

# Modèles Linéaires Appliqués

Arthur Charpentier

Automne 2020

Rappels #3.2 (statistique & maximum de vraisemblance)

## Parametric Statistical Models

A family of distributions  $\mathcal{F} = \{F_\theta, \theta \in \Theta\}$  is identifiable if the mapping  $\theta \mapsto F_\theta$  is one-to-one:

$$F_{\theta_1} = F_{\theta_2} \text{ implies } \theta_1 = \theta_2.$$

**Example** The Gaussian distribution,  $\theta = (\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}_+$ ,  
 $f_\theta(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(x - \mu)^2\right)$ . If  $f_{\theta_1} = f_{\theta_2}$  then

$$\frac{1}{\sqrt{2\pi}\sigma_1} \exp\left(-\frac{1}{2\sigma_1^2}(x - \mu_1)^2\right) = \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left(-\frac{1}{2\sigma_2^2}(x - \mu_2)^2\right)$$

$$\frac{1}{\sigma_1^2}(x - \mu_1)^2 + \ln \sigma_1 = \frac{1}{\sigma_2^2}(x - \mu_2)^2 + \ln \sigma_2$$

$$x^2 \left( \frac{1}{\sigma_1^2} - \frac{1}{\sigma_2^2} \right) - 2x \left( \frac{\mu_1}{\sigma_1^2} - \frac{\mu_2}{\sigma_2^2} \right) + \left( \frac{\mu_1^2}{\sigma_1^2} - \frac{\mu_2^2}{\sigma_2^2} + \ln \sigma_1 - \ln \sigma_2 \right) = 0 \quad \forall x \in \mathbb{R}$$

hence  $\sigma_1^2 = \sigma_2^2$  and  $\mu_1 = \mu_2$ .

# Parametric Statistical Models

**Example** Mixture of two distributions :

$$\theta = (p, \lambda, \mu)^\top \in (0, 1) \times \mathbb{R}_+ \times \mathbb{R}_+,$$

$$f_\theta(x) = p \cdot (\lambda e^{-\lambda x}) + (1 - p) \cdot (\mu e^{-\mu x})$$

Observe that  $\theta_1 = (p, \lambda, \mu)$  and  $\theta_2 = (1 - p, \mu, \lambda)$  yield the same distributions, since  $f_{\theta_1}(x) = f_{\theta_2}(x)$ ,  $\forall x \in \mathbb{R}_+$ .

It is necessary to add a (linear) constraint : either  $p > 1 - p$  or  $\lambda > \mu$ .

Here we want to solve  $\min \{ \log \mathcal{L}(\theta) \}$  for  $\theta \in \mathbb{R}^p$  subject to  $\mathbf{A}\theta \geq \mathbf{b}$  for some  $k \times p$  matrix  $\mathbf{A}$  and  $k$  dimensional vector  $\mathbf{b}$ .

## Parametric Statistical Models

$\theta = (p, \lambda, \mu)^\top \in (0, 1) \times \mathbb{R}_+ \times \mathbb{R}_+$ , and  $\lambda > \mu$

$$f_\theta(x) = p \cdot (\lambda e^{-\lambda x}) + (1 - p) \cdot (\mu e^{-\mu x})$$

$$\begin{pmatrix} 1 & 0 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & -1 \end{pmatrix} \begin{pmatrix} p \\ \lambda \\ \mu \end{pmatrix} = \begin{pmatrix} p \\ -p \\ \mu \\ \lambda - \mu \end{pmatrix} \geq \begin{pmatrix} 0 \\ -1 \\ 0 \\ 0 \end{pmatrix} \quad \text{i.e.} \quad \begin{cases} p \geq 0 \\ p \leq 1 \\ \mu \geq 0 \\ \lambda \geq \mu \end{cases}$$

```
1 logL = function(param){  
2   -sum(log(param[1]*dexp(X,param[2])+(1-param[1])*dexp(X  
   ,param[3])))  
3 }  
4 Amat = matrix(c(1,-1,0,0,0,0,0,1,0,0,1,-1), 4, 3)  
5 bvec = c(0,-1,0,0)  
6 constrOptim(c(.25,2,.5), logL, NULL, ui = Amat, ci =  
   bvec)$par
```

## Parametric Statistical Models

- Un estimateur,  $T$ , de  $\theta$  est dit sans biais, ou non biaisé, si  $\mathbb{E}(T) = \theta$ . Autrement dit, le biais,

$$\text{biais}(\theta) = \mathbb{E}(T) - \theta = 0$$

- Risque quadratique d'un estimateur  $T$  de  $\theta$ :

$$R(T, \theta) = \mathbb{E}[(T - \theta)^2]$$

On a

$$R(T, \theta) = b(\theta)^2 + \text{Var}_\theta(T)$$

(pour un estimateur sans biais,  $R(T, \theta) = \text{Var}_\theta(T)$ )

- Soient  $T_1$  et  $T_2$  deux estimateurs de  $\theta$ . On dira que  $T_1$  est plus efficace que  $T_2$  si  $R(T_1, \theta) \leq R(T_2, \theta)$ .

# Parametric Statistical Models

On dit que la suite d'estimateurs  $(T_n)_{n \geq 1}$  d'estimateurs de  $\theta$  est

- ▶ convergente, si  $T_n \xrightarrow{\mathbb{P}} \theta$  pour tout  $\theta \in \Theta$ .
- ▶ fortement convergente, si  $T_n \xrightarrow{p.s.} \theta$  pour tout  $\theta \in \Theta$ .
- ▶ asymptotiquement normale, si pour tout  $\theta \in \Theta$ , il existe une matrice de covariance  $\Sigma(\theta)$  telle que

$$\sqrt{n}(T_n - \theta) \xrightarrow{\mathcal{L}} N(0, \Sigma(\theta))$$

lorsque  $n \rightarrow \infty$ .

## Parametric Statistical Models

Soit  $X$  une v.a. continue à valeurs dans  $\mathcal{X}$ . On supposera ici

(i)  $\{x \in \mathcal{X} : f(x; \theta) > 0\}$  ne dépend pas de  $\theta \in \Theta$ .

(ii) La fonction  $\theta \mapsto f(x; \theta)$  est  $C^2$  sur  $\Theta$ .

(iii)  $\forall A \subseteq \mathcal{X}$

$$\frac{\partial}{\partial \theta} \int_A f(x; \theta) dx = \int_A \frac{\partial}{\partial \theta} f(x; \theta) dx$$

et

$$\frac{\partial^2}{\partial \theta \partial \theta^\top} \int_A f(x; \theta) dx = \int_A \frac{\partial^2}{\partial \theta \partial \theta^\top} f(x; \theta) dx.$$

(iv) La statistique  $T(\mathbf{X})$  est de carré intégrable: elle satisfait  $\mathbb{E}_\theta(T(\mathbf{X})^2) < \infty$  et

$$\frac{\partial}{\partial \theta} \mathbb{E}_\theta(T(\mathbf{X})) = \int_{\mathcal{X}^n} T(\mathbf{x}) \frac{\partial}{\partial \theta} f(\mathbf{x}; \theta) dx_1 \dots dx_n.$$