# Generalized Linear Models

Cathy Maugis-Rabusseau

INSA Toulouse / IMT
GMM 116
cathy.maugis@insa-toulouse.fr

2021-2022

1. **Introduction**
2. **Characterization of a GLM**
3. **Estimation**
4. **Asymptotic distribution of MLE**
5. **Tests**
6. **Confidence interval for $\theta_j$**
7. **Goodness of fit**

# Outline

1. Introduction

# Introduction

- Data :
  - $Y = (Y_1, \ldots, Y_n)'$ be the response vector
  - $x^{(1)}, \ldots, x^{(p)}$ be the explanatory variables (quali and/or quanti)

- Rewriting the linear model

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

$$\Longleftrightarrow$$

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

- Goal: Extend to response variables following other probability laws

  Example: binary responses, counts, ...

# Examples

- Example 1:

In a clinical experiment, one seeks to compare two procedures for a given surgical operation. The experiment is being carried out on two hospitals. We therefore have here two factors with two levels: **operating mode** and **hospital**. The response variable corresponds for each patient to the success or failure of the intervention: it is a binary variable.

- Example 2:

An insurer is interested in the number of automobile claims declared during the last ten years. He wishes to study whether this number of claims is linked to the age of the driver, the size of the car, . . . . The number of claims can be modeled by a Poisson law.

# Principle of GLM

- Modeling:

1. A **random component**: $Y$ is a random variable with expectation $\mu$

2. A **linear component** :

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

3. **A link** between $\mu$ and $\eta$ :

$$g(\mu) = \eta \text{ where } g \text{ is a link function}$$

- It is required to review the principle of parameter estimation and therefore the constructions of confidence intervals and tests

# Outline

## 2 Characterization of a GLM

- Probability law of the response $Y$
- Linear predictor
- Link function

# Outline

2. **Characterization of a GLM**
   - Probability law of the response $Y$
   - Linear predictor
   - Link function

# Exponential family of distributions

## Definition

Let $Y$ be a unidimensional random variable. The probability distribution of $Y$ belongs to an **exponential family** if the law of $Y$ is dominated by a measure $\mu$ and if the likelihood of $Y$ relative to the measure $\mu$ can be written as follows:

$$f_Y(y, \omega, \phi) = \exp\left[\frac{y\omega - b(\omega)}{\gamma(\phi)} + c(y, \phi)\right].$$

- This formulation includes most of the usual laws: Gaussian, inverse Gaussian, Gamma, Poisson, Binomial
- $\omega$ is called the **canonical parameter**
- Warning: the reference measure $\mu$ changes from one exponential family to another

# Exponential family of distributions

## Proposition

Let $Y$ be a random variable whose law belongs to the exponential family then

$$\mathbb{E}[Y] = b'(\omega) \text{ and } \mathrm{Var}(Y) = b''(\omega)\gamma(\phi).$$

**Proof:**

1. For $\mathbb{E}[Y]$, calculate $\frac{\partial}{\partial\omega} f_Y(y, \omega, \phi)$ and integrate with respect to $y$
2. For $\mathrm{Var}(Y)$, calculate $\frac{\partial^2}{\partial\omega^2} f_Y(y, \omega, \phi)$ and integrate with respect to $y$
3. For both , use that $\int f_Y(y, \omega, \phi)\ dy = 1$ and $f$ is enough regular to exchange the differentiation and the integration

# Examples in the exponential family

| Distribution | $\omega$ | $b(\omega)$ | $\gamma(\phi)$ | $\mathbb{E}[Y] = b'(\omega)$ | $\mathrm{Var}(Y) = b^{''}(\omega)\gamma(\phi)$ |
|---|---|---|---|---|---|
| Gaussian<br>$\mathcal{N}(\mu, \sigma^2)$ | $\mu$ | $\frac{\omega^2}{2}$ | $\phi = \sigma^2$ | $\mu = \omega$ | $\sigma^2$ |
| Bernoulli<br>$\mathcal{B}(p)$ | $\ln(p/1-p)$ | $\ln(1+e^{\omega})$ | 1 | $p = \frac{e^{\omega}}{1+e^{\omega}}$ | $p(1-p)$ |
| Poisson<br>$\mathcal{P}(\lambda)$ | $\ln(\lambda)$ | $\lambda = e^{\omega}$ | 1 | $\lambda = e^{\omega}$ | $\lambda = e^{\omega}$ |
| Gamma<br>$\mathcal{G}(\mu, \nu)$ | $-\frac{1}{\mu}$ | $-\ln(-\omega)$ | $\frac{1}{\nu}$ | $\mu = -\frac{1}{\omega}$ | $\frac{\mu^2}{\nu}$ |
| Inverse Gaussian<br>$IG(\mu, \sigma^2)$ | $-\frac{1}{2\mu^2}$ | $-\sqrt{-2\omega}$ | $\sigma^2$ | $\mu = (\sqrt{-2\omega})^{-1}$ | $\mu^3\sigma^2$ |

# Outline

## 2 Characterization of a GLM

- Probability law of the response $Y$
- Linear predictor
- Link function

# Linear predictor

- The explanatory variables induce a design matrix **X**
- The **linear predictor** (deterministic component of the model) is the n-dimensional vector defined by

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

- We will come back later to the parameterization for a quantitative or qualitative explanatory variable and to the over-parameterization (and therefore the link between $p$ and $k$)

# Outline

## 2 Characterization of a GLM

- Probability law of the response $Y$
- Linear predictor
- Link function

# Link function

- Functional link between the random and the systematic components:

$$\forall i = 1, \cdots, n, \; g(\mu_i) = \eta_i = (\mathbf{X}\theta)_i \text{ with } \mu_i = \mathbb{E}[Y_i]$$

- $g$ = **link function**, is assumed to be a monotonic differentiable function.

- The link function $g$ is called **canonical link function** if

$$\forall i = 1, \cdots, n, \; g(\mu_i) = \omega_i = (\mathbf{X}\theta)_i.$$

# Examples

- The canonical link function for
  - the Gaussian distribution is the identity function: $\omega_i = \mu_i$
  - the Poisson distribution is the logarithmic function: $\omega_i = \ln(\mu_i)$
  - the Bernoulli distribution is the logit function: $\omega_i = \ln\left(\frac{\mu_i}{1-\mu_i}\right)$
- For a binary response $Y$, we may also consider
  - the **probit** function: $\eta_i = g(\mu_i) = \Phi^{-1}(\mu_i)$ where $\Phi(.)$ is the Normal cumulative distribution function.
  - the complementary log-log function $\eta_i = g(\mu_i) = \ln(-\ln(1-\mu_i))$
- In the sequel, we focus on
  - **the logistic regression** : binary response + logit function
  - **the log-linear model** : Poisson distribution + log link

# Outline

## 3 Estimation

- Maximum likelihood estimation
- Newton-Raphson and Fisher-scoring algorithms
- Likelihood equations

# Estimation of $\theta$

- Since the dispersion parameter $\phi$ is not in the expectation expression, we assume for simplify that it is fixed (or preliminary estimated), only $\theta$ remains to be estimated.

- To estimate $\theta \in \mathbb{R}^k$ :
  - The least-squares estimation is not suitable in several frameworks in GLM.
  - We use the maximum likelihood estimation (MLE)

# Outline

3 **Estimation**

- Maximum likelihood estimation
- Newton-Raphson and Fisher-scoring algorithms
- Likelihood equations

# Maximum likelihood estimation (MLE)

- Log-likelihood: $\theta \mapsto l(Y;\theta)$ with

$$\theta \mapsto l(\underline{y};\theta) = \sum_{i=1}^{n} \ln[f_{Y_i}(y_i;\omega_i)],$$

where $\theta$, $\eta$, $\mu$ and $\omega$ are linked by the model.

- MLE:

$$\widehat{\theta}_{ML} \in \arg\max_{\theta} L(\underline{Y};\theta) = \arg\max_{\theta} l(\underline{Y};\theta)$$

- If $g$ is the canonical link function, we have $\omega_i = \mathbf{x}_i\theta$ and thus

$$l(\underline{y};\theta) = \sum_{i=1}^{n} \frac{y_i\mathbf{x}_i\theta - b(\mathbf{x}_i\theta)}{\gamma(\phi)} + c(y_i,\phi)$$

# Maximum likelihood estimation (MLE)

- In order to determine the MLE, we consider the score

$$S(\underline{Y};\theta) = \left(\frac{\partial}{\partial\theta_1} l(\underline{Y};\theta), \ldots, \frac{\partial}{\partial\theta_k} l(\underline{Y};\theta)\right)'.$$

- The MLE fulfills

$$S(\underline{Y};\widehat{\theta}_{ML}) = 0_k.$$

- In the particular case where $g$ is the canonical link function,

$$\forall j = 1, \ldots, k,\ \frac{\partial}{\partial\theta_j} l(\underline{y};\theta) = \sum_{i=1}^{n} \frac{1}{\gamma(\phi)} x_i^{(j)} [y_i - b'(\mathbf{x}_i\theta)] = 0$$

$$\Leftrightarrow \sum_{i=1}^{n} [y_i - b'(\mathbf{x}_i\theta)] \frac{\mathbf{x}_i}{\gamma(\phi)} = 0_k$$

- This system is very often nonlinear in $\theta$ and there is no analytical formula for this estimator.

# Outline

3 **Estimation**

- Maximum likelihood estimation
- Newton-Raphson and Fisher-scoring algorithms
- Likelihood equations

# Newton-Raphson algorithm

- The Newton-Raphson algorithm is an iterative algorithm based on the Taylor expansion to order 1 of the score
- It uses the Hessian matrix $\mathcal{J}$
- Remark: $\mathcal{J}$ must be invertible and since it depends on $\theta$, this matrix must be updated at each step of this iterative algorithm.

- Initialisation: $u^{(0)}$.
- For any integer $h$

$$u^{(h)} = u^{(h-1)} - [\mathcal{J}^{(h-1)}]^{-1} S(\underline{Y}; u^{(h-1)}). \tag{1}$$

- Stop when

$$|u^{(h)} - u^{(h-1)}| \leq \Delta.$$

- $\hat{\theta}_{ML} = u^{(h)}$.

# Fisher-scoring algorithm

- Fisher-scoring algorithm: We replace the Hessian matrix $\mathcal{J}$ by the Fisher information matrix

$$\mathcal{I}_n(\theta)_{j,\ell} = -\mathbb{E}\left[\frac{\partial^2}{\partial\theta_j\partial\theta_\ell} l(\underline{Y};\theta)\right].$$

- $\mathcal{I}_n(\theta)$ must be invertible

- May help to avoid problems of non-invertibility of the Hessian matrix

# Outline

## 3 Estimation

- Maximum likelihood estimation
- Newton-Raphson and Fisher-scoring algorithms
- Likelihood equations

# Score and Fisher information matrix

## Proposition

Let $S(\underline{Y};\theta) = (S_1, \dots, S_k)'$ with $S_j = \frac{\partial}{\partial\theta_j} l(\underline{Y};\theta)$ be the score. Then

$$\forall j \in \{1, \dots, k\}, S_j = \sum_{i=1}^{n} \frac{(Y_i - \mu_i)x_i^{(j)}}{\mathrm{Var}(Y_i)} \ \frac{\partial\mu_i}{\partial\eta_i} \text{ and } \mathbb{E}[S_j] = 0.$$

## Proposition

The Fisher information matrix can be written $\mathcal{I}_n(\theta) = \mathbf{X}'\mathbf{W}\mathbf{X}$ where $\mathbf{W}$ is a weight diagonal matrix:

$$[\mathbf{W}]_{ii} = \frac{1}{\mathrm{Var}(Y_i)} \left(\frac{\partial\mu_i}{\partial\eta_i}\right)^2.$$

Remark: $\mathcal{I}_n(\theta)$ is the variance-covariance matrix of $S(\underline{Y};\theta)$ thus $(\mathcal{I}_n(\theta))_{j\ell} = \mathbb{E}[S_j S_\ell]$.

# For the canonical link

## Corollary

If $g$ is the canonical link, we have $\eta_i = \omega_i = \mathbf{x}_i\boldsymbol{\theta}$ thus

$$\frac{\partial \mu_i}{\partial \eta_i} = \frac{\partial \mu_i}{\partial \omega_i} = b''(\omega_i) = \frac{Var(Y_i)}{\gamma(\phi)}.$$

Then,

$$S_j = \sum_{i=1}^{n} \frac{(Y_i - \mu_i)}{\gamma(\phi)} x_i^{(j)} \text{ and } W_{ii} = \frac{Var(Y_i)}{\gamma(\phi)^2}.$$

In particular, the Fisher-scoring and Newton-Raphson methods coincide.

# Outline

4 Asymptotic distribution of MLE

# Asymptotic distribution of MLE

**Theorem**

Under some conditions of regularity of the probability density, the MLE fulfills the following properties:

- $\hat{\theta}_{ML}$ converges in probability to $\theta \in \mathbb{R}^k$
- $\hat{\theta}_{ML}$ converges in law to a Gaussian distribution:

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

- The **Wald's statistics** $\mathcal{W}$ satisfies

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

Remark: the matrix $\mathcal{I}_n(\theta)$ is unknown but replacing it by $\mathcal{I}_n(\hat{\theta}_{ML})$, we have

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

# Outline

## 5 Tests

- Sub-model testing
- Wald's test $(C\theta = 0_q)$
- Test for the nullity of $\theta_j$

# Outline

5 **Tests**

- Sub-model testing
- Wald's test $(C\theta = 0_q)$
- Test for the nullity of $\theta_j$

# Context

- The sub-model testing allows to determine whether a subset of explanatory variables is sufficient to explain the response $Y$ as in the linear model framework.
- Let $M_0$ be a sub-model of $M_1$, defined by
    - model $(M_1)$: $g(\mu) = X\theta$
    - sub-model $(M_0)$ : $g(\mu) = Z\beta$ with $Im(Z) \subset Im(X)$
- We want to test if the sub-model $(M_0)$ is sufficient to explain the response variable compared to the more complex model $(M_1)$

# Likelihood ration test

- Test's statistics:

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

  where $\hat{\beta}$ and $\hat{\theta}$ are the MLE in models $M_0$ and $M_1$ respectively.

- Under $H_0$, under some conditions, we may prove that

$$T \underset{n\to+\infty}{\overset{\mathcal{L}}{\to}} \chi^2(k_1 - k_0)$$

  where $k_0 = dim([Z])$ and $k_1 = dim([X])$.

- Reject zone:

$$\mathcal{R}_\alpha = \{T > v_{1-\alpha,k_1-k_0}\}$$

  where $v_{1-\alpha,k_1-k_0}$ is the $(1-\alpha)$- quantile of $\chi^2(k_1 - k_0)$.

- This test is **asymptotic!**

# With the deviance

- The **deviance** of model $M$ is the difference between the log-likelihood of model $M$ and that of the saturated model $M_{sat}$

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

- The test's statistics $T$ can be written with the deviance of the both models:

$$T = \mathcal{D}(M_0) - \mathcal{D}(M_1) \underset{n\to+\infty}{\overset{\mathcal{L}}{\to}} \chi^2(k_1 - k_0)$$

# Outline

5 **Tests**

- Sub-model testing
- Wald's test $(C\theta = 0_q)$
- Test for the nullity of $\theta_j$

# Wald's test for $C\theta = 0_q$

- Let $C \in \mathcal{M}_{qk}(\mathbb{R})$. We want to test

$$\mathcal{H}_0 : C\theta = 0_q \text{ against } \mathcal{H}_1 : C\theta \neq 0_q$$

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

$$(C\hat{\theta}_{ML} - C\theta)' \left[C\mathcal{I}_n(\hat{\theta}_{ML})^{-1}C'\right]^{-1} (C\hat{\theta}_{ML} - C\theta) \xrightarrow[n\to+\infty]{\mathcal{L}} \chi^2(q).$$

- Reject zone:

$$\mathcal{R}_\alpha = \left\{(C\hat{\theta}_{ML})' \left[C\mathcal{I}_n(\hat{\theta}_{ML})^{-1}C'\right]^{-1} (C\hat{\theta}_{ML}) > v_{1-\alpha,q}\right\}$$

where $v_{1-\alpha,q}$ is the $(1-\alpha)$ quantile of $\chi^2(q)$.

# Wald's test for $C\theta = 0_q$

- Since $\mathcal{I}_n(\hat{\theta}) = X'WX$, we have under $\mathcal{H}_0$,

$$(C\widehat{\theta}_{ML})' \left[ C(X'WX)^{-1}C' \right]^{-1} (C\widehat{\theta}_{ML}) \underset{n\to+\infty}{\xrightarrow{\mathcal{L}}} \chi^2(q)$$

- This test is based on $(X'WX)^{-1}$ which generalizes $(X'X)^{-1}$ used in linear model

- The Wald's test and the Fisher's test "are equivalent" in the Gaussian linear model framework.

# Outline

5 **Tests**

- Sub-model testing
- Wald's test $(C\theta = 0_q)$
- Test for the nullity of $\theta_j$

# Test for the nullity of $\theta_j$ - Z-test

- $\mathcal{H}_0 : \theta_j = 0$ against $\mathcal{H}_0 : \theta_j \neq 0$
- $\theta_j$ is estimated by $(\hat{\theta}_{MLE})_j$
- Law of $(\hat{\theta}_{ML})_j$ under $\mathcal{H}_0$: Since

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

for $n$ large enough, $\hat{\theta}_{ML} - \theta \overset{\mathcal{L}}{\simeq} \mathcal{N}_k(0_k, \mathcal{I}_n(\hat{\theta}_{ML})^{-1})$ and

$$(\hat{\theta}_{ML})_j \underset{\mathcal{H}_0}{\overset{\mathcal{L}}{\simeq}} \mathcal{N}\left(0, [\mathcal{I}_n(\hat{\theta}_{ML})^{-1}]_{jj}\right).$$

- Reject zone:

$$\mathcal{R}_\alpha = \left\{ T_j := \left|(\hat{\theta}_{ML})_j\right| / \sqrt{[\mathcal{I}_n(\hat{\theta}_{ML})^{-1}]_{jj}} > z_{1-\alpha/2} \right\}$$

where $z_{1-\alpha/2}$ is the $1-\alpha/2$ quantile of $\mathcal{N}(0,1)$.

# Test for the nullity of $\theta_j$ - Z-test

- Z-test :

$$\mathcal{R}_\alpha = \left\{ T_j := \left|(\hat{\theta}_{ML})_j\right| / \sqrt{[\mathcal{I}_n(\hat{\theta}_{ML})^{-1}]_{jj}} > z_{1-\alpha/2} \right\}$$

- This test is equivalent to Wald's test for $q = 1$ and $C = I$:

$$\mathcal{R}_\alpha = \left\{ \left[(\hat{\theta}_{ML})_j\right]^2 / [\mathcal{I}_n(\hat{\theta}_{ML})^{-1}]_{jj} > v_{1-\alpha,1} \right\}$$

where $v_{1-\alpha,1}$ is the $(1-\alpha)$-quantile of $\chi^2(1)$.

# Outline

## 6 Confidence interval for $\theta_j$

- Wald's construction
- Based on the likelihood ratio

# Outline

## 6 Confidence interval for $\theta_j$

- Wald's construction
- Based on the likelihood ratio

# Using Wald

Since

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

for $n$ large enough, $\hat{\theta}_{ML} - \theta \overset{\mathcal{L}}{\simeq} \mathcal{N}_k(0_k, \mathcal{I}_n(\hat{\theta}_{ML})^{-1})$ and

$$(\hat{\theta}_{ML})_j - \theta_j \overset{\mathcal{L}}{\simeq} \mathcal{N}\left(0, [\mathcal{I}_n(\hat{\theta}_{ML})^{-1}]_{jj}\right).$$

Thus

$$\left[(\hat{\theta}_{ML})_j - \theta_j\right] / \sqrt{[\mathcal{I}_n(\hat{\theta}_{ML})^{-1}]_{jj}} \xrightarrow[n\to+\infty]{\mathcal{L}} \mathcal{N}(0,1).$$

We deduce an **asymptotic** confidence interval for $\theta_j$ with confidence level $1-\alpha$:

$$IC_{1-\alpha}(\theta_j) = \left[(\hat{\theta}_{ML})_j \pm z_{1-\alpha/2}\sqrt{[\mathcal{I}_n(\hat{\theta}_{ML})^{-1}]_{jj}}\right]$$

where $z_{1-\alpha/2}$ is the$1-\alpha/2$ quantile of $\mathcal{N}(0,1)$.

# Outline

## 6 Confidence interval for $\theta_j$

- Wald's construction
- Based on the likelihood ratio

# CI with the likelihood ratio

- La **fonction de vraisemblance profil** de $\theta_j$ est définie par

$$l^\star(\underline{Y};\theta_j) = \max_{\tilde{\theta}} \ l(\underline{Y};\tilde{\theta})$$

où $\tilde{\theta}$ est le vecteur $\theta$ avec le $j$ème élément fixé à $\theta_j$.

- Si $\theta_j$ est la vraie valeur du paramètre alors

$$2\left[l(\underline{Y};\hat{\theta}_{ML}) - l^\star(\underline{Y};\theta_j)\right] \underset{n\to+\infty}{\xrightarrow{\mathcal{L}}} \chi^2(1).$$

- Ainsi, si on considère l'ensemble

$$\mathcal{G} = \left\{u ; 2\left[l(\underline{Y};\hat{\theta}_{ML}) - l^\star(\underline{Y};u)\right] \leq v_{1-\alpha,1}\right\},$$

on obtient que

$$\mathbb{P}(\theta_j \in \mathcal{G}) \underset{n\to+\infty}{\longrightarrow} 1-\alpha.$$

Ainsi $\mathcal{G}$ est un intervalle de confiance asymptotique pour $\theta_j$ au niveau de confiance $1-\alpha$.

# Outline

7 Goodness of fit

# Pseudo $R^2$

- By analogy with the $R^2 = 1 - \frac{SSR}{SST}$ used in the linear model framework, the **pseudo**-$R^2$ is defined by

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

  where $\mathcal{D}(M_0)$ is the deviance of the null model $M_0$.

- The pseudo-$R^2 \in [0, 1]$. The closer it is to 1, the better the fit of the model.

# Summary

## Summary

- Know how to model a GLM by clearly specifying the three elements (random component, linear component, link function)
- Know how to show that a distribution belongs to the exponential family (the definition is not to be known, it will be recalled)
- Know the canonical link function for the Gaussian, Bernoulli and Poisson distributions
- Understand the general spirit to determine the MLE
- Know the theorem on the law of the estimator $\hat{\theta}_{\text{MLE}}$ in GLM
- Know how to build a test of nested models, a Wald's test and a $Z$-test
- Build a confidence interval for $\theta_j$ by Wald
- Know the definition of the pseudo- $R^2$