Estimation bayésienne

Franck Corset

Master 2 - SSD

Estimation bayésienne

Définition

Cas uni-dimensionnel

- On suppose que θ est réel et $\mathbf{x} = (x_1, \ldots, x_n)$ sont l'échantillon observé.
- On note $\pi(\theta|\mathbf{x})$, la loi a posteriori.
- On appelle estimation bayésienne de θ la moyenne a posteriori, notée $\mathbb{E}[\theta|\mathbf{x}]$.
- Cette moyenne a posteriori est

$$\mathbb{E}[\theta|(x_1, \ldots, x_n)] = \int_\Theta \theta \pi(\theta|\mathbf{x})\, d\theta = \frac{\int_\Theta \theta f(\mathbf{x}|\theta)\pi(\theta)\, d\theta}{\int_\Theta f(\mathbf{x}|\theta)\pi(\theta)\, d\theta}$$

- On note $\hat{\theta}_B$, l'estimateur bayésien.

Exemple 1 : Cas Gaussien

- $X \sim \mathcal{N}(\mu, \sigma^2)$ avec σ^2 connue.
- Loi a priori sur μ est $\pi(\mu) = \mathcal{N}(\mu_0, \tau^2)$
- La loi a posteriori est

$$\mathcal{N}(\frac{n\tau^2}{n\tau^2+\sigma^2}\bar{x}_n + \frac{\sigma^2}{n\tau^2+\sigma^2}\mu_0, \frac{\sigma^2\tau^2}{n\tau^2+\sigma^2})$$

- L'estimateur bayésien est

$$\hat{\theta}_B = \frac{n\tau^2}{n\tau^2+\sigma^2}\bar{x}_n + \frac{\sigma^2}{n\tau^2+\sigma^2}\mu_0$$

Exemple 2 : Cas Bernoulli

- X suit une loi de Bernoulli de paramètre θ
- Loi a priori : loi Beta, notée $\pi(\theta) = \mathcal{B}(a, b)$ de densité

$$\pi(\theta) = \frac{1}{B(a, b)} \theta^{a-1}(1-\theta)^{b-1}$$

où $B(a, b) = \dfrac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$ et $\Gamma(x) = \int_0^{+\infty} t^{x-1} e^{-t}\, dt$

- La loi a posteriori est une loi Beta :

$$\theta | (x_1, \dots, x_n) \sim \mathcal{B}(a + \sum_{i=1}^{n} x_i, n - \sum_{i=1}^{n} x_i + b)$$

- L'estimateur bayésien est :

$$\hat{\theta}_B = \frac{a + \sum_{i=1}^{n} x_i}{a + b + n}$$

Remarques

- Si l'on pense que la vraisemblance possède plusieurs maxima locaux, on peut choisir le mode a posteriori comme estimateur bayésien :
$$\hat{\theta}_B = argmax_{\theta \in \Theta} \pi(\theta|\mathbf{x})$$
- Dans notre cas, on ne prendra que la moyenne a posteriori comme estimateur bayésien.
- L'estimation bayésienne d'une fonction $h(\theta)$ est $\mathbb{E}[h(\theta)|\mathbf{x}]$, où l'espérance est définie à partir de la loi a posteriori de θ.

Exemple 3 : Cas Poisson

- $X \sim \mathcal{P}(\theta)$ et la loi a priori est $\theta \sim Gamma(a, b)$.
- La loi a posteriori $\pi(\theta)$ est la loi $Gamma(a + \sum_{i=1}^{n} x_i, \frac{b}{bn+1})$
- L'estimateur bayésien est donc

$$\hat{\theta}_B = (\frac{a}{n} + \bar{x}_n)\frac{bn}{bn+1}$$

Remarques

- Dans la plupart des exemples précédents, les lois a posteriori sont de la même nature que les lois a priori : On parle de lois a priori conjuguées.
- Exemple : Soit $X \sim \mathcal{E}(\lambda)$. Montrer que la loi Gamma est une loi a priori conjuguée.
- Les lois a priori étaient toutes informatives. On se propose dans la suite de définir des lois a priori non informatives (par exemple une loi uniforme pour le cas Bernoulli)

Loi a priori impropre

- Soit $\pi(\theta)$ une fonction positive telle que $\int_\Theta \pi(\theta)\, d\theta = +\infty$
- On parle de loi impropre, mais on suppose que $m(\mathbf{x}) = \int_\Theta L(\theta; \mathbf{x})\pi(\theta)\, d\theta$ est convergente.
- On considère alors la loi a posteriori

$$\pi(\theta|\mathbf{x}) = \frac{L(\theta; \mathbf{x})\pi(\theta)}{\int_\Theta L(\theta; \mathbf{x})\pi(\theta)\, d\theta}$$

- L'estimateur de Bayes est alors la moyenne a posteriori

$$\hat{\theta}_B = \int_\Theta \theta\pi(\theta|\mathbf{x})\, d\theta$$

Exemple 1 (1/3)

- On suppose que $X \sim \mathcal{B}(\theta)$ et on considère la loi impropre suivante sur $]0,1[$:

$$\pi(\theta) = \frac{1}{\theta(1-\theta)}$$

- Calculer l'estimateur de Bayes. A quelle condition est-il défini ?
- Le comparer à l'estimateur du maximum de vraisemblance.

Exemple 1 (2/3)

- Soit $s = \sum_{i=1}^{n} x_i$, la vraisemblance s'écrit

$$L(\mathbf{x};\theta) = \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i} = \theta^s(1-\theta)^{n-s}$$

- L'intégrale $\int_0^1 \theta^{s-1}(1-\theta)^{n-s-1}\,d\theta$ converge ssi $1-s<1$ et $s+1-n<1$, c'est-à-dire $0<s<n$.
- La loi a posteriori est

$$\pi(\theta|\mathbf{x}) \propto \theta^{s-1}(1-\theta)^{n-s-1}$$

Exemple 1 (3/3)

- La loi a posteriori est donc une loi Beta, $\mathcal{B}(s, n-s)$
- L'estimateur bayésien est la moyenne a posteriori

$$\hat{\theta}_B = \frac{s}{s+n-s} = \bar{x}_n = \hat{\theta}_{MLE}$$

Exemple 2 (1/3)

- On considère $X \sim \mathcal{N}(\theta, 1)$ et on considère la loi impropre suivante sur $\mathbb{R}$: $\pi(\theta) = 1$
- Calculer l'estimateur de Bayes. A quelle condition est-il défini ?
- Le comparer à l'estimateur du maximum de vraisemblance.

Exemple 2 (2/3)

- La vraisemblance s'écrit

$$
\begin{aligned}
L(\mathbf{x};\theta) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}(x_i-\theta)^2) \\
&= (\frac{1}{2\pi})^{n/2} \exp(-\frac{1}{2}\sum_{i=1}^{n}(x_i-\theta)^2)
\end{aligned}
$$

- L'estimation par maximum de vraisemblance est $\hat{\theta}_{MLE} = \bar{x}_n$
- L'intégrale $\int_{\mathbb{R}} L(\mathbf{x};\theta)\, d\theta$ est toujours convergente
- La loi a posteriori est (avec $s = \sum_{i=1}^{n} x_i$)

$$\pi(\theta|\mathbf{x}) \propto \exp(-\frac{1}{2}(n\theta^2 - 2s\theta))$$

Exemple 2 (3/3)

- La loi a posteriori est $\mathcal{N}(\bar{x}_n; 1/n)$
- L'estimateur bayésien est la moyenne a posteriori :

$$\hat{\theta}_B = \bar{x}_n = \hat{\theta}_{MLE}$$

Loi a priori non informative de Jeffreys

- Soit $\theta \in \mathbb{R}$. On appelle loi de Jeffreys, la loi (éventuellement impropre) de densité à support sur Θ :

$$\pi_J(\theta) \propto (I(\theta))^{1/2}$$

  où $I(\theta)$ est l'information de Fisher.
- L'information de Fisher est définie de la façon suivante

$$I(\theta) = E[-\frac{\partial^2}{\partial\theta^2} log(L(X;\theta)))|\theta]$$

  où $L(X;\theta)$ est la fonction de vraisemblance.
- Cette quantité représente la quantité d'information sur θ apportée par X.

Exemple : Bernoulli de paramètre θ

- On suppose que $X_i \sim \mathcal{B}(\theta)$.
- La dérivée seconde est égale à

$$\frac{\partial^2}{\partial\theta^2} log(L(X;\theta))) = -\frac{X}{\theta^2} - \frac{n-X}{(1-\theta)^2}$$

  où $X = \sum_{i=1}^n X_i$ suit une loi binomiale, notée $\mathcal{B}(n,\theta)$.
- $I(\theta) = \frac{n}{\theta} + \frac{n}{(1-\theta)} \propto \frac{1}{\theta(1-\theta)}$.
- La loi de Jeffreys est donc $\pi(\theta) = (\theta(1-\theta))^{-1/2}$, c'est-à-dire une loi Beta, $Beta(1/2, 1/2)$.

Intervalles de crédibilité

- On se place dans un cadre bayésien et on suppose que $\theta \in \mathbb{R}$.

Définition
Soit $\alpha \in]0,1[$*. Un intervalle de crédibilité I de niveau de confiance* $1-\alpha$ *est défini par*

$$P(\theta \in I|\mathbf{x}) = \int_I \pi(\theta|\mathbf{x})\, d\theta$$

Exemple 1 : Bernoulli

- On suppose que pour tout i, $X_i \sim B(\theta)$ et que la loi a priori est : $\theta \sim U_{[0,1]}$
- On suppose que l'on a observé 20 succès parmi 60 essais
- La loi a posteriori est une loi Beta, $Beta(s+1, n-s+1)$
- L'intervalle de crédibilité avec un niveau de confiance de 95% (risque de 5%) est

$$P(Beta(s+1, n-s+1) \in [q_{0.025}; q_{0.975}]) = 0.95$$

où $q_{0.025}$ et $q_{0.975}$ sont respectivement les quantiles de la loi $Beta(s+1, n-s+1)$ d'ordre 0.025 et 0.975.

Application en R

```r
s<-20;n<-60;moy.a.post<-(s+1)/(n+2)
moy.a.post
```

```
## [1] 0.3387097
```

```r
borne.inf<-qbeta(0.025,s+1,n-s+1)
borne.sup<-qbeta(0.975,s+1,n-s+1)
int<-c(borne.inf,borne.sup)
int
```

```
## [1] 0.2272576 0.4600191
```

Exemple 2 : Normale

- On suppose que pour tout i, $X_i \sim \mathcal{N}(\theta, 1)$ et que la loi a priori est : $\pi(\theta) = 1$
- La loi a posteriori est une loi Normale, $\mathcal{N}(\bar{x}_n, 1/n)$
- L'intervalle de crédibilité avec un niveau de confiance de 95% (risque de 5%) est

$$P(\hat{\theta}_B \in [q_{0.025}; q_{0.975}]) = 0.95$$

  où $q_{0.025}$ et $q_{0.975}$ sont respectivement les quantiles de la loi $\mathcal{N}(\bar{x}_n, 1/n)$ d'ordre 0.025 et 0.975.
- A comparer avec l'intervalle de confiance classique : $[\bar{x}_n - 1.96/\sqrt{n}; \bar{x}_n + 1.96/\sqrt{n}]$

Application en R

```r
theta<-2;n<-40;
ech<-rnorm(n,theta,1)
moy.a.post<-mean(ech)
borne.inf<-qnorm(0.025,mean(ech),sqrt(1/n))
borne.sup<-qnorm(0.975,mean(ech),sqrt(1/n))
int<-c(borne.inf,borne.sup)
int
```

```
## [1] 1.702825 2.322620
```