Université de Strasbourg
UFR de Mathématiques et d'Informatique
Han-Ping LI

Année 2019/2020
Statistique
Etude de cas L3

Chapitre I Notions et modèles statistiques

Variables aléatoires : notions mathématiques utilisées pour modéliser les phénomènes complexes :

-Impossible de prédire sa valeur exacte

-Possible d'avancer des propositions probabilistes sur ses valeurs

Modélisation statistique :

Modéliser, c'est simplifier et structurer , choisir quoi identifier et quoi différencier, déceler ceux qui sont essentiels et ceux qui sont accessoires, relier les évènements importants.

Population : l'ensemble des individus dont une ou plusieurs variables sont prises en considération dans l'étude.

Echantillon $\Longleftrightarrow$ n individus choisis au hasard dans la population. $X_1, \ldots, X_n$ n variables aléatoires indépendantes et de même loi.

Données : $x_1, \ldots, x_n$

Modèle : $\mathbf{x} = (x_1, \ldots, x_n)$ considéré comme une réalisation d'un échantillon $\mathbf{X} = (X_1, \ldots, X_n)$

où $X_1, \ldots, X_n$ sont n v.a. i.i.d. d'une des lois dans une famille $\left\{\mathbb{P}_\theta,\ \theta \in \Theta\right\}$ θ étant un paramètre inconnu.

Paramètres

1) Paramètres de position : nombre autour duquel se repartissent les valeurs.

• Espérance de la variable X sous la loi $\mathbb{P}_\theta$ que on appelle aussi la moyenne théorique :

$$\mu = \mathbb{E}_\theta(X) = \begin{cases} \displaystyle\sum_k x_k \mathbb{P}_\theta(X = x_k) & \text{cas discret} \\ \displaystyle\int_{-\infty}^{\infty} x\, f_\theta(x) \mathrm{d}\, x & \text{cas absolument continu} \end{cases}$$

Exemples :

• Si $\mathcal{L}(X) = \mathcal{B}(1, \theta)$, alors

$$\mathbb{E}(X) = 1 \times \mathbb{P}(X = 1) + 0 \times \mathbb{P}(X = 0) = 1.\theta + 0.(1 - \theta) = \theta.$$

• Si $\mathcal{L}(X) = \mathcal{P}(\lambda)$, alors

$$\mathbb{E}(X) = \sum_{k=0}^{\infty} k \times \mathbb{P}(X = k) = \sum_{k=0}^{\infty} k \times \frac{\lambda^k}{k!} \exp(-\lambda) = \lambda.$$

• Si $\mathcal{L}(X) = \mathcal{N}(\mu, \sigma^2)$, alors

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x \times \frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{1}{2\sigma^2}(x-\mu)^2) d\,x = \mu.$$

• Médiane de la variable X sous la loi $\mathbb{P}_\theta$ que on appelle aussi la médiane théorique :

$$m_\theta = \textbf{Médiane}_\theta(X) \text{ t. q. } \mathbb{P}_\theta(X \leq m_\theta) \geq \frac{1}{2} \text{ et } \mathbb{P}_\theta(X \geq m_\theta) \geq \frac{1}{2}$$

.

2) Paramètres de dispersion : nombre indiquant le degré d'éparpillement des valeurs.

• Variance de la variable X sous la loi $\mathbb{P}_\theta$ que on appelle aussi la variance théorique (l'écart-type, resp.) :

$$\sigma^2 = \mathbb{V}ar_\theta(X) = \mathbb{E}_\theta(X-\mu)^2 = \mathbb{E}(X^2) - \mu^2$$

• écart-type de la variable X

$$\sigma = \sqrt{\mathbb{V}ar_\theta(X)}.$$

Exemples :

• Si $\mathcal{L}(X) = \mathcal{B}(1, \theta)$, alors

$$\mathbb{V}ar(X) = \mathbb{E}(X^2) - \Big(\mathbb{E}(X)\Big)^2 = \theta(1-\theta).$$

• Si $\mathcal{L}(X) = \mathcal{P}(\lambda)$, alors

$$\mathbb{V}ar(X) = \mathbb{E}(X^2) - \Big(\mathbb{E}(X)\Big)^2 = \lambda.$$

• Si $\mathcal{L}(X) = \mathcal{N}(\mu, \sigma^2)$, alors

$$\mathbb{V}ar(X) = \int_{-\infty}^{\infty} (x-\mu)^2 \times \frac{1}{\sqrt{2\pi}\sigma} \exp(-\frac{1}{2\sigma^2}(x-\mu)^2) d\,x = \sigma^2.$$

3) Paramètre de forme

Paramètre d'asymétrie :

$$\gamma_1 = \frac{\mathbb{E}_\theta(X-\mu)^3}{\sigma^3}$$

Si étalée vers gauche $\Longrightarrow \gamma_1 < 0$

Si étalée vers droite $\Longrightarrow \gamma_1 > 0$

Remarque Il est important de différentier trois sortes de grandeurs :

1) Grandeurs théoriques : ce sont des paramètres dépendant de la loi $\mathbb{P}_\theta$ donc inconnues, par exemple :

$\mu = \mathbb{E}_\theta(X), \mathbf{Mdiane}_\theta(X), \sigma^2 = \mathbb{V}ar_\theta(X)$ et skewness(X).

2) Grandeurs d'échantillon : v.a. calculées à partir de l'échantillon $\mathbf{X} = (X_1, \ldots, X_n)$. Ce sont des Statistiques (v.a.) . Par exemple :

- La moyenne de l'échantillon

$$\overline{X} = \frac{1}{n}\sum_{j=1}^{n} X_j.$$

- La médiane de l'échantillon

On note $X_{(1)} = \min_{1\le j\le n}(X_j), X_{(2)} = \min_{1\le j\le n, j\neq(1)}(X_j) \ldots, X_{(n)} = \max_{1\le j\le n}(X_j)$

$$\text{mediane}(X) = \begin{cases} X_{\left(\frac{n+1}{2}\right)} & \text{si } n \text{ est impair} \\ \dfrac{X_{\left(\frac{n}{2}\right)} + X_{\left(\frac{n}{2}+1\right)}}{2} & \text{si } n \text{ est pair} \end{cases}$$

- La variance (corrigée) de l'échantillon,

$$S_c^2 = \frac{1}{n-1}\sum_{j=1}^{n}\left(X_j - \overline{X}\right)^2.$$

- L'écart-type (corrigé)

$$S_c = \sqrt{S_c^2}$$

.

$$\gamma_1 = \frac{\sum_{j=1}^{n}\left(X_j - \overline{X}\right)^3/n}{\left(S_c\right)^3}.$$

3) Grandeurs observés : calculées à partir des observations $\mathbf{x} = (x_1, \ldots, x_n)$. Ce sont des réalisations des Statistiques. Par exemple :

$$\overline{x} = \frac{1}{n}\sum_{j=1}^{n} x_j, \quad s_c^2 = \frac{1}{n-1}\sum_{j=1}^{n}(x_j - \overline{x})^2, \quad \ldots$$

Propriétés d'échantillon

Soit $\mathbf{X} = (X_1, \ldots, X_n)$ un échantillon aléatoire de taille n provenant d'une population $\mathbb{P}_\theta$ avec $\mu = \mathbb{E}_\theta(X) < \infty$ et $\sigma^2 = \mathbb{V}ar_\theta(X) < \infty$. On a alors :

$$\mathbb{E}_\theta(\overline{X}) = \mu, \quad \mathbb{V}\text{ar}_\theta(\overline{X}) = \frac{\sigma^2}{n};$$

$$\mathbb{E}_\theta(S_c^2) = \sigma^2.$$

Propriétés au cas d'un échantillon normal

Soit $\mathbf{X} = (X_1, \ldots, X_n)$ un échantillon provenant d'une population $\mathcal{N}(\mu, \sigma^2)$. Alors,

a. $\overline{X}$ suit une loi $\mathcal{N}(\mu, \frac{\sigma^2}{n})$

b. $\overline{X}$ et S_c^2 sont indépendantes.

c. $T = \dfrac{\sqrt{n}(\overline{X} - \mu)}{S_c}$ (*non statistique*) suit une loi de Student à (n-1) degré de liberté.

d. $K^2 = \dfrac{(n-1)S_c^2}{\sigma^2}$ (*non statistique*) suit une loi de khi-deux à (n-1) degré de liberté $\chi^2_{(n-1)}$.

1.3 Inférences

Objectifs généraux : Extraire d'informations essentielles contenues dans des données ayant des éléments inconnus et intrinsèquement imprévisibles et les interpréter. On s'intéresse par exemple aux estimations, aux tests, aux vérifications de modèle.

- *Estimer* au mieux θ
- *Tester* l'hypothèse nulle $\mathbf{H}_0 \ : \theta \in \Theta_0$ contre $\mathbf{H}_1 \ : \theta \in \Theta_1$
- *Vérifier* le modèle :

existe-t-il un $\theta^* \in \Theta$, tel que $\mathbb{P}_{\theta^*}$ modélise bien les données ?

Statistique (statistic) : Une statistique $T(\mathbf{X})$ est une fonction de l'échantillon, qui ne ne doit contenir aucun paramètre inconnu. Comme un échantillon est n variables aléatoires, une statistique, elle aussi, est une variable aléatoire.

Chapitre II Estimations

On estimer un paramètre θ en utilisant une fonction de l'échantillon bien choisie : une "bonne" statistique $T(\mathbf{X}) = T(X_1, \ldots, X_n)$.

§ 2.1 critères pour choisir les estimateurs :

- Biais de l'estimateur : $b(\theta) = \mathbb{E}_\theta\big(T(\mathbf{X})\big) - \theta$

  $T(\mathbf{X})$ est sans biais si $b(\theta) = 0,\ \forall\ \theta \in \Theta.$

- Erreur quadratique* $r_T(\theta) = \mathbb{E}_\theta\big(T(\mathbf{X}) - \theta\big)^2 = \mathbb{V}ar_\theta(T(\mathbf{X})) + b(\theta)^2$

- MiniMax $\inf_T \sup_{\theta\in\Theta} r(\theta)$

- Vitesse de convergence $(T(\mathbf{X}) - \theta) \longrightarrow 0.$

§ 2.2 Méthode du maximum de vraisemblance

La méthode du maximum de vraisemblance est la technique la plus populaire pour obtenir des estimateurs, souvent les meilleures dans les cas classiques.
Lorsque les lois de probabilités sont discrètes, on a

$$\mathbf{P}(X_i = x_i) = \mathbf{P}_\theta(x_i).$$

Par contre, si les lois sont absolument continues, on a

$$\mathbf{P}(a < X_i \leq b) = \int_a^b f_\theta(x)dx\ .$$

Définition :
La fonction de **vraisemblance** est définie par

$$\theta \longrightarrow L(\theta|\mathbf{x}) = \ L(\theta|x_1, \ldots, x_n)$$

$$\text{où } \begin{cases} L(\theta|x_1, \ldots, x_n) & = \Pi_{j=1}^n \mathbf{P}_\theta(x_j) \quad \text{si discrèt} \\ L(\theta|x_1, \ldots, x_n) & = \Pi_{j=1}^n f_\theta(x_j) \quad \text{si absolument continu} \end{cases}$$

La fonction du **logarithme de vraisemblance** est définie par

$$\theta \longrightarrow l(\theta|\mathbf{x}) = \ \ln\big(L(\theta|x_1, \ldots, x_n)big)$$

$$\text{où } \begin{cases} l(\theta|x_1, \ldots, x_n) & = \displaystyle\sum_{j=1}^n \ln\big(\mathbf{P}_\theta(x_j)\big) \quad \text{si discrèt} \\ l(\theta|x_1, \ldots, x_n) & = \displaystyle\sum_{j=1}^n \ln\big(f_\theta(x_j)\big) \quad \text{si absolument continu} \end{cases}$$

On introduit donc **estimateur du maximum de vraisemblance** comme $\hat{\theta}(x_1, \ldots, x_n)$ la fonction de $\mathbf{X} = (X_1, \ldots, X_n)$ (l'échantillon) tel que

$$\hat{\theta}(x_1, \ldots, x_n) = \arg\max_{\theta\ \in\ \Theta} L(\theta|x_1, \ldots, x_n),$$

$\forall\ \mathbf{x} = (x_1, \ldots, x_n)$
ou de manière équivalente :

$$\hat{\theta}(x_1, \ldots, x_n) = \arg\ \max_{\theta\ \in\ \Theta}\ l(\theta | x_1, \ldots, x_n),$$

$\forall\ \mathbf{x} = (x_1, \ldots, x_n)$

Exemple 3 : loi gaussienne
Soit $(x_1, \ldots, x_n)$ une réalisation d'un échantillon $(X_1, \ldots, X_n)$ de loi gaussienne $N(\mu, \sigma^2)$.
Si on note $v = \sigma^2$, on a alors

$$\mathbf{P}(a < X \leq b) = \int_a^b \frac{1}{2\pi\sqrt{v}} \exp(-\frac{1}{2v}(t-\mu)^2) \mathrm{d}\, t$$

$$L(\theta | x_1, \ldots, x_n) = \frac{1}{(2\pi)^{n/2} v^{n/2}} \exp\big(-\frac{1}{2v}\sum_{j=1}^{n}(x_j - \mu)^2\big)$$

$$l(\theta | x_1, \ldots, x_n) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(v) - \frac{1}{2v}\sum_{j=1}^{n}(x_j - \mu)^2$$

$$\frac{\partial}{\partial \mu} l\big((\mu, v) | x_1, \ldots, x_n\big) = -0 - 0 - \big(-\frac{1}{2v}\sum_{j=1}^{n}(x_j - \mu)\big)$$

$$\frac{\partial}{\partial v} l\big((\mu, v) | x_1, \ldots, x_n\big) = -0 - \frac{n}{2v} + \frac{1}{2v^2}\sum_{j=1}^{n}(x_j - \mu)^2$$

$$\frac{\partial}{\partial \mu} l\big((\mu, v) | x_1, \ldots, x_n\big) = 0 \Longrightarrow \widehat{\mu}_{MV} = \overline{X} = \frac{1}{n}\sum_{j=1}^{n} X_j$$

$$\frac{\partial}{\partial v} l\big((\mu, v) | x_1, \ldots, x_n\big) = 0 \Longrightarrow \hat{v} = \frac{1}{n}\sum_{j=1}^{n}(X_j - \hat{\mu})^2.$$

$$\text{C'est-à-dire } \widehat{\sigma^2_{MV}} = \frac{1}{n}\sum_{j=1}^{n}(X_j - \overline{X})^2.$$

Ce dernier est un estimateur biaisé, alors que l'estimateur sans biais et de variance minimale est donné par

$$\widehat{\sigma^2} = \frac{1}{n-1}\sum_{j=1}^{n}(X_j - \overline{X})^2.$$

Théorème Loi forte des grands nombres

Soit $(X_n, n \in \mathbb{N}^*)$ une suite de v.a. i. i. d. avec
$\mu = \mathbb{E}(X_j) < \infty$. On a

$$\mathbb{P}\Big(\lim_{n\to\infty}\frac{X_1+\cdots+X_n}{n}=\mathbb{E}(X_1)\Big)=1.$$

Théorème Central Limite

Soit $(X_n, n\in\mathbb{N}^*)$ une suite de v.a. i. i.d. avec $\mu=\mathbb{E}(X_j)<\infty$, $\sigma^2=\mathbb{V}ar(X_j^2)<\infty$.
On a

$$\lim_{n\to\infty}\mathcal{L}\Big(\frac{\sqrt{n}(\overline{X}-\mu)}{\sigma}\Big)=\mathcal{N}(0,1).$$