Statistique Bayésienne

Anna Simoni[2]

[2]CREST - Ensae and CNRS

Outline

Définition (Modèle classique)

On se place dans un espace probabilisé paramétrique classique :

$$(\mathscr{X}; \mathscr{B}; \{P_\theta, \theta \in \Theta\})$$

où $\mathscr{X}$ désigne l'espace des données, Θ celui des paramètres θ. Le but de l'analyse statistique est de faire de l'inférence sur θ, c'est-à-dire décrire un phénomène passé ou à venir dans un cadre probabiliste.

Un modèle Bayésien est defini par un modèle mesurable équipé d'une mesure μ sur $(\Theta; \mathscr{A})$ (*distribution a priori*).

Définition (Modèle Bayésien)

Le modèle Bayésien est defini par l'espace probabilisé paramétrique suivant :

$$(\Theta \times \mathscr{X}; \mathscr{A} \otimes \mathscr{B}; \Pi)$$

où $\Pi = \mu \otimes P_\theta = P \otimes \mu^x$, P est la probabilité marginale sur $(\mathscr{X}, \mathscr{B})$ et μ^x est une version regulière de la probabilité conditionnelle sur Θ étant donné $\mathscr{X}$ (i.e. probabilité de transition).

Définition (Modèle classique)

On se place dans un espace probabilisé paramétrique classique :

$$(\mathscr{X}; \mathscr{B}; \{P_\theta, \theta \in \Theta\})$$

où $\mathscr{X}$ désigne l'espace des données, Θ celui des paramètres θ. Le but de l'analyse statistique est de faire de l'inférence sur θ, c'est-à-dire décrire un phénomène passé ou à venir dans un cadre probabiliste.

Un modèle Bayésien est defini par un modèle mesurable équipé d'une mesure μ sur $(\Theta; \mathscr{A})$ (*distribution a priori*).

Définition (Modèle Bayésien)

Le modèle Bayésien est defini par l'espace probabilisé paramétrique suivant :

$$(\Theta \times \mathscr{X}; \mathscr{A} \otimes \mathscr{B}; \Pi)$$

où $\Pi = \mu \otimes P_\theta = P \otimes \mu^x$, P est la probabilité marginale sur $(\mathscr{X}, \mathscr{B})$ et μ^x est une version regulière de la probabilité conditionnelle sur Θ étant donné $\mathscr{X}$ (i.e. probabilité de transition).

L'inférence bayésienne. I

Définition (Modèle classique)

On se place dans un espace probabilisé paramétrique classique :

$$(\mathscr{X}; \mathscr{B}; \{P_\theta, \theta \in \Theta\})$$

où $\mathscr{X}$ désigne l'espace des données, Θ celui des paramètres θ. Le but de l'analyse statistique est de faire de l'inférence sur θ, c'est-à-dire décrire un phénomène passé ou à venir dans un cadre probabiliste.

Un modèle Bayésien est defini par un modèle mesurable équipé d'une mesure μ sur $(\Theta; \mathscr{A})$ (*distribution a priori*).

Définition (Modèle Bayésien)

Le modèle Bayésien est defini par l'espace probabilisé paramétrique suivant :

$$(\Theta \times \mathscr{X}; \mathscr{A} \otimes \mathscr{B}; \Pi)$$

où $\Pi = \mu \otimes P_\theta = P \otimes \mu^x$, P est la probabilité marginale sur $(\mathscr{X}, \mathscr{B})$ et μ^x est une version regulière de la probabilité conditionnelle sur Θ étant donné $\mathscr{X}$ (i.e. probabilité de transition).

Définition (Modèle dominé)

Le modèle est dit dominé s'il existe une mesure finie commune dominante λ, c'est-à-dire pour tout θ, P_θ admet une densité par rapport à λ : $f(X|\theta) := dP_\theta(X)/d\lambda(X)$.

Cette fonction $\ell(\theta|X) := f(X|\theta)$, vue comme une fonction de θ une fois que l'on a observé un tirage de X, est appelée *vraisemblance du modèle*.

Théorème (Théorème de Bayes)

Si A et E sont des événements tels que $P(E) \neq 0$, $P(A|E)$ et $P(E|A)$ sont reliés par

$$P(A|E) = \frac{P(E|A)P(A)}{P(E|A)P(A) + P(E|A^c)P(A^c)} = \frac{P(E|A)P(A)}{P(E)}$$

Définition (Modèle dominé)

Le modèle est dit dominé s'il existe une mesure finie commune dominante λ, c'est-à-dire pour tout θ, P_θ admet une densité par rapport à λ : $f(X|\theta) := dP_\theta(X)/d\lambda(X)$.

Cette fonction $\ell(\theta|X) := f(X|\theta)$, vue comme une fonction de θ une fois que l'on a observé un tirage de X, est appelée *vraisemblance du modèle*.

Théorème (Théorème de Bayes)

Si A et E sont des événements tels que $P(E) \neq 0$, $P(A|E)$ et $P(E|A)$ sont reliés par

$$P(A|E) = \frac{P(E|A)P(A)}{P(E|A)P(A) + P(E|A^c)P(A^c)} = \frac{P(E|A)P(A)}{P(E)}$$

Définition (Loi jointe et loi a posteriori)

La loi jointe de (X,θ) *s'écrit,* $\forall B \in \mathscr{B}$*,* $\forall A \in \mathscr{A}$ *:*

$$\Pi(B \times A) = \int_A P_\theta(B) d\mu(\theta)$$

et dans le cas d'un modèle dominé (denote $f(x|\theta) := dP_\theta / d\lambda(x)$*), alors :*

$$\Pi(B \times A) = \int_A \int_B f(x|\theta) d\lambda(x) d\mu(\theta)$$

et $\pi(x,\theta) = f(x|\theta)\mu(\theta)$*.*
La loi a posteriori est définie par sa densité :

$$\mu^x(\theta) := \mu(\theta|x) = \frac{f(x|\theta)\mu(\theta)}{\int f(x|\theta)\mu(\theta)d\theta}. \qquad (1)$$

L’inférence bayésienne. IV

- Un modèle statistique Bayésien est constitué d’un modèle statistique paramétrique, $f(x|\theta)$, et d’une distribution a priori pour les paramètres, $\mu(\theta)$.

- En termes statistiques, le théorème de Bayes actualise l’information sur θ en extrayant l’information contenue dans l’observation X.

- La quantité
$$P(X) = \int f(X|\theta)\mu(\theta)d\theta$$
est la probabilité marginale de X et est une constante de normalisation de la loi a posteriori, indépendante de θ. On peut donc travailler à une constante multiplicative près : $\mu(\theta|x) \propto f(x|\theta)\mu(\theta)$.

- La probabilité prédictive de Y où $Y \sim g(Y|\theta, x)$ est :
$P(Y|x) = \int g(Y|\theta, x)\mu(\theta|x)d\theta$.

Outline

Principes de vraisemblance et d'exhaustivité. I

Deux principes fondamentaux sont respectés par le paradigme bayésien : le principe de vraisemblance et le principe d'exhaustivité.

Définition (Exhaustivité)

Quand $X \sim f(X|\theta)$, une fonction T de X (aussi appelée statistique) est exhaustive pour $\{P_\theta, \theta \in \Theta\}$ (ou pour θ) si la distribution de X conditionnellement à $T(X)$ ne dépend pas de θ.

Théorème (Critère de factorisation)

Une condition necéssaire et suffisante pour qu'une statistique T soit exhaustive pour une famille $\{P_\theta, \theta \in \Theta\}$ de distributions de X dominées par une mesure λ est qu'ils existent deux fonctions non-negatives $g(\cdot|\theta)$ et h telles que :

$$f(x|\theta) = g(T(x)|\theta)h(x) \qquad (a.e.\ \lambda).$$

Théorème (Principe d'Exhaustivité)

Deux observations x et y donnant la même valeur d'une statistique exhaustive T, c'est-à-dire telles que $T(x) = T(y)$, doivent conduire à la même inférence sur θ.

Principes de vraisemblance et d'exhaustivité. I

Deux principes fondamentaux sont respectés par le paradigme bayésien : le principe de vraisemblance et le principe d'exhaustivité.

Définition (Exhaustivité)

Quand $X \sim f(X|\theta)$, une fonction T de X (aussi appelée statistique) est exhaustive pour $\{P_\theta, \theta \in \Theta\}$ (ou pour θ) si la distribution de X conditionnellement à $T(X)$ ne dépend pas de θ.

Théorème (Critère de factorisation)

Une condition necéssaire et suffisante pour qu'une statistique T soit exhaustive pour une famille $\{P_\theta, \theta \in \Theta\}$ de distributions de X dominées par une mesure λ est qu'ils existent deux fonctions non-negatives $g(\cdot|\theta)$ et h telles que :

$$f(x|\theta) = g(T(x)|\theta)h(x) \qquad (a.e.\ \lambda).$$

Théorème (Principe d'Exhaustivité)

Deux observations x et y donnant la même valeur d'une statistique exhaustive T, c'est-à-dire telles que $T(x) = T(y)$, doivent conduire à la même inférence sur θ.

Principes de vraisemblance et d'exhaustivité. I

Deux principes fondamentaux sont respectés par le paradigme bayésien : le principe de vraisemblance et le principe d'exhaustivité.

Définition (Exhaustivité)

Quand $X \sim f(X|\theta)$, une fonction T de X (aussi appelée statistique) est exhaustive pour $\{P_\theta, \theta \in \Theta\}$ (ou pour θ) si la distribution de X conditionnellement à $T(X)$ ne dépend pas de θ.

Théorème (Critère de factorisation)

Une condition necéssaire et suffisante pour qu'une statistique T soit exhaustive pour une famille $\{P_\theta, \theta \in \Theta\}$ de distributions de X dominées par une mesure λ est qu'ils existent deux fonctions non-negatives $g(\cdot|\theta)$ et h telles que :

$$f(x|\theta) = g(T(x)|\theta)h(x) \qquad (a.e.\ \lambda).$$

Théorème (Principe d'Exhaustivité)

Deux observations x et y donnant la même valeur d'une statistique exhaustive T, c'est-à-dire telles que $T(x) = T(y)$, doivent conduire à la même inférence sur θ.

Principes de vraisemblance et d'exhaustivité. I

Deux principes fondamentaux sont respectés par le paradigme bayésien : le principe de vraisemblance et le principe d'exhaustivité.

Définition (Exhaustivité)

Quand $X \sim f(X|\theta)$, une fonction T de X (aussi appelée statistique) est exhaustive pour $\{P_\theta, \theta \in \Theta\}$ (ou pour θ) si la distribution de X conditionnellement à $T(X)$ ne dépend pas de θ.

Théorème (Critère de factorisation)

Une condition necéssaire et suffisante pour qu'une statistique T soit exhaustive pour une famille $\{P_\theta, \theta \in \Theta\}$ de distributions de X dominées par une mesure λ est qu'ils existent deux fonctions non-negatives $g(\cdot|\theta)$ et h telles que :

$$f(x|\theta) = g(T(x)|\theta)h(x) \qquad (a.e.\ \lambda).$$

Théorème (Principe d'Exhaustivité)

Deux observations x et y donnant la même valeur d'une statistique exhaustive T, c'est-à-dire telles que $T(x) = T(y)$, doivent conduire à la même inférence sur θ.

Principes de vraisemblance et d'exhaustivité. II

Théorème (Principe de vraisemblance)

L'information apportée par une observation de x sur θ est entièrement contenue dans la fonction de vraisemblance $\ell(\theta|x)$. De plus, si x_1 et x_2 sont deux observations qui dépendent du même paramètre θ, et telles qu'il existe une constante c satisfaisant

$$\ell_1(\theta|x_1) = c\ell_2(\theta|x_2)$$

pour tout θ, elles apportent la même information sur θ et doivent conduire à la même inférence.

Le principe de vraisemblance n'est valide que lorsque

(i) l'inférence concerne le même paramètre θ et

(ii) θ prend en compte tous les facteurs inconnus du modèle.

Principes de vraisemblance et d'exhaustivité. III

- L'estimation par maximum de vraisemblance (MV) n'est qu'une façon parmi d'autres de mettre en oeuvre le principe de vraisemblance.
- Si on observe $x \sim f(x|\theta)$, l'approche par MV considère l'estimateur suivant de θ :
$$\widehat{\theta} := \arg\sup_{\theta} \ell(\theta|x). \tag{2}$$
- Proprieté de l'estimateur du MV est son *invariance par reparametrisation* : pour toute fonction $h(\theta)$, l'estimateur de MV est $h(\widehat{\theta})$ (même quand h n'est pas bijective).

La méthode de MV a aussi des défauts :

1. La maximisation de $\ell(\theta|x)$ peut être assez complexe en pratique, particulièrement dans les cas multidimensionnels ou contraints.
2. Les estimateurs du MV peuvent être numériquement instables, c'est-à-dire peuvent varier considérablement pour de petites variations des observations, du moins pour des tailles d'échantillon réduites.

Principes de vraisemblance et d'exhaustivité. IV

3. L'approche du MV n'admet pas de justifications probabiliste et décisionnelle. De fait, elle ne répond pas aux exigences d'une analyse décisionnelle et échoue ainsi à fournir des outils d'éevaluation pour les estimateurs qu'elle propose. Par exemple, il n'est pas possible de faire des tests dans un contexte de MV pur.