Université de Montpellier Année 2021-2022
HAX912X - Modèles linéaires généralisés

Examen de contrôle continu
sans document - 13 octobre 2021

Exercice 1 (4 pts)

Expliquer en quoi consiste une regression lasso ? Comment peut-on choisir le paramètre de régularisation ?

Exercice 2 (6 pts)

On considère n variables aléatoires indépendantes $y_1, \dots, y_n$ telles que y_i est distribuée suivant une loi inverse gaussienne de paramètres (μ_i, σ^2) où $\mu_i > 0$ et $\sigma^2 > 0$.
Le terme inverse ne doit pas être mal interprété, la loi est inverse dans le sens suivant : la valeur du mouvement brownien à un temps fixé est de loi normale, à l'inverse, le temps en lequel le mouvement brownien avec une dérive positive atteint une valeur fixée est de loi inverse gaussienne. Nous avons

$$f(y_i; \mu_i, \sigma^2) = \frac{1}{\sqrt{2\pi y_i^3 \sigma}} \exp\left(-\frac{(y_i - \mu_i)^2}{2(\mu_i \sigma)^2 y_i}\right) \mathbf{1}_{y_i > 0}\,.$$

1 (2 pts) Montrer que la loi inverse gausienne appartient à la famille exponentielle avec un paramètre de nuisance.

2 (2 pts) Calculer $\mathbb{E}(y_i)$ et $\mathbb{V}(y_i)$.

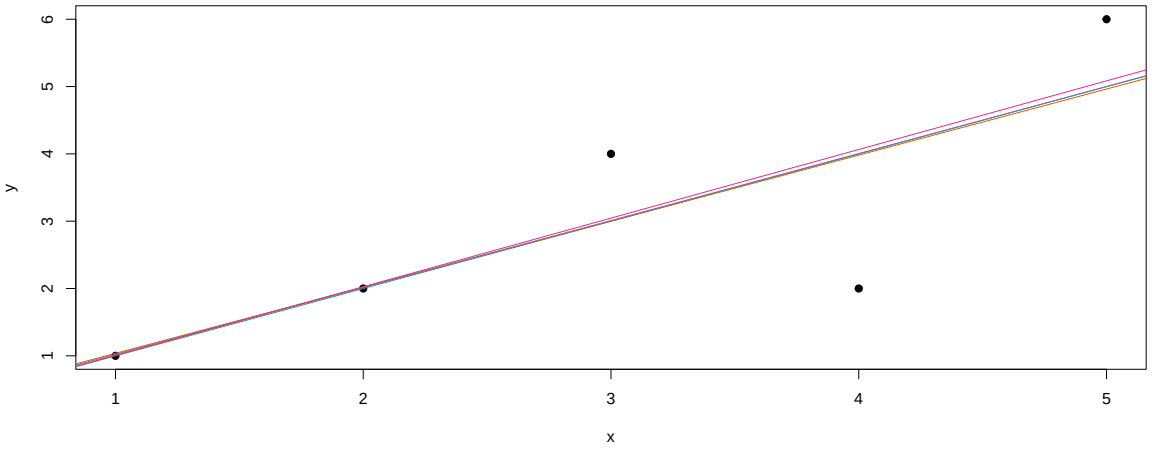
Pour tout $i = 1, \dots, n$, on suppose que $\log(\mu_i) = a + bx_i$ où $x_i \in \mathbb{R}$ est supposé connu, a et b sont des paramètres inconnus.

3 (2 pts) Montrer qu'il s'agit d'un modèle linéaire généralisé. La fonction de lien canonique a-t-elle été utilisée ?

Exercice 3 (4 pts)

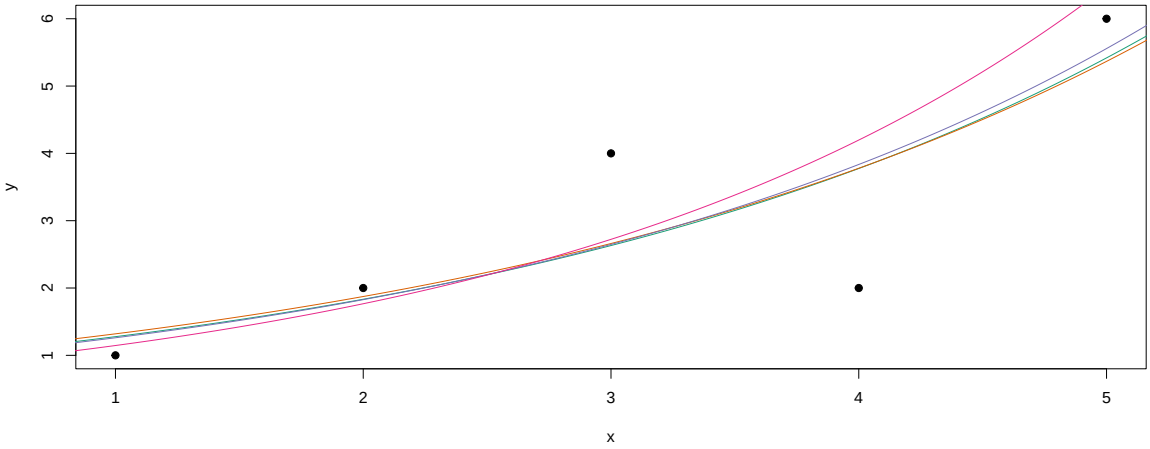
Expliquer en détails les procédures mises en oeuvre par l'intermédiaire du code R ci-dessous et expliciter clairement ce que montrent les résultats.

```
x <- c(1,2,3,4,5)
y <- c(1,2,4,2,6)
base <- data.frame(x,y)

regNId <- glm(y~x,family=gaussian(link="identity"),data=base)
regNlog <- glm(y~x,family=gaussian(link="log"),data=base)
regPId <- glm(y~x,family=poisson(link="identity"),data=base)
regPlog <- glm(y~x,family=poisson(link="log"),data=base)
regGId <- glm(y~x,family=Gamma(link="identity"),data=base)
regGlog <- glm(y~x,family=Gamma(link="log"),data=base)
regIGId <- glm(y~x,family=inverse.gaussian(link="identity"),data=base)
regIGlog <- glm(y~x,family=inverse.gaussian(link="log"),data=base

library(RColorBrewer)
darkcols <- brewer.pal(8, "Dark2")

plot(x,y,pch=19)
abline(regNId,col=darkcols[1])
abline(regPId,col=darkcols[2])
abline(regGId,col=darkcols[3])
abline(regIGId,col=darkcols[4])
```




```
plot(x,y,pch=19)
u <- seq(.8,5.2,by=.01)
lines(u,predict(regNlog,newdata=data.frame(x=u),type="response"),
col=darkcols[1])
```

```
lines(u,predict(regPlog,newdata=data.frame(x=u),type="response"),
col=darkcols[2])
lines(u,predict(regGlog,newdata=data.frame(x=u),type="response"),
col=darkcols[3])
lines(u,predict(regIGlog,newdata=data.frame(x=u),type="response"),
col=darkcols[4])
```



Exercice 4 (6 pts)

Commenter en détails le fichier R Markdown fourni en annexes. Il s'agit d'une étude sur les facteurs influençant la présence d'inondation (runoff) lors de tempêtes.

```
Data collected over a 4-year period from a Madison home.
Outcome: indicator if a rain storm produces runoff.
Multiple predictors.
```

Correction controlli continui HAX912X

13/11/2021

Esercizio 3

1) $f(y;\mu,\sigma^2) = \exp\left\{-\frac{y^2-2y\mu+\mu^2}{2\mu^2\sigma^2 y} - \frac{1}{2}\log(\sigma^2)\right\}$

$$\frac{1}{\sqrt{2\pi y^3}}\,\Pi\{y>0\}$$

$$f(y;\mu,\sigma^2) = \left\{\frac{y\left(-\frac{1}{2\mu^2}\right)+\frac{1}{\mu}}{\sigma^2} - \frac{1}{2\sigma^2 y} - \frac{1}{2}\log(\sigma^2)\right\}$$

$$r(\omega|y)$$

$$\theta = -\frac{1}{2\mu^2} \quad \text{e} \quad \phi = \sigma^2$$

$$\Rightarrow \mu = \frac{1}{\sqrt{-2\theta}} \quad \text{e} \quad b(\theta) = -\sqrt{-2\theta}$$

$$c(y,\phi) = -\frac{1}{2\phi y} - \frac{1}{2}\log(\phi)$$

2] $E(y) = b'(\theta) = -\left(\frac{1}{2}\right)\frac{(-2)}{\sqrt{-2\theta}}$

$\Leftrightarrow E(y) = \frac{1}{\sqrt{-2\theta}} = \mu$

$V(y) = \phi\, b''(\theta)$

$= \sigma^2 \frac{-(-2)}{2(-2\theta)\sqrt{-2\theta}}$

$= \sigma^2 \frac{1}{(-2\theta)^{3/2}}$

$= \sigma^2 \mu^3$

3] $\log(\mu) = \alpha + \beta x$

$\Rightarrow E(y) = \mu = e^{\alpha + \beta x}$

Il s'agit bien d'un modèle linéaire généralisé

$g(\mu) = e^{\mu} \neq b'(\mu)$

ce n'est pas le lien canonique.