

# project

Yue Zhang et G  r  mi Bridonneau

## Introduction

### 1 La transformation de Box-Cox

#### 1. (r  fl  chir un nom/titre)

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

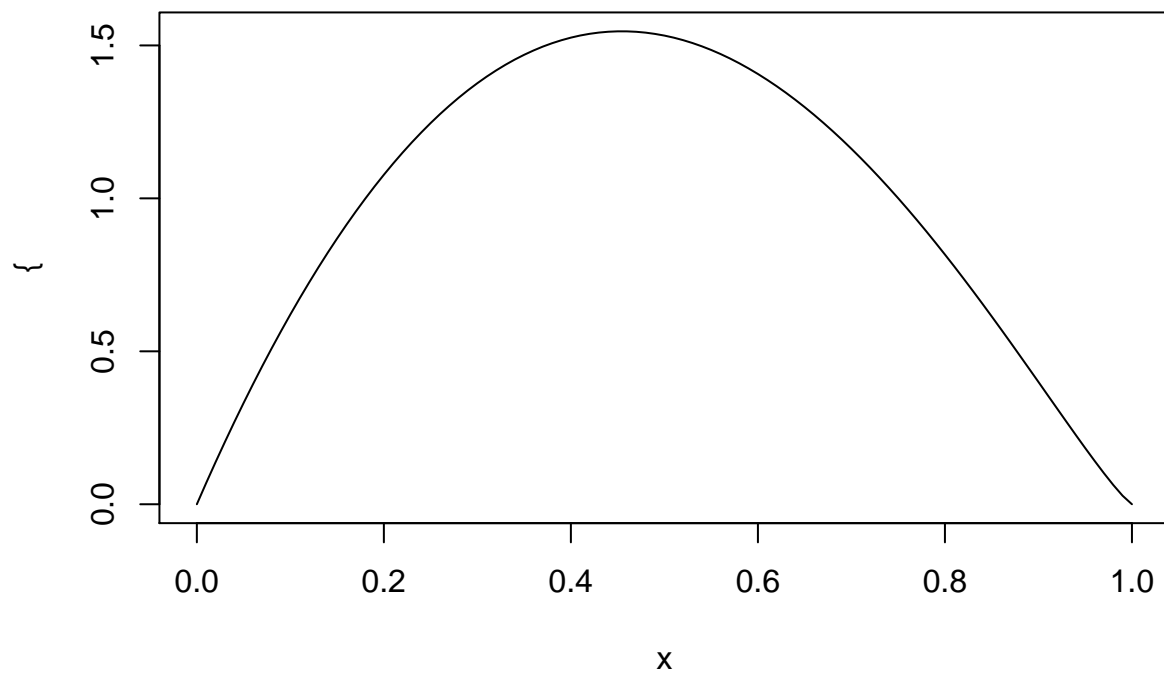
On a (1)  $h_\lambda(y) = x' \theta + \varepsilon$ ,  $\varepsilon \sim \mathcal{N}(0, \sigma I_n)$ .

– (mon id  e de solution)

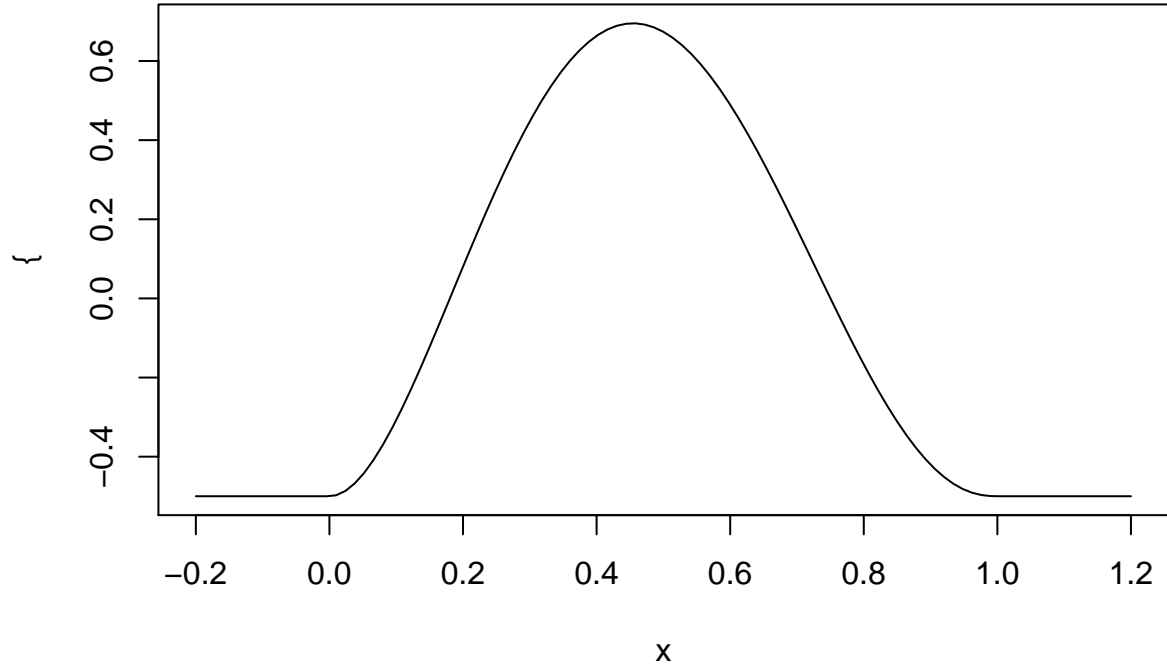
La transformation  $\tilde{h}_\lambda$  est valable seulement pour  $y > 0$ . De plus pour tout  $\lambda \neq 0$ , la transformation  $\tilde{h}_\lambda$  est born  e et donc la transformation ne peut pas   tre gaussienne. Pour  $\lambda = 0$  on n'a pas ce probl  me gr  ce    la surjectivit   du logarithme.

Si toutes les observations sont positives on peut quand m  me utiliser cette transformation car on perdra qu'une faible partie des donn  es normalement dans la queue    gauche de la r  partition. Par exemple si les donn  es ne suivent pas une loi normale mais une loi beta de param  tre  $\alpha = 2, \beta = 2.2$  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)})
```



```
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:

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

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

### 3. Estimation du maximum de vraisemblance

A  $\lambda$  fixé, on souhaite déterminer l'estimateur du maximum de vraisemblance  $\hat{\theta}(\lambda)$  et  $\hat{\sigma}^2(\lambda)$ . Donc tout d'abord, depuis l'équation 1 on calcule la log-vraisemblance.

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

Ensuite, étant donné que la log-vraisemblance  $\ell$  l'équation 2 est une transformation monotone de la vraisemblance  $L$  dans l'équation 1, on maximise la log-vraisemblance  $\ell$  respectivement pour  $\theta$ ,  $\sigma^2$  et  $\lambda$ , donc on obtient le premier ordre dérivation ci-dessous:

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

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

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

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

Pour vérifier la formule avec  $L_{max}(\lambda)$ , on remplace nos emv  $\hat{\sigma}^2$  et  $\hat{\theta}$  calculés dans les équations 3 et 4 dans la log-vraisemblance  $\ell$ :

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

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

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

Et

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

On peut bien vérifier que  $\frac{\partial \ell}{\partial \lambda}$  et  $\frac{\partial L_{max}}{\partial \lambda}$  sont égaux par calcul de l'équation maximum vraisemblance. Par l'équation 3, on sait que  $\hat{\sigma}^2(\lambda) = \frac{h_\lambda(Y)'(I_n - H)h_\lambda(Y)}{n} = \frac{SCR(\lambda)}{n}$  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 calcul en remplaçant  $\hat{\sigma}^2$ , et pour rappel  $h_\lambda(Y) = \frac{Y^\lambda - 1}{\lambda}$ :

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

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:

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

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

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

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  $\sqrt{n}(\hat{\beta} - \beta)$ , quand  $n \rightarrow \infty$ , elle converge en une loi normale.

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

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

#### 4. Distribution asymptotique de l'emv

Estimer la variance de  $\hat{\lambda}$

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

$$\begin{aligned} \sqrt{n}(\hat{\beta} - \beta) &\rightarrow \mathcal{N}(0, I_n(\beta)^{-1}) \\ \hat{\beta} &\sim \mathcal{N}(\beta, I_n(\beta)^{-1}) \end{aligned} \quad (12)$$

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

$$\begin{aligned} I_n(\beta) &= \mathbb{E}_\beta[\dot{\ell}\dot{\ell}'] \\ &= \mathbb{E}_\beta[-\ddot{\ell}] \end{aligned} \quad (13)$$

où  $\ddot{\ell}$  est la matrice Hessian  $\ddot{\ell} = \frac{\partial^2 \ell}{\partial \beta \partial \beta'}$  (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 Hessian de  $L_{max}(\lambda)$ . Dans l'équation 7, on a calculé

$$\frac{\partial L_{max}}{\partial \lambda} = - \frac{(h_\lambda(Y) - X\theta) \left| \frac{\partial h_\lambda(Y)}{\partial \lambda} \right|}{\hat{\sigma}^2} + \sum_{i=1}^n \log |Y_i|, \text{ et on obtient sans souci } \frac{\partial L_{max}}{\partial \theta} = \frac{X'(h_\lambda(Y) - X\theta)}{\hat{\sigma}^2}.$$

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

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

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

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

### Intervalle de confiance

Etant donné que l'emv est asymptotiquement normalement distribué, donc on peut construire le test

$$T = \frac{\hat{\beta} - \beta}{\sqrt{\frac{\widehat{Var}(\hat{\beta})}{n}}} \sim \mathcal{N}(0, I_n). \text{ Par définition, } P(q_{\alpha/2} < \frac{\hat{\beta} - \beta}{\sqrt{\frac{\widehat{Var}(\hat{\beta})}{n}}} < q_{1-\alpha/2}) = 1 - \alpha, \text{ donc on peut obtenir}$$

l'intervalle de confiance  $[\hat{\beta} - q_{1-\alpha/2} \sqrt{\frac{\widehat{Var}(\hat{\beta})}{n}}, \hat{\beta} + q_{\alpha/2} \sqrt{\frac{\widehat{Var}(\hat{\beta})}{n}}]$ , 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  $[\hat{\beta} - q_{1-\alpha/2} \sqrt{\frac{\widehat{Var}(\hat{\beta})}{n}}, \hat{\beta} + q_{1-\alpha/2} \sqrt{\frac{\widehat{Var}(\hat{\beta})}{n}}]$ .

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

### Test de Wald

On définit  $A = [0, 1, 0]$ ,  $\beta = (\theta, \lambda, \sigma^2)'$ ,  $\beta_0 = (\theta_0, \lambda_0, \sigma_0^2)'$

$H_0 : A(\beta - \beta_0) = 0$  contre  $H_1 : A(\beta - \beta_0) \neq 0$

Sous  $H_0$ , avec la delta méthode:

$$\sqrt{n}(A\hat{\beta} - A\beta_0) \rightarrow \mathcal{N}(0, AV_n A') \quad (16)$$

$$T_n = [AV_n A']^{-1/2} A(\hat{\beta} - \beta_0) \rightarrow \mathcal{N}(0, 1) \quad (17)$$

En utilisant la propriété de la statistique de Wald,  $W$  est la carré de la norme de  $T_n$  et sa loi asymptotique sous  $H_0$  est:

$$W = n(A\hat{\beta} - A\beta_0)(A\hat{V}_n A')^{-1}(A\hat{\beta} - A\beta_0)' \rightarrow \chi^2(1) \quad (18)$$

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

## 5. Test du rapport vraisemblance

Par le théorème asymptotique du RV, sous  $H_0$ :

$$TRV = -2\log(RV) \rightarrow \chi^2(1) \quad (19)$$

$TRV \geq 0$ , la région de rejet  $\mathcal{R} = \left\{ TRV > q_{1-\alpha}^{\chi^2(1)} \right\}$  du test de rapport de vraisemblances maximales est asymptotiquement de niveau  $\alpha$ ,  $P_{(H_0)}(\mathcal{R}) \rightarrow \alpha$ .

Par la définition de rapport de vraisemblance:

$$RV = \frac{L(\lambda_0; Y)}{L(\hat{\lambda}; Y)} \quad (20)$$

où  $L$  est la fonction de vraisemblance.

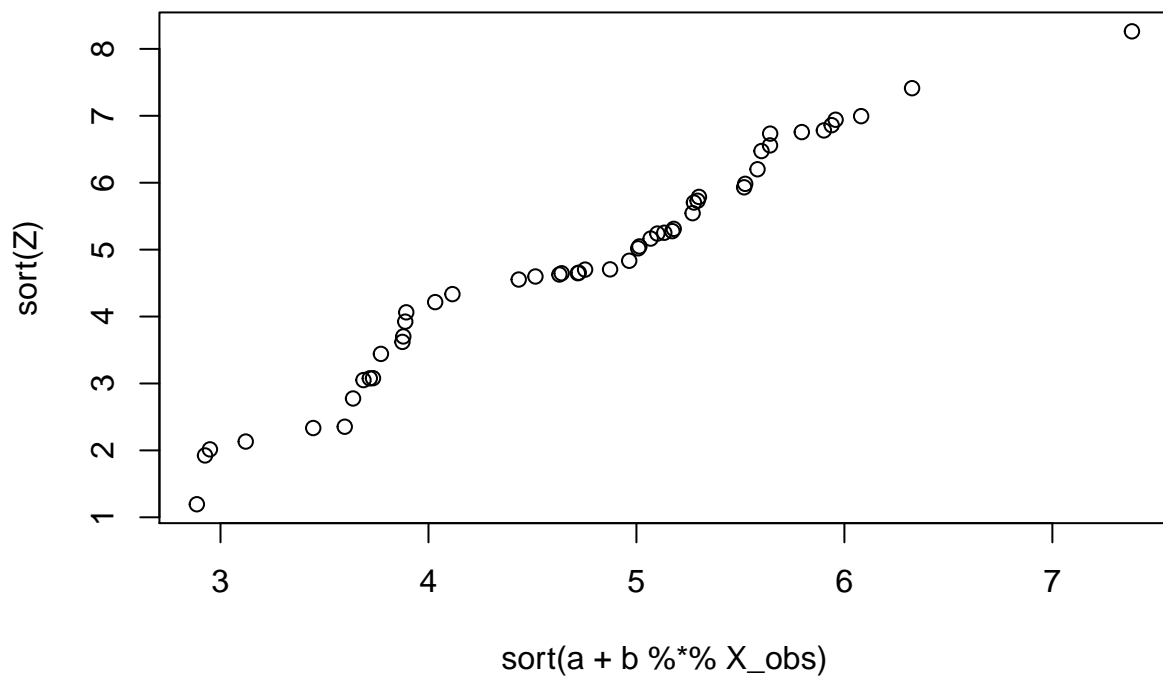
$$\begin{aligned} TRV &= -2 \log\left(\frac{L(\lambda_0; Y)}{L(\hat{\lambda}; Y)}\right) \\ &= -2(\log L(\lambda_0; Y) - \log L(\hat{\lambda}; Y)) \\ &= 2(L_{\max}(\hat{\lambda}; Y) - L_{\max}(\lambda; Y)) \\ &= 2\left(-\frac{n}{2} \log(\hat{\sigma}^2(\hat{\lambda})) + \frac{n}{2} \log(\hat{\sigma}^2(\lambda))\right) \\ &= n \log\left(\frac{\hat{\sigma}^2(\lambda)}{\hat{\sigma}^2(\hat{\lambda})}\right) \end{aligned} \quad (21)$$

## 2 Test de la méthode sur des données simulées

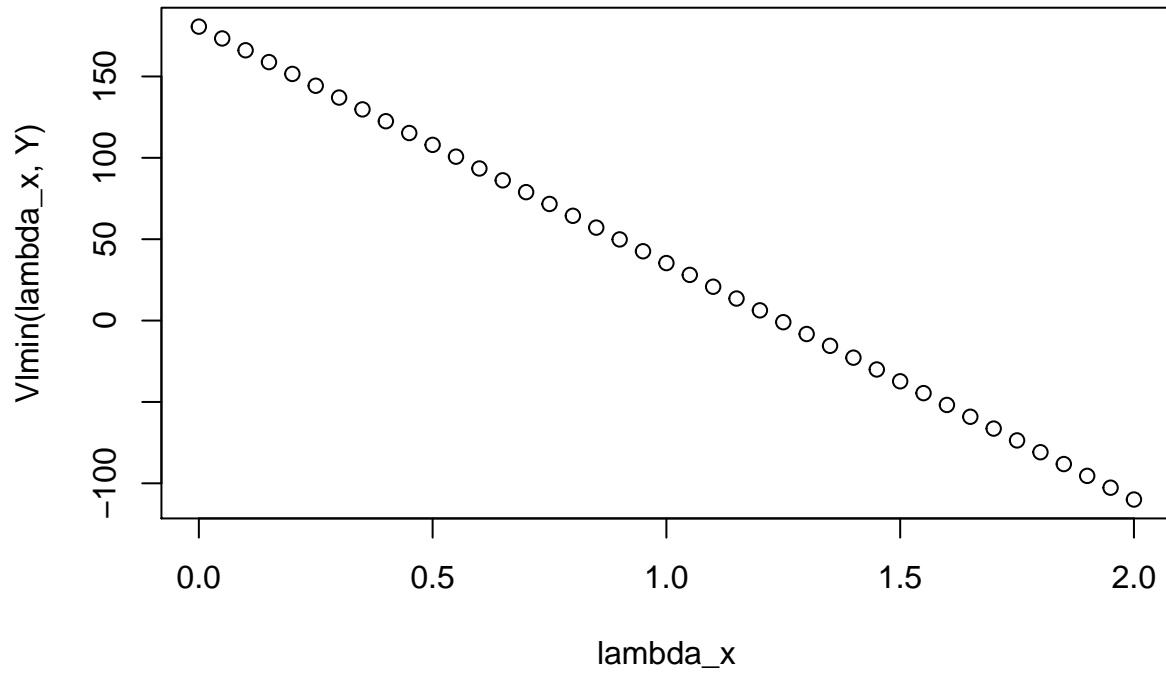
### 1.

La condition de convergence indiquée dans la section 1 est-elle vérifiée?





2.



### 3 Cas pratique

#### Conclusion