CentraleSupélec

Statistique et apprentissage

Chargés de cours (ordre alphabétique) :

Julien Bect, Gilles Faÿ, Ziad Kobeissi, Laurent Le Brusquet,
Vincent Lescarret, Arshak Minasyan, Arthur Tenenhaus† & Xujia Zhu

† Coordinateur du cours

Cours 5/9

Estimation bayésienne

Objectifs du cours 5

- Introduire la notion d'information a priori.
- Se familiariser avec les approches bayésiennes.
- Construire de nouveaux estimateurs intégrant un a priori.

Plan du cours

Plan du cours

Rappel : comparaison d'estimateurs

Risque quadratique : $R_\theta(\hat{\eta}) = \mathbb{E}_\theta\left(\|\hat{\eta} - g(\theta)\|^2\right)$.

Définitions

On dira que $\hat{\eta}'$ est préférable (au sens large) à $\hat{\eta}$ si

- $\forall \theta \in \Theta,\ R_\theta(\hat{\eta}') \leq R_\theta(\hat{\eta})$,

On dira qu'il est strictement préférable à $\hat{\eta}$ si, de plus,

- $\exists \theta \in \Theta,\ R_\theta(\hat{\eta}') < R_\theta(\hat{\eta})$,

Remarques

- La relation « préférable à » est un ordre partiel sur les risques.
- Il n'existe pas en général un estimateur optimal, càd un estimateur préférable à tous les autres (sauf à restreindre la classe d'estimateurs considérés).

Rappel : comparaison d'estimateurs

Risque quadratique : $R_\theta(\hat{\eta}) = \mathbb{E}_\theta\left(\|\hat{\eta} - g(\theta)\|^2\right)$.

Définitions

On dira que $\hat{\eta}'$ est préférable (au sens large) à $\hat{\eta}$ si

- $\forall \theta \in \Theta,\ R_\theta(\hat{\eta}') \leq R_\theta(\hat{\eta})$,

On dira qu'il est strictement préférable à $\hat{\eta}$ si, de plus,

- $\exists \theta \in \Theta,\ R_\theta(\hat{\eta}') < R_\theta(\hat{\eta})$,

Remarques

- La relation « préférable à » est un ordre partiel sur les risques.
- Il n'existe pas en général un estimateur optimal, càd un estimateur préférable à tous les autres (sauf à restreindre la classe d'estimateurs considérés).

Comparer (tous) les estimateurs : deux approches

Deux approches permettent d'affiner la comparaison
dans les cas où les fonctions de risque ne sont pas comparables :

1. approche **minimax** (ou « pire cas ») :

$$R_{\max}\left(\hat{\eta}\right) = \sup_{\theta\in\Theta} R_\theta\left(\hat{\eta}\right),$$

➠ ne sera pas discutée cette année ;

2. approche bayésienne (ou « en moyenne ») :

$$R_{\mathrm{Bayes},\pi}\left(\hat{\eta}\right) = \int_\Theta R_\theta\left(\hat{\eta}\right)\pi(\mathrm{d}\theta),$$

où π est une mesure de probabilité sur Θ, à choisir.

➠ c'est le sujet de ce cours.

Comparer (tous) les estimateurs : deux approches

Deux approches permettent d'affiner la comparaison
dans les cas où les fonctions de risque ne sont pas comparables :

1. approche minimax (ou « pire cas ») :

$$R_{\max}(\hat{\eta}) = \sup_{\theta \in \Theta} R_\theta(\hat{\eta}),$$

   ➠ ne sera pas discutée cette année ;

2. approche bayésienne (ou « en moyenne ») :

$$R_{\text{Bayes},\pi}(\hat{\eta}) = \int_\Theta R_\theta(\hat{\eta})\, \pi(\mathrm{d}\theta),$$

   où π est une mesure de probabilité sur Θ, à choisir.

   ➠ c'est le sujet de ce cours.

Comparer (tous) les estimateurs : deux approches

Deux approches permettent d'affiner la comparaison
dans les cas où les fonctions de risque ne sont pas comparables :

1. approche minimax (ou « pire cas ») :

$$R_{\max}\big(\hat{\eta}\big) \;=\; \sup_{\theta\in\Theta}\, R_\theta\big(\hat{\eta}\big),$$

➠ ne sera pas discutée cette année ;

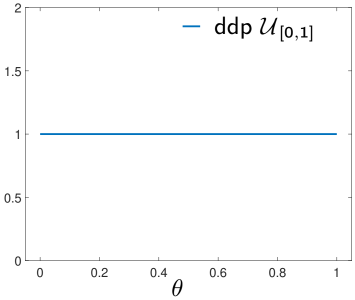
2. approche **bayésienne** (ou « en moyenne ») :

$$R_{\text{Bayes},\pi}\big(\hat{\eta}\big) \;=\; \int_\Theta R_\theta\big(\hat{\eta}\big)\, \pi(\mathrm{d}\theta),$$

où π est une mesure de probabilité sur Θ, à choisir.

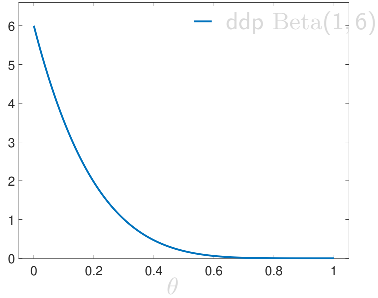
➠ c'est le sujet de ce cours.

Exemple : boules blanches / boules rouges (voir cours n°1)



Mesure π : uniforme sur $[0,1]$

$$\hat{\theta}_{\mathrm{a}} = \frac{\sum_{i=1}^{n} X_i + 1}{n+2}$$



Mesure π : $\mathrm{Beta}\,(1,6)$

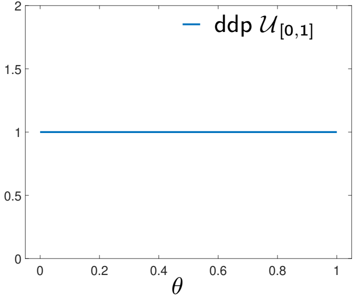
lois bêta

$$\hat{\theta}_{\mathrm{b}} = \frac{\sum_{i=1}^{n} X_i + 1}{n+7}$$

Observation : $\hat{\theta}_{\mathrm{b}} < \hat{\theta}_{\mathrm{a}}$,

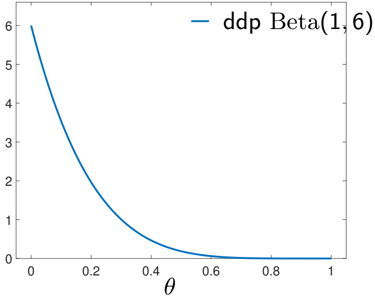
➠ le deuxième estimateur fournit des réponses plus petites.

Exemple : boules blanches / boules rouges (voir cours n°1)



Mesure π : uniforme sur $[0,1]$

$$\hat{\theta}_{\mathrm{a}} = \frac{\sum_{i=1}^{n} X_i + 1}{n+2}$$



Mesure π : $\mathrm{Beta}\,(1,6)$

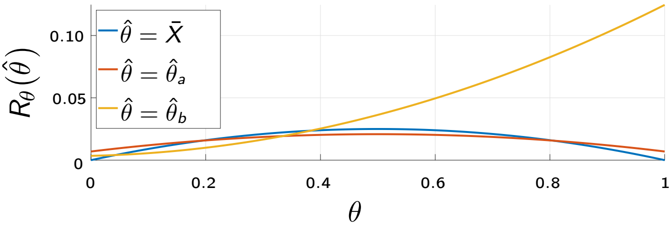
lois bêta

$$\hat{\theta}_{\mathrm{b}} = \frac{\sum_{i=1}^{n} X_i + 1}{n+7}$$

Observation : $\hat{\theta}_{\mathrm{b}} < \hat{\theta}_{\mathrm{a}}$,

➠ le deuxième estimateur fournit des réponses plus petites.

Exemple : boules blanches / boules rouges (avec $n = 10$)



	$\hat{\theta} = \bar{X}$	$\hat{\theta} = \hat{\theta}_a$	$\hat{\theta} = \hat{\theta}_b$
$R_{\max}(\hat{\theta})$	0.025 $\frac{1}{4n}$	≈ 0.0208 $\frac{1}{4(n+2)}$	≈ 0.1246 $\frac{36}{(n+7)^2}$ ❁
$R_{\text{Bayes},\pi}(\hat{\theta})$ avec $\pi \sim \mathcal{U}_{[0,1]}$	≈ 0.0167 $\frac{1}{6n}$	≈ 0.0162 $\frac{n+4}{6(n+2)^2}$	≈ 0.0456 $\frac{n+69}{6(n+7)^2}$
$R_{\text{Bayes},\pi}(\hat{\theta})$ avec $\pi \sim \text{Beta}(1,6)$	≈ 0.0107 $\frac{3}{28n}$	≈ 0.0129 $\frac{3n+22}{28(n+2)^2}$	≈ 0.0089 $\frac{3n+42}{28(n+7)^2}$

exercice 2 Établir les expressions de $R_{\mathbf{max}}$ et $R_{\mathbf{Bayes},\pi}$ pour $\hat{\theta} = \bar{X}$.

❁ valable pour $n \leq 77$

Paramètre inconnu → variable aléatoire

On supposera à partir de maintenant un modèle dominé : ddp $f_\theta(\underline{x})$.

Considérons le risque bayésien (quadratique dans ce cas)

$$\begin{aligned} R_{\text{Bayes},\pi}(\hat\eta) &= \int_\Theta R_\theta(\hat\eta)\, \pi(\mathrm{d}\theta) \\ &= \int_\Theta \mathbb{E}_\theta\left(\|\hat\eta - g(\theta)\|^2\right)\, \pi(\mathrm{d}\theta). \end{aligned}$$

On peut le ré-écrire sous la forme :

$$R_{\text{Bayes},\pi}(\hat\eta) = \iint_{\underline{\mathcal{X}}\times\Theta} \|\hat\eta(\underline{x}) - g(\theta)\|^2 \quad \underbrace{f_\theta(\underline{x})\,\nu(\mathrm{d}\underline{x})\,\pi(\mathrm{d}\theta)}_{\text{Mesure de proba sur } \underline{\mathcal{X}}\times\Theta} \quad .$$

Paramètre inconnu → variable aléatoire

On supposera à partir de maintenant un modèle dominé : ddp $f_\theta(\underline{x})$.

Considérons le risque bayésien (quadratique dans ce cas)

$$\begin{aligned} R_{\text{Bayes},\pi}(\hat{\eta}) &= \int_\Theta R_\theta(\hat{\eta})\,\pi(\mathrm{d}\theta) \\ &= \int_\Theta \mathbb{E}_\theta\left(\|\hat{\eta} - g(\theta)\|^2\right)\,\pi(\mathrm{d}\theta). \end{aligned}$$

On peut le ré-écrire sous la forme :

$$R_{\text{Bayes},\pi}(\hat{\eta}) = \iint_{\underline{\mathcal{X}}\times\Theta} \|\hat{\eta}(\underline{x}) - g(\theta)\|^2 \quad \underbrace{f_\theta(\underline{x})\,\nu(\mathrm{d}\underline{x})\,\pi(\mathrm{d}\theta)}_{\text{Mesure de proba sur } \underline{\mathcal{X}}\times\Theta} \quad .$$

Paramètre inconnu → variable aléatoire (suite)

Introduisons une nouvelle variable aléatoire ϑ, telle que

$$(\underline{X},\, \vartheta) \;\sim\; f_\theta(\underline{x})\, \nu(\mathrm{d}\underline{x})\; \pi(\mathrm{d}\theta). \qquad (\star)$$

Alors le risque bayésien se ré-écrit plus simplement :

$$R_{\mathrm{Bayes},\pi} \;=\; \mathbb{E}\left(\|\hat{\eta} - g(\vartheta)\|^2 \right),$$

l'espérance portant cette fois sur $\underline{X}$ **et** sur ϑ.

Approche bayésienne

En statistique bayésienne, le paramètre inconnu θ est (aussi) modélisé comme une variable aléatoire.

(Remarque technique : l'introduction d'une nouvelle VA ϑ telle que $(\star)$ soit vraie est toujours possible, quitte à remplacer l'espace Ω sous-jacent par $\widetilde{\Omega} = \Omega \times \Theta$, à condition que Θ soit muni d'une tribu $\mathscr{F}_\Theta$ telle que $\theta \mapsto \mathbb{P}_\theta(E)$ est $\mathscr{F}_\Theta$-mesurable pour tout $E \in \mathscr{F}$.)

Paramètre inconnu → variable aléatoire (suite)

Introduisons une nouvelle variable aléatoire ϑ, telle que

$$(\underline{X},\, \vartheta) \;\sim\; f_\theta(\underline{x})\, \nu(\mathrm{d}\underline{x})\; \pi(\mathrm{d}\theta). \qquad (\star)$$

Alors le risque bayésien se ré-écrit plus simplement :

$$R_{\text{Bayes},\pi} \;=\; \mathbb{E}\left(\|\hat{\eta} - g(\vartheta)\|^2\right),$$

l'espérance portant cette fois sur $\underline{X}$ **et** sur ϑ.

Approche bayésienne

En statistique bayésienne, le paramètre inconnu θ est (aussi) modélisé comme une variable aléatoire.

(Remarque technique : l'introduction d'une nouvelle VA ϑ telle que $(\star)$ soit vraie est toujours possible, quitte à remplacer l'espace Ω sous-jacent par $\widetilde{\Omega} = \Omega \times \Theta$, à condition que Θ soit muni d'une tribu $\mathscr{F}_\Theta$ telle que $\theta \mapsto \mathbb{P}_\theta(E)$ est $\mathscr{F}_\Theta$-mesurable pour tout $E \in \mathscr{F}$.)

Paramètre inconnu $\to$ variable aléatoire (suite)

Introduisons une nouvelle variable aléatoire ϑ, telle que

$$(\underline{X},\, \vartheta) \;\sim\; f_\theta(\underline{x})\, \nu(\mathrm{d}\underline{x})\; \pi(\mathrm{d}\theta). \qquad (\star)$$

Alors le risque bayésien se ré-écrit plus simplement :

$$R_{\mathsf{Bayes},\pi} \;=\; \mathbb{E}\left(\|\hat{\eta} - g(\vartheta)\|^2\right),$$

l'espérance portant cette fois sur $\underline{X}$ **et** sur ϑ.

Approche bayésienne

En statistique bayésienne, le paramètre inconnu θ est (aussi) modélisé comme une variable aléatoire.

(Remarque technique : l'introduction d'une nouvelle VA ϑ telle que $(\star)$ soit vraie est toujours possible, quitte à remplacer l'espace Ω sous-jacent par $\widetilde{\Omega} = \Omega \times \Theta$, à condition que Θ soit muni d'une tribu $\mathscr{F}_\Theta$ telle que $\theta \mapsto \mathbb{P}_\theta(E)$ est $\mathscr{F}_\Theta$-mesurable pour tout $E \in \mathscr{F}$.)

Plan du cours

Modèle statistique bayésien

Hypothèses techniques : on suppose à partir de maintenant que

- Θ est muni d'une tribu $\mathscr{F}_\Theta$. Par ex. : si $\Theta \subset \mathbb{R}^p$, $\mathscr{F}_\Theta = \mathcal{B}(\Theta)$;
- $\theta \mapsto \mathbb{P}_\theta(E)$ est $\mathscr{F}_\Theta$-mesurable pour tout $E \in \mathscr{F}$ (tribu sur Ω sous-jacent).

Définition

On appelle modèle statistique bayésien la donnée

- d'un modèle statistique tel que défini précédemment :

$$\left(\underline{\mathcal{X}},\ \underline{\mathscr{A}},\ \left\{\mathbb{P}_\theta^{\underline{X}},\ \theta \in \Theta\right\}\right),$$

- d'une mesure de probabilité π, dite loi a priori, sur $(\Theta, \mathscr{F}_\Theta)$.

Modèle supposé dominé $\to$ permet de définir une vraisemblance.

Modèle statistique bayésien

Hypothèses techniques : on suppose à partir de maintenant que

- Θ est muni d'une tribu $\mathscr{F}_\Theta$. Par ex. : si $\Theta \subset \mathbb{R}^p$, $\mathscr{F}_\Theta = \mathcal{B}(\Theta)$;
- $\theta \mapsto \mathbb{P}_\theta(E)$ est $\mathscr{F}_\Theta$-mesurable pour tout $E \in \mathscr{F}$ (tribu sur Ω sous-jacent).

Définition

On appelle modèle statistique bayésien la donnée

- d'un modèle statistique tel que défini précédemment :

$$\left(\underline{\mathcal{X}},\ \underline{\mathscr{A}},\ \left\{\mathbb{P}_\theta^{\underline{X}},\ \theta \in \Theta\right\}\right),$$

- d'une mesure de probabilité π, dite loi a priori, sur $(\Theta,\ \mathscr{F}_\Theta)$.

Modèle supposé dominé $\rightarrow$ permet de définir une vraisemblance.

Lois jointe, a priori et a posteriori

Rappel : on introduit une nouvelle variable aléatoire ϑ, telle que

$$(\underline{X}, \vartheta) \sim f_\theta(\underline{x})\, \nu(\mathrm{d}\underline{x})\, \pi(\mathrm{d}\theta). \qquad (\star)$$

Vocabulaire bayésien

On appelle :

- loi jointe la loi de $\underline{X}$ et ϑ, c'est-à-dire $(\star)$,
- loi a priori la loi marginale $\mathbb{P}^{\vartheta}$ de ϑ, c'est-à-dire π,
- loi a posteriori la loi $\mathbb{P}^{\vartheta|\underline{X}}$ de ϑ sachant les observations.

Interprétation (« bayésienne subjective »)

- loi a priori $\rightarrow$ connaissance de θ avant acquisition des données
- loi a posteriori $\rightarrow$ ... après acquisition des données

Lois jointe, a priori et a posteriori

Rappel : on introduit une nouvelle variable aléatoire ϑ, telle que

$$(\underline{X}, \vartheta) \sim f_\theta(\underline{x})\, \nu(\mathrm{d}\underline{x})\, \pi(\mathrm{d}\theta). \qquad (\star)$$

Vocabulaire bayésien

On appelle :

- loi jointe la loi de $\underline{X}$ et ϑ, c'est-à-dire $(\star)$,
- loi a priori la loi marginale $\mathbb{P}^\vartheta$ de ϑ, c'est-à-dire π,
- loi a posteriori la loi $\mathbb{P}^{\vartheta|\underline{X}}$ de ϑ sachant les observations.

Interprétation (« bayésienne subjective »)

- loi a priori $\rightarrow$ connaissance de θ avant acquisition des données
- loi a posteriori $\rightarrow$ ... après acquisition des données

Lois jointe, a priori et a posteriori

Rappel : on introduit une nouvelle variable aléatoire ϑ, telle que

$$(\underline{X},\, \vartheta) \;\sim\; f_\theta(\underline{x})\, \nu(\mathrm{d}\underline{x})\, \pi(\mathrm{d}\theta). \qquad (\star)$$

Vocabulaire bayésien

On appelle :

- loi jointe la loi de $\underline{X}$ et ϑ, c'est-à-dire $(\star)$,
- loi a priori la loi marginale $\mathbb{P}^{\vartheta}$ de ϑ, c'est-à-dire π,
- loi a posteriori la loi $\mathbb{P}^{\vartheta|\underline{X}}$ de ϑ sachant les observations.

Interprétation (« bayésienne subjective »)

- loi a priori $\to$ connaissance de θ avant acquisition des données
- loi a posteriori $\to$ ... après acquisition des données

A propos... qu'est-ce qu'une loi conditionnelle ?

Définition générale : hors programme !

(⇒ utilise la notion de noyau)

Soit (U, V) un couple de variables (ou vecteurs) aléatoires admettant une densité par rapport à une certaine mesure produit $\nu_1 \otimes \nu_2$.

On *définira* $\mathbb{P}^{V|U=u}$ comme la mesure admettant la densité

$$f^{V|U}(v \mid u) \;=\; \frac{f^{U,V}(u, v)}{f^{U}(u)}$$

par rapport à ν_2, pour tout u tel que $f^{U}(u) > 0$.

On a alors, pour toute fonction φ mesurable t.q. $\varphi(U, V) \in L^1$,

$$\mathbb{E}\left(\varphi(U, V) \mid U\right) \overset{\text{p.s.}}{=} \int_\Theta \varphi(U, v)\, f^{V|U}(v \mid U)\, \nu_2(\mathrm{d}v).$$

A propos... qu'est-ce qu'une loi conditionnelle ?

Définition générale : hors programme !

($\Rightarrow$ utilise la notion de noyau)

Soit (U, V) un couple de variables (ou vecteurs) aléatoires admettant une densité par rapport à une certaine mesure produit $\nu_1 \otimes \nu_2$.

On *définira* $\mathbb{P}^{V|U=u}$ comme la mesure admettant la densité

$$f^{V|U}(v \mid u) \;=\; \frac{f^{U,V}(u,v)}{f^{U}(u)}$$

par rapport à ν_2, pour tout u tel que $f^U(u) > 0$.

On a alors, pour toute fonction φ mesurable t.q. $\varphi(U, V) \in L^1$,

$$\mathbb{E}\left(\varphi(U,V) \mid U\right) \overset{\text{p.s.}}{=} \int_\Theta \varphi(U,v)\, f^{V|U}(v \mid U)\, \nu_2(\mathrm{d}v).$$

A propos... qu'est-ce qu'une loi conditionnelle ?

Définition générale : hors programme !

(⇒ utilise la notion de noyau)

Soit (U, V) un couple de variables (ou vecteurs) aléatoires admettant une densité par rapport à une certaine mesure produit $\nu_1 \otimes \nu_2$.

On *définira* $\mathbb{P}^{V|U=u}$ comme la mesure admettant la densité

$$f^{V|U}(v \mid u) \;=\; \frac{f^{U,V}(u,v)}{f^{U}(u)}$$

par rapport à ν_2, pour tout u tel que $f^U(u) > 0$.

On a alors, pour toute fonction φ mesurable t.q. $\varphi(U, V) \in L^1$,

$$\mathbb{E}\left(\varphi(U,V) \mid U\right) \;\stackrel{\text{p.s.}}{=}\; \int_\Theta \varphi(U,v)\, f^{V|U}(v \mid U)\, \nu_2(\mathrm{d}v).$$

Densité jointe et densités marginales

On supposera à partir de maintenant† que π admet une densité

- par rapport à une mesure ρ sur $(\Theta, \mathscr{F}_\Theta)$, par ex. Lebesgue,
- on notera (abusivement) : $\pi(\mathrm{d}\theta) = \pi(\theta)\, \rho(\mathrm{d}\theta)$.

Proposition

La loi jointe admet la densité jointe

$$f^{(\underline{X},\vartheta)}(\underline{x}, \theta) \;=\; f_\theta(\underline{x})\, \pi(\theta),$$

et les densité marginales associées sont

$$\begin{aligned} f^{\vartheta}(\theta) \;&=\; \pi(\theta), \\ f^{\underline{X}}(\underline{x}) \;&=\; \int f_\theta(\underline{x})\, \pi(\theta)\, \rho(\mathrm{d}\theta). \end{aligned}$$

† : Ce n'est pas vraiment une hypothèse, on peut toujours prendre $\rho = \pi$ (avec la ddp égale à 1).

Densité jointe et densités marginales

On supposera à partir de maintenant[†] que π admet une densité

- par rapport à une mesure ρ sur $(\Theta, \mathscr{F}_\Theta)$, par ex. Lebesgue,
- on notera (abusivement) : $\pi(\mathrm{d}\theta) = \pi(\theta)\,\rho(\mathrm{d}\theta)$.

Proposition

La loi jointe admet la densité jointe

$$f^{(\underline{X},\vartheta)}(\underline{x},\theta) \;=\; f_\theta(\underline{x})\,\pi(\theta),$$

et les densité marginales associées sont

$$\begin{aligned} f^{\vartheta}(\theta) \;&=\; \pi(\theta), \\ f^{\underline{X}}(\underline{x}) \;&=\; \int f_\theta(\underline{x})\,\pi(\theta)\,\rho(\mathrm{d}\theta). \end{aligned}$$

[†] : Ce n'est pas vraiment une hypothèse, on peut toujours prendre $\rho = \pi$ (avec la ddp égale à 1).

Démonstration

Densité jointe (démonstration informelle) :

$$\begin{aligned}\mathbb{P}^{(\underline{X},\vartheta)}(\mathrm{d}\underline{x}, \mathrm{d}\theta) &= f_\theta(\underline{x})\,\nu(\mathrm{d}\underline{x}) \quad \pi(\theta)\,\rho(\mathrm{d}\theta) \\ &= \underbrace{f_\theta(\underline{x})\,\pi(\theta)}_{\text{ddp jointe}} \quad \nu(\mathrm{d}\underline{x})\,\rho(\mathrm{d}\theta)\end{aligned}$$

Densités marginales $\rightarrow$ il suffit d'intégrer :

$$\begin{aligned}f^{\vartheta}(\theta) &= \int f_\theta(\underline{x})\,\pi(\theta)\,\nu(\mathrm{d}\underline{x}) = \pi(\theta), \\ f^{\underline{X}}(\underline{x}) &= \int f_\theta(\underline{x})\,\pi(\theta)\,\rho(\mathrm{d}\theta).\end{aligned}$$

□

Démonstration

Densité jointe (démonstration informelle) :

$$\begin{aligned}
\mathbb{P}^{(\underline{X},\vartheta)}(\mathrm{d}\underline{x}, \mathrm{d}\theta) &= f_\theta(\underline{x})\,\nu(\mathrm{d}\underline{x}) \quad \pi(\theta)\,\rho(\mathrm{d}\theta) \\
&= \underbrace{f_\theta(\underline{x})\,\pi(\theta)}_{\text{ddp jointe}} \quad \nu(\mathrm{d}\underline{x})\,\rho(\mathrm{d}\theta)
\end{aligned}$$

Densités marginales $\rightarrow$ il suffit d'intégrer :

$$\begin{aligned}
f^{\vartheta}(\theta) &= \int f_\theta(\underline{x})\,\pi(\theta)\,\nu(\mathrm{d}\underline{x}) = \pi(\theta), \\
f^{\underline{X}}(\underline{x}) &= \int f_\theta(\underline{x})\,\pi(\theta)\,\rho(\mathrm{d}\theta).
\end{aligned}$$

$\square$

Vraisemblance et formule de Bayes

Rappel de la densité conditionnelle :

$$f^{V|U}(v \mid u) = \frac{f^{(U,V)}(u,v)}{f^{U}(u)}, \qquad \forall u \text{ t.q. } f^{U}(u) \neq 0. \qquad (\star)$$

Proposition

i) La loi conditionnelle de $\underline{X}$ sachant ϑ admet la ddp

$$f^{\underline{X}|\vartheta}(\underline{x} \mid \theta) = f_\theta(\underline{x}) \qquad (\text{« vraisemblance »}).$$

ii) La loi a posteriori (ϑ sachant $\underline{X}$) admet la densité :

$$f^{\vartheta|\underline{X}}(\theta \mid \underline{x}) = \frac{f_\theta(\underline{x})\,\pi(\theta)}{f^{\underline{X}}(\underline{x})} \qquad (\text{formule de Bayes}).$$

Démonstration. Il suffit d'appliquer $(\star)$ à la densité jointe. □

Vraisemblance et formule de Bayes

Rappel de la densité conditionnelle :

$$f^{V|U}(v \mid u) \;=\; \frac{f^{(U,V)}(u,v)}{f^{U}(u)}, \qquad \forall u \text{ t.q. } f^{U}(u) \neq 0. \qquad (\star)$$

Proposition

i) La loi conditionnelle de $\underline{X}$ sachant ϑ admet la ddp

$$f^{\underline{X}|\vartheta}(\underline{x} \mid \theta) \;=\; f_\theta(\underline{x}) \qquad \text{(« vraisemblance »).}$$

ii) La loi a posteriori (ϑ sachant $\underline{X}$) admet la densité :

$$f^{\vartheta|\underline{X}}(\theta \mid \underline{x}) \;=\; \frac{f_\theta(\underline{x})\,\pi(\theta)}{f^{\underline{X}}(\underline{x})} \qquad \text{(formule de Bayes).}$$

Démonstration. Il suffit d'appliquer $(\star)$ à la densité jointe. □

Remarque : proportionnalité

Le terme $\frac{1}{f^{\underline{X}}(\underline{x})}$ joue le rôle d'une constante de normalisation :

$$f^{\vartheta|\underline{X}}(\theta \mid \underline{x}) \;=\; \frac{f_\theta(\underline{x})\,\pi(\theta)}{f^{\underline{X}}(\underline{x})}.$$

Notation. Le symbole « $\propto$ » indique la proportionnalité. Ainsi,

$$f^{\vartheta|\underline{X}}(\theta \mid \underline{x}) \;\propto\; f_\theta(\underline{x})\,\pi(\theta),$$

soit encore, de façon informelle,

$$\boxed{\text{ddp a posteriori} \;\propto\; \text{vraisemblance} \;\times\; \text{ddp a priori.}}$$

La « constante » $f^{\underline{X}}(\underline{x})$ est souvent difficile à calculer, mais dans certaines situations (estimateur MAP, méthodes numériques MCMC...) on peut s'en affranchir.

Remarque : proportionnalité

Le terme $\frac{1}{f^{\underline{X}}(\underline{x})}$ joue le rôle d'une constante de normalisation :

$$f^{\vartheta|\underline{X}}(\theta \mid \underline{x}) \;=\; \frac{f_{\theta}(\underline{x})\,\pi(\theta)}{f^{\underline{X}}(\underline{x})}.$$

Notation. Le symbole « $\propto$ » indique la proportionnalité. Ainsi,

$$f^{\vartheta|\underline{X}}(\theta \mid \underline{x}) \;\propto\; f_{\theta}(\underline{x})\,\pi(\theta),$$

soit encore, de façon informelle,

$$\boxed{\text{ddp a posteriori} \;\propto\; \text{vraisemblance} \;\times\; \text{ddp a priori.}}$$

La « constante » $f^{\underline{X}}(\underline{x})$ est souvent difficile à calculer, mais dans certaines situations (estimateur MAP, méthodes numériques MCMC...) on peut s'en affranchir.

Exemple : boules blanches / boules rouges (suite)

Rappel : on veut estimer $\theta = \frac{B}{R+B}$ à partir de $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \mathrm{Ber}(\theta)$.

Densité des observations :

$$f_\theta(\underline{x}) = \Pi_{i=1}^{n} \theta^{x_i} \, (1-\theta)^{1-x_i} = \theta^{N(\underline{x})} \, (1-\theta)^{n-N(\underline{x})}.$$

avec $N(\underline{x}) = \sum_{i=1}^{n} x_i$.

On se donne un a priori $\mathrm{Beta}(a_0, b_0)$ sur le paramètre θ :

$$\pi(\theta) \;\propto\; \theta^{a_0-1} \, (1-\theta)^{b_0-1} \,,$$

et on note, comme précédemment, ϑ la VA correspondante.

(La question du choix de l'a priori sera discutée plus loin.)

Exemple : boules blanches / boules rouges (suite)

Rappel : on veut estimer $\theta = \frac{B}{R+B}$ à partir de $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \mathrm{Ber}(\theta)$.

Densité des observations :

$$f_\theta(\underline{x}) = \Pi_{i=1}^{n} \theta^{x_i} (1-\theta)^{1-x_i} = \theta^{N(\underline{x})} (1-\theta)^{n-N(\underline{x})}.$$

avec $N(\underline{x}) = \sum_{i=1}^{n} x_i$.

On se donne un a priori $\mathrm{Beta}(a_0, b_0)$ sur le paramètre θ :

$$\pi(\theta) \; \propto \; \theta^{a_0 - 1} (1-\theta)^{b_0 - 1},$$

et on note, comme précédemment, ϑ la VA correspondante.

(La question du choix de l'a priori sera discutée plus loin.)

Exemple : boules blanches / boules rouges (suite)

Alors on a :

$$
\begin{aligned}
f^{\vartheta|\underline{X}}(\theta \mid \underline{x}) &\propto f_\theta(\underline{x})\,\pi(\theta) \\
&\propto \theta^{N(\underline{x})}\,(1-\theta)^{n-N(\underline{x})} \;\cdot\; \theta^{a_0-1}\,(1-\theta)^{b_0-1} \\
&= \theta^{a_0+N(\underline{x})-1}\,(1-\theta)^{b_0+n-N(\underline{x})-1}\,.
\end{aligned}
$$

On reconnaît (à une cst près) la densité de la loi $\mathrm{Beta}(a_n, b_n)$, avec

$$
\begin{cases}
a_n = a_0 + N, \\
b_n = b_0 + n - N.
\end{cases}
$$

lois bêta

Conclusion. Loi a posteriori : $\vartheta \mid \underline{X} \sim \mathrm{Beta}(a_n, b_n)$.

Exemple : boules blanches / boules rouges (suite)

Alors on a :

$$
\begin{aligned}
f^{\vartheta|\underline{X}}(\theta \mid \underline{x}) &\propto f_\theta(\underline{x})\, \pi(\theta) \\
&\propto \theta^{N(\underline{x})} (1-\theta)^{n-N(\underline{x})} \ \cdot\ \theta^{a_0-1} \, (1-\theta)^{b_0-1} \\
&= \theta^{a_0+N(\underline{x})-1} \, (1-\theta)^{b_0+n-N(\underline{x})-1} \,.
\end{aligned}
$$

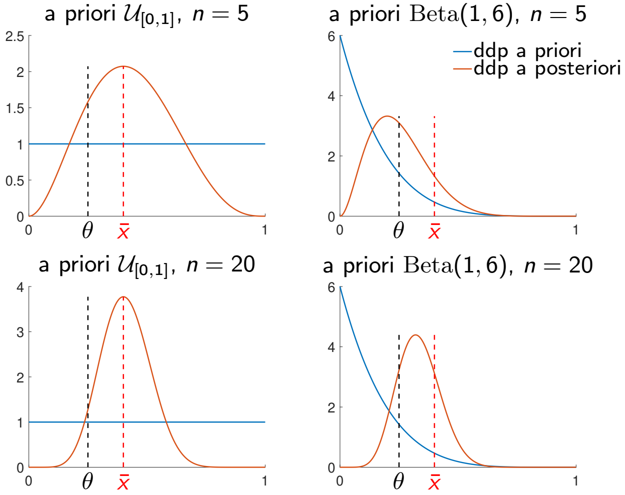
On reconnaît (à une cst près) la densité de la loi $\mathrm{Beta}(a_n, b_n)$, avec

$$
\begin{cases}
a_n = a_0 + N, \\
b_n = b_0 + n - N.
\end{cases}
$$

lois bêta

Conclusion. Loi a posteriori : $\vartheta \mid \underline{X} \sim \mathrm{Beta}(a_n, b_n)$.

Exemple : boules blanches / boules rouges (suite)



Remarque : pour $n \to \infty$, on a $\mathbb{E}(\vartheta \mid \underline{X}_n) = \bar{X}_n + O(\frac{1}{n})$ avec $\mathrm{var}(\vartheta \mid \underline{X}_n) \simeq \frac{\theta(1-\theta)}{n}$.

Exemple : fiabilité composant

Rappel : $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \mathcal{E}(\theta) = \mathcal{E}(\frac{1}{\eta})$, d'où la vraisemblance :

$$\begin{aligned}\mathcal{L}(\eta, \underline{x}_n) \;=\; f(\underline{x}_n \mid \eta) \;&=\; \prod_{i=1}^{n} \frac{1}{\eta} \exp\left(-\frac{1}{\eta} x_i\right) \\ &=\; \eta^{-n} \exp\left(-\frac{1}{\eta}\sum_{i=1}^{n} x_i\right).\end{aligned}$$

(Ici on choisit de tout paramétrer directement en η.)

On choisit (voir plus loin) un a priori $\mathcal{N}(\eta_0, \sigma_0^2)$ tronqué pour η :

$$\pi(\eta) \;\propto\; \exp\left(-\frac{(\eta-\eta_0)^2}{2\sigma_0^2}\right) \mathbb{1}_{\eta \geq 0}.$$

Abus de notation : on note simplement f la densité conditionnelle, au lieu de $f^{\underline{X}_n \mid \boldsymbol{\eta}}$, où $\boldsymbol{\eta}$ serait la variable aléatoire associée au paramètre η.

Exemple : fiabilité composant

Rappel : $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \mathcal{E}(\theta) = \mathcal{E}(\frac{1}{\eta})$, d'où la vraisemblance :

$$\begin{aligned} \mathcal{L}(\eta, \underline{x}_n) \;=\; f(\underline{x}_n \mid \eta) \;&=\; \prod_{i=1}^{n} \frac{1}{\eta} \exp\left(-\frac{1}{\eta} x_i\right) \\ &=\; \eta^{-n} \exp\left(-\frac{1}{\eta} \sum_{i=1}^{n} x_i\right). \end{aligned}$$

(Ici on choisit de tout paramétrer directement en η.)

On choisit (voir plus loin) un a priori $\mathcal{N}(\eta_0, \sigma_0^2)$ tronqué pour η :

$$\pi(\eta) \;\propto\; \exp\left(-\frac{(\eta - \eta_0)^2}{2\sigma_0^2}\right) \mathbb{1}_{\eta \geq 0}.$$

Abus de notation : on note simplement f la densité conditionnelle, au lieu de $f^{\underline{X}_n \mid \boldsymbol{\eta}}$, où $\boldsymbol{\eta}$ serait la variable aléatoire associée au paramètre η.

Exemple : fiabilité composant (suite)

Loi a posteriori de η. Par la formule de Bayes, on obtient :

$$f(\eta \mid \underline{x}_n) \;\propto\; \underbrace{\eta^{-n} \exp\left(-\frac{1}{\eta}\sum_{i=1}^{n} x_i\right)}_{\text{vraisemblance}} \cdot \underbrace{\exp\left(-\frac{(\eta-\eta_0)^2}{2\sigma_0^2}\right)}_{\text{ddp a priori}}.$$

⚠ Cette fois-ci on ne reconnaît pas une densité « connue »

➠ évaluation numérique des intégrales

$$f(\underline{x}_n) \;=\; \int_0^{+\infty} \eta^{-n}\, e^{-\frac{1}{\eta}\sum_{i=1}^{n} x_i}\; e^{-\frac{(\eta-\eta_0)^2}{2\sigma_0^2}}\, \mathrm{d}\eta$$

$$\mathbb{E}\left(\eta \mid \underline{X}_n = \underline{x}_n\right) \;=\; \frac{1}{f(\underline{x}_n)} \int_0^{+\infty} \eta \cdot \eta^{-n}\, e^{-\frac{1}{\eta}\sum_{i=1}^{n} x_i}\; e^{-\frac{(\eta-\eta_0)^2}{2\sigma_0^2}}\, \mathrm{d}\eta$$

Abus de notation (bis) : on utilise souvent la même symbole (ici η) pour noter un point de l'espace des paramètres et la VA associée au paramètre.

Exemple : fiabilité composant (suite)

Loi a posteriori de η. Par la formule de Bayes, on obtient :

$$f(\eta \mid \underline{x}_n) \;\propto\; \underbrace{\eta^{-n} \exp\left(-\frac{1}{\eta}\sum_{i=1}^{n} x_i\right)}_{\text{vraisemblance}} \cdot \underbrace{\exp\left(-\frac{(\eta-\eta_0)^2}{2\sigma_0^2}\right)}_{\text{ddp a priori}}.$$

⚠ Cette fois-ci on ne reconnaît pas une densité « connue »

➠ évaluation numérique des intégrales

$$f(\underline{x}_n) \;=\; \int_0^{+\infty} \eta^{-n}\, e^{-\frac{1}{\eta}\sum_{i=1}^{n} x_i}\; e^{-\frac{(\eta-\eta_0)^2}{2\sigma_0^2}}\; \mathrm{d}\eta$$

$$\mathbb{E}\left(\eta \mid \underline{X}_n = \underline{x}_n\right) \;=\; \frac{1}{f(\underline{x}_n)} \int_0^{+\infty} \eta \cdot \eta^{-n}\, e^{-\frac{1}{\eta}\sum_{i=1}^{n} x_i}\; e^{-\frac{(\eta-\eta_0)^2}{2\sigma_0^2}}\; \mathrm{d}\eta$$

Abus de notation (bis) : on utilise souvent la même symbole (ici η) pour noter un point de l'espace des paramètres et la VA associée au paramètre.

Exemple : fiabilité composant (suite)

Application numérique. $\eta_0 = 14.0$, $\sigma_0 = 1.0$ et valeur vraie : $\eta = 11.4$.

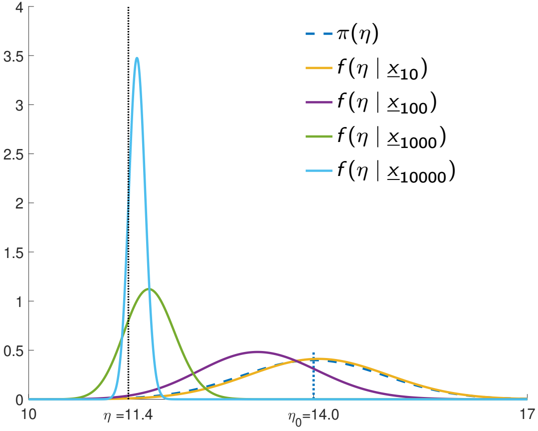


Figure – Densité a priori de η et densités a posteriori pour 4 valeurs de n.

Plan du cours

Plusieurs approches

Deux types de sources d'information a priori :

- données « historiques »,
- experts : connaissances subjectives, expertise métier, etc.

Sujets plus avancés (pas traités dans ce cours) :

- fusions de plusieurs sources d'information a priori,
- loi a priori « peu informatives » ou « objectives »,
- loi a priori la plus défavorable (cf. minimax),
- ...

Exemple : boules blanches / boules rouges (suite)

On dispose de données issues d'une première expérience :

- échantillon de $n_0 = 20$ observations,
- $N_0 = 15$ boules blanches tirées.

Choix d'un a priori

On peut décider, par exemple, de choisir une loi $\mathrm{Beta}(a_0, b_0)$, avec $a_0 = N_0 = 15$ et $b_0 = n_0 - N_0 = 5$.

Arguments en faveur de ce choix :

- la forme de la loi facilite les calculs (voir plus loin) ;
- espérance : $\frac{a_0}{a_0+b_0} = p_0$, avec $p_0 = \frac{N_0}{n_0}$;
- variance : $\frac{a_0 b_0}{(a_0+b_0)^2(a_0+b_0+1)} \approx \frac{p_0(1-p_0)}{n_0}$ ➠ variance de $\bar{X}_{n_0}$.

Exemple : boules blanches / boules rouges (suite)

On dispose de données issues d'une première expérience :

- échantillon de $n_0 = 20$ observations,
- $N_0 = 15$ boules blanches tirées.

Choix d'un a priori

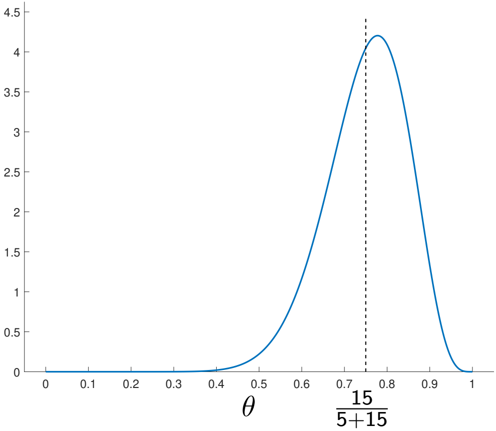
On peut décider, par exemple, de choisir une loi $\mathrm{Beta}(a_0, b_0)$, avec $a_0 = N_0 = 15$ et $b_0 = n_0 - N_0 = 5$.

Arguments en faveur de ce choix :

- la forme de la loi facilite les calculs (voir plus loin) ;
- espérance : $\frac{a_0}{a_0+b_0} = p_0$, avec $p_0 = \frac{N_0}{n_0}$;
- variance : $\frac{a_0 b_0}{(a_0+b_0)^2(a_0+b_0+1)} \approx \frac{p_0(1-p_0)}{n_0}$ ➠ variance de $\bar{X}_{n_0}$.

Exemple : boules blanches / boules rouges (suite)

Densité a priori $\mathrm{Beta}(15, 5)$



Exemple : fiabilité composant

On dispose des informations suivantes :

- Le constructeur indique que la durée de vie des composantes est de l'ordre de $\eta_0 = 6$ mois.
- Un expert du domaine estime à $\varepsilon_0 = 10\%$ la précision de l'information fournie par le constructeur.

Choix d'un a priori (élicitation)

On peut décider, par exemple, de choisir une loi $\mathcal{N}(\eta_0, \sigma_0)$, tronquée à $[0, +\infty[$, avec $\sigma_0 = \varepsilon_0 \eta_0 / 1.96$.

Arguments en faveur de ce choix :

- L'a priori est (approx.) centré sur la valeur constructeur η_0.
- $\approx$ 95% de la proba a priori est portée par l'intervalle $[0.9\eta_0,\ 1.1\eta_0]$.
- La forme choisie (gaussienne) et la valeur 95% sont arbitraires.

Exemple : fiabilité composant

On dispose des informations suivantes :

- Le constructeur indique que la durée de vie des composantes est de l'ordre de $\eta_0 = 6$ mois.
- Un expert du domaine estime à $\varepsilon_0 = 10\%$ la précision de l'information fournie par le constructeur.

Choix d'un a priori (**élicitation**)

On peut décider, par exemple, de choisir une loi $\mathcal{N}(\eta_0, \sigma_0)$, tronquée à $[0, +\infty[$, avec $\sigma_0 = \varepsilon_0 \eta_0 / 1.96$.

Arguments en faveur de ce choix :

- L'a priori est (approx.) centré sur la valeur constructeur η_0.
- $\approx$ 95% de la proba a priori est portée par l'intervalle $[0.9\eta_0,\ 1.1\eta_0]$.
- La forme choisie (gaussienne) et la valeur 95% sont arbitraires.

A priori conjugués ➠ facilitent les calculs !

Familles de lois a priori conjuguées

Une famille de lois (densités) est dite conjuguée pour un certain modèle statistique si, pour toute loi a priori π dans cette famille, la loi a posteriori $f^{\vartheta|\underline{X}}$ est encore dedans.

Exemples.

- échantillon $\mathrm{Ber}(\theta)$ + a priori bêta,
- échantillon $\mathcal{N}(\mu, \sigma^2)$ avec σ^2 connu + a priori $\mathcal{N}$ sur μ,
- échantillon $\mathcal{N}(\mu, \sigma^2)$ avec μ connu + a priori $\mathcal{IG}^\dagger$ sur σ^2,
- échantillon $\mathcal{E}(\theta)$ + a priori gamma, ➠ lois gamma
- ...

† : inverse gamma. $Z \sim \mathcal{IG}$ si $1/Z$ suit une loi gamma.

A priori conjugués ➠ facilitent les calculs !

Familles de lois a priori conjuguées

Une famille de lois (densités) est dite conjuguée pour un certain modèle statistique si, pour toute loi a priori π dans cette famille, la loi a posteriori $f^{\vartheta|\underline{X}}$ est encore dedans.

Exemples.

- échantillon $\mathrm{Ber}(\theta)$ + a priori bêta,
- échantillon $\mathcal{N}(\mu, \sigma^2)$ avec σ^2 connu + a priori $\mathcal{N}$ sur μ,
- échantillon $\mathcal{N}(\mu, \sigma^2)$ avec μ connu + a priori $\mathcal{IG}^\dagger$ sur σ^2,
- échantillon $\mathcal{E}(\theta)$ + a priori gamma, ➠ lois gamma
- ...

† : inverse gamma. $Z \sim \mathcal{IG}$ si $1/Z$ suit une loi gamma.

Plan du cours

Estimateurs bayésiens

Objectif

Construire des estimateurs de $\eta = g(\theta)$ prenant en compte

- les données $\underline{x}$,
- et la loi a priori π.

Estimateurs bayésiens

On se donne une fonction de perte $L : N \times N \to \mathbb{R}$.

- rappel : $L(\eta, \tilde{\eta})$ est la perte si l'on estime $\tilde{\eta}$ alors que la vraie valeur est η.

Définition : estimateur bayésien

Il minimise l'espérance a posteriori de la perte :

$$\hat{\eta} \;=\; \arg\min_{\tilde{\eta} \in N} J(\tilde{\eta}, \underline{X})$$

avec

$$\begin{aligned} J(\tilde{\eta}, \underline{x}) &= \mathbb{E}\left(L(g(\vartheta), \tilde{\eta}) \;\middle|\; \underline{X} = \underline{x} \right) \\ &= \int_\Theta L(g(\theta), \tilde{\eta})\, f^{\vartheta | \underline{X}}(\theta \mid \underline{x})\, \rho(\mathrm{d}\theta). \end{aligned}$$

(☞ J est bien définie pour $\mathbb{P}^{\underline{X}}$-presque tout $\underline{x}$.)

Remarque : un tel estimateur minimise le risque bayésien R_π.

Estimateurs bayésiens

On se donne une fonction de perte $L : N \times N \to \mathbb{R}$.

- rappel : $L(\eta, \tilde{\eta})$ est la perte si l'on estime $\tilde{\eta}$ alors que la vraie valeur est η.

Définition : estimateur bayésien

Il minimise l'espérance a posteriori de la perte :

$$\hat{\eta} \;=\; \arg\min_{\tilde{\eta} \in N} \; J(\tilde{\eta}, \underline{X})$$

avec

$$\begin{aligned} J(\tilde{\eta}, \underline{x}) \;&=\; \mathbb{E}\left(L\big(g(\vartheta), \tilde{\eta}\big) \;\middle|\; \underline{X} = \underline{x}\right) \\ &=\; \int_{\Theta} L\big(g(\theta), \tilde{\eta}\big)\, f^{\vartheta|\underline{X}}(\theta \mid \underline{x})\, \rho(\mathrm{d}\theta). \end{aligned}$$

(☞ J est bien définie pour $\mathbb{P}^{\underline{X}}$-presque tout $\underline{x}$.)

Remarque : un tel estimateur minimise le risque bayésien R_π.

Perte quadratique

Considérons le cas de la perte quadratique $L(\eta, \tilde{\eta}) = \|\eta - \tilde{\eta}\|^2$:

$$J(\tilde{\eta}, \underline{x}) = \int_\Theta \|g(\theta) - \tilde{\eta}\|^2 \; f^{\vartheta|\underline{X}}(\theta \mid \underline{x}) \, \rho(\mathrm{d}\theta).$$

Proposition

Dans ce cas, l'estimateur bayésien est

$$\hat{\eta} \;=\; \mathbb{E}\left(g(\vartheta) \mid \underline{X}\right) \;=\; \int_\Theta g(\theta) \, f^{\vartheta|\underline{X}}(\theta \mid \underline{X}) \, \rho(\mathrm{d}\theta).$$

➠ $\hat{\eta}$ est la moyenne a posteriori de ϑ.

Remarque : on peut aussi l'écrire

$$\hat{\eta}(\underline{x}) \;=\; \frac{\int_\Theta g(\theta) \, f_\theta(\underline{x}) \, \pi(\theta) \, \rho(\mathrm{d}\theta)}{f^{\underline{X}}(\underline{x})} \;=\; \frac{\int_\Theta g(\theta) \, f_\theta(\underline{x}) \, \pi(\theta) \, \rho(\mathrm{d}\theta)}{\int_\Theta f_\theta(\underline{x}) \, \pi(\theta) \, \rho(\mathrm{d}\theta)}.$$

Perte quadratique

Considérons le cas de la perte quadratique $L(\eta, \tilde{\eta}) = \|\eta - \tilde{\eta}\|^2$:

$$J(\tilde{\eta}, \underline{x}) = \int_\Theta \|g(\theta) - \tilde{\eta}\|^2 \; f^{\vartheta|\underline{X}}(\theta \mid \underline{x})\, \rho(\mathrm{d}\theta).$$

Proposition

Dans ce cas, l'estimateur bayésien est

$$\hat{\eta} \;=\; \mathbb{E}\left(g(\vartheta) \mid \underline{X}\right) \;=\; \int_\Theta g(\theta)\, f^{\vartheta|\underline{X}}(\theta \mid \underline{X})\, \rho(\mathrm{d}\theta).$$

➠ $\hat{\eta}$ est la moyenne a posteriori de ϑ.

Remarque : on peut aussi l'écrire

$$\hat{\eta}(\underline{x}) \;=\; \frac{\int_\Theta g(\theta)\, f_\theta(\underline{x})\, \pi(\theta)\, \rho(\mathrm{d}\theta)}{f^{\underline{X}}(\underline{x})} \;=\; \frac{\int_\Theta g(\theta)\, f_\theta(\underline{x})\, \pi(\theta)\, \rho(\mathrm{d}\theta)}{\int_\Theta f_\theta(\underline{x})\, \pi(\theta)\, \rho(\mathrm{d}\theta)}.$$

Exemple : boules blanches / boules rouges (suite)

Avec un a priori $\vartheta \sim \mathrm{Beta}(a_0, b_0)$, on a vu que :

$$\vartheta | \underline{X} \ \sim \ \mathrm{Beta}\left(N + a_0, n - N + b_0\right)$$

avec $N = \sum_{i=1}^{n} X_i$.

L'espérance d'une loi $\mathrm{Beta}(a, b)$ étant $\frac{a}{a+b}$, il vient :

$$\hat{\theta} \ = \ \mathbb{E}\left(\vartheta \mid \underline{X}\right) \ = \ \frac{N + a_0}{n + a_0 + b_0}.$$

Remarque : on retrouve les expressions de $\hat{\theta}_a$ et $\hat{\theta}_b$ (retour au slide 6).

Exemple : boules blanches / boules rouges (suite)

Avec un a priori $\vartheta \sim \mathrm{Beta}(a_0, b_0)$, on a vu que :

$$\vartheta | \underline{X} \ \sim \ \mathrm{Beta}\left(N + a_0, n - N + b_0\right)$$

avec $N = \sum_{i=1}^{n} X_i$.

L'espérance d'une loi $\mathrm{Beta}(a, b)$ étant $\frac{a}{a+b}$, il vient :

$$\hat{\theta} \ = \ \mathbb{E}\left(\vartheta \mid \underline{X}\right) \ = \ \frac{N + a_0}{n + a_0 + b_0}.$$

Remarque : on retrouve les expressions de $\hat{\theta}_a$ et $\hat{\theta}_b$ (retour au slide 6).

Autre exemple : n-échantillon gaussien (σ^2 connu)

On montre (cf. TD 5) que si $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \mathcal{N}\left(\theta, \sigma_0^2\right)$

- avec $\theta \in \mathbb{R}$ (inconnu), $\sigma_0 > 0$ (connu),
- et $\vartheta \sim \mathcal{N}\left(\mu, \tau^2\right)$,

alors

$$\vartheta \mid \underline{X} \ \sim \ \mathcal{N}\left(\frac{\tau^2 \sum_{i=1}^n X_i + \sigma_0^2 \mu}{n\tau^2 + \sigma_0^2}, \ \frac{\tau^2 \sigma_0^2}{n\tau^2 + \sigma_0^2}\right)$$

D'où l'estimateur bayésien (pour la perte quadratique) :

$$\hat{\theta} \ = \ \lambda \overline{X} + (1 - \lambda)\, \mu \qquad \text{avec } \lambda = \frac{n\tau^2}{n\tau^2 + \sigma_0^2}$$

Interprétation.

- quand $n \to \infty$, $\hat{\theta} \approx \bar{X}$ (l'a priori n'a plus d'influence)
- à n fini, quand $\dfrac{\sigma_0}{\tau} \gg 1$, $\hat{\theta} \approx \mu$ (données presque inutiles).

Autre exemple : n-échantillon gaussien (σ^2 connu)

On montre (cf. TD 5) que si $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \mathcal{N}\left(\theta, \sigma_0^2\right)$

- avec $\theta \in \mathbb{R}$ (inconnu), $\sigma_0 > 0$ (connu),
- et $\vartheta \sim \mathcal{N}\left(\mu, \tau^2\right)$,

alors

$$\vartheta \mid \underline{X} \;\sim\; \mathcal{N}\left(\frac{\tau^2 \sum_{i=1}^n X_i + \sigma_0^2 \mu}{n\tau^2 + \sigma_0^2}, \; \frac{\tau^2 \sigma_0^2}{n\tau^2 + \sigma_0^2}\right)$$

D'où l'estimateur bayésien (pour la perte quadratique) :

$$\hat{\theta} \;=\; \lambda \overline{X} + (1 - \lambda)\, \mu \qquad \text{avec } \lambda = \frac{n\tau^2}{n\tau^2 + \sigma_0^2}$$

Interprétation.

- quand $n \to \infty$, $\hat{\theta} \approx \bar{X}$ (l'a priori n'a plus d'influence)
- à n fini, quand $\dfrac{\sigma_0}{\tau} \gg 1$, $\hat{\theta} \approx \mu$ (données presque inutiles).

Autre exemple : n-échantillon gaussien (σ^2 connu)

On montre (cf. TD 5) que si $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \mathcal{N}\left(\theta, \sigma_0^2\right)$

- avec $\theta \in \mathbb{R}$ (inconnu), $\sigma_0 > 0$ (connu),
- et $\vartheta \sim \mathcal{N}\left(\mu, \tau^2\right)$,

alors

$$\vartheta \mid \underline{X} \; \sim \; \mathcal{N}\left(\frac{\tau^2 \sum_{i=1}^n X_i + \sigma_0^2 \mu}{n\tau^2 + \sigma_0^2}, \; \frac{\tau^2 \sigma_0^2}{n\tau^2 + \sigma_0^2}\right)$$

D'où l'estimateur bayésien (pour la perte quadratique) :

$$\hat{\theta} \; = \; \lambda \overline{X} + (1 - \lambda) \, \mu \qquad \text{avec } \lambda = \frac{n\tau^2}{n\tau^2 + \sigma_0^2}$$

Interprétation.

- quand $n \to \infty$, $\hat{\theta} \approx \bar{X}$ (l'a priori n'a plus d'influence)
- à n fini, quand $\dfrac{\sigma_0}{\tau} \gg 1$, $\hat{\theta} \approx \mu$ (données presque inutiles).

Perte L^1

Supposons pour simplifier que $\eta = \theta \in \mathbb{R}$.

Considérons la fonction de perte $L(\theta, \tilde{\theta}) = \left|\theta - \tilde{\theta}\right|$:

$$J(\tilde{\theta}, \underline{x}) = \int_\Theta \left|\theta - \tilde{\theta}\right| f^{\vartheta | \underline{X}}(\theta \mid \underline{x})\, \rho(\mathrm{d}\theta).$$

Proposition

Dans ce cas, l'estimateur bayésien $\hat{\theta}$ est tel que

$$\int_{-\infty}^{\hat{\theta}} f^{\vartheta | \underline{X}}(\theta \mid \underline{X})\, \rho(\mathrm{d}\theta) \;=\; \int_{\hat{\theta}}^{\infty} f^{\vartheta | \underline{X}}(\theta \mid \underline{X})\, \rho(\mathrm{d}\theta) \;=\; \frac{1}{2} \quad \mathbb{P}^{\underline{X}}\text{-p.s.}.$$

➠ $\hat{\theta}$ est une médiane de la densité a posteriori de ϑ.

Remarque : lorsque ϑ a une densité a posteriori symétrique, les deux estimateurs bayésiens (perte L^1 et perte L^2) coïncident.

Exemple : moyenne d'un n-échantillon gaussien, avec a priori gaussien.

Perte L^1

Supposons pour simplifier que $\eta = \theta \in \mathbb{R}$.

Considérons la fonction de perte $L(\theta, \tilde{\theta}) = \left|\theta - \tilde{\theta}\right|$:

$$J(\tilde{\theta}, \underline{x}) = \int_{\Theta} \left|\theta - \tilde{\theta}\right| f^{\vartheta|\underline{X}}(\theta \mid \underline{x})\, \rho(\mathrm{d}\theta).$$

Proposition

Dans ce cas, l'estimateur bayésien $\hat{\theta}$ est tel que

$$\int_{-\infty}^{\hat{\theta}} f^{\vartheta|\underline{X}}(\theta \mid \underline{X})\, \rho(\mathrm{d}\theta) \;=\; \int_{\hat{\theta}}^{\infty} f^{\vartheta|\underline{X}}(\theta \mid \underline{X})\, \rho(\mathrm{d}\theta) \;=\; \frac{1}{2} \quad \mathbb{P}^{\underline{X}}\text{-p.s..}$$

➠ $\hat{\theta}$ est une médiane de la densité a posteriori de ϑ.

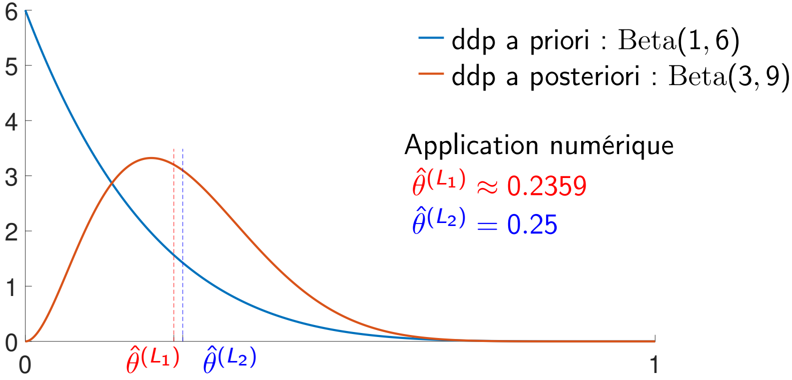
Remarque : lorsque ϑ a une densité a posteriori symétrique, les deux estimateurs bayésiens (perte L^1 et perte L^2) coïncident.

Exemple : moyenne d'un n-échantillon gaussien, avec a priori gaussien.

Exemple : boules blanches / boules rouges (suite)

Echantillon observé ($n = 5$) : $\underline{x} = (B, R, R, B, R)$.

A priori sur θ : $\vartheta \sim \mathrm{Beta}(1, 6)$, avec $\theta = \mathbb{P}(X_1 = B)$.



Application numérique

$\hat{\theta}^{(L_1)} \approx 0.2359$

$\hat{\theta}^{(L_2)} = 0.25$

Plan du cours

Plan du cours

Exercice 1 (loi exponentielle + a priori gamma)

corrigé

Soient $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \mathcal{E}(\theta)$ avec $\theta \in \Theta =]0, +\infty[$.

On munit θ d'un a priori $\mathrm{Gamma}\,(\alpha_0, \beta_0)$.

Questions

1. Montrer que l'a priori gamma est conjugué, et déterminer les paramètres α_n et β_n de la loi a posteriori.
2. Expliciter l'estimateur bayésien de θ, pour la perte quadratique.
3. Montrer que cet estimateur tend vers l'EMV si les paramètres α_0 et β_0 tendent vers une limite à préciser.

Exercice 2 (risques maximal et risque bayésien)

corrigé

Soient $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \mathrm{Ber}(\theta)$ avec $\theta \in \Theta =]0, 1[$.

On veut estimer θ. L'objectif de cet exercice est de démontrer les expressions des risques quadratiques bayésien et maximal de $\hat{\theta} = \bar{X}$, annoncés au slide 7.

Questions

i. Calculer le risque quadratique $R_\theta\left(\bar{X}\right)$, et en déduire le risque maximal $R_{\max}\left(\bar{X}\right)$.

ii. Calculer le risque bayésien $R_{\text{Bayes},\pi}\left(\bar{X}\right)$ lorsque π est une loi bêta de paramètres $a > 0$ et $b > 0$.

Plan du cours

Corrigé de l'exercice 1

retour au sujet

Remarque préliminaire : dans ce corrigé on s'autorise, comme c'est souvent le cas en pratique, à noter de la même manière le paramètre « déterministe » θ et la variable aléatoire associée, notée ϑ dans le cours.

i) On écrit la vraisemblance :

$$\mathcal{L}(\theta; \underline{x}) = f(\underline{x} \mid \theta) = \prod_{i=1}^{n} \theta e^{-\theta x_i} = \theta^n e^{-\theta \sum_{i=1}^{n} x_i},$$

et la densité a priori :

$$\pi(\theta) = \frac{\beta_0^{\alpha_0}}{\Gamma(\alpha_0)} \theta^{\alpha_0 - 1} e^{-\beta_0 \theta} \propto \theta^{\alpha_0 - 1} e^{-\beta_0 \theta}.$$

On en déduit la densité a posteriori par la formule de Bayes :

$$f(\theta \mid \underline{x}) \propto \mathcal{L}(\theta; \underline{x})\, \pi(\theta) \propto \theta^{\alpha_0 + n} e^{-\theta\left(\beta_0 + \sum_{i=1}^{n} x_i\right)}$$

Corrigé de l'exercice 1

retour au sujet

La loi de θ sachant $\underline{X}$, ou loi a posteriori, est donc une loi gamma de paramètres

- $\alpha_n = \alpha_0 + n$,
- $\beta_n = \beta_0 + \sum_{i=1}^n X_i$.

ii) L'estimateur bayésien pour la perte quadratique est donné l'espérance a posteriori de θ sachant les observations :

$$\mathbb{E}\left(\theta \mid \underline{X}\right) = \frac{\alpha_n}{\beta_n} = \frac{\alpha_0 + n}{\beta_0 + \sum_{i=1}^n X_i}.$$

iii) Cet estimateur tend vers l'EMV $1/\bar{X}_n$ lorsque les paramètres α_0 et β_0 tendent vers zéro.

retour au sujet

i) $\bar{X}$ est un estimateur sans biais de $\theta = \mathbb{E}_\theta(X_1)$, donc

$$R_\theta(\bar{X}) \;=\; \mathrm{var}_\theta\left(\bar{X}\right) \;=\; \frac{1}{n}\,\mathrm{var}_\theta(X_1) \;=\; \frac{\theta\,(1-\theta)}{n}.$$

La fonction $\theta \mapsto R_\theta\big(\hat{\theta}\big)$ est un polynôme de degré deux en θ qui atteint sont maximum en $\theta = \frac{1}{2}$, d'où :

$$R_{\max}\big(\bar{X}\big) \;=\; \frac{1}{4n}.$$

ii) Soit $B(a,b) = \Gamma(a)\,\Gamma(b)\,/\,\Gamma(a+b)$.

Corrigé de l'exercice 2

retour au sujet

Le risque bayésien pour $\pi = \mathrm{Beta}(a, b)$ vaut :

$$
\begin{aligned}
R_{\text{Bayes},\pi}\left(\bar{X}\right) &= \int R_\theta\left(\hat{\theta}\right)\pi(\mathrm{d}\theta) \\
&= \int_0^1 \frac{\theta\ (1-\theta)}{n} \cdot \frac{1}{B(a,b)}\,\theta^{a-1}\ (1-\theta)^{b-1}\ \mathrm{d}\theta \\
&= \frac{1}{n\,B(a,b)} \int_0^1 \theta^a\ (1-\theta)^b\ \mathrm{d}\theta \\
&= \frac{B\left(a+1, b+1\right)}{n\,B(a,b)} = \frac{1}{n}\,\frac{ab}{(a+b+1)(a+b)}.
\end{aligned}
$$

En particulier,

- Pour $\pi = \mathcal{U}_{[0,1]} = \mathrm{Beta}(1,1)$, $R_{\text{Bayes},\pi}\left(\bar{X}\right) = \frac{1}{6n}$.
- Pour $\pi = \mathrm{Beta}(1,6)$, $R_{\text{Bayes},\pi}\left(\bar{X}\right) = \frac{3}{28n}$.

Plan du cours

La famille des lois bêta

Soit $X \sim \mathrm{Beta}(a, b)$ avec $(a, b) = \theta \in (\mathbb{R}_\star^+)^2$. Sa densité est :

$$f_\theta(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \, x^{a-1}(1-x)^{b-1} \, \mathbb{1}_{]0,1[}(x).$$

Moments

- moyenne : $\mathbb{E}_\theta(X) = \frac{a}{a+b}$
- variance : $\mathrm{var}_\theta(X) = \frac{ab}{(a+b)^2(a+b+1)}$

Cas particulier

- $\mathcal{U}_{[0,1]} = \mathrm{Beta}(1, 1)$

Propriétés

- Si $X \sim \mathrm{Beta}(a, 1)$, alors $-\log(X) \sim \mathcal{E}\left(\frac{1}{a}\right)$.
- Si $X \sim \Gamma(a, \lambda)$, $Y \sim \Gamma(b, \lambda)$, et $X \perp\!\!\!\perp Y$, alors $\frac{X}{X+Y} \sim \mathrm{Beta}(a, b)$.

retour au slide 6 retour au slide 18 retour au slide 36

La famille des lois gamma

On dit que X suit la loi $\Gamma(p, \lambda)$, de paramètres $p > 0$ et $\lambda > 0$, si elle admet pour densité de probabilité

$$f(x) = \frac{\lambda^p}{\Gamma(p)} x^{p-1} \exp(-\lambda x) \mathbb{1}_{\mathbb{R}^+}(x).$$

Moments

- moyenne : $\mathbb{E}_\theta(X) = \frac{p}{\lambda}$
- variance : $\mathrm{var}_\theta(X) = \frac{p}{\lambda^2}$

Cas particuliers

- $\mathcal{E}(\lambda) = \Gamma(p = 1, \lambda)$
- $\Gamma(p = \frac{n}{2}, \lambda = \frac{n}{2}) = \chi^2(n)$

Propriétés

- Soit $a > 0$. Si $X \sim \Gamma(p, \lambda)$, alors $aX \sim \Gamma\left(p, \frac{\lambda}{a}\right)$.
- Si X et Y sont indépendantes, avec $X \sim \Gamma(p, \lambda)$ et $Y \sim \Gamma(q, \lambda)$, alors $X + Y \sim \Gamma(p + q, \lambda)$.

retour au slide 27