

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche
Scientifique

Université A.Mira de Béjaïa
Faculté des Sciences Exactes
Département de Mathématiques



Mémoire

En vue de l'obtention du Diplôme de

Master en Mathématiques

Option : Probabilités Statistique et Applications

Thème

Méthodes de MCMC, applications bayésiennes

Présenté par :

BOUDJETTOU Céline

Soutenu, le 24 juin 2023,
devant le jury composé de :

| | | | | |
|-----------------|-------------|--------------|------------|--------------------|
| M ^{me} | K. LAGHA | Présidente | Professeur | U. A. Mira. Béjaïa |
| M ^r | Y. BOUMZAID | Promoteur | M.C.B | U. A. Mira. Béjaïa |
| M ^{me} | L. BOURAINE | Examinatrice | Professeur | U. A. Mira. Béjaïa |

Promotion : 2022/2023.

Remerciements

Je tiens à exprimer ma profonde gratitude à Dieu pour m'avoir accordé la persévérance, la motivation et la force nécessaires pour mener à bien ce travail.

Je tiens à adresser mes remerciements les plus sincères à mon encadreur *M^r Y. BOUMZAID* pour son soutien, ses conseils éclairés précieux et son expertise tout au long de ce projet, je le remercie également pour sa disponibilité, ses encouragements ont été essentiels pour m'aider à surmonter les défis et à atteindre mes objectifs. Je lui en suis infiniment reconnaissante.

Mes remerciements vont aussi aux membres du jury qui ont accepté d'évaluer et d'examiner mon travail, pour leur temps et leur expertise.

Enfin, j'aimerais adresser mes remerciements à tous mes enseignants qui nous ont accompagné durant notre cursus universitaire, à tous ceux qui ont contribué, de près ou de loin, à la réalisation de ce mémoire.

※ *Dédicaces* ※

Je dédie ce modeste travail à ceux qui sont toujours présents dans mon coeur.

À mes très chers parents, pour leurs soutiens, leurs encouragements ont été des piliers essentiels de ma réussite tout au long de ces années d'études.

À mon cher frère SIF EDDINE.

À mes chères soeurs MANISSA et KAISSA.

À mes chers grand parents paternels et maternels.

À tous mes amis, à tous ceux que j'aime, et à toutes les personnes qui m'ont prodigué des encouragements et qui se sont donné la peine de me soutenir durant cette formation.

Table des matières

| | |
|--|----------|
| Notations | iii |
| Liste des tableaux | v |
| Table des figures | vi |
| Introduction générale | 1 |
| 1 Statistique bayésienne | 3 |
| 1.1 Rappels sur les notions de probabilité | 3 |
| 1.1.1 Espace de probabilité | 3 |
| 1.1.2 Probabilités conditionnelles | 4 |
| 1.2 Principe de la démarche bayésienne | 5 |
| 1.3 Notions de base de la statistique bayésienne | 5 |
| 1.3.1 Lois des observations | 5 |
| 1.3.2 Information a priori | 5 |
| 1.3.3 Théorème de Bayes | 5 |
| 1.3.4 Loi a posteriori | 6 |
| 1.3.5 Raisonnement proportionnel | 7 |
| 1.4 Choix de la loi a priori | 8 |
| 1.4.1 Loi conjuguée naturelles | 8 |
| 1.4.2 Modèle exponentiel | 9 |
| 1.4.3 Lois a priori subjectives | 10 |
| 1.4.4 Lois a priori impropres | 11 |
| 1.4.5 Lois a priori non-informatives | 11 |
| 1.4.6 Modèle bayésien hiérarchique | 13 |
| 1.5 La théorie de la décision | 14 |
| 1.5.1 La fonction coût | 14 |
| 1.5.2 Risque fréquentiste | 15 |
| 1.5.3 Risque a posteriori | 15 |
| 1.5.4 Risque intégré | 16 |
| 1.5.5 Risque de Bayes | 16 |
| 1.5.6 Risque mini-max | 16 |
| 1.5.7 Admissibilité | 17 |
| 1.6 L'estimation bayésienne | 18 |
| 1.6.1 Cas uni-dimensionnel | 18 |
| 1.6.2 Cas multi-dimensionnel | 19 |
| 1.7 Tests et régions de confiance bayésien | 19 |
| 1.7.1 Régions de confiance | 19 |

| | | |
|----------|--|------------|
| 1.7.2 | Tests bayésien | 20 |
| 1.8 | Incorporation de l'information a priori en pratique | 22 |
| 1.9 | Conclusion | 26 |
| 2 | Méthodes de Monte-Carlo | 27 |
| 2.1 | Intégration Monte-Carlo | 27 |
| 2.1.1 | Le paradigme bayésien | 28 |
| 2.1.2 | Erreur de l'estimation et l'intervalle de confiance | 31 |
| 2.2 | Méthodes de Monte-Carlo par chaîne de Markov | 33 |
| 2.2.1 | Généralités sur les chaînes de Markov | 33 |
| 2.2.2 | Propriétés fondamentales | 34 |
| 2.2.3 | Classification des chaînes de Markov | 34 |
| 2.2.4 | Lois des grands nombres (convergence et vitesse de convergence) | 36 |
| 2.2.5 | Méthodes de Monte-Carlo par chaîne de Markove (MCMC) | 37 |
| 2.2.6 | Les performances des méthodes de MCMC | 44 |
| 2.3 | Conclusion | 48 |
| 3 | Mise en pratique des statistiques bayésiennes et des méthodes de MCMC | 50 |
| 3.1 | Régression bayésienne | 51 |
| 3.1.1 | Définitions et domaines d'applications | 51 |
| 3.1.2 | Approche méthodologique | 52 |
| 3.1.3 | Exemple 1 : Régression simple | 52 |
| 3.1.4 | Exemple 2 : Régression multiple | 65 |
| 3.2 | Estimation de la taille d'une population | 72 |
| 3.2.1 | Définitions et domaines d'applications | 72 |
| 3.2.2 | Approche méthodologique | 72 |
| 3.2.3 | Estimation de la taille d'une population de poissons dans un lac | 73 |
| 3.2.4 | Estimation du poids moyen d'une population de poissons dans un lac | 86 |
| 3.3 | Conclusion | 99 |
| | Conclusion générale | 101 |
| A | Annexe | 103 |
| A.1 | Simulation de variables aléatoires | 103 |
| A.1.1 | Générateur de nombres uniformes | 103 |
| A.1.2 | Générateur de nombres non-uniformes | 104 |
| B | Annexe | 108 |
| B.1 | Régression bayésienne | 108 |
| B.1.1 | Méthode des moindres carrées | 108 |
| B.1.2 | La forme alternative pour moindre carrée | 109 |
| B.1.3 | Estimation de variance | 109 |
| B.1.4 | Fonction de vraisemblance de $\alpha_{\bar{x}}$ et β | 109 |
| B.1.5 | La loi a priori conjointe | 111 |
| B.1.6 | La loi a posteriori conjointe | 111 |
| | Bibliographie | 113 |

Notations

| | |
|-------------------------------------|---|
| Ω | Ensembles des résultats Possible. |
| $(\Omega, \mathcal{A}, \mathbb{P})$ | Espace de probabilités. |
| X | Variable aléatoire. |
| x | Réalisation de la variable aléatoire X . |
| \mathcal{X} | Domaine de la variable aléatoire X . |
| Y | Variable aléatoire. |
| y | Réalisation de la variable aléatoire Y . |
| U | Variable aléatoire distribuée selon $\mathcal{U}([0, 1])$. |
| u | Réalisation de la variable aléatoire U . |
| Θ | Ensemble des paramètres. |
| θ | Paramètre (discret ou continu). |
| F | Distribution cible de la densité f . |
| f | Densité de probabilité. |
| $\pi(\theta)$ | Loi a priori. |
| $\pi(\theta x)$ | Loi a posteriori. |
| $L(x \theta)$ | Loi des observations (vraisemblance). |
| \mathbb{E} | Espérance mathématique. |
| $\mathcal{U}([0, 1])$ | Distribution uniforme définie sur l'intervalle unité. |
| $\mathcal{U}([a, b])$ | Distribution uniforme définie sur l'intervalle $[a, b]$. |
| $\mathcal{N}(\mu, \sigma^2)$ | Distribution gaussienne de moyenne μ et variance σ^2 . |
| $\mathcal{N}(0, 1)$ | Distribution gaussienne centrée réduite. |
| $\text{Exp}(\lambda)$ | Distribution exponentielle de paramètre λ . |
| $\text{Be}(\alpha, \beta)$ | Distribution Beta de paramètres α et β . |
| $\mathcal{G}(a, b)$ | Distribution Gamma de paramètres a et b . |
| $\mathcal{L}(\cdot)$ | Loi d'une variable aléatoire. |
| $\mathbb{1}_A$ | Fonction indicatrice qui vaut 1 sur l'ensemble A et 0 ailleurs. |
| n | Nombre total des itérations. |
| i | Indice des itérations d'un algorithme. |
| i, j | Indices généraux. |

| | |
|----------------|--|
| x, y | Etats d'une chaîne de Markov. |
| S | Espace des états d'une chaîne de Markov. |
| Q | Noyau de transition de la loi de proposition q . |
| \mathbb{R} | Ensemble des nombres réels. |
| \mathbb{N} | Ensemble des entiers naturels. |
| \mathbb{N}^* | Ensemble des entiers naturels non nuls. |
| \mathbb{R}^d | Ensemble des nombres réels de dimension d . |
| \log | Logarithme népérien. |
| \sim | Suivre la loi, de loi.... |
| \propto | Proportionnel à. |
| v.a | Variable aléatoire. |
| i.i.d | Indépendante identiquement distribuée. |
| p.s | Presque sûrement. |
| TCL | Théorème centrale limite. |
| ssi | Si et seulement si. |
| I.C | Intervalle de confiance. |
| Var | Variance. |
| MC | Monte-Carlo. |
| MCMC | Monte-Carlo Chaîne de Markov. |

Liste des tableaux

| | | |
|------|--|----|
| 1.1 | Lois a priori conjuguées naturelles usuelles. | 9 |
| 1.2 | Interprétation du facteur de Bayes. | 21 |
| 3.1 | Le taux de décès par attaque cardiaque chez les hommes de 55 à 59 ans et les calories provenant de protéines animales. | 53 |
| 3.2 | Les taux de décès estimés (\hat{Y}) par l'application de la méthode analytique. | 55 |
| 3.3 | Les taux de décès estimés (\hat{Y}) par l'application de méthode de Monte-Carlo. | 58 |
| 3.4 | Les taux de décès estimés (\hat{Y}) par l'application de l'algorithme Metropolis-Hastings. | 62 |
| 3.5 | Les taux de décès estimés (\hat{Y}) par l'application de l'algorithme de Gibbs | 63 |
| 3.6 | Les taux de décès réels (Y) et estimés (\hat{Y}) par l'application des 4 méthodes. | 64 |
| 3.7 | Le taux de décès par attaque cardiaque chez les hommes de 55 à 59 ans et les variable X_1, X_2, X_3 | 65 |
| 3.8 | Les taux de décès estimés (\hat{Y}) par l'application de méthode MCMC. | 67 |
| 3.9 | Les taux de décès estimés (\hat{Y}) par l'application de méthode MCMC. | 69 |
| 3.10 | Les taux de décès estimés (\hat{Y}) par l'application de la méthode MCMC. | 70 |
| 3.11 | Les données observés dans le cas $n \theta \sim \mathcal{B}(n, \frac{\theta}{\mu})$ et $\theta \sim \mathcal{N}(\mu, \sigma^2)$ | 76 |
| 3.12 | La taille de population de poissons estimée par rapport au nombre de poisson observé. | 78 |
| 3.13 | Les données observés dans le cas $n \theta \sim \mathcal{B}(n, \frac{\theta}{N})$ et $\theta \sim \mathcal{U}([a, b])$ | 79 |
| 3.14 | Les données observés dans le cas $D \theta \sim \mathcal{P}(\theta)$ et $\theta \sim \mathcal{N}(\mu, \sigma^2)$ | 82 |
| 3.15 | Les données observés dans le cas $D \theta \sim \mathcal{P}(\theta)$ et $\theta \sim \mathcal{U}(a, b)$ | 85 |
| 3.16 | Les données observées et les paramètres de la loi a priori, scénario 1. | 88 |
| 3.17 | Les paramètres de la loi a priori et les données observées dans le scénario 2. | 94 |
| 3.18 | Poids moyen estimé des poissons et les intervalles de crédibilité dans les trois zones : profonde, peu profonde et dans le lac. | 95 |
| 3.19 | Poids moyen estimés des poissons et les intervalle de crédibilité dans la région profonde, peu profonde et dans le lac. | 96 |
| 3.20 | Les poids observés des poissons (D), de profondeur de l'eau en m ($prof$) et de quantité de nourriture disponible par m^3 ($nour$) dans le scénario 2. | 98 |

Table des figures

| | | |
|------|---|----|
| 1.1 | Thomas Bayes. | 6 |
| 1.2 | La représentation graphique de la distribution $\mathcal{B}e(3.4932, 14.0072)$ | 23 |
| 1.3 | La représentation graphique des distributions Beta. | 24 |
| 1.4 | La représentation graphique des distributions Gama. La courbe en bleu pour $\theta \in [0, 5]$ et la courbe rouge pour $\theta \in [1, 10]$ | 25 |
| 1.5 | La représentation graphique des distributions $\mathcal{N}(2.5505, 0.2344)$ et $\mathcal{N}(3.3457, 0.2867)$. 26 | |
| 2.1 | L'estimation de $\hat{\theta}_n$. La courbe en rouge pour $N = 5$, la courbe en vert pour $N = 10$ et la courbe en bleu pour $N = 50$ | 31 |
| 2.2 | Distribution de la loi de mélange $f(x)$ | 45 |
| 2.3 | Simulations de 10000 itérations avec $q(y x) \sim \mathcal{N}(x, 5)$, $q(y x) \sim \mathcal{N}(x, 1)$ et $q(y x) \sim \mathcal{U}([x - 10, x + 1])$ respectivement (de gauche à droite). | 46 |
| 2.4 | Simulations de 10000 itérations avec respectivement $X_1 = -20$, $X_1 = 0$ et $X_1 = 20$. 47 | |
| 2.5 | Distribution et histogramme ainsi la série chronologique de trois simulations avec respectivement $n_0 = 0$, $n_0 = 500$ et $n_0 = 1000$ | 48 |
| 3.1 | Nuage de point $(x_i, y_i), i = 1, \dots, n$ et la droite de régression par l'application de la méthodes analytique. | 55 |
| 3.2 | La représentation graphique de la distribution a priori et a posteriori de la pente. 56 | |
| 3.3 | La représentation graphique de l'histogramme des échantillons du coefficient du régression β | 57 |
| 3.4 | La représentation graphique de la densité des échantillons du coefficient du régression β | 57 |
| 3.5 | Nuage de point $(x_i, y_i), i = 1, \dots, n$ et la droite de régression obtenue par la méthode MC. | 59 |
| 3.6 | La représentation graphique de le trajectoire de la chaîne de Markov des coefficients de régression α et β par l'application de l'algorithme Metropolis-Hastings. 60 | |
| 3.7 | Trajectoire des coefficients de régression | 61 |
| 3.8 | La représentation graphique de la densité des coefficients α et β | 61 |
| 3.9 | L'histogramme des coefficients de régression α et β | 62 |
| 3.10 | Nuage de point $(x_i, y_i), i = 1, \dots, n$ et la droite de régression obtenue par l'application l'algorithme de Metropolis-Hastings. | 63 |
| 3.11 | Nuage de point $(x_i, y_i), i = 1, \dots, n$ et la droite de régression par l'application de l'algorithme de Gibbs. | 64 |
| 3.12 | L'histogramme des coefficients de régression $\alpha, \beta_1, \beta_2, \beta_3$ par l'application de méthode MCMC | 66 |
| 3.13 | Nuage de point et le trajectoire de la régression par l'application de méthode MCMC | 67 |

| | | |
|------|---|----|
| 3.14 | Trajectoire des coefficient de régression β_1 et β_2 par l'application de méthode MCMC. | 68 |
| 3.15 | Nuage de point et le trajectoire de la régression | 69 |
| 3.16 | Trajectoire et l'histogramme des coefficient de régression β_2 et β_3 par l'application de méthode MCMC. | 70 |
| 3.17 | Nuage de point et le trajectoire de régression. | 71 |
| 3.18 | Organigramme de calcule de la distribution a posteriori avec $n \theta \sim \mathcal{B}(n, \frac{\theta}{\mu})$ et $\theta \sim \mathcal{N}(\mu, \sigma^2)$ | 76 |
| 3.19 | À gauche l'histogramme de la distribution a posteriori et à droite distribution de la taille de population des poissons estimée, pour $n \theta \sim \mathcal{B}(n, \frac{\theta}{\mu})$ et $\theta \sim \mathcal{N}(\mu, \sigma^2)$ | 77 |
| 3.20 | L'évolution de la population des poissons estimée, pour $n \theta \sim \mathcal{B}(n, \frac{\theta}{\mu})$ et $\theta \sim \mathcal{N}(\mu, \sigma^2)$ | 77 |
| 3.21 | Organigramme de calcule de la distribution a posteriori avec $n \theta \sim \mathcal{B}(n, \frac{\theta}{N})$ et $\theta \sim \mathcal{U}([a, b])$ | 79 |
| 3.22 | À gauche l'histogramme de la distribution a posteriori, à droite la densité de l'estimation de taille de population des poissons, pour $n \theta \sim \mathcal{B}(n, \frac{\theta}{N})$ et $\theta \sim \mathcal{U}([a, b])$ | 80 |
| 3.23 | L'évolution de la population des poissons estimée, pour $n \theta \sim \mathcal{B}(n, \frac{\theta}{N})$ et $\pi(\theta) \sim \mathcal{U}([a, b])$ | 80 |
| 3.24 | Organigramme de calcule de la distribution a posteriori pour $D \theta \sim \mathcal{P}(\theta)$ et $\theta \sim \mathcal{N}(\mu, \sigma^2)$ | 82 |
| 3.25 | À gauche l'histogramme de la distribution a posteriori, à droite la densité de l'estimation de la taille de population des poissons, pour $D \theta \sim \mathcal{P}(\theta)$ et $\theta \sim \mathcal{N}(\mu, \sigma^2)$ | 83 |
| 3.26 | L'évolution de la population des poissons estimée, pour $D \theta \sim \mathcal{P}(\theta)$ et $\pi(\theta) \sim \mathcal{N}(\mu, \sigma^2)$ | 83 |
| 3.27 | Organigramme de calcule de la distribution a posteriori avec $D \theta \sim \mathcal{P}(\lambda)$ et $\theta \sim \mathcal{U}(a, b)$ | 84 |
| 3.28 | À gauche l'histogramme de la distribution a posteriori, à droite la densité de l'estimation de taille de population des poissons, pour $D \theta \sim \mathcal{P}(\theta)$ et $\theta \sim \mathcal{U}([a, b])$ | 85 |
| 3.29 | L'évolution de la population des poissons estimés, pour $D \theta \sim \mathcal{P}(\theta)$ et $\theta \sim \mathcal{U}([a, b])$ | 86 |
| 3.30 | Organigramme de calcule de la distribution a posteriorin, scénario 1 | 88 |
| 3.31 | À gauche l'histogramme de la distribution a posteriori, à droite la densité de l'estimation de poids moyen des poissons, pour $Y \theta \sim \mathcal{N}(\theta, \sigma^2)$ et $\theta \sim \mathcal{N}(\mu, \sigma^2)$, en utilisant l'algorithme de Metropolis-Hastings. | 89 |
| 3.32 | L'évolution de poids moyen estimé des poissons, pour $Y \theta \sim \mathcal{N}(\theta, \sigma^2)$ et $\theta \sim \mathcal{N}(\mu, \sigma^2)$, en utilisant l'algorithme de Metropolis-Hastings. | 90 |
| 3.33 | À gauche l'histogramme de la distribution a posteriori, à droite la densité de l'estimation de poids moyen des poissons, pour $Y \theta \sim \mathcal{N}(\theta, \sigma^2)$ et $\theta \sim \mathcal{N}(\mu, \sigma^2)$, en utilisant l'algorithme de Gibbs. | 91 |
| 3.34 | L'évolution de poids moyen estimés des poissons, pour $Y \theta \sim \mathcal{N}(\theta, \sigma^2)$ et $\theta \sim \mathcal{N}(\mu, \sigma^2)$, en utilisant l'algorithme de Gibbs. | 92 |
| 3.35 | Organigramme de calcule de la distribution a posteriori dans scénario 2 | 94 |
| 3.36 | L'histogramme de la distribution a posteriori, la densité de l'estimation de poids moyen des poissons et l'évolution de poids moyen estimé des poissons. À gauche dans la zone profonde, au milieu dans le Lac et à droite dans la zone peu profonde. | 95 |

| | | |
|------|---|----|
| 3.37 | L’histogramme de la distribution a posteriori, la distribution a priori et l’évolution de poids moyen estimé des poissons. À gauche dans la zone profonde, au milieu dans le Lac et à droite dans la zone peu profonde. | 96 |
| 3.38 | Organigramme de calcul de la distribution a posteriori dans scénario 3 | 98 |
| 3.39 | L’évolution de poids moyen des poissons par rapport à l’évolution des échantillons MCMC pour la profondeur et la nourriture. | 99 |

Introduction générale

La statistique bayésienne offre un cadre formel séduisant. En effet cette dernière peut-être vue comme la rationalisation ultime de la statistique classique. Aujourd’hui, la statistique bayésienne est largement utilisée dans de nombreux domaines, y compris les sciences, l’économie, la biologie, l’ingénierie, l’informatique et bien d’autres. Elle offre une approche alternative à la statistique fréquentiste traditionnelle. La statistique bayésienne permet de construire un solide pont entre la théorie et les données expérimentales, offrant ainsi une base solide pour les travaux pratiques.

La statistique bayésienne a une histoire riche qui remonte au XVIIIe siècle avec les travaux de Thomas Bayes, mais c’est au cours du XXe siècle qu’elle a été développée en tant que discipline statistique à part entière, avec des contributions significatives de nombreux chercheurs, et elle continue d’évoluer et de trouver des nouvelles applications dans de nombreux domaines.

La statistique bayésienne repose sur le même cadre de pensée que la statistique fréquentiste, mais avec une différence importante, le paramètre θ n’est plus considéré comme totalement inconnu, mais plutôt comme une variable aléatoire à laquelle on associe une loi appelée ”loi a priori”. Pour construire une loi appelée ”loi a posteriori”, on utilise un célèbre théorème connu sous le nom de ”théorème de Bayes” [1].

L’inférence bayésienne, basée sur le théorème de Bayes en l’honneur du mathématicien britannique Thomas Bayes (1702 – 1761), est une approche statistique qui permet de combiner des informations a priori avec des données observées afin d’obtenir des estimations probabilistes des paramètres inconnus d’un modèle. L’une des principales difficultés en inférence bayésienne réside dans l’évaluation des distributions a posteriori, qui représentent les connaissances mises à jour après l’observation des données. Dans de nombreux cas, ces distributions ne peuvent pas être calculées analytiquement, et c’est là que les méthodes de Monte-Carlo par chaîne de Markov entrent en jeu.

Les méthodes de Monte-Carlo par chaînes de Markov, également connues sous le nom de méthodes MCMC (Monte Carlo Markov Chain), sont de puissants outils statistiques utilisés dans les applications bayésiennes. Elles permettent d’explorer de manière efficace et pratique les distributions de probabilité complexes et multidimensionnelles inhérentes aux problèmes d’inférence bayésienne.

Les méthodes MCMC offrent une approche numérique pour générer des échantillons approximatifs à partir des distributions a posteriori. L’algorithme Metropolis-Hastings, introduit par Metropolis et al 1953 [16] et généralisé par Hastings 1970 [12], est considéré comme l’une des premières méthodes MCMC. L’idée centrale de cette dernière est de construire une chaîne de Markov dont les états convergent vers la distribution cible. Ces chaînes sont construites en utilisant des règles de transition basées sur des probabilités conditionnelles. En itérant cette

procédure, on peut obtenir une séquence de valeurs qui constitue un échantillon représentatif de la distribution cible.

L'avantage des méthodes MCMC réside dans leurs capacités à gérer des modèles complexes avec de grande dimensions et des relations non linéaires entre les variables. Elles sont particulièrement utiles lorsque les distributions a posteriori ne peuvent pas être calculées directement ou lorsqu'il est difficile de générer des échantillons indépendants à partir de ces distributions.

L'objectif de ce mémoire est d'étudier et d'appliquer les méthodes de Monte-Carlo et de Monte-Carlo par chaîne de Markov (MCMC) aux problèmes concrets de la statistique bayésienne. Pour organiser notre étude de manière cohérente, nous avons structuré notre travail en trois chapitres.

Le premier chapitre introduit les concepts fondamentaux de la statistique bayésienne, tels que la loi a posteriori et les différentes approches pour construire une loi a priori, en tenant compte de la disponibilité ou de l'absence d'informations a priori.

Le deuxième chapitre se concentre sur les méthodes de Monte-Carlo. Nous présentons d'abord les techniques classiques d'intégration Monte-Carlo utilisées pour estimer des quantités spécifiques, puis nous abordons les méthodes de Monte-Carlo par chaîne de Markov, qui sont particulièrement adaptées à la statistique bayésienne.

Le troisième chapitre est consacré à l'application des algorithmes MCMC dans le contexte bayésien. Nous illustrons leur utilisation à travers des exemples théoriques et des applications concrètes sur des données réelles, démontrant ainsi leur pertinence et leur efficacité.

En conclusion, nous récapitulons de manière synthétique les principaux points abordés tout au long de cette étude dans une conclusion générale.

Chapitre 1

Statistique bayésienne

La statistique bayésienne peut-être vue comme une généralisation de la statistique classique, son principe est de modéliser des paramètres inconnus de la distribution d'échantillonnage à travers une structure probabiliste. Son objectif principale est d'inférer sur la loi qui génère les observations afin de comprendre des phénomènes passés ou de prévoir des événements futur. Dans ce chapitre, nous introduisons les principes de base de la statistique bayésienne.

1.1 Rappels sur les notions de probabilité

1.1.1 Espace de probabilité

Dans cette sous-section, nous donnons quelques définitions qui portent sur les espaces de probabilité.

Espace des épreuves

Définition 1. *L'ensemble de tous les résultats possibles d'une expérience aléatoire est appelé l'espace des épreuves, et noté par Ω .*

Exemple 1. *Prenons un exemple simple d'une expérience aléatoire qui consiste à lancer un dé à six faces numérotées de 1 à 6. Dans ce cas, l'espace des épreuves est,*

$$\Omega = \{1, 2, 3, 4, 5, 6\}.$$

Espace des événements

Définition 2. *L'espace des événements est l'ensemble de tous les sous-ensembles possibles de l'espace des épreuves.*

Exemple 2. *Prenons l'exemple d'une expérience aléatoire, qui consiste à lancer une pièce de monnaie. L'espace des épreuves est $\Omega = \{pile, face\}$. L'espace des événements est alors l'ensemble de tous les sous-ensembles de Ω : $\{\{pile\}, \{face\}, \{pile, face\}, \emptyset\}$.*

Tribus

Une tribu peut-être vue comme un sous-ensemble des espaces des événements, qui contient uniquement certains événements satisfaisant quelques propriétés spécifiques.

Définition 3. Soit Ω un ensemble quelconque. Une famille \mathcal{A} d'éléments $\mathcal{P}(\mathcal{A})$ (l'ensemble de toutes les parties de l'ensemble Ω) est dite tribu si elle vérifie les conditions suivantes :

1. $\Omega \in \mathcal{A}$.
2. \mathcal{A} est stable par passage au complémentaire, i.e. $\forall A \in \mathcal{A}, A^c \in \mathcal{A}$.
3. \mathcal{A} est stable par réunion dénombrable, i.e. si $A_1, A_2, \dots, A_n, \dots$ une suite finie au infinie d'éléments de \mathcal{A} , alors $\cup_{n \geq 1} A_n$ est un élément de \mathcal{A} .

Remarque 1. Lorsque \mathcal{A} est une tribu sur Ω , le couple (Ω, \mathcal{A}) est appelé espace probabilisable. Une tribu est souvent utilisée pour définir la notion de mesure de probabilité.

Notion de probabilité

Définition 4. Soit (Ω, \mathcal{A}) un espace probabilisable (mesurable) et A un événement de Ω . On appelle probabilité, notée \mathbb{P} , sur (Ω, \mathcal{A}) , toutes application de \mathcal{A} , telle que,

$$\begin{aligned} \mathbb{P} : \mathcal{A} &\rightarrow [0, 1], \\ A &\mapsto \mathbb{P}(A), \end{aligned}$$

vérifiant :

1. $\mathbb{P}(\Omega) = 1$,
2. pour toute suite $(A_n)_{n \geq 1}$ d'éléments de \mathcal{A} deux à deux disjoints (incompatibles),

$$\mathbb{P}(\cup_{n \in \mathbb{N}} A_n) = \sum_{n=1}^{+\infty} \mathbb{P}(A_n).$$

Remarque 2. Si \mathbb{P} est une probabilité définie sur un espace probabilisable (Ω, \mathcal{A}) , alors le triplet $(\Omega, \mathcal{A}, \mathbb{P})$ est appelé espace de probabilité.

1.1.2 Probabilités conditionnelles

Définition 5. Soient $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité et B un événement de Ω , tel que $\mathbb{P}(B) > 0$. On appelle probabilité de A conditionnellement à B , ou sachant B , la probabilité notée $\mathbb{P}(A|B)$, définie par :

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}. \quad (1.1)$$

Remarque 3. Si A et B sont deux événements indépendants, alors :

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \frac{\mathbb{P}(A) \times \mathbb{P}(B)}{\mathbb{P}(B)} = \mathbb{P}(A). \quad (1.2)$$

1.2 Principe de la démarche bayésienne

Considérons un modèle statistique $(\Omega, \mathcal{A}, \mathbb{P})$ le but de l'analyse bayésienne paramétrique est de faire de l'inférence sur $f(\theta)$ avec f une fonction définie sur Θ . Le principe de la démarche bayésienne consiste à combiner la connaissance a priori et les données observées pour obtenir une estimation de paramètre sous-forme d'une distribution de probabilité a posteriori $\pi(\theta|x)$.

1.3 Notions de base de la statistique bayésienne

Considérons un échantillon de taille n , l'ensemble des observations noté par $x=(x_1, x_2, \dots, x_n)$. Les observations $(x_n)_{n \geq 1}$ sont donc considérées comme des réalisations de variables aléatoires notées $(X_n)_{n \geq 1}$.

1.3.1 Lois des observations

Définition 6. *On interprète la loi des observations (la fonction de vraisemblance) comme la loi conditionnelle des observations $x = (x_1, x_2, \dots, x_n)$ sachant θ , de densité notée $L(x|\theta)$ définie par :*

$$L(x|\theta) = \begin{cases} f_\theta(x) = f(x|\theta) & = \prod_{i=1}^n f(x_i, \theta) & \text{cas continu,} \\ \mathbb{P}_\theta(x) = \mathbb{P}(X = x|\theta) & = \prod_{i=1}^n \mathbb{P}(X_i = x_i, \theta) & \text{cas discret.} \end{cases}$$

1.3.2 Information a priori

Définition 7. *L'information a priori est toute information disponible sur le paramètre θ , en dehors de celle apportée par les observations. L'information a priori sur θ est entachée d'incertitude. Cette information est modélisée à travers une loi de probabilité appelée loi a priori. Sa densité est notée par $\pi(\theta)$.*

1.3.3 Théorème de Bayes

Soient $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité, et soit la famille d'événements A_1, A_2, \dots, A_n qui forme un système complet d'événements pour Ω tel que $\mathbb{P}(A_i) > 0$ pour tout $i \in \{1, 2, \dots, n\}$. Si B un événement ($B \in \mathcal{A}$) tel que $\mathbb{P}(B) > 0$, alors pour tout $j \in \{1, 2, \dots, n\}$,

$$\mathbb{P}(A_j|B) = \frac{\mathbb{P}(A_j)\mathbb{P}(B|A_j)}{\sum_{i=1}^n \mathbb{P}(A_i)\mathbb{P}(B|A_i)}. \quad (1.3)$$

L'équation (1.3), est appelée formule de Bayes (ou Théorème de Bayes [1]) en l'honneur du mathématicien britannique Thomas Bayes (1702-1761, voir figure (1.1)), qui a développé cette formule. La formule de Bayes est un mécanisme d'apprentissage qui combine l'information a priori sur les paramètres et l'information apportée par les données en une information a posteriori sur les paramètres. Précisément la formule de Bayes s'applique à la densité conjointe a priori des paramètres et renvoie la densité conjointe des paramètres a posteriori c'est-à-dire conditionnellement aux observations.



FIGURE 1.1 – Thomas Bayes.

1.3.4 Loi a posteriori

Définition 8. La loi a posteriori est la loi conditionnelle de θ sachant les observation $x=(x_1, x_2, \dots, x_n)$, sa densité est notée par $\pi(\theta|x)$. En vertu de la formule de Bayes (1.3), cette dernière est définie comme suit :

$$\pi(\theta|x) = \frac{L(x|\theta)\pi(\theta)}{\int_{\Theta} L(x|\theta)\pi(\theta)d\theta} = \frac{\varphi(\theta, x)}{f(x)}. \quad (1.4)$$

Où $\varphi(\theta, x)$ est la distribution conjointe de (x, θ) ,

$$\varphi(\theta, x) = L(x|\theta)\pi(\theta),$$

et $f(x)$ la distribution marginale de x ,

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

Exemple 3. supposons, que l'on dispose d'un n -échantillon i.i.d, $x=(X_1, X_2, \dots, X_n)$, de loi Bernoulli $\mathcal{B}(\theta)$, où $\theta > 0$ est un paramètre inconnu de loi a priori de type Beta, $\theta \sim \mathcal{Be}(a, b)$, $a > 0, b > 0$.

La vraisemblance est donnée par :

$$L(x|\theta) = \prod_{i=1}^n \mathbb{P}(X = x_i|\theta) = \theta^{\sum_{i=1}^n x_i} (1 - \theta)^{n - \sum_{i=1}^n x_i} \mathbb{1}_{\{0,1\}}(x).$$

La densité de la loi a priori est donnée par :

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

Donc la densité de la loi a posteriori est donnée par :

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

Il est facile de vérifier que :

$$\int_{\Theta} L(x|\theta)\pi(\theta)d\theta = \mathcal{B}e(\alpha, \beta).$$

$$\text{Avec : } \alpha = a + \sum_{i=1}^n x_i \text{ et } \beta = b + n - \sum_{i=1}^n x_i .$$

D'où :

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

$$D'où, \theta|x \sim \mathcal{B}e(a + \sum_{i=1}^n x_i, b + n - \sum_{i=1}^n x_i).$$

1.3.5 Raisonnement proportionnel

Soient f et g deux fonctions réelles définies sur le même espace \mathcal{X} . On dit que f et g sont proportionnelles, et on note $f \propto g$, s'il existe une constante $a > 0$ telle que [7] :

$$f(x) = a g(x), \quad \forall x \in \mathcal{X}.$$

Résultats

1. La relation \propto est une relation d'équivalence. En particulier,

$$\forall f, g \text{ et } h \quad f \propto g \text{ et } g \propto h \Rightarrow f \propto h.$$

2. Soient X et Y deux v.a. et soient f et g les densités de X et Y , respectivement. On suppose que la loi de probabilité de X est inconnue et celle de Y est connue et elle est notée par \mathbb{P} . Alors :

$$f \propto g \Rightarrow X \sim \mathbb{P}.$$

3. La proportionnalité dans le contexte bayésien est donnée par :

$$\pi(\theta|x) = \frac{\varphi(x, \theta)}{f(x)} = \frac{L(x|\theta)\pi(\theta)}{\int_{\Theta} L(x|\theta)\pi(\theta)d\theta} = \frac{1}{a} L(x|\theta)\pi(\theta),$$

où, $a = \frac{1}{f(x)} = \frac{1}{\int_{\Theta} L(x|\theta)\pi(\theta)d\theta}$, est indépendante de θ ,
on a :

$$\pi(\theta|x) \propto L(x|\theta)\pi(\theta).$$

Exemple 4. Soit un n -échantillon $x = (X_1, X_2, \dots, X_n)$ issu d'une v.a X de loi de Poisson $\mathcal{P}(\theta)$ de paramètre inconnu θ , $\theta > 0$. Considérons la loi a priori de θ de type Gamma $\mathcal{G}(\alpha, \beta)$, $\alpha > 0, \beta > 0$.

La loi de x est,

$$\mathbb{P}(X = x|\theta) = \frac{\theta^x}{x!} e^{-\theta}, \quad x \in \mathbb{N}, \theta > 0.$$

La loi a priori de θ est,

$$\pi(\theta) = \frac{\beta^\alpha e^{-\theta b}}{\Gamma(\alpha)} \theta^{\alpha-1}, \quad \theta > 0.$$

Avec $\Gamma(\alpha) = \int_0^{+\infty} x^{\alpha-1} e^{-x} dx$, $\alpha > 0$.

La vraisemblance de l'échantillon est donnée par,

$$L(x|\theta) = \prod_{i=1}^n \left(\frac{\theta^{x_i}}{x_i!} e^{-\theta} \right) = \frac{\theta^{\sum_{i=1}^n x_i} e^{-n\theta}}{\prod_{i=1}^n x_i!} \mathbb{1}_{\mathbb{N}}(x).$$

La loi a posteriori de θ est,

$$\begin{aligned} \pi(\theta|x) &\propto L(x|\theta)\pi(\theta), \\ &\propto \left(\frac{\theta^{\sum_{i=1}^n x_i} e^{-n\theta}}{\prod_{i=1}^n x_i!} \right) \left(\frac{\beta^\alpha e^{-\theta\beta}}{\Gamma(\alpha)} \theta^{\alpha-1} \right), \\ &\propto \theta^{\sum_{i=1}^n x_i + \alpha - 1} e^{-(n+\beta)\theta}. \end{aligned}$$

D'où

$$\theta|x \sim \mathcal{G} \left(\sum_{i=1}^n x_i + \alpha, n + \beta \right).$$

1.4 Choix de la loi a priori

Le choix de la loi a priori est l'aspect le plus critiqué dans l'analyse statistique bayésienne. La loi a priori est un élément clé de l'inférence bayésienne et sa détermination doit-être effectuée avec précaution pour obtenir des résultats fiables et interprétables. Comment passer des informations a priori à une loi a priori? est la question fondamentale et légitime dans la mise en oeuvre de toute analyse au sens bayésien. En pratique, il est rare que l'information a priori soit suffisamment précise pour conduire à une détermination exacte de la loi a priori au sens où plusieurs lois de probabilités soient des lois usuelles (loi normale, loi gamma,...) dites conjuguées. En absence d'information a priori on introduira la notion de loi a priori non informative qui permet de travailler dans un cadre bayésien.

1.4.1 Loi conjuguée naturelles

Une des difficultés de l'approche bayésienne est le calcul de la loi a posteriori. Ce calcul sera facilité lorsque la loi a priori et la loi a posteriori sont de la même famille. Dans ce cas, on parle de lois a prioris conjuguées. L'approche a priori conjuguée, introduite par Raiffa et Schlaifer en (1961) [18], peut-être considérée comme un point de départ pour l'élaboration de distributions a priori fondées sur des informations a priori limitées.

Définition 9. Une classe \mathcal{C} de distribution a priori $\pi(\theta)$ conjuguée naturelle par rapport à une densité $f(x|\theta)$ est un ensemble de distribution $\pi(\theta)$ tels que la loi a posteriori $\pi(\theta|x)$ appartient à \mathcal{C} . La classe \mathcal{C} est dite, alors conjuguée naturelle [18].

Le tableau (1.1), donne quelques exemples des lois a priori conjuguées naturelles usuelles.

| $f(x \theta)$ | $\pi(\theta)$ | $\pi(\theta x)$ |
|--|---------------------------------------|--|
| Normale $\mathcal{N}(\theta, \sigma^2)$ | Normale $\mathcal{N}(\mu, \tau^2)$ | $\mathcal{N}\left(\frac{x}{\sigma^2} + \frac{\mu}{\tau^2}, \left(\frac{1}{\sigma^2} + \frac{1}{\tau^2}\right)^{-1}\right)$ |
| Poisson $\mathcal{P}(\theta)$ | Gamma $\mathcal{G}(\alpha, \beta)$ | $\mathcal{G}(\alpha + x, \beta + 1)$ |
| Gamma $\mathcal{G}(\eta, \theta)$ | Gamma $\mathcal{G}(\alpha, \beta)$ | $\mathcal{G}(\alpha + \eta, \beta + x)$ |
| Binomiale $\mathcal{B}(n, \theta)$ | Beta $\mathcal{B}e(\alpha, \beta)$ | $\mathcal{B}e(\alpha + x, \beta + n - x)$ |
| Binomiale Négative $\mathcal{BN}(\eta, \theta)$ | Beta $\mathcal{B}e(\alpha, \beta)$ | $\mathcal{B}e(\alpha + \eta, \beta + x)$ |
| Normale $\mathcal{N}\left(\mu, \frac{1}{\theta}\right)$ | Gamma $\mathcal{G}(\alpha, \beta)$ | $\mathcal{G}\left(\alpha + \frac{1}{2}, \beta + \frac{(\mu-x)^2}{2}\right)$ |

TABLE 1.1 – Lois a priori conjuguées naturelles usuelles.

1.4.2 Modèle exponentiel

Définition 10. On appelle famille exponentielle à s -paramètres, toute famille de loi de distribution \mathbb{P}_θ dont la densité a la forme suivante [19],

$$f(x|\theta) = h(x) \exp\left(\sum_{i=1}^s \eta_i(\theta) T_i(x) - B(\theta)\right). \quad (1.5)$$

Où : $\eta_i(\cdot)$ et $B(\cdot)$ sont des fonctions de paramètre θ , $T_i(\cdot), i = 1, \dots, s$ sont des statistiques et $h(\cdot)$ est une fonction de x .

Exemple 5. Considérons les modèles suivants :

- Loi exponentielle $\mathcal{Exp}\left(\frac{1}{\theta}\right)$:

$$\begin{aligned} f(x|\theta) &= \frac{1}{\theta} \exp\left(-\frac{x}{\theta}\right) \mathbb{1}_{]0, +\infty[}(x), \\ &= \exp\left(-\frac{1}{\theta}x - \log \theta\right) \mathbb{1}_{]0, +\infty[}(x). \end{aligned}$$

Dans ce cas,

$$s = 1, \eta_1 = \frac{1}{\theta}, B(\theta) = \log \theta, T_1 = x, h(x) = \mathbb{1}_{]0, +\infty[}(x).$$

- La loi Binomiale $\mathcal{B}(n, \theta)$, $n \in \mathbb{N}^*, \theta \in \{0, 1\}$,

$$\begin{aligned} \mathbb{P}(X = x|\theta) &= C_n^x \theta^x (1 - \theta)^{n-x} \mathbb{1}_{\{0, \dots, n\}}(x), \\ &= C_n^x \exp(x \log \theta + (n - x) \log(1 - \theta)) \mathbb{1}_{\{0, \dots, n\}}(x), \\ &= C_n^x \exp\left(x \log\left(\frac{\theta}{1 - \theta}\right) + n \log(1 - \theta)\right) \mathbb{1}_{\{0, \dots, n\}}(x). \end{aligned}$$

Alors, dans ce cas,

$$s = 1, \eta_1(\theta) = \log\left(\frac{\theta}{1 - \theta}\right), T_1(x) = x, B(\theta) = n \log(1 - \theta) \text{ et } h(x) = C_n^x \mathbb{1}_{\{0, \dots, n\}}(x).$$

- La loi Gamma $\mathcal{G}(a, b)$:

$$f(x|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp(-\beta x), \quad \alpha, \beta > 0, \quad x > 0.$$

En remarquant que : $x^{\alpha-1} \exp(-\beta x) = \exp(-\beta x + (\alpha-1) \log(x))$, on peut écrire dans ce cas,

$$f(x|\alpha, \beta) = \exp\left(-\beta x + (\alpha-1) \log(x) - \log\left(\frac{\Gamma(\alpha)}{\beta^\alpha}\right)\right) \mathbb{1}_{\{0, \dots, n\}}(x).$$

alors on a :

$s = 2$, $\eta_1(\alpha, \beta) = -\beta$, $\eta_2(\alpha, \beta) = \alpha - 1$, $T_1(x) = x$, $T_2(x) = \log(x)$, $B(\alpha, \beta) = \log\left(\frac{\Gamma(\alpha)}{\beta^\alpha}\right)$ et $h(x) = 1$.

Il est classique d'écrire le modèle exponentiel sous la forme dite canonique en reparamétrisant : $\eta_i(\theta) = \theta_i$.

$$f(x|\theta) = \exp\left(\sum_{i=1}^s \theta_i T_i(x) - B(\theta)\right) h(x).$$

La proposition suivante donne la forme des lois naturelles conjuguées dans le cas du modèle exponentiel.

Proposition 1. [20]

Soit $f(x|\theta)$ appartenant à une famille exponentielle écrite sous la forme canonique. Alors une famille de lois a priori conjuguée pour $f(x|\theta)$ est donnée par,

$$\pi(\theta|\lambda, \mu) = K(\lambda, \mu) \exp(\theta\mu - \lambda B(\theta)). \tag{1.6}$$

Où $K(\lambda, \mu)$ est une constante de normalisation.

Et la loi a posteriori est de la forme :

$$\begin{aligned} \pi(\theta|x) &= \frac{h(x) \exp(\theta x - B(\theta)) K(\mu, \lambda) \exp(\theta\mu - \lambda B(\theta))}{\int_{\Theta} h(x) \exp(\theta x - B(\theta)) K(\mu, \lambda) \exp(\theta\mu - \lambda B(\theta)) d\theta}, \\ &\propto \frac{\exp(\theta x - B(\theta)) \exp(\theta\mu - \lambda B(\theta))}{\int_{\Theta} \exp(\theta x - B(\theta)) \exp(\theta\mu - \lambda B(\theta)) d\theta}, \\ &\propto \exp(\theta x - B(\theta)) \exp(\theta\mu - \lambda B(\theta)), \\ &\propto \exp((\mu + x)\theta - (\lambda + 1)B(\theta)). \end{aligned}$$

Remarque 4. Seuls les modèles à structure exponentielle admettent une famille conjuguée.

1.4.3 Lois a priori subjectives

Précisons tout d'abord que cette démarche n'est pas forcément facile dans la pratique. L'idée est d'utiliser les données antérieures. Dans un cas concret, il peut-être judicieux de baser son raisonnement sur l'expertise de spécialistes. Par exemple, si on fait des biostatistiques, on s'appuiera sur l'expertise des médecins et des biologistes pour déterminer une loi a priori cohérente. Si l'on a plusieurs expertises distinctes, on pourra les pondérer en utilisant un modèle hiérarchique.

1.4.4 Lois a priori impropres

La loi a priori peut-être impropre,

$$\int_{\mathbb{R}} \pi(\theta) d\theta = +\infty.$$

Le choix de type de loi n'a donc plus d'intérêt calculatoire et s'interprète difficilement. Nous verrons par la suite que la construction de lois non informative peut conduire à des lois a priori de ce type.

1.4.5 Lois a priori non-informatives

Dans certaines situations, il peut-être difficile de déterminer une loi a priori pertinente pour l'inférence bayésienne en raison d'un manque d'informations sur les paramètres ou de la difficulté à traduire ces informations en une loi de probabilité. Dans ce cas, les statisticiens peuvent utiliser des lois a priori non-informatives. Les lois a priori non-informatives sont des lois de probabilité qui n'expriment aucune information préalable sur les paramètres.

Exemple 6. *Dans le cas où l'espace des paramètres est discret et fini, $\theta = \{\theta_1, \theta_2, \dots, \theta_n\}$, alors la loi a priori non-informative est :*

$$\pi(\theta_i) = \frac{1}{n}, \quad \forall i = 1, \dots, n.$$

Dans le cas où nous avons un espace du paramètre continue et borné, $\Theta = [a, b]$, $-\infty < a < b < +\infty$, alors la loi a priori est uniforme,

$$\pi(\theta) = \frac{1}{b-a} \mathbb{1}_{a \leq \theta \leq b},$$

donc il s'agit d'une distribution non-informative pour θ .

Nous décrirons, dans ce qui suit, quelques techniques les plus importantes dans la construction de la loi a priori non informative.

Lois a priori invariantes

La loi a priori invariante s'exprime lorsque l'absence de l'information a priori se traduit par une propriété d'invariance par rapport à une famille de transformation de Θ . Le choix de la loi a priori invariante peut-être controversé car il peut conduire à des résultats imprécis ou incohérents dans certaines situations.

Exemple 7. *La famille des lois invariantes par translation, c'est-à-dire, la loi a priori doit vérifier la relation suivante :*

$$\pi(\theta) = \pi(\theta - \theta_0), \quad \theta_0 \in \Theta.$$

Par conséquent, la loi a priori solution de cette équation est la loi uniforme sur θ définie par :

$$\pi(\theta) = c.$$

Où c est une constante et θ est un paramètre de position.

La famille des lois invariantes par changement d'échelle ayant la forme,

$$\pi(\theta) = \frac{1}{c} \pi\left(\frac{\theta}{c}\right).$$

Où c est une constante strictement positive et θ est dit paramètre d'échelle. La loi a priori solution de cette équation s'écrit alors,

$$\pi(\theta) = \frac{c}{\theta}.$$

Loi a priori de Jefferys

Jefferys (1946,1961) [13], propose une approche plus globale fondée sur l'information de Fisher $I(\theta)$ qui évite de prendre en compte une structure invariante particulière de loi des observations.

Définition 11. Soit θ un paramètre réel. On appelle loi a priori non-informative de Jefferys, la loi (éventuellement impropre) définie par :

Cas uni-dimensionnel

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

Où,

$$I(\theta) = \mathbb{E} \left[\left(\frac{\partial}{\partial \theta} \log f(x|\theta) \right)^2 \right]. \quad (1.7)$$

Si le domaine de X est indépendant de θ alors,

$$I(\theta) = -\mathbb{E} \left[\frac{\partial^2}{\partial \theta^2} \log f(x|\theta) \right]. \quad (1.8)$$

Cas multi-dimensionnel

La matrice d'information de Fisher est une généralisation de (1.8),

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

Et la loi non informative de Jeffreys est définie par,

$$\pi(\theta) \propto [\det(I(\theta))]^{\frac{1}{2}}.$$

Exemple 8. Soit $X \sim \mathcal{N}(\mu, \sigma^2)$, $\theta = (\mu, \sigma)$ inconnu.

On a :

$$\begin{aligned} f(x|\theta) &= \frac{1}{\sqrt{2\pi}\sigma} \exp \left(-\frac{1}{2} \left(\frac{x-\mu}{\sigma} \right)^2 \right), \\ &\propto \frac{1}{\sigma} \exp \left(-\frac{(x-\mu)^2}{2\sigma^2} \right). \end{aligned}$$

Alors,

$$\log f(x|\theta) \propto -\frac{1}{2} \log \sigma^2 - \frac{(x-\mu)^2}{2\sigma^2}.$$

Nous avons,

$$\begin{cases} \frac{\partial}{\partial \mu} \log f(x|\theta) &= \frac{x-\mu}{\sigma^2}, \\ \frac{\partial}{\partial \sigma^2} \log f(x|\theta) &= -\frac{1}{2\sigma^2} + \frac{(x-\mu)^2}{2(\sigma^2)^2}. \end{cases}$$

Et,

$$\begin{cases} \frac{\partial^2}{\partial \mu^2} \log f(x|\theta) &= -\frac{1}{\sigma^2}, \\ \frac{\partial^2}{\partial (\sigma^2)^2} \log f(x|\theta) &= \frac{1}{2(\sigma^2)^2} - \frac{(x-\mu)^2}{(\sigma^2)^3}, \\ \frac{\partial^2}{\partial \mu \partial \sigma^2} \log f(x|\theta) &= -\frac{(x-\mu)}{2(\sigma^2)^2}. \end{cases}$$

La matrice d'information de Fisher s'obtient en calculant l'espérance mathématique des dérivées secondes.

On a,

$$\begin{cases} \mathbb{E}[X - \theta] = 0, \\ \mathbb{E}[(X - \theta)^2] = \sigma^2. \end{cases}$$

On obtient,

$$I(\theta) = \mathbb{E} \begin{bmatrix} \frac{1}{\sigma^2} & \frac{2(x-\theta)}{2(\sigma^2)^2} \\ \frac{2(x-\theta)}{2(\sigma^2)^2} & \frac{1}{2(\sigma^2)^2} - \frac{(x-\mu)^2}{(\sigma^2)^3} \end{bmatrix} = \begin{bmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{1}{2\sigma^2} \end{bmatrix}.$$

Avec,

$$\det I(\theta) = \frac{1}{2\sigma^4}.$$

Et la loi de Jeffreys est,

$$\begin{aligned} \pi(\theta) &\propto [\det I(\theta)]^{\frac{1}{2}}, \\ &\propto \frac{1}{\sigma^2}. \end{aligned}$$

1.4.6 Modèle bayésien hiérarchique

Définition 12. On appelle modèle bayésien hiérarchique un modèle statistique bayésien où la loi a priori est décomposé en distribution conditionnelle,

$$\begin{cases} X|\theta &\sim f(x|\theta), \\ \theta|\theta_1 &\sim \pi_1(\theta|\theta_1), \\ \theta_1|\theta_2 &\sim \pi_2(\theta_1|\theta_2), \\ \vdots &\vdots \\ \theta_{k-1}|\theta_k &\sim \pi_k(\theta_{k-1}|\theta_k), \\ \theta_k &\sim \pi_{k+1}(\theta_k). \end{cases}$$

Les paramètres $\theta_i, i = 1, \dots, k$ sont appelés hyperparamètres.
La loi a priori de θ est alors,

$$\pi(\theta) = \int_{\Theta_1 \times \dots \times \Theta_k} \pi_1(\theta|\theta_1)\pi_2(\theta_1|\theta_2)\dots\pi_{k+1}(\theta_k)d\theta_2\dots d\theta_{k+1}.$$

Avec : $\theta_i \in \Theta, i = 1, \dots, k$.

Exemple 9. *Considérons le modèle bayésien hiérarchique simple suivant,*

$$\begin{cases} X|\theta & \sim f(x|\theta), \\ \theta|\lambda & \sim \text{Exp}(\lambda), \\ \lambda & \sim \text{Exp}(\mu). \end{cases}$$

Où : λ est inconnu et μ est connu.

Et :

$$\pi_1(\theta|\lambda) = \lambda e^{-\lambda\theta}, \lambda > 0, \theta > 0,$$

$$\pi_2(\lambda) = \mu e^{-\mu\lambda}, \mu > 0.$$

D'où la loi a priori de θ et donnée par,

$$\begin{aligned} \pi(\theta) &= \int_0^{+\infty} \pi(\theta|\lambda)\pi_2(\lambda)d\lambda, \\ &= \int_0^{+\infty} \lambda\mu e^{-\lambda(\theta+\mu)}d\lambda, \\ &= \frac{\mu}{(\theta + \mu)^2}, \mu > 0. \end{aligned}$$

1.5 La théorie de la décision

L'objectif des études inférentielles est de fournir une décision aux statisticiens. Les différentes décisions sont comparées au moyen d'un critère d'évaluation donné par la fonction coût. Pour le modèle $X \in (\Omega, \mathcal{A}, \mathbb{P}_\theta, \theta \in \Theta)$, on définit \mathcal{D} l'ensemble des décision δ , en générale $\mathcal{D} = \Theta$. δ est un estimateur.

1.5.1 La fonction coût

On appelle la fonction de coût (ou de perte), toute fonction l de $\Omega \times \mathcal{A}$ dans \mathbb{R} . C'est à dire,

$$\begin{aligned} l : \quad \Omega \times \mathcal{A} &\rightarrow \mathbb{R}, \\ (\theta, \delta(x)) &\mapsto l(\theta, \delta(x)). \end{aligned}$$

$l(\theta, \delta(x))$ évalue le coût (ou la perte) associée à la décision $\delta(x)$ quand le paramètre vaut θ .

Fonctions de coût usuelles

Parmi les fonctions de coût usuelles, on peut citer [20] :

- **Fonction du coût quadratique**

La fonction de coût quadratique est la fonction définie par,

$$l(\theta, \delta(x)) = (\theta - \delta(x))^2.$$

Dans ce cas ce coût pénalise trop fortement les grandes erreurs.

- **Fonction du coût absolu L^1** La fonction du coût absolu L^1 est donnée par,

$$l(\theta, \delta(x)) = |\theta - \delta(x)|.$$

La fonction de coût linéaire par morceaux est définie par,

$$l_{k_1, k_2} = \begin{cases} k_2(\theta - \delta(x)), & \theta > \delta(x), \\ k_1(\delta(x) - \theta), & \theta \leq \delta(x). \end{cases}$$

k_1, k_2 sont des constantes positives.

Contrairement au coût quadratique, le coût absolu ne pénalisent pas les grandes erreurs.

- **Coût 0-1**

C'est un coût non négatif, très utilisé dans la théorie des tests. Il associe à un estimateur δ la pénalité 0 si la réponse est correcte et 1 sinon. Il est défini par,

$$l(\theta, \delta(x)) = \mathbb{1}_{\delta(x)=0} \mathbb{1}_{\theta \in \Theta_1} + \mathbb{1}_{\delta(x)=1} \mathbb{1}_{\theta \in \Theta_0},$$

avec, $\Theta = \Theta_1 \cup \Theta_0$,

et,

$$\delta(x) = \begin{cases} 1, & \text{si } \theta \in \Theta_0, \\ 0, & \text{si } \theta \in \Theta_1. \end{cases}$$

1.5.2 Risque fréquentiste

Dans le cadre de la théorie de la décision, la fonction de coût l est supposée fixée une fois pour toute, et on s'intéresse à ce qui se passe lorsque le paramètre $\theta \in \Theta$ varie ainsi que la règle de décision δ . On appelle risque fréquentiste noté par $R(\theta, \delta)$, le coût moyen (l'espérance mathématique) associé au paramètre θ et à la règle de décision δ ,

$$R(\theta, \delta) = \mathbb{E}_\theta[l(\theta, \delta(x))] = \int_{\mathcal{X}} l(\theta, \delta(x)) f(x, \theta) dx.$$

1.5.3 Risque a posteriori

On appelle le risque a posteriori, noté $\rho(\pi, \delta|x)$, la moyenne du coût par rapport à la loi a posteriori,

$$\rho(\pi, \delta|x) = \mathbb{E}^\pi[l(\theta, \delta(x))|x] = \int_{\Theta} l(\theta, \delta(x)) \pi(\theta|x) d\theta.$$

1.5.4 Risque intégré

On définit le risque intégré, qui est le risque fréquentiste moyen sur les valeurs de θ selon leurs distributions a priori, noté $r(\pi, \delta)$ par,

$$r(\pi, \delta) = \int_{\Theta} R(\theta, \delta) \pi(\theta) d\theta = \mathbb{E}^{\pi}[R(\theta, \delta)].$$

Remarque 5. Dans l'approche classique la décision optimale δ doit minimiser le risque fréquentiste $R(\theta, \delta)$ pour tout $\theta \in \Theta$.

1.5.5 Risque de Bayes

Le risque bayésien (ou risque de Bayes) est la quantité,

$$r(\pi) = r(\pi, \delta^{\pi}) = \inf_{\delta \in \mathcal{D}} r(\pi, \delta).$$

Théorème 1. Si $\exists \delta \in \mathcal{D}$, $r(\pi, \delta) < \infty$ et $\forall x \in \mathcal{X}$ $\delta^{\pi}(x) = \arg \min_{\delta \in \mathcal{D}} \rho(\pi, \delta|x)$, alors $\delta^{\pi}(x)$ est un estimateur bayésien.

Preuve 1. Nous avons,

$$\begin{aligned} r(\pi, \delta) &= \int_{\Theta} R(\theta, \delta) \pi(\theta) d\theta, \\ &= \int_{\Theta} \left(\int_{\mathcal{X}} l(\theta, \delta(x)) L(x|\theta) dx \right) \pi(\theta) d\theta, \\ &= \int_{\mathcal{X}} \left(\int_{\Theta} l(\theta, \delta(x)) L(x|\theta) \pi(\theta) d\theta \right) dx, \\ &= \int_{\mathcal{X}} (l(\theta, \delta(x)) \pi(\theta|x)) f(x) dx, \\ &= \int_{\mathcal{X}} \rho(\theta, \delta|x) f(x) dx. \end{aligned}$$

Ainsi, pour $\delta \in \mathcal{D}$, $\rho(\pi, \delta^{\pi}|x) \leq \rho(\pi, \delta|x) \Rightarrow r(\pi, \delta^{\pi}) \leq r(\pi, \delta)$.
Ce qui permet de conclure.

□

1.5.6 Risque mini-max

On appelle risque mini-max associé à la fonction de coût l , la valeur,

$$\bar{R} = \inf_{\delta \in \mathcal{D}} \sup_{\theta} R(\theta, \delta) = \inf_{\delta \in \mathcal{D}} \sup_{\theta} \mathbb{E}_{\theta}[l(\theta, \delta(x))].$$

Théorème 2. [14]

Le risque de Bayes est toujours plus petit que le risque mini-max, i.e. $\underline{R} \leq \bar{R}$.

Preuve 2. Pour toute règle de décision $\delta \in \mathcal{D}$ on a,

$$\underline{R} \leq r(\pi) = \int_{\Theta} R(\theta, \delta) \pi(\theta) d\theta \leq \sup_{\theta} R(\theta, \delta),$$

si bien que $\underline{R} \leq \sup_{\theta} R(\theta, \delta)$ pour toute règle de décision δ ,

et donc $\underline{R} \leq \inf_{\delta} \sup_{\theta} R(\theta, \delta) = \bar{R}$.

□

\underline{R} désigne le risque max-min, et \bar{R} désigne le risque mini-max.

1.5.7 Admissibilité

Le critère d'admissibilité induit un ordre partiel sur \mathcal{D} en comparant les risques fréquentistes des estimateurs $R(\theta, \delta)$.

Définition 13. Soit un modèle paramétrique et une fonction de perte l sur $\Theta \times \mathcal{D}$. on dit que $\delta \in \Theta$ est inadmissible si et seulement si $\exists \delta_0 \in \mathcal{D}, \forall \theta \in \Theta$,

$$R(\theta, \delta) \geq R(\theta, \delta_0),$$

et $\exists \theta_0 \in \Theta$,

$$R(\theta_0, \delta) > R(\theta_0, \delta_0).$$

Dans le cas contraire δ est admissible.

Proposition 2. [20]

S'il existe un unique estimateur minimax, cet estimateur est admissible.

Proposition 3. [14]

Si δ_0 est admissible de risque constant, δ_0 est l'unique estimateur minimax.

Théorème 3. [14]

Si l'estimateur bayésien δ^π associé à une fonction de perte l et une loi a priori π est unique alors il est admissible.

Preuve 3. Supposons δ^π estimateur bayésien non admissible :

$\exists \delta_0 \in \mathcal{D}, \forall \theta \in \Theta, R(\theta, \delta_0) \leq R(\theta, \delta)$ et $\exists \theta_0 \in \Theta, R(\theta_0, \delta_0) < R(\theta_0, \delta)$.

En intégrant la première inégalité :

$$\int_{\Theta} R(\theta, \delta_0) d\pi(\theta) \leq \int_{\Theta} R(\theta, \delta^\pi) d\pi(\theta) = r(\pi),$$

donc δ_0 est aussi un estimateur bayésien associé à l et π et $\delta_0 \neq \delta^\pi$, d'après la seconde inégalité.

Le théorème se déduit par contraposée.

□

Ce théorème s'applique notamment dans le cas d'un risque fini et d'une fonction de coût convexe. En outre, l'unicité de l'estimateur bayésien implique la finitude du risque :

$$r(\pi) = \int_{\Theta} R(\theta, \delta^\pi) d\pi(\theta) < +\infty,$$

(sinon, tout estimateur minimise le risque).

Proposition 4. [20]

Si un estimateur de Bayes, δ^π associé à une loi a priori (propre ou impropre) π , est tel que le risque de Bayes,

$$r(\pi) = \int_{\Theta} R(\theta, \delta^\pi) \pi(\theta) d\theta,$$

soit fini, alors δ^π est admissible.

1.6 L'estimation bayésienne

1.6.1 Cas uni-dimensionnel

Définition 14. On appelle estimateur de bayes associé à une fonction de coût $l(\theta, \delta(x))$ et à une loi a priori π , toute décision δ^π qui minimise le risque a posteriori,

$$\delta^\pi(x) = \arg \min_{\delta \in \mathcal{D}} \rho(\pi, \delta|x), x \in \mathcal{X}.$$

On peut citer :

- Estimateur MMSE (Minimum Mean Square Error) ;
- Estimateur MAP (maximum a posteriori) ;
- Médiane a posteriori ;
- Estimateur PMS (posteriori mean square).

Estimateur MMSE

L'estimateur MMSE noté par $\hat{\theta}_{MMSE}$ est la moyenne a postérieure de θ , mathématiquement on a,

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

Estimateur MAP

L'estimateur MAP de θ se fait par maximisation de la loi a posteriori. C'est à dire choisir le mode de probabilité a postérieure $\pi(\theta|x)$. L'estimateur MAP est défini par,

$$\hat{\theta}_{MAP} = \arg \max_{\theta} \pi(\theta|x) = \arg \max_{\theta} \pi(\theta) f(x|\theta).$$

Médiane a posteriori

La médiane est également un résumé possible de la distribution a posteriori. Comme son nom l'indique, il s'agit de la médiane de $\pi(\theta|x)$.

Estimateur PMS

Le carré moyen a posteriori PMS (posterior mean square) d'un estimateur $\hat{\theta}$ de paramètre θ , mesure la distance quadratique moyenne par rapport à la densité a posteriori à laquelle l'estimateur de θ est éloigné de la valeur réelle θ , il est défini par,

$$\hat{\theta}_{PMS} = \int_{\Theta} (\theta - \hat{\theta})^2 \pi(\theta|x) d\theta.$$

Propriétés d'estimateur de Bayes [6] :

1. L'estimateur de Bayes est admissible.
2. L'estimateur de Bayes est biaisé.
3. L'estimateur de Bayes est convergent en probabilité quand la taille de l'échantillon $n \rightarrow \infty$.

1.6.2 Cas multi-dimensionnel

Dans un contexte multi-dimensionnel on considère $\theta = (\theta_1, \dots, \theta_K)$, la moyenne a posteriori $\mathbb{E}[\theta|x]$ est égale au vecteur $(\mathbb{E}[\theta_k|x], k = 1, \dots, K)$.

Avec,

$$\mathbb{E}[\theta_k|x] = \int_{\Theta_k} \theta_k \pi(\theta_k|x) d\theta_k,$$

$\pi(\theta_k|x)$ est obtenu en intégrant $\pi(\theta|x)$ sur toutes les composantes de θ autres que θ_k .

1.7 Tests et régions de confiance bayésien

Les tests et régions de confiance bayésiens sont des outils statistiques puissants, qui permettent de prendre des décisions sur les paramètres d'un modèle en utilisant la théorie de la probabilité bayésienne. Les tests bayésiens permettent de comparer deux hypothèses H_0 et H_1 , en utilisant la théorie de la probabilité bayésienne, les tests bayésiens peuvent-être plus flexibles que les tests classiques, car ils permettent l'incorporation de connaissances préalables et la quantification de l'incertitude associée aux résultats. Les régions de confiance bayésiennes sont des intervalles de valeurs possibles pour un paramètre, qui prennent en compte l'incertitude associée à l'estimation des paramètres, contrairement aux intervalles de confiance classiques, qui ne prennent en compte que l'incertitude associée à l'échantillonnage.

1.7.1 Régions de confiance

Définition 15. Soit un modèle statistique bayésien $(\mathcal{X}, \mathcal{A}, \mathbb{P}_\theta, \pi(\theta), \theta \in \Theta)$, on appelle région de confiance de niveau α ou région α -crédible, tout ensemble C_x tel que,

$$\mathbb{P}^{\pi(\cdot|x)}(\theta \in C_x) \geq (1 - \alpha),$$

où $\mathbb{P}^{\pi(\cdot|x)}$ est la loi de probabilité dont la densité a posteriori est $\pi(\cdot|x)$.

On dira qu'une région est α -crédible de Plus Forte Densité a Posteriori, et on note **PFDP** (ou encore **HPD**, Highest Posterior Density), si elle s'écrit :

$$C_x^\pi = \{\theta; \pi(\theta|x) \geq k_\alpha\},$$

où k_α est la plus grande valeur telle que,

$$\mathbb{P}^{\pi(\cdot|x)}(\theta \in C_x^\pi) \geq (1 - \alpha).$$

Intervalle de confiance a priori

On appelle un intervalle de confiance a priori J , de niveau de confiance $(1 - \alpha)$, l'intervalle pour lequel on a [7],

$$\mathbb{P}(\theta \in J) = \int_J \pi(\theta) d\theta = (1 - \alpha),$$

où α est une probabilité fixée dans $]0,1[$.

Intervalle de confiance a posteriori

On appelle intervalle de confiance a posteriori I , de niveau de confiance $(1 - \alpha)$, l'intervalle pour lequel on a [7],

$$\mathbb{P}(\theta \in I|x) = \int_I \pi(\theta|x) d\theta = (1 - \alpha),$$

où α est une probabilité fixé dans $]0,1[$.

1.7.2 Tests bayésien

Une autre façon de mener l'inférence sur le paramètre θ est de proposer des hypothèses portant sur des régions de Θ . Autrement dit, on se pose la question de savoir s'il est possible d'admettre que θ appartienne à telle ou telle région de Θ et ce au vu de ce que l'on a observé. On va donc émettre une hypothèse notée H_0 dite hypothèse nulle que l'on va chercher à tester. L'observation nous autorise-t-elle à considérer cette hypothèse comme vraie ou nous conduit-elle à la rejeter? En général, on teste l'hypothèse H_0 contre une hypothèse dite hypothèse alternative notée H_1 . Le test s'écrit,

$$H_0 : \theta \in \Theta_0 \quad VS \quad H_1 : \theta \in \Theta_1,$$

où Θ_0 et Θ_1 sont des sous-ensembles de Θ .

Dans la construction du test, on cherchera à faire en sorte que ces deux erreurs soient les plus faibles possibles.

- La première de ces erreurs, rejeter H_0 alors qu'elle est vraie est appelée erreur de 1^{ère} espèce ou erreur de type I , ou parfois risque de 1^{ère} espèce. On la note, α .
- La seconde, ne pas rejeter H_0 alors qu'elle est fautive est appelée erreur de 2^{ème} espèce ou erreur de type II , ou parfois risque de 2^{ème} espèce. On la note, β .

Un test est un problème de décision. Suivant ce que l'on aura observé, on décidera de rejeter ou non H_0 . On va chercher à explorer les sous-ensembles de l'espace des observations \mathcal{X} qui contredisent l'hypothèse nulle. Ils formeront une région dite région de rejet ou région critique.

$$W = \{x_1, \dots, x_n \mid \text{on rejete } H_0\}.$$

Avec :

$$\alpha = \mathbb{P}_\theta(W \mid \theta \in \Theta_0),$$

$$\beta = 1 - \mathbb{P}_\theta(W \mid \theta \in \Theta_1).$$

L'espace des actions contient deux éléments,
 d_0 : on ne rejette pas H_0 , $\theta \in \Theta_0$,
 d_1 : on rejette H_0 .

A ces décisions vont-êre associées des coûts $l(\theta, d_0(x))$ et $l(\theta, d_1(x))$. Par définition, les décisions bayésiennes sont celles qui minimisent le coût a posteriori.

Facteur de Bayes

Dans le cadre bayésien, disposant d'une loi a priori et ayant calculé une loi a posteriori, la mesure des régions Θ_0 et Θ_1 est directe et ne présente pas de difficulté. On calculera deux probabilités que l'on comparera pour prendre une décision, rejeter ou accepter H_0 . Ces probabilités sont,

$$\alpha_0 = \mathbb{P}(H_0|X) = \int_{\Theta_0} \pi(\theta|x)d\theta \text{ et } \alpha_1 = \mathbb{P}(H_1|X) = \int_{\Theta_1} \pi(\theta|x)d\theta.$$

La décision repose donc sur la nature du rapport α_0/α_1 .

Définition 16. *Le facteur de Bayes est le rapport,*

$$B = \frac{\alpha_0/\alpha_1}{\pi_0/\pi_1}.$$

Avec

$$\pi_i = \mathbb{P}(\theta \in \Theta_i) = \int_{\Theta_i} \pi(\theta)d\theta, \quad i = 0, 1.$$

Remarque 6. *Lorsque $\Theta_0 = \{\theta_0\}$, et $\Theta_1 = \{\theta_1\}$, le facteur de Bayes n'est autre que le rapport de vraisemblance classique. En effet,*

$$B = \frac{f(x|\theta_0)}{f(x|\theta_1)},$$

puisque $\int_{\Theta_i} f(x|\theta)\pi(\theta)d\theta = f(x|\theta_i)\pi(\theta_i)$, $i = 0, 1$.

Interprétation du facteur de Bayes

Kass et Raftery (1995) [15], proposent un guide d'interprétation pour le facteur de Bayes. Le tableau (1.2), donne quelques interprétations possibles.

| | |
|--------------------------|---------------------------|
| Facteur de Bayes (B) | L'hypothèse H_0 |
| entre 1 et 3.2 | rejetée |
| entre 3.2 et 10 | fortement rejetée |
| entre 10 et 100 | très fortement rejetée |
| > 100 | rejetée de façon décisive |

TABLE 1.2 – Interprétation du facteur de Bayes.

1.8 Incorporation de l'information a priori en pratique

Nous allons nous limiter au cas d'un paramètre réel θ et envisager trois situations selon que :

- $\theta \in [0,1]$,
- $\theta \in [0, +\infty[$,
- $\theta \in]-\infty, +\infty[$.

Nous supposons par ailleurs que l'information a priori fourni par l'expert consiste en estimation θ^* de θ , et en un intervalle $I^* = [\alpha^*, \beta^*]$ contenant θ^* et tel que la probabilité que θ appartienne à I^* soit élevée (typiquement 0.95) [7].

Situation 1 : Le paramètre $\theta \in [0,1]$

Pour incorporer cette information a priori, il est commode de reparamétriser la loi $\mathcal{B}e(\alpha, \beta)$ comme suit, on pose,

$$\begin{cases} \alpha = \lambda\mu, \\ \beta = \lambda(1 - \mu). \end{cases}$$

Avec,

$$\begin{cases} \mu = \frac{\alpha}{\alpha + \beta} = \mathbb{E}[\theta], \\ \lambda = \alpha + \beta. \end{cases}$$

$$Var[\theta] = \frac{\mu(1 - \mu)}{1 + \lambda}.$$

Où μ est fixé, λ est inversement proportionnel à la variance et donc s'interprète comme la précision de l'information a priori. Autrement dit, plus λ est grand, plus $Var[\theta]$ est petit, et plus l'information a priori est précise ; à l'inverse, plus λ est petit, plus $Var[\theta]$ est grand, et moins l'information a priori est précise.

En pratique, on procède comme suit, on choisit comme loi a priori une loi $\mathcal{B}e(\alpha, \beta)$ et on pose $\mathbb{E}[\theta] = \theta^*$.

Le paramètre λ est déterminé par,

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

Avec $\theta \sim \mathcal{B}e(\lambda\theta^*, \lambda(1 - \theta^*))$.

Exemple 10. On suppose que :

$$\theta^* = 0.2, \mathbb{P}(\theta \in [0.05, 0.4]) = 0.95$$

On a donc $\mathbb{E}[\theta] = 0.2$, et $\theta \sim \mathcal{B}e(0.2\lambda, 0.8\lambda)$.

A l'aide de Matlab, on trouve $\lambda = 17.5089$;

par conséquent, la loi a priori $\pi(\theta) \sim \mathcal{B}e(3.4932, 14.0072)$ (voir figure (1.2)).

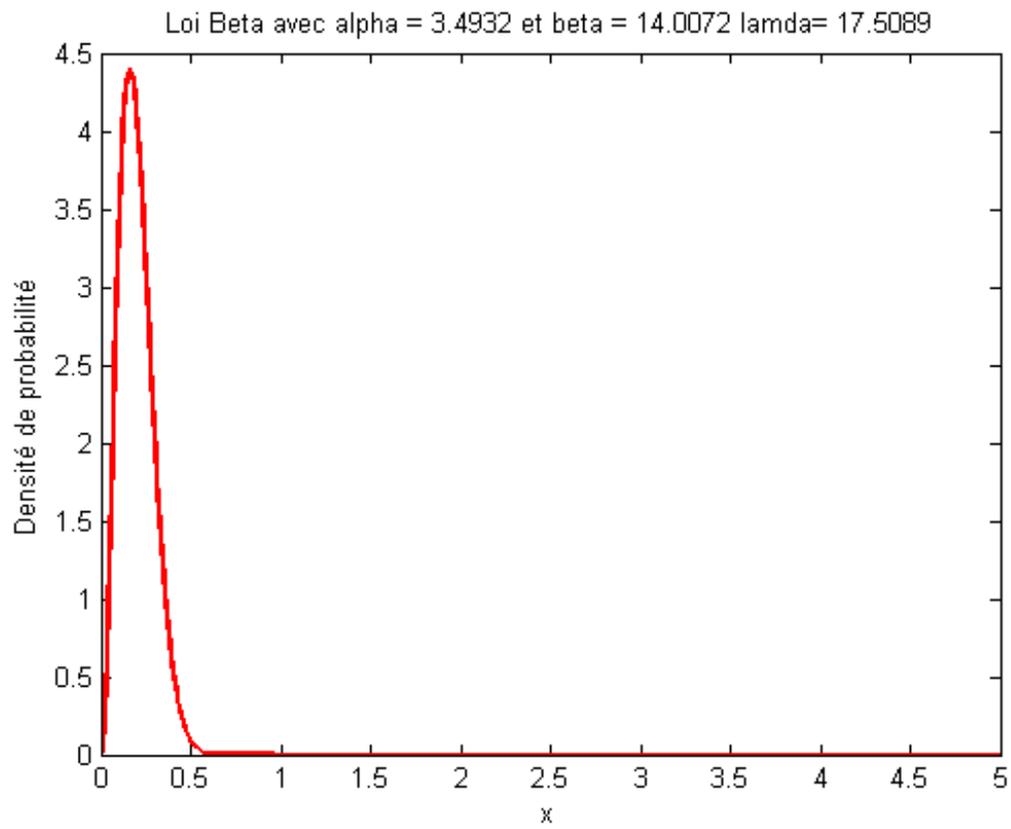


FIGURE 1.2 – La représentation graphique de la distribution $\mathcal{B}e(3.4932, 14.0072)$.

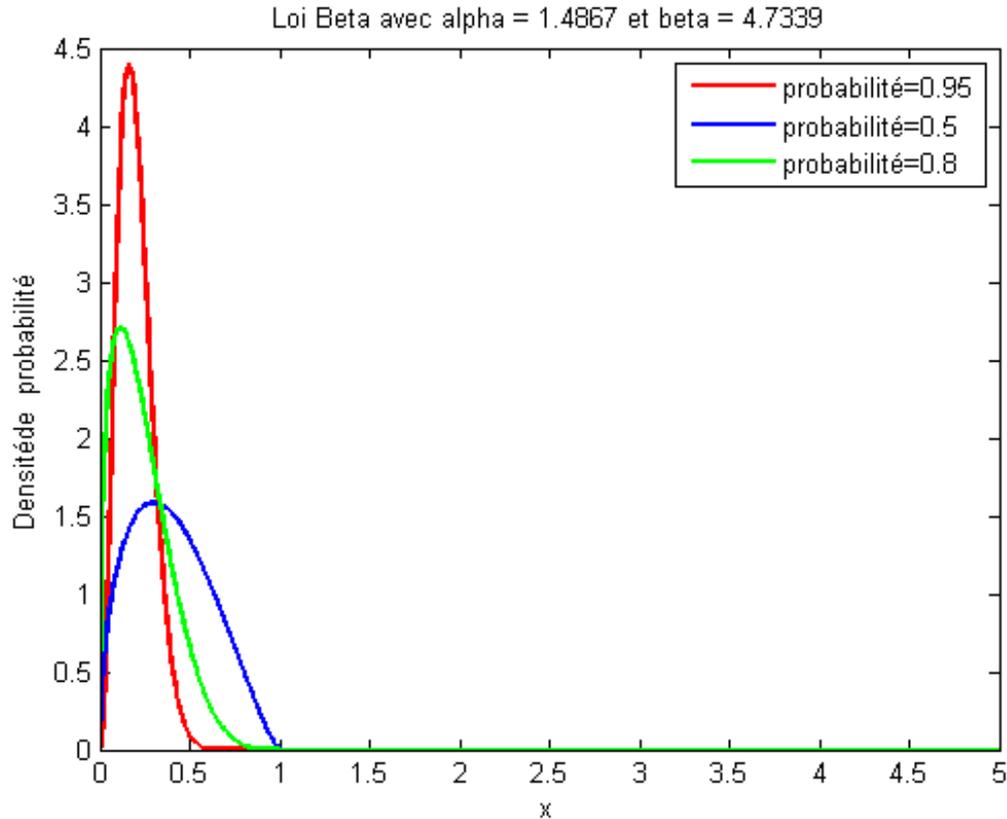


FIGURE 1.3 – La représentation graphique des distributions Beta.

D'après le programme Matlab, pour une moyenne fixée $\mu = 0.2$ et une probabilité de 0.95 on obtient une variance de 0.0086, tandis que pour une probabilité de 0.5 on obtient une variance de 0.0404, et pour une probabilité de 0.8 on obtient une variance de 0.0191. Donc plus la probabilité est grande plus la variance est petite, plus l'information a priori est précise.

Situation 2 : Le paramètre $\theta \in [0, +\infty[$

Dans ce cas on choisit comme loi a priori une loi Gamma $\mathcal{G}(\alpha, \beta)$ qui peut-être reparamétrée par sa moyenne et sa variance.

On pose,

$$\begin{cases} \alpha = \lambda\mu, \\ \beta = \lambda. \end{cases}$$

$\mu = \mathbb{E}[\theta]$
et,

$$\begin{cases} \mu = \frac{\alpha}{\beta}, \\ \lambda = \alpha. \end{cases}$$

On a,

$$\begin{cases} \mathbb{E}[\theta] = \mu = \frac{\alpha}{\beta}, \\ Var[\theta] = \frac{\alpha}{\beta^2} = \frac{\mu}{\lambda}. \end{cases}$$

En pratique, on pose $\mathbb{E}[\theta] = \theta^* = \mu$.

D'où : $\theta \sim \mathcal{G}(\lambda\theta^*, \lambda)$.

Le paramètre λ est déterminé par,

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

Exemple 11. On suppose que, $\theta^* = 0.2$, $\mathbb{P}(\theta \in [0, 5]) = 0.95$

On a donc $\mathbb{E}[\theta] = 0.2$, $\lambda = 0.7491$, et $\theta \sim \mathcal{G}(1.4074, 0.7491)$ et $\text{var}[\theta] = 0.2670$.

Et pour $\theta^* = 0.2$, $\mathbb{P}(\theta \in [4, 10]) = 0.95$ on trouve $\lambda = 0.8169$

$\theta \sim \mathcal{G}(3.2632, 0.8169)$ et $\text{var}[\theta] = 0.2448$

La représentation graphique des distributions gamma pour $\theta \sim [0, 5]$ et pour $\theta \sim [4, 10]$ sont donnés dans la figure (1.4).

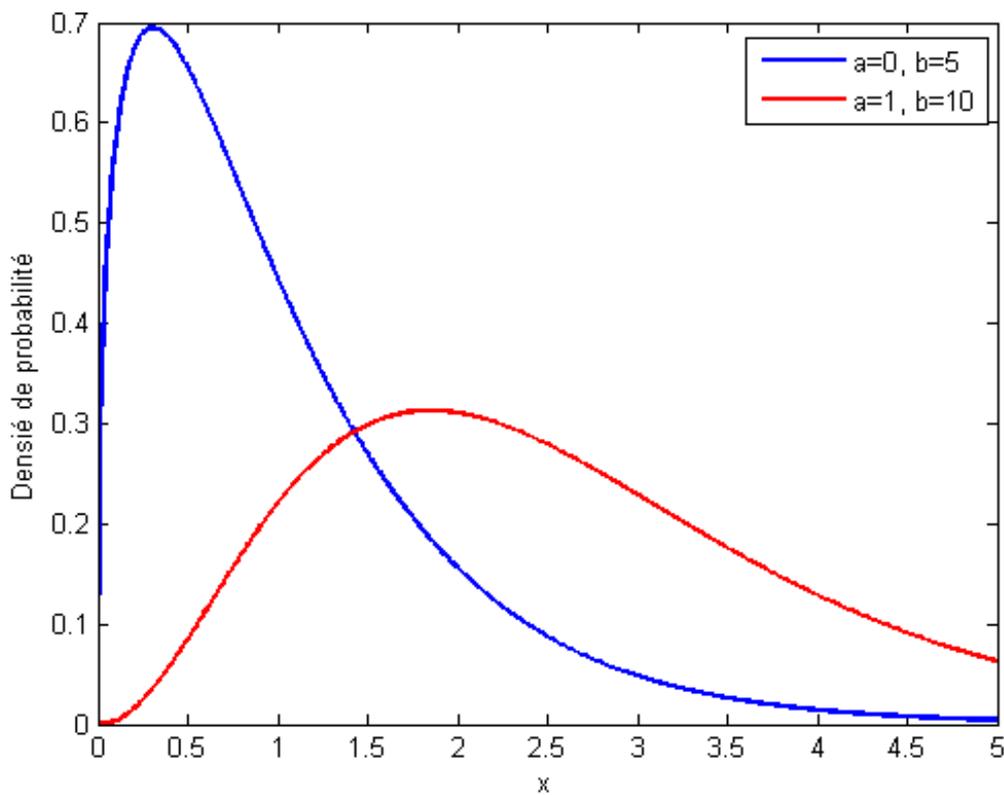


FIGURE 1.4 – La représentation graphique des distributions Gama. La courbe en bleu pour $\theta \in [0, 5]$ et la courbe rouge pour $\theta \in [1, 10]$.

D'après cette figure, plus λ est grand et plus $\text{Var}[\theta]$ est petit, plus l'information a priori est précise.

Situation 3 : Le paramètre $\theta \in]-\infty, +\infty[$

Dans ce cas on choisit, comme loi a priori, une loi Normale.

Exemple 12. On suppose que, $\theta^* = 0.2$, $\mathbb{P}(\theta \in [1, 4]) = 0.95$, on trouve $\theta \sim \mathcal{N}(2.5505, 0.2344)$.

Et pour $\mathbb{P}(\theta \in [1, 5]) = 0.95$, on trouve $\theta \sim \mathcal{N}(3.3457, 0.2867)$.

La figure (1.5), représente les distributions $\mathcal{N}(2.5505, 0.2344)$ et $\mathcal{N}(3.3457, 0.2867)$.

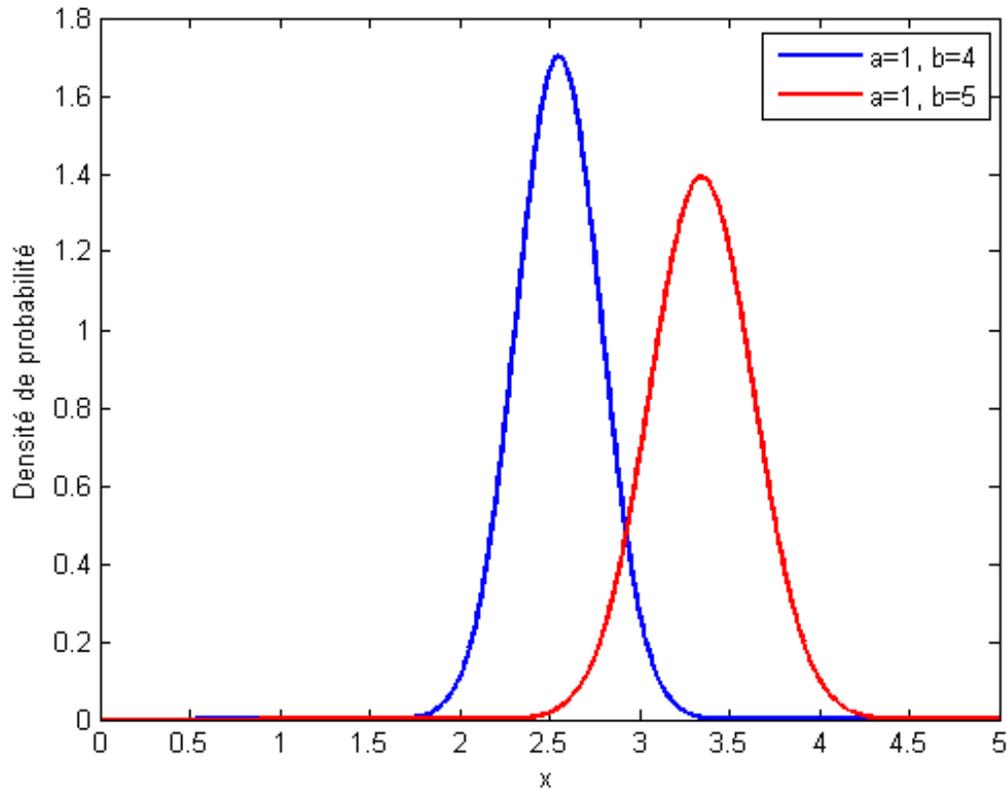


FIGURE 1.5 – La représentation graphique des distributions $\mathcal{N}(2.5505, 0.2344)$ et $\mathcal{N}(3.3457, 0.2867)$.

1.9 Conclusion

Dans ce chapitre, nous avons abordé les notions fondamentales de la statistique bayésienne, notamment les concepts de l'information a priori, de la loi a priori et de l'estimation bayésienne. Nous avons exploré les principes et les techniques clés de l'inférence bayésienne.

Dans le chapitre suivant, nous nous pencherons sur les méthodes de Monte-Carlo et les méthodes de Monte-Carlo par chaîne de Markov (MCMC). Nous examinerons en détail ces approches, y compris les algorithmes MCMC tels que l'algorithme de Metropolis-Hastings et l'échantillonneur de Gibbs.

Chapitre 2

Méthodes de Monte-Carlo

Dans ce chapitre, nous explorons les méthodes de Monte-Carlo, en mettant l'accent sur l'intégration Monte-Carlo et les méthodes de Monte-Carlo par chaîne de Markov (MCMC). Nous étudions en détail deux approches principales pour construire des chaînes de Markov avec une distribution stationnaire spécifiée : les algorithmes de Metropolis-Hastings et l'échantillonneur de Gibbs.

2.1 Intégration Monte-Carlo

Le nom de ces méthodes a été inspiré par les jeux de hasard pratiqués au casino de Monte-Carlo (Monaco). Nicholas Metropolis est crédité de l'invention de ce terme en 1947, et l'article coécrit avec Stanislaw Ulam en 1949 est la première publication à utiliser ce nom. Pour appliquer les méthodes de Monte-Carlo, il est nécessaire de savoir comment simuler des variables aléatoires selon une loi donnée. La simulation est une technique qui permet de reproduire le comportement d'un système réel en utilisant un modèle informatique (pour plus de détail voir Annexe A).

Une utilisation courante des méthodes de Monte-Carlo consiste à estimer de manière approximative des quantités de la forme,

$$I = \mathbb{E}[\varphi(X)] = \int_{\mathbb{R}^d} \varphi(x)f(x)dx \quad d \in \mathbb{N}^*. \quad (2.1)$$

Où $\varphi : \mathbb{R}^d \rightarrow \mathbb{R}$ est une fonction donnée et X un vecteur aléatoire de densité f suivant laquelle on sait simuler.

Dans ce contexte, l'estimateur Monte-Carlo fondamental est défini comme suit,

$$\hat{I}_n = \frac{1}{n} \sum_{i=1}^n \varphi(X_i). \quad (2.2)$$

Où les X_i sont générées de façon i.i.d. selon f .

L'intégration Monte-Carlo repose sur la loi forte des grands nombres.

Théorème 4. (*Convergence*)

Soit $(X_n)_{n \geq 1}$ une suite de v.a.i.i.d à valeur dans \mathbb{R}^d , $d \in \mathbb{N}^*$, $X_i \sim \mathcal{L}(f)$. On suppose que $\varphi(x)$ est une fonction mesurable, ($\mathbb{E}[\varphi(X)] < +\infty$), alors :

$$\frac{1}{n} \sum_{i=1}^n \varphi(X_i) \xrightarrow[n \rightarrow +\infty]{p.s.} \mathbb{E}[\varphi(X)].$$

Preuve 4. Pour la preuve de ce théorème voir [21].

Ce théorème donne la validité de l'approximation de Monte-Carlo et sous quelles hypothèses elle est valable.

En d'autres termes, pour estimer une quantité de type $I = \int_{\mathbb{R}^d} \varphi(x)f(x)dx$ à l'aide d'une méthode de Monte-Carlo, on suit les étapes suivantes,

Etape1 : Si φ est une fonction mesurable ($\mathbb{E}[\varphi(X)] < +\infty$), et si f est une densité, alors, on peut mettre la quantité I , sous forme d'une espérance, i.e.

$$I = \int_{\mathbb{R}^d} \varphi(x)f(x)dx = \mathbb{E}[\varphi(X)].$$

Etape2 : Si on sait simuler $\varphi(X_1), \dots, \varphi(X_n)$ alors par l'application de la loi forte des grands nombres nous pouvons approcher la quantité I par un estimateur Monte-Carlo, i.e.

$$I = \mathbb{E}[\varphi(X)] \approx \frac{1}{n} \sum_{i=1}^n \varphi(X_i).$$

En utilisant les hypothèses précédentes et les lois des grands nombres, l'algorithme qui permet d'estimer une quantité de type (2.1) est donnée par,

Algorithme

Debut

lire(n);

$s = 0$;

pour $i = 1 : n$ **faire**

générer $X(i) \sim \mathcal{L}(f)$;

$s = s + \varphi(X(i))$;

fin pour;

$\hat{I}_n \approx \frac{s}{n}$;

afficher (\hat{I}_n);

Fin.

2.1.1 Le paradigme bayésien

De façon générale, les méthodes Monte-Carlo sont d'usage constant en statistique bayésienne. En particulier, celle-ci requiert régulièrement en effet, le calcul d'intégrale de type (2.1). Le cadre typique est celui où la loi de variable X dépend d'un paramètre θ . Dans l'approche fréquentistes, θ est inconnu mais supposé avoir une valeur fixée et les observations (x_1, \dots, x_n) permettent de l'estimer par la méthode donné, par exemple au maximum de vraisemblance. L'approche bayésienne est différente, elle consiste à considérer que θ est lui-même aléatoire et suit une loi (dite a priori) donné, les observations (x_1, \dots, x_n) permettent d'affiner cette loi via sa mise à jour au vu des observations.

Plus formellement, notons π la densité de la loi a priori de θ et $f(x|\theta)$ la densité conditionnelle de x sachant θ (ou vraisemblance). Par la règle de Bayes, la densité a posteriori de θ sachant x s'écrit alors tout simplement,

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

Ceci fait, on pourra s'intéresser à une valeur moyenne par rapport à cette loi a posteriori, laquelle s'écrira donc,

$$\mathbb{E} [\varphi(\theta|x)] = \int_{\Theta} \varphi(\theta)\pi(\theta|x)d\theta = \frac{\int_{\Theta} \varphi(\theta)f(x|\theta)\pi(\theta)d\theta}{\int_{\Theta} f(x|\theta)\pi(\theta)d\theta}. \quad (2.3)$$

Il convient de garder en tête que le paramètre θ est en général multidimensionnel. Par conséquent, le calcul de l'intégrale (2.3) est, sauf cas particulier, impossible analytiquement et difficile par intégration numérique déterministe.

Supposons cependant que l'on sache simuler suivant la loi a priori $\pi(\theta)$ et, pour tout θ , évaluer la quantité $\varphi(\theta)f(x|\theta)$. On retombe alors exactement dans le cadre d'application des méthodes Monte-Carlo d'intégration, puisqu'il suffit de générer des réalisations i.i.d $(\theta_1, \dots, \theta_n)$ selon $\pi(\theta)$ pour en déduire que,

$$\frac{1}{n} \sum_{i=1}^n \varphi(\theta_i)f(x|\theta_i) \xrightarrow[n \rightarrow +\infty]{p.s.} \int \varphi(\theta)f(x|\theta)\pi(\theta)d\theta.$$

L'estimation du dénominateur de (2.3) correspond au cas particulier où $\varphi = 1$ et se traite donc de la même façon. Par conséquence, l'estimateur Monte-Carlo de $I = \mathbb{E} [\varphi(\theta)|x]$, s'écrit,

$$\hat{I}_n = \frac{\sum_{i=1}^n \varphi(\theta_i)f(x|\theta_i)}{\sum_{i=1}^n f(x|\theta_i)}.$$

En particulier, l'estimateur de Bayes,

$$\hat{\theta}(x) = \int \theta\pi(\theta|x)d\theta,$$

qui correspond à la moyenne a posteriori de θ , admet pour estimateur Monte-Carlo,

$$\hat{\theta}_n(x) = \frac{\sum_{i=1}^n \theta_i f(x|\theta_i)}{\sum_{i=1}^n f(x|\theta_i)}. \quad (2.4)$$

Exemple 13. Dans le cadre bayésienne, on suppose que la loi a priori sur θ est une loi de Cauchy standard,

$$\pi(\theta) = \frac{1}{\pi(1 + \theta^2)}, \quad \theta \in \mathbb{R}$$

supposons aussi que les variables X_i sont i.i.d selon une loi normale de moyenne θ , et de variance 1, $X_i \sim \mathcal{N}(\theta, 1)$.

La densité a posteriori s'obtient par la formule de Bayes,

$$\begin{aligned} \pi(\theta|x) &= \frac{\pi(\theta)f(x|\theta)}{f(x)}, \\ &= \frac{C_x}{1 + \theta^2} \exp\left(-\frac{1}{2} \left(\sum_{j=1}^N (x_j - \theta)^2\right)\right), \quad x \in \mathbb{R} \end{aligned}$$

où C_x est la constante de normalisation.

$$C_x = \int_{\mathbb{R}} \frac{1}{1 + \theta^2} \exp\left(-\frac{1}{2} \sum_{j=1}^N (x_j - \theta)^2\right) d\theta,$$

dépend de x mais pas de θ . La moyenne a posteriori $\theta(x)$ est par définition la moyenne de θ par cette loi a posteriori, c'est-à-dire,

$$\hat{\theta}(x) = \int_{\mathbb{R}} \theta \pi(\theta|x) d\theta = C_x \int_{\mathbb{R}} \frac{\theta}{1 + \theta^2} \exp\left(-\frac{1}{2} (x_j - \theta)^2\right) d\theta.$$

On ne dispose pas de formule analytique évidente pour calculer cette intégrale, d'où le recours aux méthodes de Monte-Carlo.

L'estimateur Monte-Carlo $\hat{\theta}_n(x)$ de $\hat{\theta}(x)$ est le même que celui donné par l'équation (2.4) à savoir,

$$\begin{aligned} \hat{\theta}_n(x) &= \frac{\sum_{i=1}^n \theta_i f(x|\theta_i)}{\sum_{i=1}^n f(x|\theta_i)}, \\ &= \frac{\sum_{i=1}^n \theta_i \exp\left(-\frac{1}{2} \sum_{j=1}^N (x_j - \theta_i)^2\right)}{\sum_{i=1}^n \exp\left(-\frac{1}{2} \sum_{j=1}^N (x_j - \theta_i)^2\right)}. \end{aligned}$$

L'algorithme qui permet d'estimer cette quantité est donné par,

Algorithme

Debut

lire (n);

lire (N);

lire(θ_0);

$s_1 = 0$;

$s_2 = 0$;

pour $i = 1 : n$ **faire**

 générer $U(i) \sim \mathcal{U}([0, 1])$;

$\theta(i) = \tan(\pi(u(i) - \frac{1}{2}))$;

$s(i) = 0$;

pour $j = 1 : N$ **faire**

 générer $V(j) \sim \mathcal{U}([0, 1])$;

 générer $W(j) \sim \mathcal{U}([0, 1])$;

$X(j) = \theta_0 + \sqrt{(-2 * \log(V(j)))} * \cos(2 * \pi * W(j))$;

$s(i) = s(i) + (X(j) - \theta(i))^2$;

fin pour;

$s_1 = s_1 + \theta(i) * \exp(-\frac{1}{2} * s(i))$;

$s_2 = s_2 + \exp(-\frac{1}{2} * s(i))$;

fin pour;

$\theta_n = s_1/s_2$

disp(θ_n);

Fin.

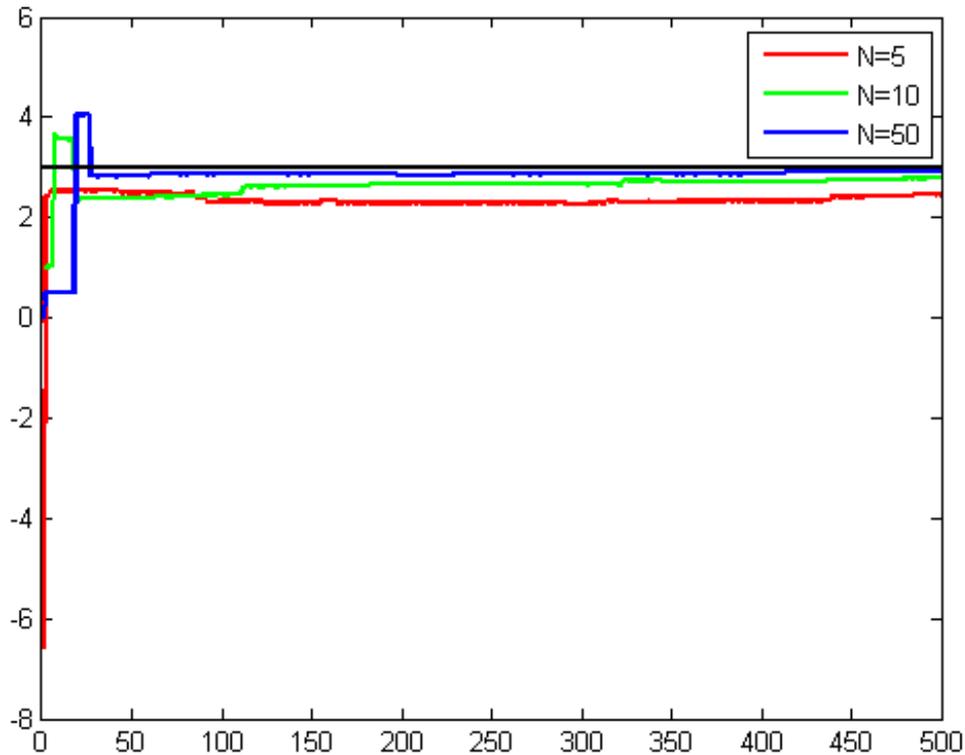


FIGURE 2.1 – L’estimation de $\hat{\theta}_n$. La courbe en rouge pour $N = 5$, la courbe en vert pour $N = 10$ et la courbe en bleu pour $N = 50$.

D’après le programme de cet algorithme et pour $\theta_0 = 3$, nous trouvons $\hat{\theta}_n = 2.9196$ pour $N = 50$, $\hat{\theta}_n = 2.7665$ pour $N = 10$, et $\hat{\theta}_n = 2.4523$ pour $N = 5$. Donc plus N est grand, plus l’estimateur $\hat{\theta}(x)$, lui même estimé par $\hat{\theta}_n(x)$, est proche de 3.

On dispose donc d’un estimateur Monte-Carlo pour estimer des quantités de type $I = \int_{\mathbb{R}^d} \varphi(x)f(x)dx$. Encore faut-il connaître la précision de ces estimateurs. C’est tout l’intérêt de la section suivante.

2.1.2 Erreur de l’estimation et l’intervalle de confiance

La méthode de Monte-Carlo pour l’estimation d’une quantité $\mathbb{E}[\varphi(X)]$ génère une erreur aléatoire, qui ne peut-être bornée. Mais, il est possible d’établir un intervalle de confiance pour le résultat en se basant sur le théorème central limite (TCL).

Théorème 5. TCL

Soit $\varphi(X_i)$, $i = 1 \dots n$, une suite de v.a.i.i.d à valeur dans \mathbb{R}^d , $d \in \mathbb{N}^*$.

On suppose que $\mathbb{E}[\varphi(X)^2] \leq +\infty$.

Soit $\text{var}[\varphi(X)] = \sigma^2$ (Variance de $\varphi(X)$), alors

$$\frac{\sqrt{n}}{\sigma} \left[\frac{1}{n} \sum_{i=1}^n \varphi(X_i) - \mathbb{E}[\varphi(X)] \right] \xrightarrow[n \rightarrow +\infty]{\text{loi}} \mathcal{N}(0, 1).$$

Preuve 5. Pour la preuve voir [21].

Notons l'erreur d'estimation $\varepsilon_n = |\hat{I}_n - I| = \left| \frac{1}{n} \sum_{i=1}^n \varphi(X_i) - \mathbb{E}[\varphi(X)] \right|$.

Sous les hypothèses du théorème (5), on cherche à approcher $\mathbb{E}[\varphi(X)]$ à β près avec une confiance de $(1 - \alpha)\%$ (L'objectif est de construire un intervalle de confiance). C'est-à-dire que l'on veut calculer

$$\mathbb{P}(|\varepsilon_n| \geq \beta) \leq \alpha,$$

ce qui est équivalent à

$$\mathbb{P}(|\varepsilon_n| \leq \beta) \geq (1 - \alpha). \quad (2.5)$$

Nous avons,

$$\begin{aligned} \mathbb{P}(|\varepsilon_n| \leq \beta) &= \mathbb{P}(-\beta \leq \varepsilon_n \leq \beta) \\ &= \mathbb{P}\left(-\beta \frac{\sqrt{n}}{\sigma} \leq \frac{\sqrt{n}}{\sigma} \varepsilon_n \leq \frac{\sqrt{n}}{\sigma} \beta\right) \\ &= \mathbb{P}\left(-\beta \frac{\sqrt{n}}{\sigma} \leq \frac{\sqrt{n}}{\sigma} \left(\frac{1}{n} \sum_{i=1}^n \varphi(X_i) - \mathbb{E}[\varphi(X)] \right) \leq \frac{\sqrt{n}}{\sigma} \beta\right). \end{aligned}$$

D'après théorème (5), $\frac{\sqrt{n}}{\sigma} \left(\frac{1}{n} \sum_{i=1}^n \varphi(X_i) - \mathbb{E}[\varphi(X)] \right) \xrightarrow[n \rightarrow +\infty]{loi} \mathcal{N}(0, 1)$.

Pour n assez grand,

$$\mathbb{P}(|\varepsilon_n| \leq \beta) \approx \int_{-\beta \frac{\sqrt{n}}{\sigma}}^{\beta \frac{\sqrt{n}}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx.$$

Par symétrie de $\mathcal{N}(0, 1)$, on trouve,

$$\mathbb{P}(|\varepsilon_n| \leq \beta) \approx 2 \int_{-\infty}^{\beta \frac{\sqrt{n}}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx - 1.$$

En remplaçant dans l'équation (2.5), on trouve,

$$\int_{-\infty}^{\beta \frac{\sqrt{n}}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \geq 1 - \alpha/2.$$

Par l'application de la fonction inverse ϕ^{-1} de la loi normale $\mathcal{N}(0, 1)$, on trouve que,

$$\begin{aligned} \frac{\sqrt{n}}{\sigma} \beta \geq \phi^{-1}(1 - \alpha/2) &\Rightarrow \sqrt{n} \geq \left(\frac{\phi^{-1}(1 - \alpha/2)}{\beta} \right) \sigma. \\ &\Rightarrow n \geq \left(\frac{\phi^{-1}(1 - \alpha/2)}{\beta} \right)^2 \sigma^2. \end{aligned}$$

On remarque que le nombre de tirage nécessaire n pour atteindre un certain niveau d'erreur avec une certaine confiance est une fonction linéaire en fonction σ^2 .

Proposition 5. Soit $\alpha \in [0, 1]$ fixé. Un intervalle de confiance de niveau asymptotique $(1 - \alpha)$ pour I est,

$$\left[\hat{I}_n - \phi^{-1}(1 - \alpha/2) \sqrt{\frac{\sigma^2}{n}}, \hat{I}_n + \phi^{-1}(1 - \alpha/2) \sqrt{\frac{\sigma^2}{n}} \right].$$

Où $\phi^{-1}(1 - \alpha/2)$ désigne le quantile d'ordre $(1 - \alpha/2)$ de la loi normale centrée réduite.

En pratique, il est possible que σ^2 ne soit pas connu, mais il est possible de l'estimer à l'aide d'une méthode de Monte-Carlo en utilisant le théorème suivant,

Théorème 6. Estimation de la variance

Si $\varphi(X_1) \dots \varphi(X_n)$ sont i.i.d avec $\mathbb{E}[\varphi(X)^2] \leq +\infty$ et $\text{var}[\varphi(X_1)] = \sigma^2$ alors,

$$\frac{1}{2n} [(\varphi(X_1) - \varphi(X_2))^2 + \dots + (\varphi(X_{2n-1}) - \varphi(X_{2n}))^2] \xrightarrow[n \rightarrow +\infty]{P.S.} \sigma^2.$$

Preuve 6. Pour une preuve de ce théorème voir [21].

Dans ce cas l'intervalle de confiance au niveau $(1 - \alpha)$ est donné par,

$$\left[\hat{I}_n - \phi^{-1}(1 - \alpha/2) \sqrt{\frac{\hat{\sigma}_n^2}{n}}, \hat{I}_n + \phi^{-1}(1 - \alpha/2) \sqrt{\frac{\hat{\sigma}_n^2}{n}} \right].$$

Où $\hat{\sigma}_n^2$ est un estimateur Monte-Carlo de la variance.

2.2 Méthodes de Monte-Carlo par chaîne de Markov

Dans la section précédente, nous avons vu que toute quantité de type $I = \mathbb{E}[\varphi(X)] = \int_{\mathbb{R}^d} \varphi(x) f(x) dx$, $d \in \mathbb{N}^*$, peut-être approchée par un estimateur Monte-Carlo ; $\hat{I}_n = \frac{1}{n} \sum_{i=1}^n \varphi(X_i)$, où la suite $(X_n)_{n \geq 1}$ est i.i.d $X_1 \sim \mathcal{L}(f)$ et $\mathbb{E}[\varphi(X)] < \infty$. Dans cette section, nous nous intéressons au cas où la suite $(X_n)_{n \geq 1}$ n'est pas i.i.d mais forme une chaîne de Markov de loi stationnaire f . Dans ce cas, nous parlerons des méthodes Monte-Carlo par chaînes de Markov (MCMC).

2.2.1 Généralités sur les chaînes de Markov

Soit $(X_n)_{n \geq 1}$ une suite de v.a à valeurs dans un ensemble S , supposé $S = \{1, 2, \dots, M\}$ un ensemble fini, ou dénombrable, appelé espace des états.

On dit que $(X_n)_{n \geq 1}$ est une chaîne de Markov (CM) si pour tout $n \geq 1$, et pour toute suite $(x_1, \dots, x_{n-1}, x, y)$ telle que,

$$\mathbb{P}(X_{n+1} = y \mid X_n = x, X_{n-1} = x_{n-1}, \dots, X_1 = x_1) > 0,$$

on a,

$$\begin{aligned} \mathbb{P}(X_{n+1} = y \mid X_n = x, X_{n-1} = x_{n-1}, \dots, X_1 = x_1) &= \mathbb{P}(X_{n+1} = y \mid X_n = x) \\ &= \mathbb{P}(X_2 = y \mid X_1 = x). \end{aligned}$$

Autrement dit, sachant le présent, le futur est indépendant du passé.

De plus si $\mathbb{P}(X_{n+1} = y \mid X_n = x) = \mathbb{P}(X_2 = y \mid X_1 = x)$ alors la chaîne de Markov $(X_n)_{n \geq 1}$ est dite homogène [4].

Probabilité de transition

On appelle probabilité de transition en une étape, la probabilité d'aller de l'état x vers l'état y la quantité,

$$p_{xy} = \mathbb{P}(X_{n+1} = y | X_n = x) = \mathbb{P}(X_2 = y | X_1 = x). \quad (2.6)$$

Matrice de transition (noyau, opérateur)

Soit $(X_n)_{n \geq 1}$ une chaîne de Markov à espace d'états S . On appelle matrice de transition ou matrice stochastique la matrice, $\mathbb{P} = (p_{xy})_{x,y \in S}$ qui vérifie les propriétés suivantes [22],

- Pour tout $x, y \in S$ $0 \leq p_{xy} \leq 1$.
- Pour tout $x \in S$ $\sum_{y \in S} p_{xy} = 1$ (somme des lignes).

Distribution stationnaire

Une probabilité π est appelée probabilité invariante ou probabilité stationnaire d'une chaîne de Markov de matrice de transition P si elle satisfait l'équation de balance globale,

$$\pi = \pi P. \quad (2.7)$$

Où $\pi = (\pi_1, \pi_2, \dots, \pi_x, \dots)$ avec $\sum_{x \in S} \pi_x = 1$.

Proposition 6. *Le fait qu'une chaîne de Markov homogène soit irréductible récurrente positive, c'est-à-dire ergodique assure l'existence d'une unique probabilité invariante (stationnaire).*

2.2.2 Propriétés fondamentales

Probabilités de transition en n étapes

On définit la probabilité de transition d'aller de l'état x vers l'état y en n coups par,

$$p_{xy}^{(n)} = \mathbb{P}(X_n = y | X_1 = x). \quad (2.8)$$

et la matrice de transition en n coups est notée,

$$P^n = (p_{xy}^{(n)})_{x,y \in S}. \quad (2.9)$$

Loi de probabilité de X_n

Soit X_1 la position initiale, X_1 peut-être décrite sous forme d'un vecteur ligne de taille M

$$\begin{aligned} \mu &= [\mu(1), \dots, \mu(i), \dots, \mu(M)] \\ &= [\mathbb{P}(X_1 = 1), \dots, \mathbb{P}(X_1 = i), \dots, \mathbb{P}(X_1 = M)]. \end{aligned}$$

Proposition 7. *Soit $(X_n)_{n \geq 1}$ une chaîne de Markov de loi initiale μ et de matrice de transition P . Alors pour tout entier naturel $n \geq 1$, la loi de X_n est donnée par*

$$\mathbb{P}(X_n) = \mu P^n. \quad (2.10)$$

2.2.3 Classification des chaînes de Markov

Soit $(X_n)_{n \geq 1}$ une chaîne de Markov d'espace d'états S .

Irréductibilité

On dit qu'une chaîne de Markov, ou sa matrice de transition, est irréductible si pour tous $x, y \in S$, la probabilité partant de x d'atteindre y est strictement positive, autrement dit : si pour tous $x, y \in S$, $\exists n \in \mathbb{N}$ telle que,

$$p_{xy}^{(n)} = \mathbb{P}(X_n = y | X_1 = x) > 0. \quad (2.11)$$

Autrement dit, une chaîne est dite irréductible si tous les états communiquent entre eux.

Récurrence et Transience

Soit $x \in S$, on définit le temps de retour à x lorsque la chaîne (X_n) part de x , par,

$$T_x = \inf \{n \geq 1, X_n = x\}. \quad (2.12)$$

Un état $x \in S$ est dit récurrent si partant de x on y revient sûrement en temps fini, i.e. :

$$\mathbb{P}(T_x < +\infty | X_1 = x) = 1. \quad (2.13)$$

L'état x est transitoire dans le cas contraire, i.e. :

$$\mathbb{P}(T_x = +\infty | X_1 = x) > 0. \quad (2.14)$$

Autrement dit, un état x est transitoire si avec une probabilité strictement positive, on peut le quitter pour ne jamais y revenir.

Périodicité

On définit la période d'un état x , $x \in S$ par,

$$d_x = \text{PGCD} \{n \geq 1, p_{xx}^{(n)} > 0\}. \quad (2.15)$$

Un état x est dit apériodique si sa période est 1, sinon il est périodique.

Une chaîne de Markov est dite apériodique si tous ses états sont apériodiques.

Ergodicité

Une chaîne de Markov homogène, irréductible, récurrente positive et apériodique est dite ergodique.

Absorption

Soit $x \in S$ on dit que x est un état absorbant si :

$$p_{xx} = 1. \quad (2.16)$$

Réversibilité(Symétrie)

On dit qu'une chaîne de Markov de matrice de transition P , est réversible(Symétrique) par rapport à la probabilité π si on a pour tous $x, y \in S$,

$$\pi_x p_{xy} = \pi_y p_{yx}. \quad (2.17)$$

En sommant (2.17) sur y on en déduit le lemme suivant.

Lemme 1. *Si une chaîne de Markov est réversible par rapport à la probabilité π , alors π est une probabilité invariante.*

Preuve 7. *Pour montrer que π est une probabilité invariante, il suffit de montrer que :*

$$\forall x \in S \quad \pi_x = \sum_{y \in S} \pi_y p_{yx}.$$

$$\begin{aligned} \sum_{y \in S} \pi_y p_{yx} &= \sum_{y \in S} \pi_x p_{xy}, \\ &= \sum_{y \in S} \pi_x \mathbb{P}(X_2 = y | X_1 = x), \\ &= \pi_x. \end{aligned}$$

□

2.2.4 Lois des grands nombres (convergence et vitesse de convergence)

Théorème 7. *(Convergence)*

Soit $(X_n)_{n \geq 1}$ une chaîne de Markov irréductible, admet une unique distribution stationnaire π , alors, pour toute fonction $\varphi : S \rightarrow \mathbb{R}$, on a :

$$\frac{1}{n} \sum_{i=1}^n \varphi(X_i) \xrightarrow[n \rightarrow +\infty]{p.s.} \mathbb{E}[\varphi(X)] = \sum_{x \in S} \varphi(x) \pi_x.$$

Dans le cas continu on a $\mathbb{E}[\varphi(X)] = \int_{x \in S} \varphi(x) \pi_{dx}$.

Preuve 8. *Pour une preuve, voir [3].*

Autrement dit, l'utilisation de ce théorème nous permet d'approcher une intégrale (cas continu) par une moyenne empirique, dans ce cas, la suite $(X_n)_{n \geq 1}$ n'est pas i.i.d mais constitue une chaîne de Markov.

Théorème 8. *Supposons que S est fini et P est une matrice de transition irréductible, apériodique et admettant une probabilité invariante π , alors :*

— *Il existe deux réels $\alpha \in [0, 1[$, $\beta \in \mathbb{R}$ tels que pour tout $x, y \in S$,*

$$|P_{xy}^n - \pi_y| \leq \beta \alpha^n.$$

— *Pour tout $x \in S$ et toute fonction $\varphi : S \rightarrow \mathbb{R}$, si on appelle $(X_n)_{n \geq 1}$ une chaîne de Markov de loi initiale π et de matrice de transition P alors,*

$$\sqrt{n} \left[\frac{1}{n} \sum_{i=1}^n \varphi(X_i) - \mathbb{E}[\varphi(X)] \right] \xrightarrow[n \rightarrow +\infty]{loi} \mathcal{N}(0, \sigma^2).$$

Avec une variance $\sigma^2 < \infty$.

Preuve 9. *Pour une preuve, voir [3].*

2.2.5 Méthodes de Monte-Carlo par chaîne de Markove (MCMC)

La méthode MCMC a été présentée pour la première fois en 1953 par Metropolis et al [16], avant d'être améliorée en 1970 par Hastings [12]. En 1984, les frères Geman [9], ont proposé l'échantillonneur de Gibbs pour la restauration bayésienne d'images, cet échantillonneur est développé par Gelfand et Smith. Au cours des années 90, après le développement et à la démocratisation de l'outil informatique, les méthodes de Monte-Carlo par chaîne de Markov ont rencontré un grand succès.

Ces méthodes consistent à créer une chaîne de Markov ergodique avec une loi stationnaire π , partant d'une valeur initiale arbitraire X_1 , on génère une chaîne $(X_n)_{n \geq 1}$ à partir d'un noyau de transition de loi stationnaire π , qui garantit de plus la convergence en loi vers π .

Dans la sous-section précédente, on a considéré une chaîne de Markov définie de manière quelconque, et nous avons cherché à déterminer sa loi stationnaire π . À présent, la démarche est inversée, nous partons d'une loi stationnaire π donnée et nous cherchons à construire une chaîne de Markov qui a π comme mesure stationnaire.

Parmi les méthodes MCMC, l'algorithme de Hastings-Metropolis (Hastings 1970) et l'algorithme de Gibbs (Geman et Geman 1984), sont les plus utilisés et ont donné lieu à de nombreux algorithmes dérivés.

Algorithme de Metropolis-Hastings

Les définitions présentées dans la section précédente nous ont permis de comprendre comment déterminer la loi stationnaire d'une chaîne de Markov donnée. Les algorithmes que nous allons présenter dans cette sous-section font l'inverse. C'est-à-dire que nous partons d'une loi stationnaire π donnée et nous cherchons à construire une chaîne de Markov dont π est sa distribution stationnaire.

Algorithme de Metropolis

L'algorithme de Metropolis est introduits par Metropolis et al 1953 [16]. Cet algorithme permet de créer une chaîne de Markov ergodique ayant une unique distribution stationnaire π . Pour ce faire, on fixe une loi π suivant laquelle on aimerait simuler ou dont on voudrait estimer une quantité de type $\int_S \varphi(x) \pi_{dx}$, nous appellerons π la loi cible.

Soit P la matrice de transition ayant π comme loi stationnaire.

Afin de construire une chaîne de Markov $(X_n)_{n \geq 1}$ on suit le même principe de la méthode de rejet (voir l'annexe A).

Soit donc $Q = (q_{xy})_{x,y \in S}$ une matrice de transition (appelée aussi matrice de proposition) telle que :

- Pour tout $(x, y) \in S^2$, on a $q_{xy} = q_{yx}$ (symétrie).
- Partant de tout état x , on sait simuler facilement selon la loi $q_x = [q_{x1}, \dots, q_{xM}]$.

Comme pour la méthode de rejet, la méthode de Metropolis fonctionne en plusieurs étapes.

1. Partant de $X_n = x$, simulons $Y = y$ selon la loi $q_x = [q_{x1}, \dots, q_{xM}]$.
2. Calculons le rapport d'acceptation rejet,

$$r_{xy} = \min \left\{ 1, \frac{\pi_y}{\pi_x} \right\}.$$

3. On tire une *v.a* qui suit loi uniforme $U \sim \mathcal{U}([0, 1])$ et on pose,

$$\begin{cases} X_{n+1} = Y = y & \text{si } U \leq r_{xy}, \\ X_{n+1} = X_n = x & \text{sinon.} \end{cases}$$

Donc l'algorithme de Metropolis s'écrit comme suit,

Algorithme Metropolis

Debut

lire (n); %nombre d'itérations;

lire($X(1)$); %la valeur initiale;

$i = 1$;

Tantque $i \leq (n - 1)$ **faire**

 générer $Y(i) \sim q_x$;

$r_{xy} = \min \left\{ 1, \frac{\pi_y(i)}{\pi_x(i)} \right\}$;

 générer $U \sim \mathcal{U}([0, 1])$;

si $U \leq r_{xy}$ **alors**

$X(i + 1) = Y(i)$;

sinon

$X(i + 1) = X(i)$;

fin si;

$i = i + 1$;

fin Tantque;

afficher (X);

Fin.

Dans la construction de la chaîne $(X_n)_{n \geq 1}$, il faut tout conserver, y compris les fois où l'on reste sur place ($X_{i+1} = X_i$).

Proposition 8. *La suite aléatoire $(X_n)_{n \geq 1}$ construite ici est une chaîne de Markov de matrice de transition P , définie par,*

$$\begin{cases} p_{xy} = q_{xy}r_{xy} & \text{si } x \neq y, \\ p_{xx} = 1 - \sum_{y \neq x} p_{xy} & \text{sinon.} \end{cases}$$

De plus π est une probabilité réversible (donc invariante) pour P .

Proposition 9. *Si la probabilité π est strictement positive et si la matrice de transition Q est telle que pour tout $x \neq y$, $Q \neq 0$, alors la matrice de transition P définie dans la proposition (8) est irréductible .*

Preuve 10. *si $x \neq y$, $p_{xy} = q_{xy}r_{xy} > 0 \Rightarrow$ tous les états communiquent.*

□

Proposition 10. *(Réversibilité de Metropolis) [11]*

Si la matrice de transition Q est irréductible, alors la chaîne $(X_n)_{n \geq 1}$ produite par l'algorithme de Metropolis est irréductible et π est réversible pour cette chaîne. En particulier, π est son unique mesure d'équilibre.

Preuve 11. Soit $P = (p_{xy})_{x,y \in S}$ la matrice de transition de la chaîne $(X_n)_{n \geq 1}$.
On veut montrer que,

$$\pi_x p_{xy} = \pi_y p_{yx},$$

pour tout $(x, y) \in S^2$, nous avons $q_{xy} = q_{yx}$ et supposons que $\pi_y \geq \pi_x$.
Nous avons d'une part,

$$p_{xy} = q_{xy} r_{xy} = q_{xy} \min \left\{ 1, \frac{\pi_y}{\pi_x} \right\} = q_{xy}.$$

D'autre part,

$$\begin{aligned} p_{yx} &= q_{yx} r_{yx} = q_{yx} \min \left(1, \frac{\pi_x}{\pi_y} \right), \\ &= q_{yx} \frac{\pi_x}{\pi_y} = q_{xy} \frac{\pi_x}{\pi_y}, \\ &= p_{xy} \frac{\pi_x}{\pi_y}. \\ \Rightarrow p_{yx} &= p_{xy} \frac{\pi_x}{\pi_y} \Rightarrow \pi_y p_{yx} = \pi_x p_{xy}. \end{aligned}$$

□

En 1970, Hastings a poursuivi les travaux de Metropolis en étendant l'application de son algorithme à des cas plus généraux. La généralisation de cet algorithme est présentée dans ce qui suit.

Généralisation méthode de Metropolis-Hastings

Dans la version originale de l'algorithme de Metropolis, la matrice de transition Q est symétrique, c'est pour quoi Hastings a proposé de généraliser l'algorithme de Metropolis. La méthode de Hastings fonctionne presque de la même façon que l'algorithme Metropolis.

Cas discret

- Partant de l'état x , on sait simuler suivant la loi $q_x = [q_{x1}, \dots, q_{xM}]$.
- Pour tout $(x, y) \in S^2$, on a $q_{xy} > 0$ implique $q_{yx} > 0$.
- pour tout $(x, y) \in S^2$, tel que $q_{xy} > 0$, on sait calculer $\frac{\pi_y q_{yx}}{\pi_x q_{xy}}$.

Donc l'algorithme de Hastings fonctionne comme suit,

1. Partant de $X_n = x$, simulons $Y = y$ selon la loi instrumentale $q_x = [q_{x1}, \dots, q_{xM}]$.
2. Calculons le rapport d'acceptation rejet,

$$r_{xy} = \min \left\{ 1, \frac{\pi_y q_{yx}}{\pi_x q_{xy}} \right\}.$$

3. On tire une suite de v.a.i.i.d $U \sim \mathcal{U}([0, 1])$ et on pose,

$$\begin{cases} X_{n+1} = Y = y & \text{si } U \leq r_{xy}, \\ X_{n+1} = X_n = x & \text{sinon.} \end{cases}$$

Alors l'algorithme de Metropolis-Hastings est donné comme suit,

Algorithme Metropolis-Hastings

Debut

lire (n); %nombre d'itérations;
lire($X(1)$); %la valeur initiale;
 $i = 1$;

Tantque $i \leq n - 1$ **faire**

générer $Y(i) \sim \mathcal{L}(q_x)$;
 $r_{xy} = \min \left\{ 1, \frac{\pi_y(i)q_{yx}(i)}{\pi_x(i)q_{xy}(i)} \right\}$;
générer $U \sim \mathcal{U}([0, 1])$;

si $U \leq r_{xy}$ **alors**

$X(i + 1) = Y(i)$;

sinon

$X(i + 1) = X(i)$;

fin si;

$i = i + 1$;

fin Tantque;

afficher (X);

Fin.

De même que la méthode de Metropolis, la suite aléatoire $(X_n)_{n \geq 1}$ ainsi générée est une chaîne de Markov de matrice de transition P avec,

$$\begin{cases} p_{xy} = q_{xy}r_{xy}, & \text{si } x \neq y, \\ p_{xx} = 1 - \sum_{y \neq x} p_{xy}, & \text{sinon.} \end{cases}$$

Par ailleurs, tout ce qui a été dit dans le cas d'un espace d'état discret se généralise à un espace d'état continu.

Cas continu

Dans le cas d'un espace d'états continu disons \mathbb{R}^d , $d \in \mathbb{N}^*$, la méthode pour choisir le noyau de transition à simuler est semblable à celle utilisée dans le cas discret. En effet, dans le cas discret, la loi cible, notée π , est remplacée dans le cas continu par la densité de probabilité f , et la distribution instrumentale q_{xy} est remplacée par $q(y|x)$. Le but est de simuler selon la densité $f(x)$. La quantité $q(y|x)$ s'interprète comme la densité de probabilité d'aller en y sachant qu'on part d'un point x .

Les hypothèses pour l'algorithme de Metropolis-Hastings dans le cas continu sont l'interprétation de celles du cas discret à savoir :

- Pour tout x , on sait simuler selon la densité $q(\cdot|x)$.
- Pour tout (x, y) , on a $q(y|x) > 0$ implique $q(x|y) > 0$.
- Pour tout (x, y) , on sait calculer $\frac{f(y)q(x|y)}{f(x)q(y|x)}$.

Alors l'algorithme de Metropolis-Hastings dans le cas continu s'écrit comme suit,

Algorithme Metropolis-Hastings

Debut

lire (n); %nombre d'itérations;

lire($X(1)$); %la valeur initiale;

$i = 1$;

Tantque $i \leq n - 1$ faire

générer $Y(i) \sim q(\cdot|X)$;

$r(X, Y) = \min \left\{ 1, \frac{f(Y(i))q(X(i)|Y(i))}{f(X(i))q(Y(i)|X(i))} \right\}$;

générer $U \sim \mathcal{U}([0, 1])$;

si $U \leq r(X, Y)$ **alors**

$X(i + 1) = Y(i)$;

sinon

$X(i + 1) = X(i)$;

fin si;

$i = i + 1$;

fin Tantque;

afficher (X);

Fin.

La méthode de Metropolis-Hastings est considérablement plus efficace en termes de calcul que la méthode d'acceptation-rejet (voir annexe A). Cependant, la rapidité avec laquelle elle converge vers la fonction à simuler dépend directement du choix de la fonction instrumentale q . En pratique, différents choix de q conduisent à plusieurs variantes de l'algorithme Metropolis-Hastings.

Typologies des algorithmes de Metropolis-Hastings

Il est important de noter que l'algorithme de Metropolis-Hastings a donné naissance à plusieurs autres méthodes, mais nous nous limiterons dans cette section à présenter les algorithmes les plus couramment rencontrés, à savoir,

- Algorithme de Metropolis-Hastings indépendant.
- Algorithme de Metropolis-Hastings à marche aléatoire.
- Algorithme de Metropolis-Hastings à une variable à la fois.

Algorithme de Metropolis-Hastings indépendant

L'algorithme de Metropolis-Hastings indépendant peut-être considéré comme une extension de l'algorithme d'acceptation-rejet, ce type d'algorithme repose sur l'utilisation d'une loi instrumentale q indépendant de X , cet algorithme peut s'écrire comme suit,

Algorithme Metropolis-Hastings indépendant

Debut

lire (n); %nombre d'itérations;

lire ($X(1)$); %la valeur initiale;

$i = 1$;

Tantque $i \leq n - 1$ **faire**

 générer $Y(i) \sim q(\cdot)$;

$r(X, Y) = \min \left\{ 1, \frac{f(Y(i))q(X(i))}{f(X(i))q(Y(i))} \right\}$;

 générer $U \sim \mathcal{U}([0, 1])$;

si $U \leq r(X, Y)$ **alors**

$X(i + 1) = Y(i)$;

sinon

$X(i + 1) = X(i)$;

fin si;

$i = i + 1$;

fin Tantque;

afficher (X);

Fin.

Même si les variables Y_i sont simulées de manière indépendante, l'échantillon obtenu n'est pas i.i.d, en particulier parce que la probabilité d'acceptation des Y_i dépend de X_i .

Algorithme de Metropolis-Hastings à marche aléatoire

Cet algorithme prend en considération la valeur précédemment simulée pour générer la valeur suivante, la loi instrumentale utilisée est une distribution qui satisfait la propriété $q(y|x) = q(y - x)$, En d'autres termes, Y_i peut s'écrire sous la forme $X_i + \varepsilon_i$, ε_i et générer par tirage indépendants d'une loi fixée facile à simuler (généralement uniforme ou normale), indépendante de X , ainsi la chaîne de Markov générée associée à q est une marche aléatoire. alors l'algorithme de Metropolis-Hastings à marche aléatoire s'écrit,

Algorithme Metropolis-Hastings à marche aléatoire

Debutlire (n); %nombre d'itérations;lire($X(1)$); %la valeur initiale; $i = 1$;**Tantque** $i \leq n - 1$ **faire**générer $\varepsilon(i) \sim q(\cdot)$; $Y(i) = X(i) + \varepsilon(i)$; $r(X, Y) = \min \left\{ 1, \frac{f(Y(i))q(X(i)-Y(i))}{f(X(i))q(Y(i)-X(i))} \right\}$;générer $U \sim \mathcal{U}([0, 1])$;**si** $U \leq r(X, Y)$ **alors** $X(i + 1) = Y(i)$;**sinon** $X(i + 1) = X(i)$;**fin si**; $i = i + 1$;**fin Tantque**;afficher (X);**Fin.**

Metropolis-Hastings à sauts réversibles

Appelé aussi à une variable à la fois, lorsque le paramètre à simuler est de grande dimension, il est souvent nécessaire de trouver une densité instrumentale multidimensionnelle pour générer une chaîne de Markov. Pour ce faire, on peut utiliser un algorithme dit à une variable à la fois. Le principe est de diviser le vecteur $X^{(n)}$ en plusieurs composantes $(X_1^{(n)}, X_2^{(n)}, \dots, X_d^{(n)})$ et de les simuler une par une. À chaque itération d'algorithme, on fait évoluer d composantes $X_i^{(n)}$ en utilisant d étapes de l'algorithme de Metropolis-Hastings, ce qui signifie que pour obtenir le nouveau vecteur $X^{(n+1)}$, il faudra utiliser d densités instrumentales $q_i(\cdot|x^{(n)})$, $i = 1, \dots, d$.

Pour simuler une composante $X_i^{(n)}$, il faut utiliser la loi instrumentale $q_i(y_i|x_{-i}^{(n)})$ et la loi cible $f_i(x_i|x_{-i}^{(n+1)})$ où $x_{-i}^{(n+1)} = (x_1^{(n+1)}, \dots, x_{i-1}^{(n+1)}, x_{i+1}^{(n+1)}, \dots, x_d^{(n+1)})$ (tel que $i = 1, \dots, d$). Alors l'algorithme de cette méthode peut s'écrire sous la forme suivante :

Algorithme Metropolis-Hastings à sauts réversibles

MH($f_1(x_1|x_{-1}^{n+1}), q_1$)MH($f_2(x_2|x_{-2}^{n+1}), q_2$)

⋮

MH($f_d(x_d|x_{-d}^{n+1}), q_d$)

Remarque 7. Les algorithmes de Metropolis-Hastings ne génèrent pas d'échantillon indépendant et identiquement distribué en particulier parce que la probabilité d'acceptation de Y_i dépend de X_i .

Remarque 8. Les algorithmes de Metropolis-Hastings peuvent-être utilisés pour simuler un vecteur aléatoire de dimension d en utilisant une densité instrumentale multidimensionnelle (Metropolis-Hastings à sauts réversibles). Mais quand d est grand, ce choix est rarement fait en pratique car la convergence d'un tel algorithme serait extrêmement lente. En effet, plus la dimension de l'espace des paramètres est grande, plus la proportion de candidats rejetés est importante. Dans ce cas il est préférable d'utiliser plutôt l'algorithme ou l'échantillonneur de Gibbs.

Échantillonneur de Gibbs

Il s'agit d'une variante spécifique de l'algorithme Metropolis-Hastings, où tous les états proposés sont acceptés. Cependant, cette approche nécessite une connaissance plus avancée de la loi cible f .

Considérons un espace d'état continu \mathbb{R}^d , $d \in \mathbb{N}^*$, supposons que la densité cible est donnée par $f(x) = f(x_1, x_2, \dots, x_d)$ pour tout indice j entre 1 et d , on note l'ensemble $x_{-j} = (x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_d)$. Dans ce cas x_{-j} correspond à x privé de sa $j^{\text{ème}}$ élément, donc la densité $f(x)$ peut-être donnée par $f(x) = f(x_j, x_{-j})$.

Tandis que $f(\cdot|x_{-j})$ est la densité conditionnelle correspondant à $(d-1)$ coordonnées gelées et $f(x)$ la densité marginale des ces $(d-1)$ coordonnées. Alors, pour tout j et pour toute suite x_{-j} , si on sait simuler suivant la densité conditionnelle $f(\cdot|x_{-j})$, alors :

1. Partant de $X_n = x$, on tire successivement des coordonnées j de 1 à d .
2. On simule Y selon la loi $f(\cdot|x_{-j})$.
3. On pose $X_{n+1} = (Y, X_{-j})$.

Donc l'algorithme de Gibbs s'écrit comme suit :

Algorithme de Gibbs

Debut

lire (n); %nombre d'itérations;

lire($X(1)$); %la valeur initiale;

$i = 1$;

Tantque $i \leq n - 1$ faire

générer $X_1^{i+1} \sim f_1(x_1|x_2^i, \dots, x_d^i)$;

générer $X_2^{i+1} \sim f_2(x_2|x_1^{i+1}, \dots, x_d^i)$;

⋮

générer $X_d^{i+1} \sim f_d(x_d|x_1^{i+1}, \dots, x_{d-1}^i)$;

$i = i + 1$;

fin Tantque;

afficher (X);

Fin.

La méthode que nous avons présentée ici est connue sous le nom d'échantillonnage de Gibbs par balayage déterministe (ou systématique) car l'indice j est choisi de manière séquentielle à chaque étape. Le balayage aléatoire consiste à sélectionner au hasard uniformément une coordonnée j entre 1 et d .

Remarque 9. *L'échantillonneur de Gibbs correspond à un algorithme de Metropolis-Hastings où toutes les transitions sont acceptées. C'est-à-dire $r(x, y) = 1$ à chaque étape.*

2.2.6 Les performances des méthodes de MCMC

Les performances des méthodes de Monte-Carlo par chaîne de Markov sont fortement influencé par les choix effectués par l'utilisateur. En effet, en plus du choix du point de départ (la valeur de l'état initial X_1), l'utilisateur doit également sélectionner une fonction instrumentale dans laquelle des échantillons sont générés afin de mieux s'ajuster à la distribution cible. Donc afin d'illustrer l'effet de ces variantes, des échantillons de variable aléatoire distribuées selon la loi de mélanges suivante,

$$f(x) = (1/3) \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} + (2/3) \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-7)^2}, \quad (2.18)$$

simulé par l'algorithme de Metropolis-Hastings avec différents choix de paramètres seront présentés à titre d'exemple.

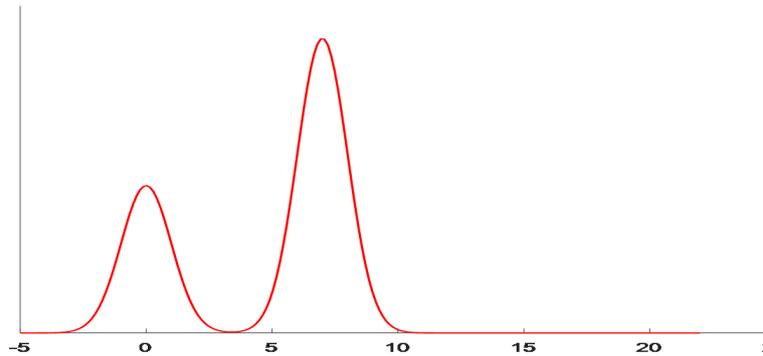


FIGURE 2.2 – Distribution de la loi de mélange $f(x)$.

Le premier paramètre à choisir est la distribution instrumentale qui a un impact considérable sur les performances de l'algorithme. Nous détaillons ceci dans la sous-section suivante.

Distribution instrumentale

Le choix de la distribution instrumentale (ou de proposition) est essentiel mais pose un défi, car il déterminera la qualité de l'algorithme de manière significative. En effet, Un choix judicieux permettra une exploration rapide de l'espace des états, ce qui accélérera la convergence de la chaîne vers la distribution cible. Donc pour des raisons de performance, on choisira évidemment une distribution q facile à simuler et surtout proche le plus possible de la densité cible. La figure (2.3) illustre l'impact du choix de la densité de proposition sur la simulation de la loi cible f .

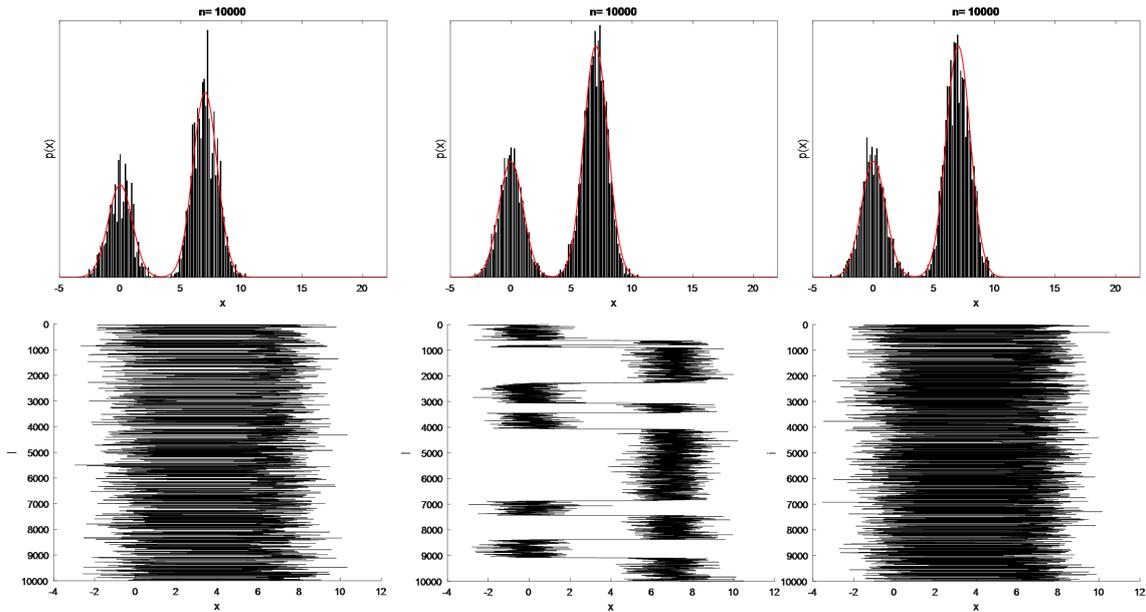


FIGURE 2.3 – Simulations de 10000 itérations avec $q(y|x) \sim \mathcal{N}(x, 5)$, $q(y|x) \sim \mathcal{N}(x, 1)$ et $q(y|x) \sim \mathcal{U}([x - 10, x + 1])$ respectivement (de gauche à droite).

À partir de la Figure (2.3), il est clair que le meilleur choix pour la loi de proposition est la loi normale $\mathcal{N}(x, 1)$, comme indiqué par le graphe du milieu. En effet, pour $n = 10000$, la loi normale $\mathcal{N}(x, 1)$ offre de meilleures approximations de la densité cible par rapport aux deux autres choix (la loi $\mathcal{N}(x, 5)$ et la loi $\mathcal{U}([x - 10, x + 1])$).

Après le choix de la distribution instrumentale, vient ensuite le choix de la valeur de l'état initial de la chaîne à générer. Ce choix aura une influence sur la vitesse de convergence de l'algorithme.

État initial

L'exécution des algorithmes de MCMC nécessite de choisir l'état initial dans lequel se trouve la chaîne. Cette valeur initiale, également appelée la graine et notée X_1 , exerce une influence significative sur la vitesse de convergence de l'algorithme. Par conséquent, il est crucial de la choisir judicieusement. Généralement, ce choix est fait au hasard, ou bien uniformément sur le support de définition (région d'intérêt) de la loi cible f , mais lorsque nous disposons de certaines informations sur la fonction cible, il est préférable de choisir la valeur de l'état initial en fonction de ces informations. La Figure (2.4), présente les résultats de trois simulations avec des valeurs initiales respectives de $X_1 = -20$, $X_1 = 0$ et $X_1 = 20$. Les graphiques de cette figure démontrent le comportement de la fonction instrumentale en fonction du choix de la valeur de l'état initial.

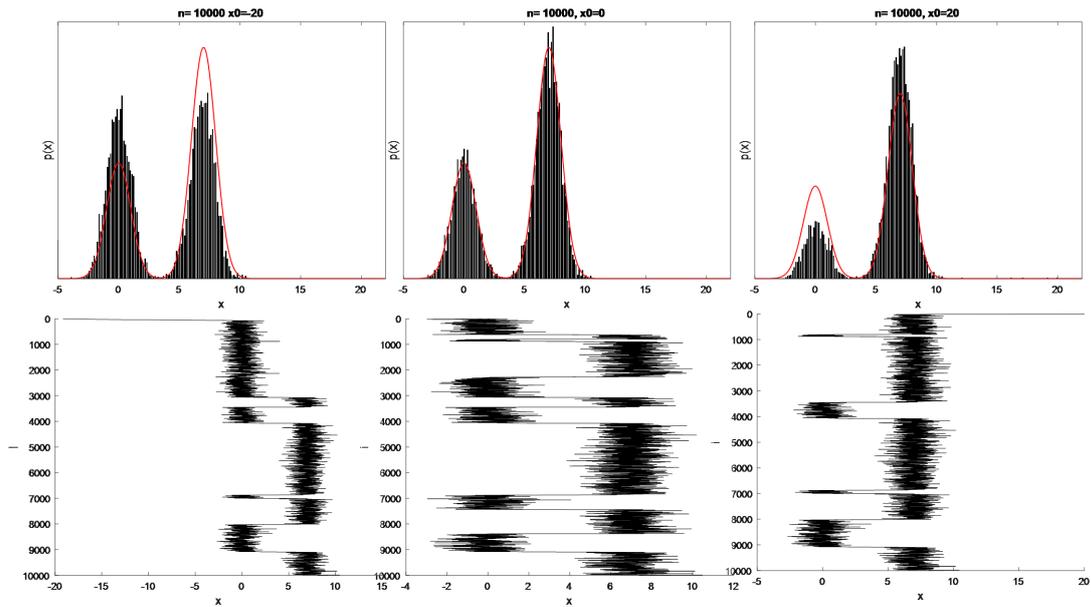


FIGURE 2.4 – Simulations de 10000 itérations avec respectivement $X_1 = -20$, $X_1 = 0$ et $X_1 = 20$.

En analysant la figure (2.4) et en comparant les résultats obtenus, on observe que la valeur optimale de l'état initial à choisir est $X_1 = 0$ (représentée par le graphique du milieu). Pour cette valeur et avec un nombre d'itérations de $n = 10000$, la chaîne simulée converge rapidement vers la distribution cible. En revanche, les valeurs $X_1 = -20$ et $X_1 = 20$ nécessitent plus de temps pour atteindre la région d'intérêt, ce qui ralentit la convergence de la chaîne vers la distribution cible.

Toute fois, pour réduire au maximum l'influence du choix de X_1 sur la simulation de la densité cible, il est d'usage de ne pas tenir compte d'un certain nombre d'itérations, ceci signifie que les premières valeurs générées par l'algorithme ne seront pas prises en compte. Les auteurs anglophones font alors référence à la période de "burn in" ou "période de chauffe".

Période de chauffe (période de transition)

Toute simulation MCMC peut-être divisée en deux périodes distinctes. La première est la période de transition, qui correspond aux premières itérations dépendant du choix de la valeur de l'état initial. Ces itérations sont souvent très instables. La seconde période est la période de simulation, où la chaîne atteint progressivement sa distribution stationnaire f . En effet, déterminer le nombre d'itérations nécessaire pour la période de chauffe, notée n_0 , est souvent difficile. Il revient donc à l'utilisateur de l'algorithme de vérifier empiriquement si la période de transition est terminée et s'il est donc possible de commencer l'échantillonnage de la distribution stationnaire f . La Figure (2.5), illustre les résultats de trois simulations de la distribution (2.5). La première simulation est réalisée sans période de chauffe, la deuxième simulation utilise une période de transition de $n_0 = 500$, et enfin la troisième simulation utilise une période de transition de $n_0 = 1000$.

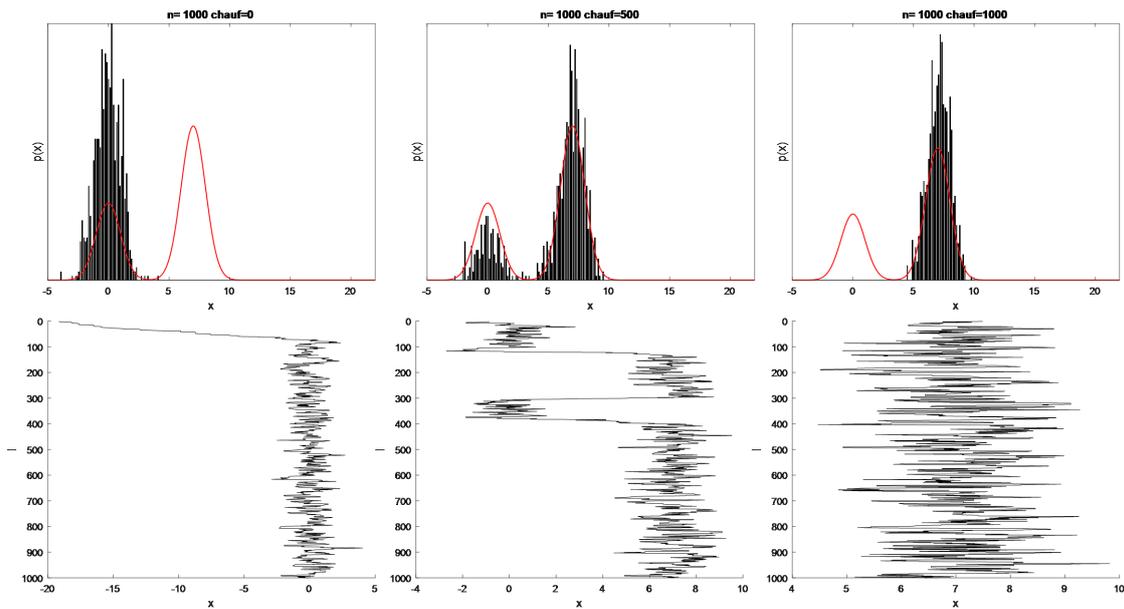


FIGURE 2.5 – Distribution et histogramme ainsi la série chronologique de trois simulations avec respectivement $n_0 = 0$, $n_0 = 500$ et $n_0 = 1000$.

Le premier graphique de la Figure (2.5), illustre la simulation de la densité f sans utilisation de période de chauffe. On peut observer que la plupart des propositions ne tombent pas dans la queue de la région d'intérêt de f . Cela est dû à l'influence de la valeur de l'état initial, qui est éloignée de la région d'intérêt. Le deuxième graphe représente la simulation de f avec la même valeur de l'état initial que dans le premier cas, mais avec une période de chauffe $n_0 = 500$. On peut observer que l'échantillon généré enveloppe bien le support de définition de f , et on remarque que l'influence de la valeur de l'état initial est réduite en utilisant une période de chauffe. Dans le troisième cas, nous avons utilisé une période de chauffe plus longue avec $n_0 = 1000$. D'après le graphique, on constate que la période de chauffe est dépassée, ce qui entraîne une perte d'information sur la loi cible, donc il est inutile de prolongée la période de chauffe.

En effet, cet exemple met en évidence l'importance de la période de chauffe. Elle permet de réduire l'influence du choix de la valeur de l'état initial X_1 sur la simulation de la densité cible. En prolongeant la période de chauffe, la chaîne MCMC a la possibilité de s'ajuster progressivement et de se rapprocher de l'équilibre. Cela contribue à obtenir des estimations plus fiables et précises de la densité cible.

2.3 Conclusion

Dans ce chapitre, nous avons exploré les méthodes de Monte-Carlo, en accordant une attention particulière à l'intégration Monte-Carlo et aux méthodes de Monte-Carlo par chaîne de Markov (MCMC). Nous avons examiné de manière détaillée deux approches clés permettant de construire des chaînes de Markov avec une distribution stationnaire spécifiée : les algorithmes de Metropolis-Hastings et l'échantillonneur de Gibbs.

Dans le chapitre suivant, nous présenterons des exemples concrets d'application de ces méthodes à la statistique bayésienne, illustrant ainsi leur pertinence et leur efficacité dans des contextes réels.

Chapitre 3

Mise en pratique des statistiques bayésiennes et des méthodes de MCMC

Les statistiques bayésiennes et les méthodes de MCMC sont étroitement liées et souvent utilisées conjointement dans de nombreuses applications. Ensemble, elles offrent une approche puissante pour l'analyse statistique et l'inférence dans des problèmes complexes.

Les statistiques bayésiennes fournissent un cadre probabiliste pour modéliser l'incertitude à l'aide de distributions de probabilité. Elles permettent d'intégrer les connaissances a priori sur les paramètres inconnus avec les données observées afin d'obtenir une distribution a posteriori mise à jour qui reflète la connaissance actuelle après avoir vu les données. L'inférence bayésienne repose sur le théorème de Bayes pour calculer la distribution a posteriori en utilisant la distribution a priori, la vraisemblance des données et une constante de normalisation.

Les méthodes de MCMC, telles que l'échantillonneur de Gibbs et l'algorithme Metropolis-Hastings, sont des outils d'échantillonnage utilisés en statistiques bayésiennes pour explorer la distribution a posteriori des paramètres. Ces méthodes permettent de générer des échantillons à partir de la distribution cible en construisant une chaîne de Markov dont la distribution d'équilibre est la distribution cible souhaitée. Les échantillons de la chaîne de Markov convergent vers la distribution cible, permettant ainsi d'obtenir des estimations des paramètres basées sur la distribution a posteriori.

L'utilisation conjointe des statistiques bayésiennes et des méthodes de MCMC permet de résoudre des problèmes statistiques complexes dans de nombreux domaines. Les applications sont vastes et incluent des domaines tels que :

1. **Modélisation statistique** : Ils permettent de modéliser des phénomènes complexes en intégrant des connaissances a priori et en mettant à jour les estimations à mesure que de nouvelles données sont disponibles.
2. **Prévision** : Ils peuvent être utilisés pour estimer les paramètres de modèles prédictifs et pour effectuer des prédictions basées sur ces modèles. Cela inclut des domaines tels que la prévision du temps etc.
3. **Économie et finance** : La statistique bayésienne est largement utilisée dans l'analyse économique et financière pour estimer les paramètres des modèles économétriques, évaluer les risques et effectuer des prévisions.

4. **Science sociale** : Ils sont utilisés pour analyser les enquêtes et les données d'études sociales, modéliser les comportements et les préférences, et faire des prédictions sur les attitudes et les opinions.
5. **Science de la santé** : Ils sont utilisés dans la modélisation et l'analyse de données cliniques pour estimer l'efficacité des traitements, évaluer les risques et prédire les résultats cliniques.
6. **Écologie et biologie** : Les méthodes de MCMC sont largement utilisées en écologie et en biologie pour estimer la taille de la population d'organismes. L'estimation de la taille de la population est essentielle pour comprendre la dynamique des populations, évaluer l'impact des perturbations environnementales, et élaborer des stratégies de conservation. Il convient de noter que la statistique bayésienne et les algorithmes MCMC ne sont pas limités à ces domaines spécifiques, mais peuvent être appliqués dans de nombreux autres domaines où l'analyse statistique est requise et où l'incertitude est une considération importante.

Dans ce chapitre, nous donnons deux applications distinctes qui illustrent l'utilité de l'approche bayésienne avec les méthodes MCMC dans des domaines différents, nous étudions comment les méthodes MCMC peuvent être adaptées et appliquées dans l'estimation bayésienne avec succès pour résoudre des problématiques spécifiques dans chacun de ces domaines. La première partie de l'application est consacrée à la régression bayésienne avec les méthodes MCMC dans le domaine médical, plus précisément sur l'étude des crises cardiaques chez les hommes de 55 à 59 ans dans différents pays. La deuxième partie de l'application, est consacrée à l'estimation de la taille de la population des poissons et de leur poids moyen dans un lac en traitant plusieurs scénarios.

3.1 Régression bayésienne

La régression en statistique bayésienne est une approche qui permet de modéliser et d'analyser les relations entre des variables en utilisant les principes de l'inférence bayésienne. Contrairement à la régression classique, qui fournit une estimation ponctuelle des paramètres du modèle, la régression bayésienne fournit une distribution de probabilité pour chaque paramètre, ce qui permet de prendre en compte l'incertitude associée à ces paramètres.

3.1.1 Définitions et domaines d'applications

La régression bayésienne est l'une des méthodes statistiques, qui est largement utilisée dans divers domaines. Voici quelques exemples de domaines d'application où la régression bayésienne peut être utilisée :

1. **Sciences médicales** : La régression bayésienne peut être utilisée pour modéliser les relations entre les variables médicales telles que l'âge, les facteurs de risque, les résultats des tests diagnostiques et les résultats cliniques. Elle peut aider à prédire des résultats tels que le risque de maladie, la survie des patients, l'efficacité des traitements, etc.
2. **Sciences sociales** : En sciences sociales, la régression bayésienne peut être utilisée pour étudier les relations entre les variables socio-économiques, démographiques, psycho-

logiques, etc. Elle peut aider à comprendre l'impact des variables sur les comportements, les attitudes, les opinions, etc.

3. **Économie et finance** : La régression bayésienne est utilisée dans l'analyse économique et financière pour modéliser les relations entre les variables macroéconomiques, les données financières et les marchés. Elle peut-être utilisée pour prévoir des variables économiques, estimer des relations causales, évaluer des politiques économiques, etc.
4. **Sciences de l'environnement** : Dans le domaine de l'environnement, la régression bayésienne peut-être utilisée pour étudier les relations entre les variables environnementales telles que la pollution, le climat, les caractéristiques du sol, etc., et les impacts sur les écosystèmes, la biodiversité, la santé humaine, etc.

3.1.2 Approche méthodologique

1. **Spécification du modèle** : On définit la structure du modèle de régression, y compris les variables indépendantes, la relation fonctionnelle entre les variables et les paramètres inconnus à estimer. Par exemple, dans une régression linéaire, le modèle peut-être défini comme $Y = \alpha + \beta X + \epsilon$, où Y est la variable dépendante, X est la variable indépendante, α et β sont les coefficients de régression à estimer, et ϵ est l'erreur résiduelle.
2. **Spécification des distributions a priori** : On attribue les distributions de probabilité a priori aux paramètres du modèle, qui expriment les connaissances ou les croyances initiales sur ces paramètres avant d'avoir accès aux données observées. Les distributions a priori peuvent être choisies de manière subjective ou en se basant sur des connaissances préalables ou des études antérieures.
3. **Estimation des paramètres** : À l'aide de méthodes d'inférence bayésienne telles que l'échantillonnage de Monte-Carlo par chaînes de Markov (MCMC), estimez les distributions a posteriori des paramètres en combinant les distributions a priori et les données observées. Cela implique de générer un grand nombre d'échantillons des paramètres à partir des distributions a posteriori.
4. **Analyse des résultats** : On analyse les distributions a posteriori des paramètres pour obtenir des informations sur les estimations, les intervalles de crédibilité et les mesures d'incertitude.

3.1.3 Exemple 1 : Régression simple

Dans cet exemple nous proposons d'appliquer la régression bayésienne dans le domaine médicale, plus précisément pour l'étude des crises cardiaques qui est directement responsable d'environ 10% de toutes les morts chaque année. Les données proposées présentent le taux de décès par attaque cardiaque chez les hommes de 55 à 59 ans dans différents pays (voir tableau (3.1)) [2].

L'objectif est de modéliser la relation entre X et Y tel que le couple (x_i, y_i) , $i = 1, \dots, n$ sont les données où la variable prédictive X et la variable de réponse Y sont indépendantes. Dans ce cas les variables sont les suivantes :

Y : taux de décès = $100 [\log(\text{nombre de décès par crise cardiaque pour 100000 hommes de 55 à 59 ans}) - 2]$.

X : calories provenant de protéines animales en pourcentage du total des calories.

| Observation | Pays | X | Y | Observation | Pays | X | Y |
|-------------|-----------|-----|-----|-------------|------------|-----|-----|
| 1 | Australie | 8 | 81 | 12 | Italie | 3 | 45 |
| 2 | Autriche | 6 | 55 | 13 | Japon | 3 | 24 |
| 3 | Canada | 8 | 80 | 14 | Mexique | 3 | 43 |
| 4 | Ceylan | 2 | 24 | 15 | Pays-Bas | 6 | 38 |
| 5 | Chili | 4 | 78 | 16 | N-Zélande | 8 | 72 |
| 6 | Danemark | 6 | 52 | 17 | Norvège | 6 | 41 |
| 7 | Finlande | 7 | 88 | 18 | Portugal | 4 | 38 |
| 8 | France | 7 | 45 | 19 | Suède | 7 | 52 |
| 9 | Allemagne | 6 | 50 | 20 | Suisse | 7 | 52 |
| 10 | Irlande | 5 | 69 | 21 | G-Bretagne | 6 | 66 |
| 11 | Brazile | 4 | 66 | 22 | Etats-Unis | 8 | 89 |

TABLE 3.1 – Le taux de décès par attaque cardiaque chez les hommes de 55 à 59 ans et les calories provenant de protéines animales.

L'objectif est d'estimer les coefficients de l'équation de régression définie par,

$$y_i = \alpha + \beta x_i + \epsilon_i. \quad (3.1)$$

Où les coefficients α et β sont respectivement l'ordonnée à l'origine et la pente de la droite de régression et ϵ_i sont indépendantes identiquement distribuées (iid), $\epsilon_i \sim \mathcal{N}(0, \sigma^2)$.

Dans ce qui suit, nous commençons par utiliser la méthode analytique, puis nous passons à la méthode de Monte-Carlo et aux méthodes de Monte-Carlo par chaîne de Markov. Enfin, nous effectuons une comparaison entre les différentes méthodes.

A. Méthode analytique

Pour la méthode analytique et afin d'estimer les paramètres des coefficients de régression α et β , on utilise la méthode des moindres carrés. Par l'application de la méthode des moindres carrés une équation alternative à l'équation (3.2) est donnée par (voir Annexe B) :

$$y_i = \alpha_{\bar{x}} + \beta(x_i - \bar{x}) + \epsilon_i \quad i = 1, \dots, n, \quad (3.2)$$

où,

$\alpha_{\bar{x}}$ la valeur moyenne pour $y(\bar{x} = x)$,

β la pente,

ϵ_i l'erreur.

Pour toutes les observations $y_i, i = 1, \dots, n$, les erreurs sont *iid* et,

$$\epsilon_i \sim \mathcal{N}(0, \sigma^2),$$

$$y_i \sim \mathcal{N}(\alpha_{\bar{x}} + \beta(x_i - \bar{x}), \sigma^2).$$

À partir de la fonction de vraisemblance de la i ème observation qui est donnée par :

$$f_i(\alpha_{\bar{x}}, \beta) \propto \exp\left(-\frac{1}{2\sigma^2}(y_i - (\alpha_{\bar{x}} + \beta(x_i - \bar{x})))^2\right),$$

il est facile de montrer (voir Annexe B) que,

$$f(\alpha_{\bar{x}}, \beta) \propto f(\alpha_{\bar{x}})f(\beta). \quad (3.3)$$

$$\alpha_{\bar{x}} \sim \mathcal{N}(A_{\bar{x}}, \frac{\sigma^2}{n}),$$

$$\beta \sim \mathcal{N}(T, \frac{\sigma^2}{SS_x}),$$

où

$$\begin{cases} A_{\bar{x}} = \bar{y}, \\ T = \frac{SS_{xy}}{SS_x}, \\ SS_x = \sum_{i=1}^n (x_i - \bar{x})^2, \\ SS_{xy} = \sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x}). \end{cases}$$

Sachant que les deux coefficients de régression $\alpha_{\bar{x}}$ et β sont indépendants, alors la loi a priori conjointe de ces deux la est donnée par :

$$\pi(\alpha_{\bar{x}}, \beta) = \pi(\alpha_{\bar{x}})\pi(\beta)$$

Dans ce cas la loi a posteriori conjointe est donnée par :

$$\pi(\alpha_{\bar{x}}, \beta|y) = \pi(\alpha_{\bar{x}}|y)\pi(\beta|y). \quad (3.4)$$

Telle que,

$$\begin{cases} \pi(\beta|y) \propto \pi(\beta)f(\beta). \\ \pi(\alpha_{\bar{x}}|y) \propto \pi(\alpha_{\bar{x}})f(\alpha_{\bar{x}}). \end{cases} \quad (3.5)$$

Il est évident que si la loi a priori $\pi(\beta) \sim \mathcal{N}(\mu_{\beta}, \sigma_{\beta}^2)$ donc la loi a posteriori

$$\pi(\beta|y) \sim \mathcal{N}(\mu'_{\beta}, (\sigma'_{\beta})^2), \quad (3.6)$$

(de même pour le paramètre $\alpha_{\bar{x}}$).

Pour le calcul des paramètres μ'_{β} et σ'_{β} voir l'Annexe B.

Pour une application, on trouve,

$$\hat{Y} = \hat{\alpha} + \hat{\beta}X = 20.74627 + 6.391791X.$$

Les taux de décès estimés (\hat{Y}) par l'application de la méthode analytique sont résumés dans le tableau (3.2).

| Observation | \hat{Y} | Observation | \hat{Y} |
|-------------|-----------|-------------|-----------|
| 1 | 71.88060 | 12 | 39.92164 |
| 2 | 59.09701 | 13 | 39.92164 |
| 3 | 71.88060 | 14 | 39.92164 |
| 4 | 33.52985 | 15 | 59.09701 |
| 5 | 46.31343 | 16 | 71.88060 |
| 6 | 59.09701 | 17 | 59.09701 |
| 7 | 65.48881 | 18 | 4631343 |
| 8 | 65.48881 | 19 | 65.48881 |
| 9 | 59.09701 | 20 | 65.48881 |
| 10 | 52.70522 | 21 | 59.09701 |
| 11 | 46.31343 | 22 | 71.88060 |

TABLE 3.2 – Les taux de décès estimés (\hat{Y}) par l’application de la méthode analytique.

La figure (3.1), représente le nuage de point $(x_i, y_i), i = 1, \dots, n$ et la droite de régression.

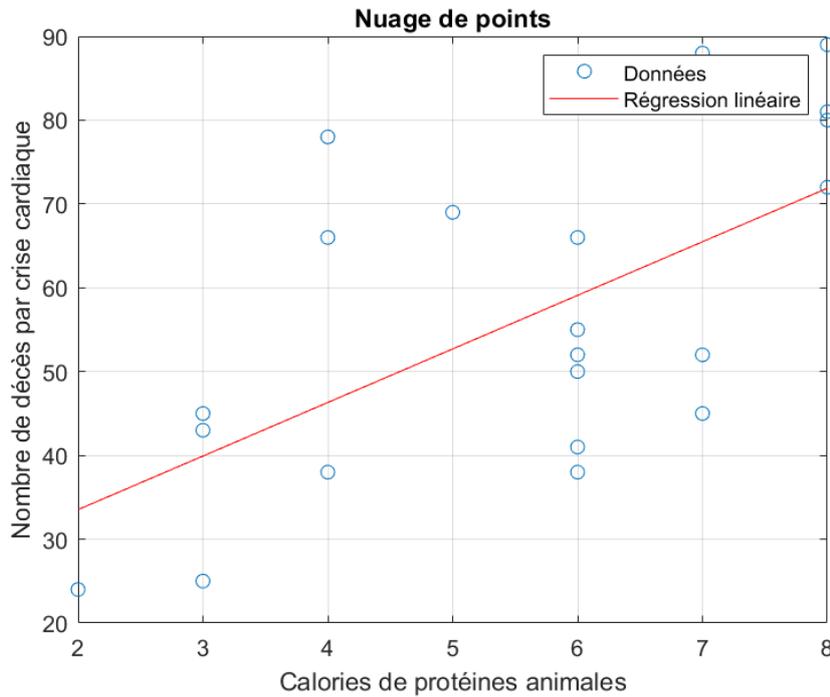


FIGURE 3.1 – Nuage de point $(x_i, y_i), i = 1, \dots, n$ et la droite de régression par l’application de la méthodes analytique.

Dans tout ce qui suit, nous supposons que $\beta \sim \mathcal{N}(1.02, (0.25)^2)$ et $\alpha_{\bar{x}} \sim \mathcal{N}(15, (1)^2)$. Dans ce cas on trouve,

La loi a posteriori $\pi(\beta|y) \sim \mathcal{N}(1.101171, (0.24764)^2)$.

La loi a posteriori $\pi(\alpha_{\bar{x}}|y) \sim \mathcal{N}(18.5226, (0.9569)^2)$.

La figure (3.2), représente la distribution a priori et a posteriori de la pente β .

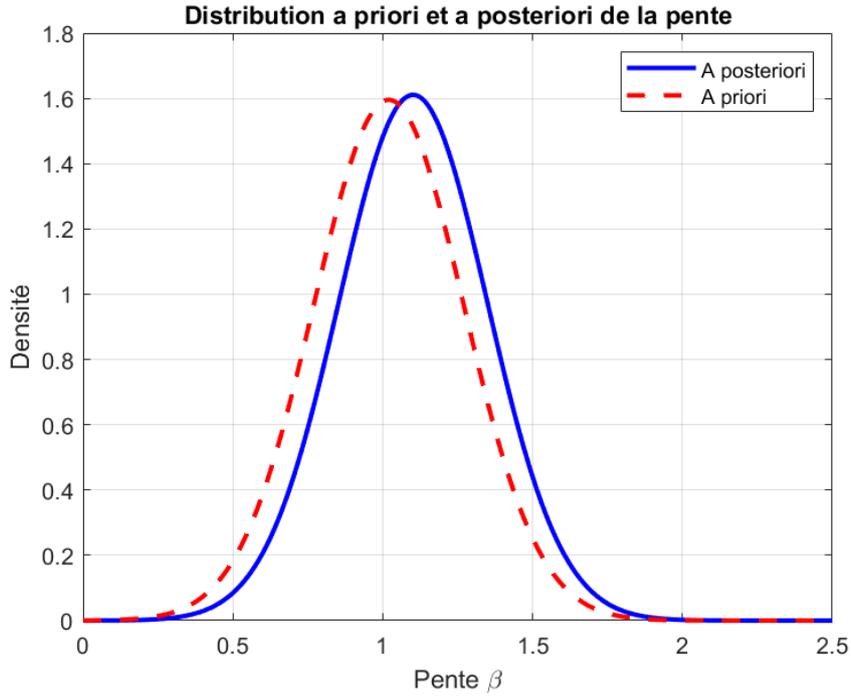


FIGURE 3.2 – La représentation graphique de la distribution a priori et a posteriori de la pente.

À partir de la Figure (3.2), on peut constater que la distribution a posteriori de la pente est cohérente avec la distribution a priori (loi normale).

B. Méthode Monte-Carlo

Cette fois ci, nous utilisons une méthode de Monte-Carlo simple pour estimer les paramètres de l'équation de régression (3.2) et les taux de décès estimés (\hat{Y}). À partir de ce qui précède (équation (3.6)), on a vu que ;

$$\beta|y \sim \mathcal{N}(\mu'_\beta, (\sigma'_\beta)^2), \quad (3.7)$$

et

$$\alpha_{\bar{x}}|y \sim \mathcal{N}(\mu'_{\alpha_{\bar{x}}}, (\sigma'_{\alpha_{\bar{x}}})^2). \quad (3.8)$$

Donc afin d'estimer les paramètres de régression $\alpha_{\bar{x}}$ et β , on doit suivre les étapes suivantes :

1. On calcule μ'_β et σ'_β à partir de la relation (B.4),
2. On calcule $\mu'_{\alpha_{\bar{x}}}$ et $\sigma'_{\alpha_{\bar{x}}}$ à partir de la relation (B.5),
3. On génère un vecteur de variables aléatoires β selon $\mathcal{N}(\mu'_\beta, (\sigma'_\beta)^2)$,
4. De même pour $\alpha_{\bar{x}}$, on génère un vecteur de variables aléatoires $\alpha_{\bar{x}}$ selon $\mathcal{N}(\mu'_{\alpha_{\bar{x}}}, (\sigma'_{\alpha_{\bar{x}}})^2)$,
La figure (3.3) et (3.4), représentent l'histogramme et la densité des échantillons du coefficient de régression β .

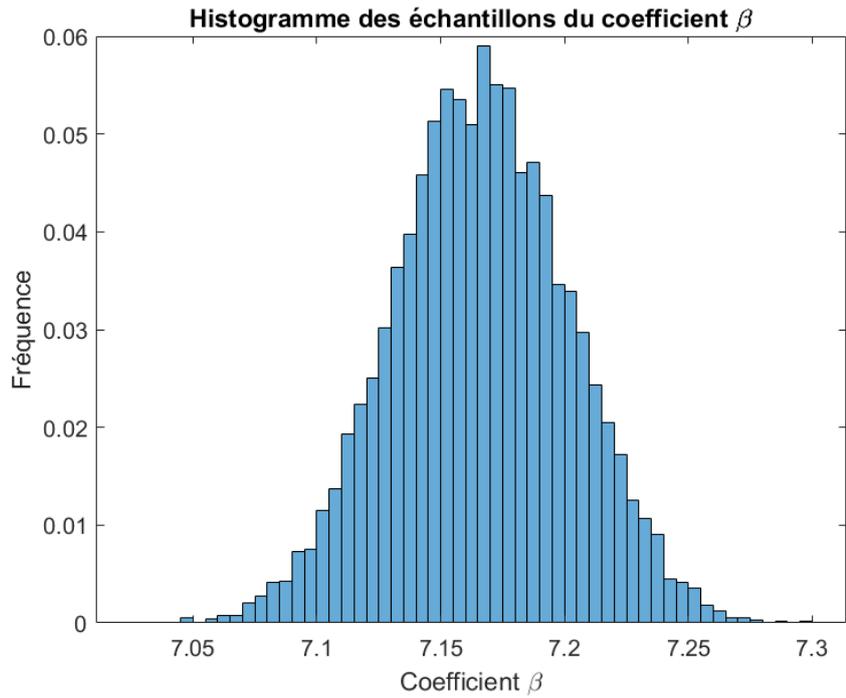


FIGURE 3.3 – La représentation graphique de l’histogramme des échantillons du coefficient de régression β .

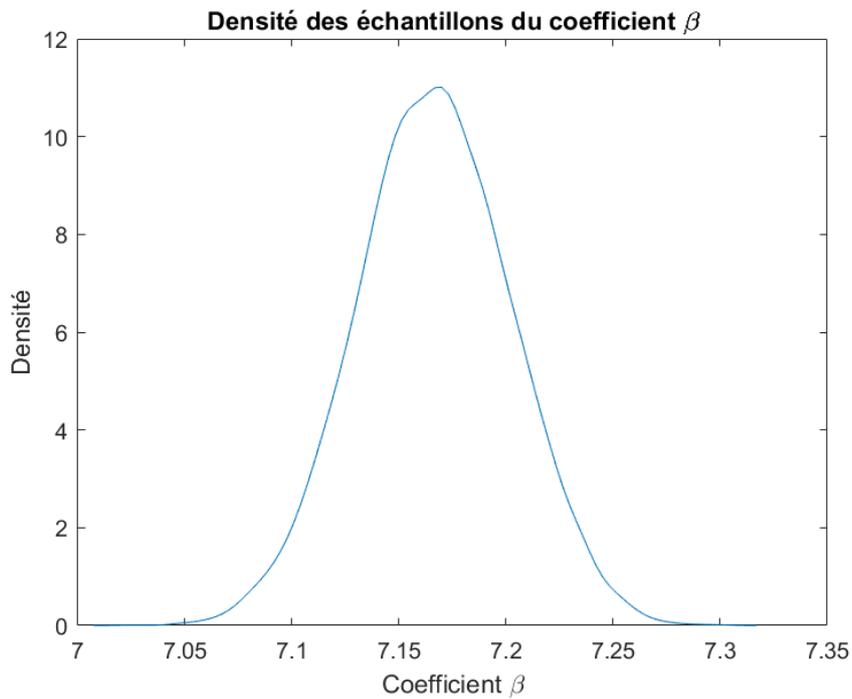


FIGURE 3.4 – La représentation graphique de la densité des échantillons du coefficient de régression β .

A partir de ces deux figures on peut voir que la densité des échantillons du coefficient de régression β présente une forme de courbe en cloche symétrique, avec une dispersion

équilibrée autour de sa moyenne. Ces caractéristiques suggèrent fortement une ressemblance avec la distribution d'une loi normale.

5. Pour finir, les valeurs des coefficients de régressions sont calculés comme la moyenne des deux vecteurs générés $\alpha_{\bar{x}}$ et β . Pour $n = 10000$ itérations on obtient $\beta = 7.1665$ avec un intervalle de confiance $[7.0952, 7.2365]$ et $\alpha = 14.9889$ avec un intervalle de confiance $[13.0177, 16.9285]$.

Les taux de décès estimés (\hat{Y}) par la méthode Monte-Carlo, sont résumés dans le tableau (3.3).

| Observation | \hat{Y} | Observation | \hat{Y} |
|-------------|-----------|-------------|-----------|
| 1 | 72.32057 | 12 | 36.48825 |
| 2 | 57.98764 | 13 | 36.48825 |
| 3 | 72.32057 | 14 | 36.48825 |
| 4 | 29.32178 | 15 | 57.98764 |
| 5 | 43.65471 | 16 | 72.32057 |
| 6 | 57.98764 | 17 | 57.98764 |
| 7 | 65.15411 | 18 | 43.65471 |
| 8 | 65.15411 | 19 | 65.15411 |
| 9 | 57.98764 | 20 | 65.15411 |
| 10 | 50.82118 | 21 | 57.98764 |
| 11 | 43.65471 | 22 | 72.32057 |

TABLE 3.3 – Les taux de décès estimés (\hat{Y}) par l'application de méthode de Monte-Carlo.

La figure (3.5), donne la représentation graphique du nuage de point $(x_i, y_i), i = 1, \dots, n$ et la droite de régression obtenue par l'application de méthode de Monte-Carlo.

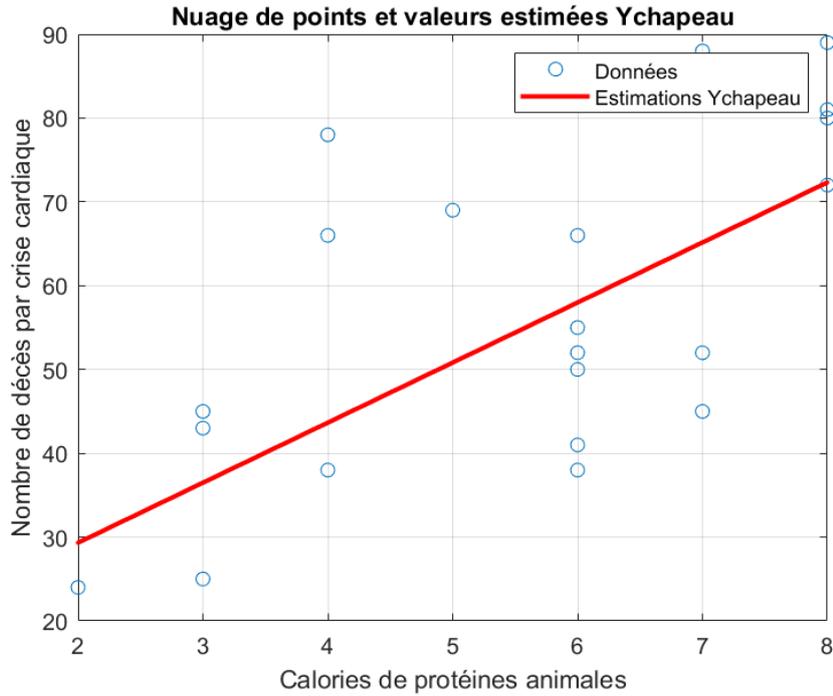


FIGURE 3.5 – Nuage de point $(x_i, y_i), i = 1, \dots, n$ et la droite de régression obtenue par la méthode MC.

C. Méthode Monte-Carlo par chaîne de Markov (algorithme Metropolis-Hastings)

Dans ce qui suit, nous appliquons l'algorithme de Metropolis-Hastings (Metropolis-Hastings) afin d'estimer les coefficients α et β de l'équation de régressions définie dans la relation (3.1).

Afin d'estimer les paramètres des coefficients de la régression, on doit suivre les étapes suivantes :

1. On suppose que $\beta \sim \mathcal{N}(1.02, (0.25)^2)$ et $\alpha_{\bar{x}} \sim \mathcal{N}(15, (1)^2)$ (les lois a priori de β et $\alpha_{\bar{x}}$),
2. On définit la fonction de vraisemblance à partir de l'équation (3.3),
3. On définit la fonction à posteriori conjointe à partir de l'équation (3.4) et le système (3.5),
4. On applique l'algorithme de Metropolis-Hastings afin de récupérer des échantillons de la distribution a posteriori qui représentent en réalité les coefficients de régression. Ce qui nous permet d'estimer les valeurs de ces coefficients et d'obtenir des informations sur leur incertitude à travers les intervalles de crédibilité.

La figure (3.6), représente l'évolution de la chaîne de Markov du coefficient de régression α et β par l'application de l'algorithme Metropolis-Hastings.

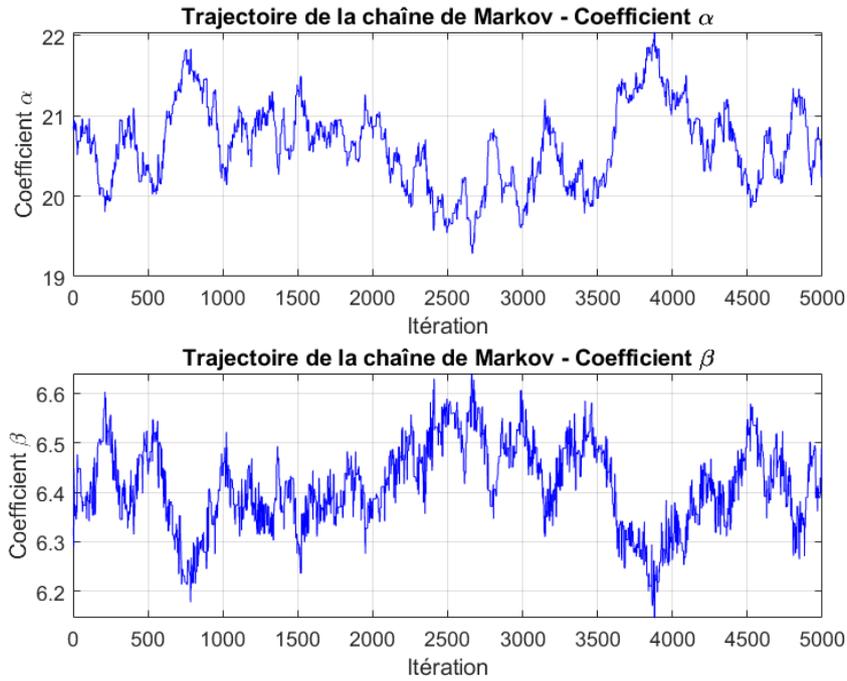


FIGURE 3.6 – La représentation graphique de le trajectoire de la chaîne de Markov des coefficients de régression α et β par l’application de l’algorithme Metropolis-Hastings.

D’après les résultats de cette étude, en effectuant 10 000 itérations avec une période de chauffe de 5000 itérations, nous avons obtenu les estimations suivantes : α est égal à 20.6308 avec un intervalle de confiance de [19.7114, 21.6845], et β est égal à 6.412 avec un intervalle de confiance de [6.235, 6.5654].

La figure (3.7) illustre les trajectoires des coefficients de régression. On peut observer que ces trajectoires sont alignées dans la même direction, ce qui suggère que les coefficients de régression évoluent de manière cohérente. Cette cohérence peut-être interprétée comme une forte relation linéaire entre les variables indépendantes et la variable dépendante.

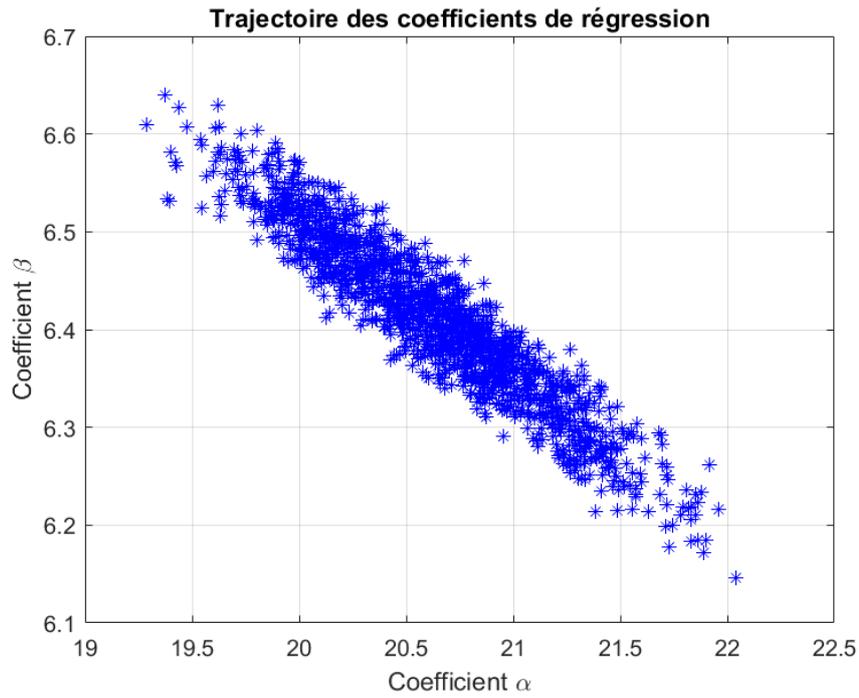


FIGURE 3.7 – Trajectoire des coefficients de régression

La figure (3.8) et (3.9), représentent la densité et l’histogramme des coefficients de régression α et β .

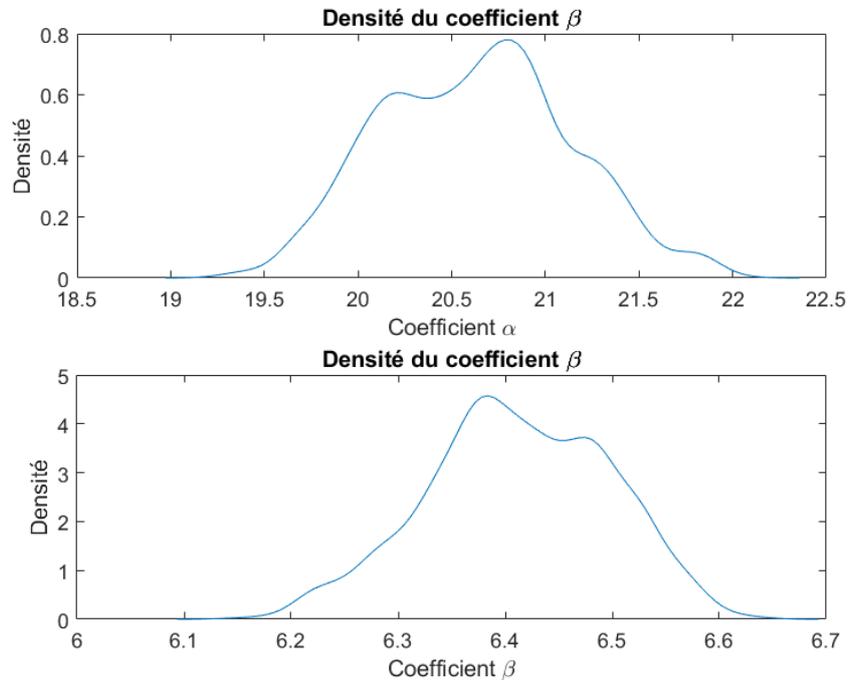


FIGURE 3.8 – La représentation graphique de la densité des coefficients α et β .

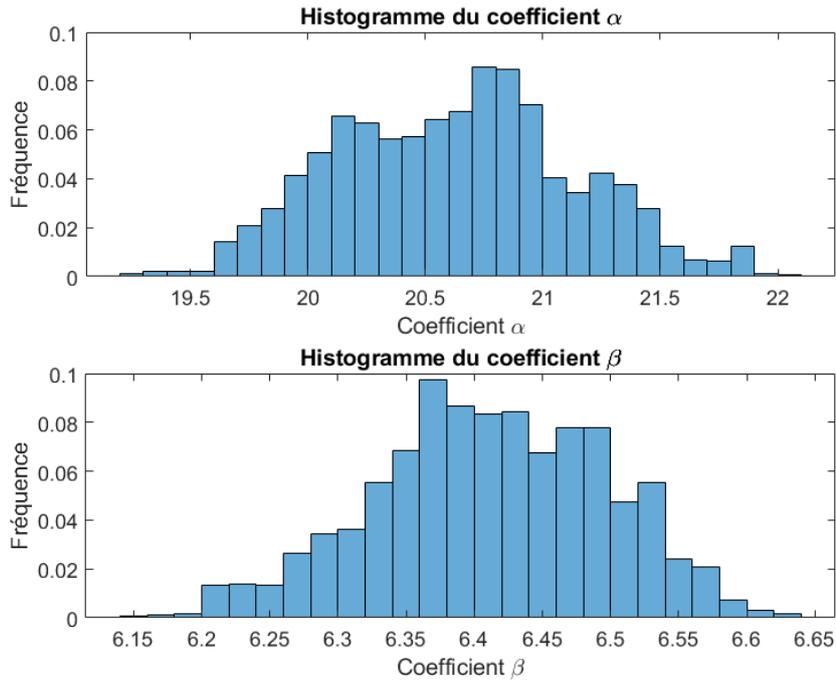


FIGURE 3.9 – L’histogramme des coefficients de régression α et β .

Les taux de décès estimés (\hat{Y}) par l’application de l’algorithme de Métropolis-Hastings sont résumés dans le tableau (3.4).

| Observation | \hat{Y} | Observation | \hat{Y} |
|-------------|-----------|-------------|-----------|
| 1 | 71.25649 | 12 | 39.19647 |
| 2 | 58.43248 | 13 | 39.19647 |
| 3 | 71.25649 | 14 | 39.19647 |
| 4 | 32.78447 | 15 | 58.43248 |
| 5 | 45.60848 | 16 | 71.25647 |
| 6 | 58.43248 | 17 | 58.43248 |
| 7 | 64.84449 | 18 | 45.60848 |
| 8 | 64.84449 | 19 | 64.84449 |
| 9 | 58.43248 | 20 | 64.84449 |
| 10 | 52.02048 | 21 | 58.43248 |
| 11 | 45.608480 | 22 | 71.25649 |

TABLE 3.4 – Les taux de décès estimés (\hat{Y}) par l’application de l’algorithme Metropolis-Hastings.

La figure (3.10), représente le nuage de point $(x_i, y_i), i = 1, \dots, n$ et la droite de régression obtenue par l’application de l’algorithme Metropolis-Hastings.

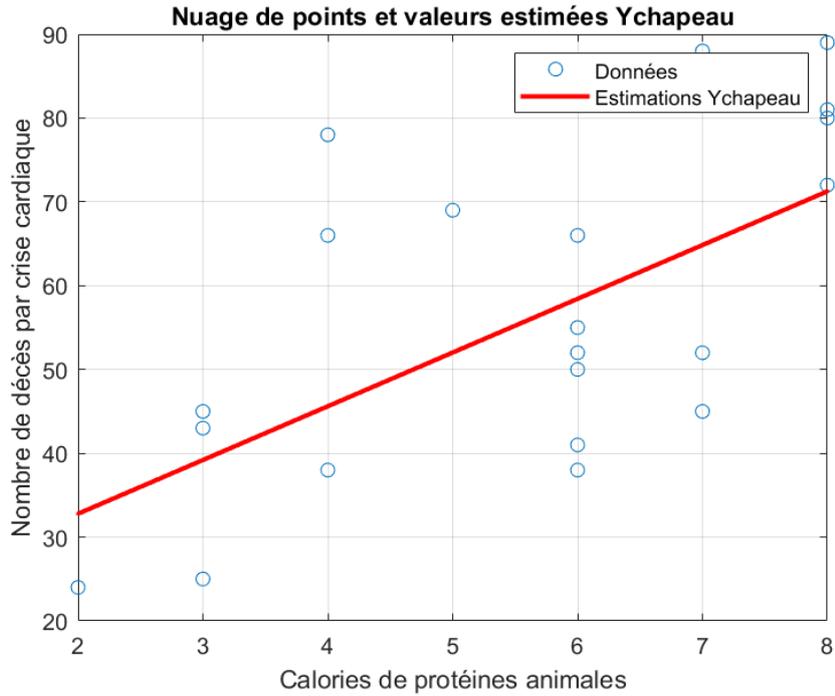


FIGURE 3.10 – Nuage de point $(x_i, y_i), i = 1, \dots, n$ et la droite de régression obtenue par l’application l’algorithme de Metropolis-Hastings.

D. Méthode Monte-Carlo par chaîne de Markov (algorithme de Gibbs)

Dans cette partie, nous appliquons l’échantillonneur de Gibbs afin d’estimer les coefficients de régression α et β , pour modéliser la relation entre X et Y .

Pour $n = 10000$ itérations on obtient $\beta = 7.3236$ et $\alpha = 14$.

Les taux de décès estimés (\hat{Y}) par l’algorithme de Gibbs sont résumés dans le tableau (3.5).

| Observation | \hat{Y} | Observation | \hat{Y} |
|-------------|-----------|-------------|-----------|
| 1 | 72.58894 | 12 | 35.97085 |
| 2 | 57.94171 | 13 | 35.97085 |
| 3 | 72.58894 | 14 | 35.97085 |
| 4 | 28.64723 | 15 | 57.94171 |
| 5 | 43.29447 | 16 | 72.58894 |
| 6 | 57.94171 | 17 | 57.94171 |
| 7 | 65.26533 | 18 | 43.29447 |
| 8 | 65.26533 | 19 | 65.26533 |
| 9 | 57.94171 | 20 | 65.26533 |
| 10 | 56.1809 | 21 | 57.94171 |
| 11 | 43.29447 | 22 | 72.58894 |

TABLE 3.5 – Les taux de décès estimés (\hat{Y}) par l’application de l’algorithme de Gibbs .

La figure (3.11), représente le nuage de point $(x_i, y_i), i = 1, \dots, n$ et la droite de régression obtenue par l’application de l’algorithme de Gibbs.

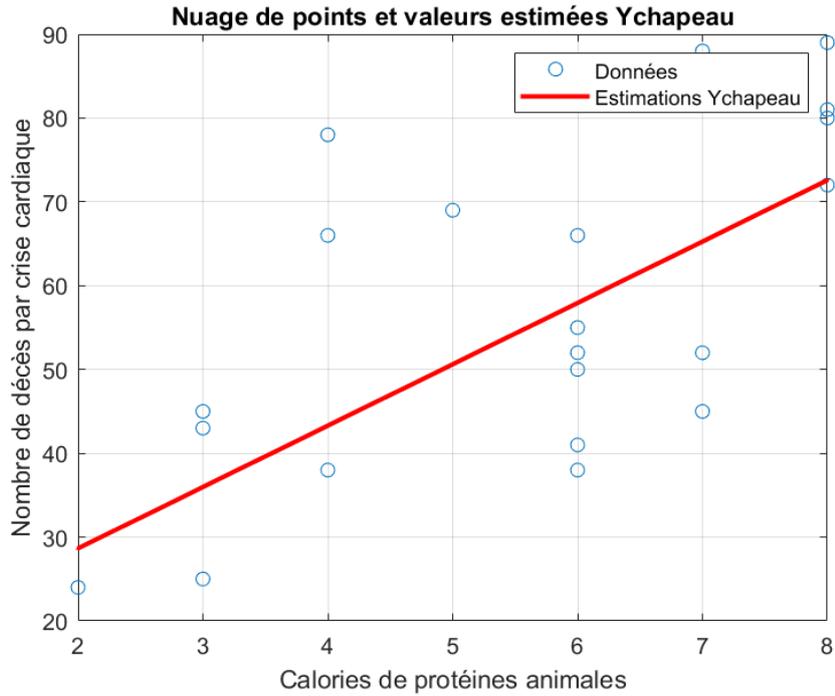


FIGURE 3.11 – Nuage de point $(x_i, y_i), i = 1, \dots, n$ et la droite de régression par l’application de l’algorithme de Gibbs.

E. Étude comparative

Dans ce qui suit nous faisons une étude comparative entre la méthode analytique et la méthode MC et MCMC pour les dix premiers pays.

Le tableaux (3.6), représente les taux des décès réels (Y) et estimés (\hat{Y}) par l’application des 4 méthodes dans les dix premiers pays.

| Observation | Y | \hat{Y} (analytique) | \hat{Y} (MC) | \hat{Y} (Metropolis-Hastings) | \hat{Y} (Gibbs) |
|-------------|-----|------------------------|----------------|---------------------------------|-------------------|
| 1 | 81 | 71.88060 | 72.32057 | 71.25649 | 72.58894 |
| 2 | 55 | 59.09701 | 57.98764 | 58.43248 | 57.94171 |
| 3 | 80 | 71.88060 | 72.32057 | 71.25649 | 72.58894 |
| 4 | 24 | 33.52985 | 29.32178 | 32.78447 | 28.64723 |
| 5 | 78 | 46.31343 | 43.65471 | 45.60848 | 43.29447 |
| 6 | 52 | 59.09701 | 57.98764 | 58.43248 | 57.94171 |
| 7 | 88 | 65.48881 | 65.15411 | 64.84449 | 65.26533 |
| 8 | 45 | 65.48881 | 65.15411 | 64.84449 | 65.26533 |
| 9 | 50 | 59.09701 | 57.98764 | 58.43248 | 57.94171 |
| 10 | 69 | 52.70522 | 50.82118 | 52.02048 | 56.1809 |

TABLE 3.6 – Les taux de décès réels (Y) et estimés (\hat{Y}) par l’application des 4 méthodes.

D’après les résultats de la régression bayésienne simple obtenues par les différente méthodes, on peut voir que les valeurs de \hat{y} sont approximativement proches les unes des autres.

3.1.4 Exemple 2 : Régression multiple

Prenons le même exemple que pour le cas simple et supposons que,

Y taux de décès = $100[\log(\text{nombre de décès par crise cardiaque pour 100000 hommes de 55 à 59 ans}) - 2]$.

X_1 nombre de café par millier d'habitants.

X_2 calories provenant des matières grasses en pourcentage du total des calories.

X_3 calories provenant de protéines animales en pourcentage du total des calories.

Les différentes données de cet exemple sont résumées dans le tableau (3.7).

L'objectif de cette modélisation est d'étudier la relation entre les variables X_1 , X_2 , X_3 et Y . Dans un premier temps, nous souhaitons modéliser la relation entre X_1 , X_2 , X_3 et Y , dans une régression multiple. Ensuite, nous nous intéressons à modéliser la relation entre X_1 , X_2 et Y et enfin entre X_2 , X_3 et Y . Dans cet exemple, nous allons voir que la méthode MCMC permet d'estimer les coefficients de régression pour chaque équation spécifique et d'évaluer l'influence des variables indépendantes sur la variable dépendante.

| Obs | Pays | X_1 | X_2 | X_3 | Y | Obs | Pays | X_1 | X_2 | X_3 | Y |
|-----|-----------|-------|-------|-------|-----|-----|------------|-------|-------|-------|-----|
| 1 | Australie | 124 | 33 | 8 | 81 | 12 | Italie | 22 | 21 | 3 | 45 |
| 2 | Autriche | 49 | 31 | 6 | 55 | 13 | Japon | 16 | 8 | 3 | 24 |
| 3 | Canada | 181 | 38 | 8 | 50 | 14 | Mexique | 10 | 23 | 3 | 43 |
| 4 | Ceylan | 4 | 17 | 2 | 24 | 15 | Pays-Bas | 63 | 37 | 6 | 38 |
| 5 | Chili | 22 | 20 | 4 | 78 | 16 | N-Zélande | 170 | 40 | 8 | 72 |
| 6 | Danemark | 152 | 39 | 6 | 52 | 17 | Norvège | 125 | 38 | 6 | 41 |
| 7 | Finlande | 75 | 30 | 7 | 88 | 18 | Portugal | 12 | 25 | 4 | 38 |
| 8 | France | 54 | 29 | 7 | 45 | 19 | Suède | 221 | 39 | 7 | 52 |
| 9 | Allemagne | 43 | 34 | 6 | 50 | 20 | Suisse | 171 | 33 | 7 | 52 |
| 10 | Irlande | 41 | 31 | 5 | 69 | 21 | G-Bretagne | 97 | 38 | 6 | 66 |
| 11 | Brazile | 17 | 23 | 4 | 66 | 22 | Etats-Unis | 254 | 39 | 8 | 89 |

TABLE 3.7 – Le taux de décès par attaque cardiaque chez les hommes de 55 à 59 ans et les variable X_1 , X_2 , X_3 .

Pour le cas multiple, nous supposons que :

- $X_{(n,p)}$ Une matrice de taille $n \times p$,
- $\beta_{(1,p)}$ Un vecteur d'une ligne et p colonnes de variable inconnues,
- $Y_{(n,1)}$ Un vecteur de n lignes et une colonne des observations,
- $\epsilon_{(n,1)}$ Un vecteur désigne l'erreur

et on considère le modèle suivant :

$$Y = X\beta + \epsilon. \quad (3.9)$$

A. Modéliser la relation entre X_1, X_2, X_3 et Y

Pour notre premier cas, nous avons,

$$\begin{aligned} X &= [X_0, X_1, X_2, X_3] \text{ (} X_0 \text{ vecteur ligne de valeur 1),} \\ \beta &= [\alpha, \beta_1, \beta_2, \beta_3], \\ Y &\text{ Un vecteur de } n \text{ lignes et une colonne des observations,} \\ \epsilon &\text{ Un vecteur de } n \text{ lignes désigne l'erreur.} \end{aligned}$$

La figure (3.12), représente l'évolution de histogramme du coefficient α, β_1, β_2 et β_3 par l'application de la méthode MCMC.

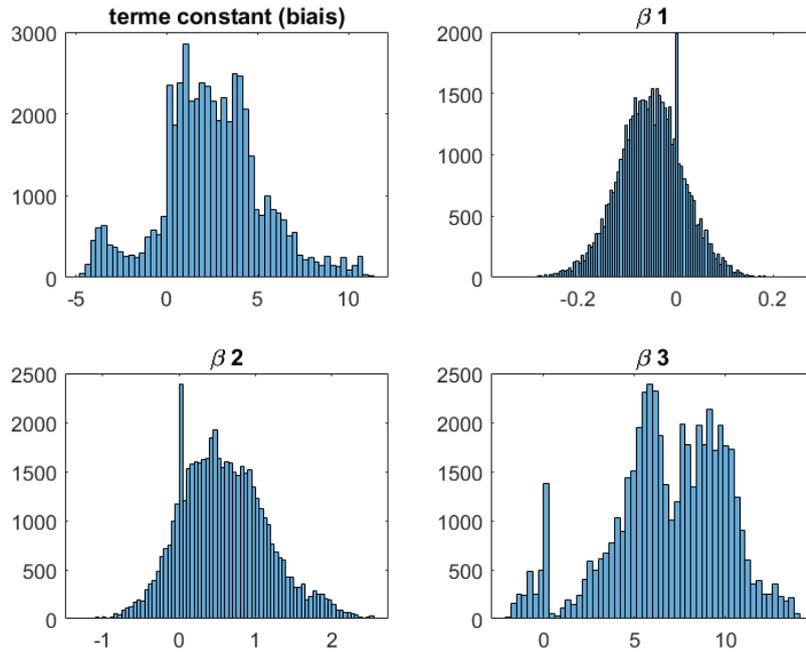


FIGURE 3.12 – L’histogramme des coefficients de régression $\alpha, \beta_1, \beta_2, \beta_3$ par l’application de méthode MCMC

Pour $n = 50000$ itérations et une période de chauffe de 1000 itérations, on obtient $\alpha = 2.5014, \beta_1 = -0.0506, \beta_2 = 0.5827$ et $\beta_3 = 6.8555$.

Les taux de décès estimés (\hat{Y}) par l’application de la méthode MCMC de Metropolis-Hastings sont résumés dans le tableau (3.8).

| Observation | \hat{Y} | Observation | \hat{Y} |
|-------------|-----------|-------------|-----------|
| 1 | 70.29503 | 12 | 34.19058 |
| 2 | 59.21682 | 13 | 26.91929 |
| 3 | 70.32187 | 14 | 35.96370 |
| 4 | 25.91582 | 15 | 62.00403 |
| 5 | 40.46340 | 16 | 72.04435 |
| 6 | 58.66218 | 17 | 59.44684 |
| 7 | 64.17292 | 18 | 43.88335 |
| 8 | 64.65372 | 19 | 62.02331 |
| 9 | 61.85149 | 20 | 61.05926 |
| 10 | 52.76644 | 21 | 60.86486 |
| 11 | 42.46473 | 22 | 67.20760 |

TABLE 3.8 – Les taux de décès estimés (\hat{Y}) par l'application de méthode MCMC.

La figure (3.13), représente le nuage de point et le trajectoire de la régression par l'application de méthode MCMC.

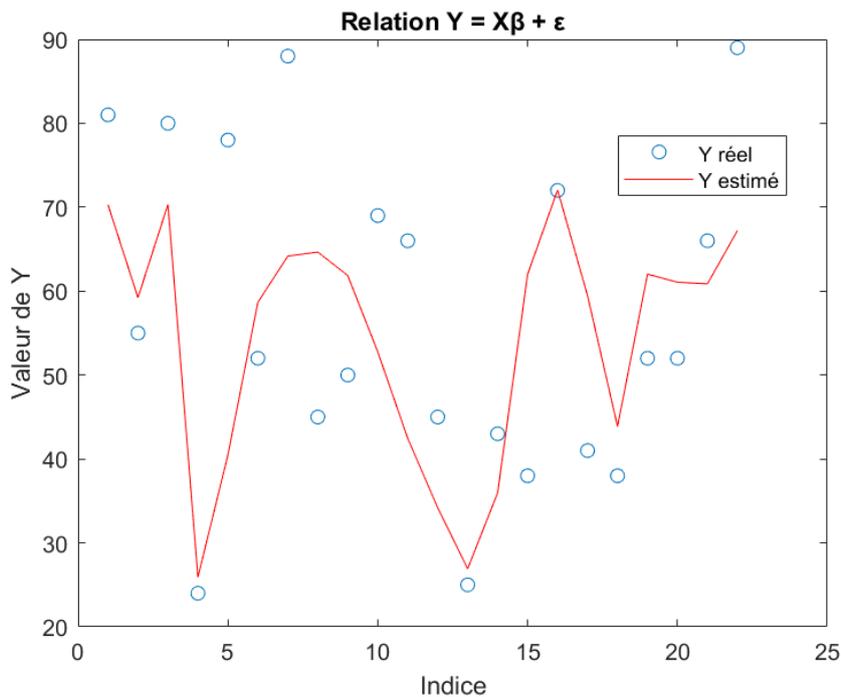


FIGURE 3.13 – Nuage de point et la trajectoire de la régression par l'application de méthode MCMC

L'étude examine le taux de décès et présente les résultats obtenus à partir de l'application de la méthode MCMC. Les résultats indiquent les relations entre différentes variables et le taux de décès.

Le coefficient β_1 a une valeur de -0.0506 , ce qui suggère un effet contraire entre la variable X_1 (café) et le taux de décès. En d'autres termes, la consommation de café est associée à une diminution du taux de décès selon cette étude.

Le coefficient β_2 a une valeur de 0.5827, ce qui suggère que la variable X_2 (calories provenant de matières grasses) a une influence négligeable sur le taux de décès. Par conséquent, la proportion de calories provenant de graisses dans l'alimentation n'a pas d'effet significatif sur le taux de décès dans le cadre de cette étude.

Le coefficient β_3 a une valeur de 6.8555, ce qui indique une relation positive et significative entre la variable X_3 (calories provenant de protéines animales) et le taux de décès. Cela suggère que l'augmentation de la proportion de calories provenant de protéines animales dans l'alimentation est associée à une augmentation du taux de décès. Cette observation peut-être attribuée à des facteurs tels que les risques pour la santé cardiovasculaire liés à une consommation excessive de protéines animales.

B. Modéliser la relation entre X_1 , X_2 et Y

Dans ce cas l'objectif est de modéliser la relation entre X_1 , X_2 et Y par l'application d'une méthode de MCMC. Pour cela on doit estimer les coefficients de régression α , β_1 et β_2 .

Pour ce cas, nous avons,

$$\begin{aligned} X &= [X_0, X_1, X_2] \text{ (} X_0 \text{ vecteur ligne de valeur 1),} \\ \beta &= [\alpha, \beta_1, \beta_2], \\ Y &\text{ Un vecteur de n lignes et une colonne des observations,} \\ \epsilon &\text{ Un vecteur de n lignes désigne l'erreur.} \end{aligned}$$

La figure (3.14), représente l'évolution du trajectoire de la chaîne MCMC pour les coefficients de régression β_1 et β_2 .

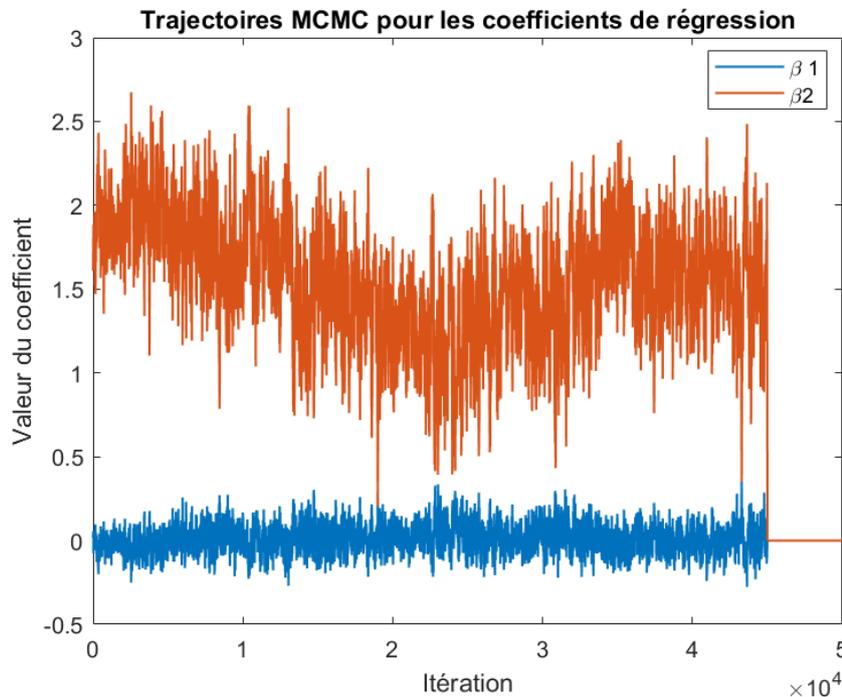


FIGURE 3.14 – Trajectoire des coefficient de régression β_1 et β_2 par l'application de méthode MCMC.

Selon les résultats de la figure et après 50000 itérations avec une période de chauffe de 5000,

nous obtenons les valeurs suivantes : $\alpha = 6.1460$, $\beta_1 = 0.0104$ et $\beta_2 = 1.3998$. On remarque clairement que les deux coefficients β_1 et β_2 ont peu d'influence sur le taux de décès.

Les taux de décès estimés (\hat{Y}) par l'application de méthode MCMC sont résumés dans le tableau (3.9).

| Observation | \hat{Y} | Observation | \hat{Y} |
|-------------|-----------|-------------|-----------|
| 1 | 53.63438 | 12 | 35.77146 |
| 2 | 50.05140 | 13 | 17.51149 |
| 3 | 61.22873 | 14 | 38.44569 |
| 4 | 29.98427 | 15 | 58.59638 |
| 5 | 34.37166 | 16 | 63.91341 |
| 6 | 62.32560 | 17 | 60.64379 |
| 7 | 48.92319 | 18 | 41.26617 |
| 8 | 47.30404 | 19 | 63.04634 |
| 9 | 55.58789 | 20 | 54.12532 |
| 10 | 49.96783 | 21 | 60.35132 |
| 11 | 38.51881 | 22 | 63.39103 |

TABLE 3.9 – Les taux de décès estimés (\hat{Y}) par l'application de méthode MCMC.

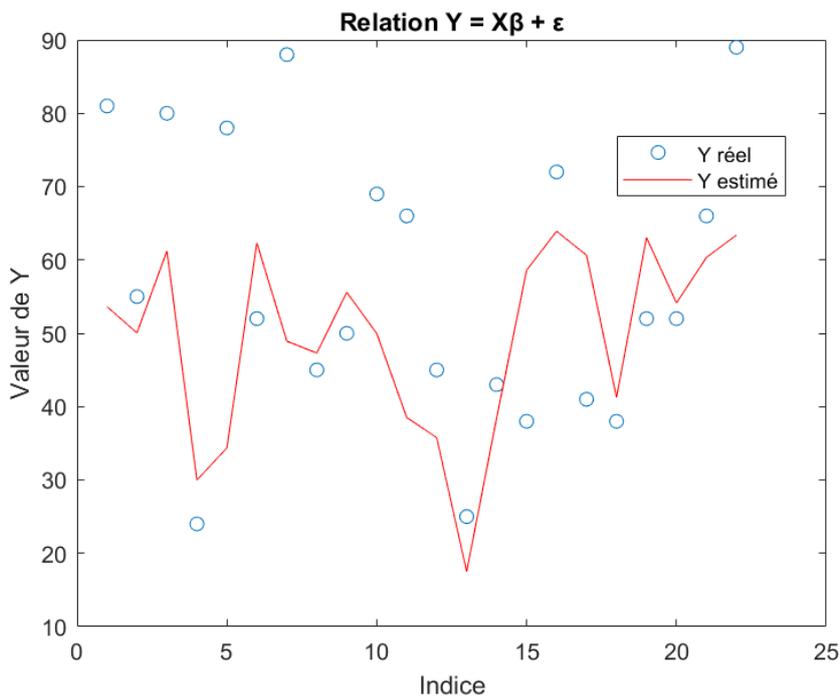


FIGURE 3.15 – Nuage de point et le trajectoire de la régression

C. Modéliser la relation entre X_2 , X_3 et Y

Dans ce cas l'objectif est de trouver la relation entre X_2 , X_3 et Y . Autrement dit, on doit estimer les coefficients de régression α , β_2 et β_3 .

La figure (3.16), représente l'évolution de la trajectoire et l'histogramme des coefficients de régression α , β_2 et β_3 par l'application de méthodes MCMC.

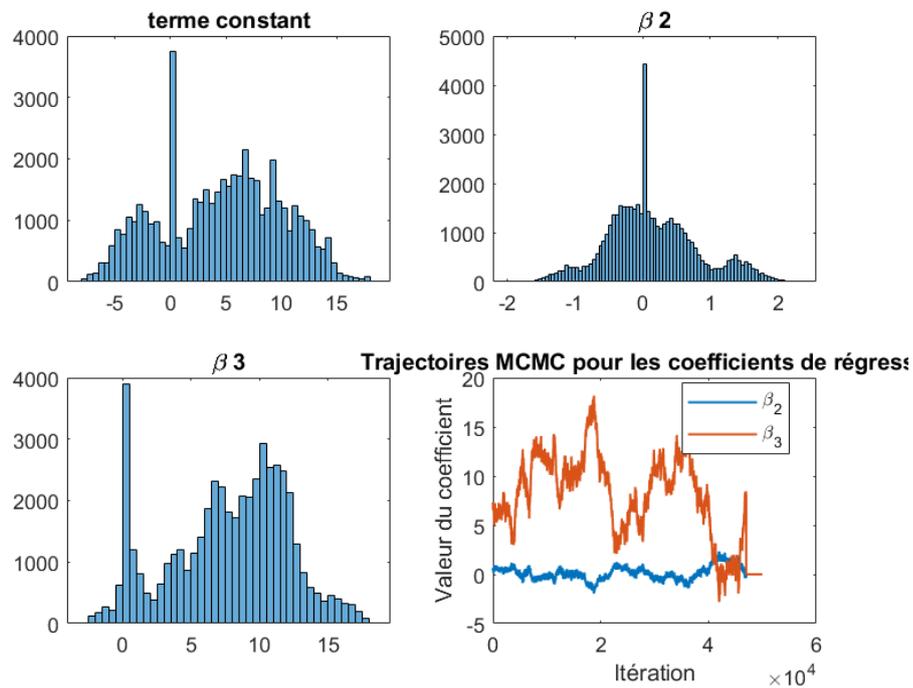


FIGURE 3.16 – Trajectoire et l'histogramme des coefficient de régression β_2 et β_3 par l'application de méthode MCMC.

Pour 50000 itérations et avec une période de chauffe de 5000 itérations on obtient $\alpha = 4.5695$, $\beta_2 = 0.1314$ et $\beta_3 = 7.7045$.

Les taux de décès estimés (\hat{Y}) par l'application de la méthode MCMC sont résumés dans le tableau (3.10).

| Observation | \hat{Y} | Observation | \hat{Y} |
|-------------|-----------|-------------|-----------|
| 1 | 70.54134 | 12 | 30.44220 |
| 2 | 54.86958 | 13 | 28.73412 |
| 3 | 71.19829 | 14 | 30.70498 |
| 4 | 22.21215 | 15 | 55.65792 |
| 5 | 38.01530 | 16 | 71.46107 |
| 6 | 55.92070 | 17 | 55.78931 |
| 7 | 62.44268 | 18 | 38.67225 |
| 8 | 62.31129 | 19 | 63.62519 |
| 9 | 55.39514 | 20 | 62.83685 |
| 10 | 47.16509 | 21 | 55.78931 |
| 11 | 38.40947 | 22 | 71.32968 |

TABLE 3.10 – Les taux de décès estimés (\hat{Y}) par l'application de la méthode MCMC.

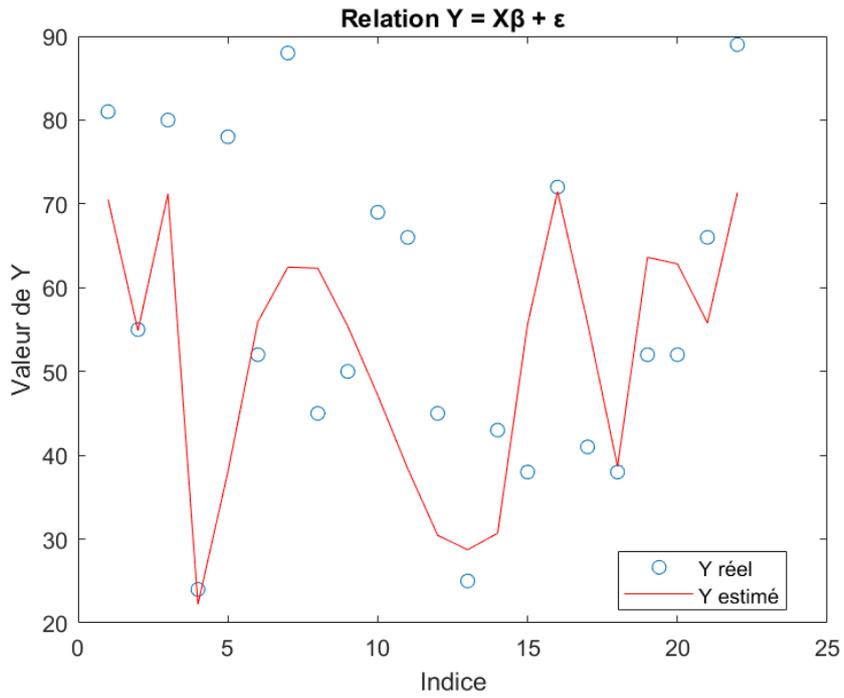


FIGURE 3.17 – Nuage de point et le trajectoire de régression.

Les résultats de cette étude, confirme ce qu'on a déjà montré dans le cas simple et multiple. En effet, ces résultats suggèrent que l'augmentation de la proportion de calories provenant de protéines animales dans l'alimentation est associée fortement à une augmentation du taux de décès.

3.2 Estimation de la taille d'une population

L'estimation de la taille d'une population est une question fondamentale dans de nombreux domaines de recherche. Qu'il s'agisse de populations animales, humaines ou d'autres organismes, connaître la taille de la population est essentiel pour la prise de décisions éclairées, la gestion des ressources et la planification des activités.

La méthode bayésienne et les algorithmes de Monte Carlo par chaîne de Markov (MCMC) offrent une approche puissante et flexible pour estimer la taille d'une population. Cette approche combine la modélisation probabiliste avec l'inférence bayésienne pour fournir des estimations probabilistes de la taille de la population, prenant en compte à la fois les connaissances a priori et les données observées.

3.2.1 Définitions et domaines d'applications

La taille d'une population peut-être définie comme le nombre total d'individus ou d'éléments constituant cette population. Cela peut s'appliquer à différents domaines, tels que :

1. **L'écologie** : Dans le domaine de l'écologie, la taille d'une population fait référence au nombre d'individus d'une espèce présents dans un écosystème spécifique. Il peut-être important d'estimer la taille d'une population pour évaluer son statut de conservation, étudier les interactions écologiques ou évaluer l'impact des perturbations environnementales.
2. **La biologie de la conservation** : Dans le contexte de la biologie de la conservation, la taille d'une population est un indicateur clé pour évaluer la viabilité et la santé d'une espèce menacée. L'estimation précise de la taille de la population est essentielle pour élaborer des stratégies de conservation et prendre des décisions éclairées en matière de gestion des populations.
3. **Les sciences sociales** : Dans les sciences sociales, la taille de la population peut se référer à la taille d'un groupe humain, comme la population d'un pays, d'une ville ou d'une communauté spécifique. L'estimation de la taille de la population est essentielle pour la planification des ressources, la formulation de politiques publiques, les études démographiques, etc.
4. **L'épidémiologie** : En épidémiologie, la taille de la population est souvent utilisée pour déterminer l'incidence et la prévalence d'une maladie. Elle permet d'estimer le nombre de personnes susceptibles d'être affectées par une maladie dans une population donnée, ce qui est crucial pour la gestion des épidémies et la prise de décisions en matière de santé publique.

3.2.2 Approche méthodologique

L'estimation de la taille d'une population à l'aide de la méthode bayésienne et des algorithmes MCMC, passe par les étapes suivantes :

1. **Formulation du modèle** : On définit un modèle probabiliste qui décrit la relation entre les données observées et la taille de la population. Le modèle peut inclure des

paramètres tels que la taille de la population et d'autres paramètres d'intérêt.

2. **Spécification des distributions a priori** : On détermine les distributions a priori pour les paramètres du modèle, y compris la taille de la population. Les distributions a priori représentent les connaissances et les croyances préalables sur les paramètres avant d'observer les données.
3. **Collecte des données** : On récupère des données sur un échantillon de la population ou sur des informations indirectes liées à la taille de la population.
4. **Inférence bayésienne** : À l'aide du théorème de Bayes, on combine les distributions a priori et les données pour obtenir la distribution a posteriori des paramètres. Dans ce cas, la distribution a posteriori représente l'incertitude quant à la taille de la population compte tenu des données observées.
5. **Échantillonnage MCMC** : On utilise un algorithme Monte-Carlo par chaîne Markov, tel que l'algorithme de Metropolis-Hastings ou l'échantillonneur de Gibbs, pour générer un échantillon de la distribution a posteriori. Les échantillons sont tirés de manière itérative et dépendent des échantillons précédents, formant une chaîne de Markov qui converge vers la distribution a posteriori.
6. **Estimation de la taille de la population** : À partir de l'échantillon généré par l'algorithme MCMC, on estime la taille de la population en utilisant différentes mesures statistiques, telles que la moyenne, la médiane ou les intervalles de crédibilité.
7. **Diagnostic et évaluation** : On évalue la convergence de la chaîne de Markov pour s'assurer que l'algorithme de MCMC a convergé vers la distribution a posteriori. Comme, on peut aussi effectuer des diagnostics supplémentaires pour évaluer la qualité des échantillons et l'adéquation du modèle aux données observées.

3.2.3 Estimation de la taille d'une population de poissons dans un lac

Supposons qu'un groupe de scientifiques souhaite estimer la taille d'une population d'un type de poisson dans un lac. Pour cela, ils ont effectué des observations en utilisant des filets de pêche pendant une période déterminée. Au total, ils ont capturé et compté un échantillon de poissons de taille N . Parmi ces N poissons pêchés, les scientifiques ont observé un nombre spécifique de poissons, représenté par n .

Modèle

- Paramètre inconnu : θ (la taille de la population de poissons dans le lac).
- Données observées : n (nombre de poissons observés dans l'échantillon).

Aspect mathématique

- Loi a priori : $\pi(\theta)$ (distribution a priori sur θ).
- Loi de vraisemblance : $L(n|\theta)$ (probabilité des données observées n).
- Loi a posteriori : $\pi(\theta|n)$ (probabilité de θ étant donné les données observées n).

Objectif

Notre objectif est d'estimer la distribution a posteriori $\pi(\theta|n)$ en utilisant l'algorithme MCMC. Cela nous permettra d'obtenir une estimation de la taille de la population de poissons dans le lac, en combinant nos connaissances a priori sur θ , la vraisemblance des données observées n , et en utilisant une approche probabiliste basée sur l'algorithme MCMC.

Illustration

Dans ce qui suit, nous donnons une illustration de ce qui précède.

1. Choix de la loi de vraisemblance

• Loi binomiale

Supposons que le groupe de scientifiques ait utilisé des filets de pêche et ait capturé un échantillon de 100 poissons ($N = 100$). Parmi ces 100 poissons, ils ont observé et compté 80 poissons spécifiques ($n = 80$) qui appartiennent au type de poisson qu'ils souhaitent étudier.

Dans ce cas, $n = 80$ signifie que les scientifiques ont pu identifier et compter précisément 80 poissons du type recherché parmi l'échantillon de 100 poissons capturés. Ces 80 poissons observés fournissent des informations précieuses sur la proportion et la répartition de ce type de poisson dans la population totale.

En résumé, n est une mesure concrète qui indique le nombre spécifique de poissons observés, identifiés et comptés parmi l'échantillon capturé. Il s'agit d'une information clé pour évaluer la composition et la proportion d'un type de poisson particulier dans la population totale, en fournissant des données tangibles sur les individus observés dans l'échantillon étudié. Dans ce cas, la loi binomiale peut être un choix approprié pour la loi de vraisemblance. En effet, la loi binomiale modélise le nombre de succès (dans ce cas, le nombre de poissons observés) dans un nombre fixe d'essais indépendants (la taille de l'échantillon). Elle est souvent utilisée lorsque les données observées sont discrètes et que chaque observation peut être classée comme un succès ou un échec.

• Loi poisson

Supposons que le groupe de scientifiques pêchent régulièrement dans le lac pour évaluer la population de poissons. Lors de chaque pêche, ils capturent un échantillon de poissons et comptent le nombre de poissons capturés.

Supposons que les données observées lors de n pêches successives sont données dans le vecteurs $[10, 12, 8, 15, \dots]$. Dans ce cas, chaque pêche est considérée comme un événement de comptage indépendant, où nous observons le nombre de poissons capturés. La loi de Poisson nous permet de modéliser cette distribution de comptage en utilisant un seul paramètre (θ), qui représente le taux moyen de capture de poissons.

En supposant que les captures de poissons lors de chaque pêche sont indépendantes les unes des autres, nous pouvons utiliser la loi de Poisson pour estimer le taux moyen de capture (représentant la taille de la population) à partir des données observées. La loi de Poisson nous permet également de prendre en compte la rareté des événements de capture, ce qui est souvent le cas dans des populations de poissons où seuls quelques individus sont capturés lors de chaque pêche.

Ainsi, en utilisant la loi de Poisson comme loi de vraisemblance, nous pouvons modéliser la probabilité d'observer les nombres de poissons capturés lors des pêches spécifiques et estimer la taille de la population de poissons à partir de ces données observées.

2. Choix de la loi a priori

Contrairement à la loi de vraisemblance, on a aucune information sur la loi à priori dans ce qui suit nous donnons quelques idées de choix de cette dernière.

• Loi uniforme

Supposons que nous menions une étude sur un lac relativement peu exploré et que nous n'ayons aucune information spécifique sur la taille de la population de poissons. Si nous supposons que le nombre de poissons se situe dans une plage donnée, par exemple, entre 10000 et 12000 poissons. Ainsi, nous pouvons choisir une loi uniforme comme loi a priori, où chaque valeur de la taille de population entre 10000 et 12000 a la même probabilité d'être observée.

• Loi normale

Supposons que nous ayons réalisé des études antérieures dans d'autres lacs similaires et que nous ayons constaté que la taille de la population de poissons tend à suivre une distribution proche d'une loi normale, avec une moyenne de 10000 poissons et un écart-type de 2000 poissons. Dans ce cas, nous pourrions utiliser une loi normale comme loi a priori, en incorporant ces connaissances antérieures. Cela nous permettrait de tenir compte de notre hypothèse selon laquelle la taille de la population est susceptible d'être concentrée autour de la moyenne de 10000 poissons.

Dans ce qui suit, nous allons examiner quatre cas différents. Le premier cas concerne une loi a priori $\pi(\theta)$ suivant une distribution normale $\mathcal{N}(\mu, \sigma^2)$, avec une vraisemblance $L(n|\theta)$ suivant une distribution binomiale $\mathcal{B}(n, \frac{\theta}{\mu})$. Le deuxième cas concerne une loi a priori $\pi(\theta)$ suivant une distribution uniforme $\mathcal{U}([a, b])$, avec une vraisemblance $L(n|\theta)$ suivant une distribution binomiale $\mathcal{B}(n, \frac{\theta}{N})$. Le troisième cas concerne une loi a priori $\pi(\theta)$ suivant une distribution normale $\mathcal{N}(\mu, \sigma^2)$, avec une vraisemblance $L(D|\theta)$ suivant une distribution de Poisson $\mathcal{P}(\theta)$. Enfin, le quatrième cas concerne une loi a priori $\pi(\theta)$ suivant une distribution uniforme $\mathcal{U}([a, b])$, avec une vraisemblance $L(D|\theta)$ suivant une distribution de Poisson $\mathcal{P}(\theta)$.

A. Cas 1 : $n|\theta \sim \mathcal{B}(n, \frac{\theta}{\mu})$ et $\theta \sim \mathcal{N}(\mu, \sigma^2)$

Pour le premier cas, nous supposons que les observations suivent la loi binomiale $\mathcal{B}(n, \frac{\theta}{\mu})$, et la loi a priori est la loi normale $\mathcal{N}(\mu, \sigma^2)$.

L'organigramme de calcul de la distribution a posteriori dans ce cas est donné dans la figure(3.18).

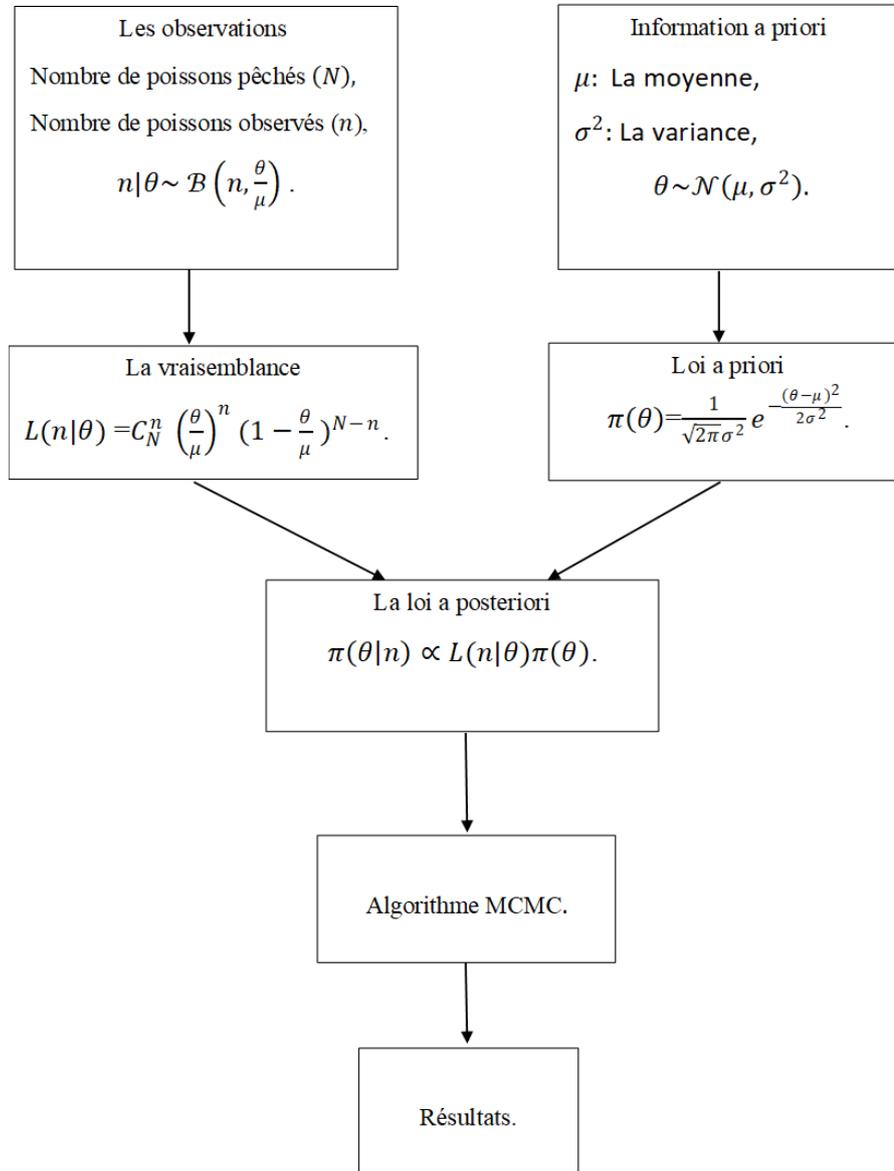


FIGURE 3.18 – Organigramme de calcul de la distribution a posteriori avec $n|\theta \sim \mathcal{B}(n, \frac{\theta}{\mu})$ et $\theta \sim \mathcal{N}(\mu, \sigma^2)$

Le tableau (3.11), représente le nombre de poissons pêchés, le nombre de poissons observées et les paramètres de la loi a priori.

| μ | σ | Nombre de poissons pêchés (N) | Nombre de poissons observées (n) |
|-------|----------|-------------------------------|----------------------------------|
| 10000 | 2000 | 100 | 80 |

TABLE 3.11 – Les données observés dans le cas $n|\theta \sim \mathcal{B}(n, \frac{\theta}{\mu})$ et $\theta \sim \mathcal{N}(\mu, \sigma^2)$.

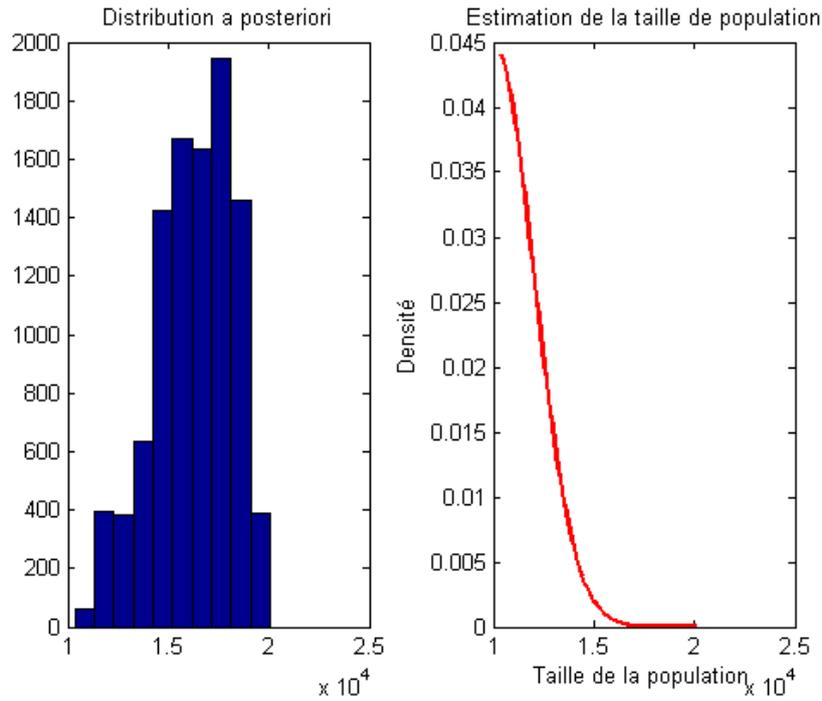


FIGURE 3.19 – À gauche l’histogramme de la distribution a posteriori et à droite distribution de la taille de population des poissons estimée, pour $n|\theta \sim \mathcal{B}(n, \frac{\theta}{\mu})$ et $\theta \sim \mathcal{N}(\mu, \sigma^2)$.

La figure (3.19), représente l’histogramme a posteriori et la distribution de la taille de population estimées à partir d’une loi de vraisemblance binomiale $\mathcal{B}(n, \frac{\theta}{\mu})$ et d’une loi a priori normale $\mathcal{N}(\mu, \sigma^2)$ en utilisant l’algorithme de Metropolis-Hastings.

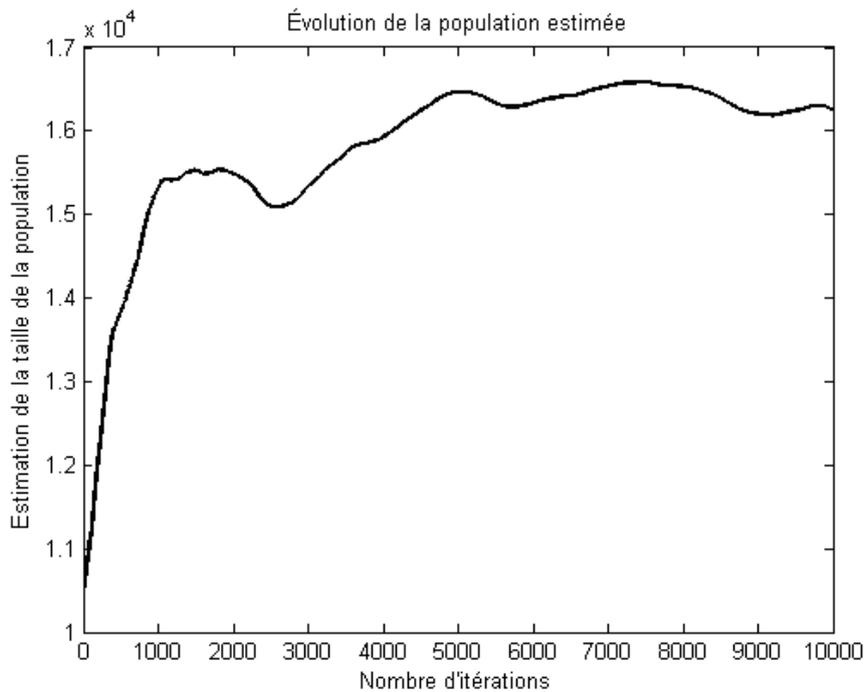


FIGURE 3.20 – L’évolution de la population des poissons estimée, pour $n|\theta \sim \mathcal{B}(n, \frac{\theta}{\mu})$ et $\theta \sim \mathcal{N}(\mu, \sigma^2)$.

La Figure (3.20), présente les résultats de l'estimation de la taille de la population des poissons dans le lac, ainsi que l'évolution de la population estimée. Cette estimation a été réalisée en utilisant l'algorithme Metropolis-Hastings. Pour 10 000 itérations, et pour une période de chauffe de 1 000 itérations, la taille de la population estimée s'élève à 16 256 poissons dans le lac. Cette valeur représente une estimation approximative de la population de poissons dans le lac, en se basant sur les informations fournies par la vraisemblance et la distribution a priori.

Le tableau (3.12) présente les résultats de deux estimations de la taille de la population des poissons. La première estimation a été réalisée en observant 100 poissons pêchés, parmi lesquels 80 spécimens spécifiques ont été identifiés. La deuxième estimation a été effectuée en observant également 100 poissons pêchés, mais cette fois-ci seulement 30 spécimens spécifiques ont été identifiés.

| Nombre de poissons pêchés (N) | Nombre de poissons observés (n) | La taille de population de poissons estimées |
|-------------------------------|---------------------------------|--|
| 100 | 80 | 16256 |
| 100 | 30 | 11205 |

TABLE 3.12 – La taille de population de poissons estimée par rapport au nombre de poisson observé.

Les résultats obtenus (voir tableau (3.12)) suggèrent une corrélation positive entre le nombre de poissons observés parmi ceux qui ont été pêchés et la taille estimée de la population de poissons. En d'autres termes, lorsque davantage de poissons sont observés parmi l'échantillon pêché, cela conduit à une estimation plus élevée de la taille de la population totale de poissons dans le lac.

Cette tendance peut s'expliquer par le fait que l'observation de plus de poissons spécifiques parmi l'échantillon pêché indique une plus grande représentativité de l'échantillon par rapport à la population totale de poissons. Lorsque davantage de poissons spécifiques sont identifiés, il est raisonnable de supposer que ces poissons spécifiques sont également présents dans la population générale du lac. Par conséquent, en extrapolant ces résultats à l'ensemble de la population, l'estimation de la taille de la population de poissons augmente.

B. Cas 2 : $n|\theta \sim \mathcal{B}(n, \frac{\theta}{N})$ et $\theta \sim \mathcal{U}([a, b])$

Pour ce cas, nous supposons que les observations suivent la loi binomiale $\mathcal{B}(n, \frac{\theta}{N})$, et la loi a priori est la loi uniforme $\theta \sim \mathcal{U}([a, b])$.

L'organigramme de calcul de la distribution a posteriori dans ce cas est donné dans la figure (3.21).

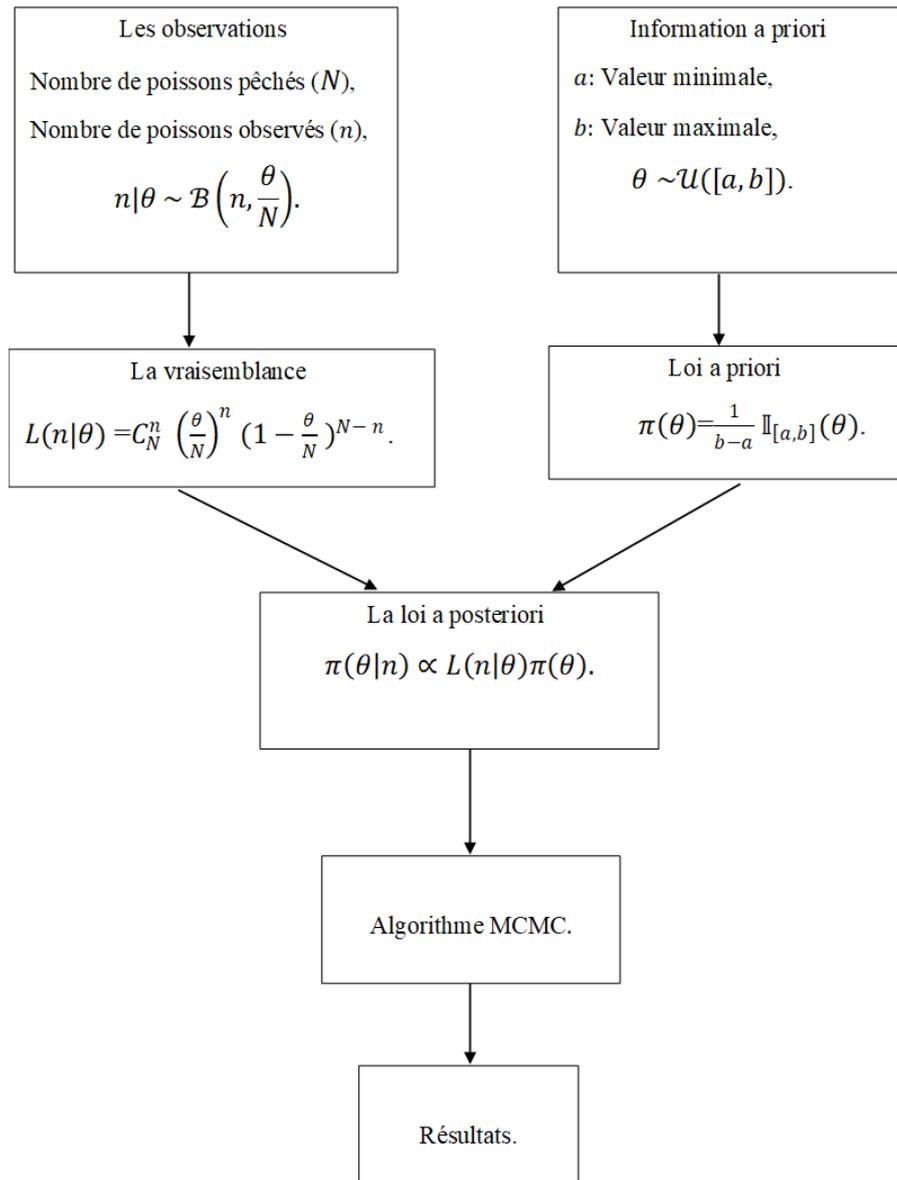


FIGURE 3.21 – Organigramme de calcul de la distribution a posteriori avec $n|\theta \sim \mathcal{B}(n, \frac{\theta}{N})$ et $\theta \sim \mathcal{U}([a, b])$.

Le tableau (3.13), représente le nombre de poissons pêchés, le nombre de poissons observées et les paramètres de la loi a priori.

| a | b | Nombre de poissons pêchés (N) | Nombre de poissons observées (n) |
|-------|-------|-------------------------------|----------------------------------|
| 10000 | 12000 | 100 | 80 |

TABLE 3.13 – Les données observées dans le cas $n|\theta \sim \mathcal{B}(n, \frac{\theta}{N})$ et $\theta \sim \mathcal{U}([a, b])$.

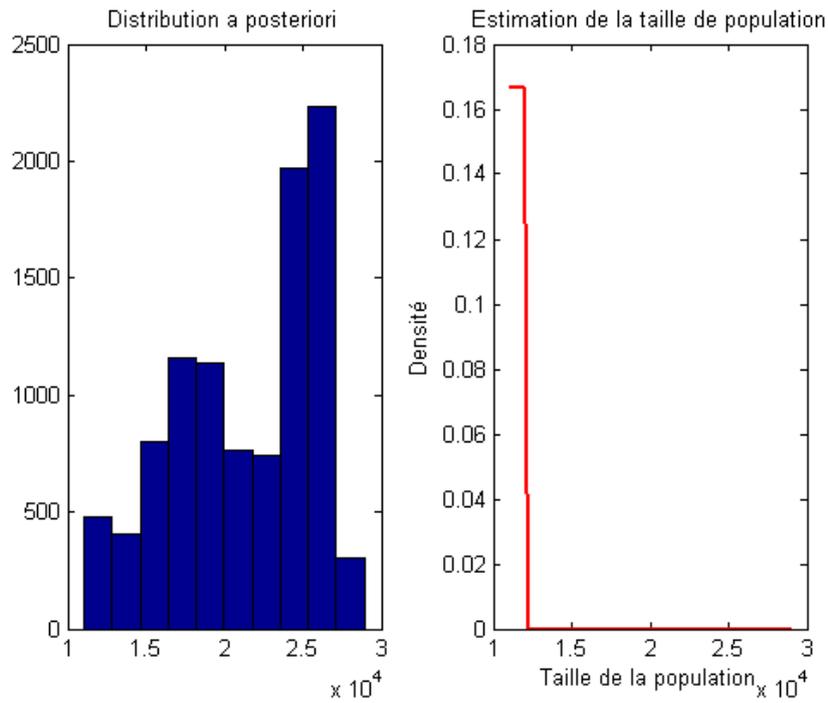


FIGURE 3.22 – À gauche l’histogramme de la distribution a posteriori, à droite la densité de l’estimation de taille de population des poissons, pour $n|\theta \sim \mathcal{B}(n, \frac{\theta}{N})$ et $\theta \sim \mathcal{U}([a, b])$.

La figure (3.22), représente l’histogramme a posteriori et la distribution de la taille de la population estimée à partir d’une loi a priori uniforme $\mathcal{U}([a, b])$ et d’une loi de vraisemblance binomiale $\mathcal{B}(n, \frac{\theta}{N})$ en utilisant l’algorithme Metropolis-Hastings.

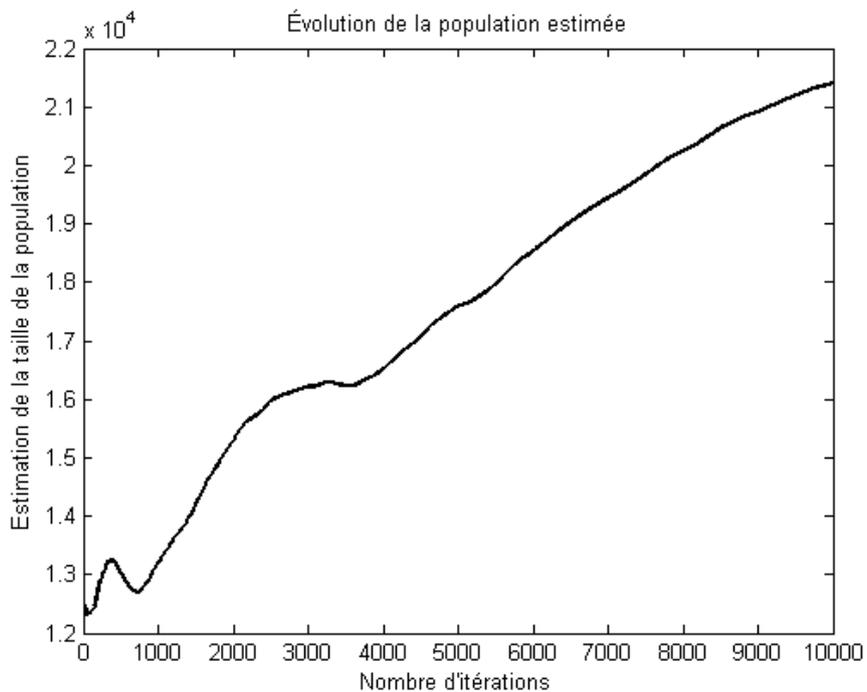


FIGURE 3.23 – L’évolution de la population des poissons estimée, pour $n|\theta \sim \mathcal{B}(n, \frac{\theta}{N})$ et $\pi(\theta) \sim \mathcal{U}([a, b])$.

La figure (3.23), illustre les résultats de l'estimation de la taille de la population des poissons dans le lac et l'évolution de la population estimée par l'utilisation de l'algorithme Metropolis-Hastings. Après 10000 itérations et avec une période de chauffe de 1000, l'estimation obtenue est de 21411 poissons dans le lac. Cela signifie que, selon les informations fournies par la vraisemblance et la distribution a priori, la population de poissons dans le lac est estimée à être d'environ 21411 individus.

C. Cas 3 : $D|\theta \sim \mathcal{P}(\theta)$ et $\theta \sim \mathcal{N}(\mu, \sigma^2)$

Pour le troisième cas, nous supposons que les observations suivent la loi de poisson $\mathcal{P}(\theta)$, et la loi a priori est la loi normale $\mathcal{N}(\mu, \sigma^2)$. L'organigramme de calcul de la distribution a posteriori dans ce cas est donné dans la figure (3.24).

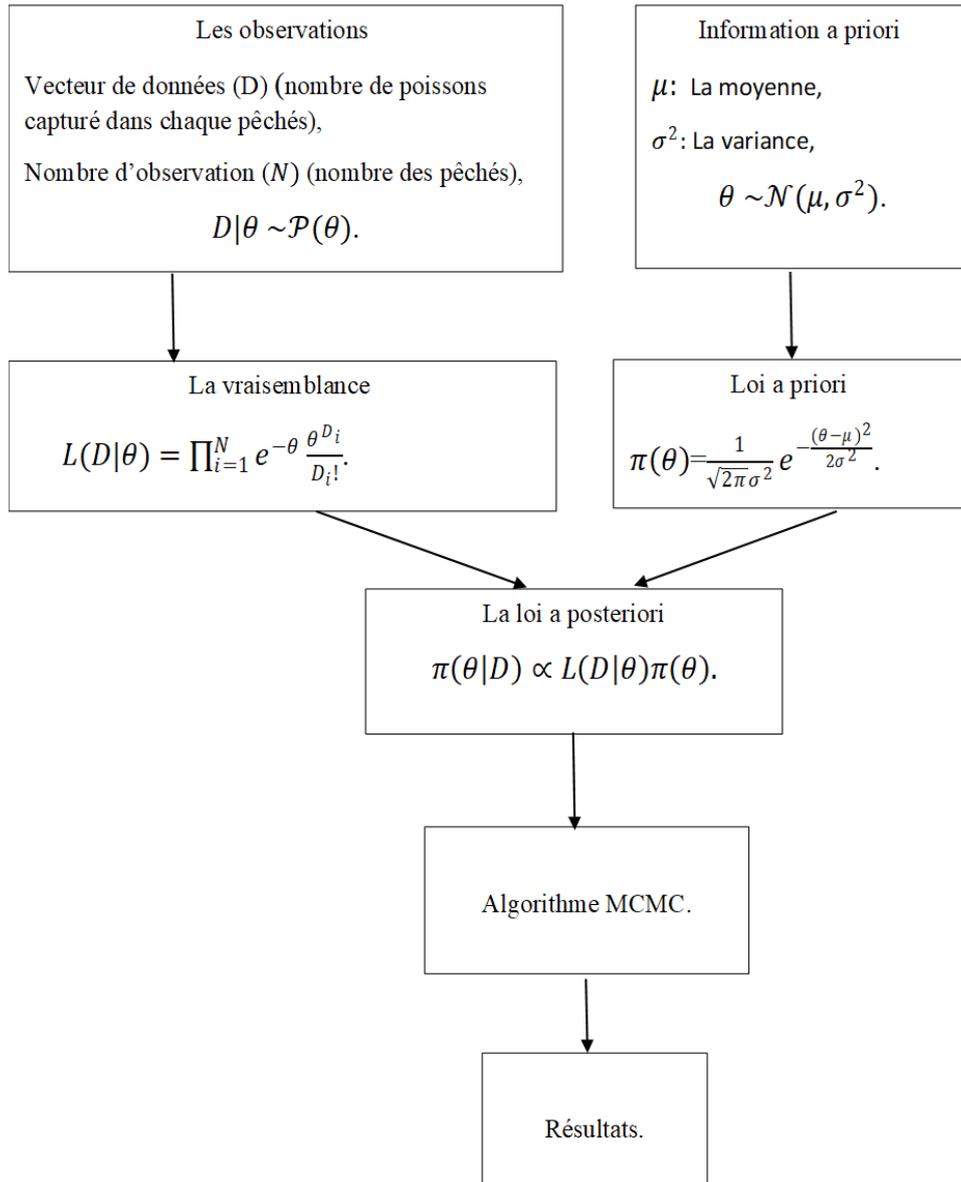


FIGURE 3.24 – Organigramme de calcul de la distribution a posteriori pour $D|\theta \sim \mathcal{P}(\theta)$ et $\theta \sim \mathcal{N}(\mu, \sigma^2)$.

Le vecteur des données observées D qui représente le nombre de poissons capturés dans chaque pêché, le nombre des observations N qui représente le nombre des pêchés et les paramètres de loi a priori sont donnés dans le tableau (3.14).

| μ | σ | Vecteur des données observées (D) | Nombre d'observations (N) |
|-------|----------|-----------------------------------|---------------------------|
| 10000 | 2000 | 50, 40, 30, 55, 47, 15 | 6 |

TABLE 3.14 – Les données observés dans le cas $D|\theta \sim \mathcal{P}(\theta)$ et $\theta \sim \mathcal{N}(\mu, \sigma^2)$.

La figure (3.25), représente l'histogramme a posteriori et la distribution de la taille de la population estimée à partir d'une loi de vraisemblance Poisson $\mathcal{P}(\theta)$ et d'une loi a priori normale $\mathcal{N}(\mu, \sigma^2)$.

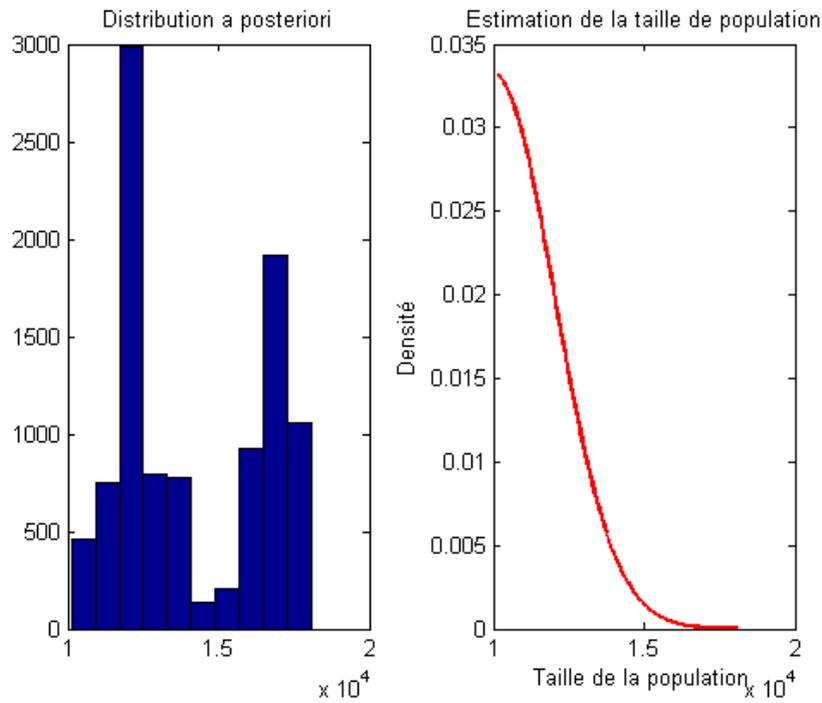


FIGURE 3.25 – À gauche l’histogramme de la distribution a posteriori, à droite la densité de l’estimation de la taille de population des poissons, pour $D|\theta \sim \mathcal{P}(\theta)$ et $\theta \sim \mathcal{N}(\mu, \sigma^2)$.

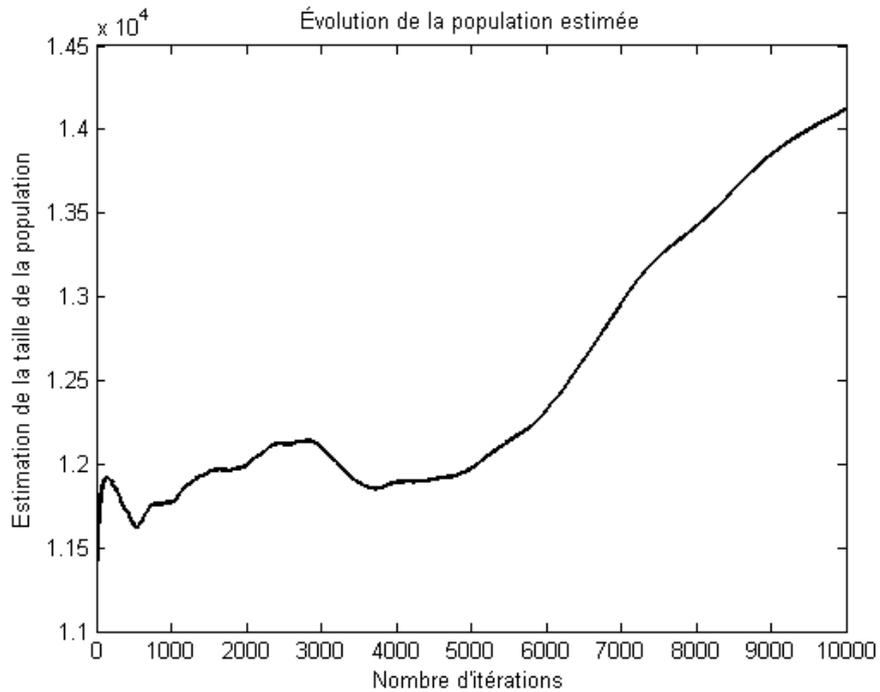


FIGURE 3.26 – L’évolution de la population des poissons estimée, pour $D|\theta \sim \mathcal{P}(\theta)$ et $\pi(\theta) \sim \mathcal{N}(\mu, \sigma^2)$.

La Figure (3.26) présente les résultats de l’estimation de la population de poissons dans un lac en utilisant l’algorithme Metropolis-Hastings. L’estimation est basée sur une vraisemblance

de Poisson $\mathcal{P}(\theta)$ et une loi a priori normale $\mathcal{N}(\mu, \sigma^2)$. La figure permet de suivre l'influence de la connaissance a priori et des nouvelles données observées sur l'estimation de la population de poissons.

Pour 10 000 itérations et avec une période de chauffe de 1 000 itérations, l'estimation obtenue est d'environ 14 123 poissons dans le lac.

D. Cas 4 : $D|\theta \sim \mathcal{P}(\lambda)$ et $\theta \sim \mathcal{U}([a, b])$

Pour le dernier cas, nous supposons que les observations suivent la loi de poisson $\mathcal{P}(\theta)$, et la loi a priori est la loi uniforme $\mathcal{U}([a, b])$.

L'organigramme de calcul de la distribution a posteriori dans ce cas est donné dans la figure (3.27).

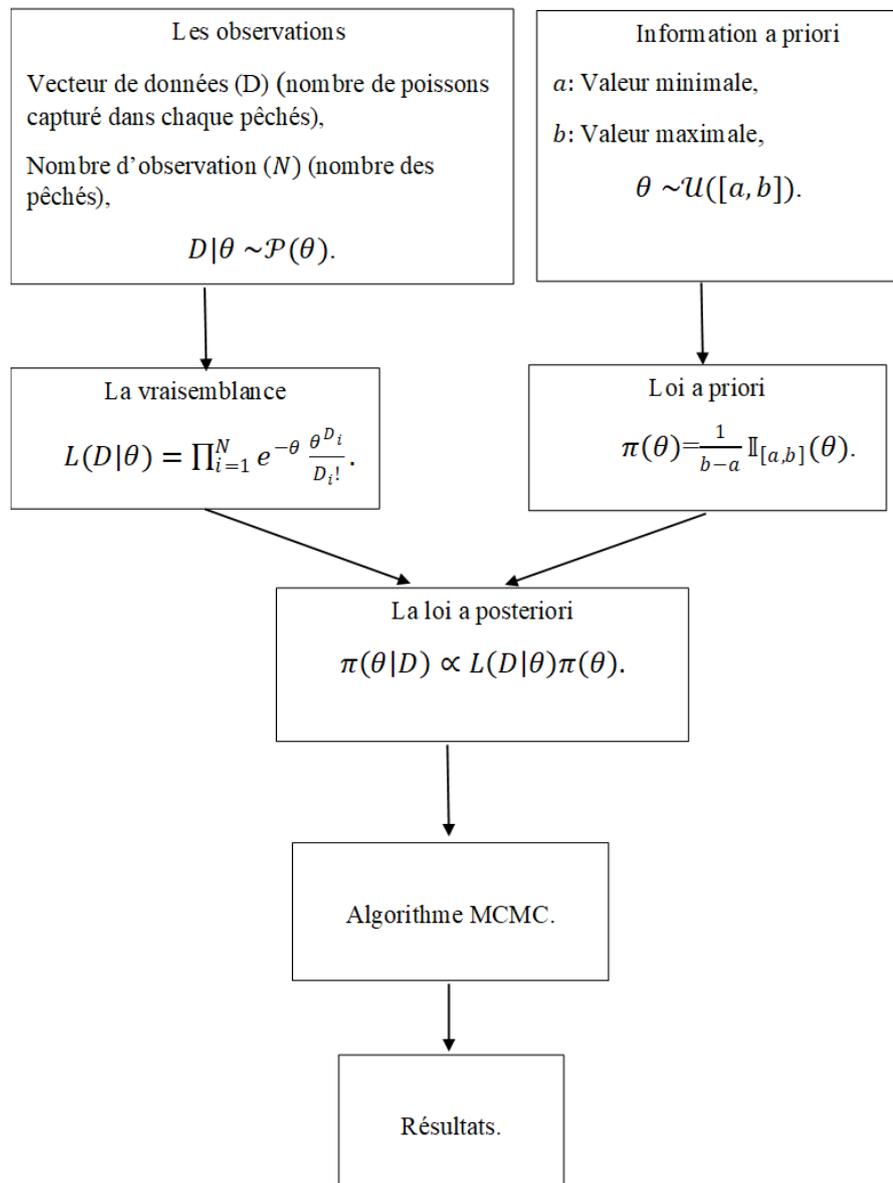


FIGURE 3.27 – Organigramme de calcul de la distribution a posteriori avec $D|\theta \sim \mathcal{P}(\lambda)$ et $\theta \sim \mathcal{U}(a, b)$.

Comme pour le cas précédent, le tableau (3.15), présente le vecteur des données observées D , le nombre des observations N et les paramètres de la loi a priori.

| a | b | Le vecteur des données observées (D) | Le nombre des observations (N) |
|-------|-------|--|------------------------------------|
| 10000 | 12000 | 50, 40, 30, 55, 47, 15 | 6 |

TABLE 3.15 – Les données observés dans le cas $D|\theta \sim \mathcal{P}(\theta)$ et $\theta \sim \mathcal{U}(a, b)$.

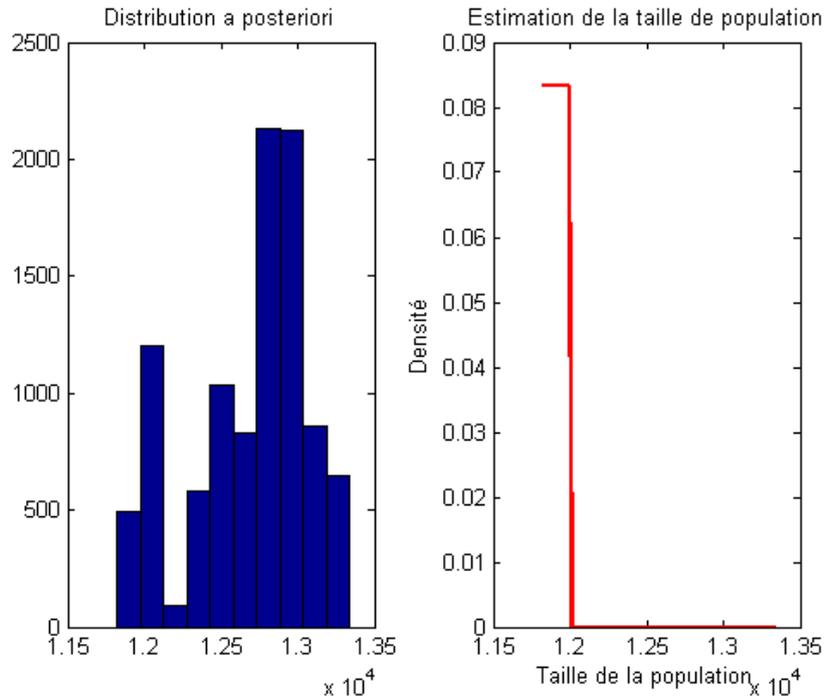


FIGURE 3.28 – À gauche l’histogramme de la distribution a posteriori, à droite la densité de l’estimation de taille de population des poissons, pour $D|\theta \sim \mathcal{P}(\theta)$ et $\theta \sim \mathcal{U}([a, b])$.

La figure (3.28), représente l’histogramme a posteriori et la distribution de la taille de la population estimée à partir d’une vraisemblance binomiale $\mathcal{P}(\theta)$ et d’une loi a priori uniforme $\mathcal{U}([a, b])$, en utilisant l’algorithme Metropolis-Hastings.

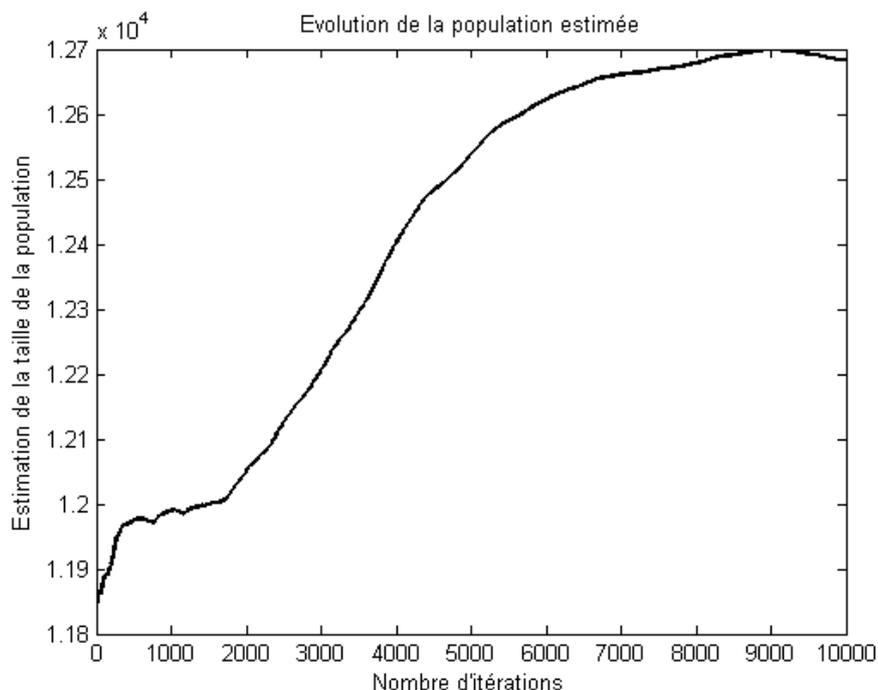


FIGURE 3.29 – L'évolution de la population des poissons estimés, pour $D|\theta \sim \mathcal{P}(\theta)$ et $\theta \sim \mathcal{U}([a, b])$.

La Figure (3.29) présente les résultats de l'estimation de la population de poissons dans un lac en utilisant l'algorithme Metropolis-Hastings. L'estimation est basée sur une vraisemblance de Poisson $\mathcal{P}(\theta)$ et une loi uniforme $\mathcal{U}([a, b])$ pour la loi a priori. La figure permet de suivre l'évolution de l'estimation de la population des poissons en fonction de la connaissance a priori et des nouvelles données observées.

Pour 10 000 itérations et pour une période de chauffe de 1 000, l'estimation obtenue est d'environ 12 682 poissons dans le lac. Cette estimation est calculée en combinant les informations fournies par la vraisemblance de Poisson et la distribution a priori uniforme. Elle suggère qu'il y a probablement environ 12 682 individus dans la population de poissons du lac, en se basant sur les données et les connaissances disponibles.

3.2.4 Estimation du poids moyen d'une population de poissons dans un lac

Dans cet exemple, nous nous intéressons à l'estimation du poids moyen d'une population de poissons dans un lac, pour cela nous proposons plusieurs scénarios différents.

===== Scénario 1 =====

Supposons que l'on veuille estimer le poids moyen d'une population de poissons dans un lac spécifique. Pour cela, nous avons collecté un ensemble d'observations (poids individuels des poissons) pour cette population.

Modèle

- Paramètre inconnu : θ (poids moyen de la population de poissons dans le lac).
- Données observées : Y (poids des poissons observés dans l'échantillon).

Aspect mathématique

- Loi a priori : $\pi(\theta)$ (distribution a priori sur θ , par exemple une distribution normale avec une moyenne et un écart-type donnés).
- Loi de vraisemblance : $L(Y|\theta)$ (probabilité des données observées Y , par exemple une distribution normale avec un écart-type donné).
- Loi a posteriori : $\pi(\theta|Y)$ (probabilité de θ étant donné les données observées Y).

Objectif

Notre objectif est d'estimer la moyenne de la distribution a posteriori $\pi(\theta|Y)$ en utilisant un des algorithmes MCMC.

L'organigramme de calculer la loi a posteriori en utilisant l'algorithme MCMC est donné dans la figure (3.30).

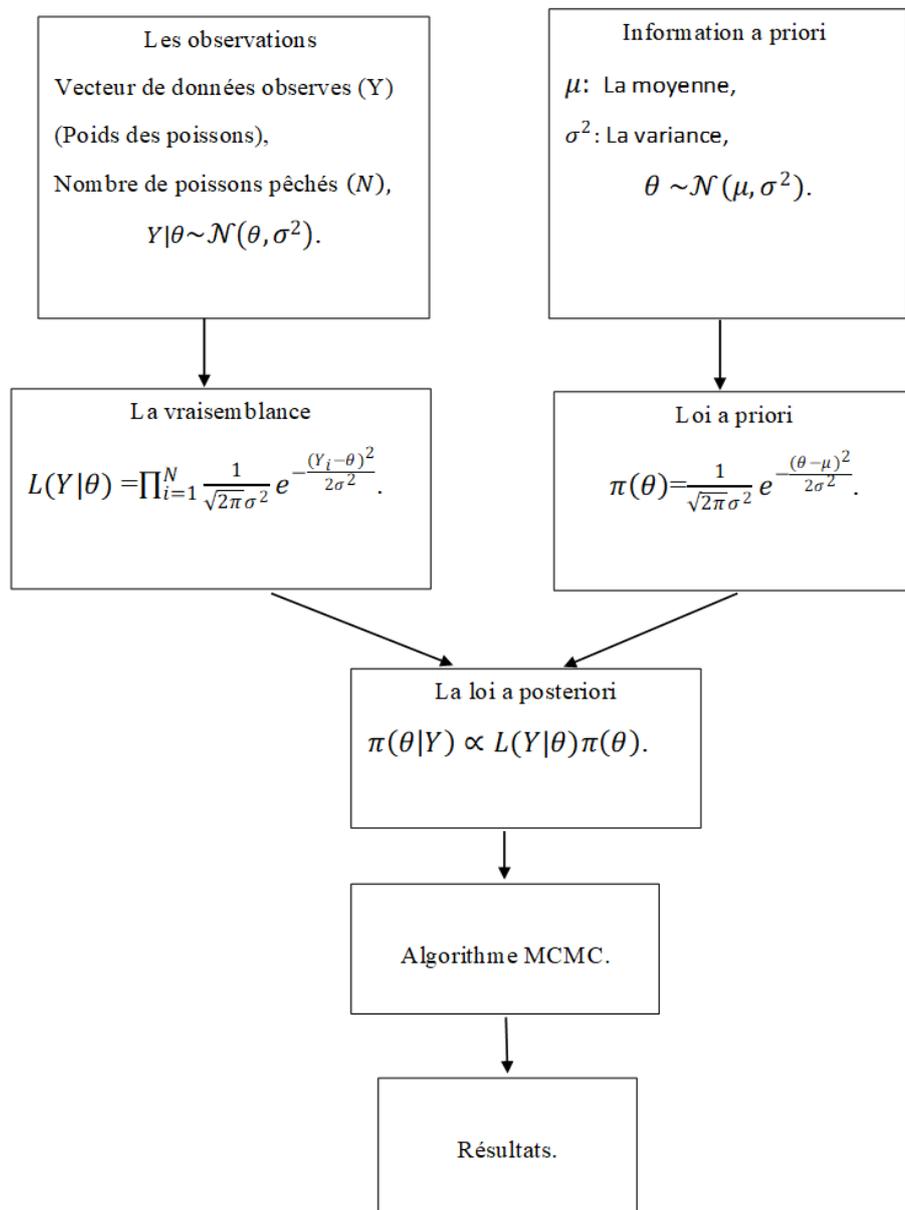


FIGURE 3.30 – Organigramme de calcul de la distribution a posteriori, scénario 1

Résultats et discussions

Le tableau (3.16), représente les données observées et les paramètres de la loi a priori.

| μ | σ | Vecteur des données observées en <i>kg</i> (Y) |
|-------|----------|--|
| 2 | 0.5 | 1.2, 1.5, 1.8, 2.1, 1.9, 1.6, 2.3, 2.5, 2.0, 2.2 |

TABLE 3.16 – Les données observées et les paramètres de la loi a priori, scénario 1.

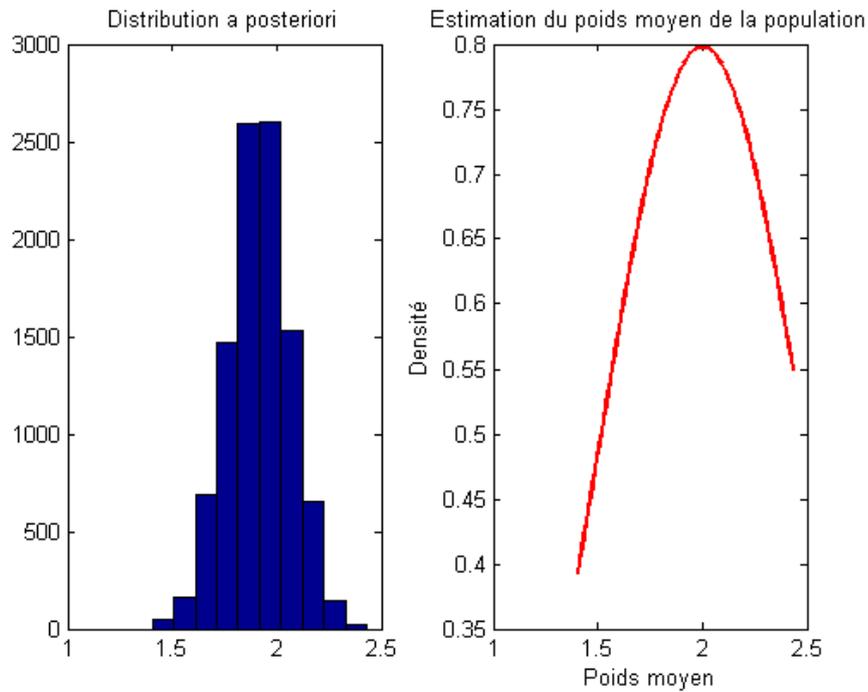


FIGURE 3.31 – À gauche l’histogramme de la distribution a posteriori, à droite la densité de l’estimation de poids moyen des poissons, pour $Y|\theta \sim \mathcal{N}(\theta, \sigma^2)$ et $\theta \sim \mathcal{N}(\mu, \sigma^2)$, en utilisant l’algorithme de Metropolis-Hastings.

La figure (3.31), représente l’histogramme a posteriori et la distribution de poids estimés à partir d’une loi de vraisemblance normale $\mathcal{N}(\theta, \sigma^2)$ et une loi a priori normale $\mathcal{N}(\mu, \sigma^2)$, en utilisant l’algorithme de Metropolis-Hastings.

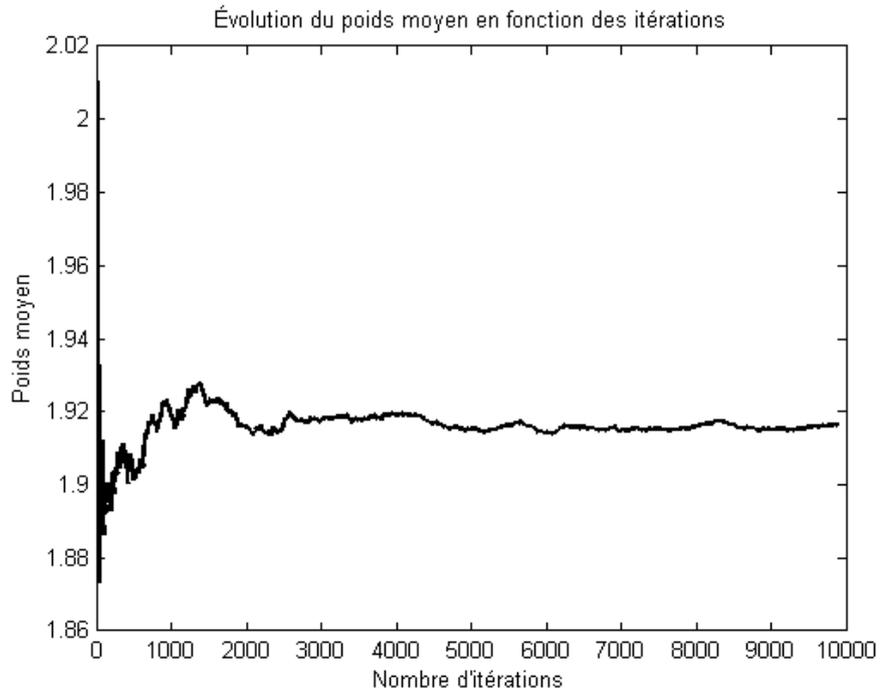


FIGURE 3.32 – L'évolution de poids moyen estimé des poissons, pour $Y|\theta \sim \mathcal{N}(\theta, \sigma^2)$ et $\theta \sim \mathcal{N}(\mu, \sigma^2)$, en utilisant l'algorithme de Metropolis-Hastings.

La figure (3.32), illustre les résultats de l'estimation du poids moyen des poissons obtenue par l'application de l'algorithme Metropolis-Hastings. Après 10000 itérations et avec une période de chauffe de 100, l'estimation obtenue du poids moyen des poissons est d'environ 1.9184 *kg*. Cela signifie que, selon les informations fournies par les données observées et l'algorithme Metropolis-Hastings, le poids moyen des poissons dans cette population est estimé à être d'environ 1.9184 *kg*.

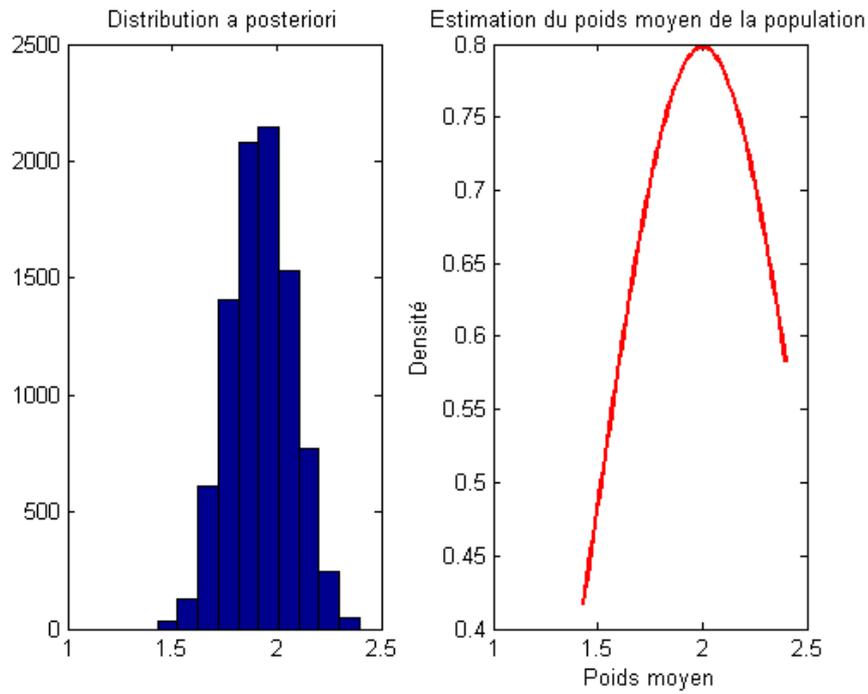


FIGURE 3.33 – À gauche l’histogramme de la distribution a posteriori, à droite la densité de l’estimation de poids moyen des poissons, pour $Y|\theta \sim \mathcal{N}(\theta, \sigma^2)$ et $\theta \sim \mathcal{N}(\mu, \sigma^2)$, en utilisant l’algorithme de Gibbs.

La figure (3.33), représente l’histogramme a posteriori et la distribution de poids estimés de poissons, en utilisant l’algorithme de Gibbs, pour $Y|\theta \sim \mathcal{N}(\theta, \sigma^2)$ et $\theta \sim \mathcal{N}(\mu, \sigma^2)$.

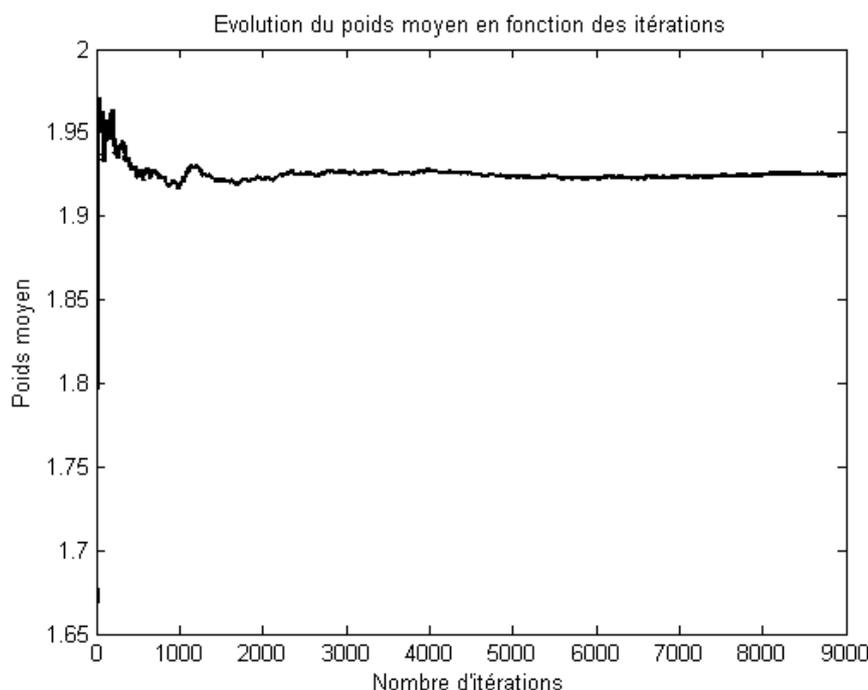


FIGURE 3.34 – L'évolution de poids moyen estimés des poissons, pour $Y|\theta \sim \mathcal{N}(\theta, \sigma^2)$ et $\theta \sim \mathcal{N}(\mu, \sigma^2)$, en utilisant l'algorithme de Gibbs.

La figure (3.34), donne les résultats de l'évolution de poids estimés de poissons, pour $Y|\theta \sim \mathcal{N}(\theta, \sigma^2)$ et $\theta \sim \mathcal{N}(\mu, \sigma^2)$. Après avoir effectué 10000 itérations de l'algorithme de Gibbs, l'estimation finale du poids moyen des poissons était de 1.9249 *kg*.

===== **Scénario 2** =====

Dans ce scénario, supposons que nous avons deux zones différentes : une zone profonde et une zone peu profonde. Nous disposons de données observées pour chaque zone, qui sont les poids des poissons capturés. L'objectif est d'estimer le poids moyen de la population de poissons dans le lac et de calculer les intervalles de crédibilité à 95% pour ces estimations.

Modèle

- Paramètre inconnu : θ_p (poids moyen de la population de poissons dans la zone profonde), θ_{pp} (poids moyen de la population de poissons dans la zone peu profonde).
- Données observées : Y_p (vecteur des poids des poissons observés dans la zone profonde), Y_{pp} (Vecteur des poids des poissons observés dans la zone peu profonde).

Aspect mathématique

- Loi a priori :
 - $\pi(\theta_p)$: distribution a priori dans la zone profonde.

- $\pi(\theta_{pp})$: distribution a priori dans la zone peu profonde.
- Loi de vraisemblance :
- $L(Y_p|\theta_p)$: loi de vraisemblance dans la zone profonde.
 - $L(Y_{pp}|\theta_{pp})$: loi de vraisemblance dans la zone peu profonde.

Objectif

Notre objectif est d'estimer la moyenne de la moyenne des deux distributions a posteriori $\pi(\theta_p|Y_p)$ et $\pi(\theta_{pp}|Y_{pp})$, ainsi que de construire les intervalles de crédibilité à 95% en utilisant l'algorithme MCMC.

L'organigramme de calcul de la distribution a posteriori est donné dans figure (3.35).

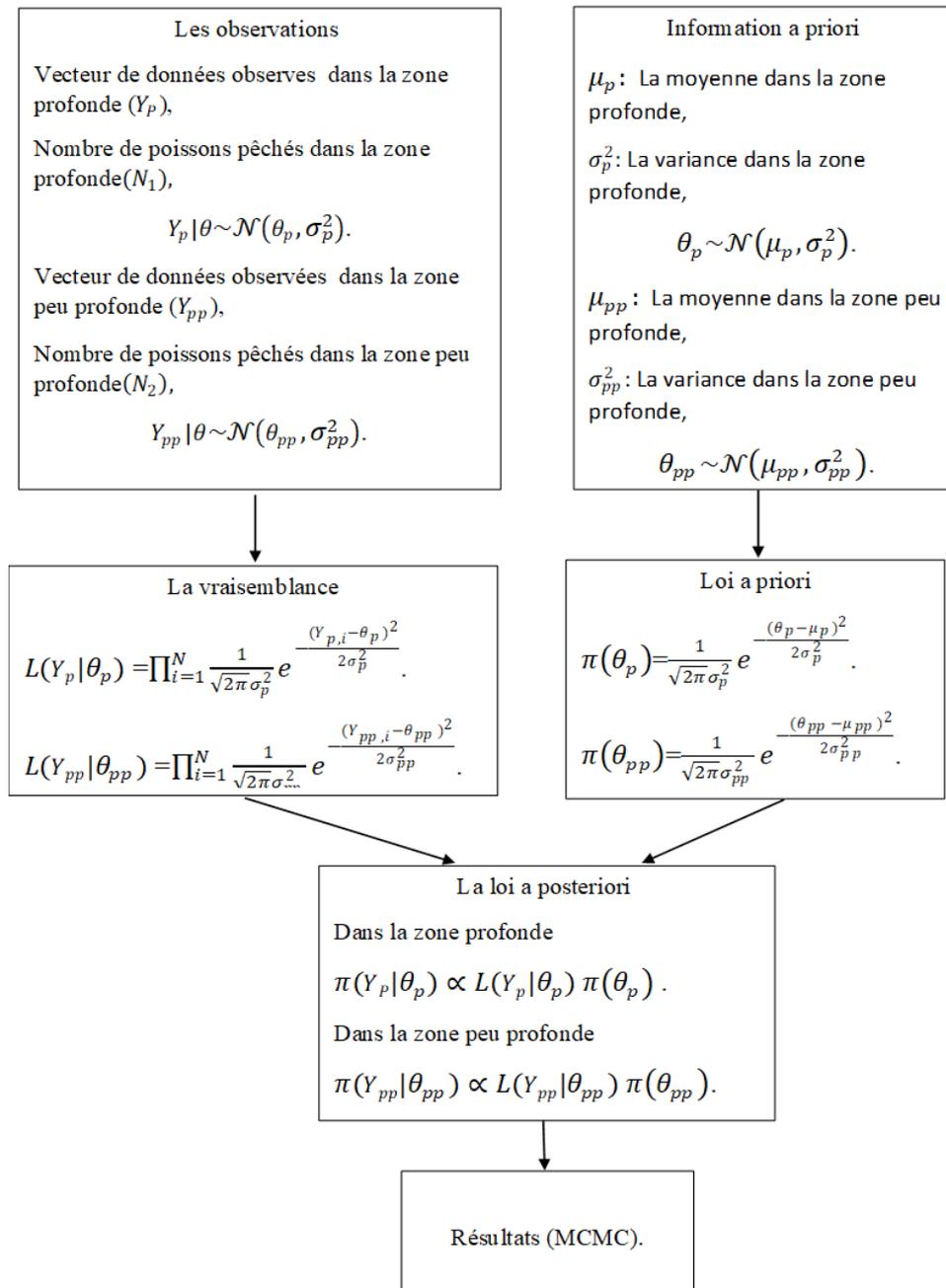


FIGURE 3.35 – Organigramme de calcul de la distribution a posteriori dans scénario 2

Résultats et discussions

La moyenne et l'écart-type de la loi a priori, ainsi que les poids observés des poissons sont donnés dans le tableau (3.17).

| Zone | La moyenne | Écart-type | Vecteur de données observées en kg (Y) |
|--------------|------------|------------|--|
| Profonde | 2 | 0.5 | 1.9, 1.6, 2.3, 2.5, 2.0, 2.2 |
| Peu profonde | 1.5 | 0.5 | 1.2, 1.5, 1.8, 2.1 |

TABLE 3.17 – Les paramètres de la loi a priori et les données observées dans le scénario 2.

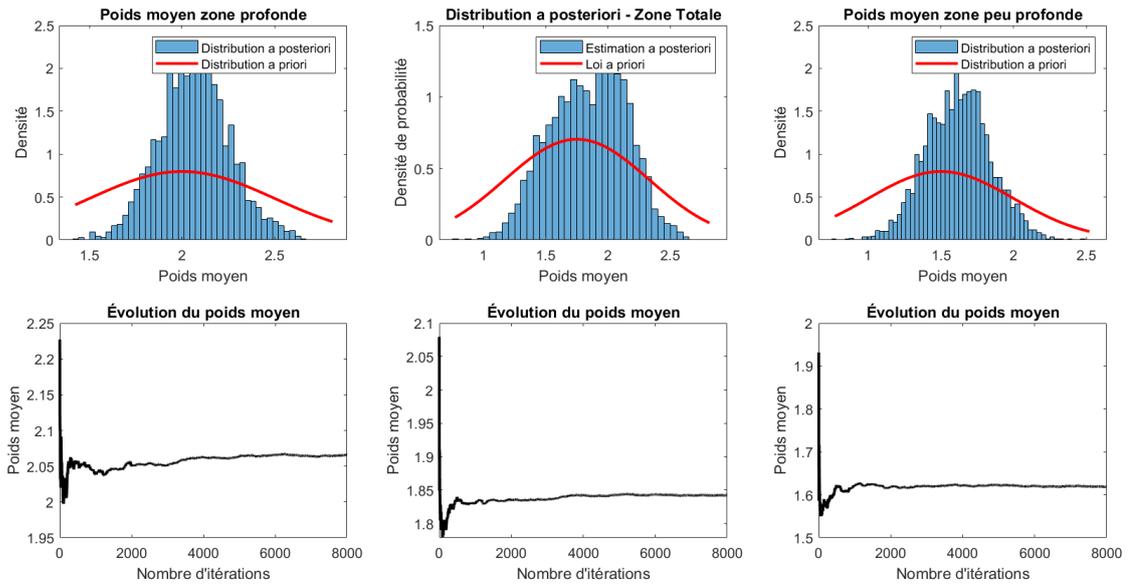


FIGURE 3.36 – L’histogramme de la distribution a posteriori, la densité de l’estimation de poids moyen des poissons et l’évolution de poids moyen estimé des poissons. À gauche dans la zone profonde, au milieu dans le Lac et à droite dans la zone peu profonde.

La figure (3.36), représente l’histogramme de la distribution a posteriori, la distribution a priori et l’évolution de poids moyen des poissons estimé dans différentes zones du lac.

Le tableau (3.18), donne les résultats de l’estimation de poids de poissons dans la zone profonde, peu profonde et dans le lac.

| Zone | Poids moyen | Intervalle de crédibilité (95%) |
|--------------|-------------|---------------------------------|
| Profonde | 2.066 | [1.703, 2.4688] |
| Peu profonde | 1.6183 | [1.1901, 2.05] |
| Lac | 1.8422 | [1.5547, 2.1237] |

TABLE 3.18 – Poids moyen estimé des poissons et les intervalles de crédibilité dans les trois zones : profonde, peu profonde et dans le lac.

Supposons maintenant que nous ne disposons pas d’informations sur les poids moyens des poissons dans les deux zones spécifiques, mais plutôt que nous avons cette information pour l’ensemble du lac. Contrairement à l’exemple précédent, cela signifie que nous disposons d’une seule information a priori.

Dans cette situation, la figure (3.37) présente l’histogramme a posteriori, la distribution des poids estimés des poissons, ainsi que l’évolution de l’estimation des poids des poissons dans la zone profonde, la zone peu profonde et dans l’ensemble du lac. Cette figure nous permet de visualiser les résultats obtenus en utilisant cette unique information a priori sur les poids moyens des poissons dans le lac.

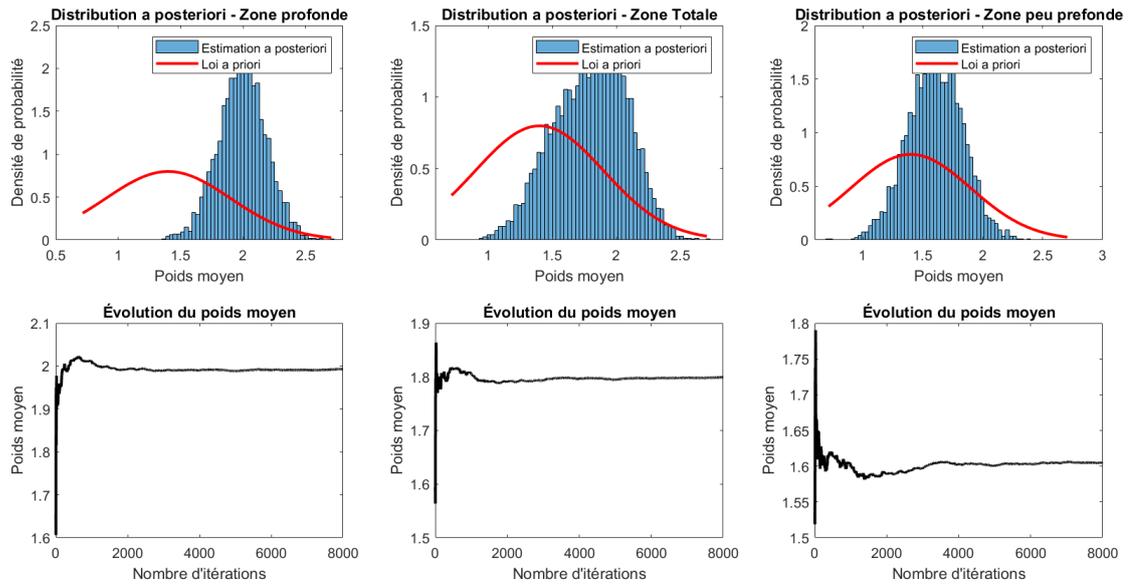


FIGURE 3.37 – L’histogramme de la distribution a posteriori, la distribution a priori et l’évolution de poids moyen estimé des poissons. À gauche dans la zone profonde, au milieu dans le Lac et à droite dans la zone peu profonde.

Les estimations des poids des poissons dans la zone profonde, la zone peu profonde et dans l’ensemble du lac sont présentées dans le tableau (3.19). Ce tableau fournit les résultats obtenus à partir de l’application d’une méthode bayésienne MCMC pour estimer les poids des poissons dans ces différentes zones.

| Zone | Poids moyen | Intervalle de crédibilité (95%) |
|--------------|-------------|---------------------------------|
| Profonde | 1.9588 | [1.5844, 2.327] |
| Peu profonde | 1.5616 | [1.1386, 2.001] |
| Lac | 1.7602 | [1.4805, 2.0371] |

TABLE 3.19 – Poids moyen estimés des poissons et les intervalle de crédibilité dans la région profonde, peu profonde et dans le lac.

En examinant de plus près les figures (3.36) et (3.37), ainsi que les résultats des tableaux (3.18) et (3.19), il devient évident que l’existence d’informations a priori dans les deux zones distinctes, c’est-à-dire la zone profonde et la zone peu profonde, a un impact positif significatif sur la précision des estimations.

Lorsque nous disposons d’informations a priori uniquement pour l’ensemble de la zone totale, nous ne pouvons pas prendre en compte les particularités et les caractéristiques propres à chaque zone spécifique. Cela limite notre capacité à modéliser précisément les variations et les comportements différents qui peuvent exister entre la zone profonde et la zone peu profonde. En conséquence, les estimations obtenues dans ce cas peuvent être moins précises et moins fiables.

En revanche, lorsque nous disposons d’informations a priori dans les deux zones distinctes, nous sommes en mesure de tenir compte des différences entre ces deux zones et de les intégrer dans notre modèle. Cela nous permet d’affiner nos estimations en prenant en considération les caractéristiques spécifiques à chaque zone. Ainsi, les résultats obtenus en incorporant ces informations a priori dans les deux zones sont généralement plus précis et plus proches de la réalité.

===== Scénario 3 =====

Dans ce scénario, on s'intéresse à l'estimation du poids moyen des poissons dans un lac en fonction de deux facteurs à savoir, la profondeur de l'eau et la disponibilité de nourriture.

Modèle

Nous considérons un modèle de régression linéaire où le poids moyen des poissons (*poids*) est modélisé en fonction de deux variables : la profondeur de l'eau (*prof*) et la quantité de nourriture disponible par mètre cube (*nour*). Le modèle peut être formulé comme suit :

$$poids = \alpha \cdot prof + \beta \cdot nour + \epsilon,$$

où α est le coefficient de régression associé à la variable (*prof*), β est le coefficient de régression associé à la variable (*nour*) et ϵ est le terme d'erreur qui suit une distribution normale avec une moyenne de zéro et une variance constante.

Aspect mathématique

— Loi a priori pour les paramètres α et β :

- $\alpha \sim \mathcal{N}(\mu_\alpha, \sigma_\alpha^2)$,
- $\beta \sim \mathcal{N}(\mu_\beta, \sigma_\beta^2)$.

— Loi de vraisemblance des données observées de poids : $(D|a, b) \sim \mathcal{N}(\alpha \cdot prof + \beta \cdot nour, \sigma^2)$.

Objectif

L'objectif est d'estimer les valeurs des coefficients de régression α et β à partir des données observées de poids moyen des poissons, de profondeur de l'eau et de quantité de nourriture disponible.

L'organigramme qui donne les lois a prioris, la loi de vraisemblance et la distribution a posteriori est donné dans la figure (3.38).

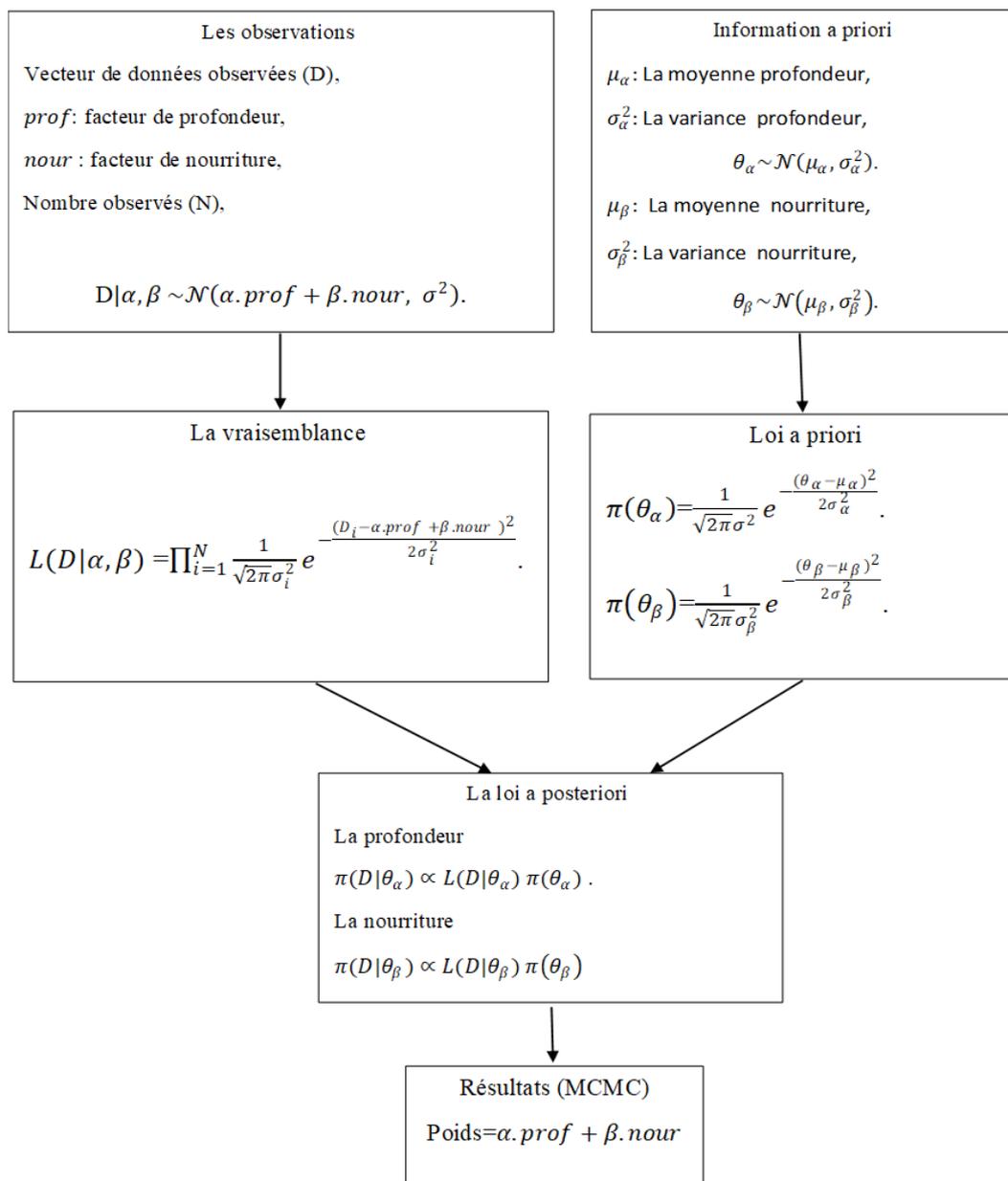


FIGURE 3.38 – Organigramme de calcul de la distribution a posteriori dans scénario 3

Résultats et discussions

Les poids observés des poissons, données observées de profondeur de l'eau en m et données observées de quantité de nourriture disponible par m^3 sont donnés dans le tableau (3.20).

| | |
|-----------------------|--|
| Données observées (D) | 2.5, 3.1, 2.8, 3.5, 2.9, 3.2, 3.8, 3.6 |
| Facteur <i>prof</i> | 10, 12, 8, 15, 10, 13, 16, 14 |
| Facteur <i>nour</i> | 0.8, 0.9, 0.7, 0.6, 0.5, 0.8, 0.9, 0.7 |

TABLE 3.20 – Les poids observés des poissons (D), de profondeur de l'eau en m (*prof*) et de quantité de nourriture disponible par m^3 (*nour*) dans le scénario 2.

En utilisant l'algorithme MCMC dans le dernier programme, nous avons effectué 10 000

itérations pour estimer le poids moyen des poissons. Les échantillons générés à partir de ces itérations nous ont permis d'obtenir une estimation du poids moyen des poissons d'environ 3.15 kg. Cela signifie que, selon notre modèle basé sur les données observées de profondeur de l'eau et de quantité de nourriture disponible, nous estimons que le poids moyen des poissons se situe autour de 3.15 kg.

De plus, nous avons également obtenu des estimations des coefficients de régression associés à la variable "profondeur" et à la variable "nourriture". Les valeurs moyennes estimées pour la variable "profondeur" sont d'environ 0.23, ce qui suggère une relation positive mais faible entre la profondeur de l'eau et le poids des poissons. Les valeurs moyennes estimées pour la variable "nourriture" sont d'environ 0.47, ce qui indique une relation positive et plus marquée entre la quantité de nourriture disponible et le poids des poissons.

La figure (3.39) illustre graphiquement l'évolution du poids moyen des poissons au fur et à mesure des itérations de l'algorithme MCMC. On peut observer que les échantillons convergent progressivement vers une valeur centrale, représentant notre estimation du poids moyen des poissons.

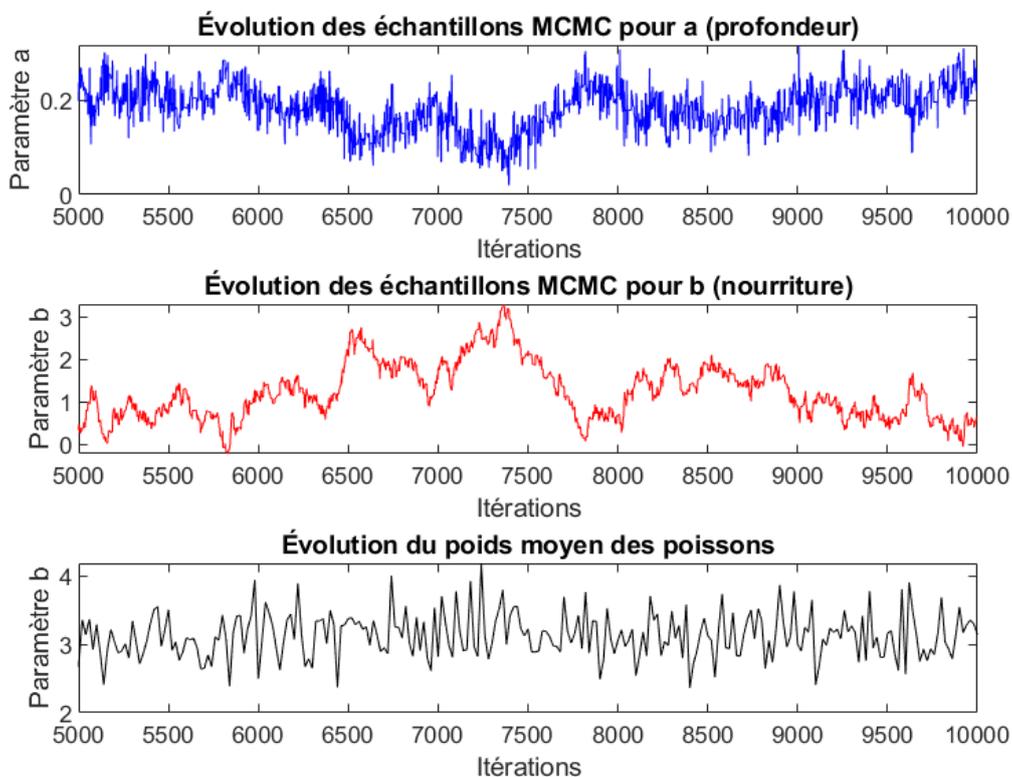


FIGURE 3.39 – L'évolution de poids moyen des poissons par rapport à l'évolution des échantillons MCMC pour la profondeur et la nourriture.

3.3 Conclusion

Dans ce chapitre, nous avons appliqué les méthodes MCMC à deux problèmes de statistique bayésienne concrets. Le premier exemple concernait la régression bayésienne pour étudier les crises cardiaques chez les hommes de 55 à 59 ans dans différents pays. Les méthodes MCMC

ont permis d'estimer les paramètres du modèle de régression et d'analyser la relation entre les variables indépendantes et la probabilité de crises cardiaques, offrant des informations précieuses dans le domaine médical.

Dans le deuxième exemple, nous avons abordé l'estimation de la taille (nombre) et du poids moyen des poissons dans un lac. En appliquant les méthodes MCMC et en tenant compte de différents scénarios, tels que les zones de pêche et les conditions environnementales, nous avons réussi à estimer ces paramètres essentiels. Ces estimations peuvent fournir des informations cruciales pour une gestion et une conservation optimales des ressources aquatiques, permettant ainsi de prendre des décisions éclairées en matière de gestion des populations de poissons dans le lac.

Ces applications concrètes démontrent l'efficacité des méthodes MCMC en statistique bayésienne pour obtenir des estimations fiables et approfondir les analyses dans des domaines variés.

Conclusion et perspectives

Conclusion

En résumé, notre étude a porté sur l'application des méthodes MCMC dans le cadre de l'approche bayésienne, avec un accent particulier sur la modélisation des problèmes complexes. Les résultats obtenus mettent en évidence le rôle crucial de ces méthodes dans la construction de modèles statistiques robustes et dans l'obtention d'estimations probabilistes précises des paramètres inconnus.

L'accent mis sur la modélisation a permis de mieux appréhender les relations entre les variables, de tenir compte des incertitudes et d'obtenir des résultats plus fiables. Les exemples concrets présentés dans notre étude, tels que la modélisation des risques de crises cardiaques et l'estimation de la taille de la population de poissons, démontrent l'efficacité des méthodes MCMC dans la résolution de problèmes réels et complexes.

En utilisant les méthodes MCMC, nous avons pu mettre en évidence l'importance de prendre en compte les connaissances a priori, d'intégrer les données observées et de générer des échantillons représentatifs de la distribution a posteriori. Cela permet de prendre des décisions éclairées et de mieux comprendre les phénomènes étudiés.

Il est clair que les méthodes MCMC ont un impact significatif dans de nombreux domaines, tels que l'écologie, la finance, les mathématiques, la physique, la biologie, la médecine et la génétique. Leur flexibilité et leur capacité à gérer des modèles complexes font d'elles un outil précieux pour les chercheurs et les praticiens.

En conclusion, l'utilisation des méthodes MCMC dans l'approche bayésienne accentue notre capacité à modéliser et à résoudre des problèmes réels de manière précise et fiable. Ces méthodes offrent une approche solide pour explorer les distributions de probabilité complexes, intégrer les connaissances a priori et améliorer la prise de décision dans divers domaines scientifiques.

Perspectives

Nous avons identifié plusieurs perspectives futures qui peuvent enrichir et étendre notre travail sur l'application des méthodes MCMC dans l'approche bayésienne. Tout d'abord, il serait intéressant d'explorer des méthodes MCMC plus avancées et adaptées aux modèles de plus grande dimension et de complexité accrue. Cela pourrait inclure des techniques de rééchantillonnage adaptatif, des algorithmes de dynamique hamiltonienne ou des approches basées sur l'apprentissage automatique bayésien.

Une autre perspective de recherche consiste à améliorer les techniques de sélection de modèles et d'estimation des hyperparamètres dans le cadre de l'inférence bayésienne. Cela permettrait d'obtenir des modèles plus précis et mieux adaptés aux données, en optimisant la balance entre complexité et ajustement.

Enfin, l'exploration de nouvelles applications dans des domaines émergents tels que l'apprentissage automatique bayésien, l'intelligence artificielle ou l'analyse des réseaux complexes ouvrirait de nouvelles perspectives de recherche.

En poursuivant ces perspectives, nous pourrions améliorer notre compréhension des phénomènes complexes, obtenir des modèles plus précis et élargir les applications de la modélisation bayésienne dans divers domaines scientifiques et appliqués.

Annexe A

Annexe

A.1 Simulation de variables aléatoires

A.1.1 Générateur de nombres uniformes

Lorsqu'on aborde la simulation de phénomènes aléatoires, on commence tout d'abord par introduire les générateurs de nombres uniformes (simulation de la loi uniforme sur l'intervalle $[0, 1]$). La simulation de la loi uniforme sur l'intervalle $[0, 1]$ est souvent la première étape dans la simulation d'autres lois de probabilité.

De nombreux algorithmes de génération de nombres pseudo-aléatoires ont pour but de produire des suites uniformément distribuées. Parmi les méthodes courantes, on trouve la méthode des congruences.

Méthode des congruences

La méthode des congruences est fondée sur un algorithme simple pour la génération de nombres pseudo-aléatoires. Cette méthode repose sur une relation de récurrence qui permet de générer une suite de nombres aléatoires à partir d'une graine initiale X_0 et de coefficients constants. En effet cette dernière est définie par :

$$X_i = (a X_{i-1} + b) \text{ mod } m, \quad i = 1, 2, \dots, n - 1, \quad (\text{A.1})$$

où a , b , m et X_0 sont des entiers positifs données. On dit que a est le multiplicateur, b est l'incrément, m le modulo, n est la taille de l'échantillon à générer.

Le nombre pseudo-aléatoire, U_i , $i = 0, \dots, n - 1$, compris entre 0 et 1, est obtenu par la relation,

$$U_i = \frac{X_i}{m}. \quad (\text{A.2})$$

La longueur d'une suite générée par une relation récursive du type $((aX_{i-1} + b) \text{ mod } m)$ est limitée à m , car il y a au plus m nombres différents modulo m . Pour obtenir une suite de longueur maximale, il est généralement recommandé de choisir une grande valeur de m . Pour garantir une période maximale de la suite, il est important que les conditions du théorème suivant soient satisfaites.

Théorème 9. (Hull et Dobell (1962)) [5]

Soit $k \in \{0, 1, \dots, m - 1\}$. Soient a , b , m , tels que :

1. b et m sont premiers entre eux ;
2. $(a - 1)$ est un multiple de chaque nombre premier qui divise m ;
3. si m est un multiple de 4 alors $(a - 1)$ l'est aussi.

Alors la suite définie par :

$$\begin{cases} X_0 = k, \\ X_i = (aX_{i-1} + b) \text{ mod } m, \end{cases}$$

a un cycle de longueur m .

A.1.2 Générateur de nombres non-uniformes

Dans la section précédente, nous avons vu la manière de simuler ou de générer des suites de nombres aléatoires qui sont distribuées uniformément sur l'intervalle $[0, 1]$. Dans cette section, nous allons présenter quelques méthodes pour générer des nombres aléatoires qui ne sont pas distribués uniformément. Dans ce cas, nous parlons de la simulation de variables aléatoires qui suivent une loi de probabilité quelconque.

Il existe plusieurs méthodes pour générer des nombres non uniformes, notamment :

- La méthode d'inversion,
- La méthode de Box-Müller,
- La méthode d'acceptation rejet.

Méthodes d'inversion

La méthode d'inversion est une méthode courante pour générer des nombres aléatoires. Comme son nom l'indique cette méthode est fondée sur l'inversion de la fonction de répartition. Cette méthode est basée sur le théorème suivant,

Théorème 10. Soit X une v.a de fonction de répartition F , continue et strictement croissante, on a,

Si $U \sim \mathcal{U}([0, 1])$ Alors $F^{-1}(U)$ à même loi que X .

Autrement dit, il suffit de simuler U suivant $\mathcal{U}([0, 1])$ puis appliquer la transformation $X = F^{-1}(U)$ pour simuler suivant la loi de X .

Preuve 12. Supposons que X continue, F croissante et F^{-1} existe.

On pose $X = F^{-1}(U)$. Alors

$$P(X \leq x) = P(F^{-1}(U) \leq x) = P(U \leq F(x)) = F(x).$$

□

L'algorithme de cette méthode est donné comme suit :

Algorithme de simulation par inversion

Debut

lire(n); %nombre d'échantillon

$i = 1$;

pour $i = 1 : n$ **faire**

 générer $U(i) \sim \mathcal{U}([0, 1])$;

$X(i) = F^{-1}(U(i))$;

fin pour;

afficher (X);

Fin.

Limite de la méthode d'inversion

L'application de la méthode d'inversion nécessite la connaissance et le calcul rapide de la fonction inverse. Ceci n'est pas toujours possible, on peut prendre l'exemple de la loi normale centrée réduite de densité :

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad x \in \mathbb{R},$$

dont la fonction de répartition n'admet pas d'expression analytique simple et encore moins pour sa fonction inverse.

Il existe plusieurs méthodes pour simuler une variable aléatoire qui suit une loi normale. Ces méthodes reposent sur :

1. L'application du Théorème Central Limite.
2. L'algorithme du rejet aux densités à support non compact.
3. La méthode de Box-Müller.

Dans ce qui suit, nous nous limiterons à la méthode de Box-Müller.

Méthode de Box-Müller

La méthode de Box-Müller permet de simuler un couple de variables aléatoires normales, centrées, réduites, et indépendantes.

Supposons que l'on veuille simuler $X \sim \mathcal{N}(0, 1)$ et $Y \sim \mathcal{N}(0, 1)$ indépendantes. On connaît la densité jointe de X et Y :

$$f(x, y) = \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}}, \quad (x, y) \in \mathbb{R}^2.$$

Proposition 11. *Si U, V sont de loi $\mathcal{U}([0, 1])$ et indépendantes. On pose,*

$$\begin{cases} X = \sqrt{-2 \log(U)} \cos(2\pi V), \\ Y = \sqrt{-2 \log(U)} \sin(2\pi V). \end{cases}$$

Alors X et Y sont indépendantes et de même loi $\mathcal{N}(0, 1)$.

Preuve 13. *Soient :*

$$X \sim \mathcal{N}(0, 1) \Rightarrow f(x)dx = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx.$$

$$Y \sim \mathcal{N}(0, 1) \Rightarrow f(y)dy = \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy.$$

Comme X et Y sont deux v.a indépendantes alors,

$$f(x, y)dxdy = \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}} dxdy.$$

On considère le changement de variables en coordonnée polaire,

$$\begin{cases} x = r \cos(\theta) \\ y = r \sin(\theta) \end{cases}$$

On a donc,

$$f(r, \theta)drd\theta = \frac{1}{2\pi} e^{-\frac{r^2}{2}} r drd\theta.$$

À partir de la dernière équation, nous avons $\frac{1}{2\pi}$ représente la densité de la loi $\Theta \sim \mathcal{U}([0, 2\pi])$. Par l'application de la méthode d'inversion vue précédemment, on aura :

$$\Theta = 2\pi V, \quad V \sim \mathcal{U}([0, 1]).$$

$re^{-\frac{r^2}{2}}$ représente la densité de la loi R . Pour déterminer la loi que suit R , nous allons étudier sa fonction de répartition $F(r)$.

$$F(r) = \mathbb{P}(R \leq r) = \int_{-\infty}^r te^{-\frac{t^2}{2}} dt = [-e^{-\frac{t^2}{2}}]_{-\infty}^r = 1 - e^{-\frac{r^2}{2}}.$$

Si on pose $R' = R^2$ alors on reconnait une loi exponentielle de paramètre $\frac{1}{2}$.

Simulation de la loi exponentielle

Pour la simulation de la loi exponentielle de paramètre $\frac{1}{2}$, on utilise la méthode d'inversion, pour $\lambda = \frac{1}{2}$, $R' = -2 \ln(U)$ avec $U \sim \mathcal{U}([0, 1])$. Donc $R = \sqrt{-2 \ln(U)}$.

D'où la relation suivante :

$$\begin{cases} X = R \cos(\Theta), \\ Y = R \sin(\Theta), \end{cases} \Rightarrow \begin{cases} X = \sqrt{-2 \ln(U)} \cos(2\pi V), \\ Y = \sqrt{-2 \ln(U)} \sin(2\pi V). \end{cases}$$

□

L'algorithme de cette méthode est donné par :

Algorithme de Box-Müller

Debut

lire(n);

pour $i = 1 : n$ **faire**

 générer $U(i) \sim \mathcal{U}([0, 1])$;

 générer $V(i) \sim \mathcal{U}([0, 1])$;

$X(i) = \sqrt{-2 \log(U(i))} \cos(2\pi V(i))$;

$Y(i) = \sqrt{-2 \log(U(i))} \sin(2\pi V(i))$;

fin pour ;

Fin.

Remarque 10. 1. Pour simuler une v.a $X \sim \mathcal{N}(0, 1)$, il suffit de :

a) Tirer deux suites v.a.i.i.d $U, V \sim \mathcal{U}([0, 1])$.

b) Poser $X = \sqrt{-2 \log(U)} \cos(2\pi V)$.

2. Pour simuler une v.a $Y \sim \mathcal{N}(\mu, \sigma^2)$, il suffit de poser $Y = \mu + \sigma X$ avec $X \sim \mathcal{N}(0, 1)$.

Méthode d'acceptation rejet

Il existe plusieurs versions de la méthode de rejet, mais nous allons ici présenter la méthode la plus générale. Supposons que nous souhaitons simuler une variable aléatoire X de densité f (dans \mathbb{R}^d) et supposons également qu'il existe une loi de densité g facile à simuler telle que :

1. On sait simuler la v.a Y de densité g .
2. Il existe une constante m telle que, pour presque tout x , $f(x) \leq mg(x)$.

Considérons deux suites indépendantes de variables aléatoires :

- (a) $(Y_n)_{n \geq 1}$ i.i.d de densité g .
- (b) $(U_n)_{n \geq 1}$ i.i.d de loi uniforme.

En d'autres termes, Y représente une proposition et U représente un tirage pile ou face pour décider si on accepte ou non cette proposition. Nous utiliserons la fonction de rapport d'acceptation r pour la pile ou face, à savoir pour tout $y \in \mathbb{R}^d$:

$$r(y) = \begin{cases} \frac{f(y)}{mg(y)} & \text{si } g(y) > 0, \\ 0 & \text{sinon.} \end{cases} \quad (\text{A.3})$$

De plus, étant donné que f et g sont des densités, la constante m intervenant dans la majoration est supérieure à 1.

La proposition suivante montre comment simuler suivant la densité f voulue.

Proposition 12. *Soit $N = \inf\{n \geq 1, U_n \leq r(Y_n)\}$ le premier instant où le tirage est accepté et $X = Y_n$ la valeur de la proposition à cet instant. Alors X a pour densité f . Par ailleurs, N suit une loi géométrique de paramètre $\frac{1}{m}$.*

Remarque 11. *La variable N étant géométrique de paramètre $\frac{1}{m}$, sa moyenne est égale à m . Concrètement, il faut donc en moyenne m essais pour obtenir une seule simulation selon la loi cible f . Par conséquent, il est important de choisir le couple (g, m) de sorte que m soit aussi proche de 1 que possible. En d'autres termes, il est préférable de choisir une densité g qui ressemble le plus possible à f .*

L'algorithme de la méthode de rejet est donné comme suit :

Algorithme d'acceptation rejet

Debut

lire(n);

$i = 1$;

$j = 1$;

Tantque($i \leq n$)**faire**

 générer $U(i) \sim \mathcal{U}([0, 1])$;

 générer $Y(i)$ selon g ;

$r(i) = \frac{f(Y(i))}{m * g(Y(i))}$;

si $U(i) \leq r(i)$

$X(j) = Y(i)$;

$j = j + 1$;

fin si;

$i = i + 1$;

fin Tantque;

afficher (X);

Fin.

Annexe B

Annexe

B.1 Régression bayésienne

B.1.1 Méthode des moindres carrées

En considérant la série statistique (x_i, y_i) pour $i = 1, \dots, n$, l'équation de régression est déterminée par,

$$y_i = \alpha + \beta x_i + \epsilon_i.$$

Où,

α L'intercepte à l'origine.

β La pente.

$\epsilon_i \sim \mathcal{N}(0, \sigma^2)$ l'erreur résiduelle (*iid*).

La somme des carrés des résidus est,

$$SS_{res} = \sum_{i=1}^n \epsilon_i^2 = \sum_{i=1}^n (y_i - (\alpha + \beta x_i))^2.$$

Les valeurs α et β sont trouvées à partir de la minimisation de SS_{res} ,

$$\begin{cases} \frac{\partial SS_{res}}{\partial \alpha} = 0, \\ \frac{\partial SS_{res}}{\partial \beta} = 0, \end{cases} \iff \begin{cases} \sum_{i=1}^n 2(y_i - (\alpha + \beta x_i))(-1) = 0, \\ \sum_{i=1}^n 2(y_i - (\alpha + \beta x_i))(-x_i) = 0, \end{cases}$$

$$\begin{aligned} \Leftrightarrow & \begin{cases} \sum_{i=1}^n y_i - \sum_{i=1}^n \alpha - \sum_{i=1}^n \beta x_i & = 0, \\ \sum_{i=1}^n x_i y_i - \sum_{i=1}^n \alpha x_i - \sum_{i=1}^n \beta x_i^2 & = 0, \end{cases} \\ \Leftrightarrow & \begin{cases} \bar{y} - \alpha - \beta \bar{x} & = 0, \\ \bar{x}\bar{y} - \alpha \bar{x} - \beta \bar{x}^2 & = 0, \end{cases} \\ \Leftrightarrow & \begin{cases} \hat{\alpha} & = \bar{y} - \hat{\beta} \bar{x}. \\ \hat{\beta} & = \frac{\bar{x}\bar{y} - \bar{x}^2}{\bar{x}^2 - \bar{x}^2}. \end{cases} \end{aligned}$$

D'où l'équation du modèle est,

$$y = \hat{\alpha} + \hat{\beta}x.$$

B.1.2 La forme alternative pour moindre carrée

La pente β et d'autre point détermine aussi l'équation du l'ajustement, on note $A_{\bar{x}}$ l'intercepte à la ligne verticale en \bar{x} ,

$$A_{\bar{x}} = \hat{\alpha} + \hat{\beta}\bar{x} = \bar{y}$$

La droite des moindres carrées passe par le point (\bar{x}, \bar{y}) . Une équation alternative est donnée par,

$$\begin{aligned} y &= \alpha + \hat{\beta}x, \\ &= \hat{\alpha} + \hat{\beta}\bar{x} - \hat{\beta}x, \\ &= \hat{\alpha} + \hat{\beta}\bar{x} + \hat{\beta}(x - \bar{x}), \\ &= A_{\bar{x}} + B(x - \bar{x}). \end{aligned}$$

Avec : $B = \hat{\beta}$ et $A_{\bar{x}} = \hat{\alpha} + \hat{\beta}\bar{x}$.

B.1.3 Estimation de variance

L'estimation de la variance σ^2 est donnée par suit,

$$\begin{aligned} \hat{\sigma}^2 &= \frac{\sum_{i=1}^n (y_i - (A_{\bar{x}} + \beta(x_i - \bar{x})))^2}{n - 2}, \\ &= \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n - 2}. \end{aligned}$$

B.1.4 Fonction de vraisemblance de $\alpha_{\bar{x}}$ et β

Considérons l'équation du modèle suivante,

$$y_i = \alpha_{\bar{x}} + \beta(x_i - \bar{x}) + \epsilon_i, \quad i = 1, \dots, n.$$

Où,

$\alpha_{\bar{x}}$: la valeur moyenne pour $y(\bar{x} = x)$.

β : la pente.

ϵ_i : l'erreur $\forall i = 1, \dots, n$.

$\epsilon_i \sim \mathcal{N}(0, \sigma^2)$ et $y_i \sim \mathcal{N}(\alpha_{\bar{x}} + \beta(x_i - \bar{x}), \sigma^2)$.

La fonction de vraisemblance conjointe à la i ème observation est la fonction de densité de deux paramètres $\alpha_{\bar{x}}$ et β lorsque (x_i, y_i) , $i = 1, \dots, n$ fixé.

$$f_i(\alpha_{\bar{x}}, \beta) \propto \exp\left(-\frac{1}{2\sigma^2}(y_i - (\alpha_{\bar{x}} + \beta(x_i - \bar{x})))^2\right).$$

D'où,

$$\begin{aligned} f(\alpha_{\bar{x}}, \beta) &= \prod_{i=1}^n \exp\left(-\frac{1}{2\sigma^2}(y_i - (\alpha_{\bar{x}} + \beta(x_i - \bar{x})))^2\right), \\ &= \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - (\alpha_{\bar{x}} + \beta(x_i - \bar{x})))^2\right). \end{aligned}$$

On pose, $S = \sum_{i=1}^n (y_i - (\alpha_{\bar{x}} + \beta(x_i - \bar{x})))^2$.

$$\begin{aligned} S &= \sum_{i=1}^n [(y_i - \bar{y} + \bar{y} - (\alpha_{\bar{x}} + \beta(x_i - \bar{x})))^2], \\ &= \sum_{i=1}^n [(y_i - \bar{y})^2 + (y_i - \bar{y})(\bar{y} - (\alpha_{\bar{x}} + \beta(x_i - \bar{x}))) + (\bar{y} - (\alpha_{\bar{x}} + \beta(x_i - \bar{x})))^2], \\ &= \sum_{i=1}^n (y_i - \bar{y})^2 + 2 \sum_{i=1}^n (y_i - \bar{y})(\bar{y} - (\alpha_{\bar{x}} + \beta(x_i - \bar{x}))) + \sum_{i=1}^n (\alpha_{\bar{x}} - \bar{y})^2 - S_1. \end{aligned}$$

Avec,

$$S_1 = 2\beta \sum_{i=1}^n (\alpha_{\bar{x}} - \bar{y})(x_i - \bar{x}) + \beta^2 \sum_{i=1}^n (x_i - \bar{x})^2.$$

D'où,

$$S = \sum_{i=1}^n (y_i - \bar{y})^2 - 2\beta \sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x}) + \beta \sum_{i=1}^n (x_i - \bar{x})^2 + n(\alpha_{\bar{x}} - \bar{y})^2.$$

Car $(\bar{y} - \alpha_{\bar{x}} = 0)$.

Donc,

$$S = SS_y - 2\beta SS_{xy} + \beta^2 SS_x + n(\alpha_{\bar{x}} - \bar{y})^2.$$

D'où,

$$\begin{aligned} f(\alpha_{\bar{x}}, \beta) &= \exp\left(-\frac{1}{2\sigma^2}(SS_y - 2\beta SS_{xy} + \beta^2 SS_x + n(\alpha_{\bar{x}} - \bar{y})^2)\right), \\ &= \exp\left(-\frac{1}{2\sigma^2}(SS_y - 2\beta SS_{xy} + \beta^2 SS_x + n(\alpha_{\bar{x}} - \bar{y})^2)\right), \\ &= \exp\left(-\frac{1}{2\sigma^2}(SS_y - 2\beta SS_{xy} + \beta^2 SS_x)\right) \exp\left(-\frac{n}{2\sigma^2}(\alpha_{\bar{x}} - \bar{y})^2\right), \\ &= \exp\left(-\frac{1}{2\sigma^2/SS_x}\left(\beta - \frac{SS_{xy}}{SS_x}\right)^2\right) \exp\left(-\frac{n}{2\sigma^2}(\alpha_{\bar{x}} - \bar{y})^2\right). \end{aligned}$$

on note : $\frac{SS_{xy}}{SS_x} = T$.

$$f(\alpha_{\bar{x}}, \beta) \propto f(\alpha_{\bar{x}})f(\beta).$$

Où :

$$f(\alpha_{\bar{x}}) = \exp\left(-\frac{n}{2\sigma^2}(\alpha_{\bar{x}} - A_{\bar{x}})^2\right).$$

Et :

$$f(\beta) = \exp\left(-\frac{SS_x}{2\sigma^2(\beta - T)^2}\right).$$

Dans l'approche bayésienne on a,

$$\alpha_{\bar{x}} \sim \mathcal{N}\left(A_{\bar{x}}, \frac{\sigma^2}{n}\right).$$

$$\beta \sim \mathcal{N}\left(T, \frac{\sigma^2}{SS_x}\right).$$

B.1.5 La loi a priori conjointe

La loi a priori conjointe entre les paramètres $\alpha_{\bar{x}}$ et β c'est le produit de deux a priori,

$$\pi(\alpha_{\bar{x}}, \beta) = \pi(\alpha_{\bar{x}})\pi(\beta). \quad (\text{B.1})$$

Car $\alpha_{\bar{x}}$ et β sont indépendant.

B.1.6 La loi a posteriori conjointe

D'après le théorème de Bayes, la loi a posteriori est donnée par,

$$\pi(\alpha_{\bar{x}}, \beta | y) = \pi(\alpha_{\bar{x}}, \beta) f(\alpha_{\bar{x}}, \beta). \quad (\text{B.2})$$

On peut la définir aussi par,

$$\pi(\alpha_{\bar{x}}, \beta | y) = \pi(\alpha_{\bar{x}} | y) \pi(\beta | y). \quad (\text{B.3})$$

Si la loi a priori $\pi(\beta) \sim \mathcal{N}(\mu_\beta, \sigma_\beta^2)$ donc la loi a posteriori $\pi(\beta | y) \sim \mathcal{N}(\mu'_\beta, (\sigma'_\beta)^2)$.

D'après le tableau (1.1), on trouve,

$$\beta | y \sim \mathcal{N}\left(\frac{1}{\frac{\sigma^2}{SS_x} + \sigma_\beta^2} \left(\frac{\sigma^2}{SS_x} \mu_\beta + \sigma_\beta^2 B\right), \left(\frac{1}{\frac{\sigma^2}{SS_x} + \sigma_\beta^2}\right) \frac{\sigma^2}{SS_x} \sigma_\beta^2\right).$$

$$\Leftrightarrow \begin{cases} \mu'_\beta &= \frac{\frac{\sigma^2}{SS_x}}{\frac{\sigma^2}{SS_x} + \sigma_\beta^2} \mu_\beta + \frac{\sigma_\beta^2}{\frac{\sigma^2}{SS_x} + \sigma_\beta^2} B = \frac{1}{1 + \sigma_\beta^2 \frac{SS_x}{\sigma^2}} \mu_\beta + \frac{1}{\frac{\sigma^2}{\sigma_\beta^2 SS_x} + 1} B. \\ (\sigma'_\beta)^2 &= \frac{\frac{\sigma^2}{SS_x} \sigma_\beta^2}{\frac{\sigma^2}{SS_x} + \sigma_\beta^2} = \frac{\sigma_\beta^2}{1 + \sigma_\beta^2 \frac{SS_x}{\sigma^2}}. \end{cases} \quad (\text{B.4})$$

$$\Leftrightarrow \begin{cases} \mu'_\beta &= \frac{\frac{1}{\sigma_\beta^2}}{\frac{1}{\sigma_\beta^2} + \frac{SS_x}{\sigma^2}} \mu_\beta + \frac{\frac{SS_x}{\sigma^2}}{\frac{1}{\sigma_\beta^2} + \frac{SS_x}{\sigma^2}} B. \\ \frac{1}{(\sigma'_\beta)^2} &= \frac{1}{\sigma_\beta^2} + \frac{SS_x}{\sigma^2}. \end{cases}$$

Si la loi a priori $\pi(\alpha_{\bar{x}}) \sim \mathcal{N}(\mu_{\alpha_{\bar{x}}}, \sigma_{\alpha_{\bar{x}}}^2)$ donc la loi a posteriori $\pi(\alpha_{\bar{x}} | y) \sim \mathcal{N}(\mu'_{\alpha_{\bar{x}}}, (\sigma'_{\alpha_{\bar{x}}})^2)$.

$$\begin{aligned}
& \left\{ \begin{aligned} \mu'_{\alpha_{\bar{x}}} &= \frac{1}{\frac{\sigma^2}{n} + \alpha_{\bar{x}}^2} \left(\frac{\sigma^2}{n} \mu_{\alpha_{\bar{x}}} + \sigma_{\alpha_{\bar{x}}}^2 A_{\bar{x}} \right) = \frac{1}{\frac{\sigma^2}{n} + \sigma_{\beta}^2} \left(\frac{\sigma^2}{n} \mu_{\alpha_{\bar{x}}} + \sigma_{\alpha_{\bar{x}}}^2 A_{\bar{x}} \right). \\ (\sigma'_{\alpha_{\bar{x}}})^2 &= \frac{1}{\frac{\sigma^2}{n} + \alpha_{\bar{x}}^2} \frac{\sigma^2}{n} \sigma_{\alpha_{\bar{x}}}^2 = \frac{1}{\frac{\sigma^2}{n} + \sigma_{\beta}^2} \frac{\sigma^2}{n} \sigma_{\alpha_{\bar{x}}}^2. \end{aligned} \right. \\
\iff & \left\{ \begin{aligned} \mu'_{\alpha_{\bar{x}}} &= \frac{\frac{1}{\sigma_{\alpha_{\bar{x}}}^2}}{\frac{1}{(\sigma'_{\alpha_{\bar{x}}})^2}} \mu_{\alpha_{\bar{x}}} + \frac{\frac{n}{\sigma_{\alpha_{\bar{x}}}^2}}{\frac{1}{(\sigma'_{\alpha_{\bar{x}}})^2}} A_{\bar{x}}. \\ \frac{1}{(\sigma'_{\alpha_{\bar{x}}})^2} &= \frac{1}{\sigma_{\alpha_{\bar{x}}}^2 + \frac{n}{\sigma^2}}. \end{aligned} \right. \tag{B.5}
\end{aligned}$$

Bibliographie

- [1] Bayes.T, *Aessay towards solving a problem in the doctrine of chances*, Philosophical Transactions of the Royal Society of London, 53 : 370-418, (1763).
- [2] Bertrand. F *Master 1 MCB.TD N2. Régression Linéaire Multiple*, 2009/2010.
- [3] Chung. K.L, *Markov chains with stationary transition probabilities*. Springer Verlag, 1967.
- [4] Coupier. D, *Processus stochastiques*, Polytech Lille, 2002.
- [5] Dodge. Y et Melfi. G, *Premiers pas en simulations*, Springer, Paris, 2008.
- [6] Drosbeke. J, Fine. J et Saporta. G, *Méthodes bayésiennes en statistique*, Editions Technip, France, 2002.
- [7] Dupuis. J, *Statistique bayésienne et algorithme mcmc*, 2007.
- [8] Fromont.M, *ISAO- Méthodes MCMC*, ENSAI deuxième année, 2007/2008.
- [9] Geman. S and Geman. D, *Stochastic relaxation, gibbes distributions, and the bayesian restoration of images*, 6 : 721-741, 1984.
- [10] Gilks. W. R, Richardson. S, Spiegelhater. D. J, *Markov Chain Monte-Carlo in practice*, Springer Science+Business Media Dordrecht, 1996.
- [11] Guyader. A, *Simulation Monte-Carlo*, Université Pierre et Marie Curie, 2015/2016.
- [12] Hastings. W.K, *Monte Carlo Sampling methods using Markov chains and their applications*, volume 57, Biometrika, pages 97-109, April 1970.
- [13] Jeffreys. H, *Theory of probabiltiy (3rd edition)*, Oxford University Press, Oxford, 1961.
- [14] Judith. R, *Statistique Bayésienne*, ENSAE Paris, 2009.
- [15] Kass. R.E and Raftery. A.E, *Bays factors*, Jornal of the American Statistical Association, 90, 773-795, 1995.
- [16] Metropolis. N, Rosenbluth. A. W, Rosenbluth. M.N, Teller. A.H, and Teller. E, *Equation of state calculations by fast computings machines*, Chemical Physics, 21(6) :1087-1092, June 1953.
- [17] Millet. A, *Méthodes de Monte-Carlo*, Master 2 ème année : Spécialité Modélisation Aléatoire Laboratoire de Probabilités et Modèles Aléatoires, Universités Paris 7 et Paris 1, 2011.
- [18] Raiffa. H et Schlaifer. R, *Applied Statistical decision Theory*, Technical report, Division of Research, Graduate School of Business Administration, Harvard univ, (1961).
- [19] Robert. C. P *L'analyse Statistique Bayésienne*, Economica, 1992.
- [20] Robert. C. P, *Le choix bayésien principe et pratique*, 1^{ere} édition, Springer-Verlag, Paris, 2006.
- [21] Robert. C, *Méthodes de Monte Carlo par chaine de Markov*, Economia, 1996.
- [22] Ruegg. A, *Processus Stochastique*, Presse Polytechniques Romandes, 1989.

=====

Résumé

Ce travail propose une étude approfondie des méthodes de Monte-Carlo par chaîne de Markov (MCMC) dans le contexte bayésien. Les MCMC sont largement utilisées pour modéliser et résoudre des problèmes complexes dans divers domaines. L'étude présente les principes fondamentaux de la statistique bayésienne, tels que les distributions a priori, le théorème de Bayes et les distributions a posteriori. Elle explore également les techniques d'intégration Monte-Carlo et se concentre sur les méthodes MCMC, notamment l'algorithme de Metropolis-Hastings et l'échantillonneur de Gibbs.

Deux applications bayésiennes sont réalisées, mettant en oeuvre des cas et des scénarios différents. La première application se focalise sur la régression bayésienne dans le domaine médical, en étudiant spécifiquement les risques de crises cardiaques. La seconde application concerne l'estimation de la taille et du poids moyen d'une population de poissons dans un lac. Ces applications démontrent l'efficacité et la pertinence des méthodes MCMC pour l'analyse bayésienne. En conclusion, ce travail met en évidence l'importance des méthodes MCMC en tant qu'outil précieux pour la modélisation et la résolution de problèmes complexes dans divers domaines.

Mots clés : Théorème de Bayes , Inférence bayésienne, Monte-Carlo, Chaîne de Markov, MCMC, Loi a priori, Loi a posteriori, Simulation, Régression bayésienne, Estimation.

=====

Abstract

This work offers an in-depth study of Markov Chain Monte Carlo (MCMC) methods in the Bayesian context. MCMC methods are widely used for modeling and solving complex problems in various domains. The study presents the fundamental principles of Bayesian statistics, such as prior distributions, Bayes' theorem, and posterior distributions. It also explores Monte Carlo integration techniques and focuses on MCMC methods, including the Metropolis-Hastings algorithm and the Gibbs sampler.

Two Bayesian applications are conducted, implementing different cases and scenarios. The first application focuses on Bayesian regression in the medical field, specifically studying the risks of heart attacks. The second application concerns the estimation of size and mean weight of a fish population in a lake. These applications demonstrate the effectiveness and relevance of MCMC methods for Bayesian analysis. In conclusion, this work highlights the importance of MCMC methods as valuable tools for modeling and solving complex problems in various domains.

Keywords : Bayes theorem, Bayesian inference, Monte Carlo, Markov chain, MCMC, Prior distribution, Posterior distribution, Simulation, Bayesian regression, Estimation.

=====