# project

Yue Zhang et Gérémi Bridonneau

## Introduction

## 1 La transformation de Box-Cox

### 1. (réfléchir un nom/titre)

Si $\lambda = 0$, $h_\lambda(y) = log\ y$, $\forall y > 0$.

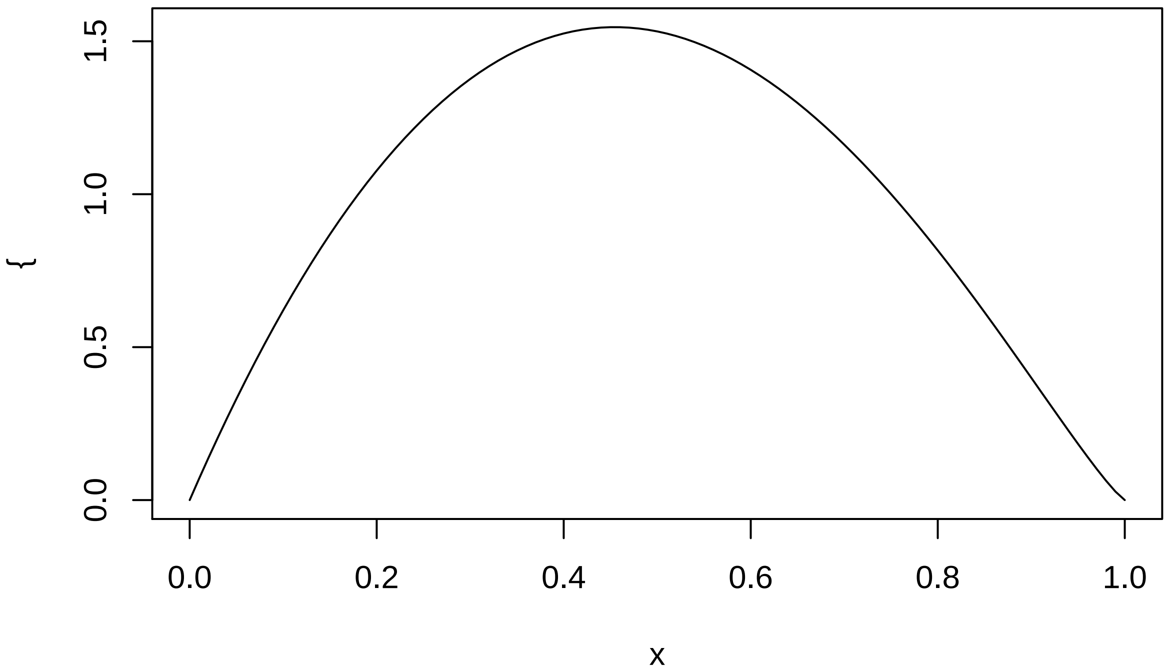
On a $(1)\ h_\lambda(y) = x^{'}\theta + \varepsilon$, $\varepsilon \sim \mathcal{N}(0, \sigma I_n)$.

– (mon idée de solution)

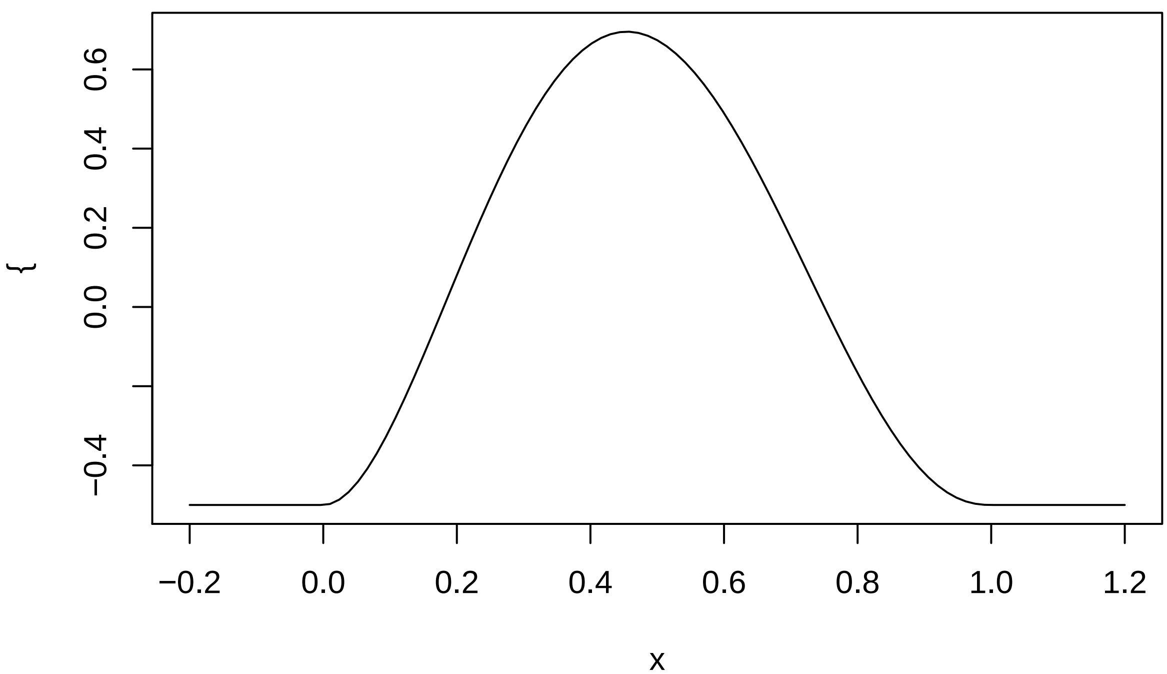
La transformation $\tilde{h}_\lambda$ est valable seulement pour $y > 0$. De plus pour tout $\lambda \neq 0$, la transformation $\tilde{h}_\lambda$ est borné et donc la transformation ne peut pas être gaussienne. Pour $\lambda = 0$ on n'a pas ce problème grâce à la surjectivité du logarithme.

Si toute les observations sont positives on peut quand même utiliser cette transformation car on perdra qu'une faible partie des données normalement dans la queue à gauche de la répartition. Par exemple si les données ne suivent pas une loi normale mais une loi beta de paramètre $\alpha = 2, \beta = 2.2$ et qu'on utilise la transformation de Box et Cox avec $\lambda = 2$ on obtient:

```
lambda <- 2
a <- 2
b <- 2.2
plot({function (x) dbeta(x, a, b)})
```




```r
plot({function (x) (dbeta(x, a, b)^lambda - 1)/lambda}, -0.2, 1.2)
```




On voit qu'on a aucune valeur négative et que donc la gaussianisation n'est pas parfaite mais cette transformation reste raisonnable.

## 2. Déterminer la fonction de vraisemblance

Supposons que pour $\beta = (\theta, \lambda, \sigma^2)'$ a $p \times 1$ vecteur de paramètres, on ait $h_\lambda(Y_i) = Z_i = x_i\theta + \varepsilon_i,\ \varepsilon_i \sim \mathcal{N}(0, \sigma^2)$, $\varepsilon_i$ suivent une loi gaussien i.i.d. Donc par la définition de vraisemblance:

$$
\begin{aligned}
L(\lambda,\theta,\sigma^2;Y) &= \prod_{i=1}^{n} \frac{\partial F}{\partial Y_i} \\
&= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{(h_\lambda(Y_i) - x_i\theta)^2}{2\sigma^2}) \left|\frac{\partial h_\lambda(Y_i)}{\partial Y_i}\right| \\
&= (\frac{1}{2\pi\sigma^2})^{\frac{n}{2}} \exp(-\frac{\sum_{i=1}^{n}(h_\lambda(Y_i) - x_i\theta)^2}{2\sigma^2}) \prod_{i=1}^{n} \left|\frac{\partial h_\lambda(Y_i)}{\partial Y_i}\right| \\
&= (\frac{1}{2\pi\sigma^2})^{\frac{n}{2}} \exp(-\frac{(h_\lambda(Y_i) - x_i\theta)'(h_\lambda(Y_i) - x_i\theta)}{2\sigma^2}) \prod_{i=1}^{n} \left|Y_i^{\lambda-1}\right|
\end{aligned} \tag{1}
$$

Donc le terme $J(\lambda;Y) = \prod_{i=1}^{n} \left|\frac{\partial h_\lambda(Y_i)}{\partial Y_i}\right| = \prod_{i=1}^{n} \left|Y_i^{\lambda-1}\right|$, est la transformation de Jacobian de $h_\lambda(Y) - X\theta$ à $Y$.

## 3. Estimation du maximum de vraisemblance

A $\lambda$ fixé, on peut déterminer l'emv par estimation du maximum de vraisemblance. Donc tout d'abord, depuis l'équation 1 on calcule le log-vraisemblance.

$$
\begin{aligned}
\ell &= \log L(\lambda,\theta,\sigma^2;Y) \\
&= -\frac{n}{2}\log(2\pi\sigma^2) - \frac{(h_\lambda(Y) - X\theta)'(h_\lambda(Y) - X\theta)}{2\sigma^2} + \sum_{i=1}^{n} \log\left|Y_i^{\lambda-1}\right| \\
&= -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{\|h_\lambda(Y) - X\theta\|^2}{2\sigma^2} + (\lambda-1)\sum_{i=1}^{n}\log|Y_i|
\end{aligned} \tag{2}
$$

En suite, étant donnée que le log-vraisemblance $\ell$ l'équation 2 est une transforamtion monotone de vralsemblance $L$ dans l'équation 1, on maximise log-vraisemblance $\ell$ en respectant $\theta$, $\sigma^2$ et $\lambda$, donc on obtient le premier ordre dérivation ci-dessous:

$$
\frac{\partial \ell}{\partial \sigma^2} = -\frac{n}{2}\frac{1}{\sigma^2} + \frac{(h_\lambda(Y) - X\theta)'(h_\lambda(Y) - X\theta)}{2\sigma^4} = 0 \tag{3}
$$

Donc on a $\widehat{\sigma}^2 = \frac{(h_\lambda(Y) - X\theta)'(h_\lambda(Y) - X\theta)}{n} = \frac{h_\lambda(Y)'(I_n - H)h_\lambda(Y)}{n}$, avec $H = X(X'X)^{-1}X'$ et $I_n$ matrice identidé.

$$
\begin{aligned}
\frac{\partial \ell}{\partial \theta} &= -\frac{2(-X)'(h_\lambda(Y) - X\theta)}{2\sigma^2} \\
&= \frac{X'(h_\lambda(Y) - X\theta)}{\sigma^2} = 0
\end{aligned} \tag{4}
$$

Donc, $\widehat{\theta} = (X'X)^{-1}X'h_\lambda(Y)$, par $X'h_\lambda(Y) = X'X\theta$.

Pour vérifier $L_{max}(\lambda)$, on remplace notre l'emv $\widehat{\sigma}^2$ et $\widehat{\theta}$ calculé par l'équation 3 et 4 dans log-vraisemblance $\ell$:

$$
\begin{aligned}
L_{max}(\lambda) &:= \ell = \log L(\lambda, \widehat{\theta}(\lambda), \widehat{\sigma}^2(\lambda)) \\
&= -\frac{n}{2}\log(\frac{\|h_\lambda(Y) - X\theta\|^2}{n}) - \frac{\|h_\lambda(Y) - X\theta\|^2 n}{2\|h_\lambda(Y) - X\theta\|^2} + (\lambda - 1)\sum_{i=1}^{n}\log|Y_i| - \frac{n}{2}\log(2\pi) \\
&= -\frac{n}{2}\log(\widehat{\sigma}^2(\lambda)) + (\lambda - 1)\sum_{i=1}^{n}\log|Y_i| - \frac{n}{2} - \frac{n}{2}\log(2\pi)
\end{aligned} \tag{5}
$$

Donc $a(n) = -\frac{n}{2} - \frac{n}{2}\log(2\pi)$ est bien une constante ne dépendant que $n$. Maintenant on calcule l'emv $\widehat{\lambda}$:

$$
\frac{\partial \ell}{\partial \lambda} = -\frac{2(h_\lambda(Y) - X\theta)\left|\frac{\partial h_\lambda(Y)}{\partial \lambda}\right|}{2\sigma^2} + \sum_{i=1}^{n}\log|Y_i| = 0 \tag{6}
$$

Et

$$
\begin{aligned}
\frac{\partial L_{max}}{\partial \lambda} &= -\frac{n}{2}\frac{1}{\widehat{\sigma}^2(\lambda)}\left|\frac{\partial \widehat{\sigma}^2(\lambda)}{\partial \lambda}\right| + \sum_{i=1}^{n}\log|Y_i| \\
&= -\frac{n}{2}\frac{1}{\widehat{\sigma}^2(\lambda)}\frac{2(h_\lambda(Y) - X\theta)\left|\frac{\partial h_\lambda(Y)}{\partial \lambda}\right|}{n} + \sum_{i=1}^{n}\log|Y_i| \\
&= -\frac{(h_\lambda(Y) - X\theta)\left|\frac{\partial h_\lambda(Y)}{\partial \lambda}\right|}{\widehat{\sigma}^2} + \sum_{i=1}^{n}\log|Y_i| = 0
\end{aligned} \tag{7}
$$

On peut bien vérifier que $\frac{\partial \ell}{\partial \lambda}$ et $\frac{\partial L_{max}}{\partial \lambda}$ sont équaux par calculation l'équation maximum vraisemblance. Par l'équation 3, on sait que $\widehat{\sigma}^2(\lambda) = \frac{h_\lambda(Y)'(I_n - H)h_\lambda(Y)}{n} = \frac{SCR(\lambda)}{n}$ avec $H = X(X'X)^{-1}X'$, est la somme de carrés résiduels de variance $h_\lambda(Y^)$ divisée par $n$. Depuis l'équation 7, on peut continuer cette calculation en remplaçant $\widehat{\sigma}^2$, et pour rappel $h_\lambda(Y) = \frac{Y^\lambda - 1}{\lambda}$:

$$
\begin{aligned}
\frac{\partial L_{max}}{\partial \lambda} &= -\frac{n}{2}\frac{n}{h_\lambda(Y)'(I_n - H)h_\lambda(Y)}\frac{2h_\lambda(Y)'(I_n - H)}{n}(\frac{Y^\lambda \log Y}{\lambda} - \frac{Y^\lambda - 1}{\lambda^2}) + \sum_{i=1}^{n}\log|Y_i| \\
&= -n\frac{h_\lambda(Y)'(I_n - H)}{h_\lambda(Y)'(I_n - H)h_\lambda(Y)}(\frac{Y^\lambda \log Y}{\lambda} - \frac{h_\lambda(Y)}{\lambda}) + \sum_{i=1}^{n}\log|Y_i| \\
&= -n\frac{h_\lambda(Y)'(I_n - H)\lambda^{-1}Y^\lambda \log Y}{h_\lambda(Y)'(I_n - H)h_\lambda(Y)} + n\frac{h_\lambda(Y)'(I_n - H)h_\lambda(Y)}{h_\lambda(Y)'(I_n - H)h_\lambda(Y)\lambda} + \sum_{i=1}^{n}\log|Y_i| \\
&= -n\frac{h_\lambda(Y)'(I_n - H)u_\lambda(Y)}{h_\lambda(Y)'(I_n - H)h_\lambda(Y)} + \frac{n}{\lambda} + \sum_{i=1}^{n}\log|Y_i|
\end{aligned} \tag{8}
$$

avec $u_\lambda(Y) = \lambda^{-1}Y^\lambda \log Y$. Le numérateur dans l'équation 8 est la somme résiduelle des produits dans l'analyse de la covariance de $h_\lambda(Y)$ et $u_\lambda(Y)$. Maintenant on utilise la transformation normalisée afin de simplifier le résultat, on définit $z_\lambda(Y)$ ci-dessous:

$$z_\lambda(Y) = \frac{h_\lambda(Y)}{J(\lambda;Y)^{1/n}}$$
$$= \frac{h_\lambda(Y)}{(\prod_{i=1}^n |Y_i|)^{\lambda-1/n}} \tag{9}$$

Donc $\widehat{\sigma}^2$ devient $\widehat{\sigma}^2(\lambda;z) = \frac{z_\lambda(Y)'(I_n - H)z_\lambda(Y)}{n} = \frac{SCR(\lambda;z)}{n}$, $SCR(\lambda;z)$ est la somme des carrées résiduelle de $z_\lambda(Y)$. De plus, $L_{max} = -\frac{n}{2}\log(\widehat{\sigma}^2(\lambda;z)) + a(n)$, donc on propose de trouver $\widehat{\lambda}$ qui maximize $L_{max}(\lambda)$, c'est à dire minimize $\widehat{\sigma}^2(\lambda;z)$. Donc on cherche l'emc (estimateur des moindres carrées)

$$\widehat{\lambda} = \arg\min_\lambda SCR(\lambda;z) \tag{10}$$

Pour répondre que l'emv est-il gaussien à distance finie? Je sais pas comment expliquer mais par le théorème du cours, l'emv est asymptotiquement normale non? emmm je vais réfléchir. genre la distribution de $\sqrt{n}(\widehat{\beta} - \beta)$, quand $n \to \infty$, elle converge en une loi normale.

$$\left[\frac{I_n(\beta)^{-1}}{n}\right]^{1/2} \sqrt{n}(\widehat{\beta} - \beta) \to \mathcal{N}(0, I_n) \tag{11}$$

$I_n(\beta)^{-1}$ est la matrice de l'information de Fisher.

## 4. Distribution asymptotique de l'emv

**Estimer la variance de $\widehat{\lambda}$**

Soit $\widehat{\beta}$ l'emv asymptotiquement normal,

$$\sqrt{n}(\widehat{\beta} - \beta) \to \mathcal{N}(0, I_n(\beta)^{-1})$$
$$\widehat{\beta} \sim \mathcal{N}(\beta, I_n(\beta)^{-1}) \tag{12}$$

Par la définition, la matrice de l'information de Fisher est écrite ci-dessous:

$$I_n(\beta) = \mathbb{E}_\beta[\dot{\ell}\dot{\ell}']$$
$$= \mathbb{E}_\beta[-\ddot{\ell}] \tag{13}$$

où $\ddot{\ell}$ est la matrice Hessien $\ddot{\ell} = \frac{\partial^2 \ell}{\partial\beta\partial\beta'}$ (pour rappelle que on a définit $\beta = (\theta, \lambda, \sigma^2)'$ a $p \times 1$ vecteur de paramètres). En particulier, on n'a pas forcément besoin d'estimer $\sigma^2$ simultanément avec $\theta$ et $\lambda$, donc pour simplifier les calculs, on décide de calculer la matrice Hessien de $L_{max}(\lambda)$. Dans l'équation 7, on a calculé $\frac{\partial L_{max}}{\partial\lambda} = -\frac{(h_\lambda(Y) - X\theta)\left|\frac{\partial h_\lambda(Y)}{\partial\lambda}\right|}{\widehat{\sigma}^2} + \sum_{i=1}^n \log|Y_i|$, et on obtient sans souci $\frac{\partial L_{max}}{\partial\theta} = \frac{X'(h_\lambda(Y) - X\theta)}{\widehat{\sigma}^2}$.

$$
\begin{aligned}
H_n(\beta) &= \frac{\partial^2 L_{max}}{\partial\beta\partial\beta'} \\
&= -\widehat{\sigma}^{-2}\begin{bmatrix} X'X & -X'\left|\frac{\partial h_\lambda(Y)}{\partial\lambda}\right| \\ -\left|\frac{\partial h_\lambda(Y)}{\partial\lambda}\right| X & \left|\frac{\partial h_\lambda(Y)}{\partial\lambda}\right|'\left|\frac{\partial h_\lambda(Y)}{\partial\lambda}\right| + \left|\frac{\partial^2 h_\lambda(Y)}{\partial^2\lambda}\right|(h_\lambda(Y) - X\theta) \end{bmatrix}
\end{aligned} \tag{14}
$$

$H_n(\beta)$ est bien une matrice définie négative. Etant donnée la distribution asymptotique normale de l'emv, on peut concluire que $\widehat{Var(\widehat{\beta})} = -[H_n(\widehat{\beta})]^{-1}$, maintenant on calcul $\widehat{Var(\widehat{\lambda})}$:

$$
\widehat{Var(\widehat{\lambda})} = -H_n(\widehat{\lambda})^{-1} \tag{15}
$$

où $H_n(\lambda) = \dfrac{\partial^2 L_{max}}{\partial^2\lambda} = \dfrac{\left|\frac{\partial h_\lambda(Y)}{\partial\lambda}\right|'\left|\frac{\partial h_\lambda(Y)}{\partial\lambda}\right| + \left|\frac{\partial^2 h_\lambda(Y)}{\partial^2\lambda}\right| h_\lambda(Y)'(I_n - H)}{-\widehat{\sigma}^2}$.

**Intervalle de confiance**

Etant donné que l'emv est asymptotiquement normalement distribué, donc on peut construire le test $T = \dfrac{\widehat{\beta}-\beta}{\sqrt{\frac{\widehat{Var(\widehat{\beta})}}{n}}} \sim \mathcal{N}(0, I_n)$. Par définition, $P(q_{\alpha/2} < \dfrac{\widehat{\beta}-\beta}{\sqrt{\frac{\widehat{Var(\widehat{\beta})}}{n}}} < q_{1-\alpha/2}) = 1-\alpha$, donc on peut obtenir l'intervalle de confiance $[\widehat{\beta} - q_{1-\alpha/2}\sqrt{\frac{\widehat{Var(\widehat{\beta})}}{n}}, \widehat{\beta} - q_{\alpha/2}\sqrt{\frac{\widehat{Var(\widehat{\beta})}}{n}}]$, où $q_{\alpha/2}$ et $q_{1-\alpha/2}$ sont quantiles d'ordre $\alpha/2$ et $1-\alpha/2$ sous la loi normale. La distribution est symétrique par rapport à 0, donc l'IC estimateur $\beta$ est également $[\widehat{\beta} - q_{1-\alpha/2}\sqrt{\frac{\widehat{Var(\widehat{\beta})}}{n}}, \widehat{\beta} + q_{1-\alpha/2}\sqrt{\frac{\widehat{Var(\widehat{\beta})}}{n}}]$.

L'intervalle de confiance de $\lambda$ donc est $[\widehat{\lambda} - q_{1-\alpha/2}\sqrt{\frac{\widehat{Var(\widehat{\lambda})}}{n}}, \widehat{\lambda} + q_{1-\alpha/2}\sqrt{\frac{\widehat{Var(\widehat{\lambda})}}{n}}]$, où $\widehat{Var(\widehat{\lambda})}$ est calculé dans l'équation 15.

**Test de Wald**

On définit $A = (0, \frac{1}{\lambda_0}, 0)$, $\beta = (\theta, \lambda, \sigma^2)'$

$H_0 : A\beta = 1$ contre $H_1 : A\beta \neq 1$

En utilisant la statistique de Wald, sous $H_0$ on a :

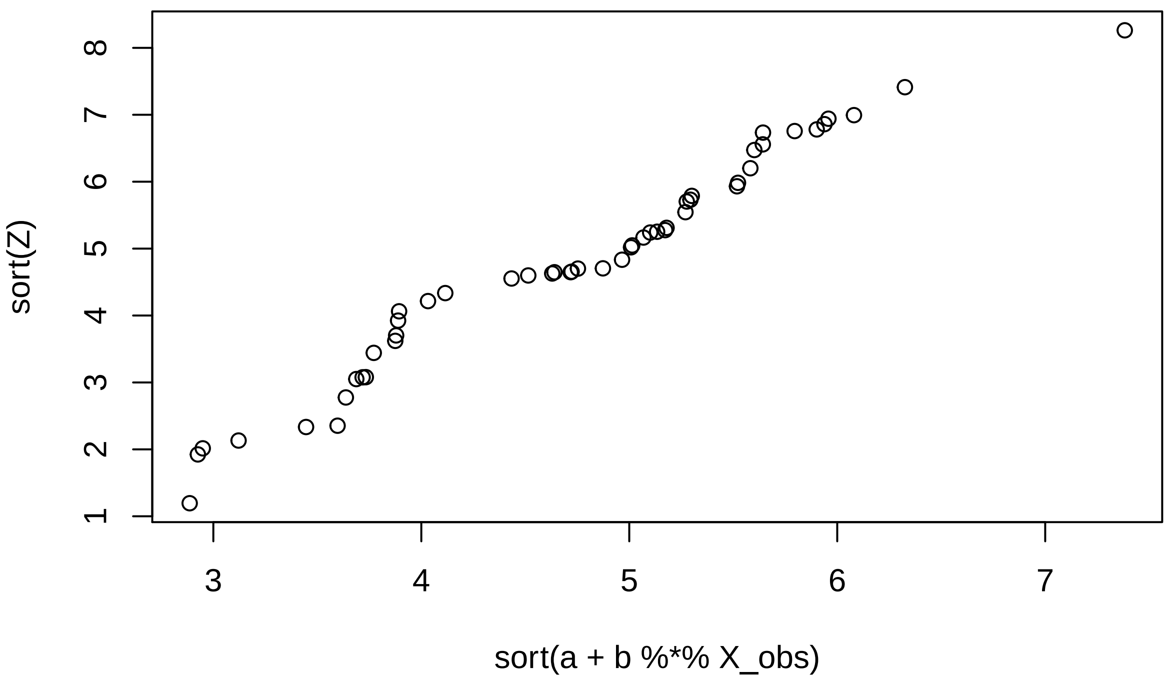
$$
W = n(A\widehat{\beta} - 1)(A\widehat{V}_n A')^{-1}(A\widehat{\beta} - 1)' \to \chi^2(1) \tag{16}
$$

où $\widehat{V}_n = I_n(\widehat{\beta})^{-1}$, et $W \geq 0$, la région de rejet est unilatère à droite de niveau asymptotique $\alpha$ donc est $\mathcal{R} = \left\{W > q_{1-\alpha}^{\chi^2(1)}\right\}$ avec $P_{(H_0)}(\mathcal{R}) \to \alpha$.

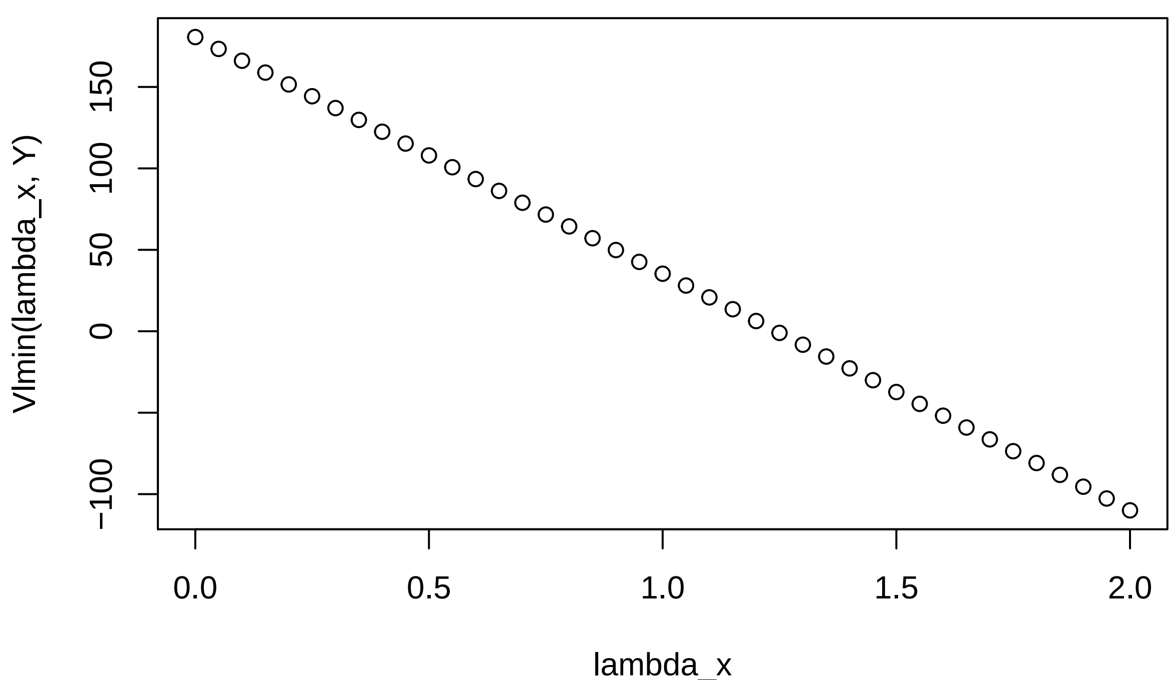
5.

## 2 Test de la méthode sur des données simulées

1.




2.



# 3 Cas pratique

# Conclusion