Some satistical notions needed for biological studies

Vivien Rossi

CIRAD - UMR Ecologie des forêts de Guyane
vivien.rossi@cirad.fr

Master 2 - Ecologie des Forêts Tropicale
AgroParisTech - Université Antilles-Guyane

Kourou, novembre 2010

objectives of the course

Give your own overview of statistical tools though *Generalized Linear Model* framework

1. reminders about some useful stuffs (usual probability distribution, data transformation . . .)
2. reminders about Linear Model
3. insights of the GLM formalism
4. Parameter estimation

outlines

the normal distribution

- names: normal, gaussian, Laplace-Gauss

the normal distribution

- names: normal, gaussian, Laplace-Gauss
- continuous probability distribution, with density

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$$

- the corner stone of the statistic:

the normal distribution

- names: normal, gaussian, Laplace-Gauss
- continuous probability distribution, with density

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$$

- the corner stone of the statistic:

CLT

Let $(x_1, \ldots, x_n)$ iid random variables, having mean μ and variance σ^2 then

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} x_i - \mu\right) \xrightarrow{d} \mathcal{N}(0, \sigma^2)$$

the normal distribution

link with other distributions

- The binomial distribution $\mathcal{B}(n, p) \approx \mathcal{N}(np, np(1-p))$ for large n and for p not too close to zero or one.
- The Poisson distribution $\mathcal{P}(\lambda) \approx \mathcal{N}(\lambda, \lambda)$ for large values of λ.
- The chi-squared distribution $\chi^2(k) \approx \mathcal{N}(k, 2k)$ for large ks.
- The Student distribution $\mathcal{T}(\nu) \approx \mathcal{N}(0, 1)$ when ν is large.

data transformation

logarithm

$$\begin{array}{rcl} \mathbb{R}+ & \rightarrow & \mathbb{R} \\ x & \mapsto & \log x \end{array}$$

sd proportional to the mean or the effects are multiplicatives usual for growth or multiplicative processes if some data are small (<1) it better to use $\log(x+1)$

square root

$$\begin{array}{rcl} \mathbb{R}+ & \rightarrow & \mathbb{R} \\ x & \mapsto & \sqrt{x} \end{array}$$

useful for variance proportional to the mean (count of rare or frequent events expressed by percentages)

data transformation

Arc sinus

$$\begin{array}{rcl} \mathbb{R}+ & \rightarrow & \mathbb{R} \\ x & \mapsto & \arcsin x \end{array}$$

useful for variance proportional to the mean (count of events expressed by percentages)

1. some reminders
2. Linear Model
3. Generalized Linear Model
4. Parameter estimation

Linear model

The simplest link between a response y and some variables $x^1,\dots,x^p$

- formalism

$$y_i = \beta_0 + \beta_1 x_i^1 + \cdots + \beta_p x_i^p + \epsilon_i \quad \text{with } \epsilon_i \sim \mathcal{N}(0, \sigma^2)$$

- other formalism

$$y_i \sim \mathcal{N}(\beta_0 + \beta_1 x_i^1 + \cdots + \beta_p x_i^p, \sigma^2)$$

- matrix formalism

$$y_i = X_i\,\beta + \epsilon_i \text{ with } X_i = (x_i^1, \dots, x_i^p) \text{ and } \beta = (\beta_1, \dots, \beta_p)$$

or

$$Y = X\,\beta + \epsilon \text{ with } Y = (y_1, \dots, y_n),\ X = [x_i^j]_{i=1:n, j=1:p} \text{ and } \epsilon \sim \mathcal{N}_n(0, \sigma^2 I_n)$$

Linear model: underling assumption

$$\underbrace{\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}}_{Y} = \underbrace{\begin{bmatrix} 1 & x_1^1 & \dots & x_1^p \\ \vdots & \vdots & & \vdots \\ 1 & x_n^1 & \dots & x_n^p \end{bmatrix}}_{X} \underbrace{\begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}}_{\beta} + \underbrace{\begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{bmatrix}}_{\epsilon}$$

assumptions

1. error $\mathcal{E}$ is independent from X or X is not random
2. ε_i i.i.d. from random variable $\mathcal{E}$: $E(\mathcal{E}) = 0$ and $V(\mathcal{E}) = \sigma^2$.
3. unknown parameters $\beta_0, \dots, \beta_p$ are constants
4. statistical inference: $\mathcal{E} \sim \mathcal{N}(0, \sigma^2 I_n)$ i.e. $\mathcal{E}_i$ i.i.d according $\mathcal{N}(0, \sigma^2)$.

LM: parameter estimation

- minimize least squares with respect to $\beta \in \mathbb{R}^{p+1}$:

$$\begin{aligned}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i^1 - \beta_2 x_i^2 - \cdots - \beta_p x_i^p)^2 &= \|\boldsymbol{y} - \boldsymbol{X}\beta\|^2 \\ &= (\boldsymbol{y} - \boldsymbol{X}\beta)'(\boldsymbol{y} - \boldsymbol{X}\beta) \\ &= \boldsymbol{y}'\boldsymbol{y} - 2\beta'\boldsymbol{X}'\boldsymbol{y} + \beta'\boldsymbol{X}'\boldsymbol{X}\beta.\end{aligned}$$

LM: parameter estimation

- minimize least squares with respect to $\beta \in \mathbb{R}^{p+1}$:

$$\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i^1 - \beta_2 x_i^2 - \cdots - \beta_p x_i^p)^2 = \|\boldsymbol{y} - \boldsymbol{X}\beta\|^2$$
$$= (\boldsymbol{y} - \boldsymbol{X}\beta)'(\boldsymbol{y} - \boldsymbol{X}\beta)$$
$$= \boldsymbol{y}'\boldsymbol{y} - 2\beta'\boldsymbol{X}'\boldsymbol{y} + \beta'\boldsymbol{X}'\boldsymbol{X}\beta.$$

- a matrix derivation of the last equation leads to the «normal equations» :

$$\boldsymbol{X}'\boldsymbol{y} - \boldsymbol{X}'\boldsymbol{X}\beta = 0$$

  the solution is a minimum because the hessian matrix $2\boldsymbol{X}'\boldsymbol{X}$ is semi-positive.

LM: parameter estimation

- minimize least squares with respect to $\beta \in \mathbb{R}^{p+1}$:

$$\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i^1 - \beta_2 x_i^2 - \cdots - \beta_p x_i^p)^2 = \|\boldsymbol{y} - \boldsymbol{X}\beta\|^2$$
$$= (\boldsymbol{y} - \boldsymbol{X}\beta)'(\boldsymbol{y} - \boldsymbol{X}\beta)$$
$$= \boldsymbol{y}'\boldsymbol{y} - 2\beta'\boldsymbol{X}'\boldsymbol{y} + \beta'\boldsymbol{X}'\boldsymbol{X}\beta.$$

- a matrix derivation of the last equation leads to the «normal equations» :

$$\boldsymbol{X}'\boldsymbol{y} - \boldsymbol{X}'\boldsymbol{X}\beta = 0$$

  the solution is a minimum because the hessian matrix $2\boldsymbol{X}'\boldsymbol{X}$ is semi-positive.

- we assume in addition that the matrix $\boldsymbol{X}'\boldsymbol{X}$ is invertible (the rank of the matrix X is $(p+1)$)

LM: parameter estimation estimation

- hence the estimate of β is:

$$b = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$$

LM: parameter estimation estimation

- hence the estimate of β is:

$$b = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$$

- the estimates of $\boldsymbol{y}$ are:

LM: parameter estimation estimation

- hence the estimate of β is:

$$b = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$$

- the estimates of $\boldsymbol{y}$ are:

$$\hat{\boldsymbol{y}} = \boldsymbol{X}b = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}\boldsymbol{y} = H\boldsymbol{y}$$

- the least squares estimate of β, b has no bias and minimize the variance

LM: parameter estimation estimation

- hence the estimate of β is:

$$b = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$$

- the estimates of $\boldsymbol{y}$ are:

$$\hat{\boldsymbol{y}} = \boldsymbol{X}b = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}\boldsymbol{y} = H\boldsymbol{y}$$

- the least squares estimate of β, b has no bias and minimize the variance
- the covariance matrix of the estimates is

$$E[(b-\beta)(b-\beta)'] = \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1},$$

LM: parameter estimation estimation

- hence the estimate of β is:

$$b = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$$

- the estimates of $\boldsymbol{y}$ are:

$$\hat{\boldsymbol{y}} = \boldsymbol{X}b = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}\boldsymbol{y} = H\boldsymbol{y}$$

- the least squares estimate of β, b has no bias and minimize the variance
- the covariance matrix of the estimates is

$$E[(b-\beta)(b-\beta)'] = \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1},$$

- the estimate of σ^2 is:

$$S^2 = \frac{\|\boldsymbol{y}-\hat{\boldsymbol{y}}\|^2}{n-p-1} = \frac{\|\boldsymbol{y}-\boldsymbol{X}b\|^2}{n-p-1} = \frac{SSE}{n-p-1}.$$

LM: quality of the model fitting

- we look at $(\hat{y}_i - y_i)^2$. To do it efficiently we consider:

LM: quality of the model fitting

- we look at $(\hat{y}_i - y_i)^2$. To do it efficiently we consider:
  - sum of squared errors
$$SSE = \|\mathbf{y} - \hat{\mathbf{y}}\|^2.$$
  - total sum of squares
$$SST = \|\mathbf{y} - \bar{y}_n\|^2 = \mathbf{y}'\mathbf{y} - n\bar{y}_n^2$$
  - regression sum of squares
$$SSR = \|\hat{\mathbf{y}} - \bar{y}_n \mathbb{1}\|^2 = \hat{\mathbf{y}}'\hat{\mathbf{y}} - n\bar{y}_n^2.$$
  - we then verify that: $SST = SSR + SSE$.

LM: quality of the model fitting

- we look at $(\hat{y}_i - y_i)^2$. To do it efficiently we consider:
  - sum of squared errors
  $$SSE = \|\mathbf{y} - \hat{\mathbf{y}}\|^2.$$
  - total sum of squares
  $$SST = \|\mathbf{y} - \bar{y}_n\|^2 = \mathbf{y}'\mathbf{y} - n\bar{y}_n^2$$
  - regression sum of squares
  $$SSR = \|\hat{\mathbf{y}} - \bar{y}_n \mathbb{1}\|^2 = \hat{\mathbf{y}}'\hat{\mathbf{y}} - n\bar{y}_n^2.$$
  - we then verify that: $SST = SSR + SSE$.

Coefficient of determination R^2

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

ratio between the variance explained by the model and the total variance

- R^2 interpretation is quite easy.

LM: quality of the model fitting

- we look at $(\hat{y}_i - y_i)^2$. To do it efficiently we consider:
  - sum of squared errors
  $$SSE = \|\mathbf{y} - \hat{\mathbf{y}}\|^2.$$
  - total sum of squares
  $$SST = \|\mathbf{y} - \bar{y}_n\|^2 = \mathbf{y}'\mathbf{y} - n\bar{y}_n^2$$
  - regression sum of squares
  $$SSR = \|\hat{\mathbf{y}} - \bar{y}_n \mathbb{1}\|^2 = \hat{\mathbf{y}}'\hat{\mathbf{y}} - n\bar{y}_n^2.$$
  - we then verify that: $SST = SSR + SSE$.

Coefficient of determination R^2

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

ratio between the variance explained by the model and the total variance

- R^2 interpretation is quite easy.

LM: model inference

- assuming the errors iid gaussians :

$$b \sim \mathcal{N}\left(\beta, \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1})\right)$$

LM: model inference

- assuming the errors iid gaussians :

$$b \sim \mathcal{N}\left(\beta, \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1})\right)$$

- result not useful practically because σ^2 unknown, but

$$\frac{(n-p-1)S^2}{\sigma^2} \sim \chi^2_{n-p-1}.$$

LM: model inference

- assuming the errors iid gaussians :

$$b \sim \mathcal{N}\left(\beta, \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1})\right)$$

- result not useful practically because σ^2 unknown, but

$$\frac{(n-p-1)S^2}{\sigma^2} \sim \chi^2_{n-p-1}.$$

- we can plug-in S^2 instead of $\sigma^2 \rightarrow$ Student distribution:

$$\frac{b_j - \beta_j}{S\sqrt{c_{jj}}} \sim Std_{n-p-1}$$

with c_{jj} the element (j, j) of the matrix $(\boldsymbol{X}'\boldsymbol{X})^{-1}$.

LM: model inference

- assuming the errors iid gaussians :

$$b \sim \mathcal{N}\left(\beta, \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1})\right)$$

- result not useful practically because σ^2 unknown, but

$$\frac{(n-p-1)S^2}{\sigma^2} \sim \chi^2_{n-p-1}.$$

- we can plug-in S^2 instead of σ^2 $\rightarrow$ Student distribution:

$$\frac{b_j - \beta_j}{S\sqrt{c_{jj}}} \sim Std_{n-p-1}$$

with c_{jj} the element (j, j) of the matrix $(\boldsymbol{X}'\boldsymbol{X})^{-1}$.

- We can test the hypothesis of the nullity of the parameter and build confidence intervals:

$$IC_\alpha(\beta_j) = \left[b_j \pm t_{n-p-1;\alpha/2} s \sqrt{c_{jj}}\right]$$

LM: comments

- there are many other tests on the parameters . . .

LM: comments

- there are many other tests on the parameters . . .
- the practical question is the choice of the variables to use in the regression

LM: comments

- there are many other tests on the parameters . . .
- the practical question is the choice of the variables to use in the regression
- model selection issue: R^2 adjusted, C_p de Mallow, AIC, BIC,. . .

LM: comments

- there are many other tests on the parameters . . .
- the practical question is the choice of the variables to use in the regression
- model selection issue: R^2 adjusted, C_p de Mallow, AIC, BIC,. . .
- the LM framework is well mastered according to the statistical point of view

LM: comments

- there are many other tests on the parameters ...
- the practical question is the choice of the variables to use in the regression
- model selection issue: R^2 adjusted, C_p de Mallow, AIC, BIC,...
- the LM framework is well mastered according to the statistical point of view
- but LM is not suitable for several type of data: binary, ordinal, counts,...

LM: comments

- there are many other tests on the parameters ...
- the practical question is the choice of the variables to use in the regression
- model selection issue: R^2 adjusted, C_p de Mallow, AIC, BIC,...
- the LM framework is well mastered according to the statistical point of view
- but LM is not suitable for several type of data: binary, ordinal, counts,...
- $\rightarrow$ generalised linear model

Generalisation of LM

LM

$$y \sim \mathcal{N}(x\beta, \sigma^2) \implies E[y] = x\beta \text{ and } V[y] = \sigma^2$$

GLM: LM is related to the response variable via a link function

$y \sim \mathcal{L}$ with $\mathcal{L}$ in the exponential family

$$E[y] = g^{-1}(x\beta) \text{ and } V[y] = f(g^{-1}(x\beta))$$

g link function and f a function inducted by g and $\mathcal{L}$

GLM unifyed several models

- LM
- logistic regression
- probit regression
- poisson regression
- ...

GLM: logistic regresssion

- data type: y binary, x quantitative or qualitative
  ex: y: 0, 1, 1, 0, 0,. . .

GLM: logistic regresssion

- data type: y binary, x quantitative or qualitative
  ex: y: 0, 1, 1, 0, 0,...
- distribution to model y: $B(p)$ Bernoulli
- How link y and x ?

GLM: logistic regresssion

- data type: y binary, x quantitative or qualitative
  ex: y: 0, 1, 1, 0, 0,. . .
- distribution to model y: $B(p)$ Bernoulli
- How link y and x ? $\rightarrow$ link x and p

GLM: logistic regresssion

- data type: y binary, x quantitative or qualitative
  ex: y: 0, 1, 1, 0, 0,. . .
- distribution to model y: $B(p)$ Bernoulli
- How link y and x ? $\rightarrow$ link x and p
- How link x and p with $p \in [0,\ 1]$?

GLM: logistic regresssion

- data type: y binary, x quantitative or qualitative
  ex: y: 0, 1, 1, 0, 0,...
- distribution to model y: $B(p)$ Bernoulli
- How link y and x ? $\rightarrow$ link x and p
- How link x and p with $p \in [0,\ 1]$?$\rightarrow$ use a bijection between $[0,\ 1]$ and $\mathbb{R}$

GLM: logistic regresssion

- data type: y binary, x quantitative or qualitative
  ex: y: 0, 1, 1, 0, 0,...
- distribution to model y: $B(p)$ Bernoulli
- How link y and x ? $\rightarrow$ link x and p
- How link x and p with $p \in [0, 1]$?$\rightarrow$ use a bijection between $[0, 1]$ and $\mathbb{R}$
- a possibility: the logit function

$$logit(p) = \log \frac{p}{1-p} \text{ with inverse } logit^{-1}(t) = \frac{1}{1+e^{-t}} \text{ the logistic function}$$

GLM: logistic regresssion

- data type: y binary, x quantitative or qualitative
  ex: y: 0, 1, 1, 0, 0,...
- distribution to model y: $B(p)$ Bernoulli
- How link y and x ? $\rightarrow$ link x and p
- How link x and p with $p \in [0,\ 1]$?$\rightarrow$ use a bijection between $[0,\ 1]$ and $\mathbb{R}$
- a possibility: the logit function

$$logit(p) = \log \frac{p}{1-p} \text{ with inverse } logit^{-1}(t) = \frac{1}{1+e^{-t}} \text{ the logistic function}$$

- $logit(p) = x\beta \iff p = logit^{-1}(x\beta) = E[y]$

GLM: probit regresssion

- data type: y binary, x quantitative or qualitative
  ex: y: 0, 1, 1, 0, 0,...
- distribution to model y: $B(p)$ Bernoulli
- How link y and x ? $\rightarrow$ link x and p
- How link x and p with $p \in [0, \ 1]$? $\rightarrow$ use a bijection between $[0, \ 1]$ and $\mathbb{R}$

GLM: probit regresssion

- data type: y binary, x quantitative or qualitative
  ex: y: 0, 1, 1, 0, 0,...
- distribution to model y: $B(p)$ Bernoulli
- How link y and x ? $\rightarrow$ link x and p
- How link x and p with $p \in [0,\ 1]$? $\rightarrow$ use a bijection between $[0,\ 1]$ and $\mathbb{R}$
- a possibility: the gaussian probability distribution

$$z = x\beta + \varepsilon \quad \text{with } \varepsilon \sim \mathcal{N}(0,1)$$

$$y = \begin{cases} 0 & \text{if } z < 0 \\ 1 & \text{if } z \geq 0 \end{cases}$$

GLM: probit regresssion

- data type: y binary, x quantitative or qualitative
  ex: y: 0, 1, 1, 0, 0,...
- distribution to model y: $B(p)$ Bernoulli
- How link y and x ? $\rightarrow$ link x and p
- How link x and p with $p \in [0,\ 1]$? $\rightarrow$ use a bijection between $[0,\ 1]$ and $\mathbb{R}$
- a possibility: the gaussian probability distribution

$$z = x\beta + \varepsilon \quad \text{with } \varepsilon \sim \mathcal{N}(0,1)$$

$$y = \begin{cases} 0 & \text{if } z < 0 \\ 1 & \text{if } z \geq 0 \end{cases}$$

- $1 - p = \int_{-\infty}^{0} \frac{1}{\sqrt{2\pi}} \exp\{-\frac{1}{2}(u - x\beta)^2\} \mathrm{d}u$

GLM: ordinal regresssion

- data type: y qualitative ordinal, x quantitative or qualitative
  ex: y: A, B, B, C, B , B, A...
- distribution to model y: $\mathcal{M}(p_a, p_b, p_c)$ Multinomial
- How link y and x ? $\rightarrow$ link x and p_a, p_b, p_c with $p_a + p_b + p_c = 1$

GLM: ordinal regresssion

- data type: y qualitative ordinal, x quantitative or qualitative
  ex: y: A, B, B, C, B , B, A...
- distribution to model y: $\mathcal{M}(p_a, p_b, p_c)$ Multinomial
- How link y and x ? $\rightarrow$ link x and p_a, p_b, p_c with $p_a + p_b + p_c = 1$
- a possibility: the gaussian probability distribution

$$z = x\beta + \varepsilon \quad \text{with } \varepsilon \sim \mathcal{N}(0, 1)$$

$$y = \begin{cases} A & \text{if } z < \alpha_0 = 0 \\ B & \text{if } \alpha_0 < z \leq \alpha_1 \\ C & \text{if } \alpha_1 < z \end{cases}$$

- $p_a = \int_{-\infty}^{\alpha_0} \frac{1}{\sqrt{2\pi}} \exp\{-\frac{1}{2}(u - x\beta)^2\} \mathrm{d}u$

GLM: poisson regresssion

- data type: y count, x quantitative or qualitative
  ex: y: 0, 3, 12, 0, 0,. . .
- distribution to model y: $\mathcal{P}(\lambda)$ Poisson
- How link y and x ? $\rightarrow$ link x and λ
- How link x and λ with $\lambda \in \mathbb{R}+$? $\rightarrow$ use a bijection between $\mathbb{R}+$ and $\mathbb{R}$

GLM: poisson regresssion

- data type: y count, x quantitative or qualitative
  ex: y: 0, 3, 12, 0, 0,...
- distribution to model y: $\mathcal{P}(\lambda)$ Poisson
- How link y and x ? $\rightarrow$ link x and λ
- How link x and λ with $\lambda \in \mathbb{R}+$? $\rightarrow$ use a bijection between $\mathbb{R}+$ and $\mathbb{R}$
- a possibility: the log function

$$\log(\lambda) = x\beta \text{ with inverse } \lambda = exp(x\beta)$$

GLM: poisson regresssion

- data type: y count, x quantitative or qualitative
  ex: y: 0, 3, 12, 0, 0,. . .
- distribution to model y: $\mathcal{P}(\lambda)$ Poisson
- How link y and x ? $\rightarrow$ link x and λ
- How link x and λ with $\lambda \in \mathbb{R}+$? $\rightarrow$ use a bijection between $\mathbb{R}+$ and $\mathbb{R}$
- a possibility: the log function

$$\log(\lambda) = x\beta \text{ with inverse } \lambda = exp(x\beta)$$

- $y \sim \mathcal{P}\big(\exp(x\beta)\big)$

GLM: poisson regresssion

- data type: y count, x quantitative or qualitative
  ex: y: 0, 3, 12, 0, 0,...
- distribution to model y: $\mathcal{P}(\lambda)$ Poisson
- How link y and x ? $\rightarrow$ link x and λ
- How link x and λ with $\lambda \in \mathbb{R}+$? $\rightarrow$ use a bijection between $\mathbb{R}+$ and $\mathbb{R}$
- a possibility: the log function

$$\log(\lambda) = x\beta \text{ with inverse } \lambda = exp(x\beta)$$

- $y \sim \mathcal{P}\big(\exp(x\beta)\big)$
- $P(y = k) = \exp^{-\exp(x\beta)} \dfrac{\exp(x\beta)^k}{k!}$

GLM: comments

- enlarge the type of data supported
- keep a quite simple formalism
- inference is more complex

GLM: comments

- enlarge the type of data supported
- keep a quite simple formalism
- inference is more complex
  - least squares doesn't work any more
  - $t - test$ for parameter significance are not reliable

GLM: comments

- enlarge the type of data supported
- keep a quite simple formalism
- inference is more complex
  - least squares doesn't work any more
  - $t - test$ for parameter significance are not reliable
  - maximize the likelihood is the appropriate method
  - likelihood ratio test for parameter significance are reliable

the likelihood

definition

Let $X_1, \ldots, X_n$ a sample that the density depends on θ. The density of this sample L, is the likelihood:

$$L(x_1, \ldots, x_n; \theta) = L(\boldsymbol{x}, \theta)$$

likelihood i.i.d. case

Assume $X_1, \ldots, X_n$ is i.i.d. according to the density distribution $f(x, \theta)$ then the likelihhod is the densities product:

$$L(\boldsymbol{x}, \theta) = \prod_{i=1}^{n} f(x_i, \theta).$$

The likelihood of a sample is highly important in statistic

Maximum likelihood estimate (mle)

- Let$X_1, \ldots, X_n$ is i.i.d. according to the density distribution $f(x, \theta)$
- Principle : take for estimate of θ the value $\hat{\theta}$ which make maximum the likelihood $L(x_1, \ldots, x_n; \theta)$ for the observations $x_1, \ldots, x_n$

Maximum likelihood estimate (mle)

- Let$X_1, \dots, X_n$ is i.i.d. according to the density distribution $f(x, \theta)$
- Principle : take for estimate of θ the value $\hat{\theta}$ which make maximum the likelihood $L(x_1, \dots, x_n; \theta)$ for the observations $x_1, \dots, x_n$

mle

A maximum likelihood estimate of θ is $\hat{\theta}$ such that

$$\hat{\theta} \in \arg\max_{\theta} L(x_1, \dots, x_n; \theta)$$

- L is the density of the sample, the mle is value the of the parameter that make the observations the more probable
- practically we maximize $\ln L(\boldsymbol{x}; \theta)$ rather than $L(\boldsymbol{x}; \theta)$

mle for several parameters

$\theta = (\theta_1, \ldots, \theta_p) \in \mathbb{R}^p$, the usual method to find the mle consist in solving:

$$\frac{\partial L(\boldsymbol{x}, \theta)}{\partial \theta_i} = 0 \quad i = 1, \ldots, p$$

mle for several parameters

$\theta = (\theta_1, \dots, \theta_p) \in \mathbb{R}^p$, the usual method to find the mle consist in solving:

$$\frac{\partial L(\boldsymbol{x}, \theta)}{\partial \theta_i} = 0 \quad i = 1, \dots, p$$

Exercice

Let $X_1, \dots, X_n$ i.i.d. according to a $\mathcal{N}(\mu, \sigma^2)$

1. Assuming that σ^2 is known, determine the mle of μ.
2. Determine the mle of μ and σ^2

mle for several parameters

$\theta = (\theta_1, \ldots, \theta_p) \in \mathbb{R}^p$, the usual method to find the mle consist in solving:

$$\frac{\partial L(\boldsymbol{x}, \theta)}{\partial \theta_i} = 0 \quad i = 1, \ldots, p$$

Exercice

Let $X_1, \ldots, X_n$ i.i.d. according to a $\mathcal{N}(\mu, \sigma^2)$

1. Assuming that σ^2 is known, determine the mle of μ.
2. Determine the mle of μ and σ^2

Remark : in the general case the mle is not tracktable analytically, we make numerical approximation with a computer

Maximum likelihood estimate properties

$\hat{\theta}$ maximum likelihood estimate of θ

- principle : estimates parameter in order to maximize model data fit
- properties :

CTL

Under regularities assumptions,

$$\frac{\hat{\theta} - \theta}{I_n(\theta)^{-1/2}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)$$

$\rightarrow$ very useful: hypothesis tests, confidence intervals

Maximum likelihood estimate properties

$\hat{\theta}$ maximum likelihood estimate of θ

- principle : estimates parameter in order to maximize model data fit
- properties :

CTL

Under regularities assumptions,

$$\frac{\hat{\theta} - \theta}{I_n(\theta)^{-1/2}} \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1)$$

$\rightarrow$ very useful: hypothesis tests, confidence intervals

- very used
  - large validity domain
  - computed numerically in the general cases

TP : Illustration of the CLT for the mle

- Write a program that excute the folowing tasks:
  1. simulate *n* realisations according to $\mathcal{N}(\mu, \sigma^2)$
  2. compute numerically $\hat{\mu}$ the mle of μ assuming σ^2 known:
     - write the likelihood function
     - use the fonction *optim* to find $\hat{\mu}$ the max of the likelihood
  3. repeat the two firsts stages N times $\rightarrow \hat{\mu}_1, \ldots, \hat{\mu}_N$
  4. plot the histogram of $\hat{\mu}_1, \ldots, \hat{\mu}_N$
  5. return the variance of $\hat{\mu}_1, \ldots, \hat{\mu}_N$

TP : Illustration of the CLT for the mle

- Write a program that excute the folowing tasks:
  1. simulate *n* realisations according to $\mathcal{N}(\mu, \sigma^2)$
  2. compute numerically $\hat{\mu}$ the mle of μ assuming σ^2 known:
     - write the likelihood function
     - use the fonction *optim* to find $\hat{\mu}$ the max of the likelihood
  3. repeat the two firsts stages *N* times $\rightarrow \hat{\mu}_1, \ldots, \hat{\mu}_N$
  4. plot the histogram of $\hat{\mu}_1, \ldots, \hat{\mu}_N$
  5. return the variance of $\hat{\mu}_1, \ldots, \hat{\mu}_N$
- Input-Outout of the program:
  - Input: μ, σ^2, *n* and *N*
  - Output: histogram and variance of the $\hat{\mu}_1, \ldots, \hat{\mu}_N$

TP : Illustration of the CLT for the mle

- Write a program that excute the folowing tasks:
  1. simulate n realisations according to $\mathcal{N}(\mu, \sigma^2)$
  2. compute numerically $\hat{\mu}$ the mle of μ assuming σ^2 known:
     - write the likelihood function
     - use the fonction *optim* to find $\hat{\mu}$ the max of the likelihood
  3. repeat the two firsts stages N times $\rightarrow \hat{\mu}_1, \ldots, \hat{\mu}_N$
  4. plot the histogram of $\hat{\mu}_1, \ldots, \hat{\mu}_N$
  5. return the variance of $\hat{\mu}_1, \ldots, \hat{\mu}_N$
- Input-Outout of the program:
  - Input: μ, σ^2, n and N
  - Output: histogram and variance of the $\hat{\mu}_1, \ldots, \hat{\mu}_N$
- Execute the prg for different input
- What is the limit $V(\hat{\mu}_1, \ldots, \hat{\mu}_N)$ for $N \rightarrow \infty$

Likelihood approache: linear model case

- linear model : $y_i = \beta x_i + \epsilon_i$

$$y_1 \sim \mathcal{N}(\beta x_1, \sigma^2), \dots, y_n \sim \mathcal{N}(\beta x_n, \sigma^2)$$

- the likelihood : evaluate how model fits data

$$\begin{aligned} L(\beta) &= Loi((y_1, x_1), \dots, (y_n, x_n)|\beta) \\ &= \prod_{i=1}^n Loi(y_i|x_i, \beta) \\ &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} \exp\{-\frac{1}{2\sigma^2}(y_i - \beta x_i)^2\} \\ &= \frac{1}{\sqrt{(2\pi)^n}} \exp\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \beta x_i)^2\} \end{aligned}$$

- maximize the likelihood ⇔ maximize the log-likelihood

$$\begin{aligned} l(\beta) &= \log L(\beta) \\ &= -\frac{n}{2} \log(2\pi) - \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \beta x_i)^2 \end{aligned}$$

- maximize the likelihood ⇔ minimize least squares