Slides statistiques bayésienne

Janvier-Mars 2021

Déroulement du cours

Séances

- Sur discord *(si ça vous va)*
- Caméra allumée pour dire bonjour (+ si vous voulez)
- Cours (slides ou tablette selon les jours) alterné avec
  - Recherche de petits exercices
  - Longues séances "TP"

Déroulement du cours

Séances

- Sur discord *(si ça vous va)*
- Caméra allumée pour dire bonjour (+ si vous voulez)
- Cours (slides ou tablette selon les jours) alterné avec
  - Recherche de petits exercices
  - Longues séances "TP"

Evaluation

Projet d'application directe du cours

- Avancé pendant les séances de TP
- Quelques rendus non évalués au fur et à mesure
- Pour la prochaine séance (28 janvier), idée de données et questions.*(doc d'une demi page max à uploader sur TOMUSS, non évalué)*
- En groupe si vous le souhaitez *(possibilité de faire le début seul, et la fin en groupe)*

Plan

Prior, posterior et prédictive

- Introduction
- Modèle
- Conjugaison de priors

Utilisation de la loi à posteriori

- Regions de crédibilité
- Estimateurs ponctuels et théorie de la décision
- Tests

Aspects numériques

- Monte-Carlo, rejet et ABC
- Metropolis-Hasting
- Priors semi-conjugués et alogirthme de Gibbs

Choix de priors et modèles Hierarchiques

Introduction

Entrée en matière classique

Avant le nouvel an, vous êtes allé·e faire un test antigénique pour le COVID (le fabriquant annonce 7% *de faux positifs et* 40% *de faux négatifs). Vous n'aviez pas de symptômes, mais n'avez pas fait spécialement attention non plus aux gestes barrières. Le taux d'incidence dans votre région était de* $1/1000$*.*
Qu'en pensez vous ?

Introduction

Entrée en matière classique

Avant le nouvel an, vous êtes allé·e faire un test antigénique pour le COVID (le fabriquant annonce 7% de faux positifs et 40% de faux négatifs). Vous n'aviez pas de symptômes, mais n'avez pas fait spécialement attention non plus aux gestes barrières. Le taux d'incidence dans votre région était de 1/1000.
Qu'en pensez vous ?

$$
\begin{aligned}
M &: \text{Avoir la maladie} \\
T &: \text{Avoir un test positif} \\
\mathbb{P}(M) = 0.001 \,; \mathbb{P}(T \mid M) &= 0.6 \,; \mathbb{P}(T \mid \bar{M}) = 0.07 \\
\mathbb{P}(M \mid T) &= \frac{\mathbb{P}(M \cap T)}{\mathbb{P}(T)} = \frac{\mathbb{P}(T \mid M)\mathbb{P}(M)}{\mathbb{P}(T \mid M)\mathbb{P}(M) + \mathbb{P}(T \mid \bar{M})\mathbb{P}(\bar{M})} \\
&\approx \frac{0.0006}{0.07} \approx 0.0085
\end{aligned}
$$

Introduction

Entrée en matière classique

Avant le nouvel an, vous êtes allé·e faire un test antigénique pour le COVID (le fabriquant annonce 7% de faux positifs et 40% de faux négatifs). Vous n'aviez pas de symptômes, mais n'avez pas fait spécialement attention non plus aux gestes barrières. Le taux d'incidence dans votre région était de 1/1000.
Qu'en pensez vous ?

$$M : \text{Avoir la maladie}$$
$$T : \text{Avoir un test positif}$$
$$\mathbb{P}(M) = 0.001 \; ; \mathbb{P}(T \mid M) = 0.6 \; ; \mathbb{P}(T \mid \bar{M}) = 0.07$$
$$\mathbb{P}(M \mid T) = \frac{\mathbb{P}(M \cap T)}{\mathbb{P}(T)} = \frac{\mathbb{P}(T \mid M)\mathbb{P}(M)}{\mathbb{P}(T \mid M)\mathbb{P}(M) + \mathbb{P}(T \mid \bar{M})\mathbb{P}(\bar{M})}$$
$$\approx \frac{0.0006}{0.07} \approx 0.0085$$

Ça va...

Grands Principes du point de vue Bayésien

3 idées importantes

1. Probabilité = Croyances
2. Données = permettent de **mettre à jour** nos croyances
3. Prise en compte de l'information a priori

Grands Principes du point de vue Bayésien

3 idées importantes

1. Probabilité = Croyances
2. Données = permettent de **mettre à jour** nos croyances
3. Prise en compte de l'information a priori

Formule essentielle "Formule de Bayes" : Si $\mathbb{P}(B) > 0$,

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B \mid A)\mathbb{P}(A)}{\mathbb{P}(B \mid A)\mathbb{P}(A) + \mathbb{P}(B \mid \bar{A})\mathbb{P}(\bar{A})}$$

Analogue pour densités : Si $f_Y(y) > 0$,

$$f_{X|Y}(x, y) = \frac{f_{X,Y}(x, y)}{f_Y(y)} = \frac{f_{Y|X}(x, y) f_X(x)}{\int_x f_{Y|X}(x, y) f_X(x) \mathrm{d}x}$$

Grands Principes du point de vue Bayésien

3 idées importantes

1. Probabilité = Croyances
2. Données = permettent de **mettre à jour** nos croyances
3. Prise en compte de l'information a priori

Formule essentielle "Formule de Bayes" : Si $\mathbb{P}(B) > 0$,

$$\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(B \mid A)\mathbb{P}(A)}{\mathbb{P}(B \mid A)\mathbb{P}(A) + \mathbb{P}(B \mid \bar{A})\mathbb{P}(\bar{A})}$$

Analogue pour densités : Si $f_Y(y) > 0$,

$$f_{X|Y}(x, y) = \frac{f_{X,Y}(x, y)}{f_Y(y)} = \frac{f_{Y|X}(x, y) f_X(x)}{\int_x f_{Y|X}(x, y) f_X(x) \mathrm{d}x}$$

Différence avec le cadre fréquentiste :

$\boxed{\text{On va considérer que les paramètres des lois des observations sont } \textit{aléatoires}}$

Cadre de travail

Modèle Bayésien

- $\mathcal{M} = \{f_{X|\theta}; \theta \in \Theta\}$

Cadre de travail

Modèle Bayésien

- $\mathcal{M} = \{f_{X|\theta}; \theta \in \Theta\}$

Remarque :

- $\mathcal{M}$ est une famille paramétrique de lois pour les observations absolument continues par rapport à une même mesure de référence (comme en fréquentiste)
- $\mathcal{M}$ peut donc être constituée de lois discrètes ou de lois continues, que l'on notera dans tous les cas $f_{X|\theta}$

Cadre de travail

Modèle Bayésien

- $\mathcal{M} = \{f_{X|\theta}; \theta \in \Theta\}$
- f_θ loi sur θ

Remarque :

- $\mathcal{M}$ est une famille paramétrique de lois pour les observations absolument continues par rapport à une même mesure de référence (comme en fréquentiste)
- $\mathcal{M}$ peut donc être constituée de lois discrètes ou de lois continues, que l'on notera dans tous les cas $f_{X|\theta}$
- f_θ est appellée *loi a priori* ou simplement *prior*
- f_θ représente l'état de nos croyance sur le paramètre **avant d'avoir vu les données.**

Du prior au posterior

Loi a posteriori

On suppose que l'on observe $X_1, \cdots X_n$ que l'on modélise i.i.d. *conditionnellement à* θ.

On appelle **loi a posteriori** ou **posterior** :

$$f_{\theta|X}(x_1, \cdots, x_n, \theta) = \frac{f_\theta(\theta)\prod_{i=1}^n f_{X|\theta}(x_i, \theta)}{\int_\theta f_\theta(\theta)\prod_{i=1}^n f_{X|\theta}(x_i, \theta)\mathrm{d}\theta}$$

Du prior au posterior

Loi a posteriori

On suppose que l'on observe $X_1, \cdots X_n$ que l'on modélise i.i.d. *conditionnellement à* θ.
On appelle **loi a posteriori** ou **posterior** :

$$f_{\theta|X}(x_1, \cdots, x_n, \theta) = \frac{f_\theta(\theta) \prod_{i=1}^n f_{X|\theta}(x_i, \theta)}{\int_\theta f_\theta(\theta) \prod_{i=1}^n f_{X|\theta}(x_i, \theta) \mathrm{d}\theta}$$

Notations

Même dans le cas à densité, on s'autorisera à noter (au choix) :

$$\begin{aligned} f_{\theta|X}(x_1, \cdots, x_n, t) &= \mathbb{P}\Big(\theta = t \mid X_1 = x_1, \cdots X_n = x_n\Big) = \mathbb{P}\Big(\theta = t \mid X = x\Big) \\ &= \mathbb{P}(t \mid x_1, \cdots x_n) = \pi(t \mid x_1, \cdots x_n) \\ &= \mathbb{P}(t \mid x) = \pi(t \mid x) \\ &= \mathbb{P}(\theta \mid x) = \pi(\theta \mid x) \end{aligned}$$

Une astuce utile

Normalisation

$$\mathbb{P}\Big(\theta = t \mid X = x\Big) = C(x)\mathbb{P}(\theta = t)\prod_{i=1}^{n}\mathbb{P}(X_i = x_i \mid \theta = t)$$

$$\frac{1}{C(x)} = \int_t \mathbb{P}(\theta = t)\prod_{i=1}^{n}\mathbb{P}(X_i = x_i \mid \theta = t)\mathrm{d}t$$

Une astuce utile

Normalisation

$$\mathbb{P}\Big(\theta = t \mid X = x\Big) = C(x)\mathbb{P}(\theta = t)\prod_{i=1}^{n}\mathbb{P}(X_i = x_i \mid \theta = t)$$

$$\frac{1}{C(x)} = \int_t \mathbb{P}(\theta = t)\prod_{i=1}^{n}\mathbb{P}(X_i = x_i \mid \theta = t)\mathrm{d}t$$

Notations

On notera $\propto$ pour "proportionnel" (à une "constante" qui dépend de X près !):

$$\mathbb{P}\Big(\theta = t \mid X = x\Big) \propto \mathbb{P}(\theta = t)\prod_{i=1}^{n}\mathbb{P}(X_i = x_i \mid \theta = t)$$

Attention : En pratique, on passe au log pour les calculs !

Super utile pour les calculs : Il n'y a plus qu'à reconnaître une loi, à constante près.

Calcul de posterior

Exercice 1

On rappelle la densité d'une loi Gamma(k, β) :

$$f_{\text{Gamma}(k,\beta)}(x) = \frac{x^{k-1}\beta^k e^{-\beta x}}{\Gamma(k)} \mathbb{1}_{x \geq 0}; \Gamma(z) = \int_0^{+\infty} t^{z-1} e^{-z} \mathrm{d}z$$

On choisit le modèle "Loi exponentielle" pour les données : $\mathcal{M} = \{\mathcal{E}(\lambda); \lambda > 0\}$ et une loi Gamma pour le prior : $\pi(\lambda) = f_{\text{Gamma}(k,\beta)}(\lambda)$

Calculer le posterior

Calcul de posterior

Exercice 1

On rappelle la densité d'une loi Gamma(k, β) :

$$f_{\text{Gamma}(k,\beta)}(x) = \frac{x^{k-1}\beta^k e^{-\beta x}}{\Gamma(k)} \mathbb{1}_{x \geq 0}; \Gamma(z) = \int_0^{+\infty} t^{z-1}e^{-z}\mathrm{d}z$$

On choisit le modèle "Loi exponentielle" pour les données : $\mathcal{M} = \{\mathcal{E}(\lambda); \lambda > 0\}$ et une loi Gamma pour le prior : $\pi(\lambda) = f_{\text{Gamma}(k,\beta)}(\lambda)$

Calculer le posterior

Exercice 2

On rappelle la densité d'une loi Beta(a, b) :

$$f_{\text{Beta}(a,b)}(x) = \frac{x^{a-1}(1-x)^{b-1}}{B(a,b)} \mathbb{1}_{x \in [0,1]}; B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

On choisit le modèle "Loi Bernoulli" pour les données : $\mathcal{M} = \{\mathcal{B}er(p); p \in [0,1]\}$ et une loi Beta pour le prior : $\pi(p) = f_{\text{Beta}(a,b)}(p)$

Calculer le posterior

Calcul de posterior

Exercice 1

On rappelle la densité d'une loi Gamma(k, β) :

$$f_{\text{Gamma}(k,\beta)}(x) = \frac{x^{k-1}\beta^k e^{-\beta x}}{\Gamma(k)} \mathbb{1}_{x \geq 0} ; \Gamma(z) = \int_0^{+\infty} t^{z-1} e^{-z} \mathrm{d}z$$

On choisit le modèle "Loi exponentielle" pour les données : $\mathcal{M} = \{\mathcal{E}(\lambda); \lambda > 0\}$ et une loi Gamma pour le prior : $\pi(\lambda) = f_{\text{Gamma}(k,\beta)}(\lambda)$

Calculer le posterior

Exercice 2

On rappelle la densité d'une loi Beta(a, b) :

$$f_{\text{Beta}(a,b)}(x) = \frac{x^{a-1}(1-x)^{b-1}}{B(a,b)} \mathbb{1}_{x \in [0,1]} ; B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

On choisit le modèle "Loi Bernoulli" pour les données : $\mathcal{M} = \{\mathcal{B}er(p); p \in [0,1]\}$ et une loi Beta pour le prior : $\pi(p) = f_{\text{Beta}(a,b)}(p)$

Calculer le posterior

Questions ?

Famille de priors conjugués

Priors conjugués

Une famille $\mathcal{F} = \{\pi_h(\theta), h \in H\}$ est dite **conjuguée** pour un modèle $\mathcal{M} = \{\pi(x \mid \theta); \theta \in \Theta\}$ si

$$\forall h \in H, \forall x_1, \cdots x_n, \exists h' \in H, \pi_h(\theta \mid x) = \pi_{h'}(\theta).$$

Famille de priors conjugués

Priors conjugués

Une famille $\mathcal{F} = \{\pi_h(\theta), h \in H\}$ est dite **conjuguée** pour un modèle $\mathcal{M} = \{\pi(x \mid \theta); \theta \in \Theta\}$ si

$$\forall h \in H, \forall x_1, \cdots x_n, \exists h' \in H, \pi_h(\theta \mid x) = \pi_{h'}(\theta).$$

Avec des notations moins ambigues :
Si $\exists h \in H, f_\theta = \pi_h$ alors pour toutes observations $x_1, \cdots x_n$,
$\exists h'(x) \in H, f_{\theta|X}(\theta, x) = \pi_{h'(x)}(\theta)$

Famille de priors conjugués

Priors conjugués

Une famille $\mathcal{F} = \{\pi_h(\theta), h \in H\}$ est dite **conjuguée** pour un modèle $\mathcal{M} = \{\pi(x \mid \theta); \theta \in \Theta\}$ si

$$\forall h \in H, \forall x_1, \cdots x_n, \exists h' \in H, \pi_h(\theta \mid x) = \pi_{h'}(\theta).$$

Avec des notations moins ambigues :
Si $\exists h \in H, f_\theta = \pi_h$ alors pour toutes observations $x_1, \cdots x_n$,
$\exists h'(x) \in H, f_{\theta|X}(\theta, x) = \pi_{h'(x)}(\theta)$

Exemples : Exercices 1,2 précédents.

Famille de priors conjugués

Priors conjugués

Une famille $\mathcal{F} = \{\pi_h(\theta), h \in H\}$ est dite **conjuguée** pour un modèle $\mathcal{M} = \{\pi(x \mid \theta); \theta \in \Theta\}$ si

$$\forall h \in H, \forall x_1, \cdots x_n, \exists h' \in H, \pi_h(\theta \mid x) = \pi_{h'}(\theta).$$

Avec des notations moins ambigues :
Si $\exists h \in H, f_\theta = \pi_h$ alors pour toutes observations $x_1, \cdots x_n$, $\exists h'(x) \in H, f_{\theta|X}(\theta, x) = \pi_{h'(x)}(\theta)$

Exemples : Exercices 1,2 précédents.

Exercice 3

Trouver la famille de priors conjugués pour les lois suivantes :

- $\mathcal{M} = \{\mathcal{P}(\lambda); \lambda > 0\}$
- $\mathcal{M} = \{\mathcal{N}\left(0, \frac{1}{\tau}\right); \tau > 0\}$

Quelques lois conjuguées

- On peut *actualiser* les paramètres du prior ("**hyperparamètres**") avec chaques nouvelles observations.
- Existe toujours pour une loi dans une famille exponentielle
- Parfois compliqué à normaliser (e;g. prior conjugué pour loi Beta ou Gamma)

Modèle $\pi(X\|\theta)$	Priors conjugués	Mise à jour
$\left\{\text{Bin}(N, p)\ ;\ p \in]0,1[\right\}$	$\left\{\text{Beta}(a, b)\ ;\ a, b > 0\right\}$	$a \to a + \sum_{i=1}^n x_i$ $b \to b + Nn - \sum_{i=1}^n x_i$
$\left\{\mathcal{E}(\lambda)\ ;\ \lambda > 0\right\}$	$\left\{\text{Gamma}(k, \beta)\ ;\ k, \beta > 0\right\}$	$k \to k + n$ $\beta \to \beta + \sum_{i=1}^n x_i$
$\left\{\mathcal{P}(\lambda)\ ;\ \lambda > 0\right\}$	$\left\{\text{Gamma}(k, \beta)\ ;\ k, \beta > 0\right\}$	$k \to k + \sum_{i=1}^n x_i$ $\beta \to \beta + n$
$\left\{\mathcal{G}(p)\ ;\ p \in]0,1[\right\}$	$\left\{\text{Beta}(a, b)\ ;\ a, b > 0\right\}$	$a \to a + n$ $b \to b + \sum_{i=1}^n x_i - n$
$\left\{\mathcal{N}(m, \frac{1}{t})\ ;\ m \in \mathbb{R}\right\}$	$\left\{\mathcal{N}(\mu, \frac{1}{\tau})\ ;\ \mu \in \mathbb{R}, \tau > 0\right\}$	$\mu \to \frac{\tau\mu + t\sum_{i=1}^n x_i}{\tau + tn}$ $\tau \to \tau + tn$
$\left\{\mathcal{N}(m, \frac{1}{t})\ ;\ t > 0\right\}$	$\left\{\text{Gamma}(k, \beta)\ ;\ k, \beta > 0\right\}$	$k \to k + \frac{n}{2}$ $\beta \to \beta + \frac{1}{2}\sum_{i=1}^n (x_i - \mu)^2$
$\left\{\mathcal{N}(m, \Lambda^{-1})\ ;\ m \in \mathbb{R}^d\right\}$	$\left\{\mathcal{N}(\mu, \Omega^{-1})\ ;\ \mu \in \mathbb{R}^d, \Omega \text{ s.d.p.}\right\}$	$\mu \to \left(\Omega + n\Lambda\right)^{-1}\left(\Omega\mu + \Lambda\sum_{i=1}^n x_i\right)$ $\Omega \to \Omega + n\Lambda$
$\left\{\mathcal{N}(m, \Lambda^{-1})\ ;\ \Lambda \text{ s.d.p}\right\}$	$\left\{\text{Wishart}(\nu, V)\ ;\ \nu > 0, V \text{ s.d.p.}\right\}$	$\nu \to \nu + n$ $V \to \left(V^{-1} + \sum_{i=1}^n (x_i - \mu)(x_i - \mu)^T\right)^{-1}$

Quelques lois conjuguées

- On peut *actualiser* les paramètres du prior (**"hyperparamètres"**) avec chaques nouvelles observations.
- Existe toujours pour une loi dans une famille exponentielle
- Parfois compliqué à normaliser (e;g. prior conjugué pour loi Beta ou Gamma)

Modèle $\pi(X\|\theta)$	Priors conjugués	Mise à jour
$\left\{\text{Bin}(N, p)\ ;\ p \in]0, 1[\right\}$	$\left\{\text{Beta}(a, b)\ ;\ a, b > 0\right\}$	$a \to a + \sum_{i=1}^n x_i$ $b \to b + Nn - \sum_{i=1}^n x_i$
$\left\{\mathcal{E}(\lambda)\ ;\ \lambda > 0\right\}$	$\left\{\text{Gamma}(k, \beta)\ ;\ k, \beta > 0\right\}$	$k \to k + n$ $\beta \to \beta + \sum_{i=1}^n x_i$
$\left\{\mathcal{P}(\lambda)\ ;\ \lambda > 0\right\}$	$\left\{\text{Gamma}(k, \beta)\ ;\ k, \beta > 0\right\}$	$k \to k + \sum_{i=1}^n x_i$ $\beta \to \beta + n$
$\left\{\mathcal{G}(p)\ ;\ p \in]0, 1[\right\}$	$\left\{\text{Beta}(a, b)\ ;\ a, b > 0\right\}$	$a \to a + n$ $b \to b + \sum_{i=1}^n x_i - n$
$\left\{\mathcal{N}(m, \frac{1}{t})\ ;\ m \in \mathbb{R}\right\}$	$\left\{\mathcal{N}(\mu, \frac{1}{\tau})\ ;\ \mu \in \mathbb{R}, \tau > 0\right\}$	$\mu \to \frac{\tau\mu + t\sum_{i=1}^n x_i}{\tau + tn}$ $\tau \to \tau + tn$
$\left\{\mathcal{N}(m, \frac{1}{t})\ ;\ t > 0\right\}$	$\left\{\text{Gamma}(k, \beta)\ ;\ k, \beta > 0\right\}$	$k \to k + \frac{n}{2}$ $\beta \to \beta + \frac{1}{2}\sum_{i=1}^n (x_i - \mu)^2$
$\left\{\mathcal{N}(m, \Lambda^{-1})\ ;\ m \in \mathbb{R}^d\right\}$	$\left\{\mathcal{N}(\mu, \Omega^{-1})\ ;\ \mu \in \mathbb{R}^d, \Omega \text{ s.d.p.}\right\}$	$\mu \to \left(\Omega + n\Lambda\right)^{-1}\left(\Omega\mu + \Lambda\sum_{i=1}^n x_i\right)$ $\Omega \to \Omega + n\Lambda$
$\left\{\mathcal{N}(m, \Lambda^{-1})\ ;\ \Lambda \text{ s.d.p}\right\}$	$\left\{\text{Wishart}(\nu, V)\ ;\ \nu > 0, V \text{ s.d.p.}\right\}$	$\nu \to \nu + n$ $V \to \left(V^{-1} + \sum_{i=1}^n (x_i - \mu)(x_i - \mu)^T\right)^{-1}$

On refait quelques exemples ?

Le nouveau collègue

Exercice 4

A. est un nouveau collègue qui a la réputation de sécher les réunions. Avant la première réunion, vous ne vous fiez pas aux rumeurs, et modélisez votre confiance sur sa probabilité p de présence par une loi $Beta(1, \frac{1}{2})$.

1. **Quel est le modèle Bayésien**
2. **Quelle est la probabilité marginale qu'il vienne à la prochaine réunion ?**

Après les 4 premières réunions, où il n'est venu qu'une seule fois, vous actualisez vos croyances.

3. **Quelle est la loi a postériori** ?
4. **Quelle est, selon vous, la probabilité qu'il vienne à la prochaine réunion, sachant les données ?**

Le nouveau collègue

Exercice 4

A. est un nouveau collègue qui a la réputation de sécher les réunions. Avant la première réunion, vous ne vous fiez pas aux rumeurs, et modélisez votre confiance sur sa probabilité p de présence par une loi $Beta(1, \frac{1}{2})$.

1. **Quel est le modèle Bayésien**
2. **Quelle est la probabilité marginale qu'il vienne à la prochaine réunion ?**

Après les 4 premières réunions, où il n'est venu qu'une seule fois, vous actualisez vos croyances.

3. **Quelle est la loi a postériori** ?
4. **Quelle est, selon vous, la probabilité qu'il vienne à la prochaine réunion, sachant les données ?**

Loi predictive

La loi d'une nouvelle donnée, intégrée sur le postérior, est appelée **loi prédictive** :

$$\begin{aligned}\pi(x_{n+1} \mid x_1, \cdots x_n) &:= f_{X_{n+1}|X_1,\cdots X_n}(x_1, \cdots, x_{n+1}) \\ &= \int_\theta \pi(x_{n+1} \mid \theta)\pi(\theta \mid x_1, \cdots, x_n)\mathrm{d}\theta\end{aligned}$$

Le nouveau collègue

Exercice 4

A. est un nouveau collègue qui a la réputation de sécher les réunions. Avant la première réunion, vous ne vous fiez pas aux rumeurs, et modélisez votre confiance sur sa probabilité p de présence par une loi $Beta(1, \frac{1}{2})$.

1. **Quel est le modèle Bayésien**
2. **Quelle est la probabilité marginale qu'il vienne à la prochaine réunion ?**

Après les 4 premières réunions, où il n'est venu qu'une seule fois, vous actualisez vos croyances.

3. **Quelle est la loi a postériori** ?
4. **Quelle est, selon vous, la probabilité qu'il vienne à la prochaine réunion, sachant les données ?**

Loi predictive

La loi d'une nouvelle donnée, intégrée sur le postérior, est appelée **loi prédictive** :

$$\begin{aligned}\pi(x_{n+1} \mid x_1, \cdots x_n) &:= f_{X_{n+1}|X_1,\cdots X_n}(x_1, \cdots, x_{n+1}) \\ &= \int_\theta \pi(x_{n+1} \mid \theta)\pi(\theta \mid x_1, \cdots, x_n)\mathrm{d}\theta\end{aligned}$$

Questions ?

Reprise : 28 janvier

Exercice 5

Calculer la loi prédictive dans un modèle avec de loi Binômiale, avec un prior Beta.
On appele cette loi la **loi beta-binômiale**.

Reprise : 28 janvier

Exercice 5

Calculer la loi prédictive dans un modèle avec de loi Binômiale, avec un prior Beta.
On appele cette loi la **loi beta-binômiale**.

$$\begin{aligned}\pi(x_i \mid p) &= \binom{N}{x_i} p^{x_i} (1-p)^{N-x_i} \\ \pi(p) &\propto p^{a-1}(1-p)^{b-1}\end{aligned}$$

Reprise : 28 janvier

Exercice 5

Calculer la loi prédictive dans un modèle avec de loi Binômiale, avec un prior Beta.
On appele cette loi la **loi beta-binômiale**.

$$
\begin{aligned}
\pi(x_i \mid p) &= \binom{N}{x_i} p^{x_i}(1-p)^{N-x_i} \\
\pi(p) &\propto p^{a-1}(1-p)^{b-1} \\
\pi(p \mid x_1, \cdots x_n) &\propto p^{a-1}(1-p)^{b-1} p^{\sum x_i}(1-p)^{Nn-\sum x_i}
\end{aligned}
$$

Reprise : 28 janvier

Exercice 5

Calculer la loi prédictive dans un modèle avec de loi Binômiale, avec un prior Beta.
On appele cette loi la **loi beta-binômiale**.

$$
\begin{aligned}
\pi(x_i \mid p) &= \binom{N}{x_i} p^{x_i}(1-p)^{N-x_i} \\
\pi(p) &\propto p^{a-1}(1-p)^{b-1} \\
\pi(p \mid x_1, \cdots x_n) &\propto p^{a-1}(1-p)^{b-1} p^{\sum x_i}(1-p)^{Nn-\sum x_i} \\
p \mid x_1, \cdots x_n &\sim \mathrm{Beta}\left(a + \sum x_i, b + Nn - \sum x_i\right) = \mathrm{Beta}\left(a_{\mathrm{post}}(x), b_{\mathrm{post}}(x)\right)
\end{aligned}
$$

Reprise : 28 janvier

Exercice 5

Calculer la loi prédictive dans un modèle avec de loi Binômiale, avec un prior Beta.
On appele cette loi la **loi beta-binômiale**.

$$
\begin{aligned}
\pi(x_i \mid p) &= \binom{N}{x_i} p^{x_i}(1-p)^{N-x_i} \\
\pi(p) &\propto p^{a-1}(1-p)^{b-1} \\
\pi(p \mid x_1, \cdots x_n) &\propto p^{a-1}(1-p)^{b-1} p^{\sum x_i}(1-p)^{Nn-\sum x_i} \\
p \mid x_1, \cdots x_n &\sim \mathrm{Beta}\left(a + \sum x_i, b + Nn - \sum x_i\right) = \mathrm{Beta}\left(a_{\text{post}}(x), b_{\text{post}}(x)\right) \\
\mathbb{P}\left(X_{n+1} = j \mid x_1, \cdots x_n\right) &= \int_{p=0}^{1} \mathbb{P}(X_{n+1} = j \mid p)\pi(p \mid x_1, \cdots x_n)\mathrm{d}p
\end{aligned}
$$

Reprise : 28 janvier

Exercice 5

Calculer la loi prédictive dans un modèle avec de loi Binômiale, avec un prior Beta.
On appele cette loi la **loi beta-binômiale**.

$$
\begin{aligned}
\pi(x_i \mid p) &= \binom{N}{x_i} p^{x_i}(1-p)^{N-x_i} \\
\pi(p) &\propto p^{a-1}(1-p)^{b-1} \\
\pi(p \mid x_1, \cdots x_n) &\propto p^{a-1}(1-p)^{b-1} p^{\sum x_i}(1-p)^{Nn-\sum x_i} \\
p \mid x_1, \cdots x_n &\sim \text{Beta}\left(a+\sum x_i, b+Nn-\sum x_i\right) = \text{Beta}\left(a_{\text{post}}(x), b_{\text{post}}(x)\right) \\
\mathbb{P}\left(X_{n+1}=j \mid x_1, \cdots x_n\right) &= \int_{p=0}^{1} \mathbb{P}(X_{n+1}=j \mid p)\pi(p \mid x_1, \cdots x_n)\mathrm{d}p \\
&= \int_{p=0}^{1} \binom{N}{j} p^{j}(1-p)^{N-j} \frac{p^{a_{\text{post}}(x)-1}(1-p)^{b_{\text{post}}(x)-1}}{B\left(a_{\text{post}}(x), b_{\text{post}}(x)\right)}\mathrm{d}p
\end{aligned}
$$

Reprise : 28 janvier

Exercice 5

Calculer la loi prédictive dans un modèle avec de loi Binômiale, avec un prior Beta.
On appele cette loi la **loi beta-binômiale**.

$$
\begin{aligned}
\pi(x_i \mid p) &= \binom{N}{x_i} p^{x_i}(1-p)^{N-x_i} \\
\pi(p) &\propto p^{a-1}(1-p)^{b-1} \\
\pi(p \mid x_1, \cdots x_n) &\propto p^{a-1}(1-p)^{b-1} p^{\sum x_i}(1-p)^{Nn-\sum x_i} \\
p \mid x_1, \cdots x_n &\sim \mathrm{Beta}\left(a + \sum x_i, b + Nn - \sum x_i\right) = \mathrm{Beta}\left(a_{\text{post}}(x), b_{\text{post}}(x)\right) \\
\mathbb{P}\left(X_{n+1} = j \mid x_1, \cdots x_n\right) &= \int_{p=0}^{1} \mathbb{P}(X_{n+1} = j \mid p)\pi(p \mid x_1, \cdots x_n)\mathrm{d}p \\
&= \int_{p=0}^{1} \binom{N}{j} p^{j}(1-p)^{N-j} \frac{p^{a_{\text{post}}(x)-1}(1-p)^{b_{\text{post}}(x)-1}}{B\left(a_{\text{post}}(x), b_{\text{post}}(x)\right)} \mathrm{d}p \\
&= \binom{N}{j} \frac{B\left(a_{\text{post}}(x) + j, b_{\text{post}}(x) + N - j\right)}{B\left(a_{\text{post}}(x), b_{\text{post}}(x)\right)} \quad \textbf{Beta-Binômiale}
\end{aligned}
$$

Résumé de l'épisode 1

Synthèse

- **Comme en fréquentiste** : Modèle pour les données $\pi(x \mid \theta)$
- **Nouveauté :** Prior $\pi(\theta)$
- **Objectif :** Posterior $\pi(\theta \mid x)$.

Résumé de l'épisode 1

Synthèse

- **Comme en fréquentiste** : Modèle pour les données $\pi(x \mid \theta)$
- **Nouveauté :** Prior $\pi(\theta)$
- **Objectif :** Posterior $\pi(\theta \mid x)$.

Et la suite ?

1. Calcul du posterior quand c'est possible *(ex : conjugués)*
2. Sinon, calcul numérique *(Comment normaliser ?)*
3. Beaucoup de paramètres : simulations du posterior *(comment ?)*
4. Que faire du posterior ? *(et de la prédictive ?)*

Résumé de l'épisode 1

Synthèse

- **Comme en fréquentiste** : Modèle pour les données $\pi(x \mid \theta)$
- **Nouveauté :** Prior $\pi(\theta)$
- **Objectif :** Posterior $\pi(\theta \mid x)$.

Et la suite ?

1. Calcul du posterior quand c'est possible *(ex : conjugués)*
2. Sinon, calcul numérique *(Comment normaliser ?)*
3. Beaucoup de paramètres : simulations du posterior *(comment ?)*
4. Que faire du posterior ? *(et de la prédictive ?)*

Actualisation du prior

On a

$$\pi(\theta \mid x_1, \cdots x_{n+1}) \propto \pi(x_{n+1} \mid \theta)\pi(\theta \mid x_1, \cdots x_n).$$

On a donc **Ancien posterior** $=$ **Nouveau prior**

Premier TP

Vous pouvez en profiter pour commencer le projet, l'idée étant de mettre en application le cours en montant progressivement en difficulté, et en réexpliquant la démarche avec vos mots.

1. On considère les données de votre choix. En utilisant une loi adaptée pour les données (on travaille pour le moment avec des données à un seul paramètre e.g. : Bernoulli) admettant un prior conjugué *que l'on sait normaliser analytiquement*.
2. On va calculer le posterior de deux façons :
   - En utlisant la conjugaison **(méthode totalement analytique)**
   - En normalisant le posterior **par intégration numérique** (par exemple par la méthode des rectangles)
3. On observera le comportement du posterior en intégrant les données une à une *(cf Exos 2 et 5 du poly)*. Faire de même pour la loi de $\frac{1}{k}\sum_{i=1}^{k} X_{n+i} \mid X_1, \cdots X_n$, à k fixé arbitrairement (par exemple $k = 100$)
4. **Bonus** : Réfléchir aux deux questions suivantes :
   4.1 Comment simuler le posterior si l'on ne connaît pas la loi explicite ?
   4.2 Quel serait l'analogue Bayésien d'un intervalle de confiance, et comment le construire ?

Bilan

Après le TP_1

Vous savez désormais

1. Montrer qu'un prior est conjugué pour une certaine famille de loi
2. Trouver un prior conjugué pour une certaine famille de loi
3. Calculer le posterior théoriquement dans le cas de priors conjugués "normalisables"
4. Calculer le posterior numériquement *lorsqu'on a peu de paramètres*
5. Simuler ou calculer la prédictive

Bilan

Après le TP_1

Vous savez désormais

1. Montrer qu'un prior est conjugué pour une certaine famille de loi
2. Trouver un prior conjugué pour une certaine famille de loi
3. Calculer le posterior théoriquement dans le cas de priors conjugués "normalisables"
4. Calculer le posterior numériquement *lorsqu'on a peu de paramètres*
5. Simuler ou calculer la prédictive

Remarque :
Il reste à voir ce qu'on pourra faire en plus grande dimension

Bilan

Après le TP_1

Vous savez désormais

1. Montrer qu'un prior est conjugué pour une certaine famille de loi
2. Trouver un prior conjugué pour une certaine famille de loi
3. Calculer le posterior théoriquement dans le cas de priors conjugués "normalisables"
4. Calculer le posterior numériquement *lorsqu'on a peu de paramètres*
5. Simuler ou calculer la prédictive

Remarque :
Il reste à voir ce qu'on pourra faire en plus grande dimension

Modèles avec beaucoup de paramètres

En général, lorsque la dimension de l'espace de paramètres est grande, on aura recours à des méthodes numériques pour *simuler sous la loi a posteriori*

Il faudra donc savoir tout faire avec un échantillon tiré sous le posterior

Bilan

Après le TP_1

Vous savez désormais

1. Montrer qu'un prior est conjugué pour une certaine famille de loi
2. Trouver un prior conjugué pour une certaine famille de loi
3. Calculer le posterior théoriquement dans le cas de priors conjugués "normalisables"
4. Calculer le posterior numériquement *lorsqu'on a peu de paramètres*
5. Simuler ou calculer la prédictive

Remarque :
Il reste à voir ce qu'on pourra faire en plus grande dimension

Modèles avec beaucoup de paramètres

En général, lorsque la dimension de l'espace de paramètres est grande, on aura recours à des méthodes numériques pour *simuler sous la loi a posteriori*

Il faudra donc savoir tout faire avec un échantillon tiré sous le posterior

$\rightsquigarrow$ Des idées ?

Premières remarques sur la simulation

En petite dimension

Lorsqu'on peut normaliser ou borner le posterior, on peut utiliser n'importe quelle méthode de simulation, par exemple méthode du rejet.

1. On se donne une loi de proposition $p(\theta)$, idéalement la plus proche possible du posterior.
2. On trouve M tel que $\sup_\theta \frac{q(\theta|X)}{p(\theta)} < M$.
3. On pose $i = 0$, et on tire θ_0 selon p, et $U_0 \sim \mathcal{U}([0,1])$.
4. **Tant que** $MU_i > \frac{q(\theta_i|X)}{p(\theta_i)}$
   - On incrémente i
   - On tire θ_i selon p et $U_i \sim \mathcal{U}([0,1])$

Premières remarques sur la simulation

En petite dimension

Lorsqu'on peut normaliser ou borner le posterior, on peut utiliser n'importe quelle méthode de simulation, par exemple méthode du rejet.

1. On se donne une loi de proposition $p(\theta)$, idéalement la plus proche possible du posterior.
2. On trouve M tel que $\sup_\theta \frac{q(\theta|X)}{p(\theta)} < M$.
3. On pose $i = 0$, et on tire θ_0 selon p, et $U_0 \sim \mathcal{U}([0,1])$.
4. **Tant que** $MU_i > \frac{q(\theta_i|X)}{p(\theta_i)}$
   - On incrémente i
   - On tire θ_i selon p et $U_i \sim \mathcal{U}([0,1])$

Méthode du rejet pour simulation exacte

On peut aussi appliquer directement la méthode du rejet pour simuler le posterior :

1. On note $\mathcal{D} = \{x_1, \cdots, x_n\}$ les observations, on pose $i = 0$, et on tire θ_0 selon $\pi(\theta)$, et $d_0 = \{y_1, \cdots, y_n\}$ i.i.d. selon $\pi(x \mid \theta_0)$.
2. **Tant que** $d_i \neq \mathcal{D}$
   - On incrémente i
   - On tire θ_i selon $\pi(\theta)$, et $d_i = \{y_1, \cdots, y_n\}$ i.i.d. selon $\pi(x \mid \theta_0)$.

Premières remarques sur la simulation

En petite dimension

Lorsqu'on peut normaliser ou borner le posterior, on peut utiliser n'importe quelle méthode de simulation, par exemple méthode du rejet.

1. On se donne une loi de proposition $p(\theta)$, idéalement la plus proche possible du posterior.
2. On trouve M tel que $\sup_\theta \frac{q(\theta|X)}{p(\theta)} < M$.
3. On pose $i = 0$, et on tire θ_0 selon p, et $U_0 \sim \mathcal{U}([0,1])$.
4. **Tant que** $MU_i > \frac{q(\theta_i|X)}{p(\theta_i)}$
   - On incrémente i
   - On tire θ_i selon p et $U_i \sim \mathcal{U}([0,1])$

Méthode du rejet pour simulation exacte

On peut aussi appliquer directement la méthode du rejet pour simuler le posterior :

1. On note $\mathcal{D} = \{x_1, \cdots, x_n\}$ les observations, on pose $i = 0$, et on tire θ_0 selon $\pi(\theta)$, et $d_0 = \{y_1, \cdots, y_n\}$ i.i.d. selon $\pi(x \mid \theta_0)$.
2. **Tant que** $d_i \neq \mathcal{D}$
   - On incrémente i
   - On tire θ_i selon $\pi(\theta)$, et $d_i = \{y_1, \cdots, y_n\}$ i.i.d. selon $\pi(x \mid \theta_0)$.

Remarque : Si beaucoup d'observations, impossible à utiliser car $\mathbb{P}(d_i = \mathcal{D}) << 1$!

Premières remarques sur la simulation

En petite dimension

Lorsqu'on peut normaliser ou borner le posterior, on peut utiliser n'importe quelle méthode de simulation, par exemple méthode du rejet.

1. On se donne une loi de proposition $p(\theta)$, idéalement la plus proche possible du posterior.
2. On trouve M tel que $\sup_\theta \frac{q(\theta|X)}{p(\theta)} < M$.
3. On pose $i = 0$, et on tire θ_0 selon p, et $U_0 \sim \mathcal{U}([0,1])$.
4. **Tant que** $MU_i > \frac{q(\theta_i|X)}{p(\theta_i)}$
   - On incrémente i
   - On tire θ_i selon p et $U_i \sim \mathcal{U}([0,1])$

Méthode du rejet pour simulation exacte

On peut aussi appliquer directement la méthode du rejet pour simuler le posterior :

1. On note $\mathcal{D} = \{x_1, \cdots, x_n\}$ les observations, on pose $i = 0$, et on tire θ_0 selon $\pi(\theta)$, et $d_0 = \{y_1, \cdots, y_n\}$ i.i.d. selon $\pi(x \mid \theta_0)$.
2. **Tant que** $d_i \neq \mathcal{D}$
   - On incrémente i
   - On tire θ_i selon $\pi(\theta)$, et $d_i = \{y_1, \cdots, y_n\}$ i.i.d. selon $\pi(x \mid \theta_0)$.

Remarque : Si beaucoup d'observations, impossible à utiliser car $\mathbb{P}(d_i = \mathcal{D}) << 1$!
$\rightsquigarrow$ *On verra d'autres méthodes (ex : ABC) plus tard*

Plan

Prior, posterior et prédictive
- Introduction
- Modèle
- Conjugaison de priors

Utilisation de la loi à posteriori
- Regions de crédibilité
- Estimateurs ponctuels et théorie de la décision
- Tests

Aspects numériques
- Monte-Carlo, rejet et ABC
- Metropolis-Hasting
- Priors semi-conjugués et alogirthme de Gibbs

Choix de priors et modèles Hierarchiques

Regions de crédibilité

Definition

Une région $R_\alpha(X) \subset \Theta$ est appelée région de crédibilité au niveau α si

$$\pi(\theta \in R_\alpha(x)|x) = 1 - \alpha.$$

Regions de crédibilité

Definition

Une région $R_\alpha(X) \subset \Theta$ est appelée région de crédibilité au niveau α si

$$\pi(\theta \in R_\alpha(x)|x) = 1 - \alpha.$$

Definition

Une région *HPD* (Highest Probability Density) est une région de la forme

$$R_h^{(HPD)}(x) = \{t \in \Theta, f_{\theta|X}(t,x) \geq h\},$$

La région *HPD* $(1-\alpha)$-crédible est donnée par $R_\alpha^{(HPD)}(x) = R_{h_\alpha}^{(HPD)}(x)$, où

$$h_\alpha = \sup\{h > 0, \pi(\theta \in R_h^{(HPD)}(x)|x) > 1 - \alpha\}.$$

Regions de crédibilité

Definition

Une région $R_\alpha(X) \subset \Theta$ est appelée région de crédibilité au niveau α si

$$\pi(\theta \in R_\alpha(x)|x) = 1 - \alpha.$$

Definition

Une région *HPD* (Highest Probability Density) est une région de la forme

$$R_h^{(HPD)}(x) = \{t \in \Theta, f_{\theta|X}(t, x) \geq h\},$$

La région *HPD* $(1-\alpha)$-crédible est donnée par $R_\alpha^{(HPD)}(x) = R_{h_\alpha}^{(HPD)}(x)$, où

$$h_\alpha = \sup\{h > 0, \pi(\theta \in R_h^{(HPD)}(x)|x) > 1 - \alpha\}.$$

- Il s'agit des région de crédibilité de niveau α de volume minimal
- Si la densité du posterior est continue et unimodale, elle sont connexes

TP 2

Exercice 6 [*En machine : TP_2*]. Voir consignes détaillées sur la page du cours

On considère les données de votre choix, pour lesquelles la loi choisie admet un prior conjugué

1. En parcourant différentes valeurs possibles pour les quantiles q_1, q_2 qui donnent des régions de crédibilité au niveau α, construire la région HPD (donc de volume minimal) au niveau $\alpha = 0.1$.
2. En parcourant différentes valeurs de seuil possibles, construire la région HPD au niveau $\alpha = 0.1$.
3. En simulant selon la loi a posteriori, et en utilisant les quantiles empiriques, construire la région HPD au niveau $\alpha = 0.1$.

TP 2

Exercice 6 [*En machine : TP_2*]. Voir consignes détaillées sur la page du cours

On considère les données de votre choix, pour lesquelles la loi choisie admet un prior conjugué

1. En parcourant différentes valeurs possibles pour les quantiles q_1, q_2 qui donnent des régions de crédibilité au niveau α, construire la région HPD (donc de volume minimal) au niveau $\alpha = 0.1$.
2. En parcourant différentes valeurs de seuil possibles, construire la région HPD au niveau $\alpha = 0.1$.
3. En simulant selon la loi a posteriori, et en utilisant les quantiles empiriques, construire la région HPD au niveau $\alpha = 0.1$.

Exercice 7

On se place dans un modèle $\mathcal{N}(m, \sigma^2)$, avec σ^2 connu, et on choisit pour m un prior $\mathcal{N}(0, 1)$.

1. Rappeler l'expression, en fonction des observations X et des quantiles d'une loi normale, l'intervalle de confiance fréquentiste au niveau $\alpha = 0.05$
2. Donner, en fonction des observations X et des quantiles d'une loi normale, la région *HPD* au niveau $\alpha = 0.05$.

TP 2

Exercice 6 [*En machine : TP_2*]. Voir consignes détaillées sur la page du cours

On considère les données de votre choix, pour lesquelles la loi choisie admet un prior conjugué

1. En parcourant différentes valeurs possibles pour les quantiles q_1, q_2 qui donnent des régions de crédibilité au niveau α, construire la région HPD (donc de volume minimal) au niveau $\alpha = 0.1$.
2. En parcourant différentes valeurs de seuil possibles, construire la région HPD au niveau $\alpha = 0.1$.
3. En simulant selon la loi a posteriori, et en utilisant les quantiles empiriques, construire la région HPD au niveau $\alpha = 0.1$.

Exercice 7

On se place dans un modèle $\mathcal{N}(m, \sigma^2)$, avec σ^2 connu, et on choisit pour m un prior $\mathcal{N}(0, 1)$.

1. Rappeler l'expression, en fonction des observations X et des quantiles d'une loi normale, l'intervalle de confiance fréquentiste au niveau $\alpha = 0.05$
2. Donner, en fonction des observations X et des quantiles d'une loi normale, la région *HPD* au niveau $\alpha = 0.05$.

Bilan provisoire

On sait désormais

1. Utiliser les données pour *actualiser le prior*
2. Construire une *région de crédibilité*
3. Calculer une *loi prédictive*

Estimation ponctuelle

Remarque : *D'un point de vue Bayésien, on préferera donner tout le posterior que le résumé en un "estimateur".*

Estimation ponctuelle

Remarque : *D'un point de vue Bayésien, on préferera donner tout le posterior que le résumé en un "estimateur".*
Cela dit, plusieurs idées semblent naturelles.

- Utiliser l'espérance a posteriori $\mathbb{E}[\theta \mid X]$
- Utiliser la médiante a posteriori Median $\left(\pi(\theta \mid X)\right)$
- Utiliser le maximum a posteriori $(MAP = \arg\max_t f_{\theta|X}(t,x))$

Estimation ponctuelle

Remarque : *D'un point de vue Bayésien, on préferera donner tout le posterior que le résumé en un "estimateur".*
Cela dit, plusieurs idées semblent naturelles.

- Utiliser l'espérance a posteriori $\mathbb{E}[\theta \mid X]$
- Utiliser la médiante a posteriori Median $\left(\pi(\theta \mid X)\right)$
- Utiliser le maximum a posteriori $(MAP = \arg\max_t f_{\theta|X}(t, x))$

Remarque : Il existe un cadre naturel pour considérer ces estimateurs, celui de la *Théorie de la décision*.

Estimation ponctuelle

Remarque : *D'un point de vue Bayésien, on préferera donner tout le posterior que le résumé en un "estimateur".*
Cela dit, plusieurs idées semblent naturelles.

- Utiliser l'espérance a posteriori $\mathbb{E}[\theta \mid X]$
- Utiliser la médiante a posteriori Median $\Big(\pi(\theta \mid X)\Big)$
- Utiliser le maximum a posteriori $(MAP = \arg\max_t f_{\theta|X}(t,x))$

Remarque : Il existe un cadre naturel pour considérer ces estimateurs, celui de la *Théorie de la décision*.

Definition

On appellera *décision* une fonction (mesurable) des observations $x \mapsto \delta(x)$ dans une espace (mesurable) $(E, \mathcal{A})$.

Fonction de perte

On se donne une fonction (mesurable) $L : \Theta \times E \to \mathbb{R}$ appelée *fonction de perte* ou *fonction de côut* (ex : $\mathbb{L}^2$, $\mathbb{L}^1$, perte $\{0,1\}$...)

Estimation ponctuelle

Remarque : *D'un point de vue Bayésien, on préferera donner tout le posterior que le résumé en un "estimateur".*
Cela dit, plusieurs idées semblent naturelles.

- Utiliser l'espérance a posteriori $\mathbb{E}[\theta \mid X]$
- Utiliser la médiante a posteriori Median $\left(\pi(\theta \mid X)\right)$
- Utiliser le maximum a posteriori $(MAP = \arg\max_t f_{\theta|X}(t, x))$

Remarque : Il existe un cadre naturel pour considérer ces estimateurs, celui de la *Théorie de la décision*.

Definition

On appellera *décision* une fonction (mesurable) des observations $x \mapsto \delta(x)$ dans une espace (mesurable) $(E, \mathcal{A})$.

Fonction de perte

On se donne une fonction (mesurable) $L : \Theta \times E \to \mathbb{R}$ appelée *fonction de perte* ou *fonction de côut* (ex : $\mathbb{L}^2$, $\mathbb{L}^1$, perte $\{0, 1\}$...)

Remarque : On cherchera alors une décision qui minimise "en moyenne", la fonction de perte, i.e. le coût moyen (*a posteriori*).

Quelques risques différents

Definition

Le **risque fréquentiste** d'une décision δ associé à une perte L s'écrit :

$$R_\delta(\theta) = \mathbb{E}[L(\theta, \delta(X))|\theta] = \int_x L(\theta, \delta(x))\pi(x|\theta)\mathrm{d}x.$$

Quelques risques différents

Definition

Le **risque fréquentiste** d'une décision δ associé à une perte L s'écrit :

$$R_\delta(\theta) = \mathbb{E}[L\big(\theta, \delta(X)\big)|\theta] = \int_x L\big(\theta, \delta(x)\big)\pi(x|\theta)\mathrm{d}x.$$

Definition

Le **risque a posteriori** d'une décision δ associé à une perte L s'écrit :

$$\rho_\delta(X) = \mathbb{E}[L\big(\theta, \delta(X)\big)|X] = \int_\theta L\big(\theta, \delta(X)\big)\pi(\theta|X)\mathrm{d}\theta$$

Quelques risques différents

Definition

Le **risque fréquentiste** d'une décision δ associé à une perte L s'écrit :

$$R_\delta(\theta) = \mathbb{E}[L(\theta, \delta(X))|\theta] = \int_x L(\theta, \delta(x))\pi(x|\theta)\mathrm{d}x.$$

Definition

Le **risque a posteriori** d'une décision δ associé à une perte L s'écrit :

$$\rho_\delta(X) = \mathbb{E}[L(\theta, \delta(X))|X] = \int_\theta L(\theta, \delta(X))\pi(\theta|X)\mathrm{d}\theta$$

Definition

Le **risque intégré** d'une décision δ associé à une perte L s'écrit :

$$\begin{aligned} r_\delta &= \mathbb{E}[L(\theta, \delta(X))] \\ &= \mathbb{E}[R_\delta(\theta)] = \int_\theta \int_x L(\theta, \delta(x))\pi(x|\theta)\mathrm{d}x\pi(\theta)\mathrm{d}\theta \\ &= \mathbb{E}[\rho_\delta(X)] = \int_x \int_\theta L(\theta, \delta(x))\pi(\theta|x)\mathrm{d}\theta\pi(x)\mathrm{d}x \end{aligned}$$

Décision bayésienne

Definition

Si $\min_\delta r(\delta) < +\infty$, on définit une **règle de décision Bayésienne** pour le prior π comme une règle de décision qui minimise le risque intégré : δ_π est une règle de décision Bayésienne si

$$\delta_\pi \in \arg\min_\delta r_\pi(\delta).$$

Décision bayésienne

Definition

Si $\min_\delta r(\delta) < +\infty$, on définit une **règle de décision Bayésienne** pour le prior π comme une règle de décision qui minimise le risque intégré : δ_π est une règle de décision Bayésienne si

$$\delta_\pi \in \arg\min_\delta r_\pi(\delta).$$

On pose

$$\delta^\star_\pi(x) \in \arg\min_\delta \rho_\delta(x).$$

Theorem

$$\boxed{\textit{La décision } \delta^\star_\pi \textit{ est une décision bayésienne.}}$$

Si de plus elle est définie de façon unique quelque soit x, alors c'est une décision admissible : Il n'existe pas de décision δ_2 telle que

$$\forall \theta \in \Theta, R^{\delta_2}(\theta) \leq R^{\delta^\star}(\theta),$$

avec inégalité stricte pour au moins un θ.

Décision bayésienne

Definition

Si $\min_\delta r(\delta) < +\infty$, on définit une **règle de décision Bayésienne** pour le prior π comme une règle de décision qui minimise le risque intégré : δ_π est une règle de décision Bayésienne si

$$\delta_\pi \in \arg\min_\delta r_\pi(\delta).$$

On pose

$$\delta^\star_\pi(x) \in \arg\min_\delta \rho_\delta(x).$$

Theorem

$$\boxed{\textit{La décision } \delta^\star_\pi \textit{ est une décision bayésienne.}}$$

Si de plus elle est définie de façon unique quelque soit x, alors c'est une décision admissible : Il n'existe pas de décision δ_2 telle que

$$\forall \theta \in \Theta, R^{\delta_2}(\theta) \leq R^{\delta^\star}(\theta),$$

avec inégalité stricte pour au moins un θ.

Fonction de perte "prédictive"

Exercice 8

1. On considère les fonction de coût données dans la partie précédente. Dans chacun des cas, donner la décision bayésienne qui minimise le risque intégré.
2. Donner la fonction de coût associé à un classifieur.

Fonction de perte "prédictive"

Exercice 8

1. On considère les fonction de coût données dans la partie précédente. Dans chacun des cas, donner la décision bayésienne qui minimise le risque intégré.
2. Donner la fonction de coût associé à un classifieur.

Remarque : On voit dans cet exercice l'intérêt d'introduire une nouvelle fonction de coût "prédictive" qui concerne des nouvelles observations.

Definition

On dira qu'une fonction $L_p : \mathbb{R} \times E \to \mathbb{R}$ est une fonction de perte **prédictive** si la fonction de perte L s'écrit

$$L(\theta, \delta) = \int_{x_{n+1}} L_p(x_{n+1}, \delta)\pi(x_{n+1} \mid \theta)\mathrm{d}x_{n+1}.$$

Fonction de perte "prédictive"

Exercice 8

1. On considère les fonction de coût données dans la partie précédente. Dans chacun des cas, donner la décision bayésienne qui minimise le risque intégré.
2. Donner la fonction de coût associé à un classifieur.

Remarque : On voit dans cet exercice l'intérêt d'introduire une nouvelle fonction de coût "prédictive" qui concerne des nouvelles observations.

Definition

On dira qu'une fonction $L_p : \mathbb{R} \times E \to \mathbb{R}$ est une fonction de perte **prédictive** si la fonction de perte L s'écrit

$$L(\theta, \delta) = \int_{x_{n+1}} L_p(x_{n+1}, \delta)\pi(x_{n+1} \mid \theta)\mathrm{d}x_{n+1}.$$

Dans ce cas, le risque a posteriori s'exprime en fonction de la loi prédictive (on note $x = (x_1, \cdots, x_n)$):

$$\begin{aligned}
\rho_\delta(x) &= \int_\theta L(\theta, \delta)\pi(\theta \mid x)\mathrm{d}\theta \\
&= \int_\theta \int_{x_{n+1}} L_p(x_{n+1}, \delta)\pi(x_{n+1} \mid \theta)\mathrm{d}x_{n+1}\pi(\theta \mid x)\mathrm{d}\theta \\
&= \int_{x_{n+1}} L_p(x_{n+1}, \delta)\pi(x_{n+1} \mid x)\mathrm{d}x_{n+1}
\end{aligned}$$

Tests et facteurs de Bayes

Perte $0 - 1$

On considère, pour $\Theta_0, \Theta_1 \subset \Theta$, la fonction de perte suivante :

$$L(\delta, \theta) = c_{fn} \mathbb{1}_{\delta=0} \mathbb{1}_{\theta \in \Theta_1} + c_{fp} \mathbb{1}_{\delta=1} \mathbb{1}_{\theta \in \Theta_0}.$$

Remarque : On supposera dans la suite deux cadres :

- Soit $\pi(\Theta_0) = \pi(\Theta_1) = 0$
- Soit $\pi(\Theta_0), \pi(\Theta_1) > 0$

Dans la suite, on $\pi(\Theta_0 \mid X)$ désignera donc soit une densité (si $\pi(\Theta_k) = 0$) soit une probabité. Attention à choisir les priors en conséquence (en particulier si l'on veut tester un singleton contre un ouvert).

Tests et facteurs de Bayes

Perte $0 - 1$

On considère, pour $\Theta_0, \Theta_1 \subset \Theta$, la fonction de perte suivante :

$$L(\delta, \theta) = c_{fn} \mathbb{1}_{\delta=0} \mathbb{1}_{\theta \in \Theta_1} + c_{fp} \mathbb{1}_{\delta=1} \mathbb{1}_{\theta \in \Theta_0}.$$

Remarque : On supposera dans la suite deux cadres :

- Soit $\pi(\Theta_0) = \pi(\Theta_1) = 0$
- Soit $\pi(\Theta_0), \pi(\Theta_1) > 0$

Dans la suite, on $\pi(\Theta_0 \mid X)$ désignera donc soit une densité (si $\pi(\Theta_k) = 0$) soit une probabité. Attention à choisir les priors en conséquence (en particulier si l'on veut tester un singleton contre un ouvert).

Décision bayésienne

$$\delta^\star(X) = \mathbb{1}_{\frac{\pi(\Theta_1 | X)}{\pi(\Theta_0 | X)} > \frac{c_{fp}}{c_{fn}}}$$

Tests et facteurs de Bayes

Perte 0 — 1

On considère, pour $\Theta_0, \Theta_1 \subset \Theta$, la fonction de perte suivante :

$$L(\delta, \theta) = c_{fn} \mathbb{1}_{\delta=0} \mathbb{1}_{\theta \in \Theta_1} + c_{fp} \mathbb{1}_{\delta=1} \mathbb{1}_{\theta \in \Theta_0}.$$

Remarque : On supposera dans la suite deux cadres :

- Soit $\pi(\Theta_0) = \pi(\Theta_1) = 0$
- Soit $\pi(\Theta_0), \pi(\Theta_1) > 0$

Dans la suite, on $\pi(\Theta_0 \mid X)$ désignera donc soit une densité (si $\pi(\Theta_k) = 0$) soit une probabité. Attention à choisir les priors en conséquence (en particulier si l'on veut tester un singleton contre un ouvert).

Décision bayésienne

$$\delta^\star(X) = \mathbb{1}_{\frac{\pi(\Theta_1 | X)}{\pi(\Theta_0 | X)} > \frac{c_{fp}}{c_{fn}}}$$

Definition (Odds)

$$\text{Odds}_{prior} = \frac{\pi(\Theta_1)}{\pi(\Theta_0)}$$

$$\text{Odds}_{post} = \frac{\pi(\Theta_1 \mid X)}{\pi(\Theta_0 \mid X)}$$

Application

Exercice 9

On reprend le cadre de l'exercice 7 : On se place dans un modèle $\mathcal{N}(m, \sigma^2)$, avec σ^2 connu, et on choisit pour m un prior $\mathcal{N}(0, 1)$.

1. On considère la fonction de perte

$$L(\theta, \delta) = (\theta - \delta)\mathbb{1}_{\theta > \delta} + 2(\delta - \theta)\mathbb{1}_{\delta > \theta}.$$

   Donner la décision bayésienne qui minimise le risque a posteriori.

2. On considère la fonction de perte prédictive

$$Lp(x_{n+1}, \delta) = (x_{n+1} - \delta)\mathbb{1}_{x_{n+1} > \delta} + 2(x_{n+1} - \theta)\mathbb{1}_{x_{n+1} > \theta}.$$

   Donner la décision bayésienne qui minimise le risque a posteriori.

Application

Exercice 9

On reprend le cadre de l'exercice 7 : On se place dans un modèle $\mathcal{N}(m, \sigma^2)$, avec σ^2 connu, et on choisit pour m un prior $\mathcal{N}(0, 1)$.

1. On considère la fonction de perte

$$L(\theta, \delta) = (\theta - \delta)\mathbb{1}_{\theta > \delta} + 2(\delta - \theta)\mathbb{1}_{\delta > \theta}.$$

   Donner la décision bayésienne qui minimise le risque a posteriori.

2. On considère la fonction de perte prédictive

$$Lp(x_{n+1}, \delta) = (x_{n+1} - \delta)\mathbb{1}_{x_{n+1} > \delta} + 2(x_{n+1} - \theta)\mathbb{1}_{x_{n+1} > \theta}.$$

   Donner la décision bayésienne qui minimise le risque a posteriori.

Remarque : Les calculs faits pour associer une fonction de perte a une décision pourront être aisément adaptés avec une fonction de perte prédictive.

Application

Exercice 10

Vous organisez une conférence, et vous devez effectuer les réservations. 100 personnes se sont inscrites. Les années passées, sur 100 inscrit-e-s, seulement 79 et 85 étaient effectivement venues.
La politique de réservation de l'hôtel est la suivante :

- En cas de sur-réservation, vous devez payer 300 euros de dédommagement par chambre sur-réservée.
- En cas de sous-réservation, cela vous coûte 100 euros de plus par chambre non réservée à l'avance.

En bon mathématicien-ne, vous décidez de formaliser le problème afin de trouver le nombre de chambres à réserver.

1. Si on note X_1 et X_2 la variable aléatoire correspondant au nombre de personnes effectivement venues les années précédentes, et X_3 celle correspondant au nombre qui vont venir cette année, par quelle loi peut-on modéliser ces variables ? (on notera N, p les paramètres).
2. En bon Bayésien-ne, vous allez considérer que X_1, X_2, X_3 sont i.i.d. conditionnellement à p, et mettre un prior sur p. Quelle famille de prior peut-on choisir pour que les calculs soient simples ? Quels hyperparamètres a_0, b_0 correspondent à un prior uniforme dans ce cadre ? À partir de maintenant, on travaille avec ce prior.
3. On note $\mathcal{D} = (X_1, X_2)$ les données observées, et $d(\mathcal{D})$ la décision correspondant au nombre de chambres réservées que vous devez prendre. Écrire la fonction Lp donnant un coût correspondant à la politique de l'hôtel, en fonction de X_3 et de la décision d.
4. Écrire la fonction L donnant, en fonction d'un paramètre p, et de la décision d, le coût prédictif.
5. Écrire la fonction rho donnant, en fonction des hyperparamètres a, b du posterior, et de la décision d, le coût à posteriori.
6. Écrire la fonction $\mathrm{minimpost}$ donnant, en fonction des hyperparamètres a, b du posterior, la décision optimale pour ce problème de décision.
7. En utilisant les données de l'énoncé, calculer les hyperparamètres du posterior, et donner la décision optimale.
8. Faire l'analyse théorique du problème : obtenir la décision optimale comme un quantile de la loi prédictive, en résolvant le problème de minimisation.
9. Vérifier la valeur numérique de $d(\mathcal{D})$ ainsi obtenue, en utilisant la fonction $\mathrm{qbetabinom}$ du package rmutil, qui donne les quantiles de la loi betabinomiale.
10. L'hôtel vous propose une nouvelle politique, plutôt que payer 300 euros par chambre en sur-réservation, il vous proposer de payer un forfait de 1000 euros en cas de sur-réservation (indépendamment du nombre de chambres). Que devient la décision optimale avec cette politique ? Quelle politique préféreriez-vous ?

Plan

Prior, posterior et prédictive
- Introduction
- Modèle
- Conjugaison de priors

Utilisation de la loi à posteriori
- Regions de crédibilité
- Estimateurs ponctuels et théorie de la décision
- Tests

Aspects numériques
- Monte-Carlo, rejet et ABC
- Metropolis-Hasting
- Priors semi-conjugués et alogirthme de Gibbs

Choix de priors et modèles Hierarchiques

Plan

Prior, posterior et prédictive

Introduction

Modèle

Conjugaison de priors

Utilisation de la loi à posteriori

Regions de crédibilité

Estimateurs ponctuels et théorie de la décision

Tests

Aspects numériques

Monte-Carlo, rejet et ABC

Metropolis-Hasting

Priors semi-conjugués et alogirthme de Gibbs

Choix de priors et modèles Hierarchiques