project

Gérémi Bridonneau et Yue Zhang

Introduction

1 La transformation de Box-Cox

1.

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

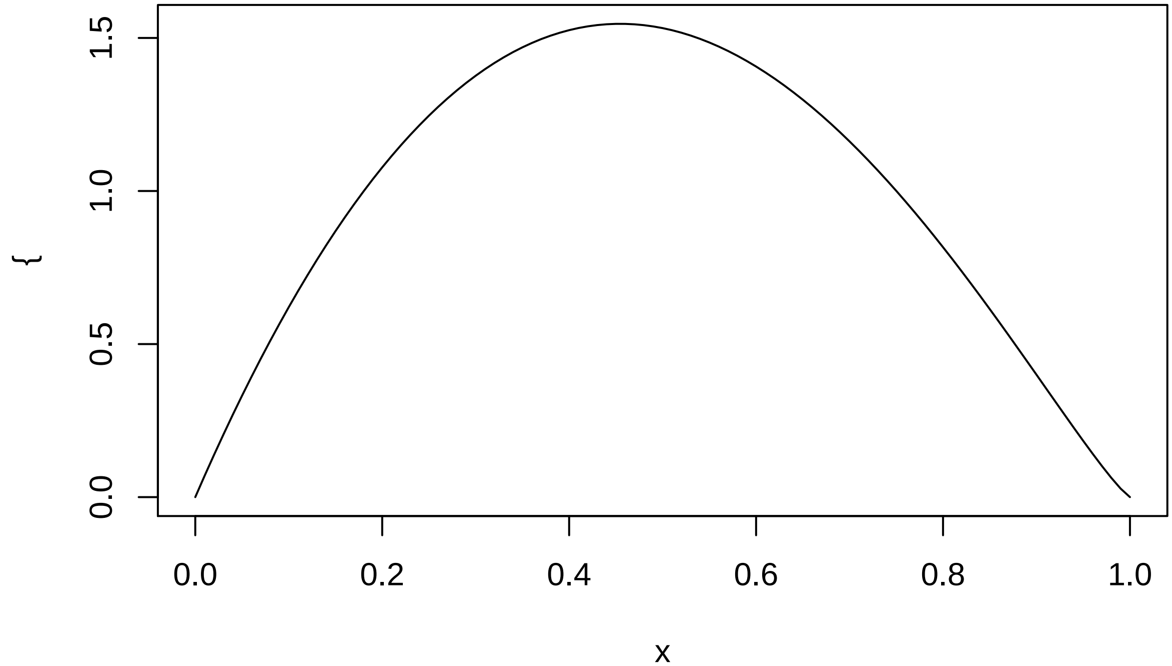
On a (1) $h_\lambda(y) = x^{'}\theta + \varepsilon$, $\varepsilon \sim \mathcal{N}(0, \sigma I_n)$.

– (mon idée de solution)

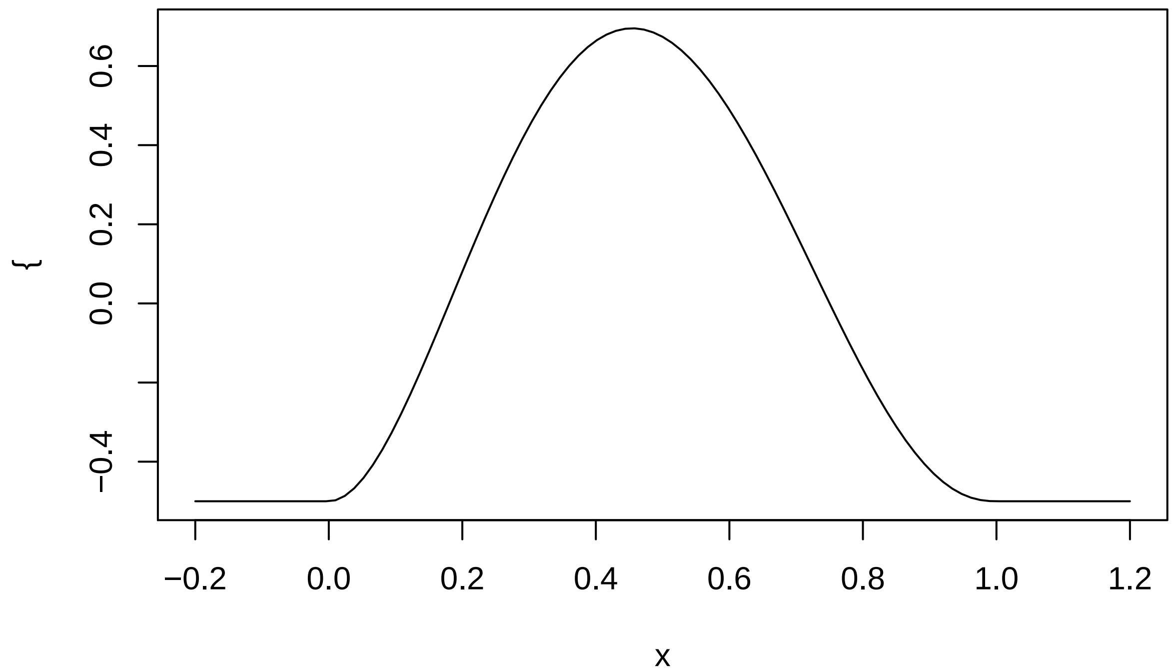
La transformation $\tilde{h}_\lambda$ est valable seulement pour $y > 0$. De plus pour tout $\lambda \neq 0$, la transformation $\tilde{h}_\lambda$ est borné et donc la transformation ne peut pas être gaussienne. Pour $\lambda = 0$ on n'a pas ce problème grâce à la surjectivité du logarithme.

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)})
```




```
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.

Supposons que pour un certain triplet $(\lambda, \theta, \sigma^2)$ on ait $h_\lambda(Y_i) = Z_i = x_i\theta + \epsilon_i,\ \varepsilon_i \equiv_{iid} \mathcal{N}(0, \sigma^2)$

3.

4.

5.

2 Test de la méthode sur des données simulées

3 Cas pratique

Conclusion