CONVERGENCES, APPROXIMATIONS ET ESTIMATION

Introduction

Nous avons pour l'instant étudié les probabilités d'un point de vue théorique : lois, espérance, transformation de variables aléatoires ... Cette théorie s'est en fait développée initialement avec (entre autres) des buts pratiques. L'un de ces buts est l'étude statistique de certains phénomènes. Pour prendre un exemple simple : peut-on, en lançant une pièce un grand nombre de fois, estimer la probabilité d'obtenir Pile ? Peut-on se prononcer sur le fait qu'elle est truquée ? Le premier résultat permettant le passage de la théorie à ces études statistiques est la loi des grands nombres. Cette loi formalise l'intuition que nous avons que, si on lance un grand nombre de fois une pièce équilibrée, on obtiendra une proportion de “Pile” proche de 1/2.
La seconde propriété importante dans ce domaine est le théorème central limite. Il précise les informations données par la loi des grands nombres en nous permettant de répondre à des questions comme la suivante : si on lance 10000 fois une pièce équilibrée, peut-on estimer la probabilité que le nombre de “Pile” se trouve entre 4900 et 5100 ? Ce théorème permet aussi de construire des intervalles de confiance pour les quantités estimées.
Les propriétés vues dans ce chapitre seront pour la plupart admises.

1 Loi faible des grands nombres

1.1 Inégalité de Bienaymé-Tchebychev

Cette inégalité est destinée à démontrer la loi des grands nombres. Elle relie l'écart à la moyenne à la variance.
On commence par un résultat qui nous servira à montrer l'inégalité de Bienaymé-Tchebychev.
Un exemple introductif tout d'abord :

Exemple 1.1. *À un devoir, la moitié de la classe a eu une note supérieure ou égale à 12. Que peut-on dire de la moyenne de classe ?*

Propriété 1.2 (Inégalité de Markov)**.** *Soit X une variable aléatoire sur un espace probabilisé $(\Omega, \mathcal{A}, P)$, à valeurs positives, et admettant un moment d'ordre 1 (c'est à dire une espérance). On a alors, pour tout $a > 0$:*

$$E(X) \geq a\ P(X \geq a)\ .$$

Démonstration dans le cas où X est à valeurs dans $\mathbb{N}$

On en déduit l'inégalité suivante :

Propriété 1.3 (Inégalité de Bienaymé-Tchebichev)**.** *Soit X une variable aléatoire sur un espace probabilisé $(\Omega, \mathcal{A}, P)$ admettant des moments d'ordre 1 et 2. On a alors, pour tout $\epsilon > 0$:*

$$P(|X - E(X)| \geq \epsilon) \leq \frac{V(X)}{\epsilon^2}.$$

Démonstration

Exemple 1.4. *Soit X une variable aléatoire suivant une loi normale centrée réduite. Majorer la probabilité que $|X|$ soit supérieure à 2 à l'aide de l'inégalité précédente.*

1.2 Loi faible des grand nombres

Ce théorème est un des plus importants de la théorie des probabilités. Il assure que si on répète une même expérience un très grand nombre de fois, alors la moyenne des résultats obtenus converge vers l'espérance du résultat d'une expérience. Nous avons déjà utilisé de manière tacite ce résultat en informatique, lorsque nous avons pris la moyenne d'un grand nombre de résultats d'une même expérience et considéré le résultat comme une valeur approchée de la moyenne théorique.

Théorème 1.5 (Loi faible des grand nombres). *Soit (X_n) une suite de variables aléatoires définies sur un espace probabilisé $(\Omega, \mathcal{A}, P)$, indépendantes et de même loi, admettant pour espérance m et pour variance σ^2. On considère, pour tout n entier naturel non nul, la suite des moyennes des n premiers X_i en posant :*

$$\overline{X_n} = \frac{1}{n} \sum_{i=1}^{n} X_i \ .$$

On a alors, pour tout $\epsilon > 0$,

$$\lim_{n \to +\infty} P\left(\left|\overline{X_n} - m\right| \geq \epsilon\right) \ = \ 0 \ .$$

Démonstration

Exemple 1.6. *On répète indéfiniment le lancer d'un dé équilibré. On modélise cette expérience aléatoire par une suite de variables aléatoires $(X_n)_{n\geq 1}$ définies sur un espace probabilisé $(\Omega, \mathcal{A}, P)$, indépendantes deux-à-deux et suivant chacune la loi uniforme sur $[[1,6]]$.*

1. *Que donne l'application de la loi des grands nombres dans cet exemple ?*
2. *Pour la même expérience aléatoire, on note, pour $n \geq 1$, Y_n la variable aléatoire qui vaut 1 si la valeur 6 sort au n-ème lancer et 0 sinon. Que donne l'application de la loi des grands nombres à la suite $(Y_n)_{n\geq 1}$?*

2 Estimation ponctuelle

2.1 Introduction

On considère dans cette partie du chapitre une question apparaissant naturellement lors des études statistiques : on suppose qu'un phénomène étudié suit une loi dont on connaît ou suppose la nature (loi uniforme, loi de Poisson, loi normale...), mais dont on ne connaît pas un paramètre (paramètre λ pour une loi de Poisson, espérance m pour une loi normale ...).

Exemple 2.1. *On possède une pièce dont sait qu'elle est truquée. On sait que le résultat d'un lancer suit une loi de Bernoulli, mais on ne connaît pas le paramètre p de cette loi. De manière intuitive, une stratégie raisonnable pour déterminer p est de lancer un grand nombre de fois la pièce et de considérer que la proportion de Pile obtenue est une valeur approchée de la probabilité d'obtenir Pile à chaque lancer.*

Le but de cette partie du chapitre est de formaliser l'intuition utilisée dans l'exemple précédent. On cherche à préciser dans quelle mesure une quantité évaluée suite à une répétition d'expériences aléatoires approche le paramètre inconnu. En particulier on cherchera à donner un intervalle dans lequelle le paramètre setrouve avec une probabilité forte.
Dans toute cette partie, on considère une variable aléatoire réelle X et un paramètre inconnu de la loi de X : θ.

2.2 n - échantillon

L'expérience aléatoire consistant dans l'exemple de l'introduction à répéter un grand nombre de fois les lancers esr formalisée par la notion suivante.

Définition 2.2. *On appelle <u>n-échantillon</u> de la loi de X un n-uplet $(X_1, ..., X_n)$ de variables aléatoires mutuellement indépendantes et de même loi que X.*

Définition 2.3. *Si $(X_1, ..., X_n)$ est un n-échantillon de la loi de X, on appelle <u>réalisation</u> de cet échantillon un n-uplet $(x_1, ..., x_n)$ où x_i est la valeur prise par la variable aléatoire X_i.*

Exemple 2.4. *Dans le cas de la pièce, les résultats des nombreux lancers de la pièce forment une réalisation de l'échantillon.*

2.3 Estimateur

Définition 2.5. *Soit $(X_1, ..., X_n)_{n\in\mathbb{N}^*}$ une suite d'échantillons de la loi de X. On appelle* estimateur *de θ une suite de variables aléatoires (T_n) où T_n est une fonction de $(X_1, ..., X_n)$ Par abus de langage, on appellera également estimateur de θ chacun des T_n.*

Remarques 2.6.
- *Les T_n ne doivent pas être explicitement fonction de θ, puisqu'on ne le connaît pas... Plus précisément, il doit être calculable à l'aide de $x_1, ..., x_n$ seulement.*
- *Cette définition est au première abord bien étonnante... En effet, on ne demande absolument pas aux T_n d'avoir un lien avec θ. En pratique on s'intéressera évidemment à des T_n permettant d'approcher celui-ci ou au moins une fonction de θ !*
  En pratique, on pourrait estimer souhaitable qu'en moyenne notre estimateur soit égal au paramètre à estimer, c'est à dire que $E(T_n) = \theta$ (un tel estimateur est dit sans biais).

Définition 2.7. *On appelle moyenne empirique de $(X_1, ..., X_n)$ la variable aléatoire :*

$$\overline{X_n} = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Propriété 2.8 (Espérance et variance de la moyenne empirique)**.** *Si X_1 a pour espérance m et pour variance σ^2 alors*

$$E(\overline{X_n}) = m \quad et \quad V(\overline{X_n}) = \frac{\sigma^2}{n}\ .$$

Démonstration :

Remarques 2.9.
- *Dans l'exemple de la pièce, on a intuitivement utilisé la moyenne empirique comme estimateur de l'espérance.*
- *Pour revenir à la fin de la remarque précédente, si on note m la moyenne à estimer, on a bien $E(\overline{X_n}) = m$ (un tel estimateur est dit sans biais).*

2.4 Convergence d'un estimateur

Définition 2.10. *Un estimateur T_n d'un paramètre θ est dit convergent lorsque :*

$$\forall\ \epsilon > 0\ ,\ \lim_{n\to+\infty} P\left(|T_n - \theta| \geq \epsilon\right)\ =\ 0\ .$$

Exemple 2.11. *La loi des grands nombres affirme exactement que la moyenne empirique est un estimateur convergent de l'espérance.*

3 Convergence en loi

La convergence en loi donne un sens précis au fait que la loi d'une variable aléatoire se rapproche d'une autre.

Définition 3.1. *Soit (X_n) une suite de variables aléatoires sur un espace probabilisé $(\Omega, \mathcal{A}, P)$. Soit X une variable aléatoire sur $(\Omega, \mathcal{A}, P)$. On note F_n la fonction de répartition de X_n et F celle de X. On dit que la suite de variables aléatoires (X_n) converge en loi vers X si pour tout $x \in \mathbb{R}$ en lequel F est continue on a :*

$$\lim_{n\to+\infty} F_n(x) = F(x)$$

On note alors :

$$X_n \quad \xrightarrow{\mathcal{L}} X.$$

Dans le cas discret, on a la caractérisation suivante :

Propriété 3.2 (Convergence en loi dans le cas discret)**.** *Si les (X_n) ainsi que X sont des variables aléatoires discrètes (à valeurs dans $\mathbb{N}$ par exemple) alors la suite (X_n) converge en loi vers X si et seulement si*

$$\forall k \in \mathbb{N}, \quad \lim_{n\to+\infty} P(X_n = k) = P(X = k)$$

On a un premier exemple de convergence en loi pour une suite de variables aléatoire suivant des lois binômiales :

Propriété 3.3. *Si les (X_n) sont des variables aléatoires discrètes de lois binomiales $\mathcal{B}(n, \lambda/n)$ alors (X_n) converge en loi vers X de loi de Poisson $\mathcal{P}(\lambda)$.*

Illustration : La convergence en loi d'une suite de variables aléatoires discrètes peut être visualisée en effectuant les diagrammes en bâtons des lois. Sur l'exemple précédent, on peu le voir (voir TP) en traçant le diagramme en bâtons d'une loi de Poisson ($\lambda = 5$ par exemple), puis, pour n de plus en plus grand, celui de la loi binomiale $\mathcal{B}(n, \lambda/n)$.

En pratique: On utilisera le fait que pour certaines valeurs de n et p ($n \geq 30$, $np \leq 15$, $p \leq 0.1$), une variable aléatoire de loi binomiale de paramètres n et p peut être approchée par une variable aléatoire de loi de Poisson de paramètre np.

4 Théorème central limite

Ce théorème précise en un certain sens la loi des grands nombres. Ces deux théorèmes sont à la base de l'utilisation de la théorie des probabilités en statistiques.

Théorème 4.1 (Théorème Central Limite)**.** *Soit (X_n) une suite de variables aléatoires définies sur un espace probabilisé $(\Omega, \mathcal{A}, P)$, indépendantes et de même loi, admettant pour espérance m et pour variance σ^2. On note $S_n = \sum_{i=1}^n X_i$ et $\overline{X}_n = \frac{1}{n} S_n$. Alors on a la convergence en loi suivante :*

$$\frac{\sqrt{n}}{\sigma}\left(\overline{X}_n - m\right) \quad \xrightarrow{\mathcal{L}} X \ ;$$

où X est une variable aléatoire suivant une loi $\mathcal{N}(0, 1)$.

Rappelons que cela signifie que, pour tout $x \in \mathbb{R}$,

$$\lim_{n\to+\infty} P\left(\sqrt{n}\left(\frac{\overline{X}_n - m}{\sigma}\right) \leq x\right) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \mathrm{dt} \ .$$

Cela a pour conséquence immédiate que, pour tous réels a et b on a :

$$\lim_{n\to+\infty} P\left(a \leq \sqrt{n}\left(\frac{\overline{X}_n - m}{\sigma}\right) \leq b\right) = \int_{a}^{b} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} \mathrm{dt}$$

Application : Revenons à notre question de l'introduction : si on lance 10000 fois une pièce équilibrée, peut-on estimer la probabilité que le nombre de "Pile" se trouve entre 4900 et 5100 ?
On modélise cette expérience aléatoire par une suite de variables aléatoires $(X_n)_{n\geq 1}$ définies sur un espace probabilisé $(\Omega, \mathcal{A}, P)$, indépendantes et suivant chacune la loi de Bernoulli de paramètre $1/2$ (X_n vaut 1 si on a obtenu Pile et 0 si on a obtenu Face). On garde les notations du théorème. La variable aléatoire S_n représente alors le nombre de fois que l'on a obtenu Pile.

1. On connaît la loi de S_{10000} (loi binomiale), donc on pourrait calculer la valeur exacte de la probabilité de l'événement $[4900 \leq S_{10000} \leq 5100]$. D'une part, cela nécessiterait des calculs difficiles à effectuer en partique ; d'autre part, si l'on n'est pas dans le cas particulier où X_n suit une loi de Bernoulli, cette méthode ne peut plus être appliquée.

2. L'inégalité de Bienaymé-Tchebichev, utilisée pour démontrer la loi des grands nombres, fournit une minoration de cette valeur.

3. Le théorème central limite fournit une méthode plus efficace pour estimer cette valeur.

En pratique : On considère que pour certaines valeurs de n et p ($n \geq 30$, $np \geq 5$, $n(1-p) \geq 5$), une loi binomiale de paramètres n et p peut être approchée par une loi $\mathcal{N}(np, np(1-p))$.

5 Estimation par intervalle de confiance

Une fois que l'on a un estimateur correct d'un paramètre, on voudrait préciser cette information en donnant un intervalle (autour de cet estimateur) dans lequel le paramètre a une probabilité forte de se trouver. C'est dans ce but que l'on introduit la notion d' intervalle de confiance.

Définition 5.1. *Soient U_n et V_n deux variables aléatoires et $\alpha \in]0,1[$. On dit que $[U_n, V_n]$ est un intervalle de confiance pour θ avec niveau de confiance $1-\alpha$ si :*

$$P(U_n \leq \theta \leq V_n) \geq 1-\alpha$$

On appelle α le risque.

Remarque 5.2. *En pratique, U_n et V_n sont des variables aléatoires dépendant d'un n-échantillon $(X_1, ..., X_n)$. L'intervalle $[u_n, v_n]$ calculé à partir d'une réalisation $x_1, ..., x_n$ de l'échantillon sera l'intervalle de confiance observé (ou réalisé).*

Propriété 5.3 (Intervalle de confiance avec Bienaymé-Tchebychev)**.** *Soit T_n un estimateur de θ tel que $E(T_n) = \theta$ et $\alpha \in]0,1[$. Alors $[T_n - \sqrt{\frac{V(T_n)}{\alpha}}, T_n + \sqrt{\frac{V(T_n)}{\alpha}}]$ est un intervalle de confiance de niveau de confiance $1-\alpha$ pour θ.*

Démonstration :

Propriété 5.4 (Intervalle de confiance avec le théorème central limite)**.** *Soit X d'espérance m et de variance σ^2. Soit $(X_1, ..., X_n)$ un échantillon de la loi de X et $\overline{X_n} = \frac{1}{n}\sum_{i=1}^{n} X_i$. Soit $t_\alpha = \Phi^{-1}(1-\alpha/2)$ (Φ désigne la fonction de répartition de la loi $\mathcal{N}(0,1)$). Alors, pour n assez grand,*

$$\left[\overline{X_n} - \frac{\sigma t_\alpha}{\sqrt{n}}, \overline{X_n} + \frac{\sigma t_\alpha}{\sqrt{n}}\right]$$

est un intervalle de confiance pour l'espérance m avec un niveau $1-\alpha$.

Démonstration :

Remarque 5.5. *Dans l'exemple précédent, on a, plutôt qu'une égalité :*

$$\lim_{n\to\infty} P\left(m \in \left[\overline{X_n} - \frac{\sigma t_\alpha}{\sqrt{n}}, \overline{X_n} + \frac{\sigma t_\alpha}{\sqrt{n}}\right]\right) \geq 1-\alpha \ .$$

On qualifie alors parfois cet intervalle d'intervalle de confiance asymptotique (mais on ne précise pas toujours en fait).

Remarque 5.6. *Dans la propriété précédente, l'intervalle de confiance dépend de σ. En pratique, pour appliquer cette méthode, il faut donc connaître σ ou avoir un moyen de l'estimer.*

Exemple 5.7. *On reprend l'exemple de la pièce dont on veut estimer la probabilité p de tomber sur pile.*

1. *Déterminer un majorant de $p(1-p)$ et en déduire un majorant de l'écart-type.*
2. *On suppose que l'on a lancé la pièce 10000 fois et obtenu 4900 fois Pile.*
   En appliquant la méthode précédente (basée sur le TCL), déterminer un intervalle de confiance au niveau de risque 5% de p.