Actuariat de l'Assurance Non-Vie # 4

A. Charpentier (Université de Rennes 1)

ENSAE ParisTech, Octobre / Décembre 2016.

http://freakonometrics.hypotheses.org

Modèles Linéaires Généralisés

Pour la régression linéaire Gaussienne, $Y|\boldsymbol{X}=\boldsymbol{x}\sim\mathcal{N}(\boldsymbol{x}^\top\boldsymbol{\beta},\sigma^2)$

Pour la régression logistique, $Y|\boldsymbol{X}=\boldsymbol{x}\sim\mathcal{B}(H(\boldsymbol{x}^\top\boldsymbol{\beta}))$ avec $H(s)=\dfrac{e^s}{1+e^s}$

Pour la régression de Poisson, $Y|\boldsymbol{X}=\boldsymbol{x}\sim\mathcal{P}(\exp(\boldsymbol{x}^\top\boldsymbol{\beta}))$

Plus généralement $Y|\boldsymbol{X}=\boldsymbol{x}\sim\mathcal{L}(\theta_x,\varphi)$ avec $\theta_x=h(\mathbb{E}[Y|\boldsymbol{X}=\boldsymbol{x}]=g(\boldsymbol{x}^\top\boldsymbol{\beta}))$

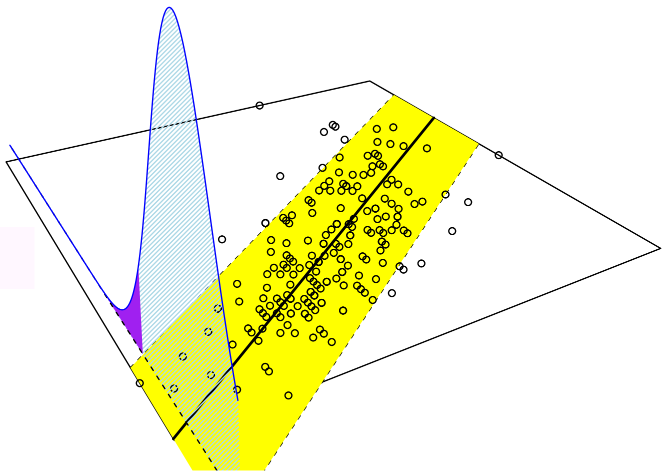
On garde un modèle linéaire, mais on va avoir une généralisation à un grand nombre de lois.

Modèles linéaires, non-linéaires et linéaires généralisés

le modèle linéaire

- $(Y|\boldsymbol{X}=\boldsymbol{x}) \sim \mathcal{N}(\theta_{\boldsymbol{x}}, \sigma^2)$
- $\mathbb{E}[Y|\boldsymbol{X}=\boldsymbol{x}] = \theta_{\boldsymbol{x}} = \boldsymbol{x}^{\mathsf{T}}\boldsymbol{\beta}$

```
> fit <- lm(y ~ x, data = df)
```

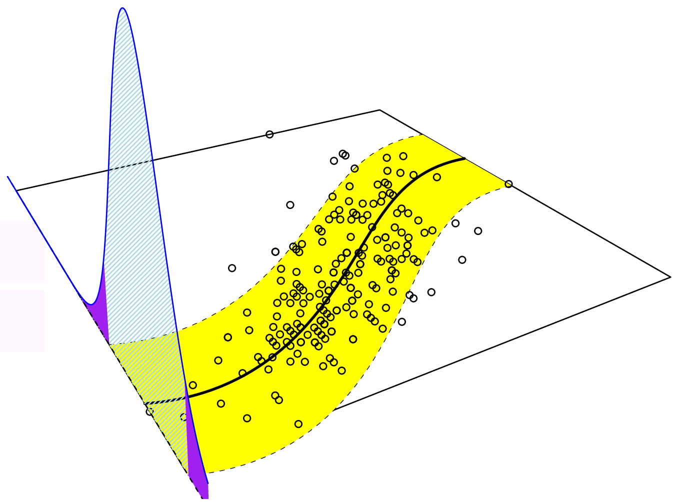


Modèles linéaires, non-linéaires et linéaires généralisés

Les modèles non-linéaires

- $(Y|\boldsymbol{X}=\boldsymbol{x}) \sim \mathcal{N}(\theta_{\boldsymbol{x}}, \sigma^2)$
- $\mathbb{E}[Y|\boldsymbol{X}=\boldsymbol{x}] = \theta_{\boldsymbol{x}} = m(\boldsymbol{x})$

```
> fit <- lm(y ~ poly(x, k), data = df)
> fit <- lm(y ~ bs(x), data = df)
```

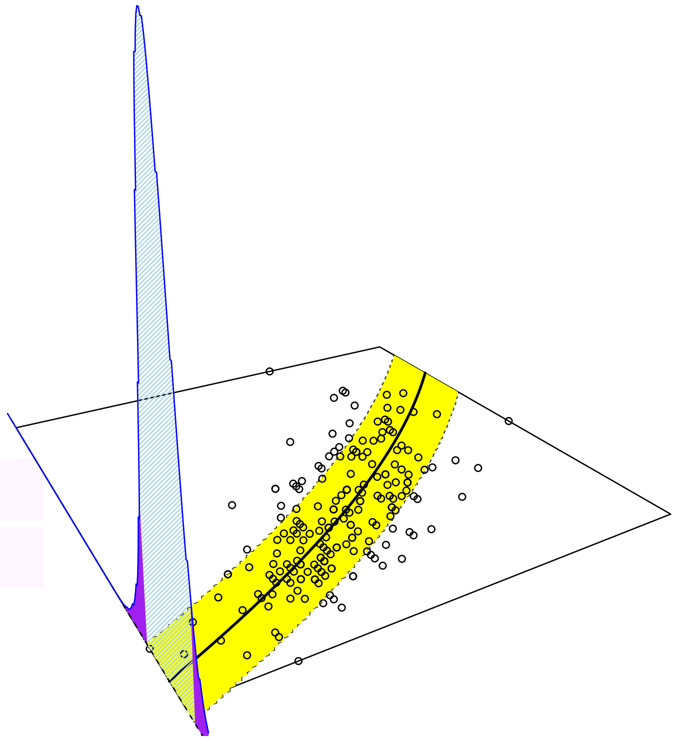


Modèles linéaires, non-linéaires et linéaires généralisés

Les modèles linéaires généralisés

- $(Y|\boldsymbol{X}=\boldsymbol{x}) \sim \mathcal{L}(\theta_{\boldsymbol{x}},\varphi)$
- $\mathbb{E}[Y|\boldsymbol{X}=\boldsymbol{x}] = h^{-1}(\theta_{\boldsymbol{x}}) = \widetilde{h}^{-1}(\boldsymbol{x}^\mathsf{T}\boldsymbol{\beta})$

```
> fit <- glm(y ~ x, data = df,
+ family = poisson(link = "log")
```



La famille exponentielle

Références: Frees (2010), chapitre 13, de Jong & Heller (2008), chapitre 5, et Denuit & Charpentier (2005), chapitre 11.

Considérons des lois de paramètres θ (et φ) dont la fonction de densité (par rapport à la mesure dominante adéquate (mesure de comptage sur $\mathbb{N}$ ou mesure de Lebesgue sur $\mathbb{R}$) s'écrit

$$f(y|\theta,\varphi) = \exp\left(\frac{y\theta - b(\theta)}{a(\varphi)} + c(y,\varphi)\right),$$

où $a(\cdot)$, $b(\cdot)$ et $c(\cdot)$ sont des fonctions, et où θ est appelé paramètre naturel. Le paramètre θ est le paramètre d'intérêt tandi que φ est considéré comme un paramètres de nuisance (et supposé connu, dans un premier temps).

La famille exponentielle

Exemple La loi Gaussienne de moyenne μ et de variance σ^2, $\mathcal{N}(\mu,\sigma^2)$ appartient à cette famille, avec $\theta=\mu$, $\varphi=\sigma^2$, $a(\varphi)=\varphi$, $b(\theta)=\theta^2/2$ et

$$c(y,\varphi)=-\frac{1}{2}\left(\frac{y^2}{\sigma^2}+\log(2\pi\sigma^2)\right),\quad y\in\mathbb{R},$$

Exemple La loi de Bernoulli de moyenne π, $\mathcal{B}(\pi)$ correspond au cas $\theta=\log\{p/(1-p)\}$, $a(\varphi)=1$, $b(\theta)=\log(1+\exp(\theta))$, $\varphi=1$ et $c(y,\varphi)=0$.

Exemple La loi binomiale de moyenne $n\pi$, $\mathcal{B}(n,\pi)$ correspond au cas $\theta=\log\{p/(1-p)\}$, $a(\varphi)=1$, $b(\theta)=n\log(1+\exp(\theta))$, $\varphi=1$ et $c(y,\varphi)=\log\binom{n}{y}$.

Exemple La loi de Poisson de moyenne λ, $\mathcal{P}(\lambda)$ appartient à cette famille,

$$f(y|\lambda)=\exp(-\lambda)\frac{\lambda^y}{y!}=\exp\Big(y\log\lambda-\lambda-\log y!\Big),\quad y\in\mathbb{N},$$

avec $\theta=\log\lambda$, $\varphi=1$, $a(\varphi)=1$, $b(\theta)=\exp\theta=\lambda$ et $c(y,\varphi)=-\log y!$.

La famille exponentielle

Exemple La loi Binomiale Négative, de paramètres r et p,

$$f(k|r,p) = \binom{y+r-1}{y}(1-p)^r p^y, \quad y \in \mathbb{N}.$$

que l'on peut écrire

$$f(k|r,p) = \exp\left(y\log p + r\log(1-p) + \log\binom{y+r-1}{y}\right)$$

soit $\theta = \log p$, $b(\theta) = -r\log p$ et $a(\varphi) = 1$

On reviendra sur cette loi dans la prochaine section du cours.

La famille exponentielle

Exemple La loi Gamma (incluant la loi exponentielle) de moyenne μ et de variance ν^{-1},

$$f(y|\mu,\nu)=\frac{1}{\Gamma(\nu)}\left(\frac{\nu}{\mu}\right)^{\nu}y^{\nu-1}\exp\left(-\frac{\nu}{\mu}y\right),\quad y\in\mathbb{R}_+,$$

est également dans la famille exponentielle. Il faut choisir $\theta=-\frac{1}{\mu}$, $a(\varphi)=\varphi$, $b(\theta)=-\log(-\theta)$, $\varphi=\nu^{-1}$ et

$$c(y,\varphi)=\left(\frac{1}{\varphi}-1\right)\log(y)-\log\left(\Gamma\left(\frac{1}{\varphi}\right)\right)$$

On reviendra sur cette loi dans la section du cours sur la modélisation des coûts de sinistres.

Espérance et variance

Pour une variable aléatoire Y dont la densité est de la forme exponentielle, alors

$$\mathbb{E}(Y) = b'(\theta) \text{ et } \mathrm{Var}(Y) = b''(\theta)\varphi,$$

i.e. la variance de Y apparaît comme le produit de deux fonctions:

- la première, $b''(\theta)$, qui dépend uniquement du paramètre θ est appelée *fonction variance*,
- la seconde est indépendante de θ et dépend uniquement de φ.

En notant $\mu = \mathbb{E}(Y)$, on voit que le paramètre θ est lié à la moyenne μ. La fonction variance peut donc être définie en fonction de μ , nous la noterons dorénavant

$$\mathrm{V}(\mu) = b''([b']^{-1}(\mu))\varphi.$$

Exemple Dans le cas de la loi normale, $\mathrm{V}(\mu) = 1$, dans le cas de la loi de Poisson, $V(\mu) = \mu$ alors que dans le cas de la loi Gamma, $V(\mu) = \mu^2$.

Espérance et fonction lien

Notons que la fonction variance caractérise complètement la loi de la famille exponentielle. Chacune des lois de la famille exponentielle possède une fonction de lien spécifique, dite *fonction de lien canonique*, permettant de relier l'espérance μ au paramètre naturel (ou canonique) θ. Le lien canonique est tel que $g_\star(\mu) = \theta$. Or, $\mu = b'(\theta)$ donc $g_\star(\cdot) = b'(\cdot)^{-1}$.

Exemple Pour la loi normale, $\theta = \mu$ (`link='identity'`),

Exemple Pour la loi de Poisson, $\theta = \log(\mu)$ (`link='log'`)

Exemple Pour la loi de Bernoulli, $\theta = logit(\mu) = \log \dfrac{\mu}{1-\mu}$, (`link='logit'`)

Exemple Pour la loi Gamma, $\theta = 1/\mu$ (`link='inverse'`)

Espérance et variance

Loi de probabilité	$V(\mu)$
Normale	1
Poisson	μ
Gamma	μ^2
Inverse gaussienne	μ^3
Binomiale	$\mu(1-\mu)$

Espérance et variance, la famille Tweedie

Tweedie (1984) a suggéré la famille suivante

$$f(y|\mu,\varphi) = A(y,\varphi)\cdot\exp\left\{\frac{1}{\varphi}\Big[y\theta(\mu)-\kappa\big(\theta(\mu)\big)\Big]\right\},$$

où

$$\theta(\mu) = \begin{cases}\dfrac{\mu^{1-\gamma}}{1-\gamma} & \gamma\neq 1\\ \log\mu & \gamma=1\end{cases} \quad \text{et} \quad \kappa\big(\theta(\mu)\big) = \begin{cases}\dfrac{\mu^{2-\gamma}}{2-\gamma} & \gamma\neq 2\\ \log\mu & \gamma=2\end{cases}$$

La loi de Y est alors une loi Poisson composée, avec des sauts Gamma,

$$Y \sim \mathcal{CP}oi\left(\mu^{2-\gamma}\varphi(2-\gamma),\mathcal{G}\left(-\frac{2-\gamma}{\varphi(1-\gamma)},\varphi(2-\gamma)\mu^{\gamma-1}\right)\right),$$

où $\gamma\in[1,2]$.

Remarque On a une mesure de Dirac en 0 avec distribution (continue) définie sur $\mathbb{R}^+$.

Espérance et variance, la famille Tweedie

On obtient alors une fonction variance de la forme $\text{V}(\mu) = \varphi\mu^\gamma$. On retrouve le modèle de Poisson quand $\gamma \rightarrow 1$ (ou $\alpha \rightarrow \infty$) et une loi Gamma quand $\gamma \rightarrow 2$ (ou $\alpha \rightarrow 0$). Il est en fait possible d'obtenir une classe beaucoup plus large, y compris dans le cas où $\gamma > 2$ en considérant des lois stables.

Paramètre naturel, et lien canonique

Le lien canonique est tel que $g(\mu_i) = \theta_i$. Or, $\mu_i = b'(\theta_i)$ d'où $g^{-1} = b'$.

Loi de probabilité	Fonction de lien canonique
Normale	$\eta = \mu$
Poisson	$\eta = \ln \mu$
Gamma	$\eta = 1/\mu$
Inverse gaussienne	$\eta = 1/\mu^2$
Binomiale	$\eta = \ln \mu - \ln(1-\mu) = \text{logit}(\mu)$

Syntaxe et programmation

La syntaxe des modèles linéaires généralisées ressemble à

```
> glm(Y~X1+X2+X3+offset(log(Z)), family = quasipoisson(link='log'),
+ data = base, weights=w)
```

qui correspond à un modèle

$$\mathbb{E}(Y_i|\boldsymbol{X}_i) = \mu_i = g^{-1}\left(\boldsymbol{X}_i'\beta + \xi_i\right) \text{ et } \text{Var}(Y_i|\boldsymbol{X}_i) = \frac{\varphi V(\mu_i)}{\omega_i}$$

où `Y` est le vecteur des Y_i que l'on cherche à modéliser (le nombre de sinistres de la police i par exemple), `X1`, `X2` et `X3` sont les variables explicatives qui peuvent être qualitatives (on parlera de facteurs) ou quantitatives, `link='log'` indique que g est la fonction log, `family=poisson` revient à choisir une fonction variance V identité, alors que `family=quasipoisson` revient à choisir une fonction variance V identité avec un paramètre de dispersion φ à estimer, `offset` correspond à la variable ξ_i, et `weights` le vecteur $\boldsymbol{\omega} = (\omega_i)$.

Contrainte(s) du logiciel

Il est possible de choisir, en théorie, n'importe quelle fonction de lien (bijective) g telle que $g(\mu) = \eta$. En colonne, la forme de la fonction lien, où $\star$ désigne le lien canonique

Loi de probabilité	μ	μ^{-1}	$\sqrt{\mu}$	$\log \mu$	μ^{-2}	$\text{logit}\mu$	$\varphi^{-1}(\mu)$
Normale	$\star$	$\star$		$\star$			
Poisson	$\star$		$\star$	$\star$			
Gamma	$\star$	$\star$		$\star$			
Inverse gaussienne	$\star$	$\star$		$\star$	$\star$		
Binomiale					$\star$	$\star$	$\star$

Les *quasi*-lois

La loi de Poisson correspondait au cas

$$f(y|\lambda) = \exp(-\lambda)\frac{\lambda^y}{y!} = \exp\Big(y\log\lambda - \lambda - \log y!\Big), \quad y \in \mathbb{N},$$

avec $\theta = \log\lambda$, $\varphi = 1$. On a alors $\text{Var}(Y) = \mathbb{E}(Y)$.

On souhaitera introduire un paramètre $\varphi \neq 1$, autorisant de la surdispersion ($\varphi > 1$). On parle alors de loi quasi-Poisson (mais ce n'est pas une vraie loi). Avec un tel modèle, on aurait $\text{Var}(Y) = \varphi \times \mathbb{E}(Y)$.

Remarque On reviendra plus longuement sur la modélisation dans le cas surdispersé dans la prochaine section du cours.

Des lois exponentielles aux GLM

Considérons des variables aléatoires indépendantes $Y_1, Y_2, \ldots, Y_n$. La densité de chacune de celles-ci est

$$f(y_i|\theta_i,\varphi) = \exp\left\{\frac{y_i\theta_i - b(\theta_i)}{a(\varphi)} + c(y_i,\varphi)\right\}$$

par rapport à la mesure dominante appropriée (mesure de comptage sur $\mathbb{N}$ ou mesure de Lebesgue sur $\mathbb{R}$). Dès lors, la vraisemblance est

$$\mathcal{L}(\boldsymbol{\theta},\varphi|\boldsymbol{y}) = \prod_{i=1}^n f(y_i|\theta_i,\varphi) = \exp\left\{\frac{\sum_{i=1}^n y_i\theta_i - \sum_{i=1}^n b(\theta_i)}{a(\varphi)} + \sum_{i=1}^n c(y_i,\varphi)\right\}.$$

On suppose que les θ_i sont fonction d'un ensemble de p paramètres $\beta_1, \beta_2, \ldots, \beta_p$, disons.

Des lois exponentielles aux GLM

Plus précisément, notant μ_i la moyenne de Y_i, on suppose que

$$g(\mu_i) = \boldsymbol{x}_i'\boldsymbol{\beta} = \eta_i$$

où

1. la fonction monotone et dérivable g est appelée fonction de lien;
2. le vecteur $\boldsymbol{x}_i$ de dimension $p \times 1$ contient des variables explicatives relatives à l'individu i;
3. le vecteur $\boldsymbol{\beta}$ de dimension $p \times 1$ contient les paramètres.

Des lois exponentielles aux GLM

Ainsi, un modèle linéaire généralisé est composé de trois éléments, à savoir

(i) de variables à expliquer $Y_1, Y_2, \ldots, Y_n$ dont la loi est dans la famille exponentielle

(ii) d'un ensemble de paramètres $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_p)'$ appartenant à un ouvert non vide de $\mathbb{R}^p$ et des variables explicatives $\boldsymbol{X} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n)'$: la matrice $n \times p$ $\boldsymbol{X}$, appelée matrice *design*, ou matrice du plan d'expérience, est supposée être de rang p, i.e. la matrice carrée $p \times p$ $\boldsymbol{X}'\boldsymbol{X}$ est inversible;

(iii) d'une fonction de lien g telle que

$$g(\mu_i) = \boldsymbol{x}_i'\boldsymbol{\beta} \text{ où } \mu_i = \mathbb{E}[Y_i]$$

qui lie le prédicteur linéaire $\eta_i = \boldsymbol{x}_i'\boldsymbol{\beta}$ à la moyenne μ_i de Y_i.

Lien, et loi

On supposera que, conditionnellement aux variables explicatives $\boldsymbol{X}$, les variables Y sont indépendantes et identiquement distribuées. En particulier, on partira d'un modèle de la forme

$$f(y_i|\theta_i,\varphi) = \exp\left(\frac{y_i\theta_i - b(\theta_i)}{a(\varphi)} + c(y_i,\varphi)\right),$$

où l'on supposera que

$$g(\mu_i) = \eta_i = \boldsymbol{X}_i'\beta,$$

pour une fonction de lien $g(\cdot)$ donnée (on gardera ainsi un score *linéaire* en les variables explicatives), et où, pour rappel,

$$\mu_i = \mathbb{E}(Y_i|\boldsymbol{X}_i).$$

La fonction lien est la fonction qui permet de lier les variables explicatives $\boldsymbol{X}$ à la prédiction μ, alors que la loi apparaît via la fonction variance, sur la forme de l'hétéroscédasticité et l'incertitude associée à la prédiction. Le petit exemple ci-dessous permet de visualiser sur un petit de données simple six régressions GLM différentes,

```
> 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"))
> regNlog <- glm(y~x,family=gaussian(link="log"))
> regPId <- glm(y~x,family=poisson(link="identity"))
> regPlog <- glm(y~x,family=poisson(link="log"))
> regGId <- glm(y~x,family=Gamma(link="identity"))
> regGlog <- glm(y~x,family=Gamma(link="log"))
```

La prédiction (ainsi qu'un intervalle de confiance) pour chacun de ces modèles est présentée sur la Figure suivante. Le code de base pour obtenir la prédiction avec un intervalle de confiance (à 95%) est simplement

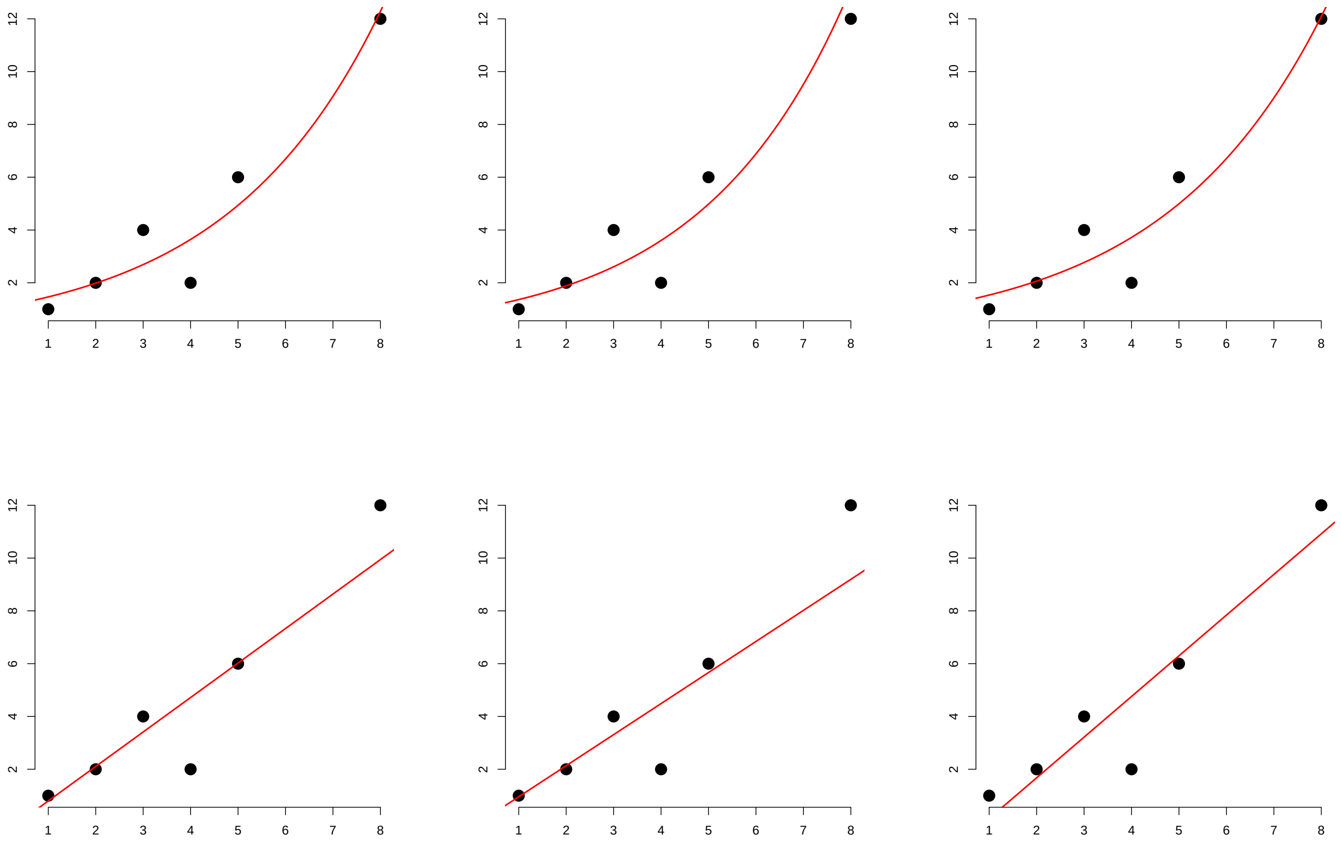
```
> visuel=function(regression,titre){
+ plot(x,y,pch=19,cex=1.5,main=titre,xlab="",ylab="")
+ abs <- seq(0,7,by=.1)
```

```
+ yp <- predict(regression,newdata=data.frame(x=abs),se.fit = TRUE,
+ type="response")
+ polygon(c(abs,rev(abs)),c(yp$fit+2*yp$se.fit,rev(yp$fit-2*yp$se.fit
    )),col="light grey",border=NA)
+ points(x,y,pch=19,cex=1.5)
+ lines(abs,yp$fit,lwd=2)
+ lines(abs,yp$fit+2*yp$se.fit,lty=2)
+ lines(abs,yp$fit-2*yp$se.fit,lty=2)}
```

Pour les 6 modèles ajustés sur le petit jeu de données,

```
> par(mfrow = c(2, 3))
> visuel(regNId,"Gaussienne, lien identite")
> visuel(regPId,"Poisson, lien identite")
> visuel(regGId,"Gamma, lien identite")
> visuel(regNlog,"Gaussienne, lien logarithmique")
> visuel(regPlog,"Poisson, lien logarithmique")
> visuel(regGlog,"Gamma, lien logarithmique")
```

Lien vs. Loi



Estimation des paramètres

La (log)-vraisemblance s'écrit dans le cas des modèles exponentiels,

$$\log \mathcal{L}(\theta_1,\ldots,\theta_n,\varphi,y_1,\ldots,y_n)=\sum_{i=1}^n\left[\frac{y_i\theta_i-b(\theta_i)}{a(\varphi)}+c(y_i,\varphi)\right].$$

On cherche les paramètres $\boldsymbol{\beta}$, il nous suffit de dériver la log-vraisemblance par rapport au paramètre $\boldsymbol{\beta}$ et d'écrire les condition du premier ordre.

Notons $\mu_i=E(Y_i)$ et $\eta_i=g(\mu_i)=X_i\beta$, le prédicteur linéaire.

Pour i et j donnés, on a

$$\frac{\partial \ln(\mathcal{L}_i)}{\partial \beta_j}=\frac{\partial \ln(\mathcal{L}_i)}{\partial \mu_i}\times\frac{\partial \mu_i}{\partial \beta_j}=\frac{\partial \mu_i}{\partial \eta_i}\times\frac{y_i-\mu_i}{\mathrm{V}(Y_i)}X_{ij}.$$

Estimation des paramétres

Ainsi on obtient les équations du score:

$$\sum_i \frac{\partial \ln(\mathcal{L}_i)}{\partial \beta_j} = \sum_i \frac{\partial \mu_i}{\partial \eta_i} \times \frac{y_i - \mu_i}{\text{V}(Y_i)} X_{ij} = 0,$$

pour tout j.

Analytiquement, on ne peut pas résoudre ces équations, mais il est toujours possible de faire une descente de gradient. Ou de reconnaître la condition du premier ordre d'une régression linéaire pondérée

$$\sum_i \frac{\partial \ln(\mathcal{L}_i)}{\partial \beta_j} = \sum_i W_i \times \frac{\partial \eta_i}{\partial \mu_i} \frac{y_i - \mu_i}{\text{V}(Y_i)} X_{ij} = 0, \text{ avec } W_i = \frac{1}{\text{V}(Y_i)} \left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2$$

pour tout j.

Estimation des paramètres

L'algorithme est le même que celui utilisé dans la régression logistique, et Poisson (mais plus général): à partir de $\boldsymbol{\beta}_k$

1. calculer le prédicteur linéaire, $\widehat{\eta}_i$, puis $\widehat{\mu}_i$
2. caluler la matrice (diagonale) de poids telle que $W^{-1} = V(\widehat{\mu})g'(\widehat{\mu})$
3. utiliser le développement de Taylor à l'ordre 1, $Z = \widehat{\eta} + (Y - \widehat{\mu})g'(\widehat{\mu})$
4. régresser Z sur les $\boldsymbol{X}$, avec des poids, pour obtenir $\boldsymbol{\beta}_{k+1}$

et on itère.

Estimation des paramètres

Par exemple, pour la régression de Poisson, on aurait un algorithme de la forme

```
> modellineaire = lm(Y~X)
> beta=coefficients(modellineaire)
> for(i in 1:101){
+ eta=predict(modellineaire)
+ mu=exp(eta)
+ w=mu
+ z=eta+(Y-mu)/mu
+ modellineaire=lm(z~X,weights=w)
+ beta=coefficients(modellineaire)
+ }
```

Estimation des paramètres

Remarque: comme pour le modèle linéaire, l'estimation de $\boldsymbol{\beta}$ se fait indépendament de φ.

On peut montrer que $\widehat{\boldsymbol{\beta}} \xrightarrow{\mathbb{P}} \boldsymbol{\beta}$ et

$$\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{\mathcal{L}} \mathcal{N}(\mathbf{0}, I(\boldsymbol{\beta})^{-1}).$$

Déviance

Rappelons que la déviance est l'écart entre la log-vraisemblance obtenue en $\boldsymbol{\beta}$, et celle obtenue avec un modèle parfait (dit saturé),

$$D = 2\varphi \times [\log \mathcal{L}(\boldsymbol{Y}) - \log \mathcal{L}(\widehat{\boldsymbol{\mu}})]$$

où $\widehat{\boldsymbol{\mu}} = g^{-1}(\boldsymbol{X}\widehat{\boldsymbol{\beta}})$. On peut aussi définir la scaled deviance,

$$D^\star = \frac{D}{\varphi} = 2 \times [\log \mathcal{L}(\boldsymbol{Y}) - \log \mathcal{L}(\widehat{\boldsymbol{\mu}})]$$

Loi de probabilité	Déviance $D^\star$
Normale	$\sum_{i=1}^n \omega_i (y_i - \mu_i)^2$
Poisson	$2\sum_{i=1}^n \omega_i \left\{ y_i \ln \frac{y_i}{\mu_i} - (y_i - \mu_i) \right\}$
Gamma	$2\sum_{i=1}^n \omega_i \left\{ -\ln \frac{y_i}{\mu_i} + \frac{y_i - \mu_i}{\mu_i} \right\}$
Inverse gaussienne	$\sum_{i=1}^n \omega_i \frac{(y_i - \mu_i)^2}{y_i \mu_i^2}$
Binomiale	$2\sum_{i=1}^n \omega_i m_i \left\{ y_i \ln \frac{y_i}{\mu_i} + (1 - y_i) \ln \frac{1-y_i}{1-\mu_i} \right\}$

Estimation du paramètre de dispersion φ

Examinons brièvement le problème de l'estimation du paramètre de dispersion φ. Pour ce faire, posons $\sigma^2 = a(\varphi)$. L'estimation de σ^2 est basée sur la déviance donnée par

$$D = \frac{1}{\sigma^2}\left\{\sum_{i=1}^n y_i\widehat{\theta}_i - \sum_{i=1}^n b(\widehat{\theta}_i)\right\}.$$

Comme $\mathbb{E}(D) \approx n - p$, on pourrait estimer σ^2 par

$$\widetilde{\sigma}^2 = \frac{1}{n-p}D.$$

Cet estimateur est toutefois peu utilisé en pratique car il est très instable. Afin d'éviter ces désagréments, on a recours à un développement de Taylor à l'ordre 2 de $L(\boldsymbol{y}|\boldsymbol{y},\sigma^2)$ qui nous donne

$$\widehat{\sigma}^2 = \frac{1}{n-p}(\boldsymbol{y}-\widehat{\boldsymbol{\mu}})'I_n(\widehat{\boldsymbol{\mu}})(\boldsymbol{y}-\widehat{\boldsymbol{\mu}});$$

cette dernière estimation est souvent appelée estimation du χ^2 de Pearson.

Estimation du paramètre de dispersion φ

On peut obtenir ce paramètre à l'aide du code suivant

```
> modelglm = glm(Y~X1+X2,family=Gamma)
> phi = summary(modeglm)$disperson
> sum(residuals(modeglm, type = "pearson")^2)/
+ modeglm$df.residual
```

Les résidus dans les GLM

Les résidus bruts sont $\widehat{\varepsilon}_i = Y_i - \widehat{\mu}_i$. Mais comme les modèles sont hétéroscédasituqes, ils n'ont qu'un intérêt limité.

Les résidus de Pearson sont

$$\widehat{\varepsilon}_{P,i} = \frac{Y_i - \widehat{\mu}_i}{\sqrt{\mathrm{V}(\widehat{\mu}_i)}}$$

Les résidus de déviance sont

$$\widehat{\varepsilon}_{D,i} = \pm\sqrt{d_i},\ \text{où } D = \sum_{i=1}^{n} d_i.$$

Example: pour un modèle Gaussien, $\widehat{\varepsilon}_{P,i} = \widehat{\varepsilon}_{D,i} = Y_i - \widehat{\mu}_i$.

Example: pour un modèle de Poisson, $\widehat{\varepsilon}_{P,i} = \dfrac{Y_i - \widehat{\mu}_i}{\sqrt{\widehat{\mu}_i}}$ et

$$\widehat{\varepsilon}_{D,i} = \pm\sqrt{|Y_i \log[Y_i/\widehat{\mu}_i] - [Y_i - \widehat{\mu}_i]|}.$$

Les résidus dans les GLM

Example: pour un modèle Gamma, $\widehat{\varepsilon}_{P,i} = \dfrac{Y_i - \widehat{\mu}_i}{\widehat{\mu}_i}$ et

$$\widehat{\varepsilon}_{D,i} = \pm\sqrt{|\log[Y_i/\widehat{\mu}_i] - [Y_i - \widehat{\mu}_i]/\widehat{\mu}_i|}.$$

La commande R est relativement simple

```
> modelglm = glm(Y~X1+X2,family=Gamma)
> residus.P = residuals(modelglm, type="pearson")
> residus.D = residuals(modelglm, type="deviance")
```

Lien, loi et lissage