Modèles Linéaires Appliqués

Arthur Charpentier

Automne 2020

Rappels #1 (probabilités)

Fonction de répartition et quantiles

Let X denote a random variable, its cumulative distribution function (c.d.f.) is

$$F(x) = \mathbb{P}(X \leq x), \text{ for all } x \in \mathbb{R}.$$

More formally, $F(x) = \mathbb{P}(\{\omega \in \Omega | X(\omega) \leq x\})$.
Observe that

- F is an increasing function on $\mathbb{R}$ with values in $[0, 1]$,
- $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to +\infty} F(x) = 1$.

Let X denote a random variable, its quantile function is

$$Q(p) = F^{-1}(p) = \inf\{x \in \mathbb{R} \text{ such that } F(x) > p\}, \text{ for all } p \in [0, 1].$$

Densité

Let X be a random variable. The density or probablity function of X is

$$f(x)=\begin{cases}\dfrac{dF(x)}{dx}=F'(x)\text{ in the (absolutely) continous case}, x\in\mathbb{R}\\ \mathbb{P}(X=x)\text{ in the discrete case}, x\in\mathbb{N}\\ dF(x),\text{ in a more general context}\end{cases}$$

Example: if $X\sim\mathcal{U}_{[0,1]}$, $f(x)=\mathbf{1}_{[0,1]}(x)$, $x\in\mathbb{R}$.

Example: if $X\sim\mathcal{U}_{[a,b]}$, $f(x)=\dfrac{1}{b-a}\mathbf{1}_{[a,b]}(x)$, $x\in\mathbb{R}$.

Moments (Espérance & Variance)

$$\mathbb{E}[Y]=\int_{\mathbb{R}} yf(y)\,dy,\text{ ou }\sum_{y\in\mathbb{N}} yf(y)\text{ dans le cas discret}$$

La version empirique, pour un échantillon $\{y_1, y_2, \cdots, y_n\}$ est

$$\overline{y}=\sum_{i=1}^{n} y_i\frac{1}{n}=\frac{1}{n}\sum_{i=1}^{n} y_i$$

$$\text{Var}[Y]=\mathbb{E}\Big[(Y-\mathbb{E}[Y])^2\Big]=\int_{\mathbb{R}}(y-\mathbb{E}[Y])^2 f(y)\,dy,$$

Note: $\text{Var}[Y]=\mathbb{E}[Y^2]-\mathbb{E}[Y]^2$.
La version empirique, pour un échantillon $\{y_1, y_2, \cdots, y_n\}$ est

$$s^2=\sum_{i=1}^{n}(y_i-\overline{y})^2\frac{1}{n}=\frac{1}{n}\sum_{i=1}^{n}(y_i-\overline{y})^2$$

Vecteur aléatoire

Let $\mathbf{Z} = (X, Y)$ be a random vector. The cumulative distribution function of $\mathbf{Z}$ is

$$F(\mathbf{z}) = F(x,y) = \mathbb{P}(X \leq x, Y \leq y), \text{ for all } \mathbf{z} = (x,y) \in \mathbb{R} \times \mathbb{R}.$$

Let $\mathbf{Z} = (X, Y)$ be a random vector. The density of $\mathbf{Z}$ is

$$f(\mathbf{z}) = f(x,y) = \begin{cases} \dfrac{\partial^2 F(x,y)}{\partial x \partial y} \text{ in the continuous case}, \mathbf{z} = (x,y) \in \mathbb{R}^2 \\ \mathbb{P}(X = x, Y = y) \text{ in the discrete case}, \mathbf{z} = (x,y) \in \mathbb{N}^2 \end{cases}$$

Note:

$$\text{Var}(\mathbf{Z}) = \mathbb{E}\Big[(\mathbf{Z} - \mathbb{E}[\mathbf{Z}])(\mathbf{Z} - \mathbb{E}[\mathbf{Z}])^\top\Big] = \begin{pmatrix} \text{Var}[X] & \text{Cov}[X,Y] \\ \text{Cov}[X,Y] & \text{Var}[Y] \end{pmatrix}$$

Lois Binomiales

$Y \sim \mathcal{B}(p)$:

$$\mathbb{P}[Y = y] = p^y(1-p)^{1-y} \begin{cases} p \text{ si } y = 1 \\ 1-p \text{ si } y = 0 \end{cases}, \text{ où } y \in \{0,1\}$$

cf loi de Bernoulli, où $p = \mathbb{P}[Y = 1] = \mathbb{E}[Y] \in [0,1]$.

$Y \sim \mathcal{B}(n, p)$:

$$\mathbb{P}[Y = y] = \binom{n}{y} p^y(1-p)^{n-y} \text{ où } y \in \{0,1,2,\ldots,n\}$$

cf loi binomiale, où $\mathbb{E}[Y] = np$.

$$Y_1, \cdots, Y_n \text{ i.i.d. } \mathcal{B}(p) \text{ alors } Y = \sum_{i=1}^{n} Y_i \sim \mathcal{B}(n,p)$$

Lois Binomiales

$Y \sim \mathcal{B}(n, p)$:

$$\mathbb{P}[Y = y] = \binom{n}{y} p^y (1-p)^{n-y}$$

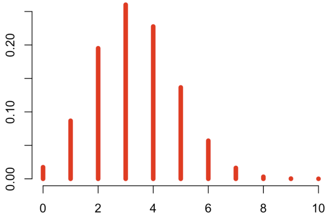
où $y \in \{0, 1, 2, \ldots, n\}$ cf loi binomiale,
où $\mathbb{E}[Y] = np$ et $\text{Var}[Y] = np(1-p)$.
Note: when $pn \sim \lambda$ and n large enough,

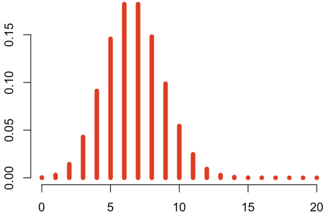
$$\mathcal{B}(n, p) \approx \mathcal{P}(\lambda), \quad \text{where } \lambda = \frac{p}{n}$$

and when $n \to \infty$,

$$\mathcal{B}(n, p) \approx \mathcal{N}(np, np(1-p))$$

(see Galton's quincunx)





Lois Binomiales & Multinomiales

$\mathbf{Y} = (Y_1, \cdots, Y_d) \sim \mathcal{M}(\mathbf{p})$ où $\mathbf{p} = (p_1, \cdots, p_d)$ si

$$Y_1 + \cdots + Y_d = 1 \text{ et } Y_j \sim \mathcal{B}(p_j),\ \forall j \in \{1, \cdots, d\}$$

i.e. $\mathbf{Y} = (\mathbf{1}_{C_1}, \mathbf{1}_{C_2}, \cdots, \mathbf{1}_{C_d})$

$\mathbf{Y} = (Y_1, \cdots, Y_d) \sim \mathcal{M}(n, \mathbf{p})$ où $\mathbf{p} = (p_1, \cdots, p_d)$ si

$$Y_1 + \cdots + Y_d = n \text{ et } Y_j \sim \mathcal{B}(n, p_j),\ \forall j \in \{1, \cdots, d\}$$

cf loi multinomiale. Pour
$(y_1, \cdots, y_d) \in \mathcal{S}_{d,n} = \{(y_1, \cdots, y_d) \in \mathbb{N}^d : (y_1 + \cdots + y_d = n\}$

$$\mathbb{P}[(Y_1, \cdots, Y_d) = (y_1, \cdots, y_d)] = \frac{n!}{y_1! \cdots y_d!} p_1^{y_1} \cdots p_d^{y_d}$$

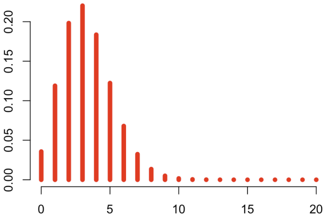
Example: $\mathbf{Y} = (Y_0, Y_1) \sim \mathcal{M}(n, \mathbf{p})$ où $\mathbf{p} = (p_0, p_1)$.

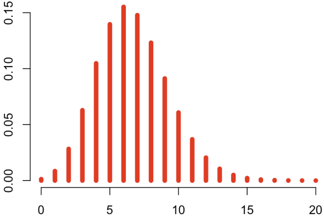
Lois de Poisson

$Y \sim \mathcal{P}(\lambda)$:

$$\mathbb{P}[Y=y] = e^{-\lambda}\frac{\lambda^y}{y!}$$

où $y \in \{0,1,2,\dots\}$ cf loi de Poisson,
où $\mathbb{E}[Y]=\lambda$ et $\text{Var}[Y]=\lambda$.
Note: si $Y \sim \mathcal{P}(\lambda)$, $\mathbb{P}[Y=0]=e^{-\lambda}$,
aussi $\lambda = -\log \mathbb{P}[Y=0]$.





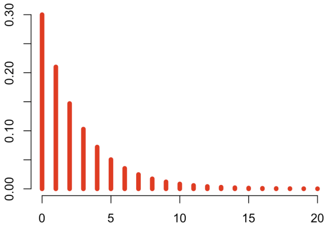
Distribution Géométrique

$Y \sim \mathcal{G}(p)$:

$$\mathbb{P}[Y = y] = p(1-p)^y$$

où $y \in \{0, 1, 2, \ldots\}$ cf loi géométrique,

$$\mathbb{E}[Y] = \frac{1}{p} - 1 \text{ and } \mathrm{Var}[Y] = \frac{1-p}{p^2}$$

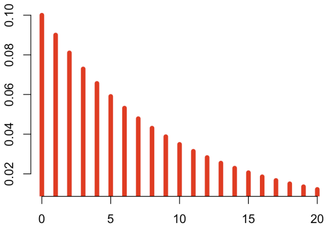


Note: possible alternative representation

$$\mathbb{P}[Y = y] = p(1-p)^{y-1}$$

où $y \in \{1, 2, \ldots\}$, et dans ce cas

$$\mathbb{E}[Y] = \frac{1}{p} \text{ and } \mathrm{Var}[Y] = \frac{1-p}{p^2}$$



Distribution Exponentielle

$Y \sim \mathcal{E}(\lambda)$:

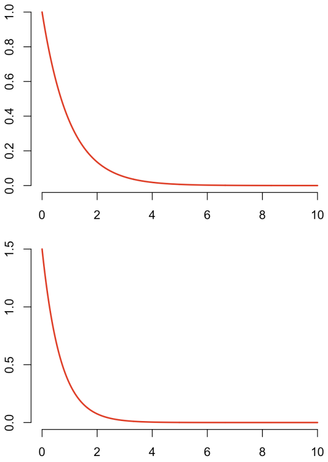
$$f(y) = \lambda e^{-\lambda y} \text{ ou } \mathbb{P}[Y > y] = e^{-\lambda y}$$

où $y \in \mathbb{R}_+$, cf loi exponentielle,

$$\mathbb{E}[Y] = \frac{1}{\lambda} \text{ and } \text{Var}[Y] = \frac{1}{\lambda^2}$$

Absence de mémoire: pour $h > 0$,

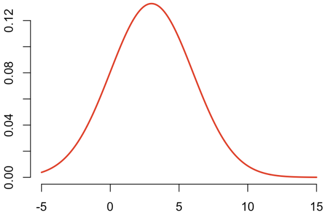
$$\mathbb{P}[Y > y + h | Y > y] = \frac{e^{-\lambda(y+h)}}{e^{-\lambda h}} \mathbb{P}[Y > h]$$



Distribution Gaussienne

$Y \sim \mathcal{N}(\mu, \sigma^2)$:

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



où $y \in \mathbb{R}$, cf loi normale,

$$\mathbb{E}[Y] = \mu \text{ and } \mathrm{Var}[Y] = \sigma^2$$

Si $Z \sim \mathcal{N}(0,1)$, $Y = \mu + \sigma Z \sim \mathcal{N}(\mu, \sigma^2)$, et $\dfrac{Y-\mu}{\sigma} \sim \mathcal{N}(0,1)$.

Note: $\mathbb{P}[Z \leq z] = \Phi(z) = \dfrac{1}{\sqrt{2\pi}} \displaystyle\int_{-\infty}^{z} \exp\left(-\frac{x^2}{2}\right) \mathrm{d}x$

```
> pnorm(1.96)
[1] 0.9750021
> qnorm(.95)
[1] 1.644854
```

Distribution Log-Normale

$Y \sim LN(\mu, \sigma^2)$:

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

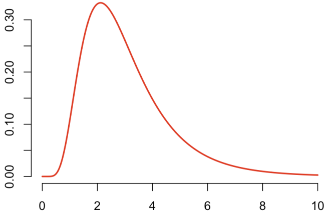
où $y \in \mathbb{R}_+$, cf loi log-normale,

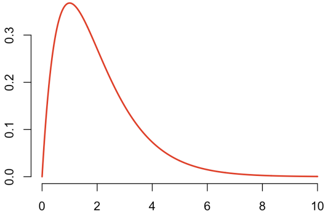
$$\mathbb{E}[Y] = \exp\left(\mu + \frac{\sigma^2}{2}\right)$$

$$\text{Var}[Y] = (e^{\sigma^2} - 1)\exp\left(2\mu + \sigma^2\right)$$

Si $Y \sim LN(\mu, \sigma^2)$, $\log Y \sim \mathcal{N}(\mu, \sigma^2)$

Note: $\mathbb{E}(\log Y) \neq \log \mathbb{E}(Y)$





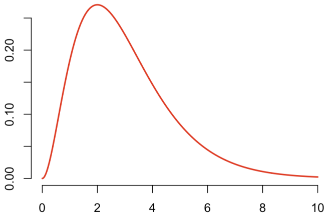
Distribution Gamma

$Y \sim \mathcal{G}(\alpha,\beta)$:

$$f(y) = y^{\alpha-1}\frac{\beta^{\alpha}\,\mathrm{e}^{-\beta x}}{\Gamma(\alpha)}$$

où $y \in \mathbb{R}_+$, cf loi gamma,

$$\mathbb{E}[Y] = \frac{\alpha}{\beta} \text{ et } \mathrm{Var}[Y] = \frac{\alpha}{\beta^2}$$



α est appelé *shape* et β *rate*
Note: si $X \sim \mathcal{G}(\nu/2, 1/2)$ alors $X \sim \chi^2(\nu)$, cf loi du chi-deux.

Convergence

The sequence (X_n) converges in probability towards X, denoted $X_n \xrightarrow{\mathbb{P}} X$, if

$$\forall \varepsilon > 0, \ \lim_{n\to\infty} \mathbb{P}\left(|X_n - X| > \varepsilon\right) = 0.$$

Let $f : \mathbb{R} \to \mathbb{R}$ be a continuous function, if $X_n \xrightarrow{\mathbb{P}} X$ then $f(X_n) \xrightarrow{\mathbb{P}} f(X)$. More generally, let $f : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ be a continuous function, if $X_n \xrightarrow{\mathbb{P}} X$ and $Y_n \xrightarrow{\mathbb{P}} y \in \mathbb{R}$ then $f(X_n, Y_n) \xrightarrow{\mathbb{P}} f(X, y)$ (Slutsky Theorem).

A sufficient condition to have $X_n \xrightarrow{\mathbb{P}} a$ is that

$$\lim_{n\to\infty} \mathbb{E}[X_n] = a \text{ and } \lim_{n\to\infty} \text{Var}(X_n) = 0$$

Suppose X_i's are i.i.d. with finite expected value $\mu = \mathbb{E}(X_i)$, then $\overline{X}_n \xrightarrow{\mathbb{P}} \mu$ as $n \to +\infty$, see Law of Large Numbers

Sequence (X_n) converges in distribution towards X, denoted $X_n \xrightarrow{\mathcal{L}} X$, if for any continuous function h

$$\lim_{n\to\infty} \mathbb{E}\left[h\left(X_n\right)\right] = \mathbb{E}\left(h\left(X\right)\right).$$

Convergence in distribution is the same as convergence of distribution function $X_n \xrightarrow{\mathcal{L}} X$ if for any $t \in \mathbb{R}$ where F_X is continuous

$$\lim_{n\to\infty} F_{X_n}\left(t\right) = F_X\left(t\right).$$

Let $X_1, X_2 \ldots$ denote i.i.d. random variables with mean μ and variance σ^2, then :

$$\frac{\overline{X}_n - \mathbb{E}(\overline{X}_n)}{\sqrt{\mathrm{Var}(\overline{X}_n)}} = \sqrt{n}\left(\frac{\overline{X}_n - \mu}{\sigma}\right) \xrightarrow{\mathcal{L}} X \text{ where } X \sim \mathcal{N}\left(0,1\right)$$

see Central Limit Theorem

Transformation

Let X be an absolutely continuous random variable with density $f(x)$. If function ϕ is a differentiable one-to-one mapping, then variable $Y = \phi(X)$ has a density g satisfying

$$g(y) = \frac{f(\phi^{-1}(y))}{\phi'(\phi^{-1}(y))}.$$

Let X be an absolutely continuous random variable with cdf F, i.e. $F(x) = \mathbb{P}(X \leq x)$. Then $Y = F(X)$ has a uniform distribution on $[0, 1]$.
Let Y be a uniform distribution on $[0, 1]$ and F denote a cdf. Then $X = F^{-1}(Y)$ is a random variable with cdf F.

Delta Method

In the univariate case, if there is a sequence of random variables X_n satisfying $\sqrt{n}[X_n - \theta] \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2)$, where θ and σ^2 are two constants, then

$$\sqrt{n}[g(X_n) - g(\theta)] \xrightarrow{\mathcal{L}} \mathcal{N}(0, \sigma^2 \cdot [g'(\theta)]^2)$$

for any function g satisfying the property that $g'(\theta)$ exists and is non-zero valued.

In the multivariate case, if there is a sequence of random vectors $\mathbf{X}_n$ satisfying $\sqrt{n}[\mathbf{X}_n - \boldsymbol{\theta}] \xrightarrow{\mathcal{L}} \mathcal{N}(\mathbf{0}, \Sigma)$, where $\boldsymbol{\theta}$ is a vector in $\mathbb{R}^d$ and Σ is a symmetric positive $d \times d$ matrix, then

$$\sqrt{n}[g(\mathbf{X}_n) - g(\boldsymbol{\theta})] \xrightarrow{\mathcal{L}} \mathcal{N}(\mathbf{0}, \nabla g(\boldsymbol{\theta})^\top \Sigma \nabla g(\boldsymbol{\theta}))$$

for any function g satisfying the property that $\nabla g(\boldsymbol{\theta})$ exists and is non-zero valued.