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

Année 2018/2019
Statistique
Etude de cas L3

Chapitre 1

Notions statistiques

1.1 Variables aléatoires

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

Exemple 1

n= 20 réalisations d'une variable aléatoire X de loi de Bernoulli $B(1, p)$ avec $p = 0.3$:

$$0, 0, 0, 1, 1, 1, 0, 1, 0, 1, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0$$

Exemple 2

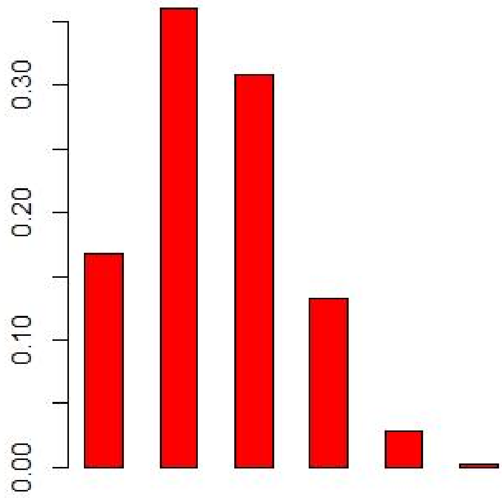
n= 20 réalisations d'une variable aléatoire X de loi binomiale $B(5, p)$ avec $p = 0.3$:

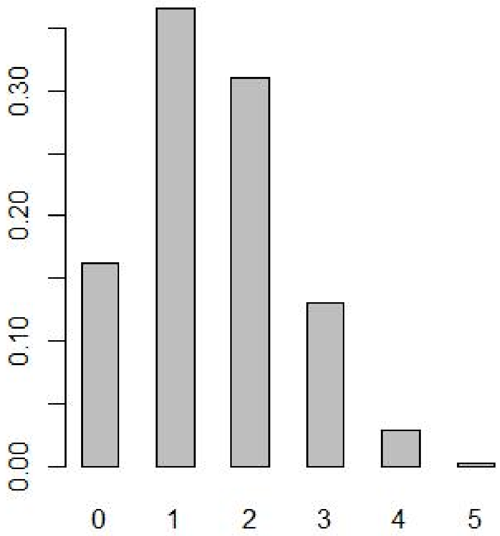
$$1, 3, 0, 2, 2, 2, 1, 1, 2, 1, 1, 1, 0, 1, 1, 3, 3, 3, 1, 2$$

n=10000 réalisations d'une v. a.X de la loi binomiale B(5,0.3) :

TABLE 1.1 – Proportion

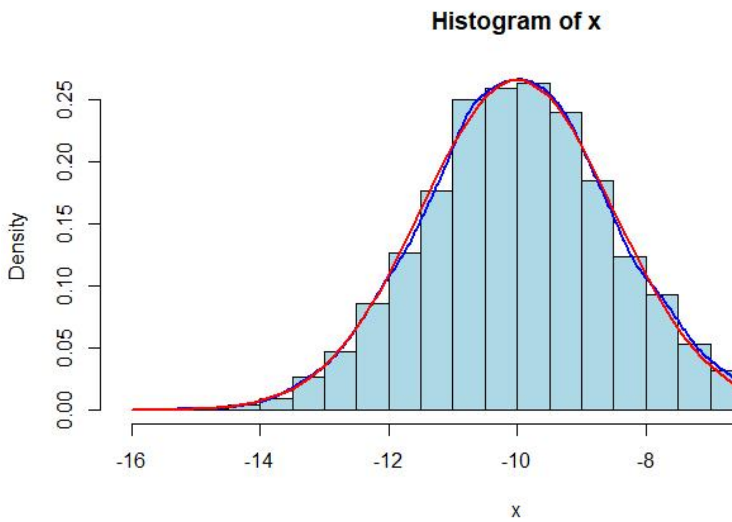
X	0	1	2	3	4	5
proportion	0.1628	0.3654	0.3099	0.1302	0.0294	0.0023
probabilité	0.16807	0.36015	0.30870	0.13230	0.02835	0.00243





Exemple 3

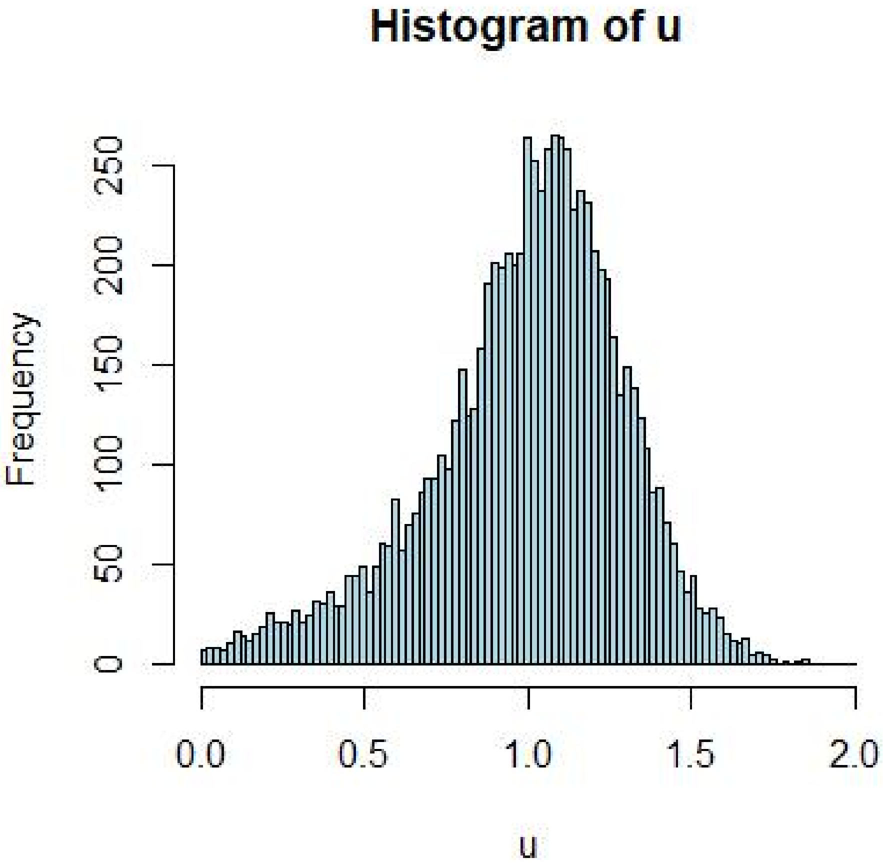
5000 réalisations d'une variable aléatoire X de loi gaus-



sienne $N(\mu, \sigma^2)$ avec $\mu = -10, \sigma = 1.5$:

Exemple 4

On s'intéresse à la largeur des anneaux de croissance des arbres. le fichier "treering" dans la librairie "datasets" avec $N = 7980$.



On prélève au hasard une valeur X dans cet ensemble de 7980 valeurs : $x = 1.004$.

Rem On note souvent X ou (Y) une variable aléatoire et x (ou y une des ses réalisations. (souvent une infinité !)

1.2 Modélisation statistique :

Modéliser, c'est structurer et simplifier, choisir quoi identifier et quoi différencier, ce qui est important et ce qui est accessoire, relier entre eux les évènements.

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

La répartition des valeurs d'une variable est associe à une loi de probabilité :

$$\text{Population} \iff \mathbb{P}_\theta, \ \big(\theta \in \Theta\big)$$

Attention : un sous-ensemble de n individus ne peut pas être "représentatif" de la population $(n << N)$..

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

<u>Exemples</u> :

1. On prélève 3 boules dans une urne contenant 7 de valeurs $5^{-1}, 10^{-1}, \ldots, 5^5$ euros.

Avec remise : 7^3 (=343)

Sans remise : choose(7,3) (=56)

L= $5^{-1:5}$

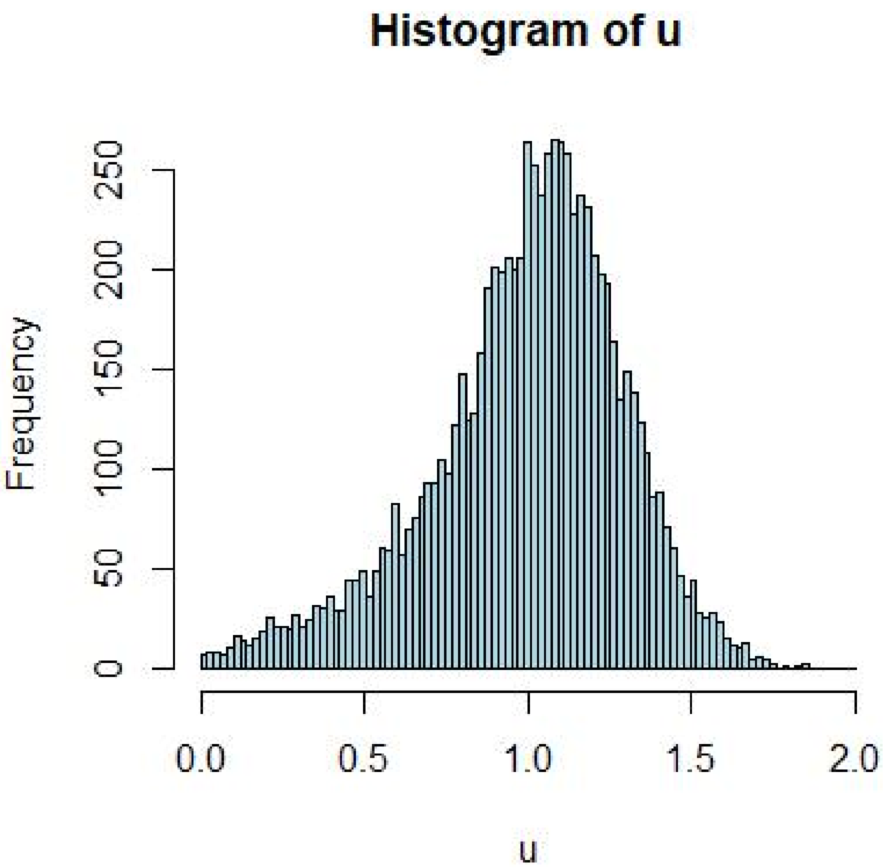
L=1 :7 ; combn(L, 3)

for (i in 1 :10){ print (sample(L, 3, replace=T))}

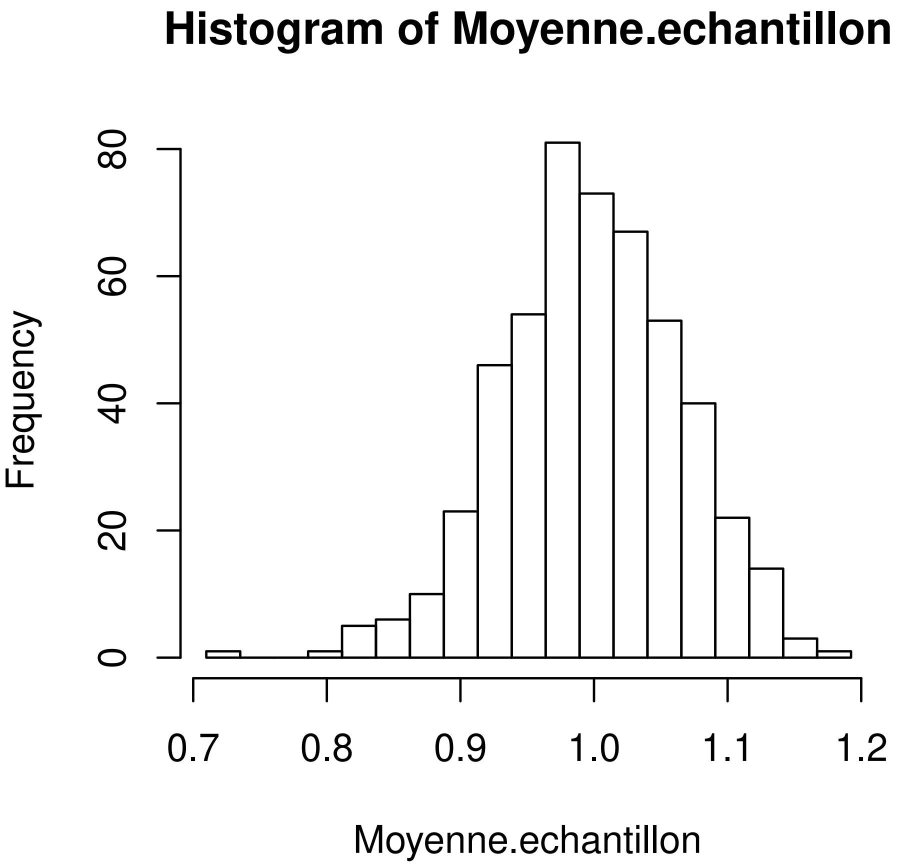
2. les résultats après avoir lancé un dé $n = 20$ fois. sample(6, 20, replace=T)

3. les poids de $n = 10$ étudiants choisis au hasard en France

4. On s'intéresse à la largeur des anneaux de croissance des arbres. le fichier "treering" dans la librairie "datasets" avec $n = 7980$.



On simule M=500 réalisations d'un échantillon de taille n=20 et on constate



Données : $x_1, \ldots, x_n$

Modèle : $\mathbf{x} = (x_1, \ldots, x_n)$ une réalisation d'un échantillon
$\mathbf{X} = (X_1, \ldots, X_n)$ (n v.a. i.i.d.) de loi $\mathbb{P}_\theta, \theta \in \Theta$ paramètre inconnu.

Paramètres

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 :

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

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

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

.

<u>Paramètres de dispersion</u> : 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.) :

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

• écart-type de la variable X

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

Paramètre de forme Paramètre d'asymétrie :

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

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

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

Rem 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 :

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

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 $X_{(1)} = \min\limits_{1 \le j \le n} (X_j), X_{(2)} = \min\limits_{1 \le j \le n, j \ne (1)} (X_j) \ldots, X_{(n)} = \max\limits_{1 \le j \le n} (X_j)$

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

• 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)

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

.

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

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

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

Propriétés

Soit $\mathbf{X} = (X_1, \ldots, X_n)$ un échantillon aléatoire de taille n provenant d'une population $f(x)$ avec variance finie σ^2. Alors,

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

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

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

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

c. $\dfrac{(n-1)S_c^2}{\sigma^2}$ suit une loi de khi-deux à (n-1) degré de liberté $\chi^2_{(n-1)}$.

d. $\dfrac{\sqrt{n}(\overline{X} - \mu)}{S_c}$ suit une loi de Student à (n-1) degré de liberté.

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 2

Estimations

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

2.1 critère

Critères pour choisir les estimateurs :

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

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

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

  l' estimateur des Erreurs quadratiques
  uniformément les plus petites
  parmi les estimateurs sans biais :
  $\Longrightarrow$ sans biais de variances uni.t minimales

- 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.

Exemple 1 : loi de Bernoulli
Un frère dispute d'un précieux tableau à sa soeur. La soeur propose d'utiliser la lancer d'une pièce pour désigner l'heureux héritier.

X une variable aléatoire de loi de Bernoulli $B(1,p)$

$$\mathbf{P}(X=k) = p^k(1-p)^{1-k},\ k \ \in \ \{0,1\}.$$

Soit $(x_1,\dots,x_n)$ une réalisation d'un échantillon $(X_1,\dots,X_n)$ de loi de Bernoulli.

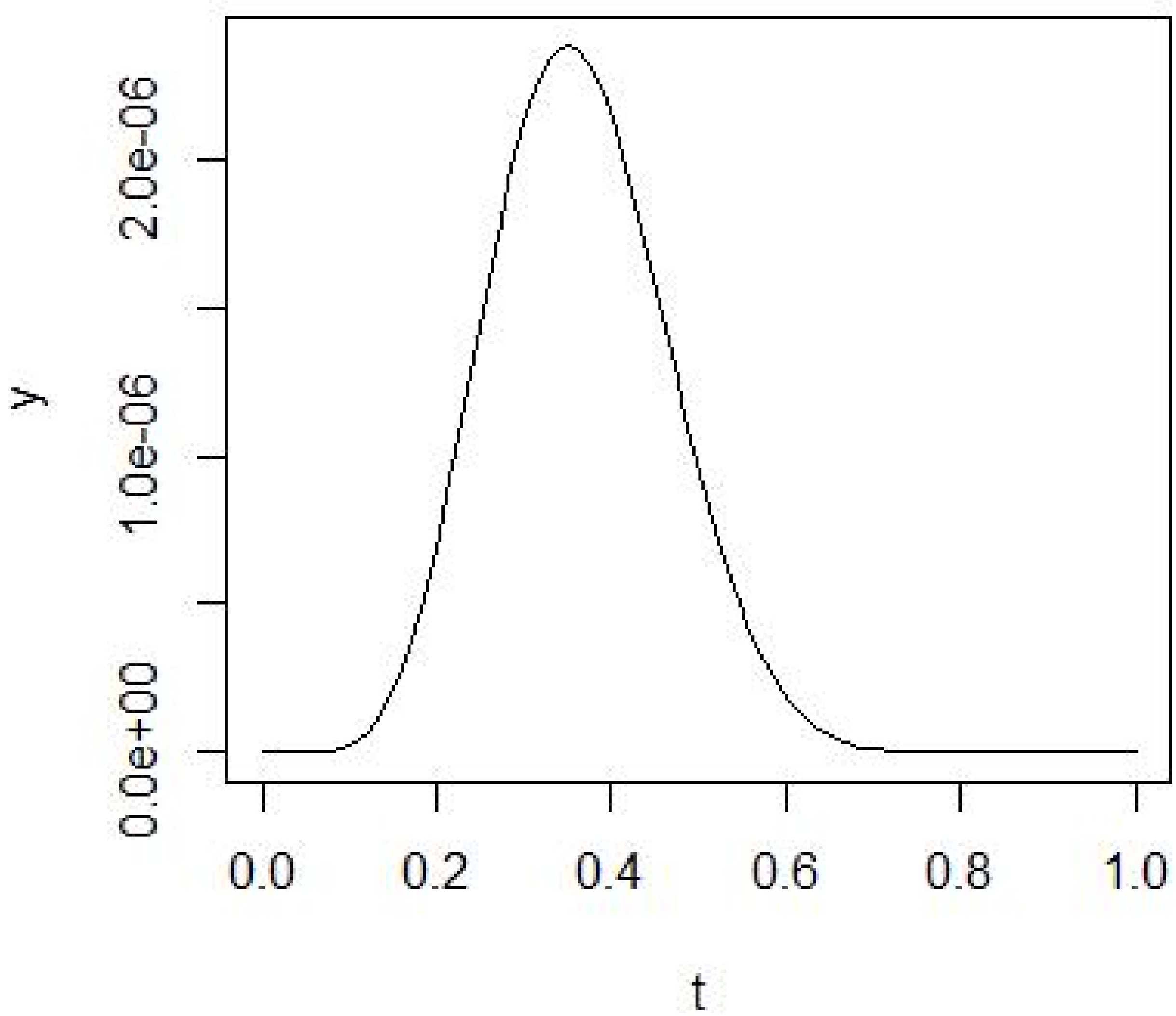
$$\begin{aligned}
&\mathbf{P}(X_1=x_1, X_2=x_2,\dots,X_n=x_n)\\
&= \mathbf{P}(X_1=x_1)\mathbf{P}(X_2=x_2)\cdots\mathbf{P}(X_n=x_n)\\
&= p^{x_1}(1-p)^{1-x_1}p^{x_2}(1-p)^{1-x_2}\cdots p^{x_n}(1-p)^{1-x_n}\\
&= p^{\sum_{j=1}^n x_j}(1-p)^{n-\sum_{j=1}^n x_j}.
\end{aligned}$$

donc $\mathbf{P}(X_1=x_1, X_2=x_2,\dots,X_n=x_n) = p^{\sum_{j=1}^n x_j}(1-p)^{n-\sum_{j=1}^n x_j}$.

Cette dernière quantité est une fonction de p, appelée fonction de vraisemblance $L(p|x_1,\dots,x_n)$:

$$L(p|x_1,\dots,x_n) = p^{\sum_{j=1}^n x_j}(1-p)^{n-\sum_{j=1}^n x_j}.$$

avec le graphique suivant :



Si on étudie son logarithme :

$$l(p|x_1,\ldots,x_n) = \log L = \sum_{j=1}^{n} x_j \log(p) + (n - \sum_{j=1}^{n} x_j)\log(1-p).$$

$$\frac{\partial}{\partial p} l(p|x_1,\ldots,x_n) = 0 \Longrightarrow p = \frac{1}{n}\sum_{j=1}^{n} x_j \;\Longrightarrow \hat{p} = \overline{X} = \frac{1}{n}\sum_{j=1}^{n} X_j.$$

Comme $\dfrac{\partial^2}{\partial p^2} l(p|x_1,\ldots,x_n) = -\dfrac{\sum_{j=1}^n X_j}{p} - \dfrac{n - \sum_{j=1}^n X_j}{1-p}$

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

Rappels :
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

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

La fonction de **vraisemblance** est définie par

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

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

On introduit donc estimateur du maximum de vraisemblance comme la fonction de $\mathbf{X}$ (l'échantillon) tel que

$$\hat{\theta}(\mathbf{x}) = \text{Argmax}_{\theta\ \in\ \Theta}\ L(\theta|\mathbf{x}),\ \forall\ \mathbf{x}$$

ou de manière équivalente :

$$\hat{\theta}(\mathbf{x}) = \text{Argmax}_{\theta\ \in\ \Theta}\ l(\theta|\mathbf{x}),\ \forall\ \mathbf{x}$$

Exemple 2 : loi de Poisson

Soit $(x_1, \ldots, x_n)$ une réalisation d'un échantillon $(X_1, \ldots, X_n)$ de loi de Poisson $P(\lambda)$. On a

$$\mathbf{P}(X = x_j) = \frac{\lambda^{x}{}_{j}}{x_j!} \exp(-\lambda) \Longrightarrow$$

$$L(\lambda | x_1, \ldots, x_n) = \frac{\lambda^{\sum_{j=1}^{n} x_j}}{\Pi_{j=1}^{n} x_j!} \exp(-n\lambda)$$

$$l(\lambda | x_1, \ldots, x_n) = \left(\sum_{j=1}^{n} x_j \right) \log(\lambda) - \sum_{j=1}^{n} \log(x_j!) - n\lambda$$

$$\frac{\partial}{\partial \lambda} l(\lambda | x_1, \ldots, x_n) = \frac{\left(\sum_{j=1}^{n} x_j \right)}{\lambda} - 0 - n.$$

L'équation $\quad \frac{\partial}{\partial \lambda} l(\lambda | x_1, \ldots, x_n) = 0 \Longrightarrow \lambda(\mathbf{x}) = \dfrac{1}{n} \displaystyle\sum_{j=1}^{n} x_j, \ \forall \ \mathbf{x}$

Comme $\quad \frac{\partial^2}{\partial \lambda^2} l(\lambda | x_1, \ldots, x_n) = -\dfrac{\left(\sum_{j=1}^{n} x_j \right)}{\lambda^2} < 0, \ \forall \ \lambda$

On en déduit $\quad \hat{\lambda}_{MV} = \overline{X} = \frac{1}{n} \sum_{j=1}^{n} X_j.$

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

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

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

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

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

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

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

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

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

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

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

3.3 Comparaison des estimateurs

Exemple 4, autrement Estimation de μ (la moyenne théorique) de largeur d'anneaux)

On s'intéresse à la largeur des anneaux de croissance des arbres. On prend le fichier "treering" dans la librairie "datasets" avec $N = 7980$ (fini) comme étant le cardinal de la population. On note $\mu = 0.9968$ la moyenne (théorique) de la population.

```
Echantillon=matrix(NA, nrow=M,ncol=n) # NA : sans valeur
# pour stocker M réalisations de l'échantillon de taille n
Moyenne.echantillon=rep(NA,M) ;
# pour stocker M réalisations de la moyenne d'échantillon
for (j in 1:M) {
    Echantillon[j, ] = sample(Population, n, replace = T)
    Moyenne.echantillon[j] = mean(Echantillon[j, ]) }
Echantillon ; Moyenne.echantillon
```

et on souhaite avoir une idée sur la grandeur du paramètre μ. Pour ce faire, on va construire $M = 100$ réalisations, basé sur un échantillon de taille $n = 65$, de l'intervalle de confiance à 95% sur le paramètre π en utilisant la méthode de Wald.

1) Sauvegarder ces 7980 valeurs dans un vecteur nommé **Population**.

2) Générer $M = 500$ réalisations de l'échantillon $(X_1, ..., X_n)$ de taille $n = 65$.

3) Appliquer la formule de l'intervalle de confiance à 95% (la méthode de Wald) sur le paramètre π (c.f. TD 5, exo-7) sur les 50 réalisations de l'échantillon pour obtenir $M = 50$ réalisations de l'intervalle de confiance à 95 %.

4) Calculer la vraie valeur de π en utilisant la totalité des 7980 valeurs de la population, puis déterminer la proportion des IC contenant cette valeur π. Commenter vos résultats.

La méthode du maximum de vraisemblance est la technique la plus populaire pour obtenir des estimateurs, souvent les meilleures dans les cas classiques.

Exemple 1 : loi de Bernoulli
Un frère dispute d'un précieux tableau à sa soeur. La soeur propose d'utiliser la lancer d'une pièce pour désigner l'heureux héritier.

X une variable aléatoire de loi de Bernoulli $B(1, p)$

$$\mathbb{P}(X = k) = p^k(1-p)^{1-k}, \ k \ \in \ \{0, 1\}.$$

Soit $(x_1, \ldots, x_n)$ une réalisation d'un échantillon $(X_1, \ldots, X_n)$ de loi de Bernoulli.

$$\begin{aligned}
&\mathbb{P}(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) \\
&= \mathbb{P}(X_1 = x_1)\mathbf{P}(X_2 = x_2) \cdots \mathbf{P}(X_n = x_n) \\
&= p^{x_1}(1-p)^{1-x_1} p^{x_2}(1-p)^{1-x_2} \cdots p^{x_n}(1-p)^{1-x_n} \\
&= p^{\sum_{j=1}^n x_j}(1-p)^{n-\sum_{j=1}^n x_j}.
\end{aligned}$$

donc $\mathbb{P}(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) = p^{\sum_{j=1}^n x_j}(1-p)^{n-\sum_{j=1}^n x_j}$.

Cette dernière quantité est une fonction de p, appelée fonction de vraisemblance $L(p|x_1,\ldots,x_n)$:

$$L(p|x_1,\ldots,x_n) = p^{\sum_{j=1}^n x_j}(1-p)^{n-\sum_{j=1}^n x_j}.$$

avec le graphique suivant :

Si on étudie son logarithme :

$$l(p|x_1,\ldots,x_n) = \log L = \sum_{j=1}^n x_j \log(p) + (n - \sum_{j=1}^n x_j)\log(1-p).$$

$$\frac{\partial}{\partial p} l(p|x_1,\ldots,x_n) = 0 \Longrightarrow p = \frac{1}{n}\sum_{j=1}^n x_j \;\Longrightarrow \hat{p} = \overline{X} = \frac{1}{n}\sum_{j=1}^n X_j.$$

$$\text{Comme } \frac{\partial^2}{\partial p^2} l(p|x_1,\ldots,x_n) = -\frac{\sum_{j=1}^n X_j}{p} - \frac{n-\sum_{j=1}^n X_j}{1-p}$$

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

Rappels :

Lorsque les lois de probabilités sont discrètes, on a

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

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

$$\mathbb{P}(a < X_i \le b) = \int_a^b f_\theta(x)dx \; .$$

La fonction de **vraisemblance** est définie par

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

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

On introduit donc estimateur du maximum de vraisemblance comme la fonction de $\mathbf{X}$ (l'échantillon) tel que

$$\hat{\theta}(\mathbf{x}) = \text{Argmax}_{\theta\, \in\, \Theta}\ L(\theta|\mathbf{x}),\ \forall\ \mathbf{x}$$

ou de manière équivalente :

$$\hat{\theta}(\mathbf{x}) = \text{Argmax}_{\theta\, \in\, \Theta}\ l(\theta|\mathbf{x}),\ \forall\ \mathbf{x}$$

Exemple 2 : loi de Poisson

Soit $(x_1, \ldots, x_n)$ une réalisation d'un échantillon $(X_1, \ldots, X_n)$ de loi de Poisson $P(\lambda)$. On a

$$\mathbf{P}(X = x_j) = \frac{\lambda^{x}{}_{j}}{x_j!} \exp(-\lambda) \Longrightarrow$$

$$L(\lambda | x_1, \ldots, x_n) = \frac{\lambda^{\sum_{j=1}^{n} x_j}}{\Pi_{j=1}^{n} x_j!} \exp(-n\lambda)$$

$$l(\lambda | x_1, \ldots, x_n) = \Big(\sum_{j=1}^{n} x_j\Big) \log(\lambda) - \sum_{j=1}^{n} \log(x_j!) - n\lambda$$

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

L'équation $\quad \frac{\partial}{\partial \lambda} l(\lambda | x_1, \ldots, x_n) = 0 \Longrightarrow \lambda(\mathbf{x}) = \dfrac{1}{n} \displaystyle\sum_{j=1}^{n} x_j, \ \forall \ \mathbf{x}$

Comme $\quad \frac{\partial^2}{\partial \lambda^2} l(\lambda | x_1, \ldots, x_n) = -\dfrac{\big(\sum_{j=1}^{n} x_j\big)}{\lambda^2} < 0, \ \forall \ \lambda$

On en déduit $\quad \hat{\lambda}_{MV} = \overline{X} = \frac{1}{n} \sum_{j=1}^{n} X_j.$

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

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

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

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

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

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

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

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

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

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

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

2.3 Comparaison des estimateurs