

## **Part 2 - Extensions of the linear model**

### **Variable selection and Generalized linear model**

Cathy Maugis-Rabusseau and Pierre Neuvial

Institut de Mathématiques de Toulouse

2021-06-26

- 1 Variable selection
- 2 Régressions régularisées
- 3 Modèle linéaire généralisé

## Section 1

### Variable selection

# Contexte

- Par soucis de simplicité, on se place ici dans le cadre de la régression linéaire multiple. Les outils présentés peuvent être utilisés dans un cadre plus général (bien souvent sans travail supplémentaire).

$$Y = X\theta + \varepsilon \text{ avec } \varepsilon \sim \mathcal{N}(0_n, \sigma^2 I_n)$$

- On se concentre sur l'étude de la matrice  $X$  i.e sur les variables explicatives (quantitatives)
- Comment choisir un modèle en adéquation avec les données mais pas trop complexe et donc éliminer certaines variables peu explicatives ?
- Présentation d'approches permettant d'affiner (de sélectionner) un modèle parmi d'autres, c'est-à-dire de déterminer quelles sont les variables les plus "explicatives", les plus importantes dans notre compréhension du modèle.

# Contexte

- Echantillon de taille  $n$  représentant des observations sur une variable réponse quantitative  $Y$  et de  $p$  variables explicatives quantitatives réelles  $\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(p)}$ .
- Une famille de modèles  $\mathcal{M}$  représentant formellement une famille de sous-ensembles de  $\{1, \dots, p\}$ .
- Pour  $m \in \mathcal{M}$ , on note  $|m|$  le cardinal de  $m$  et  $X_{(m)}$  la matrice constituée des vecteurs  $\mathbf{x}^{(j)}$  pour  $j \in m$  (supposée régulière,  $\text{rg}(X_{(m)}) = |m| + 1$ )
- Problème de **sélection de modèle**  
Compromis entre bien ajuster la réponse et avoir un modèle le plus simple possible pour l'interprétation

## Subsection 1

### Critères de sélection

# Le coefficient ajusté $\widetilde{R^2}_m$

- Coefficient  $R^2$  :

$$R^2(m) = \frac{SSE(m)}{SST} = 1 - \frac{SSR(m)}{SST}$$

croit naturellement avec  $|m|$  donc sélection du modèle le plus complexe.

- Le **coefficient  $R^2$  ajusté** :

$$\widetilde{R^2}(m) = 1 - \frac{n-1}{n-|m|-1} \cdot \frac{SSR(m)}{SST} = 1 - \frac{n-1}{n-|m|-1} \cdot \frac{\|Y - X_m \hat{\theta}_{(m)}\|^2}{\|Y - \bar{Y}\mathbf{1}_n\|^2}.$$

- Le  $\widetilde{R^2}(m)$  permet de tenir compte du nombre de régresseurs retenus et propose donc un compromis entre l'adéquation et le paramétrage du modèle.

# Le critère $C_p$ de Mallows

- Critère basé sur la minimisation du risque quadratique
- On retient le modèle minimisant le  $C_p$  de Mallows

$$C_p(m) = \|Y - \hat{Y}_{(m)}\|^2 + 2|m|\sigma^2$$

si la variance est connue.

- Dans le cas où la variance est inconnue, on utilisera  $\widehat{\sigma^2}$  l'estimateur de la variance dans le modèle complet.
- Compromis biais-variance = compromis entre **bon ajustement** et **complexité** du modèle.

# Critère AIC

- Le critère AIC (Akaike Information Criterion) est construit pour minimiser la dissemblance de Kullback-Leibler entre les 2 modèles.
- Le critère AIC consiste à sélectionner le modèle minimisant

$$\text{AIC}(m) = -2 \ln(L(Y|\hat{\theta}_m)) + 2D_m$$

- $D_m$  est la dimension du modèle  $m$  (nb de param. pour  $m$ )
- $L(Y|\hat{\theta}_m)$  = vraisemblance au maximum de vraisemblance

Dans le cas gaussien,  $\text{AIC}(m) = n \ln(\tilde{\sigma}_{(m)}^2) + 2(|m| + 2)$ .

- Ce type de critère fonctionne plutôt bien pour de petites collections de modèles. La qualité d'estimation a tendance à se dégrader lorsque  $m$  augmente. Afin de pallier ce problème, il est possible d'utiliser le critère AIC corrigé :

$$\text{AIC}_c(m) = n \ln\left(\tilde{\sigma}_{(m)}^2\right) + n \frac{n + |m| - 1}{n - |m| - 3}.$$

# Critère BIC

- Le critère BIC (Bayesian Information Criterion) est une extension du critère AIC et utilise le point de vue bayésien.
- Le critère BIC consiste à sélectionner le modèle minimisant

$$\begin{aligned} BIC(m) &= -2 \ln(L(Y|\hat{\theta}_m)) + D_m \ln(n) \\ &= n \ln(\hat{\sigma}_{(m)}^2) + D_m \ln(n). \end{aligned}$$

- Le critère BIC est asymptotiquement consistant, càd il sélectionne le "vrai" modèle (s'il est dans la collection de modèle) avec une probabilité qui tend vers 1 quand la taille de l'échantillon tend vers l'infini.

## Subsection 2

### Algorithmes

# Stratégies d'exploration

- On choisit un critère (à minimiser) pour comparer les modèles
- La détermination du “meilleur” modèle par une recherche exhaustive est impossible en raison du nombre de modèles à explorer.
- Utilisation des méthodes pas à pas :
  - algorithme descendant
  - algorithme ascendant
  - algorithme stepwise
  - ...

# Algorithme descendant

On part du **modèle complet** ( $p$  variables explicatives) et on cherche, à chaque étape de l'algorithme, la variable la plus pertinente à **retirer** selon le critère choisi. On itère ainsi l'algorithme jusqu'à atteindre l'ensemble vide. Parmi les ensembles de variables visités pendant l'algorithme, on retient le meilleur au vu du critère. Certains algorithmes s'arrêtent dès lors qu'un seuil donné est atteint.

- **Initialisation :**  $m_{[0]} = \{1, \dots, p\}$
- **Itération  $t$  :**
  - *Etape 1* : Pour tout  $j \in m_{[t]}$ , on calcule  $c_j = \text{CRIT}(m_{[t]} \setminus \{j\})$ .
  - *Etape 2* :  $\hat{j} = \arg \max_{j \in m_{[t]}} c_j$
  - *Etape 3* :  $m_{[t+1]} = m_{[t]} \setminus \{\hat{j}\}$ 
    - Si  $m_{[t+1]} \neq \emptyset$ , on retourne à l'étape 1
    - Sinon stop.

# Algorithme ascendant

On part de l'**ensemble vide** de variables et on cherche, à chaque étape de l'algorithme, la variable la plus pertinente **à ajouter** selon le critère choisi. On itère ainsi l'algorithme jusqu'à intégrer toutes les variables. Parmi les ensembles de variables visités pendant l'algorithme, on retient le meilleur au vu du critère. Certains algorithmes s'arrêtent dès lors qu'un seuil donné est atteint.

- **Initialisation :**  $m_{[0]} = \emptyset$
- **Itération  $t$  :**
  - *Etape 1* : Pour tout  $j \in \{1, \dots, p\} \setminus m_{[t]}$ , on calcule  $c_j = \text{CRIT}(m_{[t]} \cup \{j\})$ .
  - *Etape 2* :  $\hat{j} = \arg \min_j c_j$
  - *Etape 3* :  $m_{[t+1]} = m_{[t]} \cup \{\hat{j}\}$ 
    - Si  $m_{[t+1]} \neq \{1, \dots, p\}$ , on retourne à l'étape 1
    - Sinon stop.

## Autres algorithmes

- Les **méthodes stepwise** : Partant d'un modèle donné, on opère une sélection d'une nouvelle variable (comme avec une méthode ascendante), puis on cherche si on peut éliminer une des variables du modèle (comme pour une méthode descendante) et ainsi de suite. Il faut définir pour une telle méthode un critère d'entrée et un critère de sortie.
- La **méthode des “ $s$  best subsets”** : On cherche de façon exhaustive parmi tous les sous-ensembles de  $s$  variables, les  $s$  meilleures, au sens du critère considéré.

## Subsection 3

### Exemple

# Données

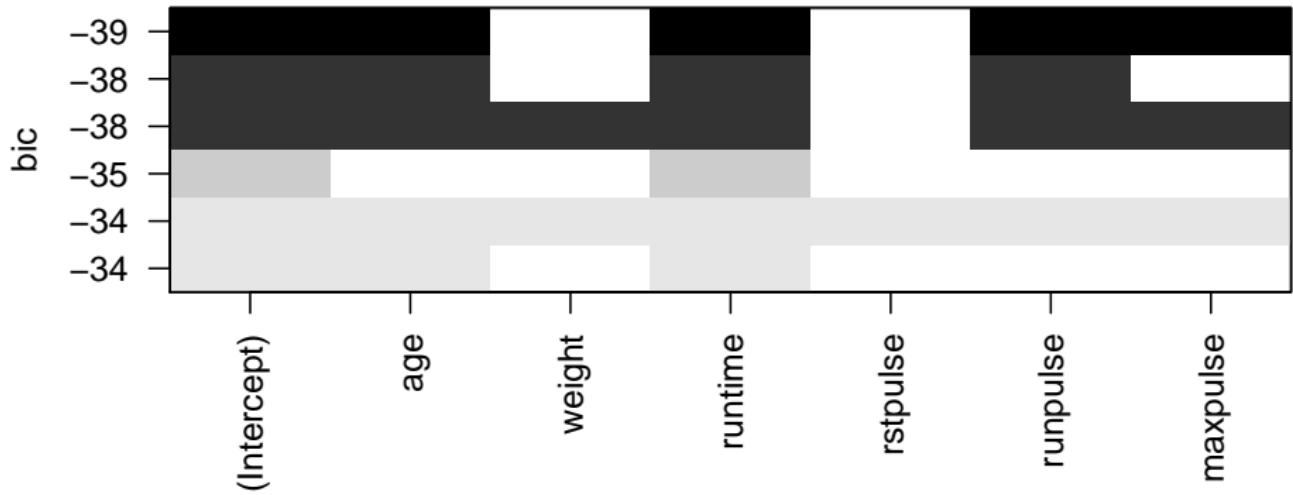
On reprend l'exemple du fitness de la partie 1

- Data collected for 31 persons during aerobic sessions
- 7 variables:
  - age (a): age
  - weight (w): weight
  - oxy (oxy): oxygen consumption
  - runtime (run): time of effort
  - rsthpulse (rst): heart rate measurement 1
  - runpulse (rp): heart rate measurement 2
  - maxpulse (maxp): heart rate measurement 3

We want to explain the consumption of oxygen (response variable  $Y=\text{oxy}$ ) according to the other quantitative variables ( $p = 6$ ).

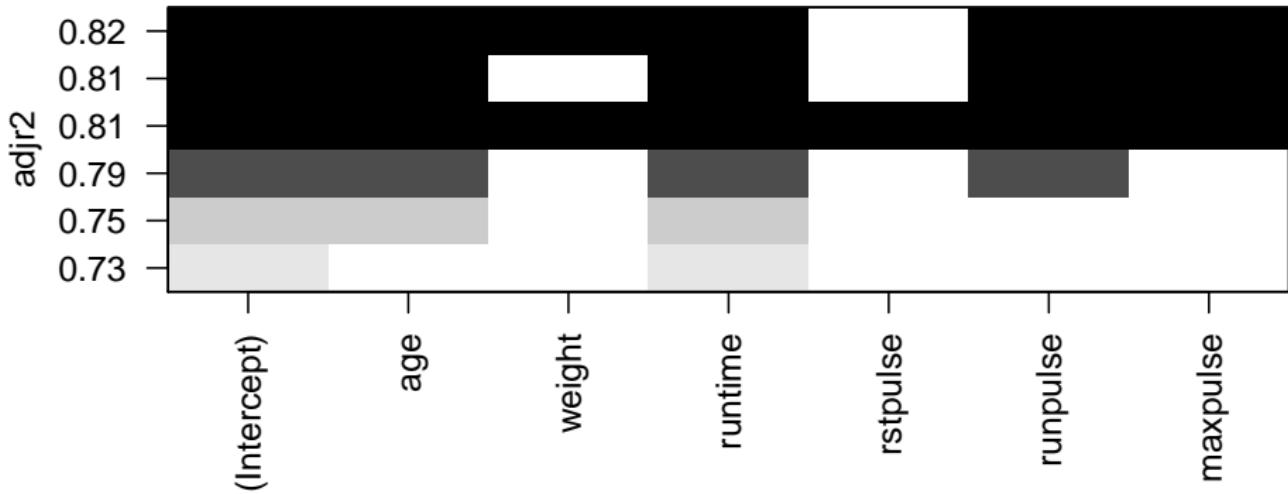
# Exemple avec *leaps()*

```
library(leaps)
choix=regsubsets(oxy~.,data=fitness,method="backward")
plot(choix,scale="bic")
```



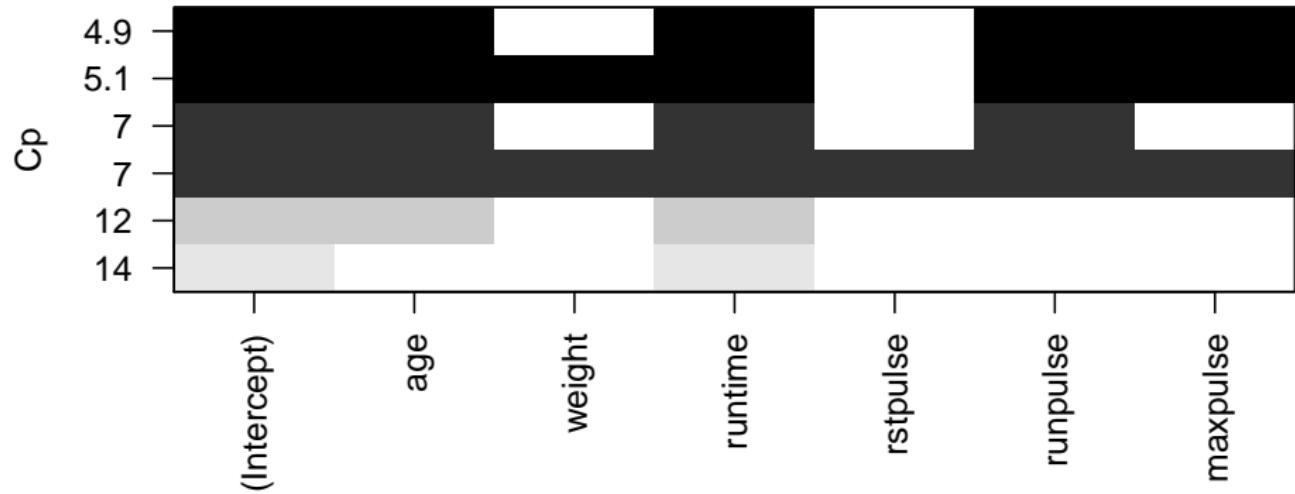
## Exemple avec *leaps()*

```
plot(choix,scale="adjr2")
```



# Exemple avec *leaps()*

```
plot(choix,scale="Cp")
```



# Exemple avec *stepAIC()*

```
reg=lm(oxy~,data=fitness)
modselect_aic=stepAIC(reg,trace=FALSE,direction=c("backward")) # AIC
modselect_bic=stepAIC(reg,trace=TRUE,direction=c("backward"),k=log(nrow(fitness))) # BIC
```

Start: AIC=68.2

oxy ~ age + weight + runtime + ristpulse + runpulse + maxpulse

	Df	Sum of Sq	RSS	AIC
- ristpulse	1	0.571	129.41	64.903
- weight	1	9.911	138.75	67.063
<none>			128.84	68.200
- maxpulse	1	26.491	155.33	70.562
- age	1	27.746	156.58	70.812
- runpulse	1	51.058	179.90	75.114
- runtime	1	250.822	379.66	98.268

Step: AIC=64.9

oxy ~ age + weight + runtime + runpulse + maxpulse

	Df	Sum of Sq	RSS	AIC
- weight	1	9.52	138.93	63.669
<none>			129.41	64.903
- maxpulse	1	26.83	156.23	67.309
- age	1	27.37	156.78	67.417
- runpulse	1	52.60	182.00	72.041
- runtime	1	320.36	449.77	100.087

Step: AIC=63.67

oxy ~ age + runtime + runpulse + maxpulse

	Df	Sum of Sq	RSS	AIC
<none>			138.93	63.669
- maxpulse	1	21.90	160.83	64.773
- age	1	22.84	161.77	64.954

# Validation

```
reg.fin<-lm(oxy~age + runtime + runpulse + maxpulse,data=fitness)
anova(reg.fin,reg)
```

## Analysis of Variance Table

```
Model 1: oxy ~ age + runtime + runpulse + maxpulse
Model 2: oxy ~ age + weight + runtime + rsthpulse + runpulse + maxpulse
Res.Df   RSS Df Sum of Sq    F Pr(>F)
1      26 138.93
2      24 128.84  2     10.092 0.94 0.4045
```

## Section 2

# Régressions régularisées

# Contexte

- Modèle singulier ( $\text{rg}(X) < k \Rightarrow X'X$  non inversible)
- Ce cas se présente quand
  - le nombre de variables explicatives est supérieur au nombre d'observations ( $n < p$ )
  - $n > p$  mais des variables sont linéairement redondantes
- $\Gamma_{\hat{\theta}} = \sigma^2(X'X)^{-1}$  donc la précision de  $\hat{\theta}$  diminue quand  $X'X$  se rapproche d'une matrice non inversible.
- En prédiction : la qualité (au sens quadratique) de la prédiction à la vraie réponse vaut biais<sup>2</sup> + variance.  
⇒ on peut préférer une augmentation légère du biais pour avoir une diminution de la variance.

# Généralités

- Méthodes de **régression dites régularisées** : minimiser un critère

$$\operatorname{argmin}_{\theta \in \mathbb{R}^k} \|Y - X\theta\|^2 + \tau \operatorname{pen}(\theta)$$

où  $\tau > 0$  à choisir et  $\operatorname{pen}(\theta)$  liée au contrôle d'une norme.

- En pratique :
  - On commence par centrer et réduire les variables explicatives  $\Rightarrow \tilde{X}$ .
  - L'intercept  $\theta_0$  assure au modèle de se positionner autour du comportement moyen de  $Y$  donc  $\tilde{Y} = Y - \bar{Y}\mathbf{1}_n$  (et on peut potentiellement le réduire)
- On va donc considérer le modèle linéaire

$$\tilde{Y} = \tilde{X}\theta + \varepsilon \text{ avec } \theta = (\theta_1, \dots, \theta_p)'$$

# Généralités

- But : minimiser le risque empirique régularisé :

$$\operatorname{argmin}_{\theta \in \mathbb{R}^k} \left\{ \|\tilde{Y} - \tilde{X}\theta\|^2 + \tau \|\theta\|_q^q \right\} \text{ où } \|\theta\|_q^q = \sum_{j=1}^p (\theta_j)^q.$$

- Les plus connues : **régression ridge** ( $q = 2$ ), **régression Lasso** ( $q = 1$ ) et **régression Elasticnet** (combine les 2)

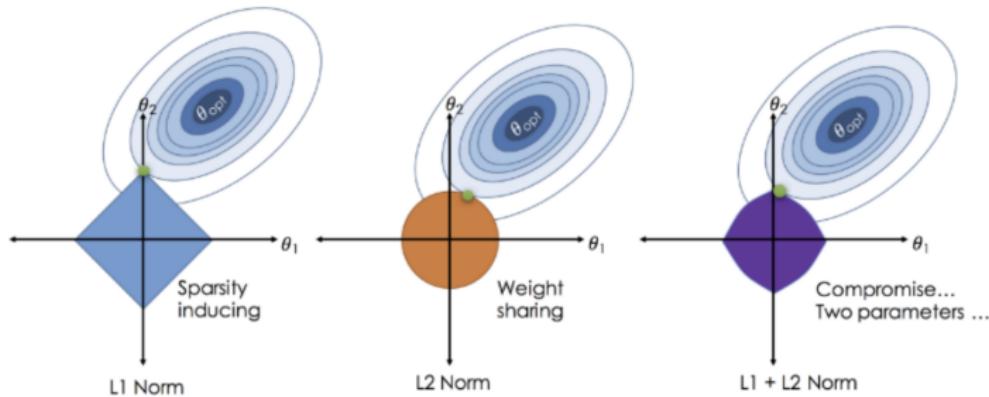
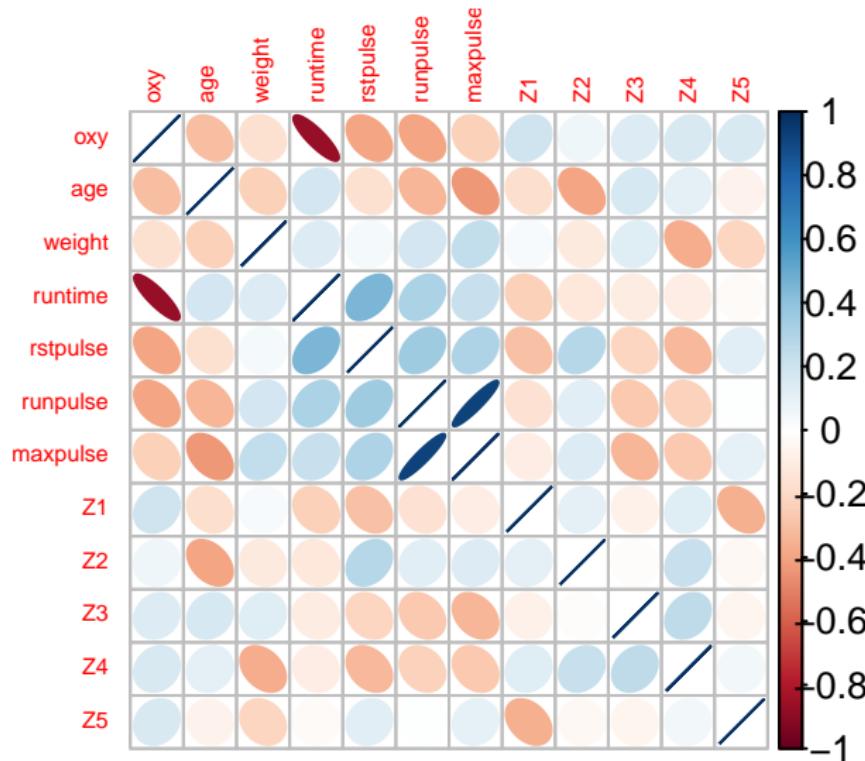


Figure 1: An image visualising how ordinary regression compares to the Lasso, the Ridge and the Elastic Net Regressors. Image Citation: Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net.

# Exemple

- Données “fitness” + 5 variables de bruit ( $\sim \mathcal{N}(0, 1)$ )



## Subsection 1

### Régression ridge

# Idée de la régression ridge

- Difficulté d'inversibilité de  $\tilde{X}'\tilde{X} \in \mathbb{R}^{p \times p}$ .
- Soit  $\tau > 0$ . Les matrices  $\tilde{X}'\tilde{X}$  et  $\tilde{X}'\tilde{X} + \tau I_p$  ont les mêmes vecteurs propres mais leur valeurs propres sont  $\{\lambda_j\}_{j \in [|1, p|]}$  et  $\{\lambda_j + \tau\}_{j \in [|1, p|]}$  respectivement.  
Ainsi,  $\det(\tilde{X}'\tilde{X} + \tau I_p) > \det(\tilde{X}'\tilde{X})$ , donc  $\tilde{X}'\tilde{X} + \tau I_p$  a "plus de chance" d'être inversible que  $\tilde{X}'\tilde{X}$ .
- Idée : remplacer  $(\tilde{X}'\tilde{X})^{-1}$  par  $(\tilde{X}'\tilde{X} + \tau I_p)^{-1}$  dans  $\hat{\theta}$

# Régression ridge

- L'estimateur ridge est donné par

$$\hat{\theta}_{\text{ridge}}(\tau) = (\tilde{X}' \tilde{X} + \tau I_p)^{-1} \tilde{X}' \tilde{Y}.$$

- L'estimateur ridge est solution du problème d'optimisation suivant

$$\operatorname{argmin}_{\theta \in \mathbb{R}^p} \|\tilde{Y} - \tilde{X}\theta\|_2^2 + \tau \|\theta\|_2^2,$$

$\Leftrightarrow$

$$\operatorname{argmin}_{\theta \in \mathbb{R}^p} \|\tilde{Y} - \tilde{X}\theta\|_2^2 \text{ sous la contrainte } \|\theta\|_2^2 \leq r(\tau)$$

où  $r(\cdot)$  est bijective.

- La régression ridge conserve toutes les variables mais elle empêche les estimateurs de prendre de trop grandes valeurs ("skrinkage") et limite ainsi la variance des prédictions.

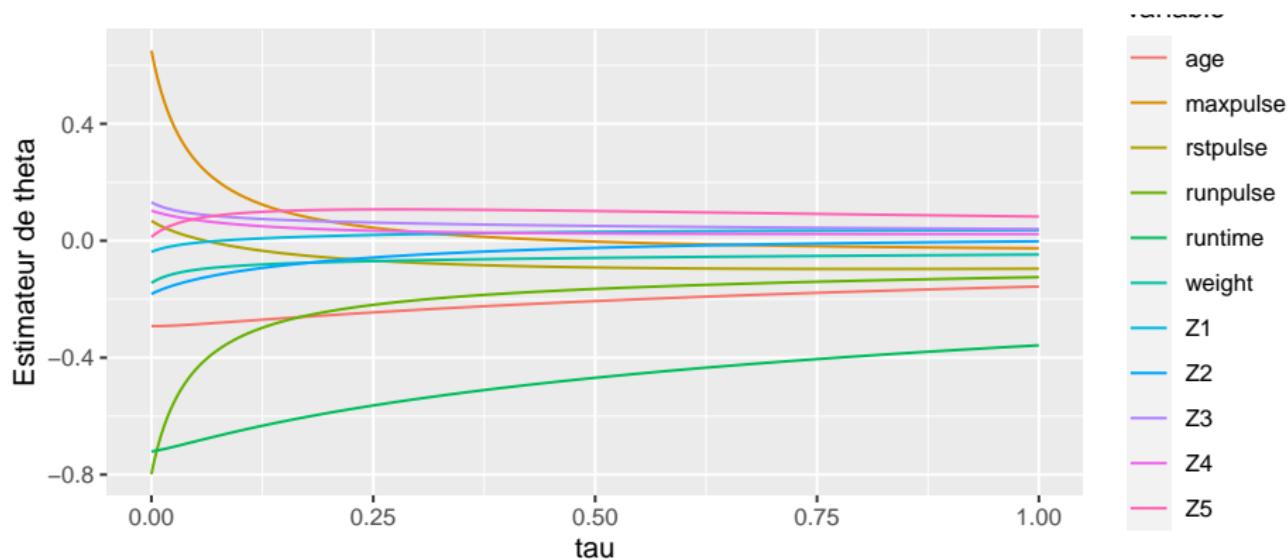
# Régression ridge

- L'estimateur  $\hat{\theta}_{\text{ridge}}(\tau)$  dépend du choix de  $\tau$  qui est un point délicat.
- Impossible de pouvoir faire ce choix a priori.
- Tracer le *chemin de régularisation* de la régression ridge

$$\tau \mapsto (\hat{\theta}_{\text{ridge}}(\tau))_j \text{ pour } j = 1, \dots, p$$

# Exemple

```
lambda_seq <- seq(0, 1, by = 0.001)
fitridge <- glmnet(tildeX,tildeY, alpha = 0, lambda  = lambda_seq,family=c("gaussian"),intercept=F)
df=data.frame(tau = rep(fitridge$lambda,ncol(tildeX)), theta=as.vector(t(fitridge$beta)),
              variable=rep(colnames(x_var),each=length(fitridge$lambda)))
g1 = ggplot(df,aes(x=tau,y=theta,col=variable))+
  geom_line()+
  ylab('Estimateur de theta')
g1
```



# Choix de $\tau$ - Validation croisée

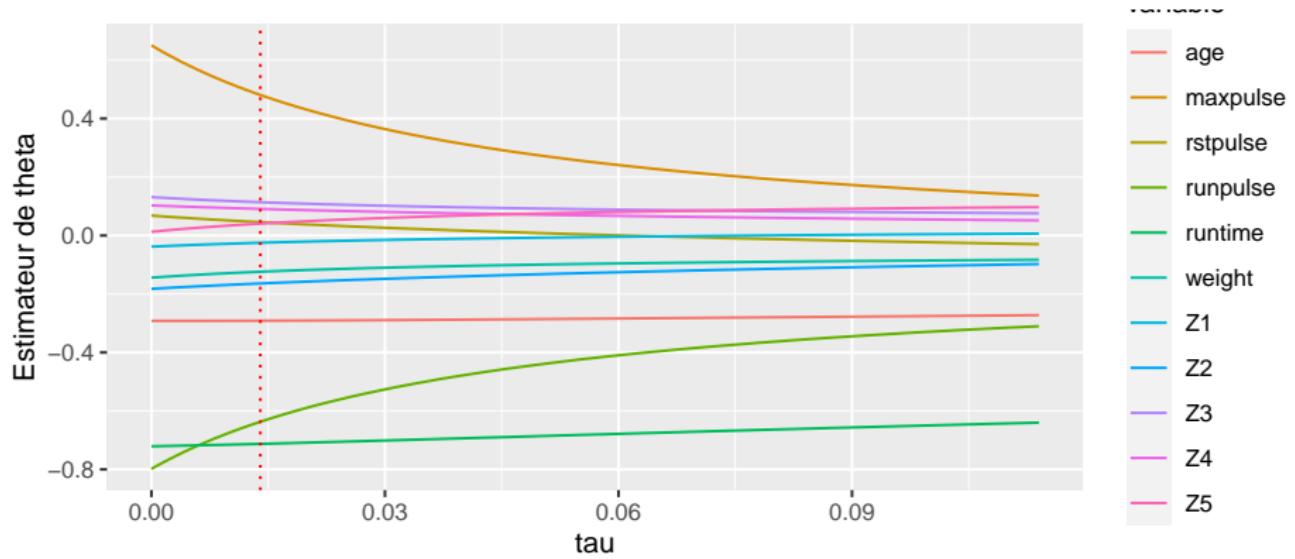
- Calibration par apprentissage / test :
  - On commence par séparer les données en un **jeu d'apprentissage** ( $Y_a, X_a$ ) et un **jeu de test** ( $Y_v, X_v$ ).
  - On estime la régression ridge sur ( $Y_a, X_a$ ) pour chaque valeur de  $\tau$  dans une grille de valeurs choisie
  - On prédit sur ( $Y_v, X_v$ ) pour chaque  $\tau$ :  $\hat{Y}_{\text{ridge}, v}(\tau)$ .
  - La qualité du modèle est alors obtenue en comparant les vraies données  $Y_v$  et les valeurs prédites  $\hat{Y}_{\text{ridge}, v}(\tau)$ . Par exemple,

$$PRESS(\tau) = \|Y_v - \hat{Y}_{\text{ridge}, v}(\tau)\|^2.$$

- Finalement on choisit la valeur de  $\tau$  qui minime ce critère.
- Validation croisée = répéter plusieurs fois le découpage entre test et apprentissage et considérer la moyenne des valeurs du critère pour chaque valeur de  $\tau$ .
- Fonction `cv.glmnet()` du package *glmnet*.

# Exemple

```
ridge_cv <- cv.glmnet(tildeX, tildeY, alpha = 0, lambda = lambda_seq, nfolds=10, type.measure=c("mse"), intercept=TRUE)
```



## Subsection 2

### Régression Lasso

# Principe

- LASSO = Least Absolute Selection and Shrinkage Operator (Tibshirani,96)
- Idée : annuler des coefficients de  $\theta$  afin d'avoir un estimateur parcimonieux (sparse en anglais)
- Induit une sélection de variables rendant le modèle plus interprétable et une matrice des variables explicatives avec de meilleures propriétés que  $X'X$ .
- L'estimateur LASSO est défini pour  $\tau > 0$  par

$$\hat{\theta}_{\text{lasso}}(\tau) \in \operatorname{argmin}_{\theta \in \mathbb{R}^p} \|\tilde{Y} - \tilde{X}\theta\|_2^2 + \tau\|\theta\|_1$$

- Ce problème de minimisation est équivalent à minimiser  $\|\tilde{Y} - \tilde{X}\theta\|_2^2$  sous la contrainte  $\|\theta\|_1 \leq r(\tau)$  avec  $r(\cdot)$  bijective.

# Principe

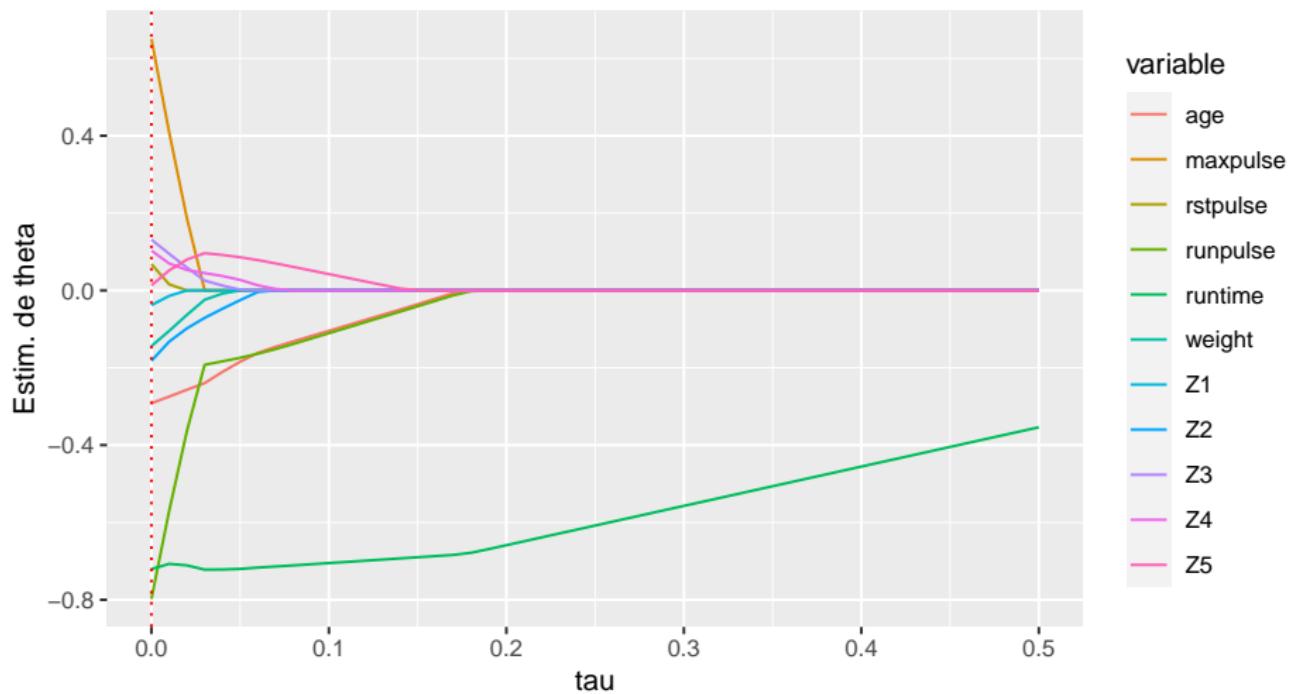
- La solution peut ne pas être unique mais le vecteur des valeurs ajustées en résultant  $\tilde{X}\hat{\theta}_{\text{lasso}}(\tau)$  est lui toujours unique.
- Choix de  $\tau$  est délicat :
  - Impossible de faire ce choix a priori
  - On peut tracer le *chemin de régularisation* de la régression Lasso

$$\tau \mapsto \hat{\theta}_{\text{lasso}}(\tau)_j \text{ pour } j = 1, \dots, p$$

- On passe par une procédure de validation croisée pour stabiliser le choix de  $\tau$

# Exemple

```
lambda_seq=seq(0,0.5,0.01)
fitlasso <- glmnet(tildeX,tildeY, alpha = 1, lambda = lambda_seq,family=c("gaussian"),intercept=F)
lasso_cv <- cv.glmnet(tildeX, tildeY, alpha = 1, lambda = lambda_seq,nfolds=10,type.measure=c("mse"),intercept=
best_lambda <- lasso_cv$lambda.min
```



### Subsection 3

## Régression Elastic-Net

## Régression Elastic-Net (Zou, 05)

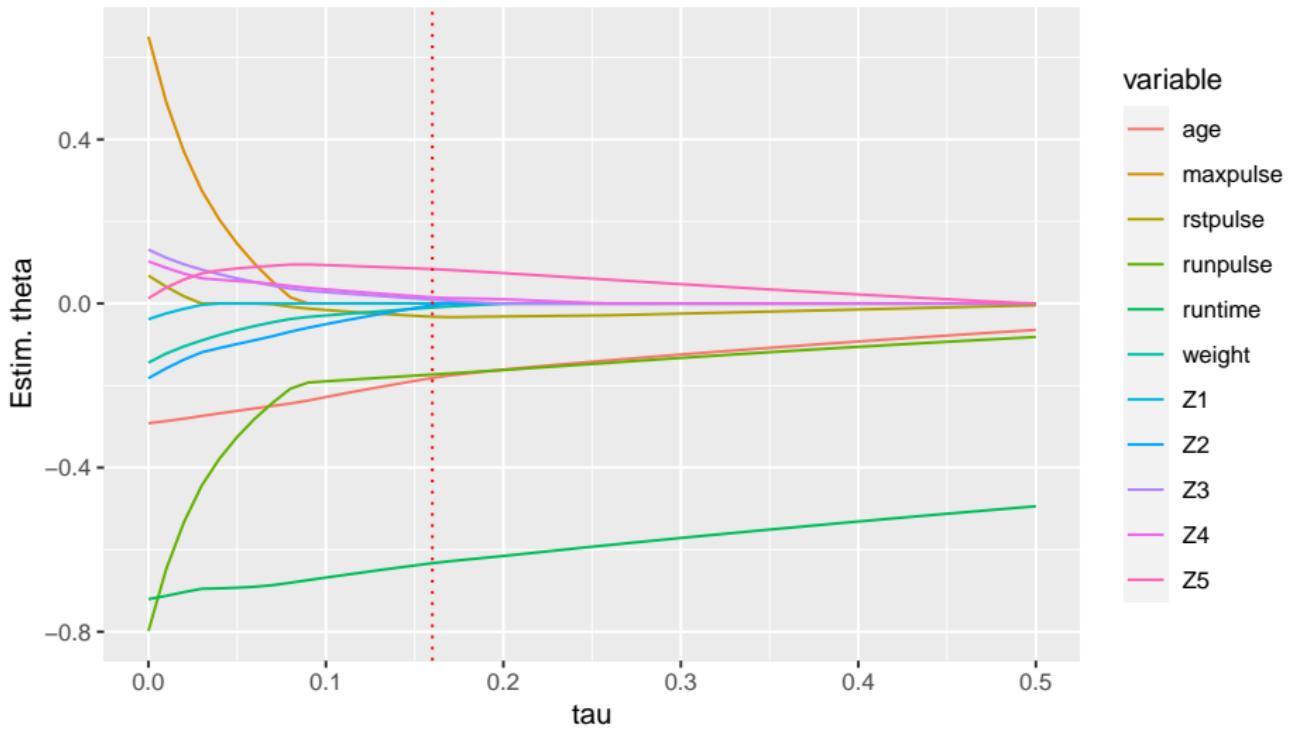
- La régression Elastic-Net combine les avantages de la régression ridge et de la régression Lasso.
- L'estimateur Elastic-Net est défini pour  $\tau > 0$  et  $\alpha > 0$  par

$$\hat{\theta}_{\text{net}}(\tau, \alpha) \in \underset{\theta \in \mathbb{R}^p}{\operatorname{argmin}} \| \tilde{Y} - \tilde{X}\theta \|_2^2 + \tau\{\alpha\|\theta\|_1 + (1 - \alpha)\|\theta\|_2^2\}$$

- Utilisation d'algorithmes d'optimisation pour déterminer  $\hat{\theta}_{\text{net}}(\tau, \alpha)$
- Calibration des seuils  $\tau$  et  $\alpha$  par validation croisée

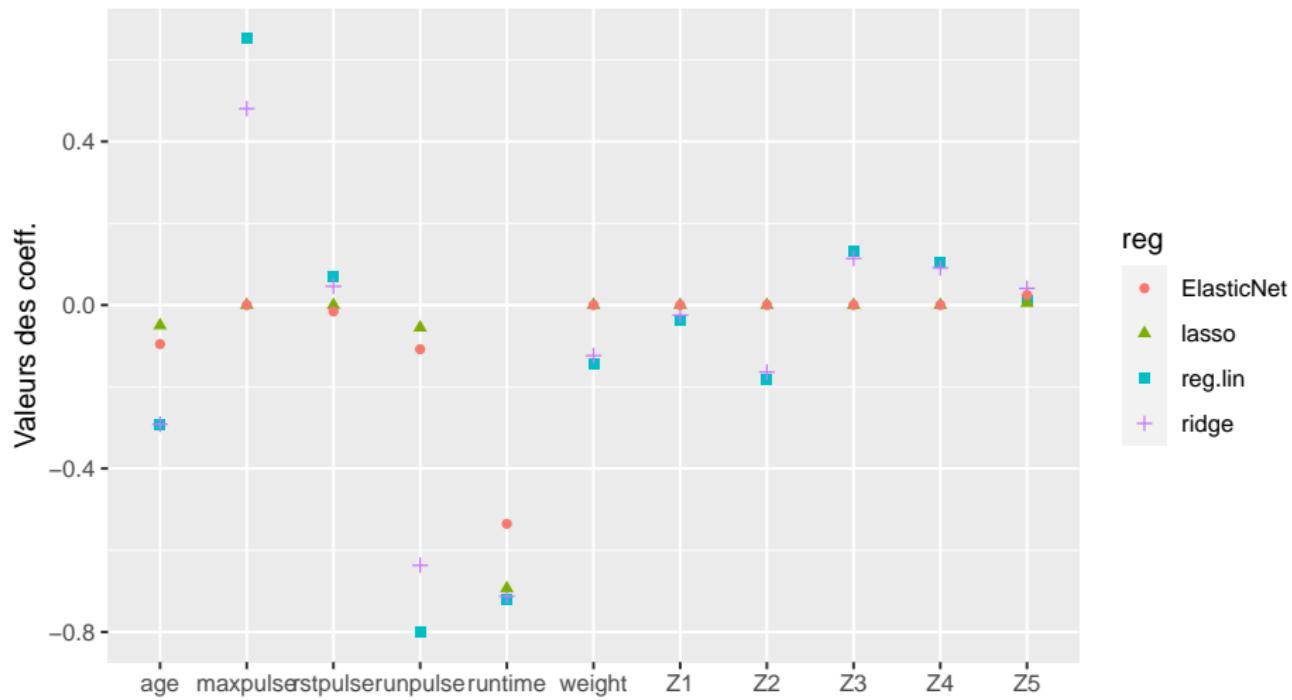
# Exemple

```
fitEN <- glmnet(tildeX,tildeY, alpha = 0.3, lambda = lambda_seq,family=c("gaussian"),intercept=F)
EN_cv <- cv.glmnet(tildeX, tildeY, alpha = 0.3, lambda = lambda_seq,nfolds=10,type.measure=c("mse"),intercept=F)
```



# Exemple

Comparaison des valeurs des coefficients :



## Section 3

# Modèle linéaire généralisé

## Subsection 1

### Principe général

# Introduction

- Données :
  - $Y = (Y_1, \dots, Y_n)'$  vecteur réponse
  - $\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(p)}$  les variables explicatives (quali et/ou quanti)
- Réécriture du modèle linéaire

$$Y = X\theta + \varepsilon \text{ avec } \varepsilon \sim \mathcal{N}_n(0_n, \sigma^2 I_n)$$



$$Y \sim \mathcal{N}_n(\mu, \sigma^2 I_n) \text{ avec } \mu := \mathbb{E}[Y] = X\theta \text{ et } \theta \in \mathbb{R}^k$$

- Objectif : Etendre à des réponses qui suivent d'autres lois de probabilités  
Ex: des réponses binaires, des comptages, ...

# Exemples

- Exemple 1 :

Dans une expérience clinique, on cherche à comparer deux modes opératoires pour une opération chirurgicale donnée. L'expérience est menée sur deux hôpitaux différents. On dispose donc ici de deux facteurs à deux modalités: *mode opératoire* et *hôpital*. La variable réponse correspond pour chaque patient au succès ou à l'échec de l'intervention : il s'agit d'une variable binaire.

- Exemple 2 :

Un assureur s'intéresse au nombre de sinistres automobile déclarés pendant ces dix dernières années. Il souhaite étudier si ce nombre de sinistres est lié à l'âge du conducteur, la taille de la voiture, .... Le nombre de sinistres peut être modélisé par une loi de Poisson.

# Modélisation d'un MLG

- Une **composante aléatoire** :

$Y$  = vecteur réponse **aléatoire**

choix d'une loi de probabilité (dans la famille exponentielle)

vecteur moyenne  $\mathbb{E}[Y] = \mu$

- Une **composante linéaire** :

$$\eta = X\theta \text{ avec } \theta \in \mathbb{R}^k$$

- Une **relation liant**  $\mu$  et  $\eta$  :  $g(\mu) = \eta$

choix d'une fonction de lien  $g$

# Exemples

Loi de proba.	Param. naturel $\omega$	Moyenne $\mu$	Dispersion $\gamma(\phi)$	Fonction de lien canonique $g$
Gaussienne $\mathcal{N}(\mu, \sigma^2)$	$\mu$	$\mu = \omega$	$\sigma^2$	$g(x) = x$
Bernoulli	$\ln\left(\frac{p}{1-p}\right)$	$p = \frac{e^\omega}{1+e^\omega}$	1	$g(x) = \ln\left(\frac{x}{1-x}\right)$
$\mathcal{B}(p)$				
Poisson $\mathcal{P}(\lambda)$	$\ln(\lambda)$	$\lambda = e^\omega$	1	$g(x) = \ln(x)$
Gamma $\mathcal{G}(\mu, \nu)$	$-\frac{1}{\mu}$	$\mu = -\frac{1}{\omega}$	$\frac{1}{\nu}$	$g(x) = -\frac{1}{x}$
Inverse Gamma $IG(\mu, \sigma^2)$	$-\frac{1}{2\mu^2}$	$\mu = \frac{1}{\sqrt{-2\omega}}$	$\sigma^2$	$g(x) = -\frac{1}{2x^2}$

- La fonction de lien canonique :  $g(\mu) = \omega$
- Rem : on peut choisir une autre fonction de lien que le lien canonique

# Estimation des paramètres

- Comme le paramètre de dispersion  $\phi$  n'apparaît pas dans l'espérance, on va supposer pour simplifier qu'il est fixé (ou estimé préalablement), seul  $\theta$  reste à estimer.
- Pour estimer  $\theta \in \mathbb{R}^k$  :
  - La méthode des moindres carrés n'est pas applicable dans un grand nombre de situations en MLG
  - On utilise la méthode d'estimation du maximum de vraisemblance ( $\hat{\theta}_{MV}$ )

# Estimation des paramètres

- On estime  $\theta$  par maximum de vraisemblance mais expression souvent non explicite  
     $\Rightarrow$  algorithmes de Newton-Raphson ou de Fisher-scoring
- Propriété : Sous certaines conditions de régularité de la densité de probabilité,  $\hat{\theta}_{MV}$  vérifie les propriétés suivantes :

- $\mathcal{I}_n(\hat{\theta}_{MV})^{1/2}(\hat{\theta}_{MV} - \theta) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0_k, I_k)$

$\mathcal{I}(\theta)$  = Information de Fisher

- La **statistique de Wald**  $\mathcal{W}$  vérifie

$$\mathcal{W} := (\hat{\theta}_{MV} - \theta)' \mathcal{I}_n(\theta) (\hat{\theta}_{MV} - \theta) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \chi^2(k).$$

# Test de nullité d'un paramètre : Z-test

- Hypothèses :  $\mathcal{H}_0 : \theta_j = 0$  contre  $\mathcal{H}_1 : \theta_j \neq 0$
- On estime  $\theta_j$  par  $(\hat{\theta}_{MV})_j$
- Loi de  $(\hat{\theta}_{MV})_j$  sous  $\mathcal{H}_0$  :

$$T_j := \frac{(\hat{\theta}_{MV})_j}{\sqrt{[\mathcal{I}_n(\hat{\theta}_{MV})^{-1}]_{jj}}} \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, 1).$$

- Zone de rejet : On va donc rejeter  $\mathcal{H}_0$  si  $\mathcal{R}_\alpha = \left\{ |T_j| > z_{1-\alpha/2} \right\}$  où  $z_{1-\alpha/2}$  est le  $1 - \alpha/2$  quantile de la loi  $\mathcal{N}(0, 1)$ .
- Pvalue asymptotique :

$$pvalue = \mathbb{P}_{\mathcal{H}_0} \left( |T_j| > |T_j^{obs}| \right) = 2(1 - \Phi(|T_j^{obs}|))$$

# Test de sous-modèle

- Soit  $(M_0)$  un sous-modèle de  $(M_1)$

$$(M_0) : g(\mathbb{E}[Y_i | \mathbf{x}_i]) = Z\beta \quad (M_1) : g(\mathbb{E}[Y_i | \mathbf{x}_i]) = X\theta$$

$$Im(Z) \subset Im(X) \quad k_0 = \dim(Im(Z)) < k_1 = \dim(Im(X))$$

- Statistique du test du rapport de vraisemblance :

$$T = -2 \ln \left[ \frac{L(\underline{Y}; \hat{\theta}_0)}{L(\underline{Y}; \hat{\theta}_1)} \right] = -2 \left[ I(\underline{Y}; \hat{\theta}_0) - I(\underline{Y}; \hat{\theta}_1) \right]$$

où  $\hat{\theta}_0$  et  $\hat{\theta}_1$  sont les EMV de  $\theta$  dans le modèle  $M_0$  et  $M_1$  resp.

- Sous  $H_0$ , sous certaines conditions, on peut montrer que

$$T \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \chi^2(k_1 - k_0)$$

- La zone de rejet :  $\mathcal{R}_\alpha = \{T > v_{1-\alpha, k_1 - k_0}\}$  où  $v_{1-\alpha, k_1 - k_0}$  est le  $(1 - \alpha)$ - quantile de la loi du  $\chi^2(k_1 - k_0)$ .

# Test de sous-modèle

- La **déviance** d'un modèle  $M$  est l'écart entre la log-vraisemblance du modèle  $M$  et celle du modèle le plus complet possible  $M_{sat}$  (modèle saturé -  $n$  paramètres)

$$\mathcal{D}(M) = -2 \left[ I(\underline{Y}; \hat{\theta}) - I(\underline{Y}; \hat{\theta}_{sat}) \right].$$

- La statistique de test  $T$  peut donc se réécrire avec la déviance :

$$T = \mathcal{D}(M_0) - \mathcal{D}(M_1) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \chi^2(k_1 - k_0)$$

# Pseudo- $R^2$

- En modèle linéaire :

$$R^2 = 1 - \frac{SSR}{SST}$$

- Par analogie, on définit le **pseudo-** $R^2$

$$pseudoR^2 = \frac{\mathcal{D}(M_0) - \mathcal{D}(M)}{\mathcal{D}(M_0)} = 1 - \frac{\mathcal{D}(M)}{\mathcal{D}(M_0)}$$

où  $\mathcal{D}(M_0)$  est la déviance du modèle nul  $M_0$  (juste un intercept).

- Ce pseudo- $R^2 \in [0, 1]$ . Plus il est proche de 1, meilleur est l'ajustement du modèle.

## Subsection 2

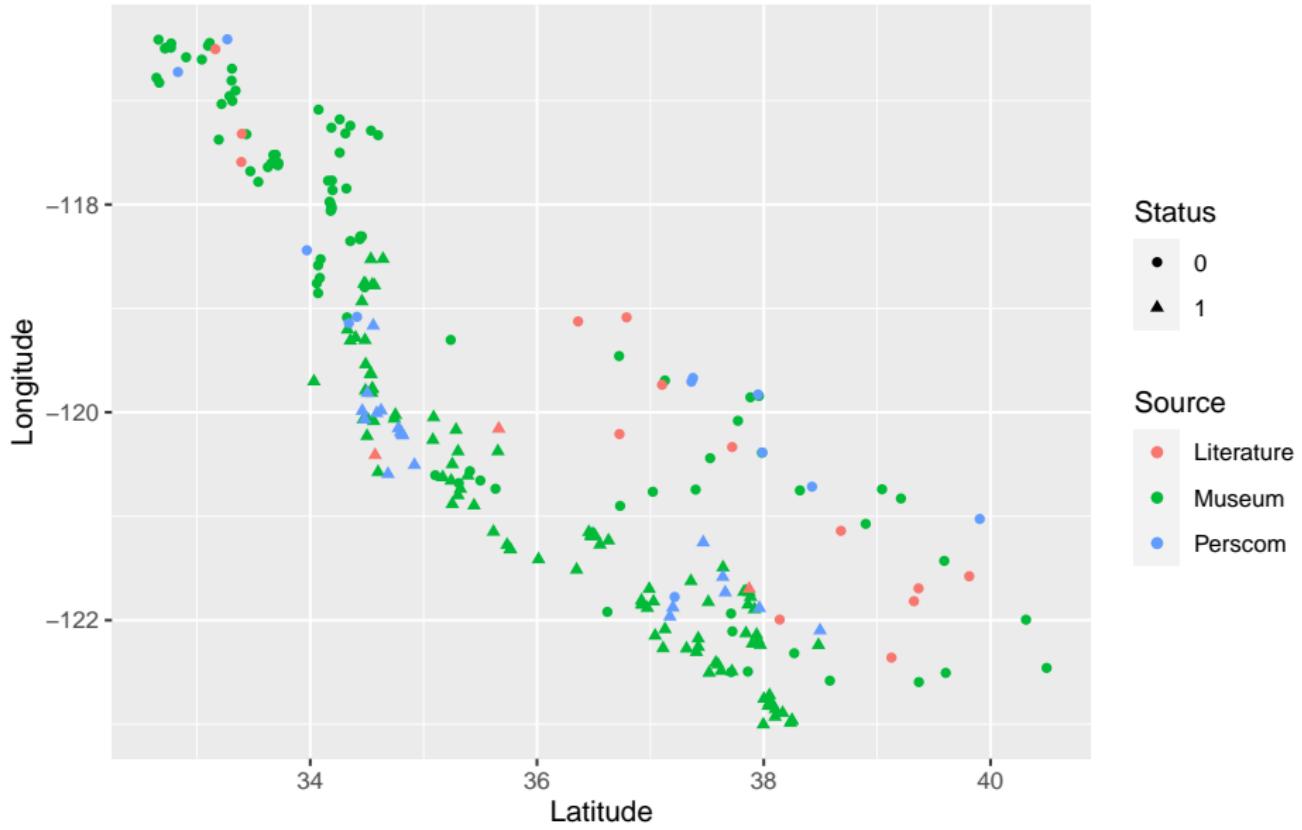
### Régression logistique

# Exemple

On étudie un jeu de données issu d'une collecte d'informations dans différentes sources concernant la présence ou l'absence de grenouilles sauvages dans 236 points d'eau de Californie. Pour chacun des sites, on dispose de sa position (longitude et latitude), de la source de l'information (Museum, Literature, PersCom) ainsi que de l'information sur la présence/absence de grenouilles.

Status	Latitude	Longitude	Source
0:113	Min. :32.64	Min. :-123.0	Literature: 17
1:123	1st Qu.:34.39	1st Qu.:-121.8	Museum :188
	Median :35.47	Median :-120.4	Perscom : 31
	Mean :35.92	Mean :-120.1	
	3rd Qu.:37.72	3rd Qu.:-118.7	
	Max. :40.49	Max. :-116.4	

# Exemple



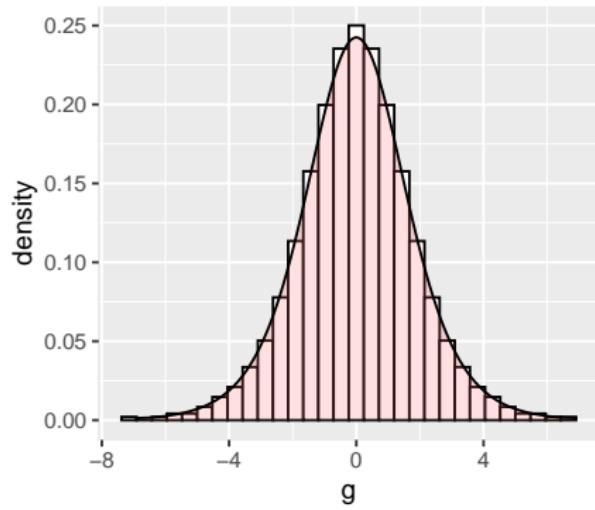
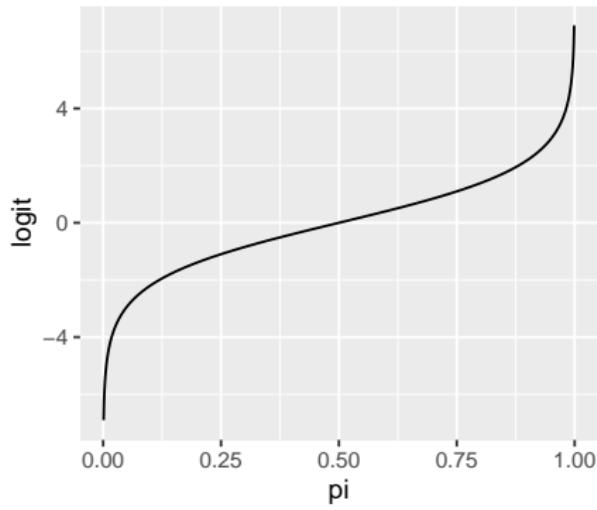
# Contexte

- Variable réponse  $Y$  : binaire
- Variables explicatives :  $\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(p)}$
- Exemple des grenouilles
  - Variable réponse binaire *Status* qui indique si présence / absence de grenouilles
  - On considère  $n = 236$  points d'eau et l'on souhaite expliquer la variable *Status* à l'aide des 3 variables suivantes :
    - Latitude (quantitative)
    - Longitude (quantitative)
    - Source (qualitative)

# Modélisation en régression logistique

- Choix de la loi de Bernoulli :  $Y_i \sim \mathcal{B}(\pi(\mathbf{x}_i))$
- Choix de la fonction de lien : la fonction logit

$$g(\pi(\mathbf{x}_i)) = \text{logit}(\mathbf{x}_i) = \ln \left( \frac{\pi(\mathbf{x}_i)}{1 - \pi(\mathbf{x}_i)} \right)$$



# Modélisation en régression logistique

- Choix de la loi de Bernoulli :  $Y_i \sim \mathcal{B}(\pi(\mathbf{x}_i))$
- Choix de la fonction de lien : la fonction logit

$$g(\pi(\mathbf{x}_i)) = \text{logit}(\mathbf{x}_i) = \ln \left( \frac{\pi(\mathbf{x}_i)}{1 - \pi(\mathbf{x}_i)} \right)$$

- Composante linéaire :  $g(\pi(\mathbf{x}_i)) = X_i\theta$
- Dans notre exemple :

$$\begin{aligned} g(\pi(\mathbf{x}_i)) &= X_i\theta \\ &= \theta_0 + \theta_1 \text{latit}_i + \theta_2 \text{long}_i \\ &\quad + \theta_3 \mathbb{1}_{\text{Source}_i=\text{Museum}} + \theta_4 \mathbb{1}_{\text{Source}_i=\text{Perscom}} \end{aligned}$$

# Exemple

```
model1=glm(Status~.,data=grenouille,family=binomial)
summary(model1)
```

Call:

```
glm(formula = Status ~ ., family = binomial, data = grenouille)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.6157	-0.2974	0.2511	0.4361	2.0621

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-261.3839	32.3230	-8.087	6.13e-16 ***
Latitude	-2.0587	0.3261	-6.314	2.72e-10 ***
Longitude	-2.7859	0.3574	-7.795	6.44e-15 ***
SourceMuseum	0.5069	0.9188	0.552	0.581
SourcePerscom	0.9491	1.0575	0.898	0.369

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 326.74 on 235 degrees of freedom

Residual deviance: 145.89 on 231 degrees of freedom

AIC: 155.89

Number of Fisher Scoring iterations: 6

# Odds et Odds Ratio

- Les odds ratio servent à mesurer l'effet d'une variable quantitative ou le contraste entre les effets d'une variable qualitative.
- L'**odds** (chance) pour un individu  $\mathbf{x}$  d'obtenir la réponse  $Y = 1$  :

$$\text{odds}(\mathbf{x}) = \exp[\text{logit}(\pi(\mathbf{x}))] = \frac{\pi(\mathbf{x})}{1 - \pi(\mathbf{x})}, \quad \text{avec } \pi(\mathbf{x}) = \mathbb{P}(Y = 1|\mathbf{x}).$$

- L'**odds ratio** (rapport de chances) entre deux individus  $\mathbf{x}$  et  $\tilde{\mathbf{x}}$  :

$$\text{OR}(\mathbf{x}, \tilde{\mathbf{x}}) = \frac{\text{odds}(\mathbf{x})}{\text{odds}(\tilde{\mathbf{x}})}.$$

# Odds et Odds Ratio

Les odds ratio peuvent être utilisés de plusieurs manières :

- Comparaison de probabilités de succès entre deux individus :

$$\begin{cases} \text{OR}(\mathbf{x}, \tilde{\mathbf{x}}) > 1 & \Leftrightarrow \pi(\mathbf{x}) > \pi(\tilde{\mathbf{x}}) \\ \text{OR}(\mathbf{x}, \tilde{\mathbf{x}}) = 1 & \Leftrightarrow \pi(\mathbf{x}) = \pi(\tilde{\mathbf{x}}) \\ \text{OR}(\mathbf{x}, \tilde{\mathbf{x}}) < 1 & \Leftrightarrow \pi(\mathbf{x}) < \pi(\tilde{\mathbf{x}}) \end{cases}$$

- Mesure de l'impact d'une variable :

$$\text{OR}(\mathbf{x}, \tilde{\mathbf{x}}) = \prod_{j=1}^p \exp[\theta_j(x^{(j)} - \tilde{x}^{(j)})].$$

Si deux individus diffèrent uniquement sur la  $j$ ème variable :

$$\text{OR}(\mathbf{x}, \tilde{\mathbf{x}}) = \exp[\theta_j(x^{(j)} - \tilde{x}^{(j)})].$$

## Exemple

$$g(\pi(x_i)) = \theta_0 + \theta_1 latit_i + \theta_2 long_i + \theta_3 \mathbb{1}_{Source_i=Museum} + \theta_4 \mathbb{1}_{Source_i=Perscom}$$

- Deux indiv.  $\mathbf{x}$  et  $\tilde{\mathbf{x}}$  qui ne diffèrent que pour *latitude* d'une unité

$$OR(\mathbf{x}, \tilde{\mathbf{x}}) = \exp[\theta_1(latit(\mathbf{x}) - latit(\tilde{\mathbf{x}}))] = \exp[\theta_1]$$

donc on a  $e^{\hat{\theta}_1} = 0.13$  plus de chance d'avoir des grenouilles quand on augmente la latitude d'une unité.

- Deux indiv.  $\mathbf{x}$  et  $\tilde{\mathbf{x}}$  qui ne diffèrent que pour *Source* ( $Source(\mathbf{x}) = \text{Museum}$  et  $Source(\tilde{\mathbf{x}}) = \text{Literature}$ )

$$OR(\mathbf{x}, \tilde{\mathbf{x}}) = \exp[\theta_3]$$

donc on a  $e^{\hat{\theta}_3} = 1.66$  fois plus de chance d'avoir des grenouilles quand la source est "Museum" que "Literature".

# Prédiction

- $\hat{\theta}_{MV}$  estimateur pour  $\theta$
- Probabilité estimée :  $\hat{\pi}(\mathbf{x}) = g^{-1}(\mathbf{x}\hat{\theta}_{MV}) = \frac{\exp(\mathbf{x}\hat{\theta}_{MV})}{1+\exp(\mathbf{x}\hat{\theta}_{MV})}$
- Prédiction :

$$\hat{Y} = \begin{cases} 1 & \text{si } \hat{\pi}(\mathbf{x}) > s \\ 0 & \text{sinon.} \end{cases}$$

Par défaut, on choisit souvent le seuil  $s = 0.5$ .

# Exemple

```
xnew<-grenouille[c(1,50,197,200,220,230),]  
pihat = predict.glm(model1,xnew,type="response")  
Yhat=pihat>0.5
```

	Status	Latitude	Longitude	Source	pihat	Yhat
1	0	37.88222	-119.855	Museum	0.007	FALSE
50	1	38.25030	-122.964	Museum	0.950	TRUE
190	0	39.81222	-121.577	Literature	0.948	TRUE
200	0	33.39652	-117.320	Literature	0.036	FALSE
210	0	33.97012	-118.440	Perscom	0.556	TRUE
230	0	38.42587	-120.715	Perscom	0.038	FALSE

Table de contingence :

	FALSE	TRUE
0	95	18
1	9	114

# Sélection de variables

```
step(model1)
```

```
Start: AIC=155.89  
Status ~ Latitude + Longitude + Source
```

	Df	Deviance	AIC
- Source	2	146.81	152.81
<none>		145.89	155.89
- Latitude	1	233.72	241.72
- Longitude	1	304.37	312.37

```
Step: AIC=152.81  
Status ~ Latitude + Longitude
```

	Df	Deviance	AIC
<none>		146.81	152.81
- Latitude	1	246.85	250.85
- Longitude	1	317.81	321.81

```
Call: glm(formula = Status ~ Latitude + Longitude, family = binomial,  
         data = grenouille)
```

Coefficients:

(Intercept)	Latitude	Longitude
-265.091	-2.116	-2.838

Degrees of Freedom: 235 Total (i.e. Null); 233 Residual

Null Deviance: 326.7

Residual Deviance: 146.8 AIC: 152.8

# Validation par sous-modèle

- Modèles à comparer :

- $(M_2) : \text{logit}(\pi(\mathbf{x}_i)) = \theta_0 + \theta_1 \text{latit}_i + \theta_2 \text{long}_i$

- $(M_1) : \text{logit}(\pi(\mathbf{x}_i)) = \theta_0 + \theta_1 \text{latit}_i + \theta_2 \text{long}_i + \theta_3 \mathbb{1}_{\text{Source}_i=\text{Museum}} + \theta_4 \mathbb{1}_{\text{Source}_i=\text{Perscom}}$

- Statistique de test :  $T = \mathcal{D}(M_2) - \mathcal{D}(M_1) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \chi^2(5 - 3)$

```
model2=glm(Status~Longitude+Latitude, data=grenouille, family=binomial)
anova(model2,model1, test="Chisq")
```

## Analysis of Deviance Table

Model 1: Status ~ Longitude + Latitude						
Model 2: Status ~ Latitude + Longitude + Source						
	Resid.	Df	Resid.	Dev Df	Deviance	Pr(>Chi)
1	233		146.81			
2	231	2	145.89	0.92211	0.6306	

# Extensions

- Variable réponse qualitative  $Y$  pouvant prendre  $M (> 2)$  modalités
- Si les modalités n'ont pas d'ordre  $\Rightarrow$  **régression multinomiale** (ou polytomique non ordonnée)  
ex: `multinom()` du package *nnet*
- Si les modalités ont un ordre  $\Rightarrow$  **régression polytomique ordonnée**  
ex: `polr()` du package *MASS*, `vglm()` du package *VGAM*

### Subsection 3

## Régression loglinéaire

# Exemple des limules

On s'intéresse à l'étude d'une population de limules vivant sur la côte Est des Etats-Unis. Jane Brockmann a réuni une base de données afin d'étudier la reproduction des limules.

Généralement une femelle est entourée d'un mâle et autour de ce couple gravitent d'autres mâles appelés satellites. Le jeu de données est composé de cinq variables :

- “couleur” : clair moyen, moyen, foncé moyen, foncé.
- “spine” : both good, one worm/broken, both worn/broken
- “width” correspond à la largeur de la carapace en cm
- “weight” correspond au poids en grammes
- “satellite” donne le nombre de satellites.

		couleur	spine	width	weight
clair moyen:	12	clair moyen	both good	: 37	Min. :21.0
foncé	:22	foncé moyen	both worn/broken	:121	1st Qu.:24.9
moyen	:44	moyen	one worm/broken	: 15	Median :26.1
	:95				Mean :26.3
					3rd Qu.:27.7
					Max. :33.5
					Min. :1200
					1st Qu.:2000
					Median :2350
					Mean :2437
					3rd Qu.:2850
					Max. :5200
		satellite			
Min.	:	0.000			
1st Qu.:	0.000				
Median :	2.000				
Mean :	2.919				
3rd Qu.:	5.000				
Max. :	15.000				

# Modélisation

- On souhaite expliquer le nombre de satellites en fonction des autres variables
- Modélisation :

- Choix de la loi de Poisson :

$$\forall i = 1, \dots, n, Y_i \sim \mathcal{P}(\lambda_i) \text{ où } \mathbb{E}[Y_i | \mathbf{x}_i] = \lambda_i$$

- Fonction de lien :  $g(.) = \ln(.)$

- Relation linéaire :

$$\ln[\lambda_i] = \theta_0 + \theta_6 width_i + \theta_7 weight_i$$

$$+ \theta_1 \mathbb{1}_{\text{couleur}_i = \text{"foncé}} + \theta_2 \mathbb{1}_{\text{couleur}_i = \text{"foncé moyen}} + \theta_3 \mathbb{1}_{\text{couleur}_i = \text{"moy}} \\ + \theta_4 \mathbb{1}_{\text{spine}_i = \text{"bothworm/broken}}} + \theta_5 \mathbb{1}_{\text{spine}_i = \text{"oneworm/broken"}}$$

# Exemple

```
modelcomplet=glm(satellite ~ ., data=limules, family=poisson(link=log))
summary(modelcomplet)
```

Call:

```
glm(formula = satellite ~ ., family = poisson(link = log), data = limules)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-3.0290	-1.8630	-0.5988	0.9331	4.9446

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-0.3618003	0.9665506	-0.374	0.70817
couleurfoncé	-0.5308601	0.2269157	-2.339	0.01931 *
couleurfoncé moyen	-0.5137051	0.1953624	-2.629	0.00855 **
couleurmoyen	-0.2648512	0.1681107	-1.575	0.11515
spineboth worn/broken	0.0872826	0.1199287	0.728	0.46674
spineone worm/broken	-0.1503718	0.2135754	-0.704	0.48139
width	0.0167487	0.0489197	0.342	0.73207
weight	0.0004965	0.0001663	2.986	0.00283 **

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 632.79 on 172 degrees of freedom  
Residual deviance: 549.59 on 165 degrees of freedom  
AIC: 920.88

Number of Fisher Scoring iterations: 6

# Interprétation des paramètres

- Deux indiv.  $\mathbf{x}$  et  $\tilde{\mathbf{x}}$  qui ne diffèrent que pour *weight* d'une unité

$$\frac{\lambda(\mathbf{x})}{\lambda(\tilde{\mathbf{x}})} = \exp[\theta_7(\text{weight}(\mathbf{x}) - \text{weight}(\tilde{\mathbf{x}}))] = \exp[\theta_7]$$

donc on a  $e^{\hat{\theta}_7} = 1$  fois plus de satellites quand on augmente le poids d'une unité.

- Deux indiv.  $\mathbf{x}$  et  $\tilde{\mathbf{x}}$  qui ne diffèrent que pour *spine* ( $\text{spine}(\mathbf{x}) = \text{bothworm/broken}$  et  $\text{spine}(\tilde{\mathbf{x}}) = \text{bothgood}$ )

$$\frac{\lambda(\mathbf{x})}{\lambda(\tilde{\mathbf{x}})} = \exp[\theta_4]$$

donc on a  $e^{\hat{\theta}_4} = 2.58$  fois plus de satellites en moyenne quand les spines sont toutes les deux abîmées que lorsqu'elles sont en bon état.

# Sélection de variables

```
step(modelcomplet)
```

Start: AIC=920.88

satellite ~ couleur + spine + width + weight

	Df	Deviance	AIC
- spine	2	551.38	918.68
- width	1	549.70	919.00
<none>		549.59	920.88
- couleur	3	558.83	924.12
- weight	1	558.63	927.93

Step: AIC=918.68

satellite ~ couleur + width + weight

	Df	Deviance	AIC
- width	1	551.80	917.10
<none>		551.38	918.68
- couleur	3	559.90	921.20
- weight	1	559.34	924.64

Step: AIC=917.1

satellite ~ couleur + weight

	Df	Deviance	AIC
<none>		551.80	917.10
- couleur	3	560.87	920.16
- weight	1	609.14	972.44

```
Call: glm(formula = satellite ~ couleur + weight, family = poisson(link = log),
          data = limules)
```

# Validation du sous-modèle

```
model1=glm(satellite ~ weight+couleur,data=limules, family=poisson(link=log))
anova(model1,modelcomplet,test="Chisq")
```

Analysis of Deviance Table

```
Model 1: satellite ~ weight + couleur
Model 2: satellite ~ couleur + spine + width + weight
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1       168      551.80
2       165      549.59  3     2.2193   0.5281
```

```
round(model1$coefficients,3)
```

(Intercept)	weight	couleurfoncé	couleurfoncé moyen
-0.050	0.001	-0.452	-0.450
couleurmoyen			
-0.205			

# Prédiction

- $\hat{\theta}_{MV}$  estimateur pour  $\theta$
- Espérance estimée (nb moyen) :

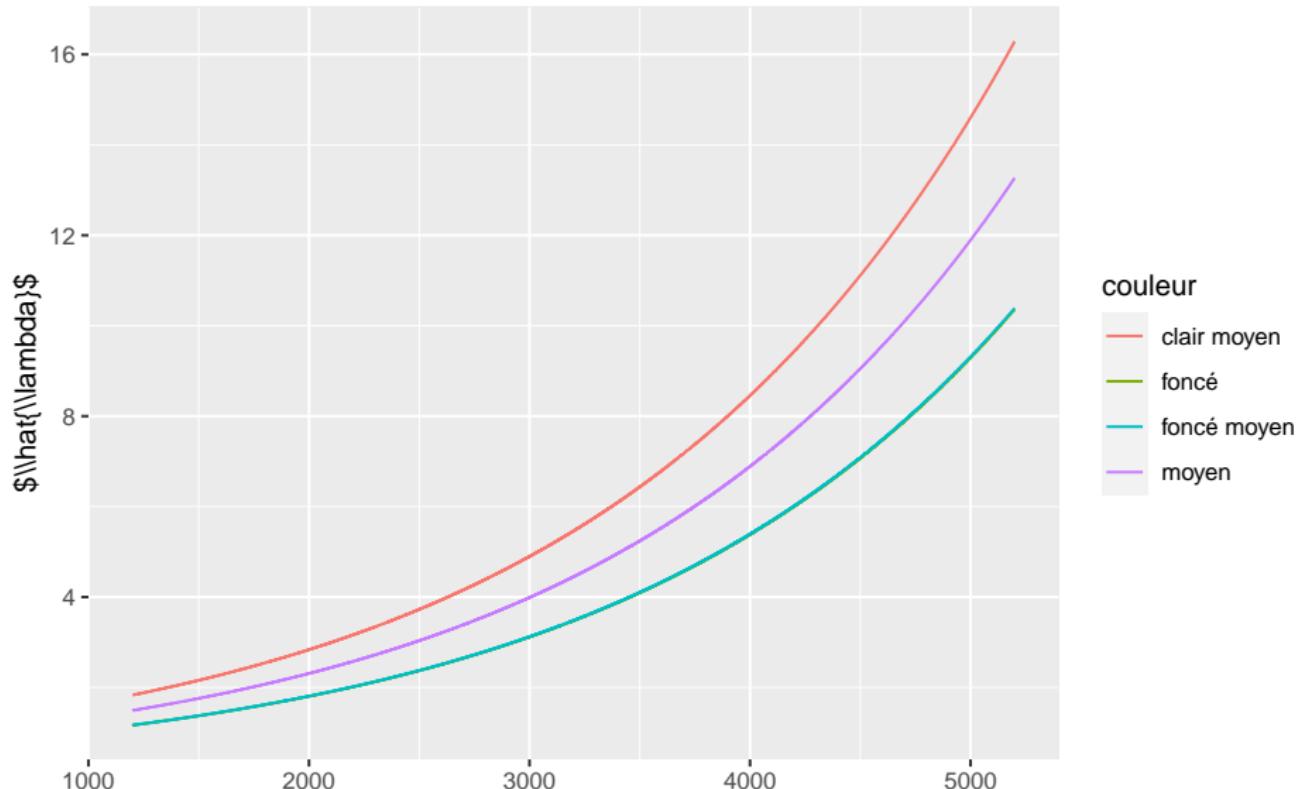
$$\hat{\lambda}(\mathbf{x}) = g^{-1}(\mathbf{x}\hat{\theta}_{MV}) = e^{\mathbf{x}\hat{\theta}_{MV}}$$

- Prédiction :

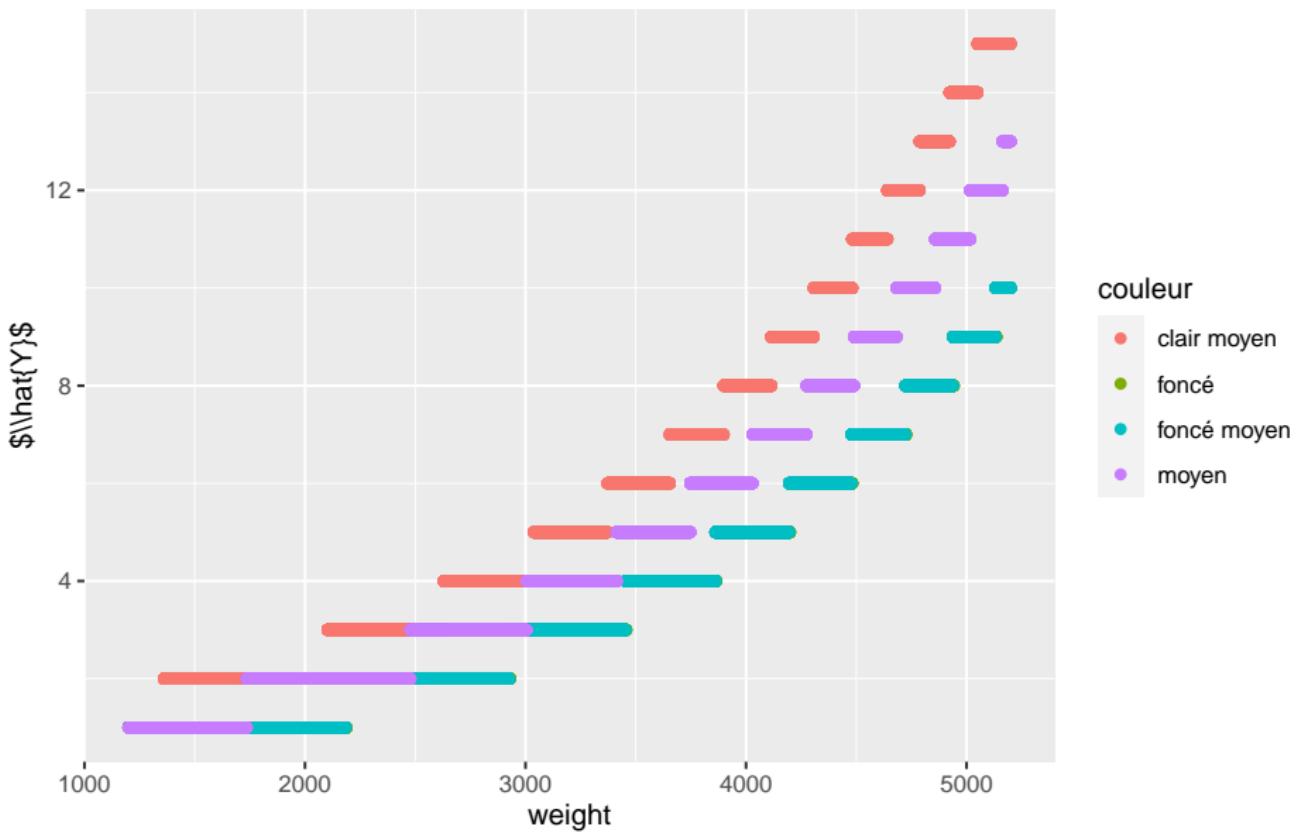
$$\hat{Y}|\mathbf{x} = \underset{k \in \mathbb{N}}{\operatorname{argmax}} \mathbb{P}(U = k) \text{ où } U \sim \mathcal{P}(\hat{\lambda}(\mathbf{x}))$$

# Prédiction

Prédiction avec le sous-modèle (variables *couleur* et *weight*)



# Prédiction



# Sur-dispersion

- Point faible de la loi de Poisson : espérance = variance  
 $\mathbb{E}[Y_i] = \text{Var}(Y_i)$
- Cependant, les données réelles présentent parfois de la **sur-dispersion** ( $\text{Var}(Y_i) > \mathbb{E}[Y_i]$ )
- Extensions possibles :
  - Quasi-poisson :  $\text{Var}(Y_i) = \phi \mathbb{E}[Y_i]$  (Poisson pour  $\phi = 1$ )
  - Binomiale négative :  $\text{Var}(Y_i) = \mathbb{E}[Y_i](1 + \sigma^2 \mathbb{E}[Y_i])$  (Poisson pour  $\sigma^2 = 0$ )
  - ...

# Sur-dispersion

```
model.disp<-glm(satellite ~ weight+couleur,data=limules, family=quasipoisson(link=log))
summary(model.disp)
```

Call:  
glm(formula = satellite ~ weight + couleur, family = quasipoisson(link = log),  
data = limules)

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.9833	-1.9272	-0.5553	0.8646	4.8270

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-0.0497806	0.4160024	-0.120	0.905
weight	0.0005462	0.0001215	4.494	1.3e-05 ***
couleurfoncé	-0.4520498	0.3719136	-1.215	0.226
couleurfoncé moyen	-0.4497964	0.3135639	-1.434	0.153
couleurmoyen	-0.2051125	0.2742671	-0.748	0.456

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for quasipoisson family taken to be 3.183714)

Null deviance: 632.79 on 172 degrees of freedom

Residual deviance: 551.80 on 168 degrees of freedom

AIC: NA

Number of Fisher Scoring iterations: 6