Statistiques mathématiques : cours 9

Guillaume Lecué

17 septembre 2018

Aujourd'hui

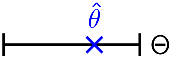
Présentation des statistiques Bayésiennes

Fréquentistes / Bayésiens

<u>Donnée</u> : X <u>Modèle</u> : $\{\mathbb{P}_\theta : \theta \in \Theta\}$

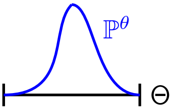
1. fréquentistes (ce que nous avons fait jusqu'ici) : à une donnée X est renvoyée un **élément de** Θ, appelé estimateur.

$$X \rightsquigarrow \hat{\theta} = \hat{\theta}(X) \in \Theta$$



2. Bayésiens (programme d'aujourd'hui) : à une donnée X est renvoyée une **mesure de probabilité sur** Θ.

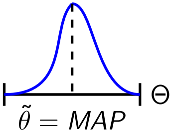
$$X \rightsquigarrow \mathbb{P}^\theta = \hat{\mathbb{P}}(X)^\theta = \text{distribution sur } \Theta$$



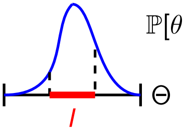
Statistiques Bayésiennes : interprétation

- Les Bayésien utilisent les données pour construire une distribution $\mathbb{P}^\theta$ sur Θ.
- Comment interpréter un résultat sous forme de mesure de probabilité ?

Les Bayésiens peuvent ne renvoyer qu'une seule valeur d'estimation de θ : le Maximum a posteriori (MAP) ou la moyenne a posteriori ou la médiane a posteriori, etc.



mais en plus ils peuvent donner des niveaux d'incertitude



$\mathbb{P}[\theta \in I] \geq 0.95$

Statistiques Bayésiennes : formalisme

L'objectif est de munir l'espace Θ d'une mesure de probabilité $\Rightarrow$

$$\boxed{\text{Le paramètre } \theta \text{ devient donc une variable aléatoire}}$$

On dispose alors d'un couple de variable aléatoires

$$(X, \theta) = (\text{donnée, paramètre})$$

ayant deux marginales (naturelles) :

- $\mathbb{P}^{X|\theta} = \mathbb{P}_\theta$: à θ fixé la loi de X est donnée par la loi du modèle $\{\mathbb{P}_\theta : \theta \in \Theta\}$ pour cette valeur de θ ! cette marginale est donc donnée par le modèle.
- $\mathbb{P}^{\theta|X}$: loi du paramètre θ étant donnée l'observation $X \rightsquigarrow$ c'est la loi qu'on cherche à construire sur Θ ! Pour la connaître, il faut soit connaître la loi du couple (X, θ) ou connaître les loi $\mathbb{P}^{X|\theta}$ et $\mathbb{P}^\theta$ et appliquer la formule de Bayes.

Formule de Bayes (1/2)

Etant donné deux événements A et B, on a (par définition)

$$\mathbb{P}[A|B] = \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[B]}$$

De même,

$$\mathbb{P}[B|A] = \frac{\mathbb{P}[A \cap B]}{\mathbb{P}[A]}$$

On en déduit la formule de Bayes :

$$\boxed{\mathbb{P}[A|B] = \frac{\mathbb{P}[B|A]\mathbb{P}[A]}{\mathbb{P}[B]}}$$

Ecrite pour des lois admettant des densités, on a

$$\boxed{\mathcal{L}(\theta|X = x_0)(\theta_0) = \frac{\mathcal{L}(X|\theta = \theta_0)(x_0) \cdot \mathcal{L}(\theta)(\theta_0)}{\mathcal{L}(X)(x_0)}}$$

où $\mathcal{L}(Z)(z)$ est “la valeur” de la densité de la loi de Z en z.

Formule de Bayes (2/2)

$$\boxed{\mathcal{L}(\theta|X = x_0)(\theta_0) = \frac{\mathcal{L}(X|\theta = \theta_0)(x_0) \cdot \mathcal{L}(\theta)(\theta_0)}{\mathcal{L}(X)(x_0)}}$$

- $\mathcal{L}(X|\theta = \theta_0)(x_0)$ est <u>connue</u> : c'est la valeur de la densité de la loi $\mathbb{P}_{\theta_0}$ en x_0 du modèle $\{\mathbb{P}_\theta : \theta \in \Theta\}$,
- $\mathcal{L}(\theta)(\theta_0)$ est <u>inconnue</u> : on va devoir se donner cette valeur <u>**a priori**</u>,
- $\mathcal{L}(X)(x_0)$ (terme indépendant de θ_0) est vu comme une constante de normalisation de la densité

$$\theta_0 \to \mathcal{L}(\theta|X = x_0)(\theta_0) \propto \mathcal{L}(X|\theta = \theta_0)(x_0) \cdot \mathcal{L}(\theta)(\theta_0)$$

donnée par

$$\mathcal{L}(X)(x_0) = \int_\Theta \mathcal{L}(X|\theta = \theta_0)(x_0) \cdot \mathcal{L}(\theta)(\theta_0) d\theta_0$$

(donc calculable à partir de $\mathcal{L}(X|\theta)$ et $\mathcal{L}(\theta)$).

Loi a priori et loi a posteriori

Definition
Une loi a priori est une mesure de probabilité $\mathbb{P}^\theta$ sur l'espace Θ qui est choisie avant l'observation des données.

En statistiques Bayésienne, on se donne donc (avant toute observation) :

- un modèle $\{\mathbb{P}^{X|\theta} : \theta \in \Theta\}$ (⚠ : $\mathbb{P}^{X|\theta}$ est une loi conditionnelle car θ est maintenant une variable aléatoire en statistiques Bayésiennes)
- une loi a priori $\mathbb{P}^\theta$ sur Θ.

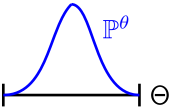
Une fois l'observation $X = x_0$ observée, on calcul la loi $\mathbb{P}^{\theta|X=x_0}$ (grâce à la formule de Bayes) :

$$\mathcal{L}(\theta|X = x_0)(\theta_0) \propto \underbrace{\mathcal{L}(X|\theta = \theta_0)(x_0)}_{\text{vraisemblance}} \cdot \underbrace{\mathcal{L}(\theta)(\theta_0)}_{\text{loi a priori}} .$$

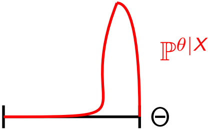
Definition
La loi $\mathcal{L}(\theta|X)$ est appelée loi a posteriori.

Méthode en statistiques Bayésiennes

1. On se donne une **loi a priori** sur Θ



2. On observe une **donnée** X
3. On modifie notre a priori sur la loi suivie par le paramètre θ en munissant Θ d'une nouvelle loi : la **loi a posteriori**



Une fois la loi a posteriori calculée (généralement à la constante absolue près $\mathcal{L}(X)(X)$ – où le "deuxième X" est la donnée alors que $\mathcal{L}(X)$ est la loi de X), on peut faire de <u>l'inférence sur θ</u>.

Calcul de loi a posteriori : exemple (1/2)

On considère :

1. le modèle $\{\mathbb{P}^{X|\theta} : \theta \in \mathbb{R}\}$ où $\mathbb{P}^{X|\theta} = \mathcal{N}(\theta, \sigma^2)$,
2. la loi a priori $\mathbb{P}^{\theta} = \mathcal{N}(\mu, \tau^2)$.

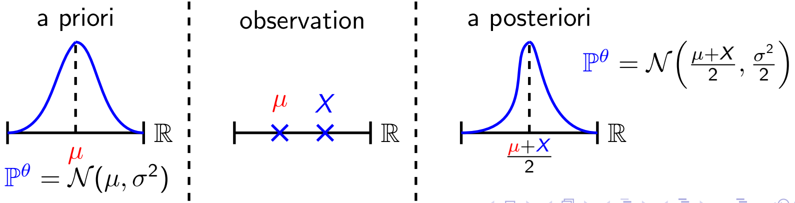
- Montrer que la loi a posteriori est

$$\mathcal{L}(\theta|X) \sim \mathcal{N}\left(\frac{\tau^2 X + \sigma^2\mu}{\tau^2 + \sigma^2}, \frac{\sigma^2\tau^2}{\tau^2 + \sigma^2}\right)$$

- Montrer que le *Maximum a posteriori*, la moyenne a posteriori et la médiane a posteriori sont tous égaux à

$$\frac{\tau^2 X + \sigma^2\mu}{2(\tau^2 + \sigma^2)}.$$

Quand $\tau = \sigma$:



Calcul de loi a posteriori : exemple (2/2)

On considère :

1. le modèle d'échantillonnage $X_1|\theta, \ldots, X_n|\theta \sim_{i.i.d.} X|\theta$ où la loi de X sachant θ a sa loi dans le modèle $\{\mathbb{P}^{X|\theta} : \theta \in \mathbb{R}\}$ où $\mathbb{P}^{X|\theta} = \mathcal{N}(\theta, 1)$,
2. la loi a priori $\mathbb{P}^{\theta} = \mathcal{N}(0, 1)$.

- la loi a posteriori est

$$\mathcal{L}(\theta|X_1, \ldots, X_n) \sim \mathcal{N}\left(\frac{X_1 + \cdots + X_n}{n+1}, \frac{1}{n+1}\right)$$

- le *Maximum a posteriori*, la *moyenne a posteriori* et la *médiane a posteriori* sont tous égaux à

$$\frac{X_1 + \ldots + X_n}{n+1}.$$

- un *intervalle de confiance Bayésien à* 95% est donné par

$$\left[\frac{X_1 + \ldots + X_n}{n+1} - \frac{1.96}{\sqrt{n+1}}, \frac{X_1 + \ldots + X_n}{n+1} + \frac{1.96}{\sqrt{n+1}}\right]$$

Propriétés asymptotiques de la loi a posteriori I

Rappels (fréquentistes) : Dans le modèle d'échantillonnage associé à un modèle régulier on a :

1. **L'EMV est consistant** : pour tout $\theta \in \Theta$,
$$\widehat{\theta}_{\mathrm{n}}^{\mathtt{mv}} \xrightarrow{\mathbb{P}_\theta} \theta.$$
2. **l'EMV est asymptotiquement normal** : pour tout $\theta \in \Theta$,
$$\sqrt{n}\left(\widehat{\theta}_{\mathrm{n}}^{\mathtt{mv}} - \theta\right) \xrightarrow{d} \mathcal{N}\left(0, \frac{1}{\mathbb{I}(\theta)}\right)$$
où $\mathbb{I}(\theta) = \mathbb{E}_\theta\left[\left(\partial_1 \log f(\theta, X)\right)^2\right]$ s'appelle l'information de Fisher de la famille $\{\mathbb{P}_\theta = f(\theta, \cdot) \cdot \mu : \theta \in \Theta\}$ au point θ.

Il existe des notions similaires en Statistiques Bayésiennes.

Propriétés asymptotiques de la loi a posteriori II

Dans le modèle d'échantillonnage :

1. $(X_n|\theta)_n$ une suite de variables aléatoires i.i.d. distribuées selon $X|\theta$ dont la loi (de X conditionnellement à θ) est dans le modèle $\{\mathbb{P}^{X|\theta} : \theta \in \Theta\}$
2. $\mathbb{P}^{\theta}$ une loi a priori sur Θ (de densité $\mathcal{L}(\theta)$)
3. la loi a posteriori $\Pi(\cdot|(X_i)_{i=1}^n)$ a pour densité

$$\begin{aligned}\theta_0 \to \mathcal{L}(\theta|(X_i)_{i=1}^n)(\theta_0) &\propto \text{ vraisemblance } \times \text{ loi a priori} \\ &\propto \mathcal{L}\left((X_i)_{i=1}^n|\theta=\theta_0\right)\left((X_i)_{i=1}^n\right) \times \mathcal{L}(\theta)(\theta_0).\end{aligned}$$

Definition

On dit que la suite de loi a posteriori $\left(\Pi(\cdot|(X_i)_{i=1}^n)\right)_n$ est **consistante** quand pour tout $\theta \in \Theta$ et tout voisinage U de θ, on a, sous $\mathbb{P}_\theta$,

$$\Pi\big(\theta \in U|(X_i)_{i=1}^n\big) \xrightarrow{\text{p.s.}} 1.$$

Rem. : Dans cette définition les X_i sont supposés de loi $\mathbb{P}_\theta$ pour un θ fixé (càd après avoir calculé la loi a posteriori, on revient au modèle fréquentiste en supposant que $X_i \sim \mathbb{P}_\theta$).

Exemple de consistance

1. Modèle d'échantillonnage de binomiale : $\mathbb{P}^{X|\theta} = \mathrm{Bin}(\theta)$ pour $\theta \in \Theta = (0,1)$
2. Loi a priori sur $\Theta = [0,1]$ est une Beta(α, β), càd de densité

$$\theta_0 \to \mathcal{L}(\theta)(\theta_0) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta_0^{\alpha-1}(1-\theta_0)^{\beta-1}.$$

Montrer que

1. la loi a posteriori suit une Beta$(S_n + \alpha, n - S_n + \beta)$ où $S_n = \sum_{i=1}^n X_i$, càd, pour $X_1^n = (X_i)_{i=1}^n$ de densité

$$\theta_0 \to \mathcal{L}(\theta|X_1^n)(\theta_0) = \frac{\Gamma(n+2)}{\Gamma(S_n+1)\Gamma(n-S_n+1)} \theta_0^{S_n}(1-\theta_0)^{n-S_n}.$$

2. les moyenne et variance a posteriori vérifient

$$\mathbb{E}[\theta|X_1^n] = \frac{S_n+\alpha}{n+\alpha+\beta} \text{ et } \mathrm{var}(\theta|X_1^n) = \frac{(S_n+\alpha)(n-S_n+\beta)}{(n+\alpha+\beta)^2(n+\alpha+\beta+1)}.$$

3. la suite des loi a posteriori est consistante.

Propriétés asymptotiques de la loi a posteriori III

Dans le modèle d'échantillonnage, $(X_n)_n$ une suite de variables aléatoires i.i.d. distribuées selon une loi dans le modèle $\{\mathbb{P}^{X|\theta} : \theta \in \Theta\}$

- π_1 une loi a priori associée à la loi a posteriori $\Pi_1(\cdot|(X_i)_{i=1}^n)$
- π_2 une loi a priori associée à la loi a posteriori $\Pi_2(\cdot|(X_i)_{i=1}^n)$

Si $\left(\Pi_1(\cdot|(X_i)_{i=1}^n)\right)_n$ et $\left(\Pi_2(\cdot|(X_i)_{i=1}^n)\right)_n$ sont consistants alors pour tout $\theta \in \Theta$, sous $\mathbb{P}_\theta$,

$$\sup_A \left|\Pi_1(A|(X_i)_{i=1}^n) - \Pi_2(A|(X_i)_{i=1}^n)\right| \xrightarrow{\text{p.s.}} 0.$$

Asymptotiquement le choix de la loi a priori n'a pas d'importance

Propriétés asymptotiques de la loi a posteriori IV

Dans un modèle régulier la loi a posteriori se comporte asymptotiquement comme

$$\Pi(\cdot|(X_i)_{i=1}^n) \approx \widehat{\theta}_{\mathrm{n}}^{\mathtt{mv}} + \frac{1}{\sqrt{n}}\mathcal{N}\left(0, \mathbb{I}(\widehat{\theta}_{\mathrm{n}}^{\mathtt{mv}})^{-1}\right) \underset{\text{sous } \mathbb{P}_{\theta_0}}{\approx} \theta_0 + \frac{1}{\sqrt{n}}\mathcal{N}\left(0, \mathbb{I}(\theta_0)^{-1}\right)$$

où

- $\widehat{\theta}_{\mathrm{n}}^{\mathtt{mv}} \in \mathrm{argmax}_{\theta_1 \in \Theta} \mathcal{L}\left((X_i)_{i=1}^n | \theta = \theta_1\right) \left((X_i)_{i=1}^n\right)$
- $\mathbb{I}(\widehat{\theta}_{\mathrm{n}}^{\mathtt{mv}})$ est l'information de Fisher en $\widehat{\theta}_{\mathrm{n}}^{\mathtt{mv}}$ pour une observation.

Théorème

Dans un modèle régulier et pour une loi a priori continue et positive en $\theta_0 \in \Theta \subset \mathbb{R}$, on a

$$\int_{\mathbb{R}} \left| \mathcal{L}(\sqrt{n}(\theta - \widehat{\theta}_{\mathrm{n}}^{\mathtt{mv}}) | X_1^n)(t) - \sqrt{\frac{\mathbb{I}(\widehat{\theta}_{\mathrm{n}}^{\mathtt{mv}})}{2\pi}} \exp(-t^2 \mathbb{I}(\widehat{\theta}_{\mathrm{n}}^{\mathtt{mv}})/2)) \right| dt \xrightarrow{\text{p.s.}} 0$$

Estimateur de Bayes

Etant donné une fonction de perte

$$\ell : \Theta \times \Theta \to \mathbb{R}^+$$

(par example, $\ell(\theta, \beta) = (\theta - \beta)^2$), l'estimateur de Bayes est

$$\hat{\theta}_b(x) \in \text{argmin}_{\beta \in \Theta} \int_{\theta \in \Theta} \ell(\theta, \beta) d\mathbb{P}^{\theta|X=x}(\theta)$$

Par exemple, pour la perte quadratique, l'estimateur de Bayes est la moyenne a posteriori :

$$\hat{\theta}_b(x) = \mathbb{E}\big[\theta|X = x\big].$$