

Chapitre 5

Information a priori et lois a priori

Le choix de la loi *a priori* est l'aspect le plus critiqué de l'analyse statistique bayésienne. En théorie, cette loi doit représenter notre information *a priori* sur les paramètres θ . Mais la notion de l'information est une notion très vague, et de toute façon, en pratique, cette information est rarement fournie sous la forme probabiliste. Plusieurs approches sont alors proposées:

- Recours à des lois usuelles d'une manière ad hoc.
- Détermination subjective de la même manière qu'on construit une fonction d'utilité ou de coût en bâtissant une échelle de vraisemblance relative pour des valeurs de θ .
- Détermination de la loi en utilisant le principe du maximum d'entropie lorsque certaines caractéristiques de la loi comme ses moments ou ses fractiles sont connues.
- Détermination de la loi en utilisant la méthode des moments lorsqu'une forme paramétrique de la loi *a priori* et un certain nombre de ses caractéristiques sont disponibles.
- Détermination de la loi en utilisant un certain nombre de principes d'invariance. Des exemples sont les lois non informatives, les lois de références, les lois conjuguées.

Dans ce chapitre nous allons détailler un peu plus ces diverses approches.

5.1 Approche pragmatique : lois usuelles

Exemple 12 Supposons que l'on a choisi pour les observations et pour les paramètres des lois normales: $x_1, \dots, x_n \sim \mathbf{N}(x|\theta, 1)$ et $\theta \sim \mathbf{N}(\theta|\mu, \tau)$. La loi *a posteriori* est alors une loi normale et si on décide de choisir comme estimateur de θ la moyenne ou le maximum *a posteriori*, on a

$$\hat{\theta} = \frac{\bar{x}\tau + \mu/n}{\tau + 1/n}.$$

On peut alors interpréter la quantité d'information apportée par un tel choix de la loi *a priori* par un équivalent échantillons de τ avec une moyenne empirique de μ . Dans ce cas l'hyperparamètre $1/\tau$ est homogène à une taille d'échantillon. En

effet l'estimateur du maximum de vraisemblance de θ basé seulement sur la loi des observations est \bar{x} , alors que l'estimateur du maximum *a posteriori* de θ basé sur les mêmes échantillons et la loi *a priori* est équivalent à son estimateur *a priori* si on avait observé $1/\tau$ échantillons supplémentaires avec la moyenne empirique μ .

5.2 Méthode des moments pour la détermination des paramètres

Si on dispose d'une forme paramétrique pour une loi et un certain nombre de ses moments, on peut essayer de déterminer ses paramètres en résolvant les équations qui les relient.

Exemple 13 Soit $X_i \sim \mathbf{Bet}(x|n_i, p_i)$, nombre de succès au baccalauréat pour une classe de n_i élèves. Des observations sur des années antérieures nous informent que la moyenne des p_i était de 0.70, avec une variance de 0.1. Supposons que l'on a choisit comme la loi *a priori* pour p_i la loi Bêta, $\mathbf{Bet}(p|\alpha, \beta)$. Il s'agit alors de déterminer les deux hyperparamètres α et β .

Connaissant les relations qui lient la moyenne et la variance d'une loi Bêta à ses paramètres

$$E[p] = \frac{\alpha}{\alpha + \beta} = 0.7, \quad \text{Var}[p] = \frac{\alpha\beta}{(\alpha + \beta)^2} = 0.1,$$

on obtient les valeurs de $\alpha = 0.77$ et $\beta = 0.33$.

La méthode des moments a un certain nombre de défauts. Par exemple, la méthode n'assure ni l'existence, ni l'unicité d'une solution. De plus, quand elle existe, la solution peut ne pas être satisfaisante (valeurs impossible de paramètre). Par exemple, elle pourra produire des valeurs négatives pour α et β dans cet exemple.

5.3 Principe du maximum d'entropie

5.4 Lois a priori non informatives

Lorsqu'on ne dispose pas d'information pertinente pour choisir une loi *a priori* et on veut cependant utiliser l'approche bayésienne on peut faire appel à des lois *a priori* non informatives. Laplace fut le premier à utiliser cette technique, puisque, ne disposant d'aucune information *a priori* sur le nombre de boules blanches dans l'urne, ou sur la proportion de ces boules, il a attribué à ces paramètres une loi *a priori* uniforme. Un tel raisonnement semble intuitivement correct, puisqu'il repose sur le principe d'indifférence et de l'équiprobabilité des événements élémentaires. Si pour un paramètre qui peut prendre un nombre fini de valeurs (Θ de dimension finie) ceci semble raisonnable et ne pose aucun problème, il n'est pas de même comme nous le verrons par la suite pour un paramètre continu.

Deux critiques ont été avancées pour cette approche. La première est que les lois uniformes n'existent pas sur des espaces non bornés et qu'il faut utiliser des lois impropres. La deuxième qui est plus fondamentale est la manque d'invariance par reparamétrisation. Si on reparamétrise la loi des observations $p(x|\theta)$ en $p(x|\phi)$ en passant par une transformation bijective $\phi = g(\theta)$, l'information *a priori* n'est pas modifiée, puisque inexistante, donc on devrait aussi utiliser une loi *a priori* non informative pour ϕ . Mais, si $p(\theta) = 1$, la loi de ϕ sera $\pi(\phi) = \left| \frac{d}{d\phi} g^{-1}(\phi) \right|$ par la formule du Jacobien. Elle est donc généralement non constante. Il y a donc un paradoxe. C'est pourquoi, pour aboutir à une notion plus acceptable des lois non informatives, il faut proposer une définition d'*invariance par reparamétrisation*. Une première possibilité est d'utiliser les structures de groupe de transformation \mathcal{G} opérant sur l'espace des observations \mathcal{X} qui induisent les groupe \mathcal{G}^* opérant sur l'espace des paramètres Θ . Deux exemples très usuels sont *invariance par translation* et *invariance par changement d'échelle*.

Proposition 16 [invariance par translation] La famille de lois $p(x|\theta) = g(x - \theta)$ est *invariant par translation*, c'est-à-dire que si $x \sim p(x|\theta)$ alors $y = x - x_0 \sim p(y|\theta - x_0)$ de la même famille; θ est alors dit *paramètre de position*. Il est alors naturel de choisir une loi *a priori* pour θ qui soit invariant par translation

$$\forall \theta, \theta_0 \in \mathbb{R}, \quad p(\theta) = p(\theta - \theta_0).$$

La solution est $p(\theta) = c$, loi uniforme sur Θ .

Proposition 17 [invariance par changement d'échelle] La famille de lois $p(x|\sigma) = \frac{1}{\sigma} g\left(\frac{x}{\sigma}\right)$ est *invariant par changement d'échelle*, c'est-à-dire que si $x \sim p(x|\sigma)$ alors $y = x/\sigma \sim p(y|1)$ de la même famille; σ est alors dit *paramètre d'échelle*. Il est alors naturel de choisir une loi *a priori* pour σ qui soit invariant par changement d'échelle

$$\forall \sigma, k \in \mathbb{R}^+, \pi(\sigma) = k\pi(k\sigma).$$

La solution est $\pi(\sigma) = \frac{c}{\sigma}$, où c est une constante positive. On peut constater que cette loi n'est plus uniforme.

Lois *a priori* non informatives de Jeffreys sont la généralisation de cette approche. En effet Jeffreys (1961) propose une approche plus globale, basée sur la matrice d'information de Fishere, qui évite de prendre en compte une structure invariante particulière de la loi des observations.

Définition 13 [Lois a priori non informatives de Jeffreys] Les lois a priori non informatives de Jeffreys (dans le cas unidimensionnel) sont définies par

$$p(\theta) \propto I^{\frac{1}{2}}(\theta),$$

où $I(\theta)$ est la quantité de l'information de Fisher

$$I(\theta) = \mathbb{E}_\theta \left\{ \frac{\partial \log p(x|\theta)}{\partial \theta} \right\}^2,$$

Ce qui, sous certaines conditions de régularité, est égale à

$$I(\theta) = -\mathbb{E}_\theta \left\{ \frac{\partial^2 \log p(x|\theta)}{\partial \theta^2} \right\}.$$

Dans le cas où $\theta \in \mathbb{R}^k$, on définit la matrice de l'information de Fisher qui a pour éléments

$$I_{ij}(\theta) = -\mathbb{E}_\theta \left\{ \frac{\partial^2 \log p(x|\theta)}{\partial \theta_i \partial \theta_j} \right\}, \quad i, j = 1, \dots, k,$$

et la loi non informative de Jeffreys est alors

$$p(\theta) \propto |I(\theta)|^{\frac{1}{2}}.$$

Notons que, si $p(x|\theta)$ définit une famille exponentielle,

$$p(x|\theta) = h(x) \exp[\boldsymbol{\theta} \mathbf{x} - \psi(\boldsymbol{\theta})],$$

on aura $I(\theta) = \nabla \nabla^t \psi(\boldsymbol{\theta})$, et

$$p(\theta) = \left(\prod_{i=1}^k \frac{\partial^2 \psi(\theta)}{\partial \theta^2} \right)^{\frac{1}{2}}.$$

Exemple 14 Soit $x \sim \mathbf{N}(\mu, \sigma^2)$ avec $\theta = (\mu, \sigma^2)$ inconnu. Dans ce cas,

$$I(\theta) = -\mathbb{E}_\theta \left\{ \begin{pmatrix} 1/\sigma^2 & 2(x-\mu)/\sigma^3 \\ 2(x-\mu)/\sigma^3 & 3(x-\mu)^2/\sigma^4 - 1/\sigma^2 \end{pmatrix} \right\} = \begin{pmatrix} 1/\sigma^2 & 0 \\ 0 & 2/\sigma^2 \end{pmatrix},$$

et la loi non informative associée est

$$p(\theta) = \pi(\mu, \sigma^2) \propto 1/\sigma^2.$$

Deux critiques ont été faites à cette approche:

- Contradiction avec le principe de vraisemblance, puisque l'information de Fisher dépend des facteurs de proportionnalité dans la vraisemblance.
- L'extension multidimensionnelle peut parfois conduire à des incohérences (paradoxe de marginalisation de Stein)

5.5 Principe d'invariance par transformation

Une première version du principe d'invariance est de considérer qu'une procédure statistique ne devrait pas dépendre de l'unité de mesure retenue. Si x et θ sont mesurés dans une unité u_1 et si y et ϕ sont les transformations de x et θ dans une nouvelle unité u_2 , l'estimateur $\delta_2(y)$ de ϕ devrait correspondre à l'estimateur $\delta_1(x)$ de θ par le même changement de mesure. La notion d'unité de mesure est à envisager au sens large : par exemple, elle peut aussi refléter le choix de l'origine ou même l'ordre suivant lequel sont enregistrées les composantes de x .

A COMPLETER

Soient $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ deux vecteurs aléatoires réels avec les densité de probabilité $p_x(\mathbf{x})$ et $p_y(\mathbf{y})$, et $\mathbf{b} \in \mathbb{R}^n$ un vecteur constant dans \mathbb{R}^n . Supposons que $p_x(\mathbf{x}) \in \mathcal{P}$ où \mathcal{P} est une famille paramétrique de lois. La question que l'on se pose est de déterminer la plus petite famille \mathcal{P} telle que pour toute transformations continues $\mathbf{y} = T(\mathbf{x})$ la densité de probabilité $p_y(\mathbf{y}) \in \mathcal{P}$.

1. **Rotation :** $\mathbf{y} = \mathbf{A}\mathbf{x}$ avec \mathbf{A} une matrice orthogonal ($\mathbf{A}^{-1} = \mathbf{A}^t, |\mathbf{A}| = 1, \mathbf{A}^t \mathbf{A} = \mathbf{I}$).
Solution : $\mathcal{P} = \{\mathbf{N}(\mathbf{x}|\mathbf{0}, \sigma^2 \mathbf{I}), \mathbf{I}$ la matrice identité $\}$.
2. **Rotation + Translation :** $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{b}$ avec \mathbf{A} une matrice orthonormale ($\mathbf{A}^{-1} = \mathbf{A}^t, |\mathbf{A}| = 1, \mathbf{A}^t \mathbf{A} = \mathbf{I}$).
Solution : $\mathcal{P} = \{\mathbf{N}(\mathbf{x}|\mathbf{m}, \sigma^2 \mathbf{I}), \mathbf{m} \in \mathbb{R}^n, \sigma^2 \mathbf{I}$ la matrice identité $\}$.
3. **Changement de base :** $\mathbf{y} = \mathbf{A}\mathbf{x}$ avec \mathbf{A} une matrice inversible.
Solution : $\mathcal{P} = \{\mathbf{N}(\mathbf{x}|\mathbf{0}, \Sigma), \Sigma$ une matrice définie positive $\}$.
4. **Rotation + Changement d'échelle + Translation :** $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{b}$ avec \mathbf{A} une matrice orthogonale ($\mathbf{A}^{-1} = \mathbf{A}^t$).
Solution : $\mathcal{P} = \{\mathbf{N}(\mathbf{x}|\mathbf{m}, \mathbf{D}), \mathbf{m} \in \mathbb{R}^n, \mathbf{D}$ une matrice diagonale définie positive $\}$.
5. **Rotation + Changement d'échelle :** $\mathbf{y} = \mathbf{A}\mathbf{x}$ avec \mathbf{A} une matrice orthogonale ($\mathbf{A}^{-1} = \mathbf{A}^t$).
Solution : $\mathcal{P} = \{\mathbf{N}(\mathbf{x}|\mathbf{0}, \mathbf{D}), \mathbf{D}$ une matrice diagonale définie positive $\}$.
6. **Changement d'échelle + Translation :** $\mathbf{y} = k\mathbf{x} + \mathbf{b}$ avec $k > 0$.
Solution : $\mathcal{P} = \{\text{famille exponentielle généralisée}\}$.
7. **Changement d'échelle :** $\mathbf{y} = k\mathbf{x}$ avec $k > 0$.
Solution : $\mathcal{P} = \{\text{famille exponentielle généralisée}\}$.
8. **Changement de base + Translation :** $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{b}$ avec \mathbf{A} une matrice inversible et $\mathbf{b} \in \mathbb{R}^n$.
Solution : $\mathcal{P} = \{\mathbf{N}(\mathbf{x}|\mathbf{m}, \Sigma), \mathbf{m} \in \mathbb{R}^n, \Sigma$ une matrice définie positive $\}$.

5.6 Famille des lois conjuguées

Pour une fonction de vraisemblance $l(\boldsymbol{\theta}) = p(\mathbf{x}|\boldsymbol{\theta})$ donnée, existe-il une famille de loi *a priori* $p(\boldsymbol{\theta})$ telle que la loi *a posteriori* appartient à la même famille que la loi *a priori*?

La réponse se trouve dans la définition des lois conjuguées.

Définition 14 [Famille des lois conjuguées] Une famille \mathcal{F} de lois *a priori* sur Θ est dite *conjuguée* (ou fermée par échantillonnage) si, pour tout $p(\boldsymbol{\theta}) \in \mathcal{F}$, la loi *a posteriori* $p(\boldsymbol{\theta}|\mathbf{x})$, appartient également à \mathcal{F} .

Raiffa et Schlaifer (1961), qui sont à l'origine des lois conjuguées, justifient leur approche par un principe d'invariance de la forme de la loi. Lorsque les observations \mathbf{x} modifient la loi *a priori* $p(\boldsymbol{\theta})$ en la loi *a posteriori* $p(\boldsymbol{\theta}|\mathbf{x})$, l'information apportée par \mathbf{x} sur $\boldsymbol{\theta}$ est limitée, donc elle ne doit pas conduire à une remise en cause de la forme de la loi $p(\boldsymbol{\theta})$, mais elle peut seulement changer (réactualiser) ses paramètres en fonction de \mathbf{x} . En d'autres termes, la modification apportée par les observations doit rester de dimension finie.

A COMPLETER

5.7 Lois conjuguées et Statistiques suffisantes

Définition 15 [Famille des lois a priori conjuguées] Supposons que $l(\boldsymbol{\theta}|\mathbf{x}) = l(\boldsymbol{\theta}|\mathbf{t})$ où $\mathbf{t} = \{n, \mathbf{s}\} = \{n, s_0, \dots, s_k\}$, un vecteur de dimension $k + 1$, est une statistique suffisante pour $p(\mathbf{x}|\boldsymbol{\theta})$.

Alors s'il existe un vecteur $\{\tau_0, \boldsymbol{\tau}\} = \{\tau_0, \tau_1, \dots, \tau_k\}$ tel que

$$p(\boldsymbol{\theta}|\boldsymbol{\tau}) = \frac{p(\mathbf{s} = (\tau_1, \dots, \tau_k)|\boldsymbol{\theta}, n = \tau_0)}{\int p(\mathbf{s} = (\tau_1, \dots, \tau_k)|\boldsymbol{\theta}, n = \tau_0) d\boldsymbol{\theta}}$$

soit bien défini, alors $p(\boldsymbol{\theta}|\mathbf{x}, \boldsymbol{\tau})$ reste dans la même famille que $p(\boldsymbol{\theta}|\boldsymbol{\tau})$. On dit alors que $p(\boldsymbol{\theta}|\boldsymbol{\tau})$ est une loi *a priori* conjuguée pour la loi des observations $p(\mathbf{x}|\boldsymbol{\theta})$ ou, d'une manière équivalente, pour la fonction de vraisemblance $l(\boldsymbol{\theta}|\mathbf{x})$

Exemple: famille exponentielle à K paramètres

5.8 Lois conjuguées des familles exponentielles

Définition 16 [Famille exponentielle] Soient, \mathcal{X} l'espace des observations, Θ l'espace des paramètres, $p(\mathbf{x})$ une fonction de \mathcal{X} dans \mathbf{R} , $g(\boldsymbol{\theta})$ une fonction de Θ dans \mathbf{R}^+ , $\phi_k(\boldsymbol{\theta})$ des fonctions de Θ dans \mathbf{R} , et $h_k(\mathbf{x})$ des fonctions de \mathcal{X} dans \mathbf{R} . Alors la famille de distributions de densité

$$\begin{aligned} p(\mathbf{x}|\boldsymbol{\theta}) &= f(\mathbf{x})g(\boldsymbol{\theta}) \exp \left[\sum_{k=1}^K \phi_k(\boldsymbol{\theta})h_k(\mathbf{x}) \right] \\ &= f(\mathbf{x})g(\boldsymbol{\theta}) \exp \left[\boldsymbol{\phi}^t(\boldsymbol{\theta})\mathbf{h}(\mathbf{x}) \right] \end{aligned}$$

est dite *famille exponentielle généralisée*. Elle est définie entièrement par $f(\mathbf{x})$, $g(\boldsymbol{\theta})$, et $\{\phi_k(\boldsymbol{\theta}), h_k(\mathbf{x})\}$, $k = 1, \dots, K$.

Notons aussi les cas particuliers suivants :

- Dans le cas d'une variable on a

$$\begin{aligned} p(x|\boldsymbol{\theta}) &= \mathbf{E}x\mathbf{f}(x|f, g, \boldsymbol{\phi}, \mathbf{h}) \\ &= f(x)g(\boldsymbol{\theta}) \exp \left[\sum_{k=1}^K \phi_k(\boldsymbol{\theta}) h_k(x) \right] \\ &= f(x)g(\boldsymbol{\theta}) \exp \left[\boldsymbol{\phi}^t(\boldsymbol{\theta}) \mathbf{h}(x) \right] \end{aligned}$$

- Si $f(\mathbf{x}) = 1$ et si $g(\boldsymbol{\theta}) = \exp[-b(\boldsymbol{\theta})]$ on obtient

$$p(\mathbf{x}|\boldsymbol{\theta}) = \exp \left[\boldsymbol{\phi}^t(\boldsymbol{\theta}) \mathbf{h}(\mathbf{x}) - b(\boldsymbol{\theta}) \right].$$

- Si $f(\mathbf{x}) = 1$ et si $g(\boldsymbol{\theta}) = \exp[-b(\boldsymbol{\theta})]$, $\mathbf{h}(\mathbf{x}) = \mathbf{x}$ et $\boldsymbol{\phi}(\boldsymbol{\theta}) = \boldsymbol{\theta}$ on obtient la famille exponentielle naturelle

$$p(\mathbf{x}|\boldsymbol{\theta}) = \exp \left[\boldsymbol{\theta}^t \mathbf{x} - b(\boldsymbol{\theta}) \right].$$

Proposition 18 [Statistique suffisante pour la famille exponentielle] Pour un n -échantillon $\{x_1, \dots, x_n\}$ d'une variable $X \sim \mathbf{E}x\mathbf{f}(x|f, g, \boldsymbol{\phi}, \mathbf{h})$ on a

$$\begin{aligned} p(\mathbf{x}|\boldsymbol{\theta}) = \prod_{j=1}^n p(x_j|\boldsymbol{\theta}) &= [g(\boldsymbol{\theta})]^n \left(\prod_{j=1}^n f(x_j) \right) \exp \left[\sum_{k=1}^K \phi_k(\boldsymbol{\theta}) \sum_{j=1}^n h_k(x_j) \right] \\ &= g^n(\boldsymbol{\theta}) f(\mathbf{x}) \exp \left[\boldsymbol{\phi}^t(\boldsymbol{\theta}) \mathbf{h}(\mathbf{x}) \right], \end{aligned}$$

où $f(\mathbf{x}) = \prod_{j=1}^n f(x_j)$. Ceci permet d'identifier

$$\mathbf{t} = \left\{ n, \sum_{j=1}^n h_1(x_j), \dots, \sum_{j=1}^n h_K(x_j) \right\}$$

comme statistique suffisante.

Proposition 19 [Lois conjuguées pour la famille exponentielle] Une famille conjuguée pour la famille exponentielle est donnée par

$$p(\boldsymbol{\theta}|\boldsymbol{\tau}) = z(\boldsymbol{\tau}) [g(\boldsymbol{\theta})]^{\tau_0} \exp \left[\sum_{k=1}^K \tau_k \phi_k(\boldsymbol{\theta}) \right]$$

La loi *a posteriori* associée est

$$p(\boldsymbol{\theta}|\mathbf{x}, \boldsymbol{\tau}) = [g(\boldsymbol{\theta})]^{n+\tau_0} \frac{\prod_{j=1}^n f(x_j)}{Z(\boldsymbol{\tau})} \exp \left[\sum_{k=1}^K c_k \phi_k(\boldsymbol{\theta}) \left(\tau_k + \sum_{j=1}^n h_k(x_j) \right) \right].$$

En notations symbolique on peut écrire :

Si

$$p(\mathbf{x}|\boldsymbol{\theta}) = \mathbf{E}x\mathbf{f}(\mathbf{x}|f, g, \boldsymbol{\phi}, \mathbf{h}),$$

alors la loi *a priori* conjuguée est

$$p(\boldsymbol{\theta}|\boldsymbol{\tau}) = \mathbf{E}x\mathbf{f}(\boldsymbol{\theta}|z, g^{\tau_0}, \boldsymbol{\phi}, \boldsymbol{\tau}),$$

et la loi *a posteriori* associée est

$$p(\boldsymbol{\theta}|\mathbf{x}, \boldsymbol{\tau}) = \mathbf{E} \mathbf{x} \mathbf{f}(\boldsymbol{\theta}|z f, g^{n+\tau_0}, \boldsymbol{\phi}, \boldsymbol{\tau}'),$$

où

$$\tau'_k = \tau_k + \sum_{j=1}^n h_k(x_j)$$

ou encore

$$\boldsymbol{\tau}' = \boldsymbol{\tau} + \bar{\mathbf{h}}, \quad \text{avec} \quad \bar{h}_k = \sum_{j=1}^n h_k(x_j).$$

Définition 17 [Famille exponentielle naturelle (forme canonique)] La famille de distributions de densité

$$p(\mathbf{x}|\boldsymbol{\theta}) = a(\mathbf{x}) \exp[\boldsymbol{\theta}^t \mathbf{x} - b(\boldsymbol{\theta})]$$

est dite *famille exponentielle naturelle*.

Proposition 20 [Lois conjuguées pour la famille exponentielle naturelle] Une famille conjuguée pour la famille exponentielle naturelle

$$p(\mathbf{x}|\boldsymbol{\theta}) = a(\mathbf{x}) \exp[\boldsymbol{\theta}^t \mathbf{x} - b(\boldsymbol{\theta})]$$

est donnée par

$$p(\boldsymbol{\theta}|\boldsymbol{\tau}_0) = c(\boldsymbol{\theta}) \exp[\boldsymbol{\tau}_0^t \boldsymbol{\theta} - d(\boldsymbol{\tau}_0)]$$

La loi *a posteriori* associée est

$$p(\boldsymbol{\theta}|\mathbf{x}, \boldsymbol{\tau}_0) = c(\boldsymbol{\theta}) \exp[\boldsymbol{\tau}_n^t \boldsymbol{\theta} - d(\boldsymbol{\tau}_n)] \quad \text{avec} \quad \boldsymbol{\tau}_n = \boldsymbol{\tau}_0 + \bar{\mathbf{x}}$$

où

$$\bar{\mathbf{x}}_n = \frac{1}{n} \sum_{j=1}^n \mathbf{x}_j,$$

et la loi prédictive associée est

$$p(\mathbf{y}|\mathbf{x}, \boldsymbol{\tau}_0) = p(\mathbf{y}|\mathbf{t}_n(\mathbf{x}), \boldsymbol{\tau}_0) = c(\boldsymbol{\theta}) \exp[\boldsymbol{\tau}_n^t \boldsymbol{\theta} - d(\boldsymbol{\tau}_n)] \quad \text{avec} \quad \boldsymbol{\tau}_n = \boldsymbol{\tau}_0 + \bar{\mathbf{x}}$$

A COMPLETER

$$p(\boldsymbol{\theta}|n_0, \boldsymbol{\tau}_0) = c(n_0, \boldsymbol{\tau}_0) \exp[n_0 \boldsymbol{\phi}_0^t \boldsymbol{\tau}_0 - n_0 b(\boldsymbol{\phi})]$$

$$p(\boldsymbol{\theta}|n_0, \boldsymbol{\tau}_0) \mapsto p(\boldsymbol{\theta}|n_0, \boldsymbol{\tau}_0, \mathbf{x}) = p(\boldsymbol{\theta}|n_0 + n, \frac{n_0 \boldsymbol{\tau}_0 + n \bar{\mathbf{y}}_n}{n_0 + n})$$

$$\mathbf{E}[\mathbf{y}|\boldsymbol{\phi}] = \mathbf{E}[\bar{\mathbf{y}}_n|\boldsymbol{\phi}] = \nabla b(\boldsymbol{\phi})$$

$$\mathbf{E}[\nabla b(\boldsymbol{\phi})|n_0, \boldsymbol{\tau}_0] = \boldsymbol{\tau}_0$$

$$\mathbf{E}[\nabla b(\boldsymbol{\phi})|n_0, \boldsymbol{\tau}_0, \mathbf{x}] = \frac{n \bar{\mathbf{y}}_n + n_0 \boldsymbol{\tau}_0}{n_0 + n} = \pi \bar{\mathbf{y}}_n + (1 - \pi) \boldsymbol{\tau}_0, \quad \text{avec} \quad \pi = \frac{n}{n_0 + n}$$

Proposition 21 [Mélange des lois conjuguées] Soit \mathcal{F} une famille conjuguée naturelle d'une famille exponentielle naturelle. Alors

$$\tilde{\mathcal{F}} = \left\{ \sum_{i=1}^M w_i p(\boldsymbol{\theta} | n_0, \boldsymbol{\tau}_0); \sum_{i=1}^M w_i = 1 \right\}$$

famille des mélanges des lois conjuguées naturelles, est aussi une famille conjuguées. De plus, la loi *a posteriori* associée est aussi un mélange,

$$p(\boldsymbol{\theta} | n_n, \boldsymbol{\tau}_n, \mathbf{x}) = \sum_{i=1}^M w'_i p(\boldsymbol{\theta} | n_n, \boldsymbol{\tau}_n, \mathbf{x})$$

avec

$$w'_i =$$

A COMPLETER

Exemple 15 Soit $x \sim \mathbf{N}(x | \theta, 1)$ et

$$p(\theta) = \sum_{i=1}^M w_i \pi_i$$

où $\pi_i = \mathbf{N}(\theta | \mu_i, \lambda_i)$. La loi *a posteriori* associée est

$$p(\boldsymbol{\theta} | n_n, \boldsymbol{\tau}_n, \mathbf{x}) = \sum_{i=1}^M w'_i p(\boldsymbol{\theta} | n_n, \boldsymbol{\tau}_n, \mathbf{x})$$

avec

$$w'_i =$$

A COMPLETER

Relation entre la loi des observations, la loi *a priori* et la loi *a posteriori*

loi des observation $p(x \theta)$	loi <i>a priori</i> $p(\theta \tau)$	loi <i>a posteriori</i> $p(\theta x, \tau) \propto p(\theta \tau)p(x \theta)$
Variables discrètes		
Binomiale Bin ($x n, \theta$)	Bêta Bet ($\theta \alpha, \beta$)	Bêta Bet ($\theta \alpha + x, \beta + n - x$)
Binomiale Bin ($x n, \theta$)	$p(\theta) = \theta(1 - \theta)^{-1}$	Bêta Bet ($\theta x, n - x$), $x \neq 0, x \neq n$
Binomiale négative NegBin ($x n, \theta$)	Bêta Bet ($\theta \alpha, \beta$)	Bêta Bet ($\theta \alpha + n, \beta + x$)
Multinomiale M_k ($x \theta_1, \dots, \theta_k$)	Dirichlet Di_k ($\theta \alpha_1, \dots, \alpha_k$)	Dirichlet Di_k ($\theta \alpha_1 + x_1, \dots, \alpha_k + x_k$)
Poisson Pn ($x \theta$)	Gamma Gam ($\theta \alpha, \beta$)	Gamma Gam ($\theta \alpha + x, \beta + 1$)
Gamma Gam ($x \nu, \theta$)	Gamma Gam ($\theta \alpha, \beta$)	Gamma Gam ($\theta \alpha + \nu, \beta + x$)
Bêta Bet ($x \alpha, \theta$)	Exponentielle Ex ($\theta \lambda$)	Exponentielle Ex ($\theta \lambda - \log(1 - x)$)
Normale N ($x \theta, \sigma^2$)	Normale N ($\theta \mu, \tau^2$)	Normale N ($\mu \frac{\mu\sigma^2 + \tau^2x}{\sigma^2 + \tau^2}, \frac{\sigma^2\tau^2}{\sigma^2 + \tau^2}$)
Variables continues		
Normale N ($x \mu, 1/\theta$)	Gamma Gam ($\theta \alpha, \beta$)	Gamma Gam ($\theta \alpha + \frac{1}{2}, \beta + \frac{1}{2}(\mu - x)^2$)
Normale N ($x \theta, \theta^2$)	Normale inverse généralisée INg ($\theta \alpha, \mu, \sigma$) \propto $ \theta ^{-\alpha} \exp\left[-\frac{1}{2\sigma^2}\left(\frac{1}{\theta} - \mu\right)^2\right]$	Normale inverse généralisée INg ($\theta \alpha_n, \mu_n, \sigma_n$)

TAB. 5.1 – Dans ce tableau la loi des observations et la loi *a posteriori* sont données pour un échantillon.

5.9 Lois a priori de référence

Relation entre vraisemblance, *a priori* et *a posteriori*

Vraisemblance $p(x \theta)$	lois conjuguée $p(\theta \tau)$	lois a posteriori $p(\theta x, \tau) \propto p(\theta \tau)p(x \theta)$	estimateur $\delta(x)$
Variables discrètes			
Bernouilli Ber ($x \theta$)	Bêta Bet ($\theta \alpha, \beta$) $\alpha = \tau_1 + 1,$ $\beta = \tau_0 - \tau_1 + 1$	Bêta Bet ($\theta \alpha_n, \beta_n$) $\alpha_n = \alpha + r_n,$ $\beta_n = \beta + r_n,$ $r_n = \sum_{j=1}^n x_j$	$\frac{\alpha+x}{\alpha+\beta+1}$
Binomiale Bin ($x n, \theta$)	Bêta Bet ($\theta \alpha, \beta$) $\alpha = \tau_1 + 1,$ $\beta = \tau_0 - \tau_1 + 1$	Bêta Bet ($\theta \alpha_n, \beta_n$) $\alpha_n = \alpha + r_n,$ $\beta_n = \beta + r_n,$ $r_n = \sum_{j=1}^n x_j$	$\frac{\alpha+x}{\alpha+\beta+n}$
Binomiale négative NegBin ($x n, \theta$)	Bêta Bet ($\theta \alpha, \beta$) $\alpha = \tau_1 + 1,$ $\beta = \tau_0 - \tau_1 + 1$	Bêta Bet ($\theta \alpha_n, \beta_n$) $\alpha_n = \alpha + r_n,$ $\beta_n = \beta + r_n,$ $r_n = \sum_{j=1}^n x_j$	$\frac{\alpha+n}{\alpha+\beta+n+x}$
Poisson Pn ($x \theta$)	Gamma Gam ($\theta \alpha, \beta$) $\alpha = \tau_1 + 1,$ $\beta = \tau_0$	Gamma Gam ($\theta \alpha_n, \beta_n$) $\alpha_n = \alpha + r_n,$ $\beta_n = \beta + r_n,$ $r_n = \sum_{j=1}^n x_j$	$\frac{\alpha+x}{\beta+1}$
Variables continues			
Gamma Gam ($x \nu, \theta$)	Gamma Gam ($\theta \alpha, \beta$) $\alpha = \tau_1 + 1,$ $\beta = \tau_0$	Gamma Gam ($\theta \alpha_n, \beta_n$) $\alpha_n = \alpha + r_n,$ $\beta_n = \beta + r_n,$ $r_n = \sum_{j=1}^n x_j$	$\frac{\alpha+\nu}{\beta+x}$
Bêta Bet ($x \alpha, \theta$)	Exponentielle Ex ($\theta \lambda$)	Exponentielle Ex ($\theta \lambda_n$) $\lambda_n = \frac{1}{\lambda - \log(1-x)}$	$\frac{1}{\lambda - \log(1-x)}$
Normale N ($x \theta, \sigma^2$)	Normale N ($\theta \mu, \tau^2$) $\mu \in \mathbb{R}, \tau > 0$	Normale N ($\mu \mu_n, \tau_n^2$) $\mu_n = \frac{\mu\sigma^2 + \tau^2 x}{\sigma^2 + \tau^2}$ $\tau_n^2 = ??,$	$\frac{\mu\sigma^2 + \tau^2 x}{\sigma^2 + \tau^2}$
Normale N ($x \mu, 1/\theta$)	Gamma Gam ($\theta \alpha, \beta$) $\alpha > 0, \beta > 0$	Gamma Gam ($\theta \alpha_n, \beta_n$) $\alpha_n =,$ $\beta_n =$	$\frac{\alpha + \frac{1}{2}}{\beta + \frac{1}{2}(\mu - x)^2}$
Normale N ($x \mu, \lambda$)	Normale-Gamma NGam ($\mu, \lambda \alpha, \beta, \gamma$) $= \mathbf{N}(\mu \gamma, \lambda(2\alpha - 1))$ $\times \mathbf{Gam}(\lambda \alpha, \beta)$ $\alpha > \frac{1}{2}, \beta > 0, \gamma \in \mathbb{R}$	Normale-Gamma NGam ($\mu, \lambda \alpha_n, \beta_n, \gamma_n$) $= \mathbf{N}(\mu \gamma_n, \lambda(2\alpha_n - 1))$ $\times \mathbf{Gam}(\lambda \alpha_n, \beta_n)$ $\alpha_n = \alpha + r_n,$ $\beta_n = \beta + r_n,$ $r_n = \sum_{j=1}^n x_j$??

TAB. 5.2 – Relation entre vraisemblance, *a priori* et *a posteriori*

Relation entre vraisemblance, *a priori* et *a posteriori*
pour des lois *a priori* dite de référence

Vraisemblance $p(x \theta)$	Reference prior $\pi(\theta)$	Posterior $\pi(\theta x) \propto p(x \theta)\pi(\theta)$
Bernouilli Ber ($x \theta$)	Bêta Bet ($\theta 1/2, 1/2$)	Bêta Bet ($\theta \alpha_n, \beta_n$) $\alpha_n = 1/2 + r_n$, $\beta = n - r_n + 1/2$, $r_n = \sum_{j=1}^n x_j$
Binomiale négative NegBin ($x r, \theta$)	Bêta Bet ($\theta 1, 1/2$)	Bêta Bet ($\theta \alpha_n, \beta_n$) $\alpha_n = r_n$, $\beta_n = x - r_n + 1/2$, $r_n = \sum_{j=1}^n x_j$

TAB. 5.3 – Relation entre vraisemblance, *a priori* et *a posteriori* pour des lois *a priori* dite de référence