Introduction à l'analyse bayésienne

Franck Corset

Master 2 - SSD

Principes de l'inférence bayésienne

Figure 1:Sir Thomas Bayes

Rappels : Modèle classique - Analyse fréquentiste (1/3)

- On se place dans un espace probabilisé paramétrique $(\mathcal{X}, \mathcal{B}, P_\theta)$ où $\theta \in \Theta$ est le vecteur des paramètres du modèle, $\mathcal{X}$ est l'espace des données et Θ est l'espace des paramètres.
- Le but de l'analyse statistique est de faire l'inférence sur le paramètre θ. Si on considère que le modèle est dominé par une mesure μ si P_θ admet une densité par rapport à μ :

$$f(X|\theta) = \frac{dP_\theta}{d\mu}$$

Rappels : Modèle classique - Analyse fréquentiste (2/3)

- Cette fonction, notée $\ell(\theta) = f(X|\theta)$, vue comme une fonction de θ, une fois qu'on a observé un tirage de X, est appelée vraisemblance du modèle. C'est la loi conditionnelle de X conditionnellement à θ.
- A partir de cette équation, il est possible de calculer l'estimateur du maximum de vraisemblance (EMV ou MLE en anglais), noté $\hat{\theta}_{ML}$, valeur qui maximise la fonction de vraisemblance.

Rappels : Modèle classique - Analyse fréquentiste (3/3)

- Ainsi, si l'on se place dans le cadre d'un échantillon i.i.d, $(X_1, \ldots, X_n)$ de même loi que X, ayant pour densité $f(x|\theta)$, où θ est paramètre inconnu que l'on cherche à estimer.
- Sachant que l'on a observé un échantillon $(x_1, \ldots, x_n)$, vu comme la réalisation de l'échantillon i.i.d. $(X_1, \ldots, X_n)$, la fonction de vraisemblance s'écrit

$$L(\theta|(x_1, \ldots, x_n)) = \prod_{i=1}^{n} f(x_i|\theta) \tag{1}$$

Analyse bayésienne

Introduction

- On va ici considérer que le paramètre θ est aléatoire.
- L'espace des paramètres Θ est muni d'une probabilité π tel que $(\Theta, \mathcal{A}, \pi)$ soit un espace probabilisé.
- On note $\theta \sim \pi$.
- π est appelé loi a priori. Cette loi détermine l'information a priori que l'on a sur le phénomène étudié, c'est-à-dire avant d'avoir observé X.
- Généralement, cette loi dépend elle-même de paramètres. Ces paramètres sont appelés hyperparamètres du modèle.

Loi jointe et loi a posteriori

- La loi jointe (X, θ) s'écrit

$$\lambda(X, \theta) = f(X|\theta) d\pi(\theta) = f(X|\theta)\pi(\theta) d\nu$$

  où ν est la mesure de Lebesgue.

- La loi a posteriori est définie par sa densité

$$d\pi(\theta|X) = \frac{f(X|\theta) d\pi(\theta)}{\int_\Theta f(X|\theta) d\pi(\theta)}$$

- La quantité $m(X) = \int_\Theta f(X|\theta) d\pi(\theta)$ est la loi marginale de X et est une constante de normalisation de la loi a posteriori, indépendante de θ.
- Par la suite, on notera $\pi(\theta|X) \propto f(X|\theta)\pi(\theta)$

Exemple : Cas gaussien (1/4)

- On suppose que $X \sim \mathcal{N}(\mu, \sigma^2)$ avec σ^2 connue.
- On chercher à inférer sur $\theta = \mu$
- On suppose que la loi a priori sur μ est $\pi(\mu) = \mathcal{N}(\mu_0, \tau^2)$
- μ_0 et τ^2 sont appelés hyperparamètres du modèle

Exemple : Cas gaussien (2/4)

- On suppose que l'on dispose d'un échantillon i.i.d. $(x_1, \ldots, x_n)$, réalisation de $(X_1, \ldots, X_n)$
- La vraisemblance du modèle est

$$l(\mu) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2}$$

- L'estimation par maximum de vraisemblance donne $\hat{\mu} = \bar{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i$

Exemple : Cas gaussien (3/4)

- La loi a posteriori $\pi(\mu|(x_1,\dots,x_n))$ est

$$\frac{\frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2} \frac{1}{\sqrt{2\pi\tau^2}} e^{-\frac{1}{2\tau^2}(\mu-\mu_0)^2}}{C}$$

  où C est la constante de normalisation.

- En ne gardant que les termes dépendants de μ, on a

$$\pi(\mu|(x_1,\dots,x_n)) \propto \exp\big(-\frac{n\mu^2}{2\sigma^2} - \frac{\mu^2}{2\tau^2} + \frac{\mu\sum_i x_i}{\sigma^2} + \frac{\mu\mu_0}{\tau^2}\big)$$

$$\propto \exp\big(-\frac{1}{2\frac{\sigma^2\tau^2}{n\tau^2+\sigma^2}}[\mu - (\frac{n\tau^2}{n\tau^2+\sigma^2}\bar{x}_n + \frac{n\tau^2}{n\tau^2+\sigma^2}\mu_0)]^2\big)$$

Exemple : Cas gaussien (4/4)

- 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'espérance est une combinaison convexe de la moyenne empirique et de la moyenne a priori
- Quand $n \to +\infty$, l'espérance tend vers $\bar{x}_n$ (et donc vers l'estimation par maximum de vraisemblance) et la variance vers 0.
- Plus on possède de données, moins l'a priori est influent !

Exemple 2 : Cas Bernoulli (1/3)

- On note θ la probabilité qu'un patient soit guéri à l'issu d'un traitement. Pour estimer ce paramètre, on dispose d'un échantillon de n patients. On note $x_i \in \{0, 1\}$, l'état du patient après avoir subi le traitement ($x_i = 1$ si le patient est guéri et $x_i = 0$ sinon).
- On suppose que l'échantillon $(x_1, \ldots, x_n)$ est la réalisation d'un échantillon i.i.d de n variables aléatoires.
- On suppose que X_i suit une loi de Bernoulli de paramètre θ
- La vraisemblance s'écrit

$$\ell(\theta) = \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i} = \theta^s(1-\theta)^{n-s}$$

où $s = \sum_{i=1}^{n} x_i$.

Exemple 2 : Cas Bernoulli (2/3)

- L'estimation par maximum de vraisemblance est

$$\hat{\theta} = \frac{s}{\theta} = \bar{x}_n$$

- Comme $\theta \in [0, 1]$, on propose la 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 $\Gamma(x) = \int_0^{+\infty} t^{x-1} e^{-t}\, dt$

Exemple 2 : Cas Bernoulli (3/3)

- La loi a posteriori est donc

$$\pi(\theta|(x_1,\ldots,x_n)) = \frac{\theta^{s+a-1}(1-\theta)^{n-s+b-1}}{\int_0^1 \theta^{s+a-1}(1-\theta)^{n-s+b-1}\,d\theta}$$

- La loi a posteriori est donc une loi Beta :

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

Exercice

- Soit le modèle statistique bayésien suivant : $X_i \sim \mathcal{P}(\theta)$ et $\theta \sim Gamma(a, b)$. Déterminer la loi a posteriori.
- La loi de Poisson de paramètre θ est

$$\forall k \in \mathbb{N},\ P(X = k) = \frac{\theta^k}{k!} e^{-\theta}$$

- La loi $Gamma(a, b)$ est caractérisée par sa densité :

$$f(x) = \frac{x^{a-1} e^{-x/b}}{\Gamma(a) b^a}$$

Application et illustration en R

- Mettre en oeuvre ces deux situations dans le logiciel R.
- Faire varier les différents paramètres des modèles.
- Comparer les lois a posteriori et les lois a priori sur un même graphique.