

# Statistique bayésienne et algorithmes MCMC

IMAT (Master 1)

**Jerôme DUPUIS** (LSP-UPS)

(Version Septembre 2007)

# Table des matières

<b>1</b>	<b>Motivations</b>	<b>3</b>
<b>2</b>	<b>Le modèle statistique paramétrique bayésien</b>	<b>3</b>
2.1	Notations, définitions et commentaires. . . . .	3
2.2	Le calcul de la loi a postériori. . . . .	5
2.2.1	une situation simple . . . . .	5
2.2.2	Le raisonnement proportionnel . . . . .	5
<b>3</b>	<b>L'estimation bayésienne</b>	<b>6</b>
3.1	Définition de l'estimateur de Bayes . . . . .	6
3.1.1	Le cas uni-dimensionnel . . . . .	6
3.1.2	Le cas multi-dimensionnel . . . . .	7
3.2	Le risque de Bayes . . . . .	7
3.3	Propriétés de l'estimateur de Bayes . . . . .	8
3.4	Loi a priori impropre et estimateur de Bayes généralisé . . . . .	8
3.5	Les intervalles de confiance bayésiens . . . . .	9
<b>4</b>	<b>L'approche bayésienne des tests</b>	<b>9</b>
<b>5</b>	<b>Modélisation de l'information a priori</b>	<b>10</b>
5.1	Lois a priori conjuguées . . . . .	10
5.2	Lois a priori non informatives . . . . .	10
5.3	Incorporation de l'information a priori en pratique. . . . .	11
5.4	Le poids de l'a priori dans la réponse bayésienne . . . . .	11
5.5	Étude de la sensibilité de la réponse bayésienne à la loi priori . . . . .	12
<b>6</b>	<b>Méthodes de simulation de Monte Carlo</b>	<b>12</b>
6.1	Méthode classique de simulation de Monte Carlo . . . . .	12
6.1.1	Motivations . . . . .	12
6.1.2	Le principe de la méthode . . . . .	13
6.2	Les méthodes de simulation de Monte Carlo par chaîne de Markov . . . . .	13
6.2.1	Motivations . . . . .	13
6.2.2	Le principe de la méthode . . . . .	14
6.2.3	L'algorithme de Gibbs . . . . .	15
6.2.4	L'algorithme de Hastings-Metropolis . . . . .	15
<b>7</b>	<b>Mise en oeuvre des méthodes de simulation de Monte Carlo</b>	<b>15</b>
7.1	Le modèle 1: étude bayésienne de l'efficacité d'un traitement . . . . .	15
7.2	Le modèle 2: contamination de l'environnement par un maïs transgénique . . . . .	15
7.3	Le modèle 3: estimation bayésienne de la taille d'une population animale. . . . .	16
<b>8</b>	<b>Une procédure bayésienne de choix de modèles</b>	<b>16</b>

# 1 Motivations

On note  $\theta$  la probabilité qu'un individu atteint d'une maladie M soit guéri à l'issu d'un traitement  $T$  (consistant en l'administration d'un certain médicament). Pour estimer le paramètre  $\theta$  inconnu, on administre à  $n$  patients le traitement  $T$ . On note  $x_i \in \{0,1\}$  l'état du patient  $i$  suite à l'administration du traitement  $T$ ;  $x_i = 1$  si le patient  $i$  est guéri et  $x_i = 0$  sinon. On considère par ailleurs que le médicament (associé au traitement  $T$ ) ne peut-être commercialisé que si  $\theta \geq 0.8$ . D'un point de vue statistique on peut formuler le problème à l'aide du test suivant:  $H_0 : \theta < 0.8$  contre  $H_1 : \theta \geq 0.8$ .

D'un point de vue statistique on se place dans le cadre du modèle d'échantillonnage. Autrement dit on suppose que les  $x_i$  sont des réalisations de variables aléatoires  $X_i$  indépendantes et de même loi (on dit que les  $X_i$  sont *i.i.d.*). Chaque  $X_i$  suit une loi de Bernoulli de paramètre  $\theta$ . Autrement dit on a:  $P(X_i = 1) = \theta$ , et du fait que les variables aléatoires sont *i.i.d.*, la vraisemblance s'écrit:

$$L(\theta; x_1, \dots, x_i, \dots, x_n) = \theta^s (1 - \theta)^{n-s} \quad \text{où} \quad s = \sum_{i=1}^n x_i$$

où  $s = \sum_{i=1}^n x_i$  représente le nombre de patients guéris. Par ailleurs, on suppose que l'on dispose d'information sur le paramètre  $\theta$  à estimer (et ce, en plus de l'information déjà apportée par les  $x_i$ ). Cette information (qu'on appellera information a priori) peut, par exemple, provenir d'une autre étude portant sur l'efficacité du traitement  $T$  et réalisé sur des patients issus, au sens statistique, de la même population (autrement dit la probabilité de guérison de tels patients est égale à  $\theta$ ). Une telle information consiste typiquement alors en une estimation  $\hat{\theta}$  de  $\theta$  et un intervalle de confiance (observé)  $I$  de niveau de confiance 0.95. Cette information peut être également fournie par un Expert; on entend par Expert, toute personne compétente sur le sujet et reconnu comme tel par ses pairs. Cette information consiste à nouveau en une estimation de  $\theta$ , notée  $\theta^*$ , et en un intervalle  $J$  contenant  $\theta^*$ , et tel que la probabilité  $\Pr(\theta \in J)$  soit élevée (0.90 ou 0.95 en pratique). A noter qu'il conviendra de donner un sens à cette dernière écriture (celle-ci n'ayant pas de sens en statistique classique).

Le problème statistique consiste alors à produire une inférence sur  $\theta$  (estimation ou test) qui repose à la fois sur les observations (c.a.d sur les  $x_i$ ) et sur l'information a priori dont on dispose. L'objet de l'approche statistique bayésienne est précisément de proposer une procédure qui réponde au problème. Rappelons que la statistique classique, quant à elle, fait reposer l'estimation de  $\theta$  uniquement sur les observations (c.a.d sur les  $x_i$ ); au cas particulier l'estimation retenue serait la moyenne des  $x_i$  (qui correspond d'ailleurs à l'estimation de  $\theta$  par maximum de vraisemblance).

## 2 Le modèle statistique paramétrique bayésien

### 2.1 Notations, définitions et commentaires.

L'ensemble des observations est noté  $\mathbf{x}$ . Dans ce cours est  $\mathbf{x} = (x_1, \dots, x_i, \dots, x_n)$ ; autrement dit, on dispose d'un échantillon de taille  $n$ . Le cadre statistique de ce cours étant celui de la statistique inférentielle, les observations  $x_i$  sont donc considérées comme des réalisations de variables aléatoires, notées  $X_i$ .

#### Définitions

- On entend par information a priori sur le paramètre  $\theta$  toute information disponible sur  $\theta$  en dehors de celle apportée par les observations.
- L'information a priori sur  $\theta$  est entachée d'incertitude (si ce n'était pas le cas, le paramètre  $\theta$  serait connu avec certitude et on n'aurait pas à l'estimer!). Il est naturel de modéliser cette information a priori au travers d'une loi de probabilité, appelée loi a priori. Sa densité est notée  $\pi(\theta)$ .
- Le modèle statistique paramétrique bayésien consiste en la donnée d'une loi a priori et de la loi des observations. On appelle loi des observations, la loi conditionnelle de  $\mathbf{X}$  sachant  $\theta$ . Sa densité est notée  $f(\mathbf{x}|\theta)$ ,

que la variable aléatoire  $\mathbf{X}$  soit discrète ou continue. Si  $X$  est discrète,  $f(\mathbf{x}|\theta)$  représente  $\Pr(X = \mathbf{x}|\theta)$ . On fera dans ce cours systématiquement l'hypothèse que, sachant  $\theta$ , les v.a.  $X_i$  sont indépendantes. Autrement dit on aura:

$$f(\mathbf{x}|\theta) = \prod_{i=1}^n f(x_i|\theta).$$

Indiquons maintenant les autres lois de probabilité qui interviennent en statistique bayésienne.

La loi a posteriori. C'est la loi conditionnelle de  $\theta$  sachant  $\mathbf{x}$ . Sa densité est notée  $\pi(\theta|\mathbf{x})$ . En vertu de la formule de Bayes, on a:

$$\pi(\theta|\mathbf{x}) = \frac{f(\mathbf{x}|\theta)\pi(\theta)}{\int_{\Theta} f(\mathbf{x}|\theta)\pi(\theta)d\theta}$$

La loi du couple  $(\theta, \mathbf{X})$ . Sa densité est notée  $h(\theta, \mathbf{x})$ . On a donc:  $h(\theta, \mathbf{x}) = f(\mathbf{x}|\theta)\pi(\theta)$ .

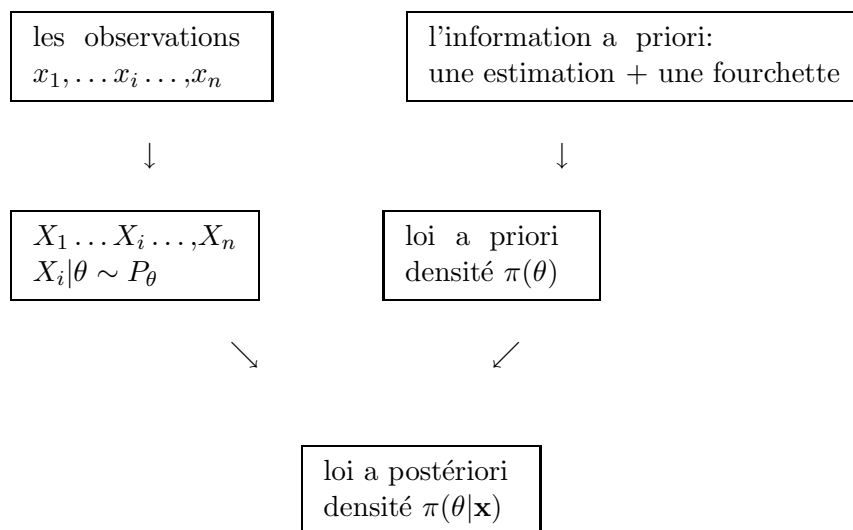
La loi marginale de  $\mathbf{X}$ . Sa densité est notée  $m(\mathbf{x})$ . On a donc:  $m(\mathbf{x}) = \int_{\Theta} f(\mathbf{x}|\theta)\pi(\theta)d\theta$ .

### Commentaires:

La philosophie de l'approche bayésienne.

Alors que la statistique classique repose sur la loi des observations, la statistique bayésienne repose sur la loi a posteriori. La loi a posteriori peut s'interpréter comme un résumé (en un sens probabiliste) de l'information disponible sur  $\theta$ , une fois  $\mathbf{x}$  observé. L'approche bayésienne réalise en quelque sorte l'actualisation de l'information a priori par l'observation  $\mathbf{x}$ , au travers de  $\pi(\theta|\mathbf{x})$ .

Le schéma ci-dessous résume la démarche bayésienne dans le cadre de la statistique paramétrique inférentielle. Il fait également apparaître, la modélisation stochastique des  $x_i$  comme étant des réalisations de variables aléatoires  $X_i$  (cette modélisation est caractéristique de la statistique inférentielle), ainsi que la modélisation stochastique de l'information a priori disponible sur le paramètre  $\theta$ , au travers de la loi a priori.



### Un abus de notation

Au vu de ce qui précède, il apparaît que la notation  $\theta$  désigne tantôt une variable aléatoire, tantôt un paramètre. Cet abus de notation qui déroute souvent le néophyte est fréquente (sinon systématique) dans les ouvrages de statistique bayésienne. On pourrait évidemment, distinguer la variable aléatoire et le paramètre par deux notations distinctes, en notant, par exemple,  $\theta_{v.a.}$  la variable aléatoire, et en réservant la notation  $\theta$  au paramètre, comme dans le schéma ci-dessous.

grandeurs aléatoires	grandeurs non aléatoires
$\theta_{v.a.}$	$\theta$ : le paramètre à estimer
$X_i   \theta \sim P_\theta$	$(x_1, \dots, x_i, \dots, x_n)$ : les observations

En pratique, on ne le fait pas; c'est le contexte qui permet de lever l'ambiguïté. Donnons quelques exemples. Quand on écrit  $\theta \sim \text{Beta}(3,7)$  ou  $\Pr(\theta \in [0.1, 0.3]) = 0.5$ , il est clair qu'on parle de la variable aléatoire. Quand on écrit  $\theta = 0.3$  il est clair qu'on parle du paramètre.

## 2.2 Le calcul de la loi a postérieure.

### 2.2.1 une situation simple

On dispose d'un vecteur d'observations:  $\mathbf{x} = (x_1, \dots, x_i, \dots, x_n)$ , et on considère le modèle bayésien suivant:  $X_i | \theta \sim \text{Bernoulli}(\theta)$  et  $\theta \sim \text{Beta}(a,b)$ . On a:

$$f(\mathbf{x} | \theta) = \prod_{i=1}^n P(X = x_i | \theta) = \theta^s (1 - \theta)^{n-s}$$

où  $s = \sum_{i=1}^n x_i$ . Comme  $\theta \sim \text{Beta}(a,b)$ , on a:

$$\pi(\theta) = \frac{1}{B(a,b)} \theta^{a-1} (1 - \theta)^{b-1} \mathbf{1}_{[0,1]}(\theta).$$

D'autre part on a:

$$\pi(\theta | \mathbf{x}) = \frac{f(\mathbf{x} | \theta) \pi(\theta)}{\int_{\Theta} f(\mathbf{x} | \theta) \pi(\theta) d\theta}.$$

Il est facile de vérifier que:

$$\int_{\Theta} f(\mathbf{x} | \theta) \pi(\theta) d\theta = B(\alpha, \beta).$$

où  $\alpha = a + s$  et  $\beta = b + n - s$ . D'où:

$$\pi(\theta | \mathbf{x}) = \frac{\theta^{\alpha-1} (1 - \theta)^{\beta-1}}{B(\alpha, \beta)} \mathbf{1}_{[0,1]}(\theta).$$

Par conséquent:

$$\theta | \mathbf{x} \sim \text{Beta}(a + \sum_{i=1}^n x_i, b + n - \sum_{i=1}^n x_i)$$

### 2.2.2 Le raisonnement proportionnel

Il est parfois possible d'éviter le calcul de l'intégrale:  $\int_{\Theta} f(\mathbf{x} | \theta) \pi(\theta) d\theta$  en raisonnant proportionnellement.

#### Une notation et une définition.

Soient deux fonctions réelles  $f$  et  $g$  définies sur le même espace  $\mathcal{Y}$ . On dit que  $f$  et  $g$  sont proportionnelles, ce qu'on note  $f \propto g$ , si il existe une constante  $a$  tel que  $f(y) = ag(y)$  pour tout  $y \in \mathcal{Y}$ . Il est clair que la relation  $\propto$  est une relation d'équivalence. En particulier:  $f \propto g$  et  $g \propto h$  entraînent  $f \propto h$ .

#### Deux remarques.

1. Soit  $f(y)$  la densité d'une variable aléatoire  $Y$  de loi inconnue. Si  $f \propto u_1 \dots \propto u_k \propto g$ , où  $u_1 \dots u_k$  désignent des fonctions réelles et  $g(y)$  est la densité d'une loi de probabilité  $P$ , alors  $Y \sim P$ .

2. Dans un contexte bayésien on a:  $\pi(\theta|\mathbf{x}) \propto f(\mathbf{x}|\theta)\pi(\theta)$ . En tant que fonctions de  $\theta$ , les deux expressions  $\pi(\theta|\mathbf{x})$  et  $f(\mathbf{x}|\theta)$  sont effectivement proportionnelles; la constante  $a$  qui apparaît dans la définition est égale ici à  $1/m(\mathbf{x})$ ; à noter que cette quantité est bien une constante, au sens où elle ne dépend pas de  $\theta$ . L'écriture  $\pi(\theta|\mathbf{x}) \propto f(\mathbf{x}|\theta)\pi(\theta)$  est souvent reformulée de la façon suivante:

$$\boxed{\pi(\theta|\mathbf{x}) \propto L(\theta; \mathbf{x})\pi(\theta)}$$

où  $L(\theta; \mathbf{x})$  désigne la vraisemblance. Rappelons que:  $L(\theta; \mathbf{x}) = f(\mathbf{x}|\theta)$  (par définition).

### Deux exemples.

1. Soit  $f(y)$  la densité d'une variable aléatoire réelle  $Y$  telle que:  $f(y) \propto \exp(-y^2/2)$  alors il est clair que, d'après la remarque No 2,  $Y$  suit une loi normale  $\mathcal{N}(0,1)$
2. Reprenons le modèle statistique bayésien précédent où  $X_i|\theta \sim \text{Bernoulli}(\theta)$  et  $\theta \sim \text{Beta}(a,b)$ . On a :

$$\pi(\theta|\mathbf{x}) \propto L(\theta; \mathbf{x})\pi(\theta).$$

Soit:

$$\pi(\theta|\mathbf{x}) \propto \theta^{\sum_{i=1}^n x_i} (1-\theta)^{n-\sum_{i=1}^n x_i} \theta^{a-1} (1-\theta)^{b-1} 1_{[0,1]}(\theta).$$

Par conséquent:

$$\pi(\theta|\mathbf{x}) \propto \theta^{a+\sum_{i=1}^n x_i-1} (1-\theta)^{b+n-\sum_{i=1}^n x_i-1} 1_{[0,1]}(\theta).$$

A une constante multiplicative près, on reconnaît dans le membre de gauche la densité d'une loi Beta de paramètres  $(a + \sum_{i=1}^n x_i$  et  $b + n - \sum_{i=1}^n x_i)$ . D'après la remarque No 2, on en déduit que:

$$\theta|\mathbf{x} \sim \text{Beta}\left(a + \sum_{i=1}^n x_i, b + n - \sum_{i=1}^n x_i\right).$$

Exercice. Soit le modèle statistique bayésien suivant:  $X_i|\theta \sim \text{Poisson}(\theta)$  et  $\theta \sim \text{Gamma}(a,b)$ . Déterminer la loi a posteriori.

## 3 L'estimation bayésienne

### 3.1 Définition de l'estimateur de Bayes

#### 3.1.1 Le cas uni-dimensionnel

On suppose dans cette section que le paramètre  $\theta$  est réel.

Rappelons que  $\pi(\theta|\mathbf{x})$  s'interprète comme un résumé de l'information disponible sur  $\theta$ , une fois  $\mathbf{x}$  observé. D'un point de vue bayésien, l'idéal serait que le statisticien, communique à son interlocuteur (le médecin, l'économiste, l'ingénieur, etc) la loi a posteriori, de façon à ne pas perdre d'information. Quand on souhaite cependant disposer d'une estimation pour  $\theta$ , on retient le plus souvent la moyenne de la loi a posteriori.

Définition. On appelle estimation bayésienne du paramètre  $\theta$  la moyenne de la loi a posteriori. Cette moyenne est notée  $E[\theta|\mathbf{x}]$ . Formellement, on a:

$$E[\theta|\mathbf{x}] = \int_{\Theta} \theta \pi(\theta|\mathbf{x}) d\theta = \frac{\int_{\Theta} \theta f(\mathbf{x}|\theta) \pi(\theta) d\theta}{\int_{\Theta} f(\mathbf{x}|\theta) \pi(\theta) d\theta}.$$

L'estimateur de Bayes de  $\theta$  est noté  $\hat{\theta}_B$ . Il est donc défini par:  $\hat{\theta}_B(\mathbf{x}) = E[\theta|\mathbf{x}]$ .

Un exemple.

On considère le modèle bayésien suivant:  $X_i|\theta \sim \text{Bernoulli}(\theta)$  et  $\theta \sim \text{Beta}(a,b)$ . Rappelons que:  $\theta|\mathbf{x} \sim \text{Beta}(\alpha,\beta)$ , où  $\alpha = a + s$  et  $\beta = b + n - s$  et  $s = \sum_{i=1}^n x_i$ . D'où:

$$E[\theta|\mathbf{x}] = \frac{a + \sum_{i=1}^n x_i}{a + b + n}.$$

Remarques.

- Il faut signaler que le mode de la loi a posteriori est parfois retenu comme estimation bayésienne de  $\theta$  si l'on suspecte que la vraisemblance est multimodale (c.a.d. comporte des maxima locaux). En ce cas, on a:

$$\hat{\theta}_B(\mathbf{x}) = \arg \max_{\theta \in \Theta} \pi(\theta|\mathbf{x}).$$

- Dans ce cours, quand on parlera par la suite d'estimation bayésienne de  $\theta$ , il s'agira toujours de la moyenne de la loi a posteriori.

- L'estimation bayésienne de  $h(\theta) \in \mathbb{R}$  est par définition  $E[h(\theta)|\mathbf{x}]$ .

Exercice. Soit le modèle statistique bayésien suivant:  $X_i|\theta \sim \text{Poisson}(\theta)$  et  $\theta \sim \text{Gamma}(a,b)$ . Déterminer l'estimateur de Bayes de  $\theta$ . Même question si:  $X_i|\theta \sim \text{Normale}(\theta,1)$  et  $\theta \sim \text{Normale}(0,1)$ .

### 3.1.2 Le cas multi-dimensionnel

Dans un contexte multi-dimensionnel où  $\theta = (\theta_j; j = 1, \dots, J)$  la moyenne a posteriori  $E[\theta|\mathbf{x}]$  est égale au vecteur  $(E[\theta_j|\mathbf{x}]; j = 1, \dots, J)$ , où

$$E[\theta_j|\mathbf{x}] = \int_{\Theta_j} \theta_j \pi(\theta_j|\mathbf{x}) d\theta_j.$$

$\pi(\theta_j|\mathbf{x})$  est obtenu en intégrant  $\pi(\theta|\mathbf{x})$  sur toutes les composantes de  $\theta$  autres que  $\theta_j$ .

Le plus souvent, les estimateurs de Bayes des  $\theta_j$  ne peuvent pas être calculés de façon explicite et il faut faire appel aux méthodes de simulation de Monte Carlo. Voici cependant une situation dans laquelle le calcul des estimateurs de Bayes ne pose aucune difficulté.

Exercice. Calculer les estimateurs de Bayes des probabilités de transition d'une chaîne de Markov (à 2 états et de longueur fixée  $T$ ) quand on dispose de  $n$  réalisations de cette même chaîne. On supposera que l'état initial de la chaîne est connu et n'est pas aléatoire. On se placera de plus dans le cadre du modèle d'échantillonnage (cf le TD No 1).

### 3.2 Le risque de Bayes

La recherche d'estimateurs de Bayes peut se faire dans le cadre de la théorie de la décision. La démarche consiste alors à se fixer une règle de préférence entre estimateurs et à chercher un estimateur optimal au sens de cette règle de préférence. Rappelons qu'en statistique classique la règle de préférence repose (le plus souvent) sur le risque quadratique, noté  $R(\hat{\theta})$ , et défini comme suit:

$$R(\hat{\theta}) = \text{Var}(\hat{\theta}_B) + [\text{biais}(\hat{\theta}_B)]^2.$$

L'approche Bayésienne fait reposer la règle de préférence sur le risque de Bayes. La densité a priori  $\pi(\theta)$  étant fixée, le risque de Bayes de  $\hat{\theta}$  est noté  $\mathcal{R}(\hat{\theta})$ . Il est défini comme suit:

$$\mathcal{R}(\hat{\theta}) = E[R(\hat{\theta})] = \int_{\Theta} R(\hat{\theta}) \pi(\theta) d\theta.$$

On dira que  $\hat{\theta}_1$  est meilleur que  $\hat{\theta}_2$  au sens du risque de Bayes, si

$$\mathcal{R}(\hat{\theta}_1) < \mathcal{R}(\hat{\theta}_2).$$

A noter que, contrairement à la règle de préférence basée sur le risque quadratique il est toujours possible de comparer deux estimateurs de  $\theta$  au sens du risque de Bayes. La densité a priori  $\pi(\theta)$  étant fixée, on montre que l'estimateur de Bayes est l'estimateur optimal de  $\theta$  au sens du risque de Bayes.

### 3.3 Propriétés de l'estimateur de Bayes

P1. L'estimateur de Bayes est admissible.

P2. L'estimateur de Bayes est biaisé.

Sous certaines hypothèses de régularité le plus souvent satisfaites en pratique, on a les deux propriétés:

P3. L'estimateur de Bayes est convergent en probabilité (quand la taille de l'échantillon  $n \rightarrow +\infty$ ).

P4. La loi a posteriori peut être asymptotiquement (c.a.d. pour de grandes valeurs de  $n$ ) approximée par une loi normale  $\mathcal{N}(E[\theta|\mathbf{x}], Var[\theta|\mathbf{x}])$ .

Cette dernière propriété est particulièrement utile pour construire des des intervalles de confiance a posteriori (cf Section 3.5).

### 3.4 Loi a priori impropre et estimateur de Bayes généralisé

Définitions.

- Soit  $\pi(\theta)$  une application de  $\Theta$  dans  $]0, +\infty[$  tel que:

$$\int_{\Theta} \pi(\theta) d\theta = +\infty;$$

on parle alors de loi a priori impropre. Cette terminologie est bien sûr un abus de langage puisque  $\pi(\theta)$  n'est pas une densité de probabilité. L'intérêt d'introduire une telle notion est fourni par la définition qui suit.

- On se donne une loi a priori impropre caractérisée par  $\pi(\theta)$ . On suppose que l'intégrale

$$\int_{\Theta} L(\theta; \mathbf{x}) \pi(\theta) d\theta$$

est convergente. On considère la densité de probabilité, en  $\theta$ , définie par:

$$\frac{L(\theta; \mathbf{x}) \pi(\theta)}{\int_{\Theta} L(\theta; \mathbf{x}) \pi(\theta) d\theta}.$$

Cette densité est notée  $\pi(\theta|\mathbf{x})$  et est appelée densité de la loi a posteriori. Dans ces conditions, on appelle estimateur de Bayes généralisé de  $\theta$ , la moyenne de cette loi a posteriori.

Exercice.  $X_i|\theta \sim$  Bernoulli  $\theta$  et  $\pi(\theta) = [\theta(1 - \theta)]^{-1} \mathbb{I}_{]0,1[}$ . Calculer l'estimateur de Bayes généralisé de  $\theta$ . A quelle condition est-il défini? Le comparer au M.L.E. Même question, quand  $X_i|\theta \sim$  Normale( $\theta,1$ ) et  $\pi(\theta) = 1$ .



### 3.5 Les intervalles de confiance bayésiens

On se donne un modèle bayésien et on suppose que  $\theta$  est un paramètre réel.

Définition. Soit  $\alpha \in ]0,1[$  fixé. Un interval  $I$  pour lequel on a:

$$P(\theta \in I|\mathbf{x}) = \int_J \pi(\theta|\mathbf{x}) d\theta = 1 - \alpha$$

est appelé un intervalle de confiance a posteriori, de niveau de confiance  $1 - \alpha$ .

On peut également définir la notion d'intervalle de confiance a priori. Un interval  $J$  pour lequel on a:

$$P(\theta \in J) = \int_J \pi(\theta) d\theta = 1 - \alpha$$

sera appelé un intervalle de confiance a priori, de niveau de confiance  $1 - \alpha$ .

Exercice 1.  $X_i|\theta \sim$  Normale  $(\theta,1)$  and  $\pi(\theta) = 1$ . Chercher un intervalle de confiance a posteriori de niveau de confiance égal à 95% centré sur  $E[\theta|\mathbf{x}]$ ; puis le comparer à l'intervalle de confiance (observé) classique, de niveau de confiance égal à 95%

Exercice 2.  $X_i|\theta \sim$  Bernouilli  $(\theta)$  et  $\theta$  suit une loi uniforme sur  $[0,1]$ . On suppose que l'on a observé 10 succès parmi  $n = 40$  essais. En utilisant l'approximation de la loi a posteriori par une loi normale, donner un intervalle de confiance a posteriori, noté  $J$ , de niveau de confiance égal à 95%. Puis donner la valeur exact de  $P(\theta \in J|\mathbf{x})$  en utilisant par exemple Maple ou Matla). L'approximation normale est-elle précise?

## 4 L'approche bayésienne des tests

Dans ce chapitre nous nous limiterons à la situation suivante. On suppose que l'espace  $\Theta$  des paramètres est partitionné en  $\Theta_0$  et  $\Theta_1$ , et que  $P(\theta \in \Theta_k) > 0$  où  $k = 0,1$ . On souhaite tester:  $H_0 : \theta \in \Theta_0$  contre l'alternative  $H_1 : \theta \in \Theta_1$ . En statistique bayésienne, la réponse à un tel test repose sur les probabilités a posteriori des hypothèses  $H_0$  et  $H_1$ :

$$P(H_0|\mathbf{x}) = P(\theta \in \Theta_0|\mathbf{x}) = \int_{\Theta_0} \pi(\theta|\mathbf{x}) d\theta$$

A noter que  $P(H_1|\mathbf{x}) = 1 - P(H_0|\mathbf{x})$ . En pratique, une hypothèse ( $H_0$  ou  $H_1$ ) est acceptée dès que sa probabilité a posteriori est jugée suffisamment forte (typiquement supérieur à 0.9 ou à 0.95). Il se peut qu'aucune des probabilités a posteriori excède 0.9; deux attitudes sont alors possibles:

- soit on ne prend pas de décision, et on décide par exemple de recueillir davantage d'observations;
- soit on choisit l'hypothèse dont la probabilité a posteriori est la plus grande;

A noter que cela a aussi un sens de définir les probabilité a priori de chacune des deux hypothèses  $H_0$  et  $H_1$ . Ainsi on a:

$$P(H_0) = P(\theta \in \Theta_0) = \int_{\Theta_0} \pi(\theta) d\theta$$

Il convient enfin de noter que les écritures  $P(H_0)$  et  $P(H_0|\mathbf{x})$  n'ont pas de sens en statistique classique.

Exercice.  $X_i|\theta \sim$  Normale  $(\theta,1)$  and  $\theta \sim$  Normale  $(0,1)$  On souhaite:  $H_0 : \theta > 0$  contre l'alternative  $H_1 : \theta \leq 0$ . Comparer la probabilité a posteriori de  $H_0$  avec la p-value.

## 5 Modélisation de l'information a priori

Le plus souvent on ne dispose pas de suffisamment d'information a priori sur le paramètre inconnu  $\theta$  pour construire la loi a priori. Dans la pratique on a recours à des lois usuelles (lois normales, lois gamma, etc) ou à des lois dites conjuguées (voir ci-dessous), l'information a priori étant alors utilisée pour déterminer les paramètres de la loi a priori, appelés hyperparamètres. En l'absence d'information a priori on introduira la notion de loi a priori non informative qui permet de rester dans un cadre bayésien, alors même que l'on ne dispose pas d'information a priori.

### 5.1 Lois a priori conjuguées

#### Définition.

La loi des observations étant supposée connue, on se donne une famille  $\mathcal{F}$  de lois de probabilité sur  $\Theta$ . On suppose que la loi a priori appartient à  $\mathcal{F}$ . Si dans ces conditions, la loi a posteriori appartient encore à  $\mathcal{F}$ , on dit que la loi a priori est conjuguée.

#### Exemple.

Soit  $\mathbf{x} = (x_1, \dots, x_i, \dots, x_n)$  l'observation. On suppose que:  $X_i|\theta$  suit une loi de Bernoulli de paramètre  $\theta$  et que la loi a priori est une loi Beta. Comme  $\theta|\mathbf{x}$  suit aussi une loi Beta, on en déduit que la loi Beta est ici conjuguée.

Exercice.  $X_i|\theta \sim$  Exponentielle ( $\theta$ ). Montrer que la loi Gamma est une loi a priori conjuguée.

### 5.2 Lois a priori non informatives

Partons d'un exemple pour introduire la notion de loi a priori non informative. On considère le modèle statistique bayésien suivant: les  $X_i|\theta$  sont *i.i.d.* et suivent une loi de Bernoulli de paramètre  $\theta \in ]0,1[$ . Il n'y a pas une unique loi a priori non informative pour le paramètre  $\theta$ . On peut en fait proposer différentes lois a priori non informatives.

- En l'absence d'information a priori sur  $\theta$ , il est naturel de proposer une loi uniforme sur  $\theta$  car elle donne une probabilité égale aux intervalles de longueur  $l$  donnée, à savoir  $l$ .
- On peut également proposer la loi a priori (impropre) de Haldane  $\pi(\theta) = [\theta(1-\theta)]^{-1} \mathbb{I}_{]0,1[}(\theta)$ , en arguant que  $E[\theta|\mathbf{x}]$  est égal à l'estimateur du maximum de vraisemblance (le vérifier).
- Une alternative a été proposée par Jeffreys en 1960.

#### Définition.

Soit  $\theta$  un paramètre réel. On appelle loi a priori non informative de Jeffreys, la loi (éventuellement impropre) de densité

$$\pi_J(\theta) \propto [I(\theta)]^{\frac{1}{2}} \mathbb{I}_{\Theta}(\theta)$$

où  $I(\theta)$  désigne l'information de Fisher apportée par  $\mathbf{x}$  sur  $\theta$ .

Justification de la loi de Jeffreys. D'une part  $I(\theta)$  s'interprète comme la quantité d'information apportée par l'observation  $\mathbf{x}$  sur  $\theta$ . D'autre part choisir  $\pi_J$  ne fait pas intervenir d'autre information que celle apportée par les observations (en fait, au travers de  $f(\mathbf{x}|\theta)$ ). En ce sens, la loi a priori de Jeffrey est non informative.

#### Exemple.

Les  $X_i|\theta$  sont *i.i.d.* et  $X_i|\theta$  suit une Bernoulli de paramètre  $\theta$ . On montre que  $I(\theta) = [\theta(1-\theta)]^{-1}$ . D'où:  $\pi_J(\theta) \propto [\theta(1-\theta)]^{-\frac{1}{2}} \mathbb{I}_{]0,1[}(\theta)$ . Autrement dit, la loi a priori est une loi Beta de paramètres  $(1/2, 1/2)$

### 5.3 Incorporation de l'information a priori en pratique.

Nous allons nous limiter au cas d'un paramètre réel  $\theta$  et envisager trois situations selon que  $\theta \in [0,1]$ ,  $[0, +\infty[$ , ou  $] -\infty, +\infty[$ . Nous supposons par ailleurs que l'information a priori fourni par l'Expert consiste en une estimation  $\theta^*$  de  $\theta$ , et en un intervalle  $I^* = [\alpha^*, \beta^*]$  contenant  $\theta^*$  et tel que la probabilité que  $\theta$  appartienne à  $I^*$  soit élevée (typiquement 0.95).

Situation No 1: le paramètre  $\theta \in [0,1]$ .

Pour incorporer cette information a priori, il est commode de re-paramétriser la loi Beta comme suit: On pose  $a = \lambda\mu$  et  $b = \lambda(1 - \mu)$  où  $\mu = \frac{a}{a+b} = \mathbb{E}[\theta]$  et  $\lambda = a + b$ . A noter que  $\text{Var } \theta = \frac{\mu(1-\mu)}{1+\lambda}$ . Donc à  $\mu$  fixé,  $\lambda$  est inversement proportionnel à la variance et donc s'interprète comme la précision de l'information a priori. Autrement dit, plus  $\lambda$  est grand, et plus  $\text{Var}(\theta)$  est petit, et plus l'information a priori est précise; à l'inverse, plus  $\lambda$  est petit, et plus  $\text{Var}(\theta)$  est grand, et moins l'information a priori est précise.

En pratique, on procède comme suit. On choisit comme loi a priori une loi Beta( $a,b$ ) et on pose:  $E(\theta) = \theta^*$ . Le paramètre  $\lambda$  est lui déterminé par

$$\Pr(\theta \in I^*) = \int_{I^*} \pi(\theta) d\theta = 0.95$$

où  $\theta \sim \text{Beta}(\lambda\theta^*, \lambda(1 - \theta^*))$ . En pratique  $\lambda$  est déterminé, soit par tâtonnement à l'aide de Matlab (cf TP No 1), soit à l'aide d'un algorithme qu'il convient de programmer (et qui procède par dichotomie). Donnons un exemple. On suppose que:  $\theta^* = 0.2$  et que  $\Pr(\theta \in [0.05, 0.4]) = 0.95$ . On a donc:  $E(\theta) = 0.2$  et  $\theta \sim \text{Beta}(0.2\lambda, 0.8\lambda)$ . A l'aide de Matlab, on trouve  $\lambda = 17.5$ ; par conséquent, la loi a priori est une loi Beta de paramètres  $(a,b) = (3.5, 14)$ .

Situation No 2: le paramètre  $\theta \in [0, +\infty[$ .

Dans ce cas on choisit comme loi a priori une loi Gamma qui peut être reparamétrisée par sa moyenne et sa variance (comme la loi Beta). Le vérifier en exercice.

Situation No 3: le paramètre  $\theta \in ] -\infty, +\infty[$ .

Dans ce cas on choisit, comme loi a priori, une loi Normale. Exercice. Incorporer l'information a priori consistant en  $\theta^* = 1$  et  $I^* = [0.8, 1.2]$

### 5.4 Le poids de l'a priori dans la réponse bayésienne

Examinons cette question sur un exemple pour comprendre comment l'information a priori et l'information contenue dans les observations se combinent l'une à l'autre pour produire la réponse bayésienne. On se donne le modèle bayésien suivant:  $X_i|\theta \sim \text{Bernoulli}(\theta)$  et  $\theta \sim \text{Beta}(a,b)$ . Il est commode de reparamétriser la loi Beta à l'aide de  $\lambda$  et  $\mu$  (comme ci-dessus) et de travailler avec la formule suivante (facile à établir):

$$E[\theta|\mathbf{x}] = \frac{\lambda}{\lambda + n} E[\theta] + \frac{n}{\lambda + n} \bar{x}.$$

L'estimation bayésienne de  $\theta$  apparaît donc comme la moyenne pondérée de  $\bar{x}$  (cad de l'estimation de  $\theta$  par maximum de vraisemblance), et de la moyenne a priori  $E[\theta]$ ; le poids de  $\bar{x}$  est la taille  $n$  de l'échantillon, et celui de  $E[\theta]$  est  $\lambda$  qui s'interprète comme la précision de l'a priori (le dénominateur commun  $\lambda + n$  a été sciemment omis, car seul compte en fait le numérateur). Géométriquement,  $E[\theta|\mathbf{x}]$  est le barycentre des points de coordonnées  $E[\theta]$  et  $\bar{x}$ , affectés respectivement des coefficients  $\frac{\lambda}{\lambda+n}$  et  $\frac{n}{\lambda+n}$ .

A noter que si  $\lambda = n$ , l'estimation bayésienne de  $\theta$  se situe exactement au milieu de l'intervalle  $[E[\theta], \bar{x}]$ , si  $\lambda > n$  cette estimation est plus proche de  $E[\theta]$  que de  $\bar{x}$ , et si  $\lambda < n$  c'est l'inverse qui se produit.

Pour examiner l'influence de l'a priori sur  $E[\theta|\mathbf{x}]$  il est édifiant de s'intéresser aux cas limites:  $\lambda \rightarrow 0$  et  $\lambda \rightarrow \infty$  (la taille  $n$  de l'échantillon étant fixé, ainsi que  $\mu = E(\theta)$ ). Dans le premier cas, le poids de l'a

priori est nul, et  $E[\theta|\mathbf{x}] \rightarrow \bar{x}$  qui est la réponse classique. Dans le second cas, le poids des données est nul, et  $E[\theta|\mathbf{x}] \rightarrow E[\theta]$  qui ne dépend plus de  $\mathbf{x}$ . Le tableau ci-dessous résume la situation.

	$\lambda \rightarrow 0$	$\lambda \rightarrow +\infty$
Var ( $\theta$ )	$\mu(1 - \mu)$ <i>maximale</i>	0 <i>minimale</i>
loi a priori	loi de Haldane	loi concentrée en $\mu$
interprétation	<i>situation non informative</i>	<i>situation extrêmement informative</i>
Estimation bayésienne	$\bar{x}$	$E[\theta]$

Il est également intéressant de regarder ce que devient  $E[\theta|\mathbf{x}]$  quand  $n \rightarrow +\infty$ ,  $\lambda$  et  $\mu$  étant fixés; dans ce cas, le poids de l'a priori devient négligeable, et la réponse bayésienne coïncide avec la réponse classique, cad  $\bar{x}$  (l'estimation de  $\theta$  par maximum de vraisemblance).

## 5.5 Étude de la sensibilité de la réponse bayésienne à la loi priori

Le choix de la loi a priori comporte une certaine part d'arbitraire, et ce, à deux niveaux. Le premier niveau réside dans le choix de la famille de probabilités retenue pour la loi a priori (une loi Beta pour un paramètre dans  $[0; 1]$ , une loi Gamma pour un paramètre dans  $[0; +\infty[$ , etc). Le second réside dans le choix des valeurs numériques communiquées par l'Expert (typiquement  $\theta^*$  et  $I^*$ ) pour déterminer les hyperparamètres de la loi a priori. Il convient par conséquent d'examiner dans quelle mesure ces choix affectent l'estimation bayésienne de  $\theta$ . Dans la pratique, on se contente d'examiner, dans quelle mesure, une petite perturbation des hyperparamètres modifie l'estimation Bayésienne de  $\theta$ . C'est ce qu'on appelle: faire l'étude de sensibilité de la réponse bayésienne à la loi a priori.

Pour conclure cette Section 5, disons qu'en pratique, on comparera toujours l'estimateur de Bayes obtenu avec la loi a priori informative, avec l'estimateur de Bayes correspondant à une loi a priori non informative; et ce, afin d'apprécier le poids de l'a priori dans la réponse Bayésienne informative.

## 6 Méthodes de simulation de Monte Carlo

### 6.1 Méthode classique de simulation de Monte Carlo

#### 6.1.1 Motivations

Afin de motiver les méthodes de simulation de Monte Carlo dans un contexte bayésien, nous allons considérer une situation dans laquelle le calcul de l'estimation bayésienne de la grandeur d'intérêt est particulièrement laborieux à la main. Il importe de préciser que l'utilisation des méthodes de simulation de Monte Carlo ne se limite pas à la statistique bayésienne (ni même à la statistique); c'est pourquoi nous allons aussi donner un autre exemple de l'utilisation de ces méthodes, et ce dans un cadre purement probabiliste.

Le modèle No 1. Etude bayésienne de l'efficacité d'un traitement.

$n = 20$  patients atteints d'une même maladie suivent tous le même traitement. L'état de chaque patient est déterminé, régulièrement, lors de visites médicales. On désigne par  $x_{(i,t)}$  l'état du patient  $i$  lors de la  $t$ -ième visite;  $x_{(i,t)} = a$  s'il présente les symptômes de la maladie  $M$ , et  $b$  sinon. Les  $X_{(i,t)}$  sont markoviens ( $i$  étant fixé). La probabilité de transition de  $a$  vers  $a$  est notée  $\alpha$  et celle de  $b$  vers  $b$  est notée  $\beta$ . Le paramètre du modèle est donc  $\theta = (\alpha, \beta)$ . Le suivi du patient  $i$  est noté  $x_i$ . Les  $X_i$  sont supposés indépendants (conditionnellement à  $\theta$ ) et suivent tous la même loi. On suppose que les paramètres  $\alpha$  et  $\beta$  sont a priori indépendants. On adopte des lois uniformes comme lois a priori.

On note  $\omega$  la probabilité qu'un patient soit guéri à l'issue du traitement (on considère que c'est le cas s'il ne présente plus les symptômes de la maladie lors de la quatrième visite). On considère que le traitement

est efficace si l'estimation de  $\omega$  dépasse 0.95. Il est facile de se convaincre que le calcul à la main de  $E[\omega|\mathbf{x}]$  est particulièrement laborieux, même pour des valeurs modérées du nombre de visites.

Le modèle No 2. Contamination de l'environnement par un maïs transgénique.

Un pied de maïs transgénique a été planté en plein champ. On note  $M$  le point de chute au sol d'une graine de ce pied de maïs. On désigne par  $X_1$  et  $X_2$  les coordonnées du point  $M$  dans un repère orthonormé ayant pour origine le pieds de maïs transgénique. L'origine du repère est notée par la suite  $O$ . On suppose que les v.a.  $X_1$  et  $X_2$  sont indépendantes et que  $X_j \sim \mathcal{N}(m_j, \sigma_j^2); j = 1, 2$  où les  $(m_j, \sigma_j^2)$  sont connus. On souhaite évaluer la distance moyenne de  $O$  à  $M$ . Il s'agit donc de calculer l'intégrale:

$$\int_{\mathcal{X}} \sqrt{x_1^2 + x_2^2} f(\mathbf{x}) d\mathbf{x}.$$

où  $\mathcal{X} = ]-\infty, +\infty[ \times ]-\infty, +\infty[$ . On souhaite également évaluer la probabilité que la distance  $OM$  n'excède pas une valeur  $a$  donnée; valeur au delà de laquelle cette graine contaminerait l'environnement. Il est clair que le calcul de cette intégrale et de cette probabilité ne sont pas faisable à la main.

### 6.1.2 Le principe de la méthode

Soit  $\mathbf{V} = (V_1, \dots, V_j, \dots, V_k)$ , où  $k \geq 1$ , un vecteur aléatoire de densité  $f(\mathbf{v})$ . On souhaite calculer  $E[h(\mathbf{V})]$  où  $h(\mathbf{V})$  est réel. On suppose que l'on sait simuler selon la loi de  $\mathbf{V}$ . Le principe de ces méthodes est de simuler une suite de variables aléatoires  $\mathbf{V}^{(1)}, \dots, \mathbf{V}^{(l)}$  i.i.d. selon la loi de  $V$ , puis d'approcher  $E[h(\mathbf{V})]$  à l'aide de la loi des grands nombres:

$$\frac{1}{L} \sum_{l=1}^L h(\mathbf{V}^{(l)}) \xrightarrow{L \rightarrow +\infty} E[h(\mathbf{V})].$$

Décrivons maintenant cette méthode de simulation classique de Monte Carlo dans un contexte bayésien. On suppose que l'on souhaite calculer  $E[h(\theta)|\mathbf{x}]$ , où  $h(\theta)$  est réel et où  $\mathbf{x}$  désigne l'ensemble des observations. On suppose que l'on sait simuler selon la loi a posteriori. Le principe de ces méthodes est de simuler une suite de variables aléatoires  $\theta^{(1)}, \dots, \theta^{(l)}$  i.i.d. selon la loi a posteriori, puis d'approcher  $E[h(\theta)|\mathbf{x}]$  à l'aide de la loi des grands nombres:

$$\frac{1}{L} \sum_{l=1}^L h(\theta^{(l)}) \xrightarrow{L \rightarrow +\infty} E[h(\theta)|\mathbf{x}].$$

## 6.2 Les méthodes de simulation de Monte Carlo par chaîne de Markov

### 6.2.1 Motivations

Comme précédemment, il importe de préciser que l'utilisation des méthodes de simulation de Monte Carlo par chaîne de Markov ne se limite pas à la statistique bayésienne (ni même à la statistique); c'est pourquoi nous allons donner un exemple d'utilisation de ces méthodes dans un cadre purement probabiliste, puis dans un cadre bayésien.

Le modèle No 2 (suite). Contamination de l'environnement par un maïs transgénique.

On reprend la situation précédente, mais on ne suppose plus cette fois que les v.a.  $X_1$  et  $X_2$  sont indépendantes. On fait simplement l'hypothèse que le couple  $(X_1, X_2)$  est distribué selon une loi normale bidimensionnelle de moyenne connue et de matrice de variance-covariance également connue. Si l'on dispose d'un logiciel permettant de simuler une v.a. distribuée selon une loi normale bidimensionnelle, il est alors possible d'utiliser la méthode classique de simulation de Monte Carlo (cf la Section précédente); mais les logiciels standards n'offre pas en réalité cette possibilité. D'où la nécessité de mettre au point des algorithmes spécifiques.

Le modèle No 3. Estimation bayésienne de la taille d'une population animale.

Pour fixer les idées, supposons que l'on cherche à estimer le nombre de poissons vivant dans un lac; ce nombre est noté par la suite  $N$ . Pour certaines régions du globe (comme au Canada, par exemple), ce problème est d'un grand intérêt économique et écologique.  $N$  est ici estimé à partir d'un échantillonnage par capture-marquage-recapture.

*Le protocole expérimental.*

Au temps  $t_1$ , un premier échantillon de  $k_1$  poissons est extrait du lac. Ceux-ci sont marqués (en appliquant, par exemple, un produit coloré et indélébile sur la nageoire dorsale); puis, ils sont tous relâchés dans le lac. Au temps  $t_2$  (en pratique quelques jours plus tard), un second échantillon de  $k_2$  poissons est extrait du lac; parmi ceux-ci, certains poissons sont marqués et d'autres non. On note simplement le nombre de poissons marqués; puis, tous sont relâchés dans le lac.

*Description des données.*

Pour un poisson  $i$  du lac, quatre situations sont possibles, et ce relativement à son appartenance (ou non) aux échantillons 1 et 2; chaque situation s'appelle une histoire de capture-recapture. L'histoire de capture-recapture du poisson  $i$  est notée  $x_i$ . Il y a en tout quatre histoires de capture-recapture possibles:  $(0,0)$ ,  $(1,0)$ ,  $(0,1)$ ,  $(1,1)$  qui sont numérotées de 0 à 3 (dans cet ordre). La première composante de  $x_i$  (resp. la seconde) renseigne sur l'appartenance du poisson  $i$  à l'échantillon 1 (resp. l'échantillon 2); l'appartenance étant codée 1 et la non-appartenance codée 0. On note  $n_h$  le nombre de poissons dont l'histoire de capture-recapture a pour numéro  $h$  (où  $h \in \{0,1,2,3\}$ ). L'ensemble des observations est  $\mathbf{y} = \{n_1, n_2, n_3\}$ . A noter que  $n_0$  n'est évidemment pas observé. A noter également que  $k_1 = n_1 + n_3$ ,  $k_2 = n_2 + n_3$  et que  $N = n_0 + n_1 + n_2 + n_3$ .

*Les hypothèses et les paramètres du modèle*

- On suppose que la population de poissons est fermée; ce qui signifie que  $N$  est constant sur l'intervalle de temps  $[t_1, t_2]$ .
- La probabilité qu'un poisson fasse partie de l'échantillon 1 est notée  $p$ . S'agissant de la probabilité de faire partie de l'échantillon 2, il faut distinguer deux cas: soit le poisson a fait partie de l'échantillon 1, et dans ce cas cette probabilité est notée  $q$ ; soit il n'en a pas fait partie, et dans ce cas, cette probabilité est égale à  $p$ . On adopte cette modélisation pour tenir compte du fait que la probabilité de capture d'un poisson marqué diffère en général de celle d'un poisson non marqué.
- Les  $X_i$  sont *i.i.d.* (conditionnellement aux paramètres du modèle).
- On suppose que les paramètres du modèle sont a priori indépendants, que  $p$  suit une loi Beta( $a, b$ ), que  $q$  suit une loi Beta( $c, d$ ) et que  $N$  suit une loi Binomiale négative de paramètres  $(r, \alpha)$ . Les hyperparamètres sont supposés connus.

Comme précédemment, le calcul à la main de  $E[N|\mathbf{y}]$  n'est pas faisable, et simuler  $N$  selon la loi de  $E[N|\mathbf{y}]$  n'est pas aisée. A nouveau, il y a donc nécessité de développer des algorithmes adéquates.

### 6.2.2 Le principe de la méthode

Soit  $\mathbf{V} = (V_1, \dots, V_j, \dots, V_k)$ , où  $k \geq 1$ , un vecteur aléatoire de densité  $f(\mathbf{v})$ . On souhaite calculer  $E[h(\mathbf{V})]$  où  $h(\mathbf{V})$  est réel. On suppose que l'on ne sait pas simuler selon la loi de  $\mathbf{V}$ . Le principe des méthodes de simulation de Monte Carlo par chaîne de Markov est de générer une chaîne de Markov  $(\mathbf{V}^{(l)}; l \geq 0)$  qui soit asymptotiquement distribuée selon la loi de  $\mathbf{V}$ , puis d'approcher  $E[h(\mathbf{V})]$  à l'aide du théorème ergodique:

$$\frac{1}{L} \sum_{l=1}^L h(\mathbf{V}^{(l)}) \xrightarrow{L \rightarrow +\infty} E[h(\mathbf{V})].$$

Précisons que  $\mathbf{V}^{(l)}$  est un vecteur de la même dimension que  $\mathbf{V}$ . Il existe deux algorithmes à même de produire une telle chaîne de Markov: il s'agit de l'algorithme de Gibbs et de l'algorithme de Hastings-Metropolis. Dans ce cours on se limitera à l'algorithme de Gibbs (l'algorithme de Hastings-Metropolis étant présenté à titre de complément).

### 6.2.3 L'algorithme de Gibbs

On suppose que  $k \geq 2$ , où  $k$  est la dimension de  $\mathbf{V}$ . On suppose, de plus que, pour tout  $j$ , on sait simuler suivant la loi de  $V_j|\bar{v}_j$ , où  $\bar{v}_j = \mathbf{v} - \{v_j\}$ . On part de  $\mathbf{V}^{(0)}$  arbitraire. A l'étape ( $l$ ) de l'algorithme de Gibbs on simule le vecteur  $\mathbf{V}^{(l)} = (V_j^{(l)}; j = 1, \dots, k)$  comme suit:

$$V_1^{(l)} \sim \text{Loi}(V_1|V_2 = v_2^{(l-1)}, \dots, V_k = v_k^{(l-1)})$$

$$V_j^{(l)} \sim \text{Loi}(V_j|V_1 = v_1^{(l)}, \dots, V_{j-1} = v_{j-1}^{(l)}, V_{j+1} = v_{j+1}^{(l-1)}, \dots, V_k = v_k^{(l-1)})$$

$$V_k^{(l)} \sim \text{Loi}(V_k|V_1 = v_1^{(l)}, \dots, V_{k-1} = v_{k-1}^{(l)})$$

Exercice. Indiquer comment mettre en oeuvre l'algorithme de Gibbs dans cadre bayésien; pour ce faire on décrira une étape ( $l$ ) de l'algorithme de Gibbs. On supposera, pour fixer les idées, que  $\theta = (\theta_1, \theta_2)$ , et on notera  $\mathbf{y}$  l'ensemble des observations. Indiquer comment obtenir  $E[\theta_k|\mathbf{y}]$ , où  $k = 1, 2$ .

### 6.2.4 L'algorithme de Hastings-Metropolis

On part de  $\mathbf{v}^{(0)}$  arbitraire. L'étape ( $l$ ) se déroule comme suit. On simule une v.a.  $\mathbf{U}^{(l)} \sim q(\mathbf{u}|\mathbf{v}^{(l-1)})$  où  $\mathbf{U}$  est un vecteur de dimension  $k$ , et où  $q(\cdot|\cdot)$  est une densité symétrique (au sens où  $q(\mathbf{u}|\mathbf{v}) = q(\mathbf{v}|\mathbf{u})$ ).

Cas 1.  $f(\mathbf{u}^{(l)}) \geq f(\mathbf{v}^{(l-1)})$ ; dans ce cas,  $\mathbf{v}^{(l)} = \mathbf{u}^{(l)}$ .

Cas 2.  $f(\mathbf{u}^{(l)}) < f(\mathbf{v}^{(l-1)})$ ; dans ce cas on accepte  $\mathbf{v}^{(l)}$  avec probabilité  $\rho^{(l)}$ , où  $\rho^{(l)} = f(\mathbf{u}^{(l)})/f(\mathbf{v}^{(l-1)})$ , sinon, on rejette cette valeur, et dans ce cas,  $\mathbf{v}^{(l)} = \mathbf{v}^{(l-1)}$

A noter que l'algorithme, permet 'd'échapper' à des maxima locaux (s'il en existe), et de visiter ainsi tout l'espace (dans lequel la v.a.  $\mathbf{V}$  prend ses valeurs).

## 7 Mise en oeuvre des méthodes de simulation de Monte Carlo

### 7.1 Le modèle 1: étude bayésienne de l'efficacité d'un traitement

Exercice. Indiquer comment obtenir, à l'aide d'un algorithme (classique) de simulation de Monte Carlo, l'estimation bayésienne de  $\omega$ . *Indication*: On rappelle que  $\alpha|\mathbf{x} \sim \text{Beta}(1 + N(a,a), 1 + N(a,b))$  et que  $\beta|\mathbf{x} \sim \text{Beta}(1 + N(b,b), 1 + N(b,a))$  où  $N(r,s)$  désigne le nombre total de transitions de  $r$  vers  $s$  entre les temps  $t$  et  $t + 1$  comptées sur le tableau de données  $\mathbf{x}$ . On rappelle également que  $\pi(\alpha, \beta|\mathbf{x}) = \pi(\alpha|\mathbf{x})\pi(\beta|\mathbf{x})$ ; autrement dit, pour simuler  $\theta$  selon la loi a posteriori il suffit de simuler les deux lois Beta ci-dessus de façon indépendantes.

### 7.2 Le modèle 2: contamination de l'environnement par un maïs transgénique

Exercice 1. On suppose que les v.a.  $X_1$  et  $X_2$  sont indépendantes (cf section précédente). Indiquer comment, à l'aide d'un algorithme (classique) de simulation de Monte Carlo, obtenir la distance moyenne de 0 à  $M$ , ainsi que la probabilité que la distance OM n'excède pas une valeur  $a$  donnée.

Exercice 2. On ne suppose plus que les v.a.  $X_1$  et  $X_2$  sont indépendantes (cf section précédente). Chercher les lois conditionnelles de  $X_1|X_2$  puis de  $X_2|X_1$ . Puis décrire une étape ( $l$ ) de l'algorithme de Gibbs. Indiquer comment, à l'aide de l'algorithme de Gibbs, obtenir la distance moyenne de 0 à  $M$ , ainsi que la probabilité que la distance OM n'excède pas une valeur  $a$  donnée.

### 7.3 Le modèle 3: estimation bayésienne de la taille d'une population animale.

Exercice. Chercher les lois conditionnelles de  $N|p,q,\mathbf{y}$ , de  $q|p,N,\mathbf{y}$ , et de  $p|q,N,\mathbf{y}$ . Puis décrire une étape ( $l$ ) de l'algorithme de Gibbs. Indiquer comment, à l'aide de cet algorithme, obtenir l'estimation bayésienne de  $N$ . Tester l'hypothèse  $H_0 : N > a$  où  $a$  est un entier donné.

## 8 Une procédure bayésienne de choix de modèles

L'ensemble des observations  $(x_1, \dots, x_i, \dots, x_n)$  est toujours noté  $\mathbf{x}$ . On part de  $J$  modèles (emboîtés ou non) notés  $M_1, \dots, M_j, \dots, M_J$ , chacun étant susceptible d'avoir généré les observations. Le problème est de proposer une règle de décision qui permette de choisir un modèle parmi tous ces modèles en compétition; ce choix devant se faire sur la base des  $x_i$ . On suppose que le modèle  $M_j$  est paramétré par  $\theta_j$ . La densité de  $\mathbf{x}$  sous le modèle  $M_j$  est notée  $f_j(\mathbf{x}|\theta_j)$ . La probabilité a priori du modèle  $M_j$  est notée  $p_j$  et la densité de la loi a priori de  $\theta_j$  est notée  $\pi_j(\theta_j)$ . La procédure bayésienne de choix de modèles repose de façon naturelle sur les probabilités a posteriori de chacun des modèles. La règle de décision consiste simplement à retenir le modèle dont la probabilité a posteriori est maximale. La probabilité a posteriori du modèle  $M_k$  où  $k \in \{1, \dots, J\}$  se déduisant de la formule de Bayes:

$$P(M_k|\mathbf{x}) = \frac{p_k \int_{\Theta_k} f_k(\mathbf{x}|\theta_k) \pi_k(\theta_k) d\theta_k}{\sum_{j=1}^J p_j \int_{\Theta_j} f_j(\mathbf{x}|\theta_j) \pi_j(\theta_j) d\theta_j}$$

Exercice. Reprendre le TD No 2 et considérer les trois modèles suivants. Le modèle  $M_1$  est caractérisé par le fait que les  $X_{i,t}$  sont indépendants; le modèle  $M_2$ , par le fait que les  $X_{i,t}$  sont Markoviens; et enfin le modèle  $M_3$ , par le fait que les  $X_{i,t}$  sont Markoviens et que les probabilités de transition dépendent du temps. Quel modèle choisissez-vous pour les données de suivis proposées dans le TD No 2.