Modèles Linéaires Appliqués

Arthur Charpentier

Automne 2020

Rappels #3.2 (statistique & inférence paramétrique)

Parametric Statistical Models

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

$$F_{\boldsymbol{\theta}_1} = F_{\boldsymbol{\theta}_2} \text{ implies } \boldsymbol{\theta}_1 = \boldsymbol{\theta}_2.$$

Example The Gaussian distribution, $\boldsymbol{\theta} = (\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}_+$, $f_{\boldsymbol{\theta}}(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right)$. If $f_{\boldsymbol{\theta}_1} = f_{\boldsymbol{\theta}_2}$ then

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

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

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

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

Parametric Statistical Models

Example Mixture of two distributions :
$\boldsymbol{\theta} = (p, \lambda, \mu)^\top \in (0,1) \times \mathbb{R}_+ \times \mathbb{R}_+$,

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

Observe that $\boldsymbol{\theta}_1 = (p, \lambda, \mu)$ and $\boldsymbol{\theta}_2 = (1-p, \mu, \lambda)$ yield the same distributions, since $f_{\boldsymbol{\theta}_1}(x) = f_{\boldsymbol{\theta}_2}(x)$, $\forall x \in \mathbb{R}_+$.
It is necessary to add a (linear) constraint : either $p > 1-p$ or $\lambda > \mu$.

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

Parametric Statistical Models

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

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

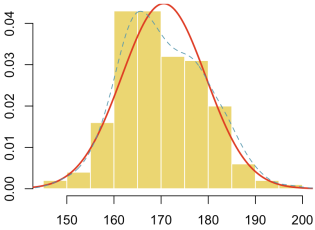
$$\begin{pmatrix} 1 & 0 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & -1 \end{pmatrix} \begin{pmatrix} p \\ \lambda \\ \mu \end{pmatrix} = \begin{pmatrix} p \\ -p \\ \mu \\ \lambda - \mu \end{pmatrix} \geq \begin{pmatrix} 0 \\ -1 \\ 0 \\ 0 \end{pmatrix} \quad \text{i.e.} \quad \begin{cases} p \geq 0 \\ p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

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

Parametric Statistical Models

Height of students 1. Gaussian model, $f(x) = \phi_{\overline{x},s^2}(x)$

```
> X = Davis$height
> library(MASS)
> hist(X, proba=TRUE)
> (param = fitdistr(X,"normal")$estimate)
     mean         sd
170.02000   11.97788
> f1 = function(x) dnorm(x,param[1],param[2])
> x = seq(100,210,by=.2)
> lines(x,f1(x),lwd=2)
> lines(density(X),lty=2)
```



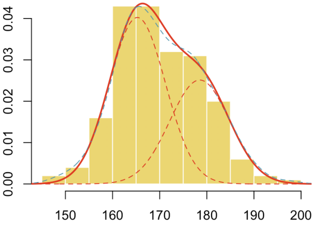

```
> logdf = function(x,p){
 p1 = p[1]
 m1 = p[2]; s1 = p[4]
 m2 = p[3]; s2 = p[5]
 log(p1*dnorm(x,m1,s1)+(1-p1)*dnorm(x,m2,s2))}
```

Parametric Statistical Models

Height of students 2. Mixture of 2 Gaussians,

$$f(x) = p\phi_{\mu_1,\sigma_1^2}(x) + (1-p)\phi_{\mu_2,\sigma_2^2}(x)$$

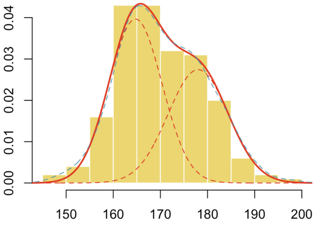
```
> logL = function(parameter) -sum(logdf(X,parameter))
> Amat = matrix(c(1,-1,0,0,0,0,0,0,0,0,1,0,0,0,0,0,0,0,0,1), 4, 5)
> bvec = c(0,-1,0,0)
> (param12 = constrOptim(c(.5,160,180,10,10), logL, NULL, ui = Amat, ci = bvec)$par)
[1]   0.5996263 165.2690084
    178.4991624   5.9447675
      6.3564746
```



Height of students 3. Conditional Gaussian

$$f(x) = \mathbb{P}(F) \cdot \phi_{\overline{x}_F, s_F^2}(x) + \mathbb{P}(M) \cdot \phi_{\overline{x}_M, s_M^2}(x)$$

```
> (pM = mean(Davis$sex=="F"))
[1] 0.56
> (paramF = fitdistr(X[Davis$sex=="F"],"normal")$estimate)
     mean        sd
163.74107  11.59183
> (paramM = fitdistr(X[Davis$sex=="M"],"normal")$estimate)
      mean          sd
178.011364   6.404001
```



Parametric Statistical Models

- Un estimateur, T, de θ 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 θ:

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

On a

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

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

- Soient T_1 et T_2 deux estimateurs de θ. 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 θ est

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

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

lorsque $n \to \infty$.

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 $\mathcal{C}^2$ sur Θ.
(iii) $\forall A \subseteq \mathcal{X}$

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

et

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

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

$$\frac{\partial}{\partial\theta}\mathbb{E}_\theta(T(\mathbf{X})) = \int_{\mathcal{X}^n} T(\mathbf{x})\frac{\partial}{\partial\theta} f(\mathbf{x};\theta)\mathrm{d}x_1\ldots\mathrm{d}x_n.$$