STT 1000 - STATISTIQVES

ARTHVR CHARPENTIER

Loi uniforme $\mathcal{U}([a,b])$

La densité et la fonction de répartition sont

$$f(x) = \frac{1}{b-a}\mathbf{1}_{[a,b]}(x) \text{ et } F(x) = \begin{cases} 0 & \text{si } x < a \\ \dfrac{x-a}{b-a} & \text{si } a \leq x \leq b \\ 1 & \text{si } x > b \end{cases}$$

$$\mathbb{E}[X] = \int_a^b xf(x)dx = \int_a^b x\left(\frac{1}{b-a}\right)dx = \left[\frac{1}{b-a}\frac{x^2}{2}\right]_a^b = \frac{a+b}{2}$$

et

$$\text{Var}(X) = \int_a^b (x-\mathbb{E}(X))^2 f(x)dx = \frac{(b-a)^2}{12}$$

En effet, $\text{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$, et, pour tout $r \in \mathbb{N}_\star$

$$\mathbb{E}\left(X^r\right) = \int_a^b x^r f(x)dx = \int_a^b x^r \frac{1}{b-a}dx$$

Loi uniforme $\mathcal{U}([a,b])$, et uniforme discrète

soit

$$\mathbb{E}\left(X^r\right) = \left[\frac{1}{b-a}\cdot\frac{x^{r+1}}{r+1}\right]_a^b = \frac{b^{r+1}-a^{r+1}}{(r+1)(b-a)}$$

On peut aussi considérer la loi uniforme sur $\{1,2,\cdots,r\}$,

$$\mathbb{E}(X) = \frac{1}{r}(1+2+\cdots+r) = \frac{1}{r}\frac{r(r+1)}{2} = \frac{r+1}{2}$$

$$\begin{aligned}\mathrm{Var}(X) &= E\left(X^2\right) - (E(X))^2 \\ &= \frac{1}{r}\left(1^2+2^2+\cdots+r^2\right) - \left(\frac{r+1}{2}\right)^2 \\ &= \frac{1}{6}(r+1)(2r+1) - \left(\frac{r+1}{2}\right)^2 = \frac{1}{12}\left(r^2-1\right)\end{aligned}$$

Loi uniforme $\mathcal{U}([0,\theta])$

$$f_\theta(x) = \frac{1}{\theta}\mathbf{1}_{[0,\theta]}(x)$$

```
> sample_x = c(3.375, 3.101, 2.895, 1.641, 1.348, 1.134, 1.551, 0.347, 1.611,
     1.418, 2.401, 2.305, 1.799, 1.161, 2.928, 0.253, 1.919)
> n = length(sample_x)
```

$$\mathcal{L}(\theta, x_1, \cdots, x_n) = \prod_{i=1}^{n} f_\theta(x) = \prod_{i=1}^{n} \frac{1}{\theta}\mathbf{1}_{[0,\theta]}(x_i) = \frac{1}{\theta^n}\prod_{i=1}^{n}\mathbf{1}_{[0,\theta]}(x_i)$$

$$\mathcal{L}(\theta, x_1, \cdots, x_n) = \begin{cases} 0 \text{ si } \theta < \max\{x_i\} - \text{i.e. } \exists i,\ x_i > \theta \\ \theta^{-n} \text{ si } \theta \geq \max\{x_i\} - \text{i.e. } \forall i,\ x_i \leq \theta \end{cases}$$

donc $\widehat{\theta}_{MLE} = \max\{x_i\}$.

Loi uniforme $\mathcal{U}([0,\theta])$

$$\mathbb{E}[X] = \int x f_\theta(x)dx = \int_0^\theta \frac{xdx}{\theta}dx = \left[\frac{x^2}{2\theta}\right]_0^\theta = \frac{\theta^2}{2\theta} - \frac{0}{2\theta} = \frac{\theta}{2}$$

Méthode des moments, $\overline{x} = \dfrac{\widehat{\theta}_{MM}}{2}$ ou $\widehat{\theta}_{MM} = 2\overline{x}$.

CALCUL

```
> max(sample_x)
[1] 3.375
> 2*mean(sample_x)
[1] 3.669059
```

Loi uniforme $\mathcal{U}([0,\theta])$

Soit $Y = \max\{X_i\}$ où $X_1, \cdots, X_n$ i.i.d. de loi f_θ,

$$\mathbb{P}[Y \leq y] = \mathbb{P}[\max\{X_i\} \leq y] = \mathbb{P}[\forall i,\ X_i \leq y] = \prod_{i=1}^{n} \mathbb{P}[X_i \leq y] = F_\theta(y)^n$$

dont la densité est (sur $[0,\theta]$)

$$g(y) = \frac{d\mathbb{P}[Y \leq y] =}{dy} = \frac{\partial F_\theta(y)^n}{\partial y} = n f_\theta(y) F_\theta(y)^{n-1} = \frac{n y^{n-1}}{\theta^n}$$

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

$$\mathbb{E}[\widehat{\theta}_{MLE}] = \theta - \underbrace{\frac{\theta}{n+1}}_{bias} < \theta$$

```
> max(sample_x)*(n+1)/n
[1] 3.573529
```

Loi uniforme $\mathcal{U}([0,\theta])$

$$\mathbb{E}[Y^2] = \int_0^\theta y^2 \frac{ny^{n-1}}{\theta^n} dy = \frac{n}{\theta^n}\int_0^\theta y^{n+1} dy = \frac{n}{\theta^n}\left[\frac{y^{n+2}}{n+2}\right]_0^\theta = \frac{n\theta^2}{n+2}$$

$$\text{Var}[Y] = \mathbb{E}[Y^2] - \mathbb{E}[Y]^2 = \frac{n\theta^2}{n+2} - \frac{n^2\theta^2}{(n+1)^2} = \frac{(n+1)^2 n - n^2(n+2)}{(n+2)(n+1)^2}\theta^2$$

$$\text{Var}[Y] = \frac{2}{(n+2)(n+1)}\theta^2$$

Le risque quadratique est ici

$$\mathbb{E}\big[(Y-\theta)^2\big] = \mathbb{E}\big[Y^2 - 2Y\theta + \theta^2\big] = \frac{n\theta^2}{n+2} - 2\frac{n\theta^2}{n+1} + \theta^2$$

$$\mathbb{E}\big[(Y-\theta)^2\big] = \frac{2\theta^2}{(n+1})(n+2)$$

Loi uniforme $\mathcal{U}([0,\theta])$

L'estimateur du maximum de vraisemblance est consistant,

$$\widehat{\theta}_{MLE} \overset{p.s.}{\to} \theta$$

En effet, si $\varepsilon > 0$,

$$\mathbb{P}\big[|\widehat{\theta}_{MLE}-\theta| > \varepsilon\big] = 1-P\big[\widehat{\theta}_{MLE} \in [\theta\pm\varepsilon]\big] = 1-\big(G(\theta+\varepsilon)-G(\theta-\varepsilon)\big)$$

$$\mathbb{P}\big[|\widehat{\theta}_{MLE} - \theta| > \varepsilon\big] = 1 - 1 + \left(\frac{\theta+\varepsilon}{\theta}\right)^n = \left(1+\frac{\varepsilon}{\theta}\right)^n$$

donc la série de terme générique $\mathbb{P}\big[|\widehat{\theta}_{MLE} - \theta| > \varepsilon\big]$ converge, et d'après Borel-Cantelli, on a la convergence p.s.

Loi uniforme $\mathcal{U}([0,\theta])$

L'estimateur du maximum de vraisemblance suit asymptotiquement une loi Exponentielle

$$n(\widehat{\theta}_{MLE}-\theta)\xrightarrow{\mathcal{L}}-\mathcal{E}(\theta^{-1})$$

En effet,

$$\mathbb{P}\big[n(\theta-Y)>t\big]=\mathbb{P}\left[Y<\theta-\frac{t}{n}\right]=\frac{1}{\theta^n}\left(\theta-\frac{t}{n}\right)^n=\left(1-\frac{t}{n\theta}\right)^n$$

$$\mathbb{P}\big[n(\theta-Y)>t\big]\to e^{-t/\theta}$$

qui est la fonction de survie d'une loi exponentielle

Loi uniforme $\mathcal{U}([0,\theta])$

$\mathbb{E}[\overline{X}] = \mathbb{E}[X_i] = \dfrac{\theta}{2}$ i.e. $\mathbb{E}[\widehat{\theta}_{MM}] = \theta$. Si on calcule la variance

$$\text{Var}[\overline{X}] = \frac{\text{Var}[X_i]}{n} = \frac{\theta}{12n}$$

donc $\text{Var}[\widehat{\theta}_{MM}] = \dfrac{\theta}{3n}$

Loi uniforme $\mathcal{U}([0,\theta])$

La statistique $\max\{x_i\}$ est sufficient et complète.
En effet, la densité jointe de $(X_1, X_2, \cdots, X_n)$ est

$$f(x_1, x_2, \cdots, x_n) = \frac{1}{\theta^n}\mathbf{1}_{x_i\in[0,\theta],\forall i} = \frac{1}{\theta^n}\mathbf{1}_{\max\{x_i\}\leq\theta}$$

donc $\max\{x_i\}$ est une statistique sufficient RAPELER DEF
Soit ψ une fonction telle que $\mathbb{E}[\psi \max\{X_i\}] = 0$, montrons que que ψ est nulle, quel que soit θ

$$\mathbb{E}[\psi(\max\{X_i\})] = \int_0^\theta \psi(y)\frac{n}{\theta^n}y^{n-1}dv = 0$$

$$\frac{n}{\theta^n}\int_0^\theta \psi(y)y^{n-1}dv = 0 \text{ i.e. } \int_0^\theta \psi(y)y^{n-1}dv = 0$$

$$\text{donc } \frac{\partial}{\partial\theta}\int_0^\theta \psi(y)y^{n-1}dv = \psi(\theta)\theta^{n-1} = 0,\ \forall\theta$$

donc $\psi = 0$ (ou $\mathbb{P}[\psi(\max\{X_i\})] = 0$) donc $\max\{x_i\}$ est une statistique complète

$\widehat{\theta} = \frac{n+1}{n}\max\{x_i\}$ est BUE, d'après le théorème de

Loi uniforme $\mathcal{U}([0,\theta])$

L'estimateur de la méthode des moments est asymptotiquement Gaussien

$$\sqrt{n}(\widehat{\theta}_{MM}-\theta) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0,\frac{\theta^2}{3}\right)$$

Loi uniforme $\mathcal{U}([-\theta,\theta])$

Posons $\widehat{\theta}=\max\{|x_i|\}$, le biais de $\widehat{\theta}$ est $-\theta/(n+1)$
Posons $y_i=|x_i|$, alors $Y_i\sim\mathcal{U}([0,\theta])$
12.8
13.4

Loi uniforme $\mathcal{U}([\theta, \theta+1])$

$$f_\theta(x) = \mathbf{1}_{[\theta,\theta+1]}(x)$$

```
> sample_y = c(2.317, 2.640, 2.374, 1.882, 2.520, 2.596, 2.492, 2.394, 1.994,
    2.299, 2.189, 2.513, 2.368, 2.474, 2.067, 2.622, 2.108, 1.882, 1.970,
    1.976, 2.215, 2.370, 2.366, 2.179, 2.042, 2.425, 2.687, 2.746)
```

$$\mathcal{L}(\theta, x_1, \cdots, x_n) = \prod_{i=1}^{n} f_\theta(x) = \mathbf{1}_{[\theta,\theta+1]}(x_i) \in [0,1]$$

$$\mathcal{L}(\theta) = 1 \Longleftrightarrow \forall i, x_i \in [\theta, \theta+1] \Longleftrightarrow \theta \leq \min\{x_i\} \leq \max\{x_i\} \leq \theta+1$$

(sinon $\mathcal{L}(\theta) = 0$)

$$\mathcal{L}(\theta, x_1, \cdots, x_n) = \begin{cases} 1 \text{ si } \theta < \min\{x_i\} \\ 0 \text{ si } \theta \in [\min\{x_i\}, \max\{x_i\}] \\ 1 \text{ si } \theta > \max\{x_i\} \end{cases}$$

Loi uniforme $\mathcal{U}([\theta, \theta+1])$

Posons $U = \min\{X_i\}$ et $V = \max\{X_i\}$, de telle sorte que

$$\mathcal{L}(\theta, x_1, \cdots, x_n) = \mathbf{1}_{u \geq \theta} \mathbf{1}_{v \leq \theta+1}$$

$$\mathcal{L}(\theta) = 1 \Longleftrightarrow \theta \leq u \leq v \leq \theta + 1 \Longleftrightarrow v - 1 \leq \theta \leq u$$

i.e. $\widehat{\theta} \in [v-1, u]$ est un maximum de vraisemlance, autrement dit,

$$\widehat{\theta} = \alpha(v-1) + (1-\alpha)u \quad \text{où } \alpha \in [0,1]$$

$$\mathbb{E}[U] = \theta + \frac{1}{n+1} \text{ et } \mathbb{E}[V] = \theta + 1 - \frac{1}{n+1}$$

$$\mathbb{E}[\widehat{\theta}] = \alpha\left(\theta - \frac{1}{n+1}\right) + (1-\alpha)\left(\theta + \frac{1}{n+1}\right) = \theta + \frac{1-2\alpha}{n+1}$$

L'estimateur est sans biais ssi $\alpha = 1/2$.

Loi uniforme $\mathcal{U}([\theta,\theta+1])$

$\text{mse}\left[\widehat{\theta}\right] = \text{Var}\left[\widehat{\theta}\right] + \mathbb{E}\left[\widehat{\theta}\right]^2$, où la variance vaut

$\text{Var}[\alpha(V-1)+(1-\alpha)U] = \left[\alpha^2+(1-\alpha)^2\right]\text{Var}[V]+2\alpha(1-\alpha)\text{cov}[U,V]$

car $\text{Var}[U]=\text{Var}[V]$ et on sait que $\text{Var}[V] = \dfrac{n}{(n+1)^2(n+2)}$

Le calcul de $\text{cov}[U,V]$ est un peu plus compliqué.

On peut montrer que la densité de (U,V) est

$$g(u,v) = n(n-1)(v-u)^{n-2}\mathbf{1}_{\theta\leq u\leq v\leq\theta+1}$$

$$\mathbb{E}\left[UV\right] = n(n-1)\int_\theta^{\theta+1} v\int_\theta^v u(v-u)^{n-2}dudv$$

$$= \int_\theta^{\theta+1} v\left(-n\left[u(v-u)^{n-1}\right]_0^v + n\int_\theta^v (v-u)^{n-1}du\right)dv$$

..

$$(\theta+1)\left(\theta+\frac{1}{n+1)}\right)\left(\frac{\theta}{n+1}+\frac{1}{n+1)(n+2)}\right)$$

Loi uniforme $\mathcal{U}([\theta, \theta+1])$

donc

$$\text{cov}[U, V] = \frac{1}{(n+1)^2(n+2)}$$

aussi le *mean squared error* vaut

$$\text{mse}[\widehat{\theta}] = \frac{6\alpha^2 - 6\alpha + 2}{(n+1)(n+2)}$$

qui est minimal quand $\alpha = 1/2$

Loi uniforme $\mathcal{U}([\theta_1,\theta_2])$

La densité jointe de $X_1, X_2, \cdots, X_n$ est

$$f(x_1, x_2, \cdots, x_n) = \frac{1}{(\theta_2-\theta_1)^n}\mathbf{1}_{\min\{x_i\}>\theta_1}\mathbf{1}_{\max\{x_i\}<\theta_2}$$

La densité jointe du couple $(\min\{X_i\}, \max\{X_i\})$ est

$$h(u,v) = \frac{n(n-1)}{(\theta_2-\theta_1)^2}\left(\frac{v-u}{(\theta_2-\theta_1)^2}\right)^{n-2}$$

pour $\theta_1 < u \leq v < \theta_2$ (et 0 sinon).

$$\mathbb{E}[V] = \int_{\theta_1}^{\theta_2} t\frac{n}{\theta_2-\theta_1}\left(\frac{t-\theta_1}{(\theta_2-\theta_1)^2}\right)^{n-1} dt$$

$$\mathbb{E}[V] = \left[t\left(\frac{t-\theta_1}{\theta_2-\theta_1}\right)^n\right]_{\theta_1}^{\theta_2} - \int_{\theta_1}^{\theta_2}\left(\frac{t-\theta_1}{(\theta_2-\theta_1)^2}\right)^n dt = \theta_2 - \frac{\theta_2-\theta_1}{n+1}$$

Loi uniforme $\mathcal{U}([\theta_1, \theta_2])$

$$\mathbb{E}[U] = \int_{\theta_1}^{\theta_2} t \frac{n}{\theta_2 - \theta_1} \left(\frac{\theta_2 - t}{(\theta_2 - \theta_1)^2} \right)^{n-1} dt$$

$$\mathbb{E}[U] = \left[t \left(\frac{\theta_2 - t}{\theta_2 - \theta_1} \right)^n \right]_{\theta_1}^{\theta_2} - \int_{\theta_1}^{\theta_2} \left(\frac{\theta_2 - t}{(\theta_2 - \theta_1)^2} \right)^n dt = \theta_1 + \frac{\theta_2 - \theta_1}{n+1}$$

De plus

$$\mathbb{E}[V - U] = \frac{n-1}{n+1}(\theta_2 - \theta_1)$$

Loi uniforme $\mathcal{U}([\theta, 2\theta])$

Dans ce cas,

$$\mathbb{E}[V] = \theta_2 - \frac{\theta_2 - \theta_1}{n+1} = 2\theta - \frac{2\theta - \theta}{n+1} = \frac{2n+1}{n+1}\theta$$

$$\mathbb{E}[U] = \theta_1 + \frac{\theta_2 - \theta_1}{n+1} = \theta + \frac{2\theta - \theta}{n+1} = \frac{n+2}{n+1}\theta$$

Notons que dans ce cas

$$\mathbb{E}\left[\frac{V}{2n+1} - \frac{U}{n+2}\right] = 0$$

autrement dit il existe une fonction h non nulle telle que $\mathbb{E}[h(U,V)] = 0$ donc (U,V) n'est pas une statistique complète.