0 \leq p(1-p) \leq \frac{1}{4}$.
   (e) Montrer que :
   $$\lim_{n\to+\infty} P\left(p - t\frac{1}{2\sqrt{n}} \leq F_n \leq p + t\frac{1}{2\sqrt{n}}\right) \geq 0,99$$
   (f) En déduire un intervalle de confiance asymptotique de p et donner le niveau de confiance associé.
   (g) Dans un échantillon observé de 100 électeurs, 64 sont déclarées favorables à A.
   Donner l'intervalle de confiance de p au niveau de confiance asymptotique $0,99$.
   A est-il certain d'être élu ?

Annexe

1. : Table de la loi normale centrée réduite

La table ci-dessous comporte les valeurs de la fonction de répartition de la loi normale centrée réduite, à savoir les valeurs de :

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left(-\frac{t^2}{2}\right) dt$$

par exemple $\Phi(0,67) = 0,7486$

x	**0,00**	**0,01**	**0,02**	**0,03**	**0,04**	**0,05**	**0,06**	**0,07**	**0,08**	**0,09**
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,758	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,975	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817

2. : Table de la loi de Poisson

Table de Poisson donnant la fonction de répartition d'une loi de Poisson de paramètre λ.

k	$\lambda = 3$	$\lambda = 4$	$\lambda = 5$	$\lambda = 6$	$\lambda = 7$
0	$0,0498$	$0,0183$	$0,0067$	$0,0025$	$0,0009$
1	$0,1991$	$0,0916$	$0,0404$	$0,0174$	$0,0073$
2	$0,4232$	$0,2381$	$0,1247$	$0,0620$	$0,0296$
3	$0,6472$	$0,4335$	$0,2650$	$0,1512$	$0,0818$
4	$0,8153$	$0,6288$	$0,4405$	$0,2851$	$0,1730$
5	$0,9161$	$0,7851$	$0,6160$	$0,4457$	$0,3007$
6	$0,9665$	$0,8893$	$0,7622$	$0,6063$	$0,4497$
7	$0,9881$	$0,9489$	$0,8666$	$0,7440$	$0,5987$
8	$0,9962$	$0,9786$	$0,9319$	$0,8472$	$0,7291$
9	$0,9989$	$0,9919$	$0,9682$	$0,9161$	$0,8305$
10	$0,9997$	$0,9972$	$0,9863$	$0,9574$	$0,9015$
11	$0,9999$	$0,9991$	$0,9945$	$0,9799$	$0,9467$
12	$1,0000$	$0,9997$	$0,9980$	$0,9912$	$0,9730$
13		$0,9999$	$0,9993$	$0,9964$	$0,9872$
14		$1,0000$	$0,9998$	$0,9986$	$0,9943$
15			$0,9999$	$0,9995$	$0,9976$
16			$1,0000$	$0,9998$	$0,9990$
17				$0,9999$	$0,9996$
18				$1,0000$	$0,9999$
19					$1,0000$
20					