STT 1000 - STATISTIQVES

ARTHVR CHARPENTIER

Maximum de Vraisemblance

Vraisemblance (Likelihood)

Soit $\boldsymbol{Y} = (y_1, \cdots, y_n)$ un échantillon i.i.d. de variables de loi f_θ. La fonction de vraisemblance est

$$\mathcal{L}(\theta|\boldsymbol{y}) = \prod_{i=1}^{n} f_\theta(y_i)$$

L'estimation du maximum de vraisemblance est

$$\widehat{\theta}(\boldsymbol{y}) = \underset{\theta\in\Theta}{\text{argmin}}\big\{\mathcal{L}(\theta|\boldsymbol{y})\big\} \text{ et } \widehat{\theta}(\boldsymbol{Y}) = \underset{\theta\in\Theta}{\text{argmin}}\big\{\mathcal{L}(\theta|\boldsymbol{Y})\big\}$$

Maximum de Vraisemblance

Log-Vraisemblance

Soit $\boldsymbol{Y} = (y_1, \cdots, y_n)$ un échantillon i.i.d. de variables de loi f_θ. La fonction de log-vraisemblance est

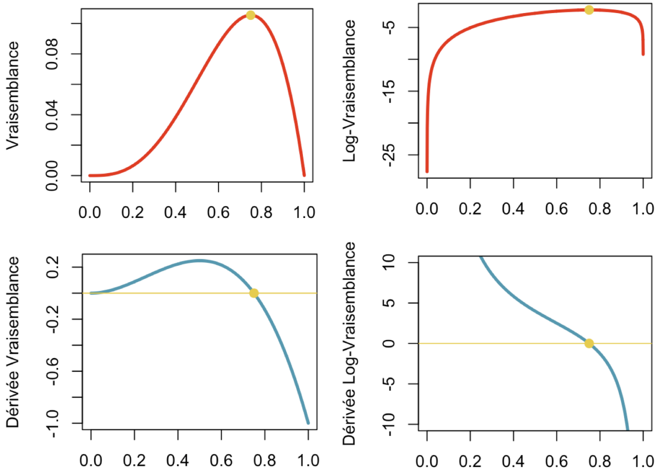
$$\log \mathcal{L}(\theta|\boldsymbol{y}) = \sum_{i=1}^{n} \log f_\theta(y_i)$$

L'estimation du maximum de vraisemblance est

$$\widehat{\theta}(\boldsymbol{y}) = \underset{\theta\in\Theta}{\text{argmin}}\left\{ \log \mathcal{L}(\theta|\boldsymbol{y})\right\} \text{ et } \widehat{\theta}(\boldsymbol{Y}) = \underset{\theta\in\Theta}{\text{argmin}}\left\{ \log \mathcal{L}(\theta|\boldsymbol{Y})\right\}$$

Vraisemblance, cas $\mathcal{B}(p)$

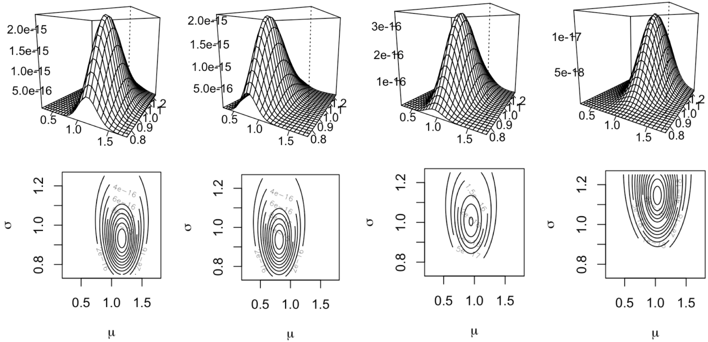
$$\boldsymbol{y} = \{0, 1, 1, 1\}, \ Y_i \sim \mathcal{B}(\theta).$$



Vraisemblance, cas $\mathcal{N}(\mu, \sigma^2)$

Pour les paramètres univariés, on peut visualiser la vraisemblance, mais c'est plus compliqué en dimension plus grande...

Vraisemblance $\mathcal{L}(\mu, \sigma^2)$ pour 4 échantillons $\boldsymbol{y}$



Vraisemblance, cas $\mathcal{B}(p)$

Example 1 : on a fait un sondage sur 15 personnes pour savoir s'ils appréciaient le cours de STT1000, quelle est l'estimation par maximum de vraisemblance de la proportion de gens satisfaits ?

- ce que nous dit la théorie

$$\mathcal{L}(p;\boldsymbol{x}) = \prod_{i=1}^{n} f(x_i;p) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} = p^{s_n}(1-p)^{n-s_n},\ s_n = \sum_{i=1}^{n} x_i$$

$$\log \mathcal{L}(p;\boldsymbol{x}) = s_n \log(p) + (n-s_n)\log(1-p)$$

$$\frac{\partial}{\partial p}\log \mathcal{L}(p;\boldsymbol{x}) = \frac{\partial}{\partial p} s_n x \log(p) + (n-s_n)\log(1-p) = \frac{s_n}{p} - \frac{n-s_n}{1-p}$$

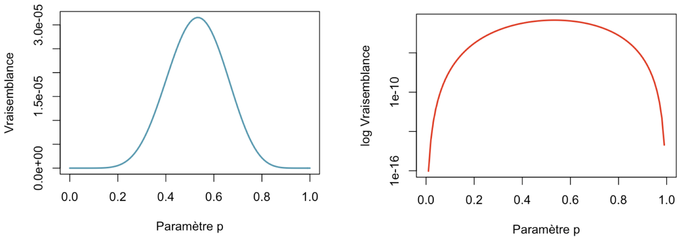
$$\left.\frac{\partial}{\partial p}\log \mathcal{L}(p;\boldsymbol{x})\right|_{p=\widehat{p}} = 0 \text{ si et seulement si } \frac{s_n}{\widehat{p}} = \frac{n-s_n}{1-\widehat{p}},\ \text{soit } \widehat{p} = \frac{s_n}{n} = \overline{x}$$

Vraisemblance, cas $\mathcal{B}(p)$

- ce que nous dit la pratique

Traçons la fonction de (log)vraisemblance $p \mapsto \mathcal{L}(p; \boldsymbol{x})$

```
> n=15
> x = c(1, 1, 1, 0, 1, 0, 0, 0, 0, 1, 1, 1, 0, 1, 0)
> vraisemblance = function(p) prod(dbinom(x,size = 1,
    prob = p))
> vect_p = seq(0,1,by=0.01)
> plot(vect_p,Vectorize(vraisemblance)(vect_p))
```



Vraisemblance, cas $\mathcal{B}(p)$

- ce que nous dit la pratique

On peut chercher le maximum de la fonction $p \mapsto \mathcal{L}(p; \boldsymbol{x})$

```
> optim(par = .5,fn = function(z) -vraisemblance(z))
$par
[1] 0.5333252

$value
[1] -3.155276e-05
```

La théorie nous avait dit que ce maximum a une forme particulière, $\widehat{p} = \overline{x}$

```
> mean(x)
[1] 0.5333333
```

Vraisemblance, cas $\mathcal{B}(p)$

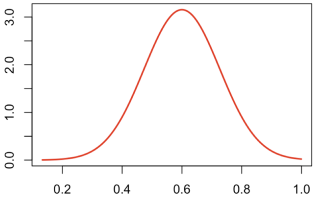
- ce que nous disent les mathématiques

Comme $\widehat{p}(\boldsymbol{x}) = \overline{x}$, on peut utiliser la loi des grands nombres,

$$Z_n = \sqrt{n}\frac{\widehat{p}(\boldsymbol{X}) - p}{\sqrt{p(1-p)}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$$

mais ici $n = 15$ (approximation Gaussienne peut être mauvaise)
Si $p = 60\%$ la distribution (approchée) de $\widehat{p}(\boldsymbol{X})$ serait

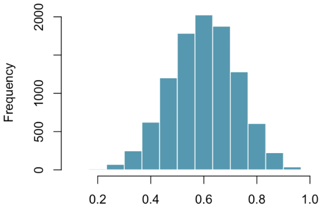
```
> u=seq(2/15,1,by=.001)
> plot(u,dnorm(u,.6,sqrt(.4*.6/15)))
```



Vraisemblance, cas $\mathcal{B}(p)$

- ce que nous disent les simulations, si on suppose $\theta = 60\%$

```
> theta=rep(NA,1e4)
> for(s in 1:1e4){
+     x=sample(0:1,size = n,prob = c(.4,.6),replace=TRUE)
+     neglogL = function(p) -sum(log(dbinom(x,size = 1,prob = p)))
+     theta[s] = optim(par = .5,fn = neglogL)$par
+ }
> hist(theta)
```



Méthode des Moments

Méthode des Moments

Soit $\boldsymbol{Y} = (y_1, \cdots, y_n)$ un échantillon i.i.d. de variables de loi f_θ. Soient $m_k(\theta) = \mathbb{E}[Y^k]$ où $Y \sim f_\theta$, et $\widehat{m}_k = \frac{1}{n}\sum_{i=1}^{n} y_i^k$ le moment empirique. Soit $\widehat{\theta} = (\widehat{\theta}_1, \cdots, \widehat{\theta}_d)$ la solution du système d'équations

$$\begin{cases} m_1(\widehat{\theta}) = \widehat{m}_1 \\ \quad\vdots \\ m_d(\widehat{\theta}) = \widehat{m}_d \end{cases}$$

Note: on peut parfois considérer les moments centrés (i.e. $\text{Var}[Y]$ au lieu de $\mathbb{E}[Y^2]$)

Méthode des moments, cas $\mathcal{B}(p)$

Example 2 : on a fait un sondage sur 15 personnes pour savoir s'ils appréciaient le cours de STT1000, quelle est l'estimation par la méthode des moments de la proportion de gens satisfaits ?

• ce que nous dit la théorie

$$\mathbb{E}(X) = m_1(p) = p$$

or $\widehat{m}_1 = \overline{x}$ donc $\widehat{p}(x) = \overline{x}$

• ce que nous disent les mathématiques
Comme $\widehat{p}(\boldsymbol{x}) = \overline{x}$, on peut utiliser la loi des grands nombres,

$$Z_n = \sqrt{n}\frac{\widehat{p}(\boldsymbol{X}) - p}{\sqrt{p(1-p)}} \overset{\mathcal{L}}{\to} \mathcal{N}(0,1)$$

mais ici $n = 15$ (approximation Gaussienne peut être mauvaise)

Méthode des moments, cas $\mathcal{B}(n,p)$

Que se passe-t-il si $Y_i \sim \mathcal{B}(n,p)$, où n est aussi inconnu ?

$$\mathbb{E}[Y] = np \text{ et } \text{Var}[Y] = np(1-p)$$

On va alors résoudre

$$\begin{cases} \widehat{n}\widehat{p} = \overline{x} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} y_i \\ \widehat{n}\widehat{p}(1-\widehat{p}) = s^2 = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} \left(y_i - \overline{y}\right)^2 \end{cases}$$

soit

$$\widehat{p} = \frac{\overline{y} - s^2}{\overline{y}} \text{ et } \widehat{n} = \frac{\overline{y}^2}{\overline{y} - s^2}$$

Note: il est possible d'avoir $\widehat{p} < 0$

Maximum de Vraisemblance vs Méthode des Moments

Example 3 : On observe des données modélisées par une loi de densité $\mapsto \theta y^{\theta-1}$ pour $y \in [0,1]$. Quels sont les estimateurs de θ?

```
> y = c(0.685, 0.754, 0.853, 0.973, 0.633, 0.97,
   0.984, 0.888, 0.876, 0.451, 0.637, 0.609, 0.898,
   0.761, 0.928, 0.819, 0.91, 0.998, 0.758, 0.931,
   0.981, 0.642, 0.885, 0.553, 0.686)
```

- méthode des moments

$$\mathbb{E}[Y] = \int_0^1 y \cdot \theta y^{\theta-1} dy = \theta \int_0^1 y^{\theta} dy = \theta \left[\frac{y^{\theta+1}}{\theta+1}\right]_0^1 = \frac{\theta}{\theta+1}$$

L'estimateur par la méthode des moments vérifie

$$\overline{y} = \frac{\widehat{\theta}}{\widehat{\theta}+1} \text{ soit } \widehat{\theta} = \frac{\overline{y}}{1-\overline{y}}$$

Maximum de Vraisemblance vs Méthode des Moments

- maximum de vraisemblance

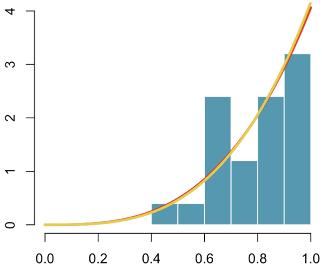
$$\log \mathcal{L}(\theta;\boldsymbol{y}) = n\log(\theta) + (\theta-1)\sum_{i=1}^{n}\log(y_i)$$

$$\frac{\partial \log \mathcal{L}(\theta;\boldsymbol{y})}{\partial\theta} = \frac{n}{\theta} + \sum_{i=1}^{n}\log(y_i)$$

$$\left.\frac{\partial \log \mathcal{L}(\theta;\boldsymbol{y})}{\partial\theta}\right| -\theta = \widehat{\theta} = 0 \text{ si } \frac{n}{\theta} = -\sum_{i=1}^{n}\log(y_i), \text{ i.e. } \widehat{\theta} = \frac{-n}{\sum_{i=1}^{n}\log(y_i)}$$

```
> (a=mean(y)/(1-mean(y)))
[1] 4.062526
> (b=-25/sum(log(y)))
[1] 4.166513
```

Les deux densités sont très proches



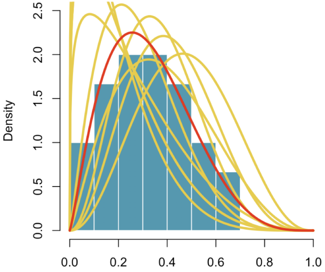
Maximum de Vraisemblance vs Méthode des Moments

Example 4 : on a fait un sondage sur 30 personnes pour savoir quelle proportion d'un livre elles ont lu. On suppose que cette proportion suit une loi Beta $\mathcal{B}(a,b)$. Proposez des estimateurs pour a et b (max vraisemblance et méthode des moments).

```
> y = c(0.175, 0.324, 0.146, 0.357, 0.148, 0.394,
   0.458, 0.228, 0.676, 0.371, 0.176, 0.027, 0.565,
   0.277, 0.282, 0.514, 0.480, 0.421, 0.507, 0.470,
   0.301, 0.038, 0.427, 0.275, 0.255, 0.075, 0.199,
   0.378, 0.632, 0.265)
```

Ici $\{Y_1,\cdots,Y_n\}$ sont i.i.d. de loi Beta, $\mathcal{B}(a,b)$,

$$f(y)=\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}x^{a-1}(1-x)^{b-1}\mathbf{1}_{[0,1]}(x)$$



Maximum de Vraisemblance vs Méthode des Moments

Conditions du premier ordre

$$\psi(\hat{a}) - \psi(\hat{a}+\hat{b}) = \frac{1}{n}\sum_{i=1}^{n}\log(y_i)$$

$$\psi(\hat{b}) - \psi(\hat{a}+\hat{b}) = \frac{1}{n}\sum_{i=1}^{n}\log(1-y_i)$$

où $\psi(z) = \dfrac{\Gamma'(z)}{\Gamma(z)}$ est la fonction digamma.

```
> logL = function(ab) -sum(log(dbeta(y,ab[1],ab[2])))
> optim(c(1,1),logL)
$par
[1] 2.026244 4.241886
```

```
> library(MASS)
> (fity = fitdistr(y, dbeta, list(shape1=1, shape2=1))
  )
    shape1       shape2
  2.0262445   4.2418876
 (0.4894734) (1.0918760)
```

Maximum de Vraisemblance vs Méthode des Moments

En notant que la moyenne empirique s'écrit

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \text{ et } \mathbb{E}[Y] = \frac{a}{a+b}$$

et que la variance empirique s'écrit

$$v = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2 \text{ et } \text{Var}[Y] = \frac{ab}{(a+b)^2(a+b+1)}$$

Les estimateurs de la méthode des moments vérifient

$$\begin{cases} [\widehat{a} + \widehat{b}]\ \overline{y} = \widehat{a} \\ (\widehat{a} + \widehat{b})^2[\widehat{a} + \widehat{b} + 1]v = \widehat{a}\widehat{b} \end{cases}$$

Maximum de Vraisemblance vs Méthode des Moments

La méthode des moments fournit les estimations suivantes:

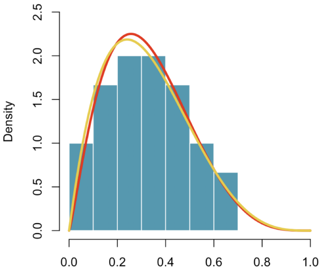
$$\widehat{a} = \bar{y}\left(\frac{\bar{y}(1-\bar{y})}{v} - 1\right),$$

$$\widehat{b} = (1-\bar{y})\left(\frac{\bar{y}(1-\bar{y})}{v} - 1\right).$$

```
> bary = mean(y)
> vary = var(y)
> (a = bary*(bary*(1-bary)/vary - 1))
[1] 2.230975
> (b = (1-bary)*(bary*(1-bary)/vary - 1))
[1] 4.569671
```

Maximum de Vraisemblance vs Méthode des Moments

Les deux estimateurs sont proches
et les densités aussi

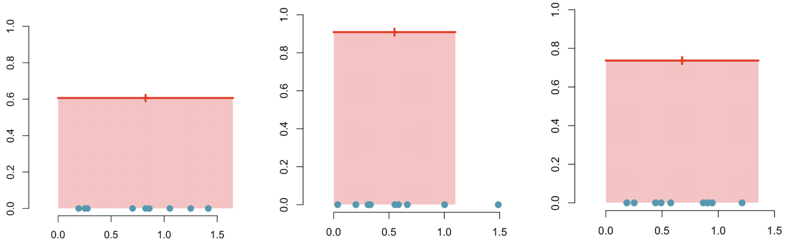


Méthode des Moments

Example: $\{y_1, \cdots, y_n\}$ de loi $\mathcal{U}([0,\theta])$, $\mathbb{E}[Y] = \theta/2$, alors

$$\overline{y} = \widehat{\theta}/2 \text{ i.e. } \widehat{\theta} = 2\overline{y}$$

Même si $y_i \leq \theta$ (par hypothèse), on peut avoir $\widehat{\theta} < y_j$



Note: estimateur du maximum de vraisemblance pour $\mathcal{U}([0,\theta])$?

Méthode des Moments

Théorème Central Limite

Soit (Y_n) une suite de variables aléatoires réelles indépendantes et de même loi admettant une espérance μ et une variance σ^2. La moyenne empirique $\overline{X}_n = \frac{1}{n}\sum_{k=1}^{n} Y_k$ centrée converge vers une loi normale :

$$\sqrt{n}[\overline{X}_n - \mu] \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2)$$

Méthode des Moments

Delta-Method

Comme $\sqrt{n}[\overline{X}_n - \mu] \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2)$, alors

$$\sqrt{n}\big[g(\overline{X}_n) - g(\mu)\big] \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2 \cdot [g'(\mu)]^2)$$

pour toute fonction g telle que $g'(\mu)$ existe et est non-nulle.

Méthode des Moments

Example: $\{Y_1, \cdots, Y_n\}$ de loi $\mathcal{U}([-\theta, \theta])$,
$\mathbb{E}[Y] = 0$ et $\text{Var}[Y] = \theta^2/3$,
On ne peut pas utiliser le premier modèle, on utilise le second,

$$s^2 = \widehat{\theta}^2/3 \text{ et } \widehat{\theta} = \sqrt{3s^2}$$

Comme $\mathbb{E}[Y] = 0$, $\mathbb{E}[Y^2] = \text{Var}[Y]$

Méthode des Moments: quel(s) moment(s) ?

Considérons un exemple (simple)

Example: $\{Y_1, \cdots, Y_n\}$ de loi lornormale $LN(\theta, 1)$,

$$\begin{cases} \mathbb{E}[Y] = e^{\mu+\sigma^2/2} \\ \text{Var}Y = (e^{\sigma^2} - 1)e^{2\mu+\sigma^2} \\ \text{skew}(Y) = (e^{\sigma^2} + 2)\sqrt{e^{\sigma^2} - 1}, \text{ etc} \end{cases}$$

etc. Comme $Y = \exp[X]$ où $X \sim \mathcal{N}(\theta, 1)$,

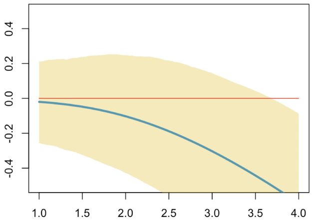
$$\mathbb{E}[Y^k] = \mathbb{E}[\exp(kX)] = M(K) = \exp\left(k\theta + \frac{k^2}{2}\right) = \psi(\theta)$$

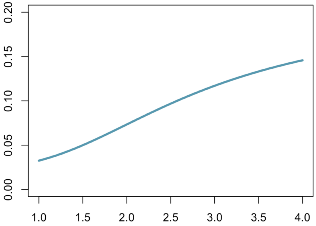
où M est la fonction génératrice des moments (de la loi normale)

$$\widehat{\theta}_k = \psi^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} y_i^k\right) = \frac{1}{k}\left[\log\left(\frac{1}{n}\sum_{i=1}^{n} y_i^k\right) - \frac{k^2}{2}\right]$$

Méthode des Moments: quel(s) moment(s) ?

Échantillons de taille $n = 50$ tirés suivant des lois $LN(0,1)$ on voit que l'on $\widehat{\theta}_k$ est un estimateur (possiblement) biaisé, et dont la variance augmente avec k





Si possible on utilise les premiers moments (moyenne, variance si deux moments sont nécessaire)