Au programme : 13.7 Estimation
Estimation ponctuelle : n-échantillon d'une variable aléatoire ; estimateur, biais d'un estimateur, estimateur asymptotiquement sans biais ; estimateur convergent, risque quadratique ; moyenne empirique, variance empirique. Estimation par un intervalle : intervalle de confiance, intervalle de confiance asymptotique. Estimation du paramètre d'une loi de Bernoulli. Application : méthode de Monte-Carlo pour le calcul approché d'une intégrale ou d'une somme de série.

Leçons :
249 Loi normale en probabilités et statistiques.
448 Exemples d'estimation en statistiques : estimation ponctuelle, estimation par intervalles de confiance.
453 Exercices illustrant l'utilisation de la loi binomiale en probabilités et en statistiques.
Et dans une moindre mesure :
220 Méthodes de calcul approché d'une intégrale. Majoration ou estimation de l'erreur.
229 Suites de variables aléatoires indépendantes de même loi de Bernoulli. Variables aléatoires de loi binomiale et approximations de la loi binomiale.
231 Espérance, variance ; loi faible des grands nombres. Applications.

Développements possibles
- Méthode de Box-Muller
- Méthode de Monte-Carlo (Chaffaï p103)
- Moivre-Laplace (Chaffaï)

1 Notes de cours

1.1 Echantillon d'une variable aléatoire

On se place sur un espace probabilisé $(\Omega, \mathcal{F}, \mathbb{P})$. On considère une variable aléatoire X à valeurs dans $\mathbb{R}^d$.
Un n-échantillon de X est la donnée de n variables aléatoires indépendantes de même loi que X.

1.2 Estimation

1.2.1 Généralités

On modélise des phénomènes à l'aide de variables aléatoires dont on ne connait pas la loi mais on sait que cette loi appartient à un ensemble de lois dépendant d'un paramètre.

Exemple 1.1.
- *Pile ou face modélisée par une Bernoulli de paramètre p.*
- *Durée de vie d'une ampoule électrique modélisée par une loi exponentielle.*
- *Nombre de clients arrivant à un guichet pendant un intervalle de temps de longueur t modélisé par une loi de Poisson de paramètre λt.*

L'estimation consiste à donner une valeur approchée du paramètre inconnu à partir des observations qu'on suppose être la réalisation d'un échantillon.
La question suivante est celle de la fiabilité de cette valeur trouvée, d'où la notion d'intervalles de confiance.

1.2.2 Estimateur

Définition 1.2. *Soit $(X_1, \cdots, X_n)$ un n-échantillon d'une variable aléatoire X sur $E = \mathbb{R}^d$ dépendant d'un paramètre $\theta \in \mathbb{R}^k$.*
T est une statistique s'il existe une application φ de E^n dans $\mathbb{R}^d$ telle que $T = \varphi(X_1, \cdots, X_n)$, la fonction φ ne dépendant pas de θ.
Un estimateur de θ sera une statistique qui servira à estimer θ.

Exemple 1.3.

Moyenne empirique

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

$\overline{X_n}$ sert à estimer l'espérance de X.

Variance empirique

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

S_n^2 sert à estimer la variance de X.
Montrer que $E[S_n^2] = \frac{n-1}{n}V(X)$.

1.2.3 Biais d'un estimateur

Définition 1.4. *Soit T un estimateur de $\theta \in \mathbb{R}$.*
Le **biais** *de T est $b_T(\theta) = E[T] - \theta$.*
Un estimateur est dit sans biais si son biais est nul.
Un estimateur est dit asymptotiquement sans biais si $b_T(\theta)$ tend vers θ quand n tend vers $+\infty$.

1.2.4 Risque quadratique

Définition 1.5. *Soit T un estimateur de $\theta \in \mathbb{R}$.*
Le **risque quadratique** *de T est $r_T(\theta) = E[(T - \theta)^2]$.*

A quoi ça sert ?

1.2.5 Estimateur convergeant

Définition 1.6. *Soit T un estimateur de $\theta \in \mathbb{R}$.*
T est un **estimateur convergeant** *si T converge en probabilités vers θ.*

1.2.6 Exemples

On considère une variable aléatoire X d'espérance m et de variance σ^2.

Estimation de m

$\overline{X_n}$ est un estimateur sans biais et convergeant de m. Son risque quadratique est donné par $\frac{\sigma^2}{n}$.

Estimation de σ^2 quand m est connu

$\frac{1}{n}\sum_{i=1}^{n}(X_i - m)^2$est un estimateur sans biais et convergeant de σ^2.

Estimation de σ^2 quand m est inconnu

S_n^2 est un estimateur asymptotiquement sans biais et convergeant de la variance de X.
$\widetilde{S_n}^2 = \frac{n}{n-1}S_n^2$ est un estimateur sans biais et convergeant de la variance de X.

Prouver les résultats précédents

1.3 Estimation par intervalle

1.3.1 Intervalle de confiance

Soit T un estimateur de θ et $0 < \alpha < 1$.

Définition 1.7. *Un intervalle de confiance pour θ de niveau $1-\alpha$ est donnée par un intervalle aléatoire I tel que $\mathbb{P}(\theta \in I) \geq 1-\alpha$.*
Si $\mathbb{P}(\theta \in I)$ tend vers $1-\alpha$, on parlera d'un intervalle de confiance asymptotique de niveau $1-\alpha$.

Comment trouver un tel intervalle ?
Si la loi de T est connue il suffit de considerer t_2 donné par $F_T(t_2) = 1-\alpha+\beta$ et t_1 donné par $F_T(t_1) = \beta$, où $0 < \beta < \alpha$. F_T est la fonction de répartition de T.
En effet on aura alors $I =]t_1, t_2]$ qui convient car

$$\mathbb{P}(t_1 < T \leq t_2) = F_T(t_2) - F_T(t_1) = 1-\alpha$$

On peut aussi utiliser l'inégalité de Bienaimé-Tchebichev.

1.3.2 Exemples

On considère une variable aléatoire X d'espérance m et de variance σ^2.

Estimation de m quand σ est connu

1. En utilisant Bienaimé-Tchebichev.

$$\mathbb{P}(|\overline{X_n} - m| > \epsilon) \leq \frac{\sigma^2}{n\epsilon^2}$$

D'où l'intervalle de confiance de niveau $1-\alpha$: ?

2. Intervalle de confiance asymptotique
   On utilise le théorème central limite qui dit que $\frac{\sqrt{n}(\overline{X_n}-m)}{\sigma}$ converge en loi vers une loi normale centrée réduite $\mathcal{N}(0,1)$.
   Ainsi si t_1, t_2 sont donnés par $F(t_2) - F(t_1) = 1 - \alpha$, $\mathbb{P}(t_1 \leq \frac{\sqrt{n}(\overline{X_n}-m)}{\sigma} \leq t_2)$ tend vers $1 - \alpha$.
   On en déduit donc un intervalle de confiance asymptotique de niveau $1 - \alpha$ donné par ?
   On peut prouver que le meilleur (?) choix pour t_1, t_2, est donné par $t_2 = -t_1$ et donc $F(t_2) = 1 - \frac{\alpha}{2}$.

Estimation de m quand σ est inconnu

On va construire un intervalle de confiance asymptotique en utilisant le théorème central limite. le problème est que l'intervalle de confiance ainsi construit dépend de σ qui est inconnu.
On utilise alors le fait que S_n^2 tend en probabilité (et même p.s.) vers σ^2. On pourrait ainsi prouver que $\frac{\sqrt{n}(\overline{X_n}-m)}{\sqrt{S_n^2}}$ converge en loi vers une loi normale centrée réduite $\mathcal{N}(0,1)$.

Cas où X suit la loi $\mathcal{N}(m, \sigma^2)$

On peut dans ce cas trouver des intervalles de confiance non asymptotiques car plusieurs lois sont connues.
Il y a quatre cas à examiner :

Estimation de m quand σ est connu

$\overline{X_n}$ a pour loi $\mathcal{N}(m, \frac{\sigma^2}{n})$.

Estimation de m quand σ est inconnu

$\frac{\sqrt{n}(\overline{X_n}-m)}{\sqrt{\widetilde{S_n}^2}}$ suit la loi de Student de paramètre $n-1$.

Estimation de σ quand m est connu

$\frac{nY_n}{\sigma^2}$ suit la loi du chi2 de paramètre n si $Y_n = \frac{1}{n}\displaystyle\sum_{k=1}^{n}(X_k - m)^2$.

Estimation de σ quand m est inconnu

$\frac{(n-1)\widetilde{S_n}^2}{\sigma^2}$ suit la loi du chi2 de paramètre $n-1$.

Rappels La loi du chi2 à de paramètres n est la loi de $Z_1^2 + \cdots + Z_n^2$ où les variables aléatoires $Z_1, \ldots, Z_n$ sont indépendantes et suivent la loi $\mathcal{N}(0,1)$, c'est la loi $\gamma(\frac{1}{2}, \frac{n}{2})$.
La loi de Student de paramètre n est la loi de $\frac{X}{\sqrt{Y/n}}$ où X est une variable aléatoire de loi normale centrée et réduite et Y suit une loi du chi2 à n degrés de paramètre n, X et Y étant indépendantes.
Donner des intervalles de confiance de niveau 95% dans les quatre cas précédents en utilisant un logiciel pour trouver les valeurs t_1 et t_2.
Faîtes des simulations des lois de Student et du chi2

2 Exercices

Voir feuille d'exercices.