



**HAL**  
open science

# Échantillonnage préférentiel adaptatif et méthodes bayésiennes approchées appliquées à la génétique des populations

Mohammed Sedki

► **To cite this version:**

Mohammed Sedki. Échantillonnage préférentiel adaptatif et méthodes bayésiennes approchées appliquées à la génétique des populations. Statistiques [math.ST]. Université Montpellier II - Sciences et Techniques du Languedoc, 2012. Français. NNT: . tel-00769095

**HAL Id: tel-00769095**

**<https://theses.hal.science/tel-00769095>**

Submitted on 28 Dec 2012

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

**THÈSE**

pour obtenir le titre de

**Docteur de l'Université Montpellier 2**

École Doctorale : Information, Structures et Systèmes

Formation Doctorale : Biostatistique

présentée par

**Mohammed Amechtoh Sedki**

**ÉCHANTILLONNAGE PRÉFÉRENTIEL ADAPTATIF ET MÉTHODES  
BAYÉSIENNES APPROCHÉES APPLIQUÉES À LA GÉNÉTIQUE DES  
POPULATIONS**

Thèse dirigée par **Jean-Michel Marin et Pierre Pudlo**

Soutenue le 31 octobre 2012 devant le jury composé de :

Arnaud DOUCET	Professeur à l'Université d'Oxford
Nicolas CHOPIN	Administrateur hors classe (ENSAE) CREST
Christophe ABRAHAM	Professeur à SupAgro Montpellier
Denis POMMERET	Professeur à l'Université d'Aix-Marseille
Jean-Michel MARIN	Professeur à l'Université Montpellier 2
Pierre PUDLO	Maître de conférences à l'Université Montpellier 2

Au vu des rapports de Arnaud DOUCET et Nicolas CHOPIN.



ÉCHANTILLONNAGE PRÉFÉRENTIEL ADAPTATIF ET MÉTHODES  
BAYÉSIENNES APPROCHÉES APPLIQUÉES À LA GÉNÉTIQUE DES  
POPULATIONS

Mohammed A. Sedki



## REMERCIEMENTS

Mes premiers remerciements s'adressent naturellement à mes directeurs de thèse, Jean-Michel Marin et Pierre Pudlo. Ils sont à l'origine de mon intérêt pour les sujets traités dans cette thèse, ils ont dirigés, encadrés, accompagnés de très près chaque étape de ces années de thèse. Je tiens par ces lignes à leur exprimer toute ma gratitude.

Je suis ensuite très reconnaissant envers mes rapporteurs de thèse : Arnaud Doucet et Nicolas Chopin qui ont accepté ce travail de lecture attentive malgré leurs emplois du temps plus que surchargés.

Ça a été un plaisir de rencontrer Denys Pommeret durant ma thèse et j'ai eu la chance de suivre le cours de M2 de Christophe Abraham. Je les remercie d'avoir accepté de faire partie de mon jury de thèse.

J'ai bénéficié durant ces trois années de l'accueil de l'équipe EPS dont je remercie tous les membres pour leur aide, gentillesse et disponibilité.

Le soutien financier du projet ANR EMILE m'a permis de multiples déplacements; j'en suis reconnaissant aux organisateurs, Jean-Marie Cornuet et Renaud Vitalis.

Je tiens également à exprimer mes remerciements à Mathieu Ribatet, qui a su m'écouter et me conseiller durant ma troisième année de thèse.

Je remercie aussi Raphael Leblois pour sa pédagogie, les nombreuses discussions sur la génétique des populations et ses nombreux conseils sur l'utilisation des logiciels MIGRAINE et IBDSIM.

Durant ma première année j'ai occupé l'ancien bureau de Romain et j'ai eu la chance d'avoir comme co-bureaux Hilde et Afaf qui m'ont données un surnom. Je les remercie pour la bonne humeur et les nombreux conseils dont j'ai pu bénéficier. Je remercie également Thomas et Julien avec qui j'ai partagé le bureau 122 en deuxième année. Je garderai un excellent souvenir de toutes les scéances d'imitations partagées avec Julien sans oublier le rire de Thomas et sa sagesse légendaire. Ma troisième année a vu l'arrivée de Björn avec son passé de biologiste, ses excellents jeux de mots et son "presque partout docteur Sedki" et Walid avec sa discrétion. Je les remercie pour les nombreuses discussions parfois politiques avec Walid, génétiques et existentielles avec Björn.

Je tiens également à remercier tous les autres doctorants que j'ai côtoyé depuis ma première année de thèse, des plus anciens aux nouveaux arrivants: Guillaume B., Anthony, Pierre, Mathieu S., Mathieu C., Claudia, Daria, Carine, Christophe, Jimmy, Jean, Lounes, Yousri, Arnaud, Ilias, Elsa, Benjamin, Tutu, Christian, Nahla, Angelina, Julien, Myriam. Je garderai un excellent souvenir des nombreuses discussions que j'ai pu avoir avec eux au sem-doc et en pause café. J'ajoute une pensée aux ATER et aux anciens thésards: Olivier, Guillaume D., Benoît, Jojo, Bruno et Vincent avec son rire légendaire et ses pieds sur la table de la salle de repos.

Je remercie les jeunes permanents de I3M qui ont pu nous intégrer via les nombreuses soirées, je pense à Boris, Ioan, Gwladys et Vanessa.

Enfin, je remercie ma famille pour leur affection et leur soutien, et tout particulièrement Arezki et Martine, grâce à qui j'ai réussi mon "atterissage".

# Table des matières

<b>I</b>	<b>Introduction et Motivation</b>	<b>11</b>
<b>1</b>	<b>Introduction</b>	<b>13</b>
1.1	Échantillonneur de Gibbs . . . . .	14
1.2	Échantillonnage préférentiel adaptatif et multiple . . . . .	17
1.3	Méthodes bayésiennes approchées et applications à la génétique des populations . . . . .	20
<b>2</b>	<b>Modèles de génétique des populations</b>	<b>25</b>
2.1	Introduction . . . . .	25
2.2	Données . . . . .	26
2.3	Généalogie d'échantillons . . . . .	27
2.3.1	Une seule population : Coalescent de Kingman . . . . .	28
2.3.2	Plusieurs populations structurées . . . . .	29
2.4	Processus mutationnels . . . . .	32
2.5	Conclusion . . . . .	34
<b>II</b>	<b>Échantillonnage Préférentiel Adaptatif et Multiple</b>	<b>41</b>
<b>3</b>	<b>Échantillonnage Préférentiel</b>	<b>43</b>
3.1	Généralités sur l'échantillonnage préférentiel . . . . .	43
3.2	Mesures d'efficacité d'un système de particules pondérées . . . . .	45
3.3	Quelques variantes de l'échantillonnage préférentiel . . . . .	46
3.3.1	Échantillonnage préférentiel auto-normalisé . . . . .	46



3.3.2	Échantillonnage préférentiel avec variables de contrôle . . . . .	48
<b>4</b>	<b>Théorèmes limites sur les tableaux triangulaires</b>	<b>49</b>
4.1	Loi des grands nombres . . . . .	49
4.2	Théorème central limite . . . . .	50
<b>5</b>	<b>Échantillonnage Préférentiel Adaptatif et Multiple</b>	<b>53</b>
5.1	Introduction et motivation . . . . .	53
5.2	État de l'art des méthodes d'échantillonnage préférentiel multiple . . . . .	56
5.2.1	Stratégies de combinaison d'échantillons . . . . .	57
5.2.2	Optimalité de l'heuristique de la balance . . . . .	60
5.2.3	Échantillonnage préférentiel multiple avec variables de contrôle . . . . .	62
5.3	Algorithme AMIS d'échantillonnage préférentiel adaptatif et multiple . . . . .	65
5.3.1	Schéma adaptatif de l'algorithme AMIS . . . . .	65
5.3.2	Difficultés liées à l'étude de la convergence de l'estimateur AMIS . . . . .	68
5.4	Convergence d'un estimateur AMIS modifié . . . . .	68
5.4.1	Nouvel algorithme avec schéma adaptatif simplifié . . . . .	68
5.4.2	Hypothèses et résultats . . . . .	70
5.4.3	Vérification des différentes hypothèses sur un exemple jouet . . . . .	72
5.4.4	Démonstration des résultats sur $\hat{\theta}_t$ . . . . .	77
5.4.5	Démonstration de la consistance de l'algorithme (théorème 5.6) . . . . .	80
5.5	Applications numériques . . . . .	85
5.5.1	Simulations et comparaisons sur une cible jouet . . . . .	85
5.5.2	Simulations et comparaisons sur un modèle de génétique des populations . . . . .	87

<b>III</b>	<b>Méthodes bayésiennes approchées</b>	<b>91</b>
<b>6</b>	<b>Méthodes bayésiennes approchées</b>	<b>93</b>
6.1	Introduction . . . . .	93
6.2	Schémas sans vraisemblance séquentiels . . . . .	95
6.3	Notre proposition auto-calibrée . . . . .	98
6.4	Expérimentations numériques . . . . .	102
6.4.1	Exemple jouet . . . . .	102
6.4.2	Exemple de génétique des populations . . . . .	103
6.5	Comparaison de stratégies de parallélisation dans ABC . . . . .	106
<b>7</b>	<b>Discussion et perspectives</b>	<b>115</b>
7.1	Questions autour de l'algorithme AMIS . . . . .	115
7.2	Quelques perspectives autour des méthodes ABC . . . . .	115



# **Première partie**

## **Introduction et Motivation**



# 1 Introduction

Mener une inférence statistique lorsque la vraisemblance du modèle est inaccessible ou très coûteuse en temps de calcul suscite divers problèmes numériques. En effet, dans de nombreuses situations, la vraisemblance est sous forme intégrale

$$\ell(\mathbf{y}|\boldsymbol{\theta}) = \int f(\mathbf{y}, \mathbf{u}|\boldsymbol{\theta}) \, d\mathbf{u}, \quad (1.1)$$

où la fonction  $f(\cdot, \cdot | \cdot) \geq 0$  représente la vraisemblance complète des données dont l'évaluation est accessible pour tout triplet  $(\mathbf{y}, \mathbf{u}, \boldsymbol{\theta})$ . Le vecteur  $\mathbf{y} \in \mathbb{R}^n$  représente le jeu de données observé ou la partie des données dont on connaît la valeur, tandis que  $\mathbf{u}$  représente la partie non observée du jeu de données appelée aussi donnée manquante, variable cachée ou variable latente dans la littérature. On adoptera la dernière appellation dans toute la suite. L'introduction des variables latentes constitue une méthodologie très générale pour modéliser des problèmes avec des structures de données très complexes. Le problème d'inférence statistique dans ces modèles est à l'origine de nombreuses techniques de probabilités numériques et d'optimisation. On distingue les algorithmes de recherche du maximum de vraisemblance dans le cadre de l'approche fréquentielle et les schémas MCMC pour l'approximation de la loi *a posteriori* pour l'approche bayésienne. Dans cette thèse, on considère les cas où l'espace de la variable latente  $\mathbf{u}$  est de très grande dimension et comporte des directions de différentes natures (discrètes et continues), ce qui rend l'intégrale (1.1) incalculable. On s'intéresse à des solutions bayésiennes au problème d'inférence dans ce cas.

Rappelons que pour l'approche fréquentielle, la question de la maximisation de la vraisemblance (1.1) en  $\boldsymbol{\theta}$  a été au cœur de nombreux développements. Parmi les plus utilisés, citons l'algorithme EM de Dempster et al. [33] et ses différentes variantes stochastiques développées pour remplacer les calculs analytiques à l'étape E, impossibles dans de nombreuses situations (e.g. SEM de Celeux and Diebolt [18], SAEM de Delyon et al. [32] et MCEM de Wei and Tanner [103]). L'ensemble de ces techniques constitue une boîte à outils indispensable pour l'approche fréquentielle.

Le champs d'application privilégié de cette thèse est l'inférence dans les modèles de génétique des populations auxquels nous consacrons le chapitre 2. Un modèle de génétique des populations est contraint par un scénario démographique qui décrit l'histoire

évolutive de l'espèce d'intérêt par une suite d'événements. Le paramètre d'un modèle de génétique des populations est composé des différentes dates d'événements démographiques, paramètres de remise en contact de populations, et des tailles de populations. Les généticiens des populations s'intéressent à l'inférence de ces quantités. En effet, obtenir des informations sur les paramètres qui agissent sur la diversité génétique permet d'améliorer et de conserver les espèces. Pour mener leurs études, les généticiens des populations se basent sur l'information génétique extraite des populations du présent. Cette information représente la variable observée  $y$ . L'information incluant l'histoire spatiale et temporelle de l'espèce considérée est inaccessible en général et représente la variable latente  $u$  du modèle.

Résumons le contenu de ce chapitre introductif. Nous commençons par rappeler l'échantillonneur de Gibbs [68, 48] qui constitue l'outil standard pour l'approximation des lois *a posteriori* dans les modèles auxquels on s'intéresse. Nous décrivons brièvement les travaux basés sur l'échantillonneur de Gibbs et consacrés aux modèles de génétique des populations. Nous évoquons quelques difficultés liées à sa mise en oeuvre sur ce type de modèles. Les deux autres parties de ce chapitre sont consacrées à l'introduction de nos contributions aux problèmes d'inférence bayésienne. Dans un premier temps nous supposons que la vraisemblance peut être approchée à l'aide d'une boîte à outils de calcul avec un certain coût numérique. Nous montrons la convergence du schéma d'échantillonnage préférentiel introduit par Cornuet et al. [21] visant à minimiser le nombre d'appels au calcul de la vraisemblance. Puis, nous nous intéressons à des modèles dont la structure latente est hautement complexe et pour lesquels l'approximation de la vraisemblance est impossible. En supposant que la simulation suivant le modèle est accessible, nous avons introduit un algorithme ABC séquentiel [6, 75] qui permet d'approcher la loi *a posteriori*.

## 1.1 Échantillonneur de Gibbs

Habituellement, l'approximation de la loi *a posteriori*

$$\pi(\theta|y) = \frac{\pi(\theta)\ell(y|\theta)}{\int \ell(y|\theta')\pi(\theta') d\theta'}, \quad (\text{où } \pi(\theta) \text{ est la loi } a \text{ priori}) \quad (1.2)$$

fait appel aux méthodes MCMC [89, 77] munies de différents mécanismes d'exploration de l'espace des paramètres.

Contrairement aux techniques Hastings-Metropolis qui nécessitent une évaluation de la vraisemblance à chaque itération, l'algorithme dit de "données augmentées" (D-A pour Data Augmentation en anglais) de Tanner and Wong [99] alterne des simulations de  $\theta^{(t)}$  et de  $u^{(t)}$  respectivement suivant les lois  $f(\theta|u, y)$  et  $f(u|\theta, y)$  un grand nombre de fois. Ainsi, la suite générée  $\theta^{(1)}, \dots, \theta^{(t)}, \dots$  forme une trajectoire d'une chaîne de Markov de loi invariante, la loi *a posteriori*  $\pi(\theta|y)$ . L'idée de Tanner and Wong [99] est basée sur le fait qu'on peut écrire la loi *a posteriori* comme une intégrale contre la loi prédictive de la variable latente  $u$  sachant les données  $y$ , *i.e.*

$$\pi(\theta|y) = \int f(\theta|u, y) f(u|y) du,$$

et la densité prédictive  $f(\mathbf{u}|\mathbf{y})$  à son tour comme une intégrale contre la loi *a posteriori*, i.e.

$$f(\mathbf{u}|\mathbf{y}) = \int f(\mathbf{u}|\boldsymbol{\theta}, \mathbf{y})\pi(\boldsymbol{\theta}|\mathbf{y}) d\boldsymbol{\theta}.$$

Les mouvements de la chaîne générée par l'algorithme D-A de [99] sont donnés par le noyau

$$\mathcal{K}(\boldsymbol{\theta}, \boldsymbol{\theta}') = \int f(\boldsymbol{\theta}|\mathbf{u}, \mathbf{y})f(\mathbf{u}|\boldsymbol{\theta}', \mathbf{y}) d\mathbf{u}.$$

Cette technique est un cas particulier de l'échantillonneur de Gibbs [46, 45, 66, 70, 68, 12] avec deux composantes ou deux étapes.

Notons  $\mathbf{z} = (\mathbf{z}_1, \dots, \mathbf{z}_d)$  le couple  $(\boldsymbol{\theta}, \mathbf{u})$ . Un pas de la chaîne explorante de l'échantillonneur de Gibbs est donné par le noyau de transition

$$\mathcal{K}(\mathbf{z}, \mathbf{z}'_i) = f(\mathbf{z}'_i|\mathbf{z}_{-i}),$$

où  $\mathbf{z}'_i$  est la  $i^{\text{ème}}$  composante du vecteur  $\mathbf{z}'$  et  $\mathbf{z}_{-i}$  correspond au vecteur  $\mathbf{z}$  privé de sa  $i^{\text{ème}}$  composante. Résumons dans l'algorithme 1, une itération de l'échantillonneur de Gibbs. La mise à jour décrite dans l'algorithme 1 correspond à la version appelée *systematic scan*

---

### Algorithme 1 Échantillonneur de Gibbs

---

**ENTRÉES :** Un point d'initialisation  $\mathbf{z}^{(1)}$ .

Mettre à jour l'échantillonneur de  $\mathbf{z}^{(j)}$  à  $\mathbf{z}^{(j+1)}$ :

**Pour**  $i = 1 \rightarrow d$  **faire**

– Simuler  $\mathbf{z}_i^{(j+1)}$  à partir de  $f(\mathbf{z}_i|\mathbf{z}_1^{(j+1)}, \dots, \mathbf{z}_{i-1}^{(j+1)}, \mathbf{z}_{i+1}^{(j)}, \dots, \mathbf{z}_d^{(j)})$

**Fin pour**

---

*Gibbs Sampler*. On peut envisager de choisir aléatoirement la composante à mettre à jour (*random scan Gibbs Sampler*).

On retrouve la vraisemblance sous la forme (1.1) dans de nombreux modèles statistiques couramment utilisés. Ces modèles impliquent naturellement des structures latentes [77, 48]. Soulignons que le logiciel *WINBUGS* de Lunn et al. [72] décrit dans Ntzoufras [82] fournit diverses implémentations de l'échantillonneur de Gibbs pour de nombreux problèmes d'inférence bayésienne et plus généralement pour les modèles hiérarchiques.

En dépit de la simplicité apparente dans l'implémentation de l'algorithme 1, de nombreux modèles impliquant des variables latentes avec des composantes à fortes corrélations montrent que la qualité d'approximation est peu satisfaisante. Les chaînes explorent très mal l'espace de la variable latente et les temps de mélange deviennent très élevés. Un grand nombre d'améliorations ont été apportées à l'échantillonneur de Gibbs [66, 71]. Des méthodes MCMC hybrides ont été développées ainsi que des extensions sous forme de stratégies de décomposition de la variable latente en différents blocs. En génétique des populations, la simulation suivant les lois conditionnelles est impossible dès que le modèle atteint une certaine complexité. Par conséquent, l'implémentation de l'échantillonneur de Gibbs est inconcevable dans ce genre de situations. Pour cela, des schémas MCMC hybrides (Hastings-Metropolis et échantillonneur de Gibbs) sont proposés, c'est-à-dire



qu'une itération de Hastings-Metropolis remplace la simulation suivant l'une ou plusieurs de ces lois conditionnelles [91]. Un grand nombre de techniques d'inférence bayésienne sont destinées aux modèles de génétique des populations. Principalement, ces techniques sont basées sur l'échantillonneur de Gibbs. Ainsi, on trouve l'implémentation dans le logiciel *IM* (Isolation with Migration) des travaux de Hey and Nielsen [53, 54] et Nielsen and Wakeley [81], où ils construisent une chaîne de Markov de loi stationnaire  $f(\mathbf{u}, \theta | \mathbf{y})$ . Les simulations suivant les lois conditionnelles sur ces modèles restent très complexes et sont décrites dans Nielsen and Wakeley [81]. Comme alternative, Kuhner et al. [62] remplacent les lois conditionnelles par des schémas Hastings-Metropolis. Les travaux de Kuhner et al. [62] ont été développés dans un contexte d'estimation par maximum de vraisemblance et sont largement repris sur différents modèles [63, 9, 43]. La prise en compte de modèles de plus en plus complexes rend l'exploration de l'espace des valeurs prises par le paramètre et la variable latente de plus en plus difficile. La forte corrélation entre les composantes de la variable latente implique des pertes en efficacité. Le mélange de la chaîne explorante est de très mauvaise qualité et les temps de calculs deviennent extrêmement lents. En dépit de nombreuses tentatives d'améliorations (voir *e.g.* pour le logiciel *IM*: lancement de plusieurs chaînes en parallèle couplé avec des techniques de "tempering" de Geyer and Thompson [47]), la qualité d'approximation obtenue par ces techniques reste insatisfaisante et les temps de calcul demeurent lents [64]. Ce constat se retrouve dans les récents travaux théoriques de Woodard and Rosenthal [105] sur la vitesse de convergence de l'échantillonneur de Gibbs dans le contexte de détection de motifs dans les séquences d'ADN.

Étant donné les difficultés énumérées précédemment, notre première contribution suppose que la vraisemblance peut être évaluée via une approximation numériquement coûteuse. Le schéma d'échantillonnage préférentiel adaptatif et multiple (AMIS pour Adaptive Multiple Importance Sampling) de Cornuet et al. [21] nécessite peu d'appels au calcul de la vraisemblance et recycle ces évaluations. L'algorithme AMIS approche la loi *a posteriori* de densité

$$\pi(\boldsymbol{\theta}|\mathbf{y}) \propto \pi(\boldsymbol{\theta})\ell(\mathbf{y}|\boldsymbol{\theta}), \quad (1.3)$$

par un système de particules pondérées. Cette technique est conçue pour pouvoir recycler les simulations obtenues par le processus itératif (la construction séquentielle d'une suite de lois d'importance). Dans les nombreux tests numériques effectués sur des modèles de génétique des populations [21, 94], l'algorithme AMIS a montré des performances numériques très prometteuses en terme de stabilité. Ces propriétés numériques sont particulièrement adéquates pour notre contexte. Toutefois, dans Cornuet et al. [21], la question de la convergence des estimateurs obtenus par cette technique reste largement ouverte. Dans cette thèse, nous montrons des résultats de convergence d'une version légèrement modifiée de l'algorithme AMIS. Sur des simulations, nous montrons que ses qualités numériques sont identiques à celles du schéma original.

## 1.2 Échantillonnage préférentiel adaptatif et multiple

Dans cette partie, nous évaluons la vraisemblance à l'aide d'une technique d'estimation sans biais. Nous considérons cet estimateur comme une boîte à outils à laquelle nous faisons appel pour calculer la vraisemblance  $\ell(\mathbf{y}|\boldsymbol{\theta})$  de notre jeu de données  $\mathbf{y}$  pour différentes valeurs du paramètre  $\boldsymbol{\theta}$ . Citons une technique d'estimation de la vraisemblance basée sur l'échantillonnage préférentiel et destinée aux modèles de génétique des populations détaillés dans le chapitre 2. Pour ces modèles, il existe une formule de récurrence [41, 49] qui donne un lien entre la vraisemblance  $\ell(\mathbf{y}|\boldsymbol{\theta})$  et les vraisemblances  $\ell(\mathbf{y}'|\boldsymbol{\theta})$ , où  $\mathbf{y}'$  est un jeu de données de taille inférieure ou égale à celle de  $\mathbf{y}$ . Stephens and Donnelly [97], De Iorio and Griffiths [25, 26] construisent des lois d'importance qui sont solutions d'un système de récurrence simplifié. Ainsi, ils approchent la vraisemblance par échantillonnage préférentiel. Leur approche consiste à simuler un grand nombre de variables  $\mathbf{u}_1, \dots, \mathbf{u}_N$  suivant une loi d'importance de densité  $g$ . L'approximation de la vraisemblance par cette technique est donnée par l'estimateur

$$\widehat{\ell(\mathbf{y}|\boldsymbol{\theta})} = \frac{1}{N} \sum_{i=1}^N \frac{f(\mathbf{y}, \mathbf{u}_i|\boldsymbol{\theta})}{g(\mathbf{u}_i)}, \quad (1.4)$$

où  $\mathbf{y}$  est le jeu de données observé et  $\boldsymbol{\theta}$  un paramètre fixé. Notons que sur ces modèles, Beaumont [5] a proposé un algorithme Hastings-Metropolis où le calcul de la probabilité d'acceptation provient d'une estimation sans biais de la vraisemblance. Les propriétés théoriques de cet algorithme sont étudiées dans Andrieu and Roberts [2]. On retrouve ce type d'idées dans le schéma PMCMC (Particle Monte-Carlo Markov Chain en anglais) de Andrieu et al. [1]. Ici, on ne s'intéressera pas aux algorithmes Hastings-Metropolis mais seulement aux schémas d'échantillonnage préférentiel adaptatif et multiple.

Le point de départ dans cette partie est le suivant: en remarquant que l'expression la densité *a posteriori* donnée en (1.2) peut être approchée par échantillonnage préférentiel en considérant le produit  $\ell(\mathbf{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})$  comme cible. L'idée est de construire un système de particules pondérées  $\{\omega_e, \mathbf{x}_e\}_{e \in E}$  indexé par l'ensemble  $E$  et d'approcher la loi *a posteriori*  $\Pi$  de la manière suivante

$$\Pi(\widehat{\cdot}|\mathbf{y}) = \sum_{e \in E} \omega_e \delta_{\mathbf{x}_e}(\cdot)$$

telle qu'on a la convergence

$$\sum_{e \in E} \omega_e \psi(\mathbf{x}_e) \longrightarrow \int \psi(\boldsymbol{\theta}) \pi(\boldsymbol{\theta}|\mathbf{y}) d\boldsymbol{\theta}$$

lorsque le nombre de particules tend vers l'infini et pour une large classe de fonctions  $\psi$ . Le calcul du poids  $\omega_e$  nécessite l'évaluation de la vraisemblance pour  $\boldsymbol{\theta} = \mathbf{x}_e$ . L'approximation donnée par la formule (1.4) peut se révéler lente dès qu'on vise une bonne précision. Afin de minimiser le nombre d'appels à la boîte à outils qui évalue la vraisemblance, Cornuet et al. [21] ont introduit l'algorithme d'échantillonnage préférentiel adaptatif et multiple AMIS (Adaptive Multiple Importance Sampling) qui a montré un gain considérable en temps consacré à l'approximation de la vraisemblance grâce à son processus de recyclage des particules.

Il existe de nombreuses stratégies différentes pour implémenter un algorithme d'échantillonnage préférentiel adaptatif [69, 85, 89, 93]. Par exemple, le schéma PMC de Cappé et al. [13], généralisé dans le paradigme D-kernel dans Douc et al. [35], Cappé et al. [14], Douc et al. [36] vise à construire un mélange de lois minimisant un critère d'optimalité (Kullback-Leibler ou minimisation de la variance) par rapport à une loi cible. Ces techniques d'échantillonnage préférentiel adaptatif sont étroitement liées aux techniques particulières de Doucet et al. [39] et aux méthodes SMC (Sequential Monte Carlo) de Del Moral et al. [29]. L'algorithme AMIS de Cornuet et al. [21] reprend l'approche de Cappé et al. [14] pour construire un système de particules  $\mathbf{x}_1^1, \dots, \mathbf{x}_{N_1}^1, \dots, \mathbf{x}_1^T, \dots, \mathbf{x}_{N_T}^T$  issues d'une suite de lois de propositions (ou instrumentales)  $Q_1, \dots, Q_T$  de densités  $q_1, \dots, q_T$ . L'originalité de cette technique est dans la construction de la loi de proposition  $Q_{t+1}$  à chaque itération  $t$  qui repose sur le recyclage de tous les  $t$  échantillons précédents cette itération. Pour effectuer ce recyclage, Cornuet et al. [21] associe un système de poids d'importance approprié. Ainsi pour  $i = 1, \dots, N_k, k = 1, \dots, t$ , nous avons le poids d'importance associé à la particule  $\mathbf{x}_i^k$  donné par

$$\omega_i^k = \pi(\mathbf{x}_i^k|\mathbf{y}) \Big/ \frac{1}{\sum_{\ell=1}^t N_\ell} \sum_{j=1}^t N_j q_j(\mathbf{x}_i^k).$$

Cette technique de pondération est inspirée des stratégies de mélange d'échantillons appelées échantillonnage préférentiel multiple, introduites dans Veach and Guibas [101] et reprises dans Owen and Zhou [83]. Ainsi, au lieu d'échantillonner encore sur la dernière loi de proposition  $Q_T$  obtenue à la fin du processus itératif, AMIS renvoie le système final de particules recyclées avec un poids d'importance final donné par

$$\pi(\mathbf{x}_i^t|\mathbf{y}) \Big/ \frac{1}{\sum_{\ell=1}^T N_\ell} \sum_{j=1}^T N_j q_j(\mathbf{x}_i^t) \quad \text{pour tout } (1 \leq i \leq N_t, 1 \leq t \leq T). \quad (1.5)$$

Soulignons que le calcul du numérateur dans l'expression précédente n'est effectué qu'une seule fois et que sa réutilisation nécessite seulement un stockage dans la mémoire. Sur les différents tests numériques réalisés dans Cornuet et al. [21] sur un modèle réel de génétique des populations, l'algorithme AMIS apporte un gain considérable en réduisant le nombre d'évaluations de la vraisemblance grâce à la stratégie de recyclage. On retrouve aussi dans Sirén et al. [94], une application de l'algorithme AMIS sur un problème d'inférence bayésienne (en génétique des populations) sur un modèle impliquant des données SNP (Single Nucleotide Polymorphism), et dans laquelle, ils insistent sur l'efficacité de l'algorithme AMIS pour ce genre de problèmes.

En dépit de ces bonnes propriétés numériques, la question de la consistance de l'algorithme AMIS de Cornuet et al. [21] ou de la convergence des estimateurs obtenus par le schéma AMIS est restée largement ouverte. Les systèmes de poids dus aux stratégies de recyclage des échantillons, rendent la structure de dépendance dans le système de particules très compliquée. L'application des théorèmes limites classiques sur des suites de variables aléatoires dépendantes ou indépendantes est impossible.

Sur cette partie, l'apport principal de cette thèse est la démonstration de la convergence d'un schéma AMIS avec une légère simplification dans le processus itératif. Plus précisément, en choisissant la suite  $N_1, \dots, N_T$  croissante, nous avons restreint la construction de la nouvelle loi de proposition  $Q_{t+1}$  à chaque itération  $t$  à l'échantillon  $x_1^t, \dots, x_{N_t}^t$  avec un système de poids basique. Ainsi, à la fin du processus itératif, par la même stratégie employée dans la version originale de AMIS, nous recyclons l'ensemble des  $T$  échantillons avec le système de poids (1.5). Sur ce nouveau schéma AMIS et à partir des résultats asymptotiques sur les tableaux triangulaires [19, 15, 37], nous avons obtenu une loi des grands nombres sur l'estimateur  $\sum_{t=1}^T \sum_{i=1}^{N_t} \omega_i^t \psi(x_i^t)$  pour une certaine classe de fonctions  $\psi$ . À partir des tests numériques effectués sur un exemple jouet et un modèle de génétique des populations, les différentes comparaisons ont montré que nos simplifications dans le schéma AMIS n'impactent pas les qualités numériques de la version originale.

La seconde contribution de cette thèse est aussi consacrée à l'inférence bayésienne lorsque la vraisemblance est de la forme (1.1). On renonce à l'estimation de la vraisemblance et on supposera seulement que la simulation suivant le modèle (suivant la vraisemblance) est possible. Notre apport est un algorithme ABC séquentiel (Approximate Bayesian Computation [6, 75]). On contourne le calcul de la vraisemblance par une approximation ABC de la loi *a posteriori* qui nécessite un grand nombre de simulations suivant le modèle. Sur les modèles de la génétique des populations que nous considérons dans le chapitre 2, cette méthode peut se révéler lente lorsqu'on vise une approximation précise de la loi *a posteriori*. L'algorithme que nous proposons est une amélioration de l'algorithme ABC-SMC de Del Moral et al. [31] que nous optimisons en nombre d'appels aux simulations suivant la vraisemblance, et que nous munissons d'un mécanisme de choix de niveaux d'acceptations auto-calibré. Nous implémentons notre algorithme pour inférer les paramètres d'un scénario évolutif réel et complexe de génétique des populations. Nous montrons que pour la même qualité d'approximation, notre algorithme nécessite deux fois moins de simulations par rapport à la méthode ABC avec acceptation couramment utilisée.

### 1.3 Méthodes bayésiennes approchées et applications à la génétique des populations

La prise en compte de modèles avec un grand nombre de paramètres augmente la complexité de l'espace de la variable latente. Lorsque un modèle de génétique des populations (voir chapitre 2) est composé d'un grand nombre d'événements démographiques, l'approximation de la vraisemblance est impossible. La méthode ABC est l'une des techniques les plus utilisées au cours de la dernière décennie pour contourner le problème du calcul de la vraisemblance. Dans cette thèse, nous avons proposé sous forme d'algorithme séquentiel une accélération de la version basique de ABC basée sur l'acceptation-rejet. Notre algorithme est muni d'un schéma d'apprentissage pour le calibrer du niveau d'acceptation. Il constitue une amélioration auto-calibrée du schéma ABC-SMC proposé dans Del Moral et al. [31]. Quant aux techniques de post-traitement destinées à améliorer la précision de l'approximation ABC, nous nous sommes intéressés à la sélection des statistiques résumées pour le choix des modèles et l'estimation des paramètres.

Les méthodes bayésiennes approchées (Approximate Bayesian Computation ABC) sont intensivement utilisées pour approcher une loi *a posteriori*  $\pi(\theta|\mathbf{y})$  lorsque l'évaluation de la vraisemblance  $\ell(\mathbf{y}|\theta)$  est impossible. L'algorithme ABC d'acceptation-rejet simule un grand nombre de couples iid  $(\theta_i, \mathbf{x}_i)_{1 \leq i \leq N}$ . Nous décrivons dans l'algorithme 2 la version acceptation-rejet de ABC proposée par Beaumont et al. [4].

La loi *a posteriori* est approchée par

$$\pi_\varepsilon(\theta|\mathbf{y}) \propto \int \pi(\theta) \ell(\mathbf{x}|\theta) \mathbf{1}\{d\{\eta(\mathbf{x}), \eta(\mathbf{y})\} < \varepsilon\} dx,$$

où  $\varepsilon$  représente un niveau de précision (couramment appelé niveau d'acceptation) et  $\eta$  est une application qui désigne l'ensemble des statistiques résumées du modèle. Pour que

**Algorithme 2** Algorithme ABC acceptance-rejet de Beaumont et al. [4]

**ENTRÉES :** Le jeu de données observé, la loi *a priori*  $\pi$ , une application  $\eta$  définissant l'ensemble des statistiques résumées, une distance  $d$  sur l'image de l'application  $\eta$  et un niveau d'acceptation  $\varepsilon$ .

$i \leftarrow 1$

**Tantque**  $i \leq N$  **faire**

1: Simuler  $\theta_i$  suivant la loi *a priori*  $\pi(\cdot)$ .

2: Simuler  $x_i$  suivant la vraisemblance  $\ell(\cdot | \theta_i)$

3: **si**  $d\{\eta(x_i), \eta(y)\} < \varepsilon$  **alors** accepter le couple  $(\theta_i, x_i)$  et  $i \leftarrow i + 1$ .

**Fin tantque**

cette approximation soit précise, il est indispensable que le niveau d'acceptation  $\varepsilon$  soit significativement bas. Ce qui implique des rejets récurrents dans la procédure et donc une augmentation du temps de simulation des couples  $(x_i, \theta_i)$ .

Comme amélioration de la technique d'acceptation-rejet, les algorithmes séquentiels [31, 7, 3] proposent de décroître progressivement le niveau d'acceptation  $\varepsilon$ . L'algorithme Monte-Carlo séquentiel (SMC) proposé par Del Moral et al. [31] alterne une étape d'acceptation-rejet en assurant la décroissance du niveau d'acceptation  $\varepsilon$ , avec une étape Metropolis-Hastings avec  $\pi_\varepsilon$  invariante. Cependant le schéma adaptatif proposé dans cet article n'est pas optimisé pour le temps de calcul lorsque celui-ci est essentiellement dû aux simulations des couples  $(x_i, \theta_i)$ .

Nous avons proposé un nouveau schéma auto-calibré dont le coût de calcul est significativement inférieur à celui de l'algorithme d'acceptation-rejet, et cela lorsque le temps de calcul est entièrement consacré à la simulations des jeux de données, et au calcul des distances  $d\{\eta(x), \eta(y)\} < \varepsilon$ .

Ce nouvel algorithme a été testé sur un scénario évolutif complexe de l'histoire des abeilles en Europe. Sur ce problème, les différentes composantes du paramètre  $\theta$  représentent les dates de divergences et d'admixture, les tailles de populations efficaces et les taux de mutations. Le jeu de données est composé de 4 échantillons de 4 région d'Europe. Ces données représentent les génotypes d'une vingtaine d'individus par population. En parallélisant le calcul sur 8 cœurs, notre algorithme produit une approximation satisfaisante de la loi *a posteriori* avec un temps de calcul de 2 heures alors que l'algorithme d'acceptation-rejet nécessite 4 heures pour le même niveau d'approximation.

Par ailleurs, nous avons étudié dans [76] différentes stratégies de parallélisation de l'algorithme ABC séquentiel. Cette étude a été menée en comparaison avec l'algorithme ABC acceptance-rejet standard. Nous avons conclu sur les différentes comparaisons que les techniques ABC séquentielles ne sont pas recommandées sur des machines avec un très grand nombre de cœurs. Le temps de gestion des différents threads rend les techniques séquentielles inefficaces. Ainsi, nous avons déduit qu'il est préférable d'implémenter l'algorithme ABC acceptance-rejet standard avec parallélisation triviale sur ce genre de machines.

La précision des techniques ABC est étroitement liée au choix des statistiques résu-

mées. Les procédures de post-traitement en ABC sont destinées à améliorer la précision dans l'approximation de la loi *a posteriori* [10]. Dans les modèles de génétique des populations introduits dans le chapitre 2 et particulièrement ceux qui sont traités par le logiciel *DIYABC* de Cornuet et al. [22], le nombre de statistiques résumées est de l'ordre de la centaine. Beaumont et al. [4] proposent le post-traitement par la régression locale pour améliorer le niveau d'approximation de ABC. Cette technique vise à rapprocher les jeux de données simulés à l'observé. Nous avons proposé dans [87] d'inclure une étape de sélection de variables dans la régression locale de Beaumont et al. [4]. Notre idée consiste à appliquer le critère BIC pour sélectionner les statistiques résumées pertinentes. Les résultats numériques obtenus dans [87] montrent que notre proposition est prometteuse. Nous présentons ici les résultats obtenus. Nous nous sommes intéressés à deux exemples: un exemple jouet d'estimation d'un paramètre dans une famille gaussienne et un problème de choix de modèles sur le premier exemple de génétique des populations (voir chapitre 2) étudié dans Robert et al. [90].

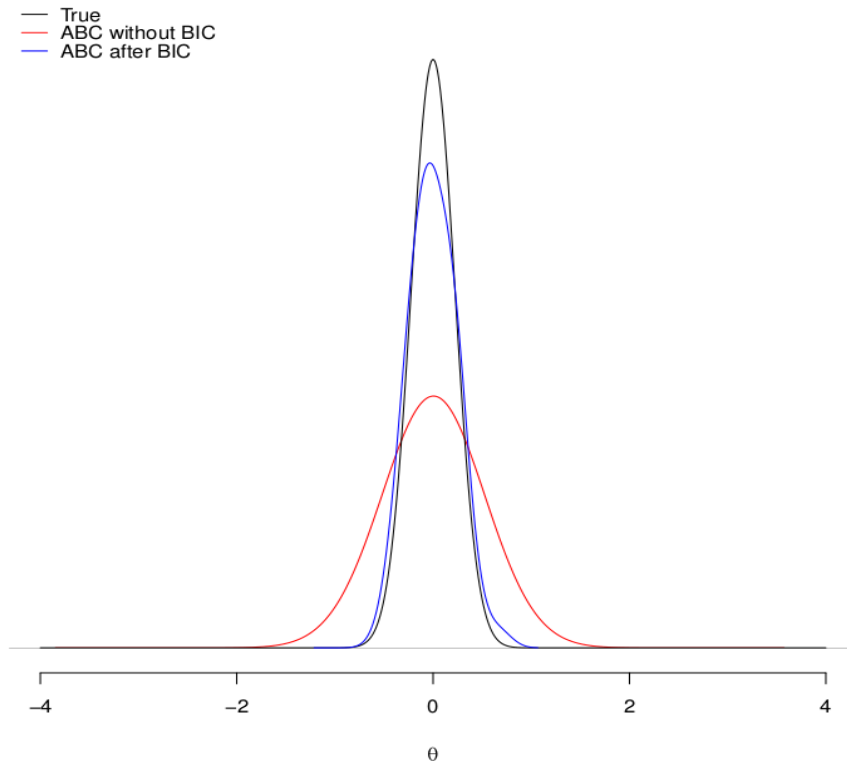


FIGURE 1.1: Estimation de paramètre dans une famille gaussienne (densité *a posteriori*).

**Estimation d'un paramètre sur un exemple jouet** Sur la figure 1.1, on présente les résultats obtenus sur un modèle jouet où  $\mathbf{x}$  est un vecteur aléatoire à valeurs dans  $\mathbb{R}^{20}$  tel que

$$x_i | \theta \sim \mathcal{N}(\theta, 1), \quad i = 1, \dots, 20$$

et le paramètre  $\theta$  est de loi *a priori* uniforme sur  $[-5, 5]$ . On s'intéresse à l'approximation de la loi *a posteriori* de  $\theta | \mathbf{x}_{\text{obs}}$  où  $\mathbf{x}_{\text{obs}}$  est fixé à  $(0, 0, \dots, 0)$ . Pour obtenir une approximation ABC, nous considérons les deux statistiques résumées  $S_1 = \sum_{i=1}^{20} \mathbf{x}_i / 20$ ,  $S_2 = \text{median}\{\mathbf{x}_1, \dots, \mathbf{x}_{20}\}$ . Pour mesurer la capacité du critère BIC à distinguer les statistiques résumées pertinentes du bruit, nous avons ajoutés deux statistiques supplémentaires  $S_3 \sim \text{Unif}(-5, 5)$  et  $S_4 \sim \mathcal{N}(0, 1)$ . Nous présentons l'estimation ABC (en bleu sur la figure 1.1) de *a posteriori* en considérant toutes les statistiques précédentes. Le critère BIC améliore l'estimation de la densité *a posteriori* (courbe noire sur la figure 1.1) en éliminant les statistiques  $S_3$  and  $S_4$ .

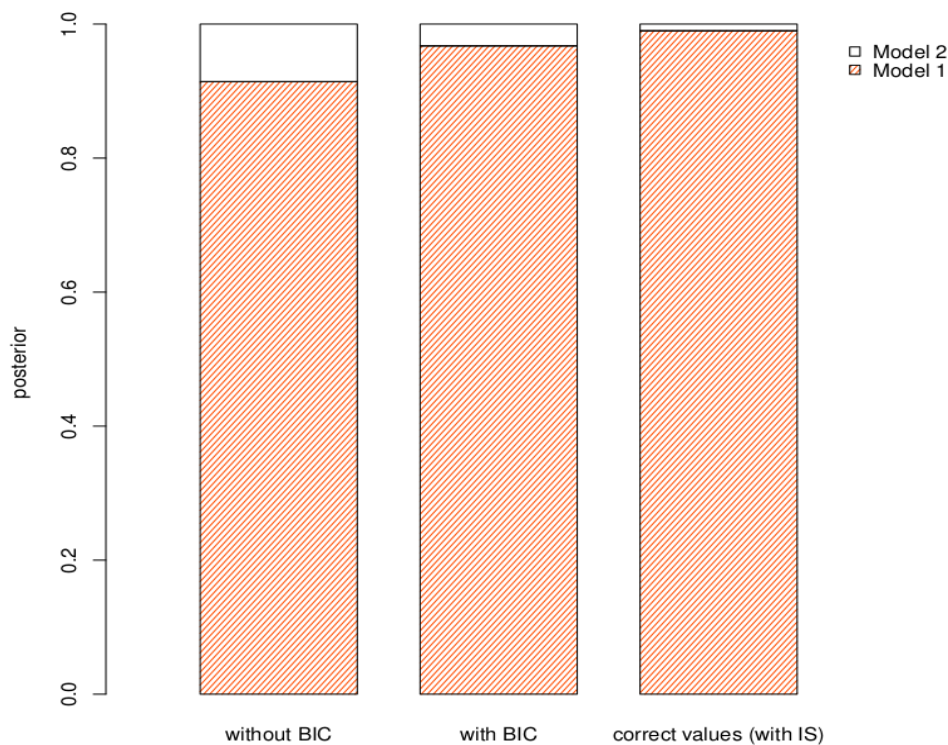


FIGURE 1.2: Sélection de modèles en génétique des populations (probabilités *a posteriori*).

**Problème de choix de scénarios dans un modèle de génétique des populations** Résumons le problème de choix de modèle étudié et dont les résultats sont présentés sur la figure 1.2. En considérant trois populations, nous voulons décider si la population 3 provient de la population 1 (Model 2) ou 2 (Model 1). Le jeu de données (génotype voir chapitre 2) considéré comme observé est simulé suivant le premier modèle (Model 1). Nous avons calculé via la régression logisitique pondérée, les approximations ABC des probabilités *a posteriori* des deux modèles. Parmi les 24 statistiques résumées, le critère BIC retient les deux statistiques LIK31 et LIK32 (voir Tab. S1 de Robert et al. [90]) pour cette régression. Ces statistiques sélectionnées estiment les similarités génétiques entre la population 3 et les deux restantes. L'estimation de la probabilité *a posteriori* du modèle



1 s'améliore et s'approche de la vraie valeur en utilisant seulement les deux statistiques résumées sélectionnées.

# 2

## Modèles de génétique des populations

### 2.1 Introduction

Un des développements principaux de la modélisation en génétique des populations est l'utilisation des méthodes dites coalescentes ou généalogiques [42, 102, 100]. Le but est de reconstruire des éléments de l'histoire de ces populations. Pour examiner la structure des données génétiques, ces méthodes utilisent l'arbre généalogique des gènes. La formulation d'un modèle est contrainte par un scénario évolutif qui imite la réalité historique et démographique de l'espèce. Un tel scénario résume l'histoire évolutive des populations par une suite d'événements démographiques depuis une population ancestrale. Ces événements sont constitués de divergences avec ou sans remises en contact, des migrations et des variations de tailles entre les populations.

Nos jeux de données sont constitués d'informations génétiques issues de plusieurs locus. Il existe plusieurs options pour modéliser le lien entre ces différents locus : généalogie commune (liaison totale), généalogies partiellement partagées puis recombinaison ou généalogies indépendantes (aucune liaison). Ici, nous nous restreignons à l'hypothèse d'indépendance. Cette hypothèse est justifiée dès que les locus sont suffisamment éloignés dans le génome nucléaire.

La formulation de tels modèles est la première étape vers l'inférence statistique. Ici, on s'intéresse à la description d'une classe de modèles probabilistes constitués d'événements inter-populationnels comme la divergence, l'admixture et la migration. Les modèles [34] que nous décrivons sont sous l'hypothèse de neutralité [57, 58]. Cette hypothèse implique l'absence d'effet de sélection. Alors, le polymorphisme expliqué par les modèles évolutifs est le résultat de la superposition des mutations génétiques sur la généalogie des individus.

Avec ces modèles, nous pouvons répondre à de nombreuses questions d'intérêt biologique. Par exemple, nous pouvons dater des divergences, quantifier des réductions ou des augmentations de tailles efficaces de populations, inférer des taux de migration, *etc*, en mettant en place une procédure d'inférence des paramètres du modèle. Nous pouvons également déterminer de quelle sources ancestrale provient une population récente, dé-

crire des voies d'invasion de populations, *etc.* Il faut alors utiliser une procédure de choix de modèle, chaque hypothèse correspond à un scénario démographique.

Résumons le contenu de ce chapitre. Un scénario évolutif est une série d'événements spatio-temporels ordonnés du plus récent au plus ancien (figure 2.1). Ces modèles expliquent des jeux de données à l'aide d'un processus latent. La composante latente inclut une généalogie et un ensemble de génotypes ancestraux. La généalogie (section 2.3) est représentée par un dendrogramme dont une lignée correspond à l'ascendance d'un individu depuis l'ancêtre commun le plus récent (MRCA) des échantillons observés. L'évolution des lignées d'une généalogie est régie par les différents événements inter-populationnels présents dans le scénario évolutif (section 2.3). Les génotypes d'un jeu de données dérivent de celui du MRCA via une série de mutations (section 2.4). Une réalisation d'un processus ponctuel positionne les mutations sur la généalogie. Selon le type de données de l'étude (section 2.2), un modèle mutationnel (section 2.4) génère le génotype du MRCA et applique les mutations sur les branches du dendrogramme. La composante latente ajoute une dimension temporelle au modèle évolutif. Ainsi, l'interprétation des paramètres de ce modèle pose des problèmes d'échelle de temps (section 2.5). Indépendamment de l'approche d'inférence statistique, le calcul de la vraisemblance  $\ell(\mathbf{x}|\phi)$  nécessite une marginalisation. Celle-ci s'obtient par une intégration sur l'espace de la composante latente (section 2.5).

## 2.2 Données

Pour mener nos études, nous utilisons des données génétiques que nous introduisons ici. Nous présentons également quelques éléments simplifiés de génétique pour expliquer la constitution de ces jeux de données et leur intérêt.

Le jeu de données est constitué de différents échantillons d'individus. Chaque échantillon correspond à une **population** géographique (appelée parfois **colonie** ou **dème**). Nous numérotions les populations de 1 à  $D$  et leur donnons les labels Pop1 à Pop $D$ . Contrairement à de nombreux problèmes en statistique, l'appartenance des individus échantillonnés aux populations d'intérêt est ici connue. Par population, la taille typique d'un échantillon varie entre une quarantaine et une centaine d'individus. La taille de l'échantillon issu de la population Pop $i$  est notée  $n_i$ .

La plupart des espèces sont **diploïdes**. Dans ce cas, les individus portent l'information génétique nucléaire en double : une copie issue de la **gamète** maternelle, une copie issue de la gamète paternelle. Cet appariement n'apporte aucune information lorsque l'équilibre d'Hardy-Weinberg [23, chapitre 2] est réalisé. On peut donc assimiler un individu diploïde aux deux gamètes qui l'ont engendré, c'est-à-dire à deux individus haploïdes. Dans toute la suite, nous considérons donc que les individus sont **haploïdes**.

Notons que le génome de nombreuses espèces n'est pas entièrement inclus dans le noyau de la cellule. Il existe par exemple un brin d'ADN dans la mitochondrie. Le processus de transmission de ce génome est plus simple (il est entièrement hérité de la mère).

C'est pourquoi il a beaucoup été utilisé dans les premières études de génétique des populations construites sur des modèles, voir par exemple les questions d'Ève mitochondriale [73, 50]. Pour simplifier l'exposé, on ne traitera pas ce cas ici.

Pour chacun des individus, l'information génétique que l'on considère dans l'étude est limitée. On ne s'intéresse qu'à quelques positions particulières du génome appelées **locus**. À ces locus, la séquence d'ADN peut varier d'un individu à l'autre, à cause des **mutations** au cours de l'évolution de l'espèce. On parle alors de **polymorphisme génétique**. Les différentes variantes s'appellent des **allèles** ou des états alléliques. La constitution de notre jeu de données a nécessité de déterminer l'allèle que porte chacun des individus pour tous les locus de l'étude. Cette opération de **génotypage** fait appel à de la micro-biologie que nous ne décrivons pas ici, et nous admettons que le résultat fourni est exact (sans bruit ou erreur de mesure).

Les modèles probabilistes que nous présentons ici expliquent le polymorphisme en retraçant l'évolution de l'espèce et les mutations au cours du temps. Nous devons donc détailler quelles sont les formes possibles d'information génétique pour chaque locus. En fait, il existe trois types de locus : **microsatellite**, séquence ou **SNP** (Single Nucleotide Polymorphism). Nous nous concentrons ici sur les locus microsatellites. Il s'agit d'une partie de l'ADN où un court motif (de 1 à 4 paires de base) est répété en de nombreux exemplaires. L'information que l'on retrouve dans notre jeu de données est alors la longueur totale de cette séquence (en nombre de paires de base). À cause de ces répétitions, cette portion d'ADN est fortement variable (fort polymorphisme) et a donc été beaucoup utilisé en génétique des populations. Nous décrivons deux modèles mutationnels classiques sur ce type de locus dans la section 2.4.

Enfin, dans les modèles que nous utilisons, nous considérons que les locus sont indépendants. Notre vraisemblance s'écrit donc comme un produit de vraisemblances du jeu de données restreint à un locus donné. Cette hypothèse cache en fait un modèle de vraisemblance composite [55, 98]. Grâce au brassage dû à la **recombinaison génétique**, cette approximation est valable dès lors que les locus sont suffisamment éloignés sur le génome (en particulier pour que les généalogies d'un locus à l'autre soient indépendantes).

## 2.3 Généalogie d'échantillons

Cette section est composée de deux parties. Premièrement, on se restreint à un scénario très simple avec une seule population fermée à l'équilibre. On introduit l'outil fondamental pour simuler une généalogie qui est le processus appelé coalescent de Kingman [59, 61, 60]. Ensuite, on s'intéresse à la généalogie de populations géographiquement structurées. Ces populations sont structurées par des événements inter-populationnels. On décrit l'évolution des lignées d'une généalogie en présence pour d'événements inter-populationnels instantanés, qui sont la divergence et l'admixture, et d'événements persistants de migration.

Certains modèles (Wright-Fisher et Moran voir [102, chapitre 3]) proposent de simuler

l'évolution de la population entière, du passé au présent, puis d'échantillonner la dernière génération. Pour une population de grande taille, la simulation suivant ce type de modèles est très lente. D'où la méthodologie que l'on adoptera et qui consiste à considérer seulement l'échantillon d'intérêt au lieu de la totalité de la population. On s'intéresse seulement à l'évolution des ascendants des individus de notre échantillon en remontant le temps.

### 2.3.1 Une seule population : Coalescent de Kingman

---

#### Algorithme 3 Coalescent de Kingman en temps naturel

---

ENTRÉES : La taille de l'échantillon  $k$ , la taille efficace de la population  $N_e$ .

**Tantque**  $k \geq 2$  **faire**

- 1) Simuler le temps inter-coalescent  $T_k$  suivant une loi exponentielle de paramètre  $\frac{k(k-1)}{2N_e}$ .
- 2) Augmenter les longueurs des  $k$  lignées de  $T_k$ .
- 3) Parmi les  $k$  lignées, choisir aléatoirement deux lignées à regrouper pour former un nœud du dendrogramme.
- 4)  $k \leftarrow k - 1$ .

**Fin tantque**

---

Introduisons le coalescent de Kingman [59, 61, 60], qui est l'outil fondamental pour la simulation des généalogies. Par souci de clarté, nous nous contentons d'un scénario basique formé d'une seule population fermée à l'équilibre. Nous supposons que cette population n'est pas soumise à un flux extérieur de gènes et ne subit aucune variation démographique interne.

La généalogie d'un échantillon d'individus est représenté par un dendrogramme (figure 2.2). On génère des lignées ancestrales jusqu'à l'ancêtre commun le plus récent (MRC-A en anglais). Chaque lignée d'un individu passe par une série d'ancêtres. Un évènement de coalescence se produit lorsque les lignées de deux individus se rejoignent en un noeud du dendrogramme (figure 2.2). La généalogie d'un échantillon de  $k$  individus est donc composée de  $k - 1$  évènements de coalescence. Chaque évènement décroît le nombre de lignées ancestrales de 1 jusqu'à la dernière lignée (racine du dendrogramme) qui correspond au MRCA.

Les variables  $T_k, \dots, T_2$  représentent les durées entre les évènements de coalescences (figure 2.2) successifs. La loi de la généalogie de  $k$  individus est entièrement caractérisée par la loi du choix des lignées à chaque évènement de coalescence et la loi des durées entre évènements  $T_k, \dots, T_2$ . Pour le coalescent de Kingman, les durées entres évènements de coalescences  $T_k, \dots, T_2$  sont indépendantes et  $T_k$  suit la loi exponentielle de paramètre  $k(k-1)/2$ .

Ici, nous détaillons la réalisation d'un évènement de coalescence dans un échantillon ancestral de taille  $k$ . Pour chacun des  $\binom{k}{2}$  couples de lignées en compétition pour une coalescence, on associe une horloge exponentielle de paramètre 1. L'horloge qui réalise le

temps minimal désigne le couple de lignées à coalescer. En pratique, la réalisation des  $\binom{k}{2}$  variables exponentielles de paramètre 1 est lent. Nous pouvons diminuer le nombre de variables aléatoires exponentielles à réaliser avec le lemme ci-dessous.

**Lemme 2.1.** Soient  $T_1, \dots, T_\ell$  des variables aléatoires exponentielles indépendantes de paramètres respectifs  $\lambda_1, \dots, \lambda_\ell$ . Alors la variable aléatoire  $\inf_{1 \leq i \leq \ell} T_i$  suit une loi exponentielle de paramètre  $\sum_{i=1}^{\ell} \lambda_i$ . Et la variable aléatoire  $T_k$  réalise cet infimum avec probabilité  $\lambda_k / \sum_{i=1}^{\ell} \lambda_i$ .

Ainsi, à l'aide du lemme 2.1, la réalisation d'un événement de coalescence revient donc à augmenter les  $k$  lignées du dendrogramme d'une réalisation exponentielle de paramètre  $\binom{k}{2}$ . Le couple de lignées à coalescer est choisi aléatoirement parmi les  $\binom{k}{2}$  couples.

Une unité de temps coalescent s'interprète comme  $Ne$  générations, où  $Ne$  est un paramètre du modèle qui s'appelle taille efficace de la population (plus de détails en section 2.5). On décrit dans l'algorithme 3 le coalescent de Kingman sur un échantillon à  $k$  individus d'une population de taille efficace  $Ne$ . Dans cet algorithme, le temps est à l'échelle naturelle, et donc le taux de coalescence dans la généalogie est linéaire en  $Ne$ .

Enfin, notons que pour ne pas croiser les lignées du dendrogramme, l'ordre des individus sur les feuilles est différent l'ordre des individus dans le jeu de données. Sur l'exemple de généalogie donné dans la figure 2.2, les numéros en bas du dendrogramme correspondent aux numéros des individus dans le jeu de données.

### 2.3.2 Plusieurs populations structurées

---

#### Algorithme 4 Généalogie (partielle) dans une population indépendante

---

**ENTRÉES :** La taille de l'échantillon ancestral notée  $k$  à la date  $t$  dans la population d'intérêt. La taille efficace  $Ne$  de la population et les dates des événements  $t$  et  $t'$ .

Simuler  $T_k$  suivant une loi exponentielle de paramètre  $k(k-1)/2Ne$ .

**Tantque**  $(t + T_k) \leq t'$  **faire**

- 1) Augmenter les longueurs des  $k$  lignées d'une longueur  $T_k$ .
- 2) Choisir aléatoirement parmi les  $k$  lignées, deux lignées à regrouper pour former un nœud du dendrogramme.
- 3)  $k \leftarrow k - 1$ .
- 4) Simuler la durée inter-coalescence  $T_k$  suivant une loi exponentielle de paramètre  $k(k-1)/2Ne$ .

**Fin tantque**

**Si**  $(t + T_k) > t'$  **alors**

Augmenter les lignées restantes jusqu'à la hauteur  $t'$ .

**Fin si**

---

Décrivons la loi d'une généalogie pour un scénario évolutif dont la structure géographique est gouvernée par des événements inter-populationnels. On combine ces événements avec le coalescent de Kingman qui décrit la généalogie intra-populationnel. Présentons succinctement les trois types d'événements inter-populationnels.

**Algorithme 5** Généalogie de populations avec migration en temps coalescent

ENTRÉES : Les tailles des échantillons ancestraux:  $k_1, \dots, k_D$  et les taux de migration  $m_{ij}$ .

**Pour**  $i=1 \rightarrow D$  **faire**

- 1) Associer une horloge exponentielle de paramètre  $1/Ne_i$  pour chaque couple d'individus de la population  $i$  qui correspond à une coalescence potentielle.
- 2) Associer  $D - 1$  horloges exponentielles de paramètres  $m_{ij}, 1 \leq j \neq i \leq D$  pour chaque individu de la population  $i$  qui correspondent à des migrations potentielles.

**Fin pour**

Parmi toutes les horloges en compétition, celle qui sonne en premier gagne. Si cette horloge correspond à un couple d'individus, on fait coalescer ces deux individus. Si c'est l'horloge d'un seul individu, et celle-ci est de paramètre  $m_{ij}$ , on déplace la lignée de cet individu de la population  $i$  vers la population  $j$ .

- La divergence (figure 2.3, (a)) est la fusion de deux populations.
- L'admixture (figure 2.3, (b)) est le partage d'une population en deux parties à l'instant de l'événement. Les lignées sont envoyées dans les deux autres populations. La destination d'une lignée est contrôlée par un paramètre du modèle appelé, taux d'admixture.
- La migration (figure 2.3, (c)) autorise le déplacement des lignées d'une population à l'autre sur une période donnée, suivant des taux par unité de temps et par gène.

La divergence et l'admixture sont des événements instantanés alors que la migration est persistante sur une période de temps.

La procédure de simulation de la généalogie dans ces trois situations simule l'évolution des lignées du dendrogramme entre deux dates d'événements inter-populationnels successifs et les changements instantanés dans les lignées de la généalogie pour ces événements instantanés. En cas de migration sur une période donnée, l'évolution des lignées du dendrogramme sur cette période est différente des situations précédentes. Les événements de coalescences et de déplacements de lignées sont en concurrence. L'algorithme 7 résume les étapes de simulation d'une généalogie contrainte par un modèle démographique avec différents événements inter-populationnels.

**Généalogie entre événements inter-populationnels instantanés** L'évolution de la généalogie intra-populations entre deux dates (notées  $t$  et  $t'$  où  $t' > t$  dans la figure 2.3 (a) et (b)) d'événements suit un coalescent de Kingman indépendant sur les lignées présentes dans chaque population. L'algorithme 4 décrit la généalogie à l'intérieur de chacune des populations. Il diffère légèrement de l'algorithme 3 décrivant l'évolution d'une population fermée à l'équilibre jusqu'au MRCA. En effet, l'algorithme 4 décrit le coalescent de Kingman sur une période  $t' - t$  entre deux événements inter-populationnels. En cas de présence d'un changement de taille efficace (figure 2.1) dans la population à une date, il suffit de changer l'échelle de temps après cette date et donc de remplacer  $Ne$  par  $Ne'$  dans la simulation des durées inter-coalescences  $T_k$ .

**Algorithme 6** Généalogie de 2 populations avec migration

**ENTRÉES :** Les tailles des deux échantillons ancestraux  $k_1$  et  $k_2$ , les taux de migration  $m_{12}$  et  $m_{21}$  et les tailles efficaces des deux populations  $Ne_1$  et  $Ne_2$ .

1) On choisit la population numéro  $i$  parmi les deux avec probabilité

$$\frac{k_i m_{ij} + [k_i(k_i - 1)/2Ne_i]}{k_1 m_{12} + [k_1(k_1 - 1)/2Ne_1] + k_2 m_{21} + [k_2(k_2 - 1)/2Ne_2]}.$$

2) On choisit le type d'évènement : soit une coalescence avec probabilité

$$\frac{k_i(k_i - 1)/2Ne_i}{k_i m_{ij} + [k_i(k_i - 1)/2Ne_i]},$$

soit une migration avec probabilité

$$\frac{k_i m_{ij}}{k_i m_{ij} + [k_i(k_i - 1)/2Ne_i]}.$$

3) Pour simuler une coalescence dans la population  $i$ :

- On simule  $T_c$  l'instant de coalescence de suivant une loi exponentielle de paramètre  $k_i(k_i - 1)/2Ne_i$ .
- On augmente les lignées de la généalogie d'une longueur  $T_c$ .
- On tire les deux lignées à joindre uniformément parmi les  $k_i$  lignées de la population  $i$  et on applique la coalescence.
- $k_i \leftarrow k_i - 1$ , et revenir en 1.

4) Pour simuler une migration de la population  $i$  à la population  $j$ :

- On simule  $T_m$  l'instant de migration de suivant une loi exponentielle de paramètre  $k_i m_{ij}/Ne_i$ .
- On augmente les lignées de la généalogie d'une longueur  $T_m$ .
- On migre une lignée choisie uniformément dans la population  $i$  vers l'autre population  $j \neq i$ .
- $k_i \leftarrow k_i - 1$  et  $k_j \leftarrow k_j + 1$ , et revenir en 1.

**Divergence** À l'instant de la divergence, les lignées présentes dans les deux populations (Pop1 et Pop2 de la figure 2.3 (a)) sont regroupées pour former une seule population (Pop1 de la figure 2.3 (a)).

**Admixture** À l'instant de l'admixture, l'échantillon ancestral de Pop3 (voir figure 2.3, (b)) est partagé sur les deux autres populations ainsi: une lignée de la population Pop3 est envoyée dans Pop1 avec probabilité  $r$  et dans Pop2 avec probabilité  $1 - r$ , où  $r$  est un paramètre du modèle appelé taux d'admixture.

**Migration** Nous décrivons dans l'algorithme 5 l'évolution des lignées en présence d'une migration entre  $D$  populations. La migration est paramétrée dans le modèle par les taux de migration de la population  $i$  vers la population  $j$ , notés  $m_{ij}$ . L'algorithme 5 détaille toutes les horloges en compétition dont les lois sont exponentielles avec les paramètres



appropriés. L'horloge qui réalise le temps minimal de ce système désigne le type d'événement, les populations et les individus concernés par l'événement. En pratique, la simulation d'une généalogie de populations structurées par une migration nécessite beaucoup moins d'horloges exponentielles que dans l'algorithme 5. Comme dans le cas de la partie 2.3.1, nous pouvons simplifier cet algorithme à l'aide du lemme 2.1. Nous donnons dans l'algorithme 6 un exemple avec  $D = 2$  populations, et nous renvoyons le lecteur à Wakeley [102, chapitre 5] pour le cas  $D > 2$ . La généalogie décrite dans cet algorithme correspond au scénario de la figure 2.3 (c). Cet algorithme est répété sur les populations Pop1 et Pop2 jusqu'à ce que les lignées du dendrogramme soient à la hauteur  $t'$ .

En conclusion, nous donnons le schéma général de simulation d'un dendrogramme contraint par de tels événements dans l'algorithme 7. Le coalescent de Kingman est bien la pierre de base de ces généalogies de gènes, qui est utilisé par morceaux, ou en compétition avec de la migration. De l'évolution des populations d'intérêt, nous avons extrait le seul processus utile au regard de l'échantillon grâce à cette généalogie. Reste à modéliser la dérive génétique le long de ses branches.

---

#### Algorithme 7 Généalogie contrainte par un scénario

---

Trier les événements inter-populationnels du plus récent au plus ancien.

**Pour**  $t$  allant de l'événement le plus récent au plus ancien **faire**

- 1) Simuler la généalogie intra-populationnelle: un coalescent de Kingman indépendant par population jusqu'à  $t$  ou combiner le coalescent de Kingman avec migration dans le cas d'une migration.
- 2) Appliquer l'événement inter-populationnel instantané à la date  $t$ .

**Fin pour**

Simuler un (ou des) coalescent(s) de Kingman (avec migrations) sur la (les) dernière(s) population(s) jusqu'au MRCA.

---

## 2.4 Processus mutationnels

Voyons maintenant quelle est la loi des génotypes de l'échantillon conditionnellement à une généalogie. Nous décrivons les positions des mutations comme une réalisation d'un processus ponctuel de Poisson sur les branches du dendrogramme. Ensuite nous introduisons deux modèles mutationnels sur locus microsatellite. L'application des mutations revient à faire des pas d'une chaîne de Markov associée au modèle mutationnel.

**Positions des Mutations** Le taux de mutation par unité de temps naturel et par individu diploïde est le paramètre  $\mu$ . Ainsi, conditionnellement à une généalogie, les positions des mutations sont données par un processus ponctuel de Poisson d'intensité  $\mu/2$  sur le dendrogramme. Autrement dit, sur une branche de longueur  $t$ , le nombre  $N$  de mutations suit une loi de Poisson de paramètre  $\mu t/2$ , et les  $N$  mutations sont uniformément réparties sur cette branche.

**Algorithme 8** Processus mutationnel le long d'une généalogie

ENTRÉES : Une généalogie  $\mathcal{G}$ , un taux de mutation  $\mu$ , et chaîne de Markov de matrice de transition  $Q$  et de la stationnaire  $\nu$ .

- 1) Appliquer un processus ponctuel de Poisson sur les branches de  $\mathcal{G}$ .
- 2) Trier les événements de mutation sur  $\mathcal{G}$  du plus ancien au plus récent.
- 3) Simuler suivant  $\nu$ , le génotype du MRCA.
- 4) Parcourir  $\mathcal{G}$  du MRCA jusqu'aux feuilles: construire les génotypes le long des branches.
  - si événement de coalescence **alors** dupliquer le génotype en haut du noeud du dendrogramme.
  - si événement de mutation **alors** appliquer un pas de la chaîne  $Q$  sur la branche (*i.e* l'individu qui porte la mutation).

**Modèles Mutationnels sur locus microsatellite** Présentons maintenant le processus d'évolution à chaque mutation. Nous introduisons ici deux modèles mutationnels (voir [104] et [20]) SMM (Stepwise Mutation Model) et GSM (Generalized Mutation Model) spécifiques au locus microsatellite. Ces deux modèles mutationnels ont l'avantage d'avoir une paramétrisation simple. Les chaînes de Markov associées aux modèles GSM et SMM sont des marches aléatoires symétriques sur un intervalle de nombres entiers  $\llbracket a; b \rrbracket$  de  $\mathbb{N}$ . Appliquer un pas de la chaîne de Markov GSM, revient à modifier le locus d'une longueur  $\pm mG$ , où  $m$  est la longueur du motif répété (supposée connue),  $G$  est une variable aléatoire de loi géométrique de paramètre  $p$  et  $\pm$  un signe aléatoire. En pratique, le paramètre  $p$  est de l'ordre de 0.2. Quant au modèle SMM, une mutation revient à diminuer ou augmenter (avec probabilités 1/2 et 1/2 respectivement) le locus d'une longueur de  $m$  paires de base où  $m$  est la longueur du motif répété. Il arrive qu'en appliquant les mutations, le génotype dépasse les bornes  $a$  et  $b$  de l'ensemble des états alléliques. Dans ce cas, le génotype de l'ancêtre prend comme valeur la plus proche des deux bornes  $a, b$  de l'espace des états alléliques.

Pour simuler les génotypes de l'échantillon en un locus donné, il suffit de faire évoluer le génotype du MRCA le long de la généalogie jusqu'au présent en appliquant les mutations (algorithme 8). Le génotype du MRCA provient de la loi stationnaire du modèle mutationnel. Observons le processus qui génère les génotypes des individus (figure 2.4). L'évolution de l'ascendance d'un individu le long d'une lignée est un processus markovien de sauts purs sur  $\llbracket a; b \rrbracket$ . Ici, la chaîne de Markov du modèle mutationnel est la chaîne incluse de ce processus. La chaîne incluse et le processus markovien ont la même loi stationnaire car le taux de mutation (taux de saut) ne dépend pas de l'état allélique. Les génotypes des individus observés sont donnés par les fins de trajectoires de ce processus (figure 2.4). Une trajectoire correspond à une lignée de la figure 2.4. Ces trajectoires sont corrélées et cette dépendance est décrite par les branches partagées par différentes lignées. Puisque le génotype du MRCA provient de la loi stationnaire, on déduit que la loi marginale du génotype d'un individu est cette même loi stationnaire. En revanche, la loi jointe des génotypes de tous les individus est très complexe à décrire. La structure de corrélation entre les trajectoires du processus est donnée par les branches partagées par les lignées du dendrogramme. Par exemple, sur la figure 2.4, les processus markoviens

expliquent les génotypes des gènes 2 et 4 (représentés en rouge et vert respectivement sur la figure) ont en commun une grande partie de leurs trajectoires. La loi marginale de l'échantillon (*i.e* loi déconditionnée) n'a pas d'écriture plus simple qu'une intégrale contre la loi de la généalogie.

## 2.5 Conclusion

**Vraisemblance** Notons  $f_\phi(\mathcal{G})$  la densité de la loi de la généalogie de gènes par rapport à une mesure de référence  $d\mathcal{G}$ . Rappelons que nous avons décrit cette loi à partir du coalescent de Kingman dans la section 2.3. De même, notons  $f_\phi(\mathcal{M}|\mathcal{G})$  la densité du processus mutationnel  $\mathcal{M}$  sachant la généalogie  $\mathcal{G}$  décrit dans la section 2.4. La vraisemblance du jeu de données complet  $\mathbf{x}$  s'écrit donc

$$\ell(\mathbf{x}|\phi) = \prod_{i \in \{\text{locus}\}} \int_{\mathcal{M}_i \rightarrow \mathbf{x}_i} f_\phi(\mathcal{M}_i|\mathcal{G}_i) f_\phi(\mathcal{G}_i) d\mathcal{G}_i d\mathcal{M}_i, \quad (2.1)$$

où  $\mathbf{x}_i$  est l'ensemble des données au locus  $i$  et  $\mathcal{M}_i \rightarrow \mathbf{x}_i$  désigne l'ensemble des génotypes sur le dendrogramme dont les feuilles correspondent à l'échantillon observé. Nous faisons ici l'hypothèse d'indépendance entre les différents locus, voir section 2.2.

Cette vraisemblance ne se calcule pas facilement. L'intégrale précédente est sur l'espace des couples  $(\mathcal{G}_i, \mathcal{M}_i)$  compatibles avec l'échantillon  $\mathbf{x}_i$ . Cet espace est de très grande dimension et comporte des directions discrètes comme les génotypes des ancêtres et des parties continues comme les différentes hauteurs dans la généalogie. En dépit de la simplicité du coalescent de Kingman et du processus mutationnel, on ne peut espérer aucune simplification formelle dans cette intégrale.

Il existe une formule de récurrence [41, 49] reliant la vraisemblance  $\ell(\mathbf{x}_i|\phi)$  d'un locus à des vraisemblances  $\ell(\mathbf{y}|\phi)$ , où  $\mathbf{y}$  est un échantillon mono-locus de taille inférieure ou égale à celle de  $\mathbf{x}_i$ . Par exemple, dans le cas d'une population fermée à l'équilibre, De Iorio and Griffiths [25, équation (3)] et, dans le cas de plusieurs populations avec migration, De Iorio and Griffiths [26, équation (2)]. En pratique, l'utilisation de cette formule de récurrence pour calculer la vraisemblance des données est impossible. On se retrouve à devoir calculer la vraisemblance de tous les échantillons possibles de taille inférieure ou égale à celle du jeu de données. L'algorithme est alors de complexité exponentielle en la taille du jeu de données et la taille de l'espace allélique : c'est un phénomène bien connu d'explosion combinatoire.

Des techniques d'échantillonnage préférentiel ont été proposées dans Stephens and Donnelly [97], De Iorio and Griffiths [25, 26], De Iorio et al. [27] pour approcher le calcul de la vraisemblance. Leurs lois d'importance sont les solutions d'un système de récurrence simplifié. Le calcul de la vraisemblance avec une certaine précision (1% de variabilité) nécessite un temps relativement élevé. Sur un modèle simple de deux populations avec migration, le temps de calcul de la vraisemblance  $\ell(\mathbf{x}_i|\phi)$  effectué par le logiciel *Migraine* de Rousset and Leblois [92] pour une valeur du paramètre  $\phi$  est de l'ordre de la minute.

**Simulation et estimation** Les techniques développées dans cette thèse s’appliquent parfaitement au contexte où la vraisemblance (2.1) est difficile à calculer. Les outils disponibles dans la littérature sont nombreux mais ne recouvrent pas toutes les questions posées en pratique. Citons quelques exemples représentatifs de logiciels de simulation et d’estimation.

- *Ms* de Hudson [56] simule par coalescence des scénarios démographiques avec migration entre populations.
- *IBDSim* de Leblois et al. [65] simule des scénarios en présence de migration (isolation par la distance) en temps discret, génération par génération, et échantillonne la dernière génération.
- *DIYABC* de Cornuet et al. [22] simule des scénarios démographiques avec des événements de divergence et d’admixture mais sans présence de migration. Notons que *DIYABC* fournit une multitude de procédures d’estimation et de sélection de modèles par ABC.
- *Migraine* de Rousset and Leblois [92] utilise l’échantillonnage préférentiel décrit ci-dessus pour estimer des surfaces de vraisemblance.
- *IM*, *IMa* de Nielsen and Wakeley [81], Hey and Nielsen [53, 54] analysent des modèles démographiques d’isolation par migration. Ils incluent les deux approches bayésienne et fréquentielle. L’approximation de la loi *a posteriori* et de la surface de vraisemblance se fait par des méthodes MCMC.

**Échelle de temps** Le processus latent  $(\mathcal{G}, \mathcal{M})$  introduit une dimension temporelle pour expliquer les données échantillonnées. La plupart des coordonnées du paramètre de  $\phi$  sont exprimés dans l’échelle de cet axe. Mais les données, toutes collectées à l’époque actuelle, ne contiennent aucune information sur cette échelle. En effet, une homothétie de rapport  $\lambda$  sur l’axe temporel ne change pas la loi marginale des données  $\mathbf{x}$  si on effectue les transformations suivantes sur les composantes de  $\phi$ :

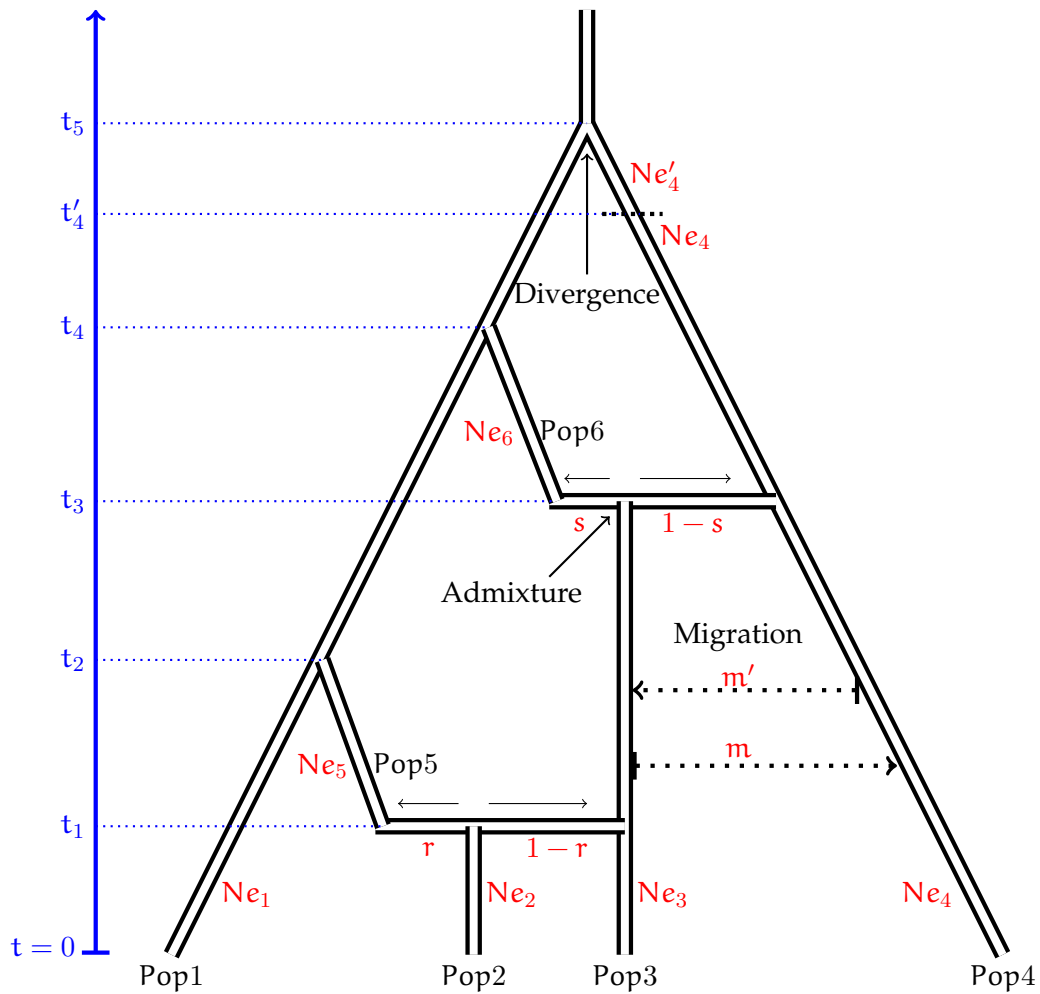
- les paramètres  $t$ , de type date, sont changés en  $\lambda t$ ,
- les paramètres  $Ne$ , de type taille efficace, sont changés en  $\lambda Ne$ .
- les paramètres  $\mu$ , de type taux de mutations, sont changés en  $\mu/\lambda$  et
- les paramètres  $m$ , de type taux de migrations, sont changés en  $m/\lambda$ .

L’analyse fréquentielle se heurte donc à un problème d’identifiabilité. On peut re-paramétriser le modèle de telle sorte que ce problème disparaisse. Pour cela, il est commode d’introduire une taille efficace de référence, notée  $Ne_{REF}$ , combinaison linéaire des tailles efficaces en échelle naturelle. Les paramètres identifiables sont alors:

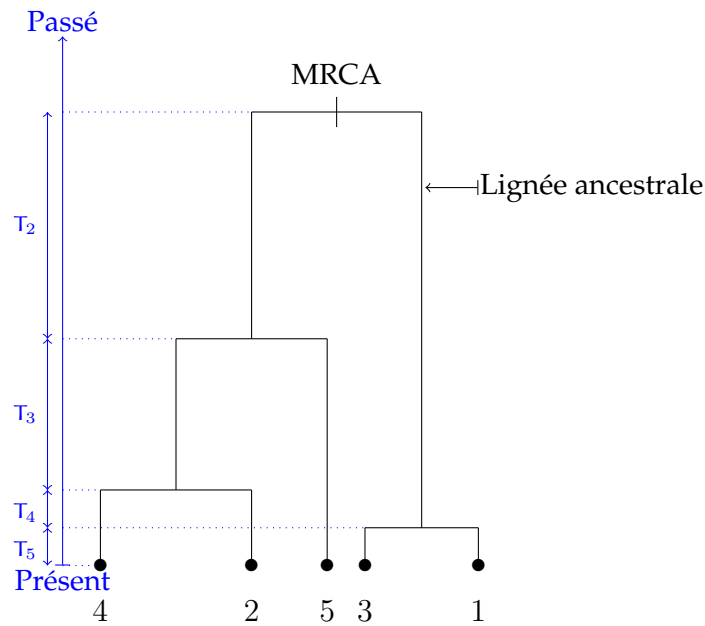
- $\overline{\tau}_i = t_i/Ne_{REF}$ ,
- $\overline{Ne}_i = Ne_i/Ne_{REF}$ ,
- $\theta_i = 4Ne_{REF} \mu_i$  et
- $\overline{m}_{ij} = Ne_{REF} m_{ij}$ ,

quitte à supprimer l’une des coordonnées de type taille efficace de  $\phi$ . Il est courant de choisir comme valeur de  $Ne_{REF}$  la somme des tailles efficaces des populations échantillonnées. Par exemple, on choisirait  $Ne_{REF} = Ne_1 + Ne_2 + Ne_3 + Ne_4$  sur l’exemple de la figure 2.1. L’interprétation de ces paramètres identifiables pour le biologiste nécessite de choisir “*au doigt mouillé*” une échelle de temps. Ce problème ne se pose pas lorsque l’on mène une analyse bayésienne. La loi *a posteriori* sur  $\phi$  est bien définie même si ce vecteur de paramètres n’est pas identifiable. L’échelle de temps est alors choisie à l’aide de la loi *a priori*.

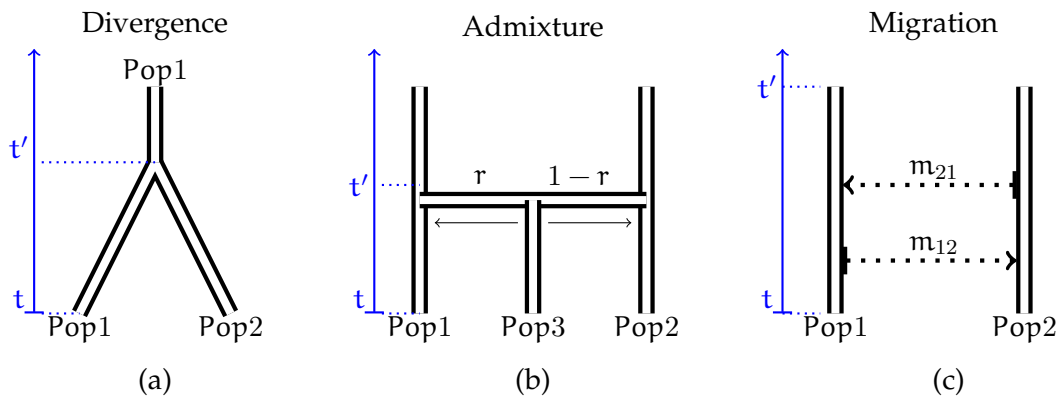
Cela permet d'introduire une variabilité sur le choix arbitraire de cette échelle, codée au travers d'une loi sur des paramètres directement interprétables. C'est un argument en faveur d'une approche bayésienne, clairement mis en avant par Beaumont and Rannala [8].



**FIGURE 2.1:** Exemple d'un scénario évolutif complexe composé d'évènements inter-populationnels. Ce scénario implique quatre populations échantillonnées Pop1, ..., Pop4 et deux autres populations non-observées Pop5 et Pop6. Les branches de ce schéma sont des "tubes" et le scénario démographique contraint la généalogie à rester à l'intérieur de ces "tubes". La migration entre les populations Pop3 et Pop4 sur la période  $[0, t_3]$  est paramétrée par les taux de migration  $m$  et  $m'$ . Les deux évènements d'admixture sont paramétrés par les dates  $t_1$  et  $t_3$  ainsi que les taux d'admixture respectifs  $r$  et  $s$ . Les trois évènements restants sont des divergences, respectivement en  $t_2$ ,  $t_4$  et  $t_5$ . L'évènement en  $t'_4$  correspond à un changement de taille efficace dans la population Pop4.

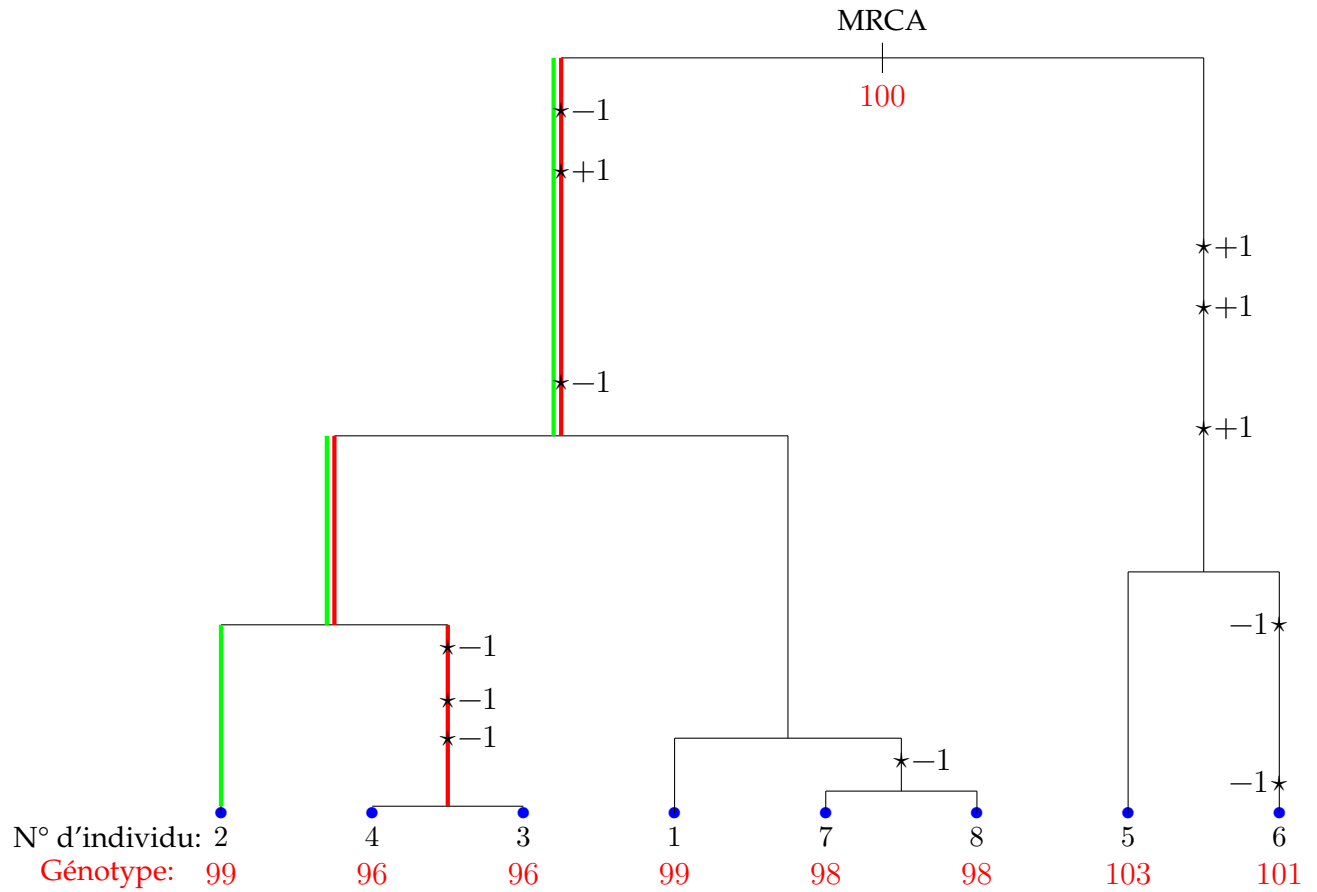


**FIGURE 2.2:** Exemple de généalogie de cinq individus issus d'une seule population fermée à l'équilibre. Les individus échantillonnés sont représentés par les feuilles du dendrogramme, les durées inter-coalescences  $T_2, \dots, T_5$  sont indépendantes, et  $T_k$  est de loi exponentielle de paramètre  $k(k-1)/2$ .



**FIGURE 2.3:** Représentations graphiques des trois types d'évènements inter-populationnels d'un scénario démographique. Il existe deux familles d'évènements inter-populationnels. La première famille est simple, elle correspond aux évènements inter-populationnels instantanés. C'est le cas d'une divergence ou d'une admixture. (a) Deux populations qui évoluent pour se fusionner dans le cas d'une divergence. (b) Trois populations qui évoluent en parallèle pour une admixture. Pour cette situation, chacun des tubes représente (on peut imaginer qu'il porte à l'intérieur) la généalogie de la population qui évolue indépendamment des autres suivant un coalescent de Kingman.

La deuxième correspond à la présence d'une migration. (c) Cette situation est légèrement plus compliquée que la précédente à cause des flux de gènes (plus d'indépendance). Ici, un seul processus évolutif gouverne les deux populations réunies. La présence de migrations entre les populations Pop1 et Pop2 implique des déplacements de lignées d'une population à l'autre et ainsi la concurrence entre les évènements de coalescence et de migration.



**FIGURE 2.4:** Exemple de simulation des génotypes d'un échantillon de huit individus en un locus microsatellite. Les mutations sont données par le modèle SMM. Les positions des mutations sont représentées par les étoiles (\*) sur les lignées du dendrogramme. Le génotype d'un individu (représenté par un nombre entier en rouge sur le dendrogramme) est obtenu en appliquant les mutations à partir du génotype du MRCA (100), ici, le long de la lignée de l'individu.





## **Deuxième partie**

# **Échantillonnage Préférentiel Adaptatif et Multiple**



# 3

## Échantillonnage Préférentiel

Nous rappelons dans ce chapitre l'échantillonnage préférentiel (Importance Sampling en anglais) basique introduit par Marshall [79]. Cette technique contourne la simulation suivant la cible (inaccessible en général) par un tirage suivant une loi de proposition (dite aussi loi d'importance) et associe à chaque particule issue de ce tirage un poids d'importance. Le résultat de cette technique est un système de particules pondérées dont la loi empirique approche la loi cible.

### 3.1 Généralités sur l'échantillonnage préférentiel

**Échantillonnage préférentiel basique** Étant donnée une loi cible  $\Pi$  de densité  $\pi$ , l'échantillonnage préférentiel vise à approcher des intégrales de type

$$\Pi(\psi) = \int \psi(\mathbf{x})\pi(\mathbf{x}) \, d\mathbf{x}, \quad (3.1)$$

pour une certaine fonction test  $\psi \in \mathbb{L}^1(\Pi)$ . Pour cela on échantillonne une loi de proposition  $Q$  de densité  $q$ , c'est-à-dire, on tire  $N$  particules indépendantes

$$\mathbf{x}_1, \dots, \mathbf{x}_N \quad \text{de même loi } Q(\mathbf{x} \in d\mathbf{x}).$$

Et on approche l'intégrale (3.1) par l'estimateur

$$\hat{\Pi}_N^{\text{IS}}(\psi) = \frac{1}{N} \sum_{i=1}^N \omega_i \psi(\mathbf{x}_i), \quad (3.2)$$

avec  $\omega_i$  le poids d'importance associé à la particule  $\mathbf{x}_i$  défini comme le rapport

$$\omega_i = \frac{\pi(\mathbf{x}_i)}{q(\mathbf{x}_i)}. \quad (3.3)$$

Il suffit que  $\text{supp}(\pi) \subset \text{supp}(q)$  pour que l'estimateur  $\hat{\Pi}_N^{\text{IS}}(\psi)$  soit sans biais pour l'intégrale  $\Pi(\psi)$ , et nous avons

$$\hat{\Pi}_N^{\text{IS}}(\psi) \longrightarrow \Pi(\psi),$$

presque sûrement, quand  $N \rightarrow +\infty$ , pour toute fonction  $\psi \in \mathbb{L}^1(\Pi)$ . Lorsque la fonction  $\psi$  est telle que

$$\int \frac{\pi^2(\mathbf{x})\psi^2(\mathbf{x})}{q(\mathbf{x})} d\mathbf{x} < +\infty.$$

En appliquant le théorème central limite, nous obtenons

$$\sqrt{N} \left\{ \widehat{\Pi}_N^{\text{IS}}(\psi) - \Pi(\psi) \right\} \longrightarrow \mathcal{N}(0, \sigma_{\text{IS}}^2(\psi)), \quad (3.4)$$

en loi lorsque  $N \rightarrow +\infty$ , et la variance asymptotique  $\sigma_{\text{IS}}^2$  est donnée par

$$\sigma_{\text{IS}}^2(\psi) = \int \frac{\pi^2(\mathbf{x})\psi^2(\mathbf{x})}{q(\mathbf{x})} d\mathbf{x} - \Pi(\psi)^2. \quad (3.5)$$

**Loi de proposition optimale pour une fonction test fixée** Pour une fonction test  $\psi \in \mathbb{L}^1(\Pi)$ , il existe une loi de proposition optimale qui minimise la variance asymptotique (3.5), cette densité est proportionnelle à  $|\psi(\mathbf{x})|\pi(\mathbf{x})$ .

**Théorème 3.1.** *Robert and Casella [89]* La loi de proposition  $Q^*$  qui minimise la variance (3.5) est de densité

$$q^*(\mathbf{x}) = \frac{|\psi(\mathbf{x})|\pi(\mathbf{x})}{\int |\psi(\mathbf{y})|\pi(\mathbf{y}) d\mathbf{y}}. \quad (3.6)$$

*Démonstration.* On cherche la densité  $q(\mathbf{x})$  qui minimise

$$\int \frac{\pi^2(\mathbf{x})\psi^2(\mathbf{x})}{q(\mathbf{x})} d\mathbf{x} - \Pi(\psi)^2 \quad \text{sous la contrainte} \quad \int q(\mathbf{x}) d\mathbf{x} = 1.$$

ce qui revient à minimiser

$$\int \frac{\pi^2(\mathbf{x})\psi^2(\mathbf{x})}{q(\mathbf{x})} d\mathbf{x} \quad \text{sous la même contrainte} \quad \int q(\mathbf{x}) d\mathbf{x} = 1.$$

Pour cela nous allons utiliser la méthode des multiplicateurs de Lagrange, donc nous avons l'expression du Lagrangien

$$\int \frac{\pi(\mathbf{x})\psi^2(\mathbf{x})}{q(\mathbf{x})} \Pi(d\mathbf{x}) - \lambda \left( \int q(\mathbf{x}) d\mathbf{x} - 1 \right).$$

Nous avons donc,

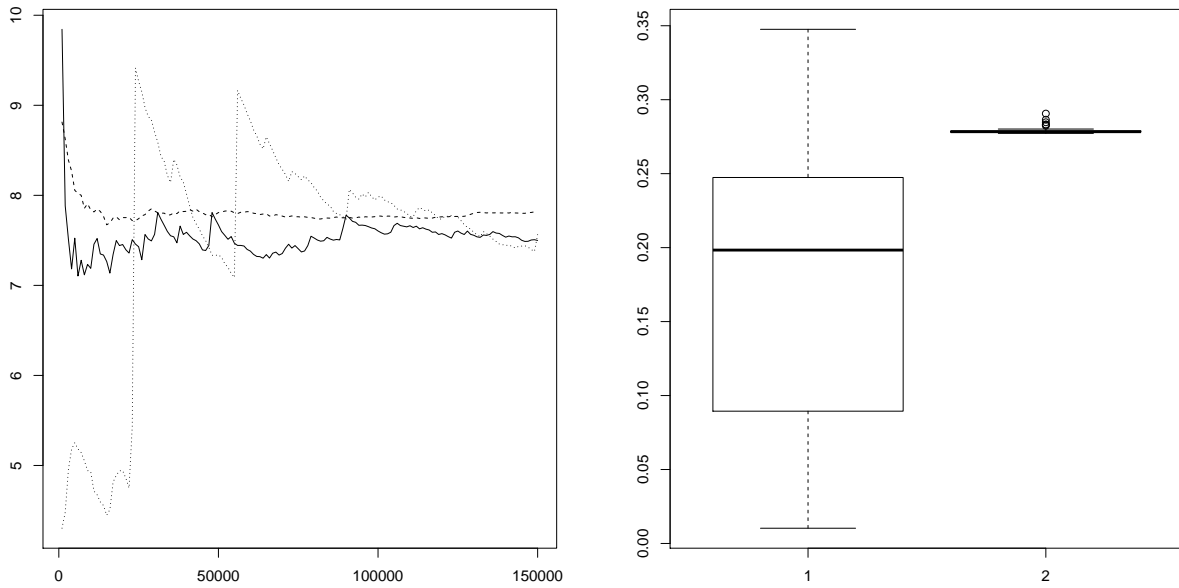
$$\frac{\psi^2(\mathbf{x})\pi^2(\mathbf{x})}{q(\mathbf{x})} - \lambda q(\mathbf{x}) = 0,$$

d'où

$$\lambda = \frac{\psi^2(\mathbf{x})\pi^2(\mathbf{x})}{q^2(\mathbf{x})}.$$

L'expression précédente montre que  $\lambda > 0$  et comme  $q(\mathbf{x})$  est une densité, nous obtenons

$$q(\mathbf{x}) = \frac{|\psi(\mathbf{x})|\pi(\mathbf{x})}{\sqrt{\lambda}}.$$



**FIGURE 3.1:** Comparaison de différents estimateurs IS construits à partir de différentes lois de propositions: Comme loi cible, une loi Student à 12 degrés de liberté. À gauche nous avons la comparaison de trois estimateurs par échantillonnage préférentiel (3.2) de  $\Pi(\psi)$  avec  $\psi(\mathbf{x}) = \mathbf{x}^5 \mathbf{1}\{\mathbf{x} > 0\}$ , chaque courbe représente une loi de proposition particulière. La courbe (en pointillés) correspond à loi  $\mathcal{N}(0, 1)$ , la courbe en (ligne discontinue) correspond à la loi de Cauchy et la courbe (ligne continue) correspond à un échantillonnage suivant la loi cible. À droite: le comportement de la variance des poids d'importance pour la loi normale en 1 et la loi de Cauchy en 2.

Sous la contrainte  $\int q(\mathbf{x}) \, d\mathbf{x} = 1$ , nous avons  $\sqrt{\lambda} = \int |\psi(\mathbf{x})| \pi(\mathbf{x}) \, d\mathbf{x}$ , d'où la densité optimale

$$q^*(\mathbf{x}) = \frac{|\psi(\mathbf{x})| \pi(\mathbf{x})}{\int |\psi(\mathbf{y})| \pi(\mathbf{y}) \, d\mathbf{y}}. \quad \square$$

## 3.2 Mesures d'efficacité d'un système de particules pondérées

Rappelons que l'échantillonnage préférentiel renvoie comme sortie, un système de particules pondérées

$$(\omega_1, \mathbf{x}_1), \dots, (\omega_N, \mathbf{x}_N) \quad (3.7)$$

pour approcher des intégrales de type

$$\int \psi(\mathbf{x}) \pi(\mathbf{x}) \, d\mathbf{x},$$

pour une large classe de fonctions test  $\psi$ . On retrouve dans Robert and Casella [89] un exemple jouet dans lequel les qualités d'estimation par échantillonnage préférentiel se

détérioreront lorsque la variance du poids d'importance (3.3) est infinie, plus précisément lorsque l'intégrale

$$\int \frac{\pi^2(\mathbf{x})}{q(\mathbf{x})} d\mathbf{x}$$

n'est pas finie.

**Effective Sample Size (ESS)** Afin de mesurer la capacité d'un système de particules pondérées à approcher une loi cible, Liu [67] introduit l'indicateur ESS (Effective Sample Size) pour mesurer la proximité entre l'échantillonnage d'une loi de proposition et l'échantillonnage de la loi cible. Pour cela Liu [67] propose l'approximation basée sur un développement de Taylor de moments de variables aléatoires suivante

$$\frac{\text{Var} [h(\mathbf{y})]}{\text{Var} [\omega\psi(\mathbf{x})]} \approx \frac{1}{1 + \text{Var}(\omega)}, \quad (3.8)$$

avec  $\mathbf{y} \sim \Pi$ ,  $\mathbf{x} \sim Q$  et  $\omega = \frac{\pi(\mathbf{x})}{q(\mathbf{x})}$ . On définit l'indicateur ESS associé au système (3.7) par

$$\text{ESS} = \frac{N}{1 + \text{Var}(\omega)}.$$

Sachant que, par la loi forte des grands nombres  $\frac{1}{N} \sum_{i=1}^N \omega_i \approx 1$  lorsque  $N$  est grand, et en remplaçant  $\text{Var}(\omega)$  par la variance empirique des poids  $\omega_1, \dots, \omega_N$ , nous avons

$$\widehat{\text{ESS}} = \frac{1}{\sum_{i=1}^N \omega_i^2}. \quad (3.9)$$

**Entropie de Shannon** Une autre mesure de qualité d'un système de particules pondérées est l'entropie de Shannon définie par

$$\text{SEnt} = - \sum_{i=1}^N \frac{\omega_i}{\sum_{j=1}^N \omega_j} \log_2 \left( \frac{\omega_i}{\sum_{j=1}^N \omega_j} \right).$$

L'entropie de Shannon peut être interprétée comme une mesure d'équilibre dans la pondération d'un système de particules. Lorsque tous les poids normalisés sont nuls sauf un seul, l'entropie de Shannon est nulle. Et lorsque tous les poids sont égaux à  $1/N$ , nous avons  $\text{SEnt}$  qui vaut  $\log_2 N$ .

### 3.3 Quelques variantes de l'échantillonnage préférentiel

#### 3.3.1 Échantillonnage préférentiel auto-normalisé

On retrouve dans Liu [68] et Robert and Casella [89] une variante de l'estimateur (3.2) initialement proposée pour le cas où la densité  $\pi$  associée à la loi cible  $\Pi$  est connue qu'à

une constante multiplicative près, ce qu'on rencontre dans les modèles bayésiens où la loi cible est la loi *a posteriori*. Cet estimateur s'obtient en normalisant le poids d'importance (3.3), d'où l'estimateur par échantillonnage préférentiel auto-normalisé donné par

$$\tilde{\Pi}_N^{\text{IS}}(\psi) = \frac{\sum_{i=1}^N \omega_i \psi(\mathbf{x}_i)}{\sum_{j=1}^N \omega_j}. \quad (3.10)$$

Lorsque la densité  $\pi$  n'est connue qu'à une constante près, *i.e*  $\int \pi(\mathbf{x}) d\mathbf{x} = c \neq 1$ , nous avons

$$\frac{1}{N} \sum_{i=1}^N \omega_i \rightarrow c$$

presque sûrement quand  $N \rightarrow +\infty$ . L'estimateur (3.10) est biaisé, cependant nous avons la convergence

$$\tilde{\Pi}_N^{\text{IS}}(\psi) \rightarrow \Pi(\psi)$$

presque sûrement quand  $N \rightarrow +\infty$ , et par le théorème de Slutsky, nous avons

$$\sqrt{N} \left\{ \tilde{\Pi}_N^{\text{IS}}(\psi) - \Pi(\psi) \right\} \rightarrow \mathcal{N}(0, \sigma_{\text{IS}}^2(\psi)),$$

en loi lorsque  $N \rightarrow +\infty$ . Casella and Robert [17] montrent qu'en général, l'estimateur (3.10) peut constituer une amélioration de l'estimateur (3.2) en terme d'erreur quadratique moyenne. En appliquant un développement de Taylor sur  $\mathbb{E} \left[ \tilde{\Pi}_N^{\text{IS}}(\psi) \right]$  et  $\text{Var} \left[ \tilde{\Pi}_N^{\text{IS}}(\psi) \right]$ , nous avons les approximations suivantes:

$$\mathbb{E} \left[ \tilde{\Pi}_N^{\text{IS}}(\psi) \right] \approx \Pi(\psi) - \frac{\text{Cov}(\omega\psi(\mathbf{x}), \omega)}{N} + \frac{\Pi(\psi) \text{Var}(\omega)}{N}, \quad (3.11)$$

et

$$\text{Var} \left[ \tilde{\Pi}_N^{\text{IS}}(\psi) \right] \approx \frac{1}{N} \left\{ \Pi(\psi)^2 \text{Var}(\omega) + \text{Var}[\omega\psi(\mathbf{x})] - 2\Pi(\psi) \text{Cov}[\omega, \omega\psi(\mathbf{x})] \right\}, \quad (3.12)$$

avec  $\mathbf{x} \sim Q$ ,  $\omega = \frac{\pi(\mathbf{x})}{q(\mathbf{x})}$  et  $\int \frac{\pi^2(\mathbf{x})\psi^2(\mathbf{x})}{q(\mathbf{x})} d\mathbf{x} < +\infty$ . Nous avons

$$\text{MSE} \left[ \hat{\Pi}_N^{\text{IS}}(\psi) \right] = \frac{\text{Var}[\omega\psi(\mathbf{x})]}{N},$$

et puisque l'estimateur  $\tilde{\Pi}_N^{\text{IS}}(\psi)$  est biaisé, en substituant les approximations (3.11) et (3.12) dans  $\text{MSE} \left[ \tilde{\Pi}_N^{\text{IS}}(\psi) \right]$ , nous avons

$$\begin{aligned} \text{MSE} \left[ \tilde{\Pi}_N^{\text{IS}}(\psi) \right] &= \left\{ \mathbb{E} \left[ \tilde{\Pi}_N^{\text{IS}}(\psi) \right] - \Pi(\psi) \right\}^2 + \text{Var} \left[ \tilde{\Pi}_N^{\text{IS}}(\psi) \right] \\ &\approx \left\{ \frac{\Pi(\psi) \text{Var}(\omega)}{N} - \frac{\text{Cov}[\omega\psi(\mathbf{x}), \omega]}{N} \right\}^2 \\ &\quad + \frac{1}{N} \left\{ \Pi(\psi)^2 \text{Var}(\omega) + \text{Var}[\omega\psi(\mathbf{x})] - 2\Pi(\psi) \text{Cov}[\omega, \omega\psi(\mathbf{x})] \right\} \\ &\approx \frac{\text{MSE} \left[ \hat{\Pi}_N^{\text{IS}}(\psi) \right]}{N} + \frac{1}{N} \left\{ \Pi(\psi)^2 \text{Var}(\omega) - 2\Pi(\psi) \text{Cov}[\omega, \omega\psi(\mathbf{x})] \right\} + \mathcal{O} \left( \frac{1}{N^2} \right). \end{aligned}$$



On déduit de l'approximation du MSE  $\left[ \widehat{\Pi}_N^{IS}(\psi) \right]$ , qu'asymptotiquement

$$\text{MSE} \left[ \widetilde{\Pi}_N^{IS}(\psi) \right] \leq \text{MSE} \left[ \widehat{\Pi}_N^{IS}(\psi) \right].$$

### 3.3.2 Échantillonnage préférentiel avec variables de contrôle

On peut améliorer l'estimateur (3.2) en terme de variance en exploitant la connaissance de l'intégrale

$$\mathbb{E}(\omega) = \int \pi(\mathbf{x}) \, d\mathbf{x} = 1.$$

Cela revient à introduire  $\omega_1, \dots, \omega_N$  comme variables de contrôle dans (3.2), et donc considérer l'estimateur par échantillonnage préférentiel avec variables de contrôles de (3.1)

$$\widehat{\Pi}_{IS}^\beta(\psi) = \frac{1}{N} \sum_{i=1}^N \{\omega_i \psi(\mathbf{x}_i) - \beta \omega_i\} + \beta. \quad (3.13)$$

L'estimateur (3.13) est sans biais et la proposition suivante montre que qu'il est une amélioration en terme de variance de l'estimateur (3.2).

**Proposition 3.2.** *En posant  $\beta = \frac{\text{Cov}[\omega\psi(\mathbf{x}), \omega]}{\text{Var}(\omega)}$ , nous avons*

$$\text{Var} \left[ \widehat{\Pi}_{IS}^\beta(\psi) \right] \leq \text{Var} \left[ \widehat{\Pi}_N^{IS}(\psi) \right]. \quad (3.14)$$

*Démonstration.* Nous avons

$$\begin{aligned} \text{Var} \left[ \widehat{\Pi}_{IS}^\beta(\psi) \right] &= \frac{1}{N^2} \sum_{i=1}^N \left\{ \text{Var}[\omega_i \psi(\mathbf{x}_i)] + \beta^2 \text{Var}(\omega_i) - 2\beta \text{Cov}[\omega_i, \omega_i \psi(\mathbf{x}_i)] \right\} \\ &= \text{Var} \left[ \widehat{\Pi}_N^{IS}(\psi) \right] + \beta^2 \frac{\text{Var}(\omega)}{N} - 2\beta \frac{\text{Cov}[\omega, \omega\psi(\mathbf{x})]}{N} \\ &= \{1 - \rho^2(\omega\psi(\mathbf{x}), \omega)\} \text{Var} \left[ \widehat{\Pi}_N^{IS}(\psi) \right] \\ &\leq \text{Var} \left[ \widehat{\Pi}_N^{IS}(\psi) \right], \end{aligned}$$

avec  $\rho(\omega\psi(\mathbf{x}), \omega)$  la corrélation entre  $\omega\psi(\mathbf{x})$  et  $\omega$ . □

En général, le calcul de  $\text{Cov}[\omega\psi(\mathbf{x}), \omega]$  et  $\text{Var}(\omega)$  n'est pas accessible. En pratique la valeur de  $\beta$  est estimée comme le coefficient de régression des variables  $\omega_1\psi(\mathbf{x}_1), \dots, \omega_N\psi(\mathbf{x}_N)$  comme variables à expliquer et des  $\omega_1, \dots, \omega_N$  comme variables explicatives. Owen and Zhou [83] généralisent la technique des variables de contrôle à l'échantillonnage préférentiel multiple et montrent que la consistance de l'estimation de  $\beta$  par régression.

# 4

## Théorèmes limites sur les tableaux triangulaires

Nous présentons dans ce chapitre deux résultats asymptotiques sur les tableaux triangulaires auxquels nous faisons référence dans cette thèse. Ce sont des généralisations de théorèmes énoncés dans le cas unidimensionnel par [15] (Proposition 9.5.7 et 9.5.12). On supposera que toutes les variables aléatoires sont définies sur un espace de probabilité commun  $(\Xi, \mathcal{F}, \mathbb{P})$ .

### 4.1 Loi des grands nombres

Commençons par rappeler la définition suivante.

**Définition 4.1** (Tension). Une famille de vecteurs aléatoires  $(U_N)$  est dite tendue si

$$\lim_{C \rightarrow \infty} \sup_N \mathbb{P} (\|U_N\| \geq C) = 0.$$

Soit  $\{(\mathbf{V}_{t,i})_{1 \leq i \leq N_t}, t \geq 1\}$  un tableau triangulaire de vecteurs aléatoires de  $\mathbb{R}^d$ , et  $(\mathcal{F}_t)_{t \geq 1}$  une suite de tribus. Énonçons une loi faible des grands nombres, et un théorème central limite.

**Théorème 4.1.** *Supposons que les conditions suivantes soient vérifiées.*

- (i) *Conditionnellement à la tribu  $\mathcal{F}_t$ , les vecteurs aléatoires  $\mathbf{V}_{t,1}, \dots, \mathbf{V}_{t,N_t}$  sont indépendants et  $\mathbb{E} \left[ \|\mathbf{V}_{t,i}\| \middle| \mathcal{F}_t \right] < +\infty$  pour tout  $t, i$ .*
- (ii) *La suite  $\left( \sum_{i=1}^{N_t} \mathbb{E} \left[ \|\mathbf{V}_{t,i}\| \middle| \mathcal{F}_t \right] \right)_{t \geq 1}$  est tendue.*
- (iii) *Pour tout  $\eta$  positif,*

$$\sum_{i=1}^{N_t} \mathbb{E} \left[ \|\mathbf{V}_{t,i}\| \mathbf{1}_{\{\|\mathbf{V}_{t,i}\| > \eta\}} \middle| \mathcal{F}_t \right] \rightarrow 0, \quad \text{en probabilité.}$$

Alors,

$$\sum_{i=1}^{N_t} \left\{ \mathbf{v}_{t,i} - \mathbb{E}[\mathbf{v}_{t,i} | \mathcal{F}_t] \right\} \longrightarrow 0, \quad \text{en probabilité.}$$

## 4.2 Théorème central limite

**Théorème 4.2.** *Supposons que les conditions suivantes soient vérifiées.*

- (i) *Conditionnellement à la tribu  $\mathcal{F}_t$ , les vecteurs aléatoires  $\mathbf{V}_{t,1}, \dots, \mathbf{V}_{t,N_t}$  sont indépendants.*
- (ii) *Pour tout  $t \geq 1$  et  $i = 1, \dots, N_t$ ,  $\mathbb{E}[\|\mathbf{V}_{t,i}\|^2 | \mathcal{F}_t] < +\infty$ .*
- (iii) *Pour tout  $\mathbf{b} \in \mathbb{R}^d$ , il existe une constante  $\sigma_{\mathbf{b}}^2 > 0$  telle que*

$$\sum_{i=1}^{N_t} \left\{ \mathbb{E}[\langle \mathbf{b}, \mathbf{V}_{t,i} \rangle^2 | \mathcal{F}_t] - \left( \mathbb{E}[\langle \mathbf{b}, \mathbf{V}_{t,i} \rangle | \mathcal{F}_t] \right)^2 \right\} \longrightarrow \sigma_{\mathbf{b}}^2, \quad \text{en probabilité.}$$

- (iv) *Pour tout vecteur  $\mathbf{e}$  de la base canonique de  $\mathbb{R}^d$  et pour tout  $\eta > 0$*

$$\sum_{i=1}^{N_t} \mathbb{E} \left[ \langle \mathbf{e}, \mathbf{V}_{t,i} \rangle^2 \mathbf{1}_{\left\{ |\langle \mathbf{e}, \mathbf{V}_{t,i} \rangle| > \eta \right\}} \middle| \mathcal{F}_t \right] \longrightarrow 0, \quad \text{en probabilité.}$$

Alors il existe une matrice  $\Sigma$  caractérisée par  $\mathbf{b}^t \Sigma \mathbf{b} = \sigma_{\mathbf{b}}^2$  pour tout  $\mathbf{b} \in \mathbb{R}^d$  telle que la suite

$$\sum_{j=1}^{N_t} \left\{ \mathbf{v}_{t,j} - \mathbb{E}[\mathbf{v}_{t,j} | \mathcal{F}_t] \right\}$$

converge en loi vers  $\mathcal{N}_d(0, \Sigma)$  quand  $t \rightarrow +\infty$ .

*Démonstration.* Pour montrer que la suite

$$\sum_{j=1}^{N_t} \left\{ \mathbf{v}_{t,j} - \mathbb{E}[\mathbf{v}_{t,j} | \mathcal{F}_t] \right\}$$

converge en loi vers une gaussienne  $\mathcal{N}_d(0, \Sigma)$ , il suffit de montrer que pour tout  $\mathbf{b} \in \mathbb{R}^d$ , il existe une constante  $\sigma_{\mathbf{b}}^2 > 0$  telle que

$$\left\langle \mathbf{b}, \sum_{j=1}^{N_t} \left\{ \mathbf{v}_{t,j} - \mathbb{E}[\mathbf{v}_{t,j} | \mathcal{F}_t] \right\} \right\rangle \longrightarrow \mathcal{N}(0, \sigma_{\mathbf{b}}^2), \quad (4.1)$$

en loi quand  $t$  tend vers l'infini. Pour montrer (4.1) pour tout  $\mathbf{b} \in \mathbb{R}^d$ , nous allons appliquer la proposition 9.5.12 de [15] sur le tableau triangulaire

$$\left\{ \left( \langle \mathbf{b}, \mathbf{V}_{t,i} \rangle \right)_{1 \leq i \leq N_t}, t \geq 1 \right\}. \quad (4.2)$$

Pour cela, vérifions les conditions (i),(ii), (iii) et (iv) de la proposition 9.5.12 de [15].

1. Pour tout  $\mathbf{b} \in \mathbb{R}^d$ , conditionnellement à la tribu  $\mathcal{F}_t$ , l'indépendance des variables aléatoires

$$\langle \mathbf{b}, \mathbf{V}_{t,1} \rangle, \dots, \langle \mathbf{b}, \mathbf{V}_{t,N_t} \rangle$$

est immédiate par hypothèse.

2. Pour tout  $\mathbf{b} \in \mathbb{R}^d$ , pour  $t \geq 1$  et  $i = 1, \dots, N_t$ , comme conséquence de la condition (ii) du théorème 4.2, nous avons

$$\begin{aligned} \mathbb{E} \left[ \langle \mathbf{b}, \mathbf{V}_{t,i} \rangle^2 \middle| \mathcal{F}_t \right] &= \sum_{k=1}^d \mathbf{b}_k^2 \mathbb{E} \left[ \left( \mathbf{V}_{t,i}^k \right)^2 \middle| \mathcal{F}_t \right] + 2 \sum_{1 \leq k < \ell \leq d} \mathbf{b}_k \mathbf{b}_\ell \mathbb{E} \left[ \mathbf{V}_{t,i}^k \mathbf{V}_{t,i}^\ell \middle| \mathcal{F}_t \right] \\ &< +\infty. \end{aligned}$$

3. Pour tout  $\mathbf{b} \in \mathbb{R}^d$ , la condition (iii) du théorème 4.2 revient à supposer qu'il existe une constante  $\sigma_{\mathbf{b}}^2 > 0$  telle que

$$\sum_{i=1}^{N_t} \left\{ \mathbb{E} \left[ \langle \mathbf{b}, \mathbf{V}_{t,i} \rangle^2 \middle| \mathcal{F}_t \right] - \left( \mathbb{E} \left[ \langle \mathbf{b}, \mathbf{V}_{t,i} \rangle \middle| \mathcal{F}_t \right] \right)^2 \right\} \longrightarrow \sigma_{\mathbf{b}}^2,$$

en probabilité.

4. Sur le tableau triangulaire (4.2) pour tout  $\mathbf{b} \in \mathbb{R}^d$ , posons

$$A_t(\mathbf{b}, \eta) = \sum_{i=1}^{N_t} \mathbb{E} \left[ \langle \mathbf{b}, \mathbf{V}_{t,i} \rangle^2 \mathbf{1} \left\{ \left| \langle \mathbf{b}, \mathbf{V}_{t,i} \rangle \right| > \eta \right\} \middle| \mathcal{F}_t \right],$$

et montrons que l'ensemble

$$A = \left\{ \mathbf{b} \in \mathbb{R}^d, \quad \forall \eta > 0 \quad A_t(\mathbf{b}, \eta) = o_p(1) \right\},$$

est un sous espace linéaire de  $\mathbb{R}^d$ .

Nous avons clairement  $0 \in A$ , et pour tout scalaire  $\alpha \neq 0$ ,

$$A_t(\alpha \mathbf{b}, \eta) = \alpha^2 A_t \left( \mathbf{b}, \frac{\eta}{|\alpha|} \right),$$

et par conséquence  $\alpha \mathbf{b} \in A$  si  $\mathbf{b} \in A$ . Nous avons aussi

$$\left| \langle \mathbf{u} + \mathbf{v}, \mathbf{V}_{t,i} \rangle \right| \leq \left| \langle \mathbf{u}, \mathbf{V}_{t,i} \rangle \right| + \left| \langle \mathbf{v}, \mathbf{V}_{t,i} \rangle \right| \leq 2 \max \left\{ \left| \langle \mathbf{u}, \mathbf{V}_{t,i} \rangle \right|, \left| \langle \mathbf{v}, \mathbf{V}_{t,i} \rangle \right| \right\},$$

et puisque

$$\max(\mathbf{x}, \mathbf{y})^2 \mathbf{1} \left\{ \max(\mathbf{x}, \mathbf{y}) > \eta \right\} \leq \mathbf{x}^2 \mathbf{1} \left\{ \mathbf{x} > \eta \right\} + \mathbf{y}^2 \mathbf{1} \left\{ \mathbf{y} > \eta \right\},$$

on déduit

$$A_t(\mathbf{u} + \mathbf{v}, \eta) \leq 4A_t \left( \mathbf{u}, \frac{\eta}{2} \right) + 4A_t \left( \mathbf{v}, \frac{\eta}{2} \right).$$

d'où  $u + v \in A$  si  $u \in A$  et  $v \in A$ , et donc l'ensemble  $A$  est un sous espace linéaire. Maintenant, pour montrer que  $A = \mathbb{R}^d$ , il nous reste qu'à vérifier que l'ensemble  $A$  contient une base de  $\mathbb{R}^d$ , pour cela soit  $\{e_1, \dots, e_d\}$  la base canonique de  $\mathbb{R}^d$ , la condition (iv) du théorème 4.2 montre que pour tout  $k = 1, \dots, d$  et pour tout  $\eta > 0$

$$\begin{aligned} A_t(e_k, \eta) &= \sum_{i=1}^{N_t} \mathbb{E} \left[ \langle e_k, \mathbf{V}_{t,i} \rangle^2 \mathbf{1}_{\{|\langle e_k, \mathbf{V}_{t,i} \rangle| > \eta\}} \mid \mathcal{F}_t \right] \\ &= \sum_{i=1}^{N_t} \mathbb{E} \left[ (\mathbf{V}_{t,i}^k)^2 \mathbf{1}_{\{|\mathbf{V}_{t,i}^k| > \eta\}} \mid \mathcal{F}_t \right] \\ &\rightarrow 0, \end{aligned}$$

en probabilité.

Comme conséquence de la proposition 9.5.12 de [15], nous avons, pour tout  $\mathbf{b} \in \mathbb{R}^d$ , il existe une constante  $\sigma_{\mathbf{b}}^2 > 0$  telle que pour tout  $u \in \mathbb{R}$

$$\mathbb{E} \left[ \exp \left\{ iu \left\langle \mathbf{b}, \sum_{j=1}^{N_t} (\mathbf{V}_{t,j} - \mathbb{E}[\mathbf{V}_{t,j} \mid \mathcal{F}_t]) \right\rangle \right\} \mid \mathcal{F}_t \right] \rightarrow \exp \left( -\frac{u^2 \sigma_{\mathbf{b}}^2}{2} \right),$$

en probabilité, et donc par le théorème de convergence dominée, nous avons

$$\mathbb{E} \left[ \exp \left\{ iu \left\langle \mathbf{b}, \sum_{j=1}^{N_t} (\mathbf{V}_{t,j} - \mathbb{E}[\mathbf{V}_{t,j} \mid \mathcal{F}_t]) \right\rangle \right\} \right] \rightarrow \exp \left( -\frac{u^2 \sigma_{\mathbf{b}}^2}{2} \right).$$

On déduit que pour tout  $\mathbf{b} \in \mathbb{R}^d$

$$\left\langle \mathbf{b}, \sum_{j=1}^{N_t} (\mathbf{V}_{t,j} - \mathbb{E}[\mathbf{V}_{t,j} \mid \mathcal{F}_t]) \right\rangle \rightarrow \mathcal{N}(0, \sigma_{\mathbf{b}}^2),$$

en loi quand  $t \rightarrow +\infty$  d'où le théorème centrale limite

$$\sum_{j=1}^{N_t} \{(\mathbf{V}_{t,j} - \mathbb{E}[\mathbf{V}_{t,j} \mid \mathcal{F}_t])\} \rightarrow \mathcal{N}_d(0, \boldsymbol{\Sigma}),$$

avec la matrice  $\boldsymbol{\Sigma}$  donnée par la connaissance de  $\mathbf{b}^t \boldsymbol{\Sigma} \mathbf{b} = \sigma_{\mathbf{b}}^2$  pour tout  $\mathbf{b} \in \mathbb{R}^d$ . □

# 5

## Échantillonnage Préférentiel Adaptatif et Multiple

### 5.1 Introduction et motivation

Les méthodes de Monte-Carlo et Monte-Carlo par chaîne de Markov proposées dans Robert and Casella [89] consistent à construire un échantillon pondéré pour approcher une loi cible  $\Pi$  de densité  $\pi$ . Autrement dit, obtenir un échantillon  $\mathbf{x}_\ell$  d'indices  $\ell = 1, \dots, n$  de poids respectifs  $\omega_\ell$  dont la mesure empirique associée

$$\chi_n = \sum_{\ell=1}^n \omega_\ell \delta_{\mathbf{x}_\ell}$$

approche la loi  $\Pi$  dans le sens, où, pour une large classe de fonctions  $\psi$ , l'intégrale empirique

$$\int \psi(\mathbf{x}) \chi_n(\mathbf{x}) = \sum_{\ell=1}^n \omega_\ell \psi(\mathbf{x}_\ell) \quad (5.1)$$

$$\longrightarrow \int \psi(\mathbf{x}) \Pi(\mathbf{x}), \quad (5.2)$$

quand  $n \rightarrow +\infty$ .

Parmi les méthodes Monte-Carlo, l'échantillonnage préférentiel (Importance Sampling) popularisé dans Ripley [88] et présenté au chapitre 3 revient à échantillonner  $\mathbf{x}_\ell$  à partir d'une loi de proposition  $Q$  de densité  $q$  et considérer le poids  $\omega_\ell = \pi(\mathbf{x}_\ell)/q(\mathbf{x}_\ell)$  appelé poids d'importance. La qualité des approximations obtenues par échantillonnage préférentiel est insatisfaisante lorsque la variance des poids d'importance  $\omega_\ell$  est élevée. Ce qui est sensiblement lié au choix de la loi de proposition  $Q$  et à la dimension de  $\mathbf{x}_\ell$ . La calibration de la loi de proposition  $Q$  sur la loi cible  $\Pi$  demeure difficile en pratique. Pour construire séquentiellement une loi de proposition adaptée à la cible, on retrouve un grand nombre de techniques itératives dans la littérature [89, 69, 85, 93]. Parmi les techniques

adaptatives les plus récentes, on s'intéresse particulièrement à celle proposée dans Cappé et al. [14], où il est proposé de se placer dans une famille paramétrique de lois

$$\left\{ Q(\theta), \theta \in \Theta \right\} \quad \text{de densités} \quad \left\{ q(\mathbf{x}, \theta), \theta \in \Theta \right\}.$$

Ce qui simplifie la construction d'une loi de proposition en un choix de paramètre adéquat  $\theta^*$ . Autrement dit, à partir d'un paramètre  $\hat{\theta}_0$  fixé arbitrairement, il est construit une suite de paramètres  $\hat{\theta}_1, \dots, \hat{\theta}_T$  de manière itérative. Pour cela, une suite d'échantillons

$$\mathbf{x}_1^1, \dots, \mathbf{x}_{N_1}^1, \dots, \dots, \mathbf{x}_1^T, \dots, \mathbf{x}_{N_T}^T,$$

simulés suivant les lois

$$Q(\hat{\theta}_1), \dots, Q(\hat{\theta}_T),$$

respectivement est utilisée. Une itération de la stratégie adaptative de [14] est donnée par les deux étapes suivantes :

- 1) Simulation des particules  $\mathbf{x}_1^t, \dots, \mathbf{x}_{N_t}^t$  indépendantes suivant la loi  $Q(\hat{\theta}_t)$ .
- 2) Calibration du paramètre  $\hat{\theta}_{t+1}$  sur l'échantillon  $\mathbf{x}_1^t, \dots, \mathbf{x}_{N_t}^t$  en minimisant un critère donné.

Cette description générique de l'itération  $t$  montre que l'adaptation du paramètre  $\hat{\theta}_{t+1}$  ne se fait que sur les particules simulées au temps  $t$ . Et la calibration du nouveau paramètre à chaque itération repose sur la minimisation d'un critère, par exemple la variance d'un estimateur ou la divergence de Kullback-Leibler entre la loi  $Q(\theta)$  et la cible  $\Pi$ . L'objectif des techniques adaptatives est la construction d'une loi de proposition adéquate  $Q(\theta^*)$ , et l'utilisation de cette loi pour faire de l'échantillonnage préférentiel pour l'estimation dans le problème de départ de calcul de l'intégrale

$$\Pi(\psi) = \int \psi(\mathbf{x}) \Pi(d\mathbf{x}). \quad (5.3)$$

Toutefois, aucune procédure de recyclage des particules employées au cours des itérations du processus adaptatif n'est proposée dans Cappé et al. [14]. Autrement dit, le processus adaptatif est entièrement destiné à la construction du paramètre  $\hat{\theta}_T$  et l'estimation de l'intégrale (5.3) est donnée dans Cappé et al. [14] par l'estimateur

$$\hat{\Pi}_{\text{IS}}(\psi) = \frac{1}{N} \sum_{i=1}^N \frac{\pi(\mathbf{x}_i)}{q(\mathbf{x}_i, \hat{\theta}_T)} \psi(\mathbf{x}_i)$$

avec  $\mathbf{x}_1, \dots, \mathbf{x}_N$  des particules iid issues de loi d'importance  $Q(\hat{\theta}_T)$ .

Sur des problèmes d'inférence en génétique des populations où les coûts de calcul sont élevés, le recyclage des simulations obtenues au cours du processus d'apprentissage est indispensable. Pour cela, Cornuet et al. [21] proposent l'algorithme d'échantillonnage préférentiel adaptatif et multiple, appelé AMIS (Adaptive Multiple Importance Sampling) et basé sur les méthodes d'échantillonnage préférentiel multiple et les techniques adaptatives. Dans l'algorithme AMIS, on retrouve les stratégies adaptatives de Cappé et al. [14] consacrées à la construction d'une suite de lois

$$Q(\hat{\theta}_1), \dots, Q(\hat{\theta}_T),$$

à partir d'une famille paramétrique donnée

$$\{Q(\theta), \theta \in \Theta\}.$$

La contribution majeure de l'algorithme AMIS de Cornuet et al. [21] est l'introduction d'un processus de recyclage de tous les échantillons passés pour construire le nouveau paramètre  $\hat{\theta}_t$  à l'itération  $t$ . À la  $t^{\text{ème}}$  itération, la  $i^{\text{ème}}$  particule qui avait été simulée pendant l'itération  $k$  ( $k \leq t$ ), notée  $\mathbf{x}_i^k$ , est utilisée pour calculer  $\hat{\theta}_t$  avec le poids d'importance

$$\frac{\pi(\mathbf{x}_i^k)}{\sum_{\ell=1}^{t-1} \frac{N_\ell}{\Omega_{t-1}} q(\mathbf{x}_i^k, \hat{\theta}_\ell)}. \quad (5.4)$$

où  $\hat{\theta}_1, \dots, \hat{\theta}_{t-1}$  sont les paramètres obtenus au cours des  $t-1$  itérations passées et  $\Omega_{t-1} = N_1 + N_2 + \dots + N_{t-1}$ . Le choix du poids d'importance (5.4) pour chaque particule est inspiré des techniques proposées et étudiées par Veach and Guibas [101], Owen and Zhou [83]. Plus précisément, l'expression du dénominateur dans le poids d'importance (5.4) correspond à l'heuristique de la balance de Veach and Guibas [101]. Lorsque les  $\theta_i$  ne sont pas calibrés en ligne, mais fixés *a priori*, Veach and Guibas [101] montrent que cette heuristique de la balance est consistante en terme de variance parmi une classe de stratégies de pondération sans biais. Toujours dans le cadre où les lois d'importance sont fixées *a priori*, Owen and Zhou [83] montrent que les estimateurs obtenus par cette heuristique héritent des qualités qu'on aurait obtenues en échantillonnant seulement avec la meilleure loi de proposition parmi celles utilisées.

Sur diverses expérimentations numériques, Cornuet et al. [21] insistent sur la stabilité de l'algorithme AMIS et le gain en terme de ESS (Effective Sample Size). Le recyclage de tous les échantillons et leur combinaison par l'heuristique de la balance s'accommodent particulièrement au cadre bayésien sur des modèles de génétique des populations. Dans ce cas, la cible  $\Pi$  est la loi *a posteriori* dont l'évaluation est particulièrement coûteuse en temps de calcul. Ceci a incité Sirén et al. [94] à appliquer l'algorithme AMIS sur des problèmes d'inférence d'histoire des populations à partir de données SNP (Single-Nucleotide Polymorphism).

Cornuet et al. [21] affirment qu'aucun résultat de convergence sur l'estimateur AMIS ne peut être obtenu simplement sans se restreindre au cas trivial où les densités  $\pi$  et  $\{q(\mathbf{x}, \theta), \theta \in \Theta\}$  sont à support compact et bornées supérieurement. La complexité de la structure de dépendance, conséquence de la stratégie de recyclage dans le schéma adaptatif, ne permet pas d'utiliser les outils classiques (chaînes de Markov, martingales, ...).

Comme contribution à ce chapitre, nous proposons de montrer la convergence de l'estimateur AMIS sous une légère simplification du schéma adaptatif et une hypothèse de croissance de la suite  $(N_t)_{t \geq 1}$ . Autrement dit, nous renonçons au recyclage, c'est-à-dire à la combinaison en ligne des  $t-1$  échantillons passés dans les étapes de calibration de  $\theta$  et nous nous contentons de calculer  $\hat{\theta}_t$  sur l'échantillon  $t-1$ . Mais, après cette calibration des différentes lois de proposition

$$Q(\hat{\theta}_1), \dots, Q(\hat{\theta}_T)$$



par la donnée de la suite des paramètres

$$\hat{\theta}_1, \dots, \hat{\theta}_T,$$

on combine la totalité des  $T$  échantillons pour former un système final de particules pondérées de taille  $\Omega_T = N_1 + N_2 + \dots + N_T$ . Ainsi toute particule  $\mathbf{x}_i^t$  est pondérée par

$$\omega_i^t = \frac{\pi(\mathbf{x}_i^t)}{\sum_{k=1}^T \frac{N_k}{\Omega_T} q(\mathbf{x}_i^t, \hat{\theta}_k)}. \quad (5.5)$$

La première partie de ce chapitre résume les techniques de combinaison d'échantillons introduites dans Veach and Guibas [101]. Principalement, elle décrit la stratégie de recyclage employée dans le schéma adaptatif de l'algorithme AMIS et appelée heuristique de la balance par Veach and Guibas [101]. Nous rappellerons les différentes conclusions de Veach and Guibas [101] basées sur des simulations numériques dans lesquelles l'utilisation de l'heuristique de la balance est recommandée. À la fin de cette première partie, nous introduisons les techniques d'échantillonnage préférentiel multiple par variables de contrôle de Owen and Zhou [83] ainsi que leurs résultats d'optimalité sur l'heuristique de la balance.

La deuxième partie est consacrée à la présentation de l'algorithme AMIS de Cornuet et al. [21] ainsi que le fonctionnement de son schéma adaptatif avec la procédure de recyclage, et nous expliquons les difficultés liées à l'étude de la convergence de l'estimateur AMIS.

L'algorithme AMIS avec schéma adaptatif simplifié est décrit à la troisième partie ainsi que notre contribution principale à ce chapitre dans laquelle nous énonçons un résultat de convergence de l'estimateur AMIS sur le schéma adaptatif modifié.

En dernière partie nous montrons sur des simulations numériques effectuées sur une cible jouet et une cible issue d'un modèle de génétique des populations, que les performances numérique de l'algorithme AMIS avec schéma adaptatif simplifié sont comparables à celles de la version originale de Cornuet et al. [21].

## 5.2 État de l'art des méthodes d'échantillonnage préférentiel multiple

Nous avons mentionné dans l'introduction les avantages de l'algorithme AMIS de Cornuet et al. [21] dus à la stratégie de recyclage de tout le passé des simulations dans son schéma adaptatif à chaque itération. Cette technique de recyclage est représentée par le système de poids (5.4) qui correspond à une stratégie de combinaison d'échantillons. Nous consacrons cette partie à la construction des techniques de combinaison d'échantillons issus de différentes lois de proposition. Nous commençons par rappeler les différentes techniques d'échantillonnage préférentiel multiple introduites et étudiées dans

Veach and Guibas [101]. Nous montrons que le choix du système de poids de l'algorithme AMIS correspond à l'heuristique de la balance. Nous justifions le choix d'usage de l'heuristique de la balance dans Cornuet et al. [21] par le résultat d'optimalité en terme de variance obtenu dans Veach and Guibas [101]. Nous terminons par le rappel de la classe des estimateurs par échantillonnage préférentiel multiple par variables de contrôle introduites dans Owen and Zhou [83] ainsi qu'un résultat d'optimalité de l'heuristique de la balance sous forme d'une borne supérieure sur la variance de l'estimateur par échantillonnage préférentiel multiple avec variables de contrôle.

### 5.2.1 Stratégies de combinaison d'échantillons

À partir d'une collection de lois de proposition

$$Q_1, \dots, Q_T,$$

fixées *a priori* de densités  $q_1(\mathbf{x}), \dots, q_T(\mathbf{x})$ , les méthodes d'échantillonnage préférentiel multiple consistent à combiner les différents échantillons

$$\begin{aligned} \mathbf{x}_1^1, \mathbf{x}_2^1, \dots, \mathbf{x}_{N_1}^1 &\sim Q_1 \\ \mathbf{x}_1^2, \mathbf{x}_2^2, \dots, \mathbf{x}_{N_2}^2 &\sim Q_2 \\ &\vdots \\ \mathbf{x}_1^T, \mathbf{x}_2^T, \dots, \mathbf{x}_{N_T}^T &\sim Q_T \end{aligned}$$

issus de ces différentes lois de propositions pour construire un estimateur de l'intégrale (5.3). Plus précisément, combiner ces différents échantillons revient à choisir une stratégie efficace pour pondérer chaque particule  $\mathbf{x}_i^t$  ( $i \leq N_t, t \leq T$ ) par un poids d'importance  $\omega_i^t$  et considérer l'estimateur

$$\hat{\Pi}(\psi) = \frac{1}{\Omega_T} \sum_{t=1}^T \sum_{i=1}^{N_t} \omega_i^t \psi(\mathbf{x}_i^t), \quad (5.6)$$

avec  $\Omega_T = N_1 + N_2 + \dots + N_T$ . Une stratégie basique de combinaison de ces différents échantillons pour estimer l'intégrale (5.3) revient à considérer l'estimateur

$$\hat{\Pi}_{\text{IS}}(\psi) = \frac{1}{\Omega_T} \sum_{t=1}^T \sum_{i=1}^{N_t} \frac{\pi(\mathbf{x}_i^t)}{q_t(\mathbf{x}_i^t)} \psi(\mathbf{x}_i^t). \quad (5.7)$$

Cet estimateur est sans biais pour (5.3) et revient à fusionner les différents échantillons en conservant le poids d'importance obtenu en faisant de l'échantillonnage préférentiel avec chaque loi de proposition  $Q_t$  individuellement. Toutefois, l'existence d'une loi  $Q_t$  parmi

$$Q_1, \dots, Q_T$$

telle que

$$\int \frac{\pi^2(\mathbf{x})}{q_t(\mathbf{x})} d\mathbf{x} = +\infty,$$

détériorer violemment l'estimateur (5.7), ce qui est contre-intuitif à l'idée d'utiliser plusieurs lois de proposition. Veach and Guibas [101] proposent une écriture généralisée de l'estimateur par échantillonnage préférentiel multiple (MIS). Ils introduisent une collection de fonctions poids

$$w_1(\mathbf{x}), \dots, w_T(\mathbf{x})$$

pour paramétrer une stratégie de combinaison des différents échantillons

$$\mathbf{x}_1^1, \dots, \mathbf{x}_{N_1}^1, \dots, \dots, \mathbf{x}_1^T, \dots, \mathbf{x}_{N_T}^T.$$

Autrement dit, chaque technique adoptée pour combiner ces différents échantillons est représentée par un choix de ces fonctions poids. Ce qui définit l'estimateur par échantillonnage préférentiel multiple

$$\widehat{\Pi}_{\text{MIS}}(\psi) = \sum_{t=1}^T \frac{1}{N_t} \sum_{i=1}^{N_t} w_t(\mathbf{x}_i^t) \frac{\pi(\mathbf{x}_i^t)}{q_t(\mathbf{x}_i^t)} \psi(\mathbf{x}_i^t). \quad (5.8)$$

Avant d'introduire les différentes heuristiques de combinaison étudiées dans Veach and Guibas [101], donnons quelques remarques sur l'estimateur (5.8).

### Quelques remarques sur l'estimateur $\widehat{\Pi}_{\text{MIS}}(\psi)$

(I) Rappelons que les lois

$$Q_1, \dots, Q_T,$$

sont connues *a priori* ainsi que les effectifs  $N_1, \dots, N_T$  des différents échantillons

$$\mathbf{x}_1^1, \dots, \mathbf{x}_{N_1}^1, \dots, \dots, \mathbf{x}_1^T, \dots, \mathbf{x}_{N_T}^T,$$

et pour  $t = 1, \dots, T$ , les particules  $\mathbf{x}_1^t, \dots, \mathbf{x}_{N_t}^t$  sont indépendantes de loi  $Q_t$ . Pour que l'estimateur (5.8) soit sans biais pour (5.3), il suffit que

$$\sum_{t=1}^T w_t(\mathbf{x}) = 1 \quad \text{et} \quad w_t(\mathbf{x}) = 0 \quad \text{lorsque} \quad q_t(\mathbf{x}) = 0.$$

(II) À noter qu'il ne faut pas confondre les fonctions poids  $w_1(\mathbf{x}), \dots, w_T(\mathbf{x})$  introduites dans Veach and Guibas [101] pour paramétrer une stratégie de combinaison d'échantillons, avec le poids d'importance  $\omega_i^t$  associé à une particule  $\mathbf{x}_i^t$  dans (5.6).

(III) Nous remarquons qu'il suffit de choisir les fonctions poids

$$w_t(\mathbf{x}) = \frac{N_t}{\Omega_T} \quad \text{pour } t = 1, \dots, T,$$

dans (5.8) pour obtenir l'estimateur (5.7) et ainsi le poids d'importance basique

$$\omega_i^t = \frac{\pi(\mathbf{x}_i^t)}{q(\mathbf{x}_i^t)}.$$

Les conditions mentionnées dans la remarque (I) sur l'estimateur (5.8) simplifient considérablement la construction de stratégies de combinaison en choisissant les fonctions poids appropriées. Dans ce paragraphe, nous rappelons les heuristiques étudiées dans Veach and Guibas [101] ainsi que les différents liens entre ces stratégies. Pour cela nous donnons l'expression d'une fonction poids pour chaque heuristique pour  $t = 1, \dots, T$ .

### Heuristique de la balance

Ici,

$$\mathbf{w}_t(\mathbf{x}) = \frac{N_t q_t(\mathbf{x})}{\sum_{k=1}^T N_k q_k(\mathbf{x})}.$$

En substituant les fonctions poids définies par l'heuristique de la balance dans l'expression (5.8), nous obtenons l'estimateur de l'intégrale (5.3) donné par

$$\hat{\Pi}_{\text{HB}}(\psi) = \frac{1}{\Omega_T} \sum_{t=1}^T \sum_{i=1}^{N_t} \frac{\pi(\mathbf{x}_i^t)}{\sum_{\ell=1}^T \frac{N_\ell}{\Omega_T} q_\ell(\mathbf{x}_i^t)} \psi(\mathbf{x}_i^t), \quad (5.9)$$

avec  $\Omega_T = N_1 + N_2 + \dots + N_T$  et le poids d'importance est donné par

$$\omega_i^t = \frac{\pi(\mathbf{x}_i^t)}{\sum_{\ell=1}^T \frac{N_\ell}{\Omega_T} q_\ell(\mathbf{x}_i^t)}. \quad (5.10)$$

L'expression du dénominateur dans (5.10) permet d'interpréter l'heuristique de la balance comme une stratégie pour fusionner les différents échantillons

$$\mathbf{x}_1^1, \dots, \mathbf{x}_{N_1}^1, \dots, \dots, \mathbf{x}_1^T, \dots, \mathbf{x}_{N_T}^T,$$

en un seul échantillon avec une seule loi de proposition de densité

$$q(\mathbf{x}) = \sum_{t=1}^T \frac{N_t}{\Omega_T} q_t(\mathbf{x}).$$

Cette loi correspond à une loi d'un mélange dont les poids sont déterministes. On peut interpréter l'estimateur comme une version Rao-Blackwellisée de l'estimateur par échantillonnage préférentiel qu'on aurait obtenu en simulant  $\Omega_T$  particules à partir de la seule loi de proposition de densité  $q(\mathbf{x})$ .

### Heuristique cutoff

Ici,

$$\mathbf{w}_t(\mathbf{x}) = \begin{cases} 0, & \text{si } N_t q_t(\mathbf{x}) < \alpha \max_{1 \leq k \leq T} N_k q_k(\mathbf{x}) \\ \frac{N_t q_t(\mathbf{x})}{\sum_{\ell=1}^T \left\{ N_\ell q_\ell(\mathbf{x}) \mid N_\ell q_\ell(\mathbf{x}) \geq \alpha \max_{1 \leq k \leq T} N_k q_k(\mathbf{x}) \right\}}, & \text{sinon.} \end{cases}$$

L'heuristique cutoff est un raffinement de l'heuristique de la balance. L'idée est de réduire l'effet des particules de faible densité, c'est-à-dire éliminer une particule  $\mathbf{x}_i^t$  en lui associant un poids nul lorsque la densité  $q_t(\mathbf{x}_i^t)$  est inférieure à un certain seuil. L'élimination des particules est contrôlée par le paramètre  $0 \leq \alpha \leq 1$  et le cas particulier  $\alpha = 0$  correspond à l'heuristique de la balance.

### Heuristique de la puissance

$$w_t(\mathbf{x}) = \frac{N_t^\beta q_t^\beta(\mathbf{x})}{\sum_{k=1}^T N_k^\beta q_k^\beta(\mathbf{x})} \quad \text{où } \beta \geq 1.$$

Comme pour l'heuristique cutoff, on retrouve dans l'heuristique de la puissance une généralisation de l'heuristique de la balance. L'introduction de la puissance  $\beta$  favorise des particules  $x_t^t$  simulées dans une zone de forte probabilité pour la loi  $Q_t$  en augmentant leurs poids, et affaiblit le poids des particules simulées dans les zones de faibles probabilités. Le cas particulier  $\beta = 1$  correspond à l'heuristique de la balance.

### Heuristique du maximum

Ici,

$$w_t(\mathbf{x}) = \begin{cases} 1, & \text{si } N_t q_t(\mathbf{x}) = \max_{1 \leq k \leq T} N_k q_k(\mathbf{x}), \\ 0, & \text{sinon.} \end{cases}$$

L'heuristique du maximum est aussi un cas limite de l'heuristique cutoff et de l'heuristique de la puissance, on obtient les fonctions poids associées avec  $\alpha = 1$  et  $\beta = +\infty$  respectivement. L'idée de cette stratégie, est de partager le support de la densité  $\pi$  en  $T$  domaines, dont chacun est donné par l'ensemble  $\{\mathbf{x}, w_t(\mathbf{x}) = 1\}$  et correspond à une loi  $Q_t$  parmi les  $T$  lois de proposition. La proportion de particules dont le poids est nul par cette stratégie est particulièrement élevée.

## 5.2.2 Optimalité de l'heuristique de la balance

Veach and Guibas [101] considèrent l'heuristique cutoff et l'heuristique de la puissance comme des généralisations sophistiquées de l'heuristique de la balance. Ils montrent sur divers simulations pour des problèmes d'infographie que l'heuristique de la puissance avec  $\beta = 2$  et l'heuristique cutoff avec  $\alpha = 0.5$  apportent des améliorations en terme de variance par rapport à l'heuristique de la balance et cela dans un cas particulier où une des différentes lois de proposition est nettement meilleure que les autres. Mais de manière générale, l'heuristique de la balance montre des résultats plus satisfaisants en terme de variance par rapport aux autres heuristiques. Nous avons effectués des comparaisons de ces différentes heuristiques (balance, puissance avec  $\beta = 2$ , cutoff avec  $\alpha = 0.5$ , maximum et échantillonnage préférentiel brut comme en (5.7)) sur un exemple jouet avec une loi cible qui est un mélange de gaussiennes bivariées, et nous avons obtenus les mêmes conclusions que celles de Veach and Guibas [101]. Autrement dit, les résultats de nos simulations ont montrés que l'heuristique de la balance fournit des estimateurs avec des variances légèrement meilleures que celles obtenues avec les heuristiques de la puissance et cutoff qui ont des qualités comparables en terme de variance. Quant aux variances des estimateurs obtenus par l'heuristique du maximum et par la stratégie basique (5.7), elles

sont nettement plus élevées que les autres. La représentation un peu formelle par des fonctions poids de l'estimateur par échantillonnage préférentiel multiple permet d'obtenir le théorème 5.1 qui donne une borne sur la différence en la variance de l'estimateur par la stratégie de l'heuristique de la balance et la variance d'un estimateur quelconque de la classe (5.8).

**Théorème 5.1.** [101] Nous avons l'inégalité suivante

$$\text{Var} \left[ \widehat{\Pi}_{HB}(\psi) \right] - \text{Var} \left[ \widehat{\Pi}_{MIS}(\psi) \right] \leq \left( \frac{1}{\min_{1 \leq t \leq T} N_t} - \frac{1}{\Omega_T} \right) \Pi(\psi)^2, \quad (5.11)$$

avec  $\Omega_T = N_1 + N_2 + \dots + N_T$ .

*Démonstration.* Nous allons traiter le cas où la fonction  $\psi$  est positive. Posons

$$I_t = \int \mathbf{w}_t(\mathbf{x}) \psi(\mathbf{x}) \pi(\mathbf{x}) \, d\mathbf{x}.$$

Ainsi  $\Pi(\psi) = \sum_{t=1}^T I_t$  et

$$\text{Var} \left[ \widehat{\Pi}_{MIS}(\psi) \right] = \left( \int \sum_{t=1}^T \frac{\mathbf{w}_t^2(\mathbf{x})}{N_t q_t(\mathbf{x})} \psi^2(\mathbf{x}) \pi^2(\mathbf{x}) \, d\mathbf{x} \right) - \left( \sum_{t=1}^T \frac{I_t^2}{N_t} \right). \quad (5.12)$$

On cherche à minorer et majorer l'expression précédente. Pour cela, commençons par minimiser son premier terme, plus précisément, cherchons les expressions des fonctions  $\mathbf{w}(\mathbf{x})_1, \dots, \mathbf{w}_T(\mathbf{x})$  qui minimisent l'intégrale

$$\int \sum_{t=1}^T \frac{\mathbf{w}_t^2(\mathbf{x})}{N_t q_t(\mathbf{x})} \psi^2(\mathbf{x}) \pi^2(\mathbf{x}) \, d\mathbf{x}.$$

Pour cela nous allons minimiser la fonction à l'intérieur du signe intégrale en tout point  $\mathbf{x}$ . Fixons un  $\mathbf{x}$ , en considérant  $\psi^2(\mathbf{x}) \pi^2(\mathbf{x})$  comme un terme constant et en notant  $\mathbf{u}_t = \mathbf{w}_t(\mathbf{x})$ , cela revient à minimiser en  $\mathbf{u}_1, \dots, \mathbf{u}_T$  l'expression

$$\sum_{t=1}^T \frac{\mathbf{u}_t^2}{N_t q_t(\mathbf{x})} \quad \text{sous la contrainte} \quad \sum_{t=1}^T \mathbf{u}_t = 1.$$

En appliquant la méthode des multiplicateurs de Lagrange, le minimum est atteint lorsque les  $T + 1$  dérivées partielles en  $\mathbf{u}_1, \dots, \mathbf{u}_T$  et  $\lambda$  de la fonction

$$\sum_{t=1}^T \frac{\mathbf{u}_t}{N_t q_t(\mathbf{x})} + \lambda \left( \sum_{t=1}^T \mathbf{u}_t - 1 \right)$$

sont toutes égales à zéro. Ce qui nous donne le système à  $T$  équations

$$-2\mathbf{u}_t = N_t q_t(\mathbf{x}) \lambda \quad \text{pour } t = 1, \dots, T \quad \text{sous la contrainte} \quad \sum_{t=1}^T \mathbf{u}_t = 1.$$

D'où la solution

$$\mathbf{u}_t = \frac{N_t q_t(\mathbf{x})}{\sum_{\ell=1}^T N_\ell q_\ell(\mathbf{x})}.$$

Maintenant, majorons et minimisons en  $I_1, \dots, I_T$  le deuxième terme  $\sum_{t=1}^T I_t^2/N_t$ . Sous l'hypothèse que  $\psi$  est positive, nous avons les quantités  $I_t \geq 0$  pour tout  $t = 1, \dots, T$ , d'où la majoration

$$\sum_{t=1}^T \frac{I_t^2}{N_t} \leq \frac{1}{\min_{1 \leq t \leq T} N_t} \sum_{t=1}^T I_t^2 \leq \frac{1}{\min_{1 \leq t \leq T} N_t} \left( \sum_{t=1}^T I_t \right)^2 = \frac{\Pi(\psi)^2}{\min_{1 \leq t \leq T} N_t}.$$

Pour minimiser

$$\sum_{t=1}^T \frac{I_t^2}{N_t} \quad \text{en } I_1, \dots, I_T, \text{ sous la contrainte } \sum_{t=1}^T I_t = \Pi(\psi),$$

nous utilisons aussi la méthode des multiplicateurs de Lagrange. La solution des  $T + 1$  équations nous donne le minimum en  $I_t = N_t / \{\Pi(\psi)\Omega_T\}$ , donc le minimum du second terme de (5.12) est

$$\sum_{t=1}^T \frac{1}{N_t} \left\{ \frac{N_t}{\Omega_T} \Pi(\psi) \right\}^2 = \frac{\Pi(\psi)^2}{\Omega_T}.$$

Nous déduisons que

$$\text{Var} \left[ \widehat{\Pi}_{\text{HB}}(\psi) \right] + \Pi(\psi)^2 - \frac{\Pi(\psi)^2}{\min_{1 \leq t \leq T} N_t} \leq \text{Var} \left[ \widehat{\Pi}_{\text{MIS}}(\psi) \right] \leq \text{Var} \left[ \widehat{\Pi}_{\text{MIS}}(\psi) \right] + \Pi(\psi)^2 - \frac{\Pi(\psi)^2}{\Omega_T}. \quad \square$$

Le théorème 5.1 montre qu'une autre stratégie de combinaison d'échantillons donnée par ses fonctions poids ne peut faire beaucoup mieux que l'heuristique de la balance en terme de variance. Ce résultat ne peut être considéré comme un résultat d'optimalité, mais il renforce les conclusions de Veach and Guibas [101] obtenues sur les diverses simulations numériques effectuées sur des problèmes d'infographie, qui recommandent l'utilisation de l'heuristique de la balance.

### 5.2.3 Échantillonnage préférentiel multiple avec variables de contrôle

Les résultats numériques et les conclusions de Veach and Guibas [101] motivent et justifient l'adoption de l'heuristique de la balance dans Cornuet et al. [21] pour recycler les échantillons passés dans le schéma adaptatif de l'algorithme AMIS. En introduisant les variables de contrôle dans les techniques d'échantillonnage préférentiel multiple, Owen and Zhou [83] exploitent l'interprétation de l'heuristique de la balance comme une fusion des différents échantillons en un seul issu d'une seule loi de proposition de densité

$$\sum_{t=1}^T \frac{N_t}{\Omega_T} q_t(\mathbf{x})$$

pour construire une extension de l'estimateur (5.9) en introduisant des variables de contrôle. L'intérêt principal de cette extension est l'obtention d'un résultat d'optimalité sous forme d'une borne supérieure sur la variance de cet estimateur. La technique des variables de contrôle de Ripley [88] exploite la connaissance d'une ou plusieurs valeurs d'intégrales pour obtenir une réduction de variance. Le contexte d'échantillonnage préférentiel multiple est favorable à l'utilisation des variables de contrôle, plus précisément, l'idée de Owen and Zhou [83] est d'utiliser le fait que

$$\int q_t(\mathbf{x}) \, d\mathbf{x} = 1 \quad \text{pour } t = 1, \dots, T,$$

comme des variables de contrôle. Soit  $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_T)$  un vecteur de  $\mathbb{R}^T$ , l'estimateur de l'intégrale (5.3) par échantillonnage préférentiel multiple avec variables de contrôle de Owen and Zhou [83] est défini comme

$$\Pi_{\text{HB}}^{\boldsymbol{\beta}}(\psi) = \frac{1}{\Omega_T} \left\{ \sum_{t=1}^T \sum_{i=1}^{N_t} \frac{\psi(\mathbf{x}_i^t) \pi(\mathbf{x}_i^t) - \sum_{k=1}^T \boldsymbol{\beta}_k q_k(\mathbf{x}_i^t)}{\sum_{\ell=1}^T \frac{N_\ell}{\Omega_T} q_\ell(\mathbf{x}_i^t)} \right\} + \sum_{t=1}^T \boldsymbol{\beta}_t, \quad (5.13)$$

Avant d'énoncer le résultat obtenu dans Owen and Zhou [83] sur la variance de l'estimateur (5.13), nous avons besoin d'introduire un estimateur par échantillonnage préférentiel avec variables de contrôle avec un seul échantillon de taille  $\Omega_T$  d'une loi de proposition  $Q_\alpha$  de densité  $q_\alpha(\mathbf{x}) = \sum \alpha_t q_t(\mathbf{x})$  avec  $\alpha_t = N_t/\Omega_T$ , c'est à dire l'estimateur

$$\Pi_{\text{ISCV}}^{\boldsymbol{\beta}}(\psi) = \frac{1}{\Omega_T} \left\{ \sum_{t=1}^T \sum_{i=1}^{N_t} \frac{\psi(\tilde{\mathbf{x}}_i) \pi(\tilde{\mathbf{x}}_i) - \sum_{k=1}^T \boldsymbol{\beta}_k q_k(\tilde{\mathbf{x}}_i)}{\sum_{\ell=1}^T \alpha_\ell q_\ell(\tilde{\mathbf{x}}_i)} \right\} + \sum_{t=1}^T \boldsymbol{\beta}_t, \quad (5.14)$$

lorsque les particules  $\tilde{\mathbf{x}}_i$  sont indépendantes et de même loi  $Q_\alpha$ , pour  $i = 1, \dots, N_t$  et  $t = 1, \dots, T$ . On remarque qu'on peut remplacer la double somme  $\sum_{t=1}^T \sum_{i=1}^{N_t}$  dans (5.14) par la somme  $\sum_{i=1}^{\Omega_T}$  car toutes les particules sont de même loi  $Q_\alpha$  et donc l'indice  $t$  n'est pas nécessaire mais nous conservons la notation avec la double somme pour ne pas s'éloigner des notations de l'estimateur (5.13). Définissons la variance asymptotique de l'estimateur (5.14) comme

$$\sigma_{\boldsymbol{\beta}}^2 = \int \left\{ \frac{\pi(\mathbf{x}) \psi(\mathbf{x})}{\sum_{t=1}^T \alpha_t q_t(\mathbf{x})} + \sum_{k=1}^T \boldsymbol{\beta}_k - \Pi(\psi) \right\}^2 Q_\alpha(d\mathbf{x}), \quad (5.15)$$

pour toute fonction  $\psi \in L^1(\Pi)$  telle que la variance asymptotique de l'estimateur (5.15) est finie. La technique d'estimation par variables de contrôle revient à chercher le vecteur  $\boldsymbol{\beta}^*$  qui minimise la variance (5.15). L'expression de la variance asymptotique  $\sigma_{\boldsymbol{\beta}}^2$  simplifie la recherche de  $\boldsymbol{\beta}^*$  en un problème de régression multiple. Ce qui revient à considérer l'estimateur  $\hat{\boldsymbol{\beta}}$  de  $\boldsymbol{\beta}^*$  dans la régression multiple en prenant

$$\frac{\pi(\mathbf{x}_i^t) \psi(\mathbf{x}_i^t)}{\sum_{t=1}^T \alpha_t q_t(\mathbf{x}_i^t)}, \quad i = 1, \dots, N_t, \quad t = 1, \dots, T,$$

comme variables à expliquer et

$$\left\{ \frac{q_1(\mathbf{x}_i^t)}{\sum_{t=1}^T \alpha_t q_t(\mathbf{x}_i^t)}, \dots, \frac{q_T(\mathbf{x}_i^t)}{\sum_{t=1}^T \alpha_t q_t(\mathbf{x}_i^t)} \right\}, \quad i = 1, \dots, N_t, \quad t = 1, \dots, T,$$



comme variables explicatives. Après l'introduction des différentes notations, et pour tout ( $1 \leq t \leq T$ ), rappelons que la variance asymptotique de l'estimateur basique par échantillonnage préférentiel de l'intégrale (5.3) de loi de proposition  $Q_t$  de densité  $q_t(\mathbf{x})$ , est donnée par

$$\sigma_t^2 = \int \frac{\pi^2(\mathbf{x})\psi^2(\mathbf{x})}{q_t(\mathbf{x})} d\mathbf{x} - \Pi(\psi)^2.$$

Notons  $\Pi_{\text{HB}}^{\beta^*}(\psi)$ , l'estimateur obtenu en remplaçant  $\beta$  par  $\beta^*$  dans (5.13). Nous énonçons dans le théorème 5.2, le résultat de Owen and Zhou [83].

**Théorème 5.2.** [83]

- (1) Pour tout  $\beta$ , l'estimateur  $\Pi_{\text{HB}}^{\beta}(\psi)$  est sans biais pour l'intégrale  $\Pi(\psi)$ .
- (2) Pour tout  $\beta^*$  minimisant la variance asymptotique (5.15), nous avons

$$\text{Var} \left[ \Pi_{\text{HB}}^{\beta^*}(\psi) \right] \leq \min_{1 \leq t \leq T} \frac{\sigma_t^2}{N_t}. \quad (5.16)$$

*Démonstration.* L'énoncé (1) s'obtient à partir d'un calcul simple d'espérance. Pour l'énoncé (2), nous avons l'inégalité du §6.1.3 de [52] qui montre que pour tout  $\beta$

$$\text{Var} \left[ \Pi_{\text{HB}}^{\beta}(\psi) \right] \leq \text{Var} \left[ \Pi_{\text{ISCV}}^{\beta}(\psi) \right],$$

donc, il nous reste qu'à majorer

$$\text{Var} \left[ \Pi_{\text{ISCV}}^{\beta}(\psi) \right] = \frac{\sigma_{\beta}^2}{\Omega_T},$$

pour cela nous allons montrer par construction que pour tout  $t = 1, \dots, T$  nous avons

$$\sigma_{\beta}^2 \leq \frac{\Omega_T \sigma_t^2}{N_t}.$$

Commençons par le cas  $t = 1$ , pour cela construisons  $\beta = (\beta_1, \dots, \beta_T)$  tel que  $\beta_1 = 0$  et  $\beta_t = \frac{\Pi(\psi)\alpha_t}{\alpha_1}$  pour tout  $t = 2, \dots, T$ . Nous avons

$$\begin{aligned} \sigma_{\beta}^2 &= \int \left\{ \frac{\pi(\mathbf{x})\psi(\mathbf{x}) + \Pi(\psi)\alpha_1^{-1} \sum_{t=2}^T \alpha_t q_t(\mathbf{x})}{\sum_{\ell=1}^T \alpha_{\ell} q_{\ell}(\mathbf{x})} - \Pi(\psi) - \Pi(\psi)\alpha_1^{-1} \sum_{t=2}^T \alpha_t \right\}^2 Q_{\alpha}(d\mathbf{x}) \\ &= \int \left\{ \frac{\pi(\mathbf{x})\psi(\mathbf{x}) + \Pi(\psi)\alpha_1^{-1} (\sum_{k=1}^T q_k(\mathbf{x}) - \alpha_1 q_1(\mathbf{x}))}{\sum_{\ell=1}^T \alpha_{\ell} q_{\ell}(\mathbf{x})} - \Pi(\psi)\alpha_1^{-1} \right\}^2 Q_{\alpha}(d\mathbf{x}) \\ &= \int \frac{\{\pi(\mathbf{x})\psi(\mathbf{x}) - \Pi(\psi)q_1(\mathbf{x})\}^2}{\sum_{\ell=1}^T \alpha_{\ell} q_{\ell}(\mathbf{x})} d\mathbf{x} \\ &\leq \frac{1}{\alpha_1} \int \frac{\{\pi(\mathbf{x})\psi(\mathbf{x}) - \Pi(\psi)q_1(\mathbf{x})\}^2}{q_1(\mathbf{x})} d\mathbf{x} \\ &= \frac{\sigma_1^2}{\alpha_1} = \frac{\Omega_T \sigma_t^2}{N_t}. \end{aligned}$$

Pour montrer que pour tout  $t = 2, \dots, T$ , il existe  $\beta$  tel que  $\sigma_\beta^2 \leq \frac{\Omega_T \sigma_t^2}{N_t}$ , nous utilisons la même construction et le même procédé que dans le cas  $t = 1$ . Pour tout  $\beta^*$  minimisant  $\sigma_\beta^2$ , et en remplaçant  $\beta$  par  $\beta^*$  dans l'estimateur (5.14), nous avons l'estimateur  $\Pi_{IS}^{\beta^*}(\psi)$  dont la variance vérifie

$$\text{Var} \left[ \Pi_{ISCV}^{\beta^*}(\psi) \right] = \frac{\sigma_{\beta^*}^2}{\Omega_T} \leq \min_{1 \leq t \leq T} \frac{\sigma_t^2}{N_t}. \quad \square$$

Par le biais du théorème 5.2, [83] montrent que l'extension de l'échantillonnage préférentiel multiple en ajoutant les variables de contrôle permet de montrer de manière très intuitive l'optimalité de l'heuristique de la balance comme stratégie de combinaison d'échantillons. Plus précisément, en supposant que nous avons un vecteur des coefficients  $\beta$  optimal, la variance de l'estimateur  $\tilde{\Pi}_{HB}(\psi)$  est inférieure à la variance de l'estimateur par échantillonnage préférentiel avec la meilleure loi en terme de variance parmi  $Q_1, \dots, Q_T$ . Cependant cette borne supérieure de la variance est sous condition d'optimalité du vecteur des coefficients  $\beta$ . Or en pratique on possède une estimation par régression du vecteur  $\beta^*$ . Le résultat énoncé dans le théorème 1 de [83] montre que pour un estimateur  $\hat{\beta}$  de  $\beta^*$  obtenu par régression, nous avons

$$\hat{\beta} = \beta^* + O_p(\Omega_T^{-1/2}),$$

ce qui permet de justifier l'utilisation en pratique d'un estimateur de  $\beta^*$ .

## 5.3 Algorithme AMIS d'échantillonnage préférentiel adaptatif et multiple

Dans cette partie, nous présentons en premier la version originale de l'algorithme AMIS proposé dans [21] et nous détaillerons les différentes étapes du processus itératif associé. Deuxièmement, nous présentons les différents outils du processus d'adaptation de la loi de proposition ainsi qu'un aperçu sur les différents choix de la famille paramétrique utilisés dans [21].

### 5.3.1 Schéma adaptatif de l'algorithme AMIS

L'originalité de l'algorithme AMIS de [21] par rapport aux techniques de calibration des lois de proposition sur une loi cible de [14], est dans l'introduction de la procédure de recyclage de tout le passé des simulations dans le schéma adaptatif. Autrement dit, à chaque itération  $t$  du schéma adaptatif, la calibration du nouveau paramètre  $\hat{\theta}_{t+1}$  se fait en combinant l'ensemble des échantillons

$$\mathbf{x}_1^1, \dots, \mathbf{x}_{N_1}^1, \dots, \dots, \mathbf{x}_1^t, \dots, \mathbf{x}_{N_t}^t.$$

Le recyclage des  $t$  échantillons est réalisé à l'étape 3 de l'algorithme 9 par la mise à jour du poids d'importance de chaque particule  $\mathbf{x}_i^k$  ( $1 \leq i \leq N_k, 1 \leq k \leq t-1$ ). Cette mise

**Algorithme 9** Échantillonnage préférentiel adaptatif et multiple (AMIS)

**ENTRÉES :** Une suite de tailles d'échantillons  $N_1, \dots, N_T$  et un paramètre initial noté  $\hat{\theta}_1$ .

À l'itération  $t = 1$ ,

1) Simuler les particules  $\mathbf{x}_1^1, \dots, \mathbf{x}_{N_1}^1$  indépendantes de loi  $Q(\hat{\theta}_1)$

2)

– **Pour**  $i = 1 \rightarrow N_1$  **faire**

Calculer le poids d'importance

$$\omega_i^1 \leftarrow \frac{\pi(\mathbf{x}_i^1)}{q(\mathbf{x}_i^1, \hat{\theta}_1)} \quad \text{et le dénominateur} \quad \delta_i^1 \leftarrow N_1 q(\mathbf{x}_i^1, \hat{\theta}_1).$$

– **Fin pour**

3) Calibrer  $\hat{\theta}_2$  à partir des particules pondérées  $(\mathbf{x}_1^1, \omega_1^1), \dots, (\mathbf{x}_{N_1}^1, \omega_{N_1}^1)$  selon un critère choisi *a priori*.

**Pour**  $t = 2 \rightarrow T$  **faire**

1) Simuler les particules  $\mathbf{x}_1^t, \dots, \mathbf{x}_{N_t}^t$  indépendantes de loi  $Q(\hat{\theta}_t)$

2)

– **Pour**  $i = 1 \rightarrow N_t$  **faire**

Calculer le dénominateur

$$\delta_i^t \leftarrow \sum_{\ell=1}^t N_\ell q(\mathbf{x}_i^t, \hat{\theta}_\ell) \quad \text{et le poids d'importance} \quad \omega_i^t \leftarrow \frac{\pi(\mathbf{x}_i^t)}{\left[ \delta_i^t / \sum_{\ell=1}^t N_\ell \right]}.$$

– **Fin pour**

3) Mettre à jour les anciens poids d'importance,

– **Pour**  $\ell = 1 \rightarrow t - 1$  et  $i = 1 \rightarrow N_\ell$  **faire**

$$\delta_i^\ell \leftarrow \delta_i^\ell + N_t q(\mathbf{x}_i^\ell, \hat{\theta}_t) \quad \text{et} \quad \omega_i^\ell \leftarrow \frac{\pi(\mathbf{x}_i^\ell)}{\left[ \delta_i^\ell / \sum_{j=1}^t N_j \right]}.$$

– **Fin pour**

4) Calibrer le nouveau paramètre  $\hat{\theta}_{t+1}$  à partir de toutes les particules pondérées  $(\mathbf{x}_1^1, \omega_1^1), \dots, (\mathbf{x}_{N_1}^1, \omega_{N_1}^1), \dots, (\mathbf{x}_1^t, \omega_1^t), \dots, (\mathbf{x}_{N_t}^t, \omega_{N_t}^t)$  selon le même critère.

**Fin pour**

**SORTIES :** Le système de particules pondérées final

$$(\mathbf{x}_1^1, \omega_1^1), \dots, (\mathbf{x}_{N_1}^1, \omega_{N_1}^1), \dots, (\mathbf{x}_1^t, \omega_1^t), \dots, (\mathbf{x}_{N_T}^T, \omega_{N_T}^T).$$

à jour des poids correspond à l'utilisation de l'heuristique de la balance comme stratégie de combinaison. Nous avons justifié l'utilisation de cette heuristique dans la partie 5.2 par les résultats numériques et des différentes comparaisons de [101] et par le résultat d'optimalité de [83]. On propose de détailler l'étape 4 de l'algorithme 9, c'est-à-dire la procédure de calibration du paramètre de la nouvelle loi de proposition à chaque itération

de l'algorithme AMIS. À la fin de l'étape 3 de l'algorithme 9, on est en possession d'un système de particules pondérées

$$(\mathbf{x}_1^1, \omega_1^1), \dots, (\mathbf{x}_{N_1}^1, \omega_{N_1}^1), \dots, (\mathbf{x}_1^t, \omega_1^t), \dots, (\mathbf{x}_{N_t}^t, \omega_{N_t}^t) \quad (5.17)$$

Pour calibrer le paramètre  $\theta_{t+1}$  à partir du système 5.17, [21] reprennent les techniques basées sur la minimisation d'une entropie entre la loi  $Q(\theta_{t+1})$  et la loi cible  $\Pi$ , présentées dans [14]. Plus précisément, on propose d'utiliser comme mesure d'entropie, la divergence de Kullback-Leibler définie comme

$$\int \log \left\{ \frac{\pi(\mathbf{x})}{q(\mathbf{x}, \theta)} \right\} \Pi(d\mathbf{x}). \quad (5.18)$$

Minimiser (5.18) en  $\theta$  à partir du système de particules 5.17 revient à maximiser

$$\int \log \{q(\mathbf{x}, \theta)\} \Pi(d\mathbf{x}), \quad (5.19)$$

ce qui revient à faire de l'estimation par maximum de vraisemblance avec une intégrale par rapport à  $\Pi$  qu'on remplacera par une intégrale empirique par rapport à la mesure

$$\sum_{k=1}^t \sum_{i=1}^{N_k} \omega_i^k \delta_{\mathbf{x}_i^k}.$$

Dans le problème de génétique des populations traité dans [21], la loi de proposition  $Q(\theta)$  est une Student multivariée  $\mathcal{T}_3(\boldsymbol{\mu}, \boldsymbol{\Sigma})$  à trois degrés de liberté, de moyenne  $\boldsymbol{\mu}$  et de matrice de covariance  $\boldsymbol{\Sigma}$ . La calibration du nouveau paramètre  $\hat{\theta}_{t+1} = (\hat{\boldsymbol{\mu}}_{t+1}, \hat{\boldsymbol{\Sigma}}_{t+1})$  est donnée par la méthode dite de substitution des moments c'est-à-dire l'estimation de  $\hat{\boldsymbol{\mu}}_{t+1}$  et  $\hat{\boldsymbol{\Sigma}}_{t+1}$  à partir du système (5.17) revient à calculer

$$\hat{\boldsymbol{\mu}}_{t+1} = \frac{1}{\Omega_t} \sum_{k=1}^t \sum_{i=1}^{N_k} \omega_i^k \mathbf{x}_i^k, \quad (5.20)$$

et

$$\hat{\boldsymbol{\Sigma}}_{t+1} = \frac{1}{\Omega_t} \sum_{k=1}^t \sum_{i=1}^{N_k} \omega_i^k (\mathbf{x}_i^k - \hat{\boldsymbol{\mu}}_{t+1})(\mathbf{x}_i^k - \hat{\boldsymbol{\mu}}_{t+1})^T, \quad (5.21)$$

avec  $\Omega_t = N_1 + N_2 + \dots + N_t$ . Sur les expérimentations numériques, [21] proposent de choisir la famille de loi  $\{Q(\theta), \theta \in \Theta\}$  telle que chaque loi  $Q(\theta)$  est un mélange de gaussiennes multivariées, *i.e* pour tout  $\theta$

$$q(\mathbf{x}, \theta) = \sum_{\ell=1}^k \rho_\ell \varphi(\mathbf{x}, \boldsymbol{\mu}_\ell, \boldsymbol{\Sigma}_\ell),$$

avec  $\sum_{\ell=1}^k \rho_\ell = 1$  et  $\varphi(\mathbf{x}, \boldsymbol{\mu}, \boldsymbol{\Sigma})$  la densité d'une gaussienne de moyenne  $\boldsymbol{\mu}$  et de matrice de covariance  $\boldsymbol{\Sigma}$ . Ce qui permet de calibrer  $\theta_{t+1} = (\boldsymbol{\rho}_1, \dots, \boldsymbol{\rho}_k, \boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_1, \dots, \boldsymbol{\Sigma}_k)$  en

utilisant l'algorithme EM, à partir du système 5.17. Dans le contexte de l'algorithme PMC, [14] utilisent un mélange de lois de Student multivariées

$$\sum_{\ell=1}^k \rho_{\ell} t_{\nu}(\mathbf{x}, \boldsymbol{\mu}_{\ell}, \boldsymbol{\Sigma}_{\ell}),$$

avec un degré de liberté fixé  $\nu$  et  $(\boldsymbol{\mu}_{\ell}, \boldsymbol{\Sigma}_{\ell})$  des paramètres de position et d'échelle. Ils proposent de calibrer  $(\rho_1, \dots, \rho_k, \boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_1, \dots, \boldsymbol{\Sigma}_k)$  à chaque itération en utilisant l'extension de l'algorithme EM proposée par [84].

### 5.3.2 Difficultés liées à l'étude de la convergence de l'estimateur AMIS

Rappelons que la nouveauté de l'algorithme AMIS est dans l'introduction de la procédure de recyclage à toutes les étapes et le système final constitué de toutes les particules utilisées dans les  $T$  itérations du processus adaptatif. À partir de ce système final, on définit l'estimateur AMIS de l'intégrale (5.3) par

$$\hat{\Pi}_T^{\text{AMIS}}(\psi) = \frac{1}{\Omega_T} \sum_{t=1}^T \sum_{i=1}^{N_t} \omega_i^t \psi(\mathbf{x}_i^t), \quad (5.22)$$

$\Omega_T = N_1 + N_2 + \dots + N_T$ . Sur les différentes expérimentations numériques sur des problèmes de génétique des populations, effectuées dans [21] et [94] dans le cadre bayésien où la loi cible  $\Pi$  représente la loi *a posteriori*, le processus de recyclage de l'algorithme AMIS réduit considérablement le nombre de simulations, coûteuses en temps de calcul dans ce genre de problèmes. Cependant, l'estimateur (5.22) n'est pas sans biais de l'intégrale (5.3) et l'adoption de l'heuristique de la balance de [101] pour mélanger les échantillons dans le schéma adaptatif rend la manipulation de la structure de dépendance dans le système final de particules par les outils classiques (martingales, chaînes de Markov) compliquée et inaccessible. La convergence quand le nombre d'itérations  $T$  tend vers l'infini de l'estimateur (5.22), est difficile à obtenir, ce qui rend indispensable, une simplification du schéma adaptatif de l'algorithme AMIS. Pour illustrer cette difficulté liée à l'étude de la convergence des estimateurs obtenus par l'algorithme 9, rappelons qu'après la mise à jour des poids d'importance au temps  $t > 2$ , le poids d'importance  $\omega_i^t$  d'une particule simulée au cours des itérations antérieures  $\mathbf{x}_i^k$  pour  $(i \leq N_k, k < t)$  dépend des  $\hat{\theta}_j$  ( $j > k$ ) et donc des particules obtenues après l'itération  $k$ .

## 5.4 Convergence d'un estimateur AMIS modifié

### 5.4.1 Nouvel algorithme avec schéma adaptatif simplifié

Dans cette partie, nous montrons la convergence de l'estimateur AMIS en apportant des légères modifications sur le schéma adaptatif de l'algorithme 9. Nous proposons une

---

**Algorithme 10** Échantillonnage préférentiel adaptatif et multiple (AMIS) avec schéma adaptatif simplifié.

---

**ENTRÉES :** Une suite croissante de tailles d'échantillons  $N_1, \dots, N_T$  et un paramètre initial noté  $\hat{\theta}_1$ .

**Pour**  $t = 1 \rightarrow T$  **faire**

- 1) Simuler les particules  $\mathbf{x}_1^t, \dots, \mathbf{x}_{N_t}^t$  indépendantes de loi  $Q(\hat{\theta}_t)$
- 2) Pour  $i = 1 \rightarrow N_t$ , calculer le poids d'importance

$$\omega_i^t \leftarrow \frac{\pi(\mathbf{x}_i^t)}{q(\mathbf{x}_i^t, \hat{\theta}_t)}.$$

- 4) Calibrer le nouveau paramètre  $\hat{\theta}_{t+1}$  à partir des particules pondérées  $(\mathbf{x}_1^t, \omega_1^t), \dots, (\mathbf{x}_{N_t}^t, \omega_{N_t}^t)$  i.e calculer

$$\hat{\theta}_{t+1} \leftarrow \frac{1}{N_t} \sum_{i=1}^{N_t} \omega_i^t h(\mathbf{x}_i^t).$$

**Fin pour**

**Pour**  $i = 1 \rightarrow N_t$  et  $t = 1 \rightarrow T$  **faire**

$$\omega_i^t \leftarrow \frac{\pi(\mathbf{x}_i^t)}{\sum_{k=1}^T \frac{N_k}{\Omega_T} q(\mathbf{x}_i^t, \hat{\theta}_k)}.$$

**Fin pour**

**SORTIES :** Le système de particules pondérées

$$(\mathbf{x}_1^1, \omega_1^1), \dots, (\mathbf{x}_{N_1}^1, \omega_{N_1}^1), \dots, (\mathbf{x}_1^t, \omega_1^t), \dots, (\mathbf{x}_{N_t}^t, \omega_{N_t}^t), \dots, (\mathbf{x}_{N_T}^T, \omega_{N_T}^T).$$


---

hypothèse de croissance sur la suite des entiers  $(N_t)_{t \geq 1}$  qui représentent les tailles des différents échantillons simulés au cours du processus itératif. En supposant que  $N_t$  tend vers l'infini, l'expression du poids d'importance (5.4) montre que le recyclage de tout le passé des simulations à chaque itération  $t$  du processus adaptatif devient moins important et nous augmentons l'influence du dernier échantillon  $\mathbf{x}_1^{t-1}, \dots, \mathbf{x}_{N_{t-1}}^{t-1}$ . Plus précisément, l'expression du dénominateur du poids d'importance (5.5)

$$\sum_{k=1}^T \frac{N_k}{\Omega_T} q(\cdot, \hat{\theta}_k),$$

montre que l'hypothèse de croissance de la suite  $(N_t)_{t \geq 1}$  rend négligeable, le poids des densités  $q(\cdot, \hat{\theta}_1), \dots, q(\cdot, \hat{\theta}_{T-1})$  dans le mélange précédent. Ce qui justifie notre simplification du schéma itératif de l'algorithme AMIS. Concrètement, nous renonçons au recyclage de tous les échantillons passés pour calibrer le nouveau paramètre  $\hat{\theta}_t$  et nous

nous contentons du seul échantillon

$$\mathbf{x}_1^{t-1}, \dots, \mathbf{x}_{N_{t-1}}^{t-1}.$$

Avec cette simplification du schéma adaptatif, nous nous retrouvons avec une procédure de calibration semblable à celle du schéma PMC de [13]. Après la calibration des différentes lois de proposition

$$Q(\hat{\theta}_1), \dots, Q(\hat{\theta}_T)$$

avec le schéma adaptatif simplifié que nous proposons et cela par la donnée de la suite des paramètres  $\hat{\theta}_1, \dots, \hat{\theta}_T$ , on recycle la totalité des  $T$  échantillons pour former un système final de particules pondérées de taille  $\Omega_T = N_1 + N_2 + \dots + N_T$ . Ainsi toute particule  $\mathbf{x}_i^t$  ( $1 \leq i \leq N_t, 1 \leq t \leq T$ ) est pondérée par

$$\omega_i^t = \frac{\pi(\mathbf{x}_i^t)}{\sum_{k=1}^T \frac{N_k}{\Omega_T} q(\mathbf{x}_i^t, \hat{\theta}_k)}. \quad (5.23)$$

Ce qui nous permet de conserver l'originalité de l'algorithme 9 de [21]. Ces légères modifications que nous apportons à l'algorithme AMIS nous permettent de retrouver une structure de dépendance simplifiée. Pour montrer la convergence presque sûre de la suite  $\hat{\theta}_t$  vers un paramètre  $\theta^*$  correspondant à une loi optimale  $Q(\theta^*)$  ainsi qu'un théorème central limite, nous nous appuyons sur une version multidimensionnelle des résultats de [15] et [37] sur les tableaux triangulaires que nous présentons au chapitre 4. Ensuite, en exploitant l'expression du dénominateur du poids d'importance (5.23) comme une moyenne de Césaro, nous obtenons la convergence de l'estimateur AMIS (5.22) vers l'intégrale (5.3).

### 5.4.2 Hypothèses et résultats

Présentons les hypothèses sur la famille paramétrique  $\{Q(\theta), \theta \in \Theta\}$  et sur la fonction  $\psi$  dont on souhaite calculer l'intégrale. Nous considérons l'espace paramétrique  $\Theta \subseteq \mathbb{R}^d$  et pour tout  $\theta \in \Theta$ , on note  $q(\mathbf{x}, \theta) : \mathcal{X} \times \Theta \mapsto \mathbb{R}^+$  la densité associée à la loi  $Q(\theta)$  par rapport à la mesure de Lebesgue. Rappelons que l'on apprend  $\theta$  au cours de l'algorithme en minimisant un critère d'écart entre la proposition et la cible. On suppose que ce critère est donné par la divergence de Kullback-Leibler, et qu'il existe un unique paramètre  $\theta^*$  dans  $\Theta$  tel que

$$\theta^* = \arg \min_{\theta \in \Theta} \int \log \left\{ \frac{\pi(\mathbf{x})}{q(\mathbf{x}, \theta)} \right\} \pi(\mathbf{x}) \, d\mathbf{x}.$$

De plus, on suppose l'existence d'une fonction  $h : \mathcal{X} \rightarrow \Theta$  telle que

$$\theta^* = \int h(\mathbf{x}) \pi(\mathbf{x}) \, d\mathbf{x}, \quad \text{et} \quad \int \frac{\pi(\mathbf{x}) \|h(\mathbf{x})\|}{q(\mathbf{x}, \theta^*)} \, d\mathbf{x} < +\infty.$$

Considérons les hypothèses suivantes

(H1) Pour tout compact  $K \subset \mathbb{R}^d$  nous avons

$$\lim_{\eta \rightarrow +\infty} \sup_{\theta \in K} \int \pi(\mathbf{x}) \|\mathbf{h}(\mathbf{x})\| \mathbf{1} \left\{ \pi(\mathbf{x}) \|\mathbf{h}(\mathbf{x})\| > \eta q(\mathbf{x}, \theta) \right\} d\mathbf{x} = 0.$$

(H2) L'intégrale  $\int \pi^2(\mathbf{x}) \|\mathbf{h}(\mathbf{x})\|^2 / q(\mathbf{x}, \theta) d\mathbf{x}$  est finie pour tout  $\theta$  et pour tout  $\eta \geq 0$ , l'application

$$\theta \mapsto \int \frac{\pi^2(\mathbf{x}) \|\mathbf{h}(\mathbf{x})\|^2}{q(\mathbf{x}, \theta)} \mathbf{1} \left\{ \pi(\mathbf{x}) \|\mathbf{h}(\mathbf{x})\| > \eta q(\mathbf{x}, \theta) \right\} d\mathbf{x}$$

est continue sur  $\Theta$ .

Et on note  $\Sigma$  la matrice dont l'élément  $\Sigma_{i,j} = \int \pi^2(\mathbf{x}) h_i(\mathbf{x}) h_j(\mathbf{x}) / q(\mathbf{x}, \theta^*) d\mathbf{x} - \theta_i^* \theta_j^*$ .

Les théorèmes 4.1 et 4.2 énoncés au chapitre 4 donnent une loi des grands nombres et un théorème central limite sur la suite  $\widehat{\theta}_t$  que nous formulons dans les deux théorèmes suivants.

**Théorème 5.3.** *Sous l'hypothèse (H1), la suite  $\widehat{\theta}_t$  converge en probabilité vers  $\theta^*$ .*

**Théorème 5.4.** *Sous les hypothèses (H1) et (H2), la suite  $\sqrt{N_t}(\widehat{\theta}_t - \theta^*)$  converge en loi vers une distribution gaussienne  $\mathcal{N}_d(0, \Sigma)$ .*

Nous démontrons ce théorème dans la section 5.4.4. Donnons ici une conséquence immédiate du théorème 5.4.

**Corollaire 5.5.** *Sous (H1) et (H2), si  $\sum_{t=1}^{+\infty} 1/N_t < +\infty$ , alors  $\widehat{\theta}_t$  converge presque sûrement vers  $\theta^*$ .*

Ajoutons les hypothèses suivantes.

(H3) Soit la fonction  $\psi \in \mathbb{L}^1(\Pi)$ , l'intégrale  $\int \left\{ \pi(\mathbf{x}) |\psi(\mathbf{x})| / q(\mathbf{x}, \theta^*) \right\} d\mathbf{x}$  est finie, et l'application  $I_{\psi}^*(\theta)$  définie par  $I_{\psi}^*(\theta) = \int \frac{\pi(\mathbf{x}) \psi(\mathbf{x})}{q(\mathbf{x}, \theta^*)} q(\mathbf{x}, \theta) d\mathbf{x}$  est continue sur  $\Theta$ .

(H4) Pour tout  $\theta \in \Theta$ , l'intégrale  $\int \left\{ \pi^2(\mathbf{x}) \psi^2(\mathbf{x}) / q^2(\mathbf{x}, \theta^*) \right\} q(\mathbf{x}, \theta) d\mathbf{x}$  est finie, et on pose  $\sigma_{\psi}^2 = \int \left\{ \pi^2(\mathbf{x}) \psi^2(\mathbf{x}) / q(\mathbf{x}, \theta^*) \right\} d\mathbf{x} - (\Pi(\psi))^2$ .

(H5) Pour tout  $\eta \geq 0$ , l'application

$$\theta \mapsto \int \frac{\pi^2(\mathbf{x}) \psi^2(\mathbf{x})}{q^2(\mathbf{x}, \theta^*)} q(\mathbf{x}, \theta) \mathbf{1} \left\{ \pi(\mathbf{x}) |\psi(\mathbf{x})| > \eta q(\mathbf{x}, \theta^*) \right\} d\mathbf{x},$$

est continue sur  $\Theta$ .

(H6) Pour toute fonction  $\psi \in \mathbb{L}^1(\Pi)$ , l'application  $R_{\psi}$  définie par

$$R_{\psi}(\theta) = \left\| \pi(\cdot) \psi(\cdot) / q(\cdot, \theta) q(\cdot, \theta^*) \right\|_{\infty} \quad \text{est continue sur } \Theta.$$



(H7) L'application  $G$ , définie par  $G(\boldsymbol{\theta}) = \|\mathbf{q}(\cdot, \boldsymbol{\theta}) - \mathbf{q}(\cdot, \boldsymbol{\theta}^*)\|_\infty$ , est continue.

Nous obtenons la consistance de l'algorithme 10.

**Théorème 5.6.** *Sous l'hypothèse (H1), (H3) et (H7) et pour toute fonction  $\psi \in \mathbb{L}^1(\Pi)$  vérifiant les hypothèses (H3), (H4), (H5) et (H6), l'estimateur  $\widehat{\Pi}_T^{AMIS}(\psi)$  converge presque sûrement vers l'intégrale  $\Pi(\psi)$ .*

### 5.4.3 Vérification des différentes hypothèses sur un exemple jouet

Vérifions ici, les différentes hypothèses sur un exemple. On considère une densité cible  $\pi$  à décroissance rapide et la famille de densités  $\{\mathbf{q}(\mathbf{x}, \boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta\}$  où  $1/\mathbf{q}(\mathbf{x}, \boldsymbol{\theta})$  est à croissance lente. Plus précisément, le produit de  $\pi$  par tout polynôme est borné et il existe un polynôme  $P(\mathbf{x}, \boldsymbol{\theta})$  tel que

$$\frac{1}{\mathbf{q}(\mathbf{x}, \boldsymbol{\theta})} \leq P(\mathbf{x}, \boldsymbol{\theta}).$$

Pour cela, on considère une loi cible gaussienne de densité  $\pi(\mathbf{x}) \propto e^{-(\mathbf{x}-m)^2/2}$  avec  $\mathcal{X} = \mathbb{R}$ . La loi de proposition  $Q(\boldsymbol{\theta})$  est une Student à trois degrés de liberté centrée en  $\boldsymbol{\theta}$ , ainsi

$$\mathbf{q}(\mathbf{x}, \boldsymbol{\theta}) = C \left(1 + \frac{(\mathbf{x} - \boldsymbol{\theta})^2}{3}\right)^{-2} \quad \text{avec } C = \left(\Gamma\left(\frac{3}{2}\right) \sqrt{3\pi}\right)^{-1} \quad \text{et } (\mathbf{x}, \boldsymbol{\theta}) \in \mathbb{R}^2.$$

Commençons par montrer que le minimum de la divergence de Kullback-Leibler entre la loi cible  $\Pi$  et la famille paramétrique  $\{Q(\boldsymbol{\theta}), \boldsymbol{\theta} \in \mathbb{R}\}$  est réalisé en  $\boldsymbol{\theta}^* = m$ . Pour cela, montrons que la fonction

$$\text{KL}(\boldsymbol{\theta}) = \int \log \left\{ \pi(\mathbf{x}) / \mathbf{q}(\mathbf{x}, \boldsymbol{\theta}) \right\} \pi(\mathbf{x}) \, d\mathbf{x}$$

est minimale en  $\boldsymbol{\theta}^* = m$ . Calculons  $\frac{\partial \text{KL}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$  et  $\frac{\partial^2 \text{KL}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^2}$  et montrons que

$$\left. \frac{\partial \text{KL}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right|_{\boldsymbol{\theta}^*} = 0 \quad \text{et} \quad \left. \frac{\partial^2 \text{KL}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^2} \right|_{\boldsymbol{\theta}^*} > 0.$$

Pour appliquer le théorème de dérivabilité sous le signe intégrale, nous remarquons que pour tout  $\boldsymbol{\theta}_0$  et pour tout  $\varepsilon > 0$ ,

$$\left| \frac{\partial}{\partial \boldsymbol{\theta}} \left[ \log \left\{ \pi(\mathbf{x}) / \mathbf{q}(\mathbf{x}, \boldsymbol{\theta}) \right\} \pi(\mathbf{x}) \right] \right| \leq 4 |\boldsymbol{\theta}_0 + \varepsilon - \mathbf{x}| \pi(\mathbf{x}),$$

et

$$\left| \frac{\partial^2}{\partial \boldsymbol{\theta}^2} \left[ \log \left\{ \pi(\mathbf{x}) / \mathbf{q}(\mathbf{x}, \boldsymbol{\theta}) \right\} \pi(\mathbf{x}) \right] \right| \leq 4 \left[ 1 + \frac{(\boldsymbol{\theta}_0 + \varepsilon - \mathbf{x})^2 + (\boldsymbol{\theta}_0 - \varepsilon - \mathbf{x})^2}{6} \right] \pi(\mathbf{x}),$$

pour tout  $\mathbf{x} \in [\theta_0 - \varepsilon, \theta_0 + \varepsilon]$ . On déduit par le théorème de dérivation sous le signe intégrale que

$$\frac{\partial \text{KL}(\theta)}{\partial \theta} = \int \frac{\partial}{\partial \theta} \left[ \log \left\{ \frac{\pi(\mathbf{x})}{q(\mathbf{x}, \theta)} \right\} \pi(\mathbf{x}) \right] d\mathbf{x},$$

et

$$\frac{\partial^2 \text{KL}(\theta)}{\partial \theta^2} = \int \frac{\partial^2}{\partial \theta^2} \left[ \log \left\{ \frac{\pi(\mathbf{x})}{q(\mathbf{x}, \theta)} \right\} \pi(\mathbf{x}) \right] d\mathbf{x}.$$

Un simple calcul d'intégrales nous donne

$$\frac{\partial \text{KL}(\theta)}{\partial \theta} \Big|_{\theta=m} = \frac{4}{\sqrt{2\pi}} \int \frac{u}{u^2 + 3} e^{-\frac{u^2}{2}} du = 0,$$

et

$$\frac{\partial^2 \text{KL}(\theta)}{\partial \theta^2} \Big|_{\theta=m} = -\frac{4}{\sqrt{2\pi}} \int \frac{u^2 - 3}{u^2 + 3} e^{-\frac{u^2}{2}} du > 0.$$

On déduit que  $\theta^* = m$  minimise la divergence de Kullback-Leibler entre la cible  $\Pi$  et la famille de lois  $\{Q(\theta), \theta \in \Theta\}$ . Ainsi donc la fonction  $h$  est définie par  $h(\mathbf{x}) = \mathbf{x}$ .

On traite le cas particulier où la fonction test  $\psi(\mathbf{x}) = \mathbf{x}^k$  et  $k$  un entier.

1. L'hypothèse (H1) peut être interprétée comme une uniforme intégrabilité de la famille de variables aléatoires  $H_K$  indexée par le compact  $K \subset \mathbb{R}^d$  et définie par

$$H_K = \left\{ \xi_\theta = \frac{|h(\mathbf{X}_\theta)| \pi(\mathbf{X}_\theta)}{q(\mathbf{X}_\theta, \theta)} \quad \mathbf{X}_\theta \sim Q(\theta), \quad \theta \in K \right\}.$$

C'est-à-dire

$$\lim_{c \rightarrow +\infty} \sup_{\theta \in K} \mathbb{E} \left[ \xi_\theta \mathbf{1}\{\xi_\theta > c\} \right] = 0. \quad (5.24)$$

En général, il est difficile de vérifier l'uniforme intégrabilité par un calcul direct. Pour montrer (5.24), on va utiliser le théorème de La Vallée-Poussin. Ainsi, il suffit de montrer qu'il existe une fonction croissante  $g : \mathbb{R}^+ \rightarrow \mathbb{R}^+$  telle que

$$\lim_{x \rightarrow +\infty} \frac{g(x)}{x} = +\infty \quad \text{et} \quad \sup_{\theta \in K} \mathbb{E} \left[ g(|\xi_\theta|) \right] < +\infty.$$

Et il suffit de choisir la fonction  $g$  définie par  $x \mapsto x |\log(x)|$  ou  $x \mapsto x^\alpha$  où  $\alpha > 1$ . Avec les fonctions  $\pi(\cdot)$ ,  $q(\cdot, \cdot)$ ,  $h(\cdot)$  et  $\psi(\cdot)$  définies précédemment, la condition (H1) se vérifie par un simple calcul avec cette méthode.

2. Montrons que pour tout  $\eta > 0$ , la fonction  $F$  définie par

$$F(\theta) = \int g(\mathbf{x}, \theta) d\mathbf{x},$$

est continue en tout  $\theta$ , où

$$g(\mathbf{x}, \theta) = \frac{\pi^2(\mathbf{x}) h^2(\mathbf{x})}{q(\mathbf{x}, \theta)} \mathbf{1}\left\{ \frac{\pi(\mathbf{x}) |h(\mathbf{x})|}{q(\mathbf{x}, \theta)} > \eta \right\}.$$

Soient  $\theta_0 \in \mathbb{R}$  et  $\varepsilon > 0$ , montrons qu'il existe une fonction  $\varphi \in \mathbb{L}^1$  positive indépendante de  $\theta$  telle que  $|g(\mathbf{x}, \theta)| \leq \varphi(\mathbf{x})$ , pour tout  $\theta \in [\theta_0 - \varepsilon, \theta_0 + \varepsilon]$  presque partout en  $\mathbf{x}$ . Fixons un  $\mathbf{x}$ , nous avons pour tout  $\theta \in [\theta_0 - \varepsilon, \theta_0 + \varepsilon]$  la majoration

$$\frac{\pi(\mathbf{x})|h(\mathbf{x})|}{q(\mathbf{x}, \theta)} \leq \frac{\pi(\mathbf{x})|h(\mathbf{x})|}{C} \left[ 1 + \frac{(\theta_0 + \varepsilon - \mathbf{x})^2 + (\theta_0 - \varepsilon - \mathbf{x})^2}{3} \right]^2.$$

Puisque la densité  $\pi$  est à décroissance rapide, on a

$$\lim_{x \rightarrow \infty} \frac{\pi(\mathbf{x})|h(\mathbf{x})|}{C} \left[ 1 + \frac{(\theta_0 + \varepsilon - \mathbf{x})^2 + (\theta_0 - \varepsilon - \mathbf{x})^2}{3} \right]^2 = 0.$$

Donc pour  $\eta > 0$ , il existe deux réels  $a, b$  dépendants de  $\theta_0$  et  $\varepsilon$  tels que

$$\mathbf{1}\left\{\pi(\mathbf{x})|h(\mathbf{x})| > \eta q(\mathbf{x}, \theta)\right\}(\mathbf{x}) \leq \mathbf{1}[a, b](\mathbf{x}).$$

Et de la continuité de la fonction  $(\mathbf{x}, \theta) \mapsto \pi^2(\mathbf{x})h^2(\mathbf{x})/q(\mathbf{x}, \theta)$  sur  $[a, b] \times [\theta_0 - \varepsilon, \theta_0 + \varepsilon]$ , on déduit l'existence d'une constante  $M > 0$  telle que

$$\frac{\pi^2(\mathbf{x})h^2(\mathbf{x})}{q(\mathbf{x}, \theta)} \mathbf{1}\left\{\pi(\mathbf{x})|h(\mathbf{x})| > \eta q(\mathbf{x}, \theta)\right\} \leq M \mathbf{1}[a, b](\mathbf{x}).$$

Ainsi, nous avons obtenu une domination de la fonction  $g(\mathbf{x}, \theta)$  par une fonction intégrable indépendante de  $\theta$ . Il nous reste à vérifier la continuité de la fonction  $g(\mathbf{x}, \theta)$  en tout  $\theta \in \mathbb{R}$  pour presque tout  $\mathbf{x}$ . On a pour tout  $\eta > 0$  et tout  $\theta_0 \in \mathbb{R}$ , la continuité de la fonction  $\pi^2(\mathbf{x})h^2(\mathbf{x})/q(\mathbf{x}, \theta)$  qui provient de la continuité des fonctions  $h(\mathbf{x}), \pi(\mathbf{x}), q(\mathbf{x}, \theta)$ . Maintenant, il suffit seulement de vérifier la continuité de la fonction

$$\mathbf{1}\left\{\pi(\mathbf{x})|h(\mathbf{x})| > \eta q(\mathbf{x}, \theta)\right\}(\mathbf{x}) \quad \text{en } \theta_0 \quad \text{pour presque tout } \mathbf{x}.$$

Soit  $\mathbf{x} \in \mathbb{R}$ , si  $\left[\pi(\mathbf{x})|h(\mathbf{x})|/q(\mathbf{x}, \theta_0)\right] > \eta$  alors il existe  $\alpha > 0$  tel que

$$\mathbf{1}\left\{\pi(\mathbf{x})|h(\mathbf{x})| > \eta q(\mathbf{x}, \theta)\right\}(\mathbf{x}) = 1 \quad \text{pour tout } \theta \in ]\theta_0 - \alpha, \theta_0 + \alpha[.$$

Si  $\left[\pi(\mathbf{x})|h(\mathbf{x})|/q(\mathbf{x}, \theta_0)\right] < \eta$  alors il existe  $\alpha > 0$  tel que

$$\mathbf{1}\left\{\pi(\mathbf{x})|h(\mathbf{x})| > \eta q(\mathbf{x}, \theta)\right\}(\mathbf{x}) = 0 \quad \text{pour tout } \theta \in ]\theta_0 - \alpha, \theta_0 + \alpha[.$$

La régularité des fonctions  $h(\mathbf{x}), \pi(\mathbf{x}), q(\mathbf{x}, \theta)$  implique que

$$\lambda \left( \left\{ \frac{\pi(\mathbf{x})|h(\mathbf{x})|}{q(\mathbf{x}, \theta_0)} = \eta \right\} \right) = 0,$$

où  $\lambda$  est la mesure de Lebesgue. On déduit la continuité de la fonction  $g(\mathbf{x}, \theta)$  en  $\theta_0$  pour presque tout  $\mathbf{x}$ . Pour  $\eta = 0$ , nous avons la majoration

$$\frac{\pi^2(\mathbf{x})h^2(\mathbf{x})}{q(\mathbf{x}, \theta)} \leq \frac{\pi^2(\mathbf{x})h^2(\mathbf{x})}{C} \left[ 1 + \frac{(\theta_0 + \varepsilon - \mathbf{x})^2 + (\theta_0 - \varepsilon - \mathbf{x})^2}{3} \right]^2,$$

sur  $[\theta_0 - \varepsilon, \theta_0 + \varepsilon]$ . D'où la domination de la fonction  $\pi^2(\mathbf{x})h^2(\mathbf{x})/q(\mathbf{x}, \theta)$  par une fonction intégrable indépendante de  $\theta$ .

3. De la décroissance rapide de la densité cible  $\pi(\mathbf{x})$ , on a clairement

$$\int \frac{\pi(\mathbf{x})|\psi(\mathbf{x})|}{q(\mathbf{x}, \theta^*)} d\mathbf{x} < +\infty.$$

Pour tout  $\theta \in \mathbb{R}$ , nous avons la domination suivante

$$\frac{\pi(\mathbf{x})\psi(\mathbf{x})}{q(\mathbf{x}, \theta^*)} q(\mathbf{x}, \theta) \leq C \frac{\pi(\mathbf{x})\psi(\mathbf{x})}{q(\mathbf{x}, \theta^*)}.$$

D'où la continuité de la fonction  $I_\psi^*(\theta)$ .

4. Pour tout  $\theta \in \mathbb{R}$

$$\int \frac{\pi^2(\mathbf{x})\psi^2(\mathbf{x})}{q^2(\mathbf{x}, \theta^*)} q(\mathbf{x}, \theta) d\mathbf{x} \leq C \int \frac{\pi^2(\mathbf{x})\psi^2(\mathbf{x})}{q^2(\mathbf{x}, \theta^*)} d\mathbf{x} < +\infty.$$

On déduit aussi l'existence de  $\sigma_\psi^2$ .

5. Pour tout  $\eta \geq 0$ , posons  $A_\eta = \left\{ \mathbf{x} \in \mathbb{R} : \pi(\mathbf{x})|\psi(\mathbf{x})| > \eta q(\mathbf{x}, \theta^*) \right\}$ , nous avons la domination

$$\frac{\pi^2(\mathbf{x})\psi^2(\mathbf{x})}{q^2(\mathbf{x}, \theta^*)} q(\mathbf{x}, \theta) \leq C \frac{\pi^2(\mathbf{x})\psi^2(\mathbf{x})}{q^2(\mathbf{x}, \theta^*)}.$$

D'où la continuité de la fonction

$$\theta \mapsto \int_{A_\eta} \frac{\pi^2(\mathbf{x})\psi^2(\mathbf{x})}{q^2(\mathbf{x}, \theta^*)} q(\mathbf{x}, \theta) d\mathbf{x}.$$

6. Vérifions la continuité de la fonction  $R_\psi$  définie par

$$R_\psi(\theta) = \left\| \frac{\pi(\cdot)\psi(\cdot)}{q(\cdot, \theta)q(\cdot, \theta^*)} \right\|_\infty.$$

Pour cela, on définit la fonction  $f(\mathbf{x}, \theta)$

$$f(\mathbf{x}, \theta) = \frac{\pi(\mathbf{x})\psi(\mathbf{x})}{q(\mathbf{x}, \theta)q(\mathbf{x}, \theta^*)}$$

et le polynôme  $P(\mathbf{x}, \theta)$  par  $P(\mathbf{x}, \theta) = \psi(\mathbf{x}) \left( q(\mathbf{x}, \theta)q(\mathbf{x}, \theta^*) \right)^{-1}$ . Pour  $|\theta_1 - \theta_2| \leq \varepsilon$ , nous avons

$$\left| \left\| f(\cdot, \theta_1) \right\|_\infty - \left\| f(\cdot, \theta_2) \right\|_\infty \right| \leq \left\| f(\cdot, \theta_1) - f(\cdot, \theta_2) \right\|_\infty,$$

et pour tout  $\mathbf{x} \in \mathbb{R}$

$$\begin{aligned} \left| f(\mathbf{x}, \theta_1) - f(\mathbf{x}, \theta_2) \right| &= \pi(\mathbf{x}) \left| P(\mathbf{x}, \theta_1) - P(\mathbf{x}, \theta_2) \right| \\ &\leq \pi(\mathbf{x}) \sup_{\theta \in [\theta_1 - \varepsilon, \theta_1 + \varepsilon]} \left| \frac{\partial P(\mathbf{x}, \theta)}{\partial \theta} \right| |\theta_1 - \theta_2| \\ &\leq \pi(\mathbf{x}) c(1 + |\mathbf{x}|)^{7+k} \sup_{\theta \in [\theta_1 - \varepsilon, \theta_1 + \varepsilon]} (1 + |\theta|)^3 |\theta_1 - \theta_2|, \end{aligned}$$

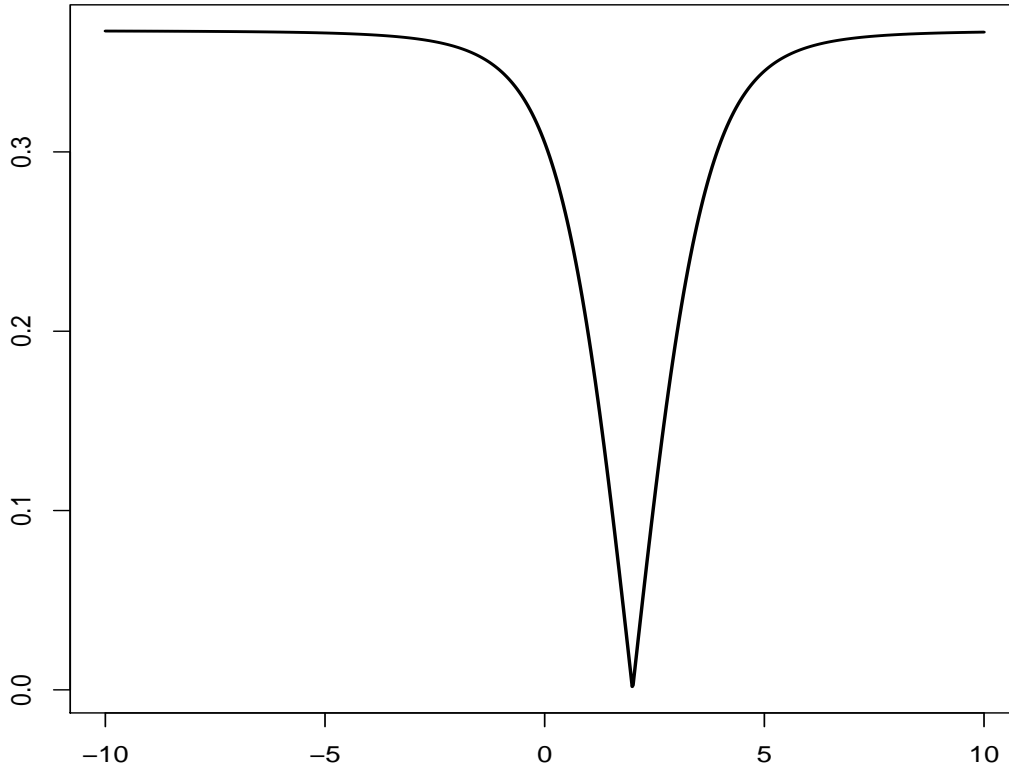


FIGURE 5.1: La fonction  $G(\theta)$  avec  $\theta^* = 2$ .

où  $c$  est une constante indépendante de  $x$  et  $\theta$ . De la décroissance rapide la fonction  $\pi$ , on déduit

$$\begin{aligned} \left\| f(\cdot, \theta_1) - f(\cdot, \theta_2) \right\|_{\infty} &\leq M \max \left\{ (1 + |\theta_1 + \varepsilon|)^3, (1 + |\theta_1 - \varepsilon|)^3 \right\} |\theta_1 - \theta_2| \\ &\rightarrow 0 \quad \text{quand } \varepsilon \text{ tend vers } 0. \end{aligned}$$

D'où la continuité de la fonction  $R_{\psi}$ .

7. Vérifions la continuité de la fonction  $G$  définie par

$$G(\theta) = \left\| q(\cdot, \theta) - q(\cdot, \theta^*) \right\|_{\infty}.$$

En posant  $F(x, \theta) = q(x, \theta) - q(x, \theta^*)$ , après un calcul explicite, la solution  $x(\theta)$  de l'équation

$$\frac{\partial F(x, \theta)}{\partial x} = 0,$$

est la racine d'un polynôme dont les coefficients sont en  $\theta$  et  $\theta^*$ . Nous donnons sur la figure 5.1, la fonction  $G(\theta)$  pour une valeur de  $\theta^*$  fixée à 2.

### 5.4.4 Démonstration des résultats sur $\widehat{\theta}_t$

Dans les démonstrations qui suivent, nous faisons le choix de noter  $\mathbf{X}_i^t$  au lieu de  $\mathbf{x}_i^t$  pour distinguer les variables aléatoires des autres variables et la suite de tribus  $\mathcal{F}_t$  est définie par

$$\mathcal{F}_t = \sigma(\mathbf{X}_1^1, \dots, \mathbf{X}_{N_1}^1, \dots, \mathbf{X}_1^{t-1}, \dots, \mathbf{X}_{N_{t-1}}^{t-1}).$$

La loi des grands nombres sur la suite  $\widehat{\theta}_t$  énoncée dans le théorème 5.3 est une conséquence du théorème 4.1. Le théorème central limite donné dans le théorème 5.4 est conséquence du théorème 4.2. Pour montrer les théorèmes 5.3 et 5.4, nous allons vérifier respectivement les conditions des théorèmes 4.1 et 4.2. La corollaire 5.5 est une conséquence directe du théorème central limite sur la suite  $\widehat{\theta}_t$ .

**Démonstration du théorème 5.3.** Posons

$$\mathbf{V}_{t,i} = \frac{\pi(\mathbf{X}_i^t) \mathbf{h}(\mathbf{X}_i^t)}{N_t q(\mathbf{X}_i^t, \widehat{\theta}_{t-1})},$$

et vérifions les conditions du théorème 4.1

1. La condition (i) est vérifiée par construction des vecteurs aléatoires  $\mathbf{V}_{t,1}, \dots, \mathbf{V}_{t,N_t}$ .
2. On a pour tout  $t \geq 1$

$$\sum_{i=1}^{N_t} \mathbb{E} \left[ \|\mathbf{V}_{t,i}\| \middle| \mathcal{F}_t \right] = \int \|\mathbf{h}(\mathbf{x})\| \pi(\mathbf{x}) \, d\mathbf{x}.$$

D'où la tension de la suite  $\sum_{i=1}^{N_t} \mathbb{E} \left[ \|\mathbf{V}_{t,i}\| \middle| \mathcal{F}_t \right]$  et donc la condition (ii).

3. Montrons que pour tout  $\eta > 0$

$$\begin{aligned} \sum_{i=1}^{N_t} \mathbb{E} \left[ \|\mathbf{V}_{t,i}\| \mathbf{1} \left\{ \|\mathbf{V}_{t,i}\| > \eta \right\} \middle| \mathcal{F}_t \right] &= \int \pi(\mathbf{x}) \|\mathbf{h}(\mathbf{x})\| \mathbf{1} \left\{ \frac{\pi(\mathbf{x}) \|\mathbf{h}(\mathbf{x})\|}{q(\mathbf{x}, \widehat{\theta}_{t-1})} > N_t \eta \right\} \, d\mathbf{x} \\ &\rightarrow 0 \quad \text{en probabilité.} \end{aligned}$$

Pour cela, commençons par montrer que la suite  $\widehat{\theta}_t$  est tendue, notons que

$$\|\widehat{\theta}_t\| \leq \frac{1}{N_t} \sum_{i=1}^{N_t} \frac{\pi(\mathbf{X}_i^t) \|\mathbf{h}(\mathbf{X}_i^t)\|}{q(\mathbf{X}_i^t, \widehat{\theta}_{t-1})}.$$

De plus

$$\mathbb{E} \left( \frac{\pi(\mathbf{X}_i^t) \|\mathbf{h}(\mathbf{X}_i^t)\|}{q(\mathbf{X}_i^t, \widehat{\theta}_{t-1})} \middle| \mathcal{F}_t \right) = \int \pi(\mathbf{x}) \|\mathbf{h}(\mathbf{x})\| \, d\mathbf{x}.$$

Donc  $\mathbb{E} \left( \|\widehat{\theta}_t\| \right) \leq \Pi(\|\mathbf{h}\|)$ . Et par l'inégalité de Markov,

$$\lim_{c \rightarrow +\infty} \sup_t \mathbb{P}(\|\widehat{\theta}_t\| > c) \leq \lim_{c \rightarrow +\infty} \frac{\Pi(\|\mathbf{h}\|)}{c} = 0.$$

D'où la tension de la suite  $\widehat{\boldsymbol{\theta}}_t$ . Introduisons la notation

$$I(\eta, \boldsymbol{\theta}) = \int \pi(\mathbf{x}) \|\mathbf{h}(\mathbf{x})\| \mathbf{1} \left\{ \pi(\mathbf{x}) \|\mathbf{h}(\mathbf{x})\| > \eta q(\mathbf{x}, \boldsymbol{\theta}) \right\} d\mathbf{x},$$

et montrons que sous l'hypothèse (H1),  $I(N_t \eta, \widehat{\boldsymbol{\theta}}_t)$  converge en probabilité vers 0 pour tout  $\eta > 0$ . Pour  $\varepsilon > 0$ , la tension de suite  $\widehat{\boldsymbol{\theta}}_t$  implique qu'il existe un compact  $K_\varepsilon$  tel que

$$\mathbb{P}(\widehat{\boldsymbol{\theta}}_t \notin K_\varepsilon) \leq \varepsilon.$$

Notons que

$$\mathbb{E} \left[ I(N_t \eta, \widehat{\boldsymbol{\theta}}_t) \right] = \mathbb{E} \left[ I(N_t \eta, \widehat{\boldsymbol{\theta}}_t) \mathbf{1} \left\{ \widehat{\boldsymbol{\theta}}_t \in K_\varepsilon \right\} \right] + \mathbb{E} \left[ I(N_t \eta, \widehat{\boldsymbol{\theta}}_t) \mathbf{1} \left\{ \widehat{\boldsymbol{\theta}}_t \notin K_\varepsilon \right\} \right].$$

D'où

$$\begin{aligned} \limsup_t \mathbb{E} \left[ I(N_t \eta, \widehat{\boldsymbol{\theta}}_t) \right] &\leq (1 - \varepsilon) \limsup_{t \rightarrow +\infty} \int_{\boldsymbol{\theta} \in K_\varepsilon} \pi(\mathbf{x}) \|\mathbf{h}(\mathbf{x})\| \mathbf{1} \left\{ \pi(\mathbf{x}) \|\mathbf{h}(\mathbf{x})\| > N_t \eta q(\mathbf{x}, \boldsymbol{\theta}) \right\} d\mathbf{x} \\ &\quad + \varepsilon \Pi(\|\mathbf{h}\|). \end{aligned}$$

Puisque  $\varepsilon$  est arbitraire et de l'hypothèse (H1), on déduit que le second membre de l'inégalité précédente tend vers 0. On en déduit que  $I(N_t \eta, \widehat{\boldsymbol{\theta}}_t)$  tend vers 0 dans  $\mathbb{L}^1(\mathbb{P})$  pour tout  $\eta > 0$ , et donc en probabilité. D'où la condition (iii).

Et donc, l'application du théorème 4.1 conclut la preuve.  $\square$

**Démonstration du théorème 5.4.** Posons

$$\mathbf{V}_{t,i} = \frac{\pi(\mathbf{X}_i^t) \mathbf{h}(\mathbf{X}_i^t)}{\sqrt{N_t} q(\mathbf{X}_i^t, \widehat{\boldsymbol{\theta}}_{t-1})}$$

et vérifions les conditions du théorème 4.2

1. La condition (i) est vérifiée par construction.
2. Pour tout  $t$  et  $i$ , sous l'hypothèse (H2) on a

$$\mathbb{E} \left[ \|\mathbf{V}_{t,i}\|^2 \middle| \mathcal{F}_t \right] = \int \frac{\pi^2(\mathbf{x}) \|\mathbf{h}(\mathbf{x})\|^2}{N_t q(\mathbf{x}, \widehat{\boldsymbol{\theta}}_{t-1})} d\mathbf{x} < +\infty,$$

d'où la condition (ii).

3. Soit  $\mathbf{b} \in \mathbb{R}^d$ , montrons qu'il existe une constante  $\sigma_{\mathbf{b}}^2 > 0$  telle que

$$\sum_{i=1}^{N_t} \left\{ \mathbb{E} \left[ \langle \mathbf{b}, \mathbf{V}_{t,i} \rangle^2 \middle| \mathcal{F}_t \right] - \left( \mathbb{E} \left[ \langle \mathbf{b}, \mathbf{V}_{t,i} \rangle \middle| \mathcal{F}_t \right] \right)^2 \right\} \longrightarrow \sigma_{\mathbf{b}}^2, \quad \text{en probabilité.}$$

Notons  $\mathbf{V}_{t,i}^k$  la  $k$ ème composante du vecteur  $\mathbf{V}_{t,i}$ , et remarquons que

$$\mathbb{E} \left[ \langle \mathbf{b}, \mathbf{V}_{t,i} \rangle^2 \middle| \mathcal{F}_t \right] = \sum_{k=1}^d \mathbf{b}_k^2 \mathbb{E} \left[ (\mathbf{V}_{t,i}^k)^2 \middle| \mathcal{F}_t \right] + 2 \sum_{1 \leq k < \ell \leq d} \mathbf{b}_k \mathbf{b}_\ell \mathbb{E} \left[ \mathbf{V}_{t,i}^k \mathbf{V}_{t,i}^\ell \middle| \mathcal{F}_t \right],$$

et

$$\sum_{i=1}^{N_t} \left( \mathbb{E} \left[ \langle \mathbf{b}, \mathbf{V}_{t,i} \rangle \middle| \mathcal{F}_t \right] \right)^2 = \sum_{k=1}^d \mathbf{b}_k^2 (\boldsymbol{\theta}_k^*)^2 + 2 \sum_{1 \leq k < \ell \leq d} \mathbf{b}_k \mathbf{b}_\ell \boldsymbol{\theta}_k^* \boldsymbol{\theta}_\ell^*.$$

Par ailleurs, l'hypothèse de continuité dans (H2) et la convergence en probabilité de la suite  $\widehat{\boldsymbol{\theta}}_t$  implique que pour tout couple  $(\mathbf{V}_{t,i}^k, \mathbf{V}_{t,i}^\ell)$  de composantes de  $\mathbf{V}_{t,i}$ , nous avons

$$\begin{aligned} \sum_{i=1}^{N_t} \mathbb{E} \left[ \mathbf{V}_{t,i}^k \mathbf{V}_{t,i}^\ell \middle| \mathcal{F}_t \right] &= \int \frac{\pi^2(\mathbf{x}) h_k(\mathbf{x}) h_\ell(\mathbf{x})}{q(\mathbf{x}, \widehat{\boldsymbol{\theta}}_{t-1})} d\mathbf{x} \\ &\longrightarrow \int \frac{\pi^2(\mathbf{x}) h_k(\mathbf{x}) h_\ell(\mathbf{x})}{q(\mathbf{x}, \boldsymbol{\theta}^*)} d\mathbf{x} \quad \text{en probabilité.} \end{aligned}$$

On déduit que

$$\sum_{i=1}^{N_t} \left\{ \mathbb{E} \left[ \langle \mathbf{b}, \mathbf{V}_{t,i} \rangle^2 \middle| \mathcal{F}_t \right] - \left( \mathbb{E} \left[ \langle \mathbf{b}, \mathbf{V}_{t,i} \rangle \middle| \mathcal{F}_t \right] \right)^2 \right\} \longrightarrow \mathbf{b}^t \boldsymbol{\Sigma} \mathbf{b}$$

en probabilité, d'où la condition (iii).

4. Montrons que pour tout vecteur  $\mathbf{e}$  de la base canonique de  $\mathbb{R}^d$  et pour tout  $\eta > 0$

$$\begin{aligned} \sum_{i=1}^{N_t} \mathbb{E} \left[ \langle \mathbf{e}, \mathbf{V}_{t,i} \rangle^2 \mathbf{1} \left\{ |\langle \mathbf{e}, \mathbf{V}_{t,i} \rangle| > \eta \right\} \middle| \mathcal{F}_t \right] &\leq \int \frac{\pi^2(\mathbf{x}) \|\mathbf{h}(\mathbf{x})\|^2}{q(\mathbf{x}, \widehat{\boldsymbol{\theta}}_{t-1})} \mathbf{1} \left\{ \pi(\mathbf{x}) \|\mathbf{h}(\mathbf{x})\| > \sqrt{N_t} \eta q(\mathbf{x}, \widehat{\boldsymbol{\theta}}_{t-1}) \right\} d\mathbf{x} \\ &\longrightarrow 0, \quad \text{en probabilité.} \end{aligned}$$

Pour cela introduisons la notation

$$M(\eta, \boldsymbol{\theta}) = \int \frac{\pi^2(\mathbf{x}) \|\mathbf{h}(\mathbf{x})\|^2}{q(\mathbf{x}, \boldsymbol{\theta})} \mathbf{1} \left\{ \pi(\mathbf{x}) \|\mathbf{h}(\mathbf{x})\| > \eta q(\mathbf{x}, \boldsymbol{\theta}) \right\} d\mathbf{x},$$

et montrons que sous l'hypothèse (H2),  $M(\sqrt{N_t} \eta, \widehat{\boldsymbol{\theta}}_t)$  converge en probabilité vers 0, pour tout  $\eta > 0$ . Notons que les conditions d'intégrabilité des fonctions

$$\frac{\pi^2(\mathbf{x}) \|\mathbf{h}(\mathbf{x})\|^2}{q(\mathbf{x}, \boldsymbol{\theta}^*)} \quad \text{et} \quad \frac{\pi(\mathbf{x}) \|\mathbf{h}(\mathbf{x})\|}{q(\mathbf{x}, \boldsymbol{\theta}^*)},$$

implique que  $M(\eta, \boldsymbol{\theta}^*)$  converge vers 0 lorsque  $\eta$  tend vers l'infini. Et nous avons aussi, la convergence en probabilité de  $M(\eta, \widehat{\boldsymbol{\theta}}_t)$  vers  $M(\eta, \boldsymbol{\theta}^*)$  pour tout  $\eta > 0$  qui découle de l'hypothèse (H2) et de la convergence en probabilité de la suite  $\widehat{\boldsymbol{\theta}}_t$ . Pour  $\varepsilon > 0$ , il existe  $\eta_0 > 0$ , tel que  $M(\eta_0, \boldsymbol{\theta}^*) \leq \varepsilon$  et pour  $t$  tel que  $\sqrt{N_t} \eta \geq \eta_0$  et donc  $M(\sqrt{N_t} \eta, \widehat{\boldsymbol{\theta}}_t) \leq M(\eta_0, \widehat{\boldsymbol{\theta}}_t)$ . Nous avons

$$\mathbb{P} \left[ M(\sqrt{N_t} \eta, \widehat{\boldsymbol{\theta}}_t) \geq 2\varepsilon \right] \leq \mathbb{P} \left[ M(\eta_0, \widehat{\boldsymbol{\theta}}_t) \geq 2\varepsilon \right] \leq \mathbb{P} \left[ M(\eta_0, \widehat{\boldsymbol{\theta}}_t) - M(\eta_0, \boldsymbol{\theta}^*) \geq \varepsilon \right].$$

On déduit la convergence en probabilité de  $M(\sqrt{N_t} \eta, \widehat{\boldsymbol{\theta}}_t)$  vers 0 pour tout  $\eta > 0$ . D'où la condition (iv) du théorème 4.2, et donc la convergence en loi de la suite  $\sqrt{N_t}(\widehat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*)$  vers une distribution gaussienne  $\mathcal{N}_d(0, \boldsymbol{\Sigma})$ .



□

**Démonstrations du corollaire 5.5.** Le théorème 5.3 implique que la constante

$$K := \sup_t N_t \mathbb{E} \left[ \|\hat{\theta}_t - \theta^*\|^2 \right] \text{ est finie.}$$

En utilisant cette majoration dans l'inégalité de Tchebychev, il vient

$$\sum_{t=1}^{+\infty} \mathbb{P} \left( \|\hat{\theta}_t - \theta^*\| > \varepsilon \right) \leq \varepsilon^{-2} K \sum_{t=1}^{+\infty} 1/N_t < +\infty,$$

et d'après le lemme de Borel-Cantelli, nous avons la convergence presque sûre de  $\hat{\theta}_t$  vers  $\theta^*$ . □

### 5.4.5 Démonstration de la consistance de l'algorithme (théorème 5.6)

Le théorème 5.6 donne la convergence de l'estimateur  $\hat{\Pi}_T^{\text{AMIS}}(\psi)$  vers l'intégrale  $\Pi(\psi)$  pour une certaine classe de fonctions  $\psi$  vérifiant les hypothèses précédentes. Afin de démontrer ce théorème, nous introduisons la suite

$$\hat{\Pi}_T^*(\psi) = \frac{1}{\Omega_T} \sum_{t=1}^T \sum_{i=1}^{N_t} \frac{\pi(\mathbf{x}_i^t) \psi(\mathbf{x}_i^t)}{q(\mathbf{x}_i^t, \theta^*)},$$

(qu'on ne peut calculer en pratique car  $\theta^*$  est inconnu). En remarquant que  $\hat{\Pi}_T^*(\psi)$  s'écrit comme une moyenne de Cesàro de la suite  $\hat{\pi}_t^*(\psi)$  qu'on définit comme

$$\hat{\pi}_t^*(\psi) = \frac{1}{N_t} \sum_{i=1}^{N_t} \frac{\pi(\mathbf{x}_i^t) \psi(\mathbf{x}_i^t)}{q(\mathbf{x}_i^t, \theta^*)}.$$

Le plan de la preuve du théorème 5.6 est comme suit. À partir du corollaire 5.5 et des hypothèses (H6) et (H7), nous montrons la convergence vers 0 de la différence entre l'estimateur  $\hat{\Pi}_T^{\text{AMIS}}(\psi)$  et la suite  $\hat{\Pi}_T^*(\psi)$ . Ensuite, nous montrons que la suite  $\hat{\pi}_t^*(\psi)$  converge vers l'intégrale  $\Pi(\psi)$  pour obtenir la convergence de  $\hat{\Pi}_T^*(\psi)$  vers  $\Pi(\psi)$ .

Au cours des différentes étapes de la démonstration du théorème 5.6, nous faisons appel au lemme de Cesàro que nous formulons dans le lemme suivant.

**Lemme 5.7** (Cesàro pondéré). *Si  $U_n$  est une suite convergente vers  $U_\infty$  et  $b_n$  une suite de nombres réels positifs telle que  $b_1 + \dots + b_n$  tend vers l'infini, alors la suite  $\sum_{k=1}^n b_k U_k / \sum_{\ell=1}^n b_\ell$  converge vers  $U_\infty$ .*

*Démonstration.* Soit  $\varepsilon > 0$ , il existe un entier  $n_0 > 0$  tel que pour tout  $n \geq n_0$ , on a

$$|U_n - U_\infty| \leq \varepsilon$$

et nous avons aussi

$$\begin{aligned} \left| \frac{\sum_{k=1}^n b_k U_k}{\sum_{\ell=1}^n b_\ell} - U_\infty \right| &\leq \frac{\sum_{k=1}^n b_k |U_k - U_\infty|}{\sum_{\ell=1}^n b_\ell} \\ &= \frac{\sum_{k=1}^{n_0} b_k |U_k - U_\infty|}{\sum_{\ell=1}^n b_\ell} + \frac{\sum_{k=n_0+1}^n b_k |U_k - U_\infty|}{\sum_{\ell=1}^n b_\ell}. \end{aligned}$$

La divergence de la série  $\sum_k b_k$  implique qu'il existe  $n_1 \geq n_0$  tel que pour tout  $n \geq n_1$

$$\frac{\sum_{k=1}^{n_0} b_k |U_k - U_\infty|}{\sum_{\ell=1}^n b_\ell} \leq \varepsilon$$

et donc

$$\left| \frac{\sum_{k=1}^n b_k U_k}{\sum_{\ell=1}^n b_\ell} - U_\infty \right| \leq 2\varepsilon. \quad \square$$

La proposition 5.9 donne la convergence vers 0 de la différence  $\widehat{\Pi}_T^{\text{AMIS}}(\psi) - \widehat{\Pi}_T^*(\psi)$ . Pour cela, on définit pour tout  $T \geq 1$ , la fonction  $D(\cdot, \widehat{\theta}_{1:T}) : \mathcal{X} \mapsto \mathbb{R}_+$  telle que pour  $\mathbf{x} \in \mathcal{X}$

$$D(\mathbf{x}, \widehat{\theta}_{1:T}) = \frac{\sum_{k=1}^T N_k q(\mathbf{x}, \widehat{\theta}_k)}{\Omega_T},$$

avec  $\Omega_T = N_1 + N_2 + \dots + N_T$ . Nous avons le lemme suivant

**Lemme 5.8.** *Sous les hypothèses (H1), (H2) et (H7), et si  $\sum_{t=1}^{+\infty} 1/N_t < +\infty$ , alors*

$$\lim_{T \rightarrow +\infty} \left\| D(\cdot, \widehat{\theta}_{1:T}) - q(\cdot, \theta^*) \right\|_\infty = 0 \quad \text{presque sûrement.}$$

*Démonstration.* Montrons que la suite  $G(\widehat{\theta}_t)$  converge vers 0 presque sûrement. Sous l'hypothèse (H7) de continuité de la fonction  $G(\theta)$  et du corollaire 5.5, nous avons

$$\lim_{t \rightarrow +\infty} G(\widehat{\theta}_t) = 0 \quad \text{presque sûrement.}$$

Nous remarquons que pour tout  $T \geq 1$ , nous avons

$$\begin{aligned} \left\| D(\cdot, \widehat{\theta}_{1:T}) - q(\cdot, \theta^*) \right\|_\infty &\leq \sum_{k=1}^T \frac{N_k \left\| q(\cdot, \widehat{\theta}_k) - q(\cdot, \theta^*) \right\|_\infty}{\Omega_T} \\ &= \sum_{k=1}^T \frac{N_k}{\Omega_T} G(\widehat{\theta}_k) \rightarrow 0 \quad \text{presque sûrement} \end{aligned}$$

par le lemme 5.7. □

**Proposition 5.9.** *Pour toute fonction  $\psi \in \mathbb{L}^1(\Pi)$  qui vérifie l'hypothèse (H6), sous les hypothèses (H1), (H2) et (H7), et si  $\sum_{t=1}^{+\infty} 1/N_t < +\infty$ , alors*

$$\lim_{T \rightarrow +\infty} \widehat{\Pi}_T^{\text{AMIS}}(\psi) - \widehat{\Pi}_T^*(\psi) = 0 \quad \text{presque sûrement.}$$

*Démonstration.* Du corollaire 5.5, on a la convergence presque sûre de la suite  $\widehat{\theta}_t$  vers  $\theta^*$ , et de la continuité de la fonction  $R_\psi$ , pour toute fonction  $\psi \in \mathbb{L}^1(\Pi)$  qui vérifie l'hypothèse (H6), nous avons

$$\lim_{t \rightarrow +\infty} R_\psi(\widehat{\theta}_t) = R_\psi(\theta^*) \quad \text{presque sûrement.}$$

Du lemme 5.7, on déduit

$$\lim_{T \rightarrow +\infty} \sum_{k=1}^T \frac{N_k R_\psi(\widehat{\theta}_k)}{\Omega_T} = R_\psi(\theta^*) \quad \text{presque sûrement.}$$

Donc, avec le lemme 5.8,

$$\lim_{T \rightarrow +\infty} \left( \sum_{k=1}^T \frac{N_k R_\psi(\widehat{\theta}_k)}{\Omega_T} \right) \left\| D(\cdot, \widehat{\theta}_{1:T}) - q(\cdot, \theta^*) \right\|_\infty = 0 \quad \text{presque sûrement.}$$

Nous avons pour tout  $T \geq 1$

$$\begin{aligned} \left| \widehat{\Pi}_T^{\text{AMIS}}(\psi) - \widehat{\Pi}_T^*(\psi) \right| &\leq \frac{1}{\Omega_T} \sum_{t=1}^T \sum_{i=1}^{N_t} \left| \pi(\mathbf{X}_i^t) \psi(\mathbf{X}_i^t) \right| \left| \frac{1}{D(\mathbf{X}_i^t, \widehat{\theta}_{1:T})} - \frac{1}{q(\mathbf{X}_i^t, \theta^*)} \right| \\ &= \frac{1}{\Omega_T} \sum_{t=1}^T \sum_{i=1}^{N_t} \left| \frac{\pi(\mathbf{X}_i^t) \psi(\mathbf{X}_i^t)}{D(\mathbf{X}_i^t, \widehat{\theta}_{1:T}) q(\mathbf{X}_i^t, \theta^*)} \right| \left| D(\mathbf{X}_i^t, \widehat{\theta}_{1:T}) - q(\mathbf{X}_i^t, \theta^*) \right|. \end{aligned}$$

En utilisant la convexité de la fonction  $x \mapsto 1/x$  sur  $\mathbb{R}_+$ , nous obtenons la majoration

$$\frac{\pi(\mathbf{X}_i^t) |\psi(\mathbf{X}_i^t)|}{D(\mathbf{X}_i^t, \widehat{\theta}_{1:T}) q(\mathbf{X}_i^t, \theta^*)} \left| D(\mathbf{X}_i^t, \widehat{\theta}_{1:T}) - q(\mathbf{X}_i^t, \theta^*) \right| \leq \left( \sum_{k=1}^T \frac{N_k R_\psi(\widehat{\theta}_k)}{\Omega_T} \right) \left\| D(\cdot, \widehat{\theta}_{1:T}) - q(\cdot, \theta^*) \right\|_\infty.$$

On déduit la majoration

$$\left| \widehat{\Pi}_T^{\text{AMIS}}(\psi) - \widehat{\Pi}_T^*(\psi) \right| \leq \frac{1}{\Omega_T} \sum_{t=1}^T \sum_{i=1}^{N_t} \left( \sum_{k=1}^T \frac{N_k R_\psi(\widehat{\theta}_k)}{\Omega_T} \right) \left\| D(\cdot, \widehat{\theta}_{1:T}) - q(\cdot, \theta^*) \right\|_\infty,$$

et par le lemme 5.7, le second membre de cette majoration tend presque sûrement vers 0, ce qui conclut la preuve de la proposition 5.9.  $\square$

Le théorème 5.10 donne la convergence de la suite  $\widehat{\pi}_t^*(\psi)$  vers l'intégrale  $\Pi(\psi)$ , et ainsi permet d'obtenir la convergence de la suite  $\widehat{\Pi}_T^*(\psi)$  vers l'intégrale  $\Pi(\psi)$  que nous formulons dans le corollaire 5.11.

**Théorème 5.10.** *Sous les hypothèses (H1) et (H2), et pour toute fonction  $\psi \in \mathbb{L}^1(\Pi)$  vérifiant les hypothèses (H3), (H4) et (H5), alors la suite  $\sqrt{N_t} \{ \widehat{\Pi}_t^*(\psi) - I_\psi^*(\widehat{\theta}_{t-1}) \}$  converge en loi vers une distribution gaussienne  $\mathcal{N}(0, \sigma_\psi^2)$ .*

**Corollaire 5.11.** *Sous (H1),(H2),(H3), (H4) et (H5), si  $\sum_{t=1}^{+\infty} 1/N_t < +\infty$ , alors  $\widehat{\Pi}_T^*(\psi)$  converge presque sûrement vers  $\Pi(\psi)$ .*

La démonstration du théorème 5.10 est semblable à celle du théorème 5.4, c'est une application directe du théorème 4.2.

*Démonstration du théorème 5.10.* Posons

$$\mathbf{V}_{t,i} = \frac{\pi(\mathbf{X}_i^t)\psi(\mathbf{X}_i^t)}{\sqrt{N_t}q(\mathbf{X}_i^t, \theta^*)}$$

et vérifions les conditions du théorème 4.2.

1. La condition (i) est donnée par construction.
2. Sous l'hypothèse (H4)

$$\mathbb{E}[|\mathbf{V}_{t,i}|^2 | \mathcal{F}_t] = \int \frac{\pi^2(\mathbf{x})\psi^2(\mathbf{x})}{N_t q^2(\mathbf{x}, \theta^*)} q(\mathbf{x}, \widehat{\theta}_{t-1}) d\mathbf{x} < +\infty$$

et donc nous avons vérifié la condition (ii).

3. Par des arguments de continuité issus des hypothèses (H3) et (H5), et de la convergence en probabilité de la suite  $\widehat{\theta}_t$ , nous avons

$$\begin{aligned} \sum_{i=1}^{N_t} \left\{ \mathbb{E}[|\mathbf{V}_{t,i}|^2 | \mathcal{F}_t] - \left( \mathbb{E}[\mathbf{V}_{t,i} | \mathcal{F}_t] \right)^2 \right\} &= \int \frac{\pi^2(\mathbf{x})\psi^2(\mathbf{x})}{q^2(\mathbf{x}, \theta^*)} q(\mathbf{x}, \widehat{\theta}_{t-1}) d\mathbf{x} \\ &\quad - \left( \int \frac{\pi(\mathbf{x})\psi(\mathbf{x})}{q(\mathbf{x}, \theta^*)} q(\mathbf{x}, \widehat{\theta}_{t-1}) d\mathbf{x} \right)^2 \\ &\rightarrow \sigma_\psi^2 \quad \text{en probabilité,} \end{aligned}$$

ce qui nous permet de vérifier la condition (iii).

4. Soit  $\eta > 0$ , montrons que

$$\begin{aligned} \sum_{i=1}^{N_t} \mathbb{E} \left[ \mathbf{V}_{t,i}^2 \mathbf{1}\{|\mathbf{V}_{t,i}| > \eta\} | \mathcal{F}_t \right] &= \int \frac{\pi^2(\mathbf{x})\psi^2(\mathbf{x})}{q^2(\mathbf{x}, \theta^*)} q(\mathbf{x}, \widehat{\theta}_{t-1}) \mathbf{1}\left\{ \frac{\pi(\mathbf{x})|\psi(\mathbf{x})|}{q(\mathbf{x}, \theta^*)} > \sqrt{N_t}\eta \right\} d\mathbf{x} \\ &\rightarrow 0 \quad \text{en probabilité.} \end{aligned}$$

Pour cela, posons

$$M^*(\eta, \theta) = \int \frac{\pi^2(\mathbf{x})\psi^2(\mathbf{x})}{q^2(\mathbf{x}, \theta^*)} q(\mathbf{x}, \theta) \mathbf{1}\left\{ \pi(\mathbf{x})|\psi(\mathbf{x})| > \eta q(\mathbf{x}, \theta^*) \right\} d\mathbf{x},$$

et montrons que pour tout  $\eta > 0$ ,  $M^*(\sqrt{N_t}\eta, \widehat{\theta}_{t-1})$  converge vers 0 en probabilité. Sous l'hypothèse (H3), l'intégrabilité de la fonction  $\pi(\mathbf{x})|\psi(\mathbf{x})|/q(\mathbf{x}, \theta^*)$  implique

$$M^*(\eta, \theta^*) \rightarrow 0 \quad \text{quand } \eta \text{ tend vers l'infini.}$$

Et sous l'hypothèse de continuité en (H3), nous avons la convergence en probabilité de  $M^*(\eta, \hat{\theta}_{t-1})$  vers  $M^*(\eta, \theta^*)$  pour tout  $\eta > 0$ . Pour  $\varepsilon > 0$ , il existe  $\eta_0 > 0$  tel que  $M^*(\eta_0, \theta) \leq \varepsilon$  et pour tout  $t$  tel que  $\sqrt{N_t}\eta \geq \eta_0$  et donc

$$M^*(\sqrt{N_t}\eta, \hat{\theta}_{t-1}) \leq M^*(\eta_0, \hat{\theta}_{t-1}).$$

On a alors

$$\mathbb{P} \left[ M^*(\sqrt{N_t}\eta, \hat{\theta}_{t-1}) \geq 2\varepsilon \right] \leq \mathbb{P} \left[ M^*(\eta_0, \hat{\theta}_{t-1}) \geq 2\varepsilon \right] \leq \mathbb{P} \left[ M^*(\eta_0, \hat{\theta}_{t-1}) - M^*(\eta_0, \theta^*) \geq \varepsilon \right].$$

D'où la convergence de  $M^*(\sqrt{N_t}\eta, \hat{\theta}_{t-1})$  vers 0 en probabilité. On déduit la condition (iv).

Et donc nous avons la convergence en loi de

$$\sqrt{N_t} \{ \hat{\pi}_t^*(\psi) - I_\psi^*(\hat{\theta}_{t-1}) \}$$

vers une gaussienne  $\mathcal{N}(0, \sigma_\psi^2)$ . □

*Démonstration du corollaire 5.11.* Montrons que la suite  $\hat{\pi}_t^*(\psi)$  converge presque sûrement vers l'intégrale  $\Pi(\psi)$ . Le théorème 5.4 implique

$$K := \sup_{t \geq 1} N_t \mathbb{E} \left[ \{ \hat{\pi}_t^*(\psi) - I_\psi^*(\hat{\theta}_{t-1}) \}^2 \right] < +\infty.$$

Et pour  $\varepsilon > 0$ , de l'inégalité de Tchebychev, nous avons

$$\sum_{t=1}^{+\infty} \mathbb{P} \left( \left| \hat{\pi}_t^*(\psi) - I_\psi^*(\hat{\theta}_{t-1}) \right| > \varepsilon \right) \leq \varepsilon^{-2} K \sum_{t=1}^{+\infty} \frac{1}{N_t} \leq +\infty$$

et par le lemme de Borel-Cantelli, nous avons la convergence presque sûre de  $\hat{\pi}_t^*(\psi) - I_\psi^*(\hat{\theta}_{t-1})$  vers 0. Par ailleurs, le corollaire 5.5 montre que la suite  $\hat{\theta}_t$  converge presque sûrement vers  $\theta^*$ . On déduit de la continuité de la fonction  $I_\psi^*(\theta)$  la convergence presque sûre de  $I_\psi^*(\hat{\theta}_{t-1})$  vers l'intégrale  $\Pi(\psi)$ , et donc

$$\lim_{t \rightarrow +\infty} \hat{\pi}_t^*(\psi) = \Pi(\psi) \quad \text{presque sûrement.}$$

Par le lemme 5.7

$$\lim_{T \rightarrow +\infty} \hat{\Pi}_T^*(\psi) = \Pi(\psi) \quad \text{presque sûrement.} \quad \square$$

### Bilan et conclusion (démonstration du théorème 5.6)

Le corollaire 5.11 donne la convergence presque sûre de la suite  $\hat{\Pi}_T^*(\psi)$  vers l'intégrale  $\Pi(\psi)$ , et la proposition 5.9 donne la convergence presque sûre de la différence entre la suite  $\hat{\Pi}_T^{\text{AMIS}}(\psi)$  et la suite  $\hat{\Pi}_T^*(\psi)$  vers 0, ce qui donne la convergence presque sûre de l'estimateur  $\hat{\Pi}_T^{\text{AMIS}}(\psi)$  vers l'intégrale  $\Pi(\psi)$ , d'où la preuve du théorème 5.6.

## 5.5 Applications numériques

Dans cette partie, nous présentons les résultats de deux expérimentations numériques. Nous montrons à travers ces résultats que les bonnes propriétés numériques du schéma AMIS original de Cornuet et al. [21] sont conservées dans le schéma simplifié que nous avons introduit dans la partie 5.4.1. Nous montrons le gain en ESS de notre schéma sur la cible banana-shape de Haario et al. [51] ainsi que des comparaisons d'approximation de la loi *a posteriori* sur un modèle de génétique des populations.

### 5.5.1 Simulations et comparaisons sur une cible jouet

La cible banana-shape introduite dans Haario et al. [51] est la densité d'un vecteur aléatoire obtenu par une transformation non-linéaire d'un vecteur gaussien multivarié. Plus précisément, on applique la transformation  $X_2 = X_2 - b(X_1^2 - \sigma^2)$  sur la deuxième composante d'un vecteur gaussien  $\mathbf{X}$ . Le vecteur  $\mathbf{X} = (X_1, \dots, X_d)$  est de loi  $\mathcal{N}(0_d, \Sigma_d)$  où  $\Sigma_d = \text{diag}(\sigma^2, 1, \dots, 1)$ . Par cette transformation, on obtient un vecteur aléatoire de densité

$$\pi(\mathbf{x}) = f_{\mathcal{N}(0_d, \Sigma_d)}(x_1, x_2 + b(x_1^2 - \sigma^2), x_3, \dots, x_d), \quad (5.25)$$

où  $f_{\mathcal{N}(0_d, \Sigma_d)}(\cdot)$  est la densité associée au vecteur aléatoire  $\mathbf{X}$ . Dans les expérimentations numériques que nous effectuons sur cette cible, nous fixons le paramètre de non-linéarité  $b = 0.03$  et  $\sigma = 10$ . Nous proposons de comparer le schéma original AMIS de Cornuet et al. [21] (voir algorithme 9), le schéma AMIS modifié (voir algorithme 10) que nous avons introduit dans 5.4.1 et qu'on notera AMISII et un troisième schéma d'échantillonnage préférentiel basique qu'on notera IS. Le schéma IS reprend toutes les particules simulées dans les différentes itérations en conservant leurs poids initial  $\pi(\mathbf{x}_i^t) / q(\mathbf{x}_i^t, \hat{\theta}_t)$ . Sur chacun des trois schémas, on construit une suite de lois de proposition  $Q(\hat{\theta}_1), \dots, Q(\hat{\theta}_T)$  via la suite de paramètres  $\hat{\theta}_1, \dots, \hat{\theta}_T$ . Chaque loi de proposition  $Q(\theta)$  est un mélange de lois de Student multivariées *i.e* la densité associée  $q(\mathbf{x}, \theta)$  est donnée par

$$q(\mathbf{x}, \theta) = \sum_{i=1}^k \rho_i t_{\nu_i}(\mathbf{x}, \mu_i, \Sigma_i), \quad (5.26)$$

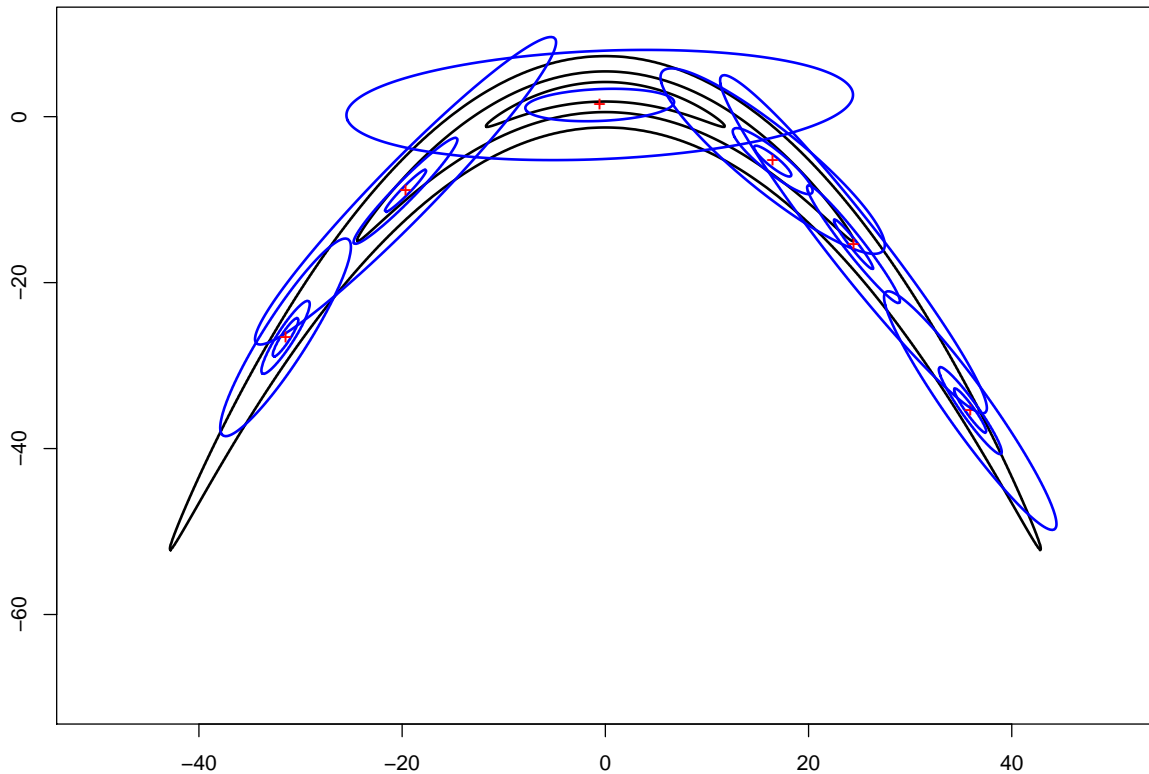
avec  $t_{\nu}(\mathbf{x}, \mu, \Sigma)$ , la densité d'une loi de Student multivariée centrée en  $\mu$  de matrice d'échelle  $\Sigma$  et de degrés de liberté  $\nu$ . L'estimation du paramètre

$$\theta = \left\{ \{\rho_1, \dots, \rho_k\} \subset \mathbb{R}_+ : \sum_{i=1}^k \rho_i = 1, \{\mu_1, \dots, \mu_k\} \subset \mathbb{R}^d, \{\Sigma_1, \dots, \Sigma_k\} \subset \mathcal{M}_+^{d \times d}(\mathbb{R}) \right\},$$

se fait par maximum de vraisemblance à chaque itération de l'un des schémas étudiés. Pour ajuster le mélange (5.26) sur des particules pondérées, nous faisons appel à une version de algorithme EM introduite dans Peel and McLachlan [84].

On présente sur la figure 5.2, la cible (5.25) en dimension deux ainsi qu'une illustration de l'initialisation adoptée dans nos expérimentations. Nous avons fixé les paramètres

$b = 0.03$  et  $\sigma = 10$ . Les trois contours en noir définissent les zones contenant respectivement 50%, 95% et 99.99% de la masse totale. On remarque qu'avec ces paramètres, la cible (5.25) possède de lourdes queues de distributions dans la direction négative de la deuxième composante. D'où la difficulté d'ajuster des mélanges de lois de Student avec des degrés de liberté élevés. Pour avoir un bon recouvrement de la cible, nous avons fixé  $\nu = 3$  et  $k = 6$ . Pour initialiser le paramètre  $\theta$ , nous avons adopté la procédure d'initialisation de Cornuet et al. [21] pour construire un échantillon d'initialisation. Plus précisément, on simule un échantillon initial (de taille 1000 sur l'exemple de la figure 5.2) à partir d'une loi logistique. Cette loi destinée à l'initialisation est calibrée (paramètre d'échelle) dans chaque direction en maximisant le ESS. On applique sur cet échantillon initial une procédure de classification hiérarchique pour extraire  $k$  groupes. Pour constituer les composantes du mélange initial, on calcule les moyennes et les matrices de covariance intra-groupes. Les poids du mélange initial sont tous égaux à  $1/k$ . Nous avons représenté en bleu sur la figure 5.2, les contours des 6 composantes (centrées sur les points en rouge) du mélange initial.



**FIGURE 5.2:** Représentation de la cible banana-shape de Haario et al. [51] avec  $b = 0.03$  et  $\sigma = 10$  et les 6 composantes du paramètre initial.

On présente sur la figure 5.3 les trois comparaisons des schémas AMIS, AMISII et IS sur la cible (5.25) en dimension 2. Sur 100 répliquions de chaque schéma, on compare sur la colonne de gauche: la variance dans la première direction dont la valeur théorique

est 100. Sur la colonne du milieu: la variance dans la deuxième direction dont la valeur théorique est 19. Sur la colonne de droite: la valeur du ESS du système final contenant l'ensemble des particules simulées. Les paramètres à chaque itération sont les suivants :  $k = 6$ ,  $v_1 = v_2 = \dots = 3$  et  $N_t = 5000 + 1000 \times t$  où  $T = 36$ .

On remarque un léger avantage en terme de biais pour le schéma AMIS sur les estimations obtenues sur les colonnes de gauche et du milieu mais les résultats des trois schémas restent comparables en terme de précision. Sur la colonne de droite, nous remarquons que le schéma AMISII améliore la stabilité en terme de ESS par rapport au schéma AMIS de Cornuet et al. [21].

Nous avons restreint l'illustration des expérimentations numériques à la dimension 2. Les comparaisons que nous avons effectuées en dimension 5 sur la cible (5.25) avec les trois schémas précédents sont identiques aux résultats présentés sur la figure 5.3 en dimension deux. Quant aux dimensions supérieures à 8, la calibration du mélange à chaque itération de l'un des schémas est difficile.

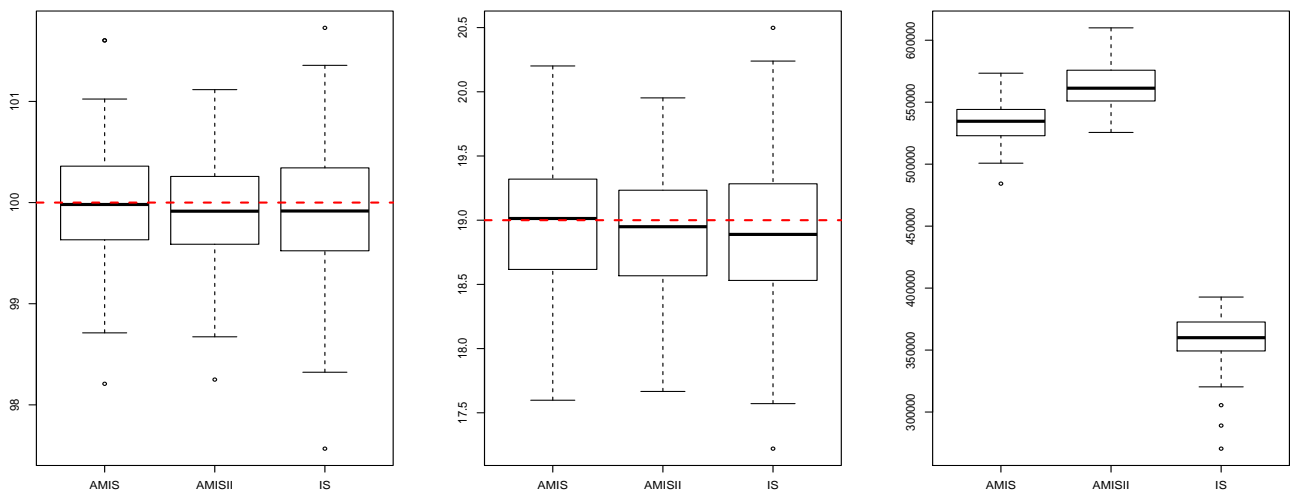


FIGURE 5.3: Comparaison des trois schémas AMIS, AMISII et IS.

## 5.5.2 Simulations et comparaisons sur un modèle de génétique des populations

Ici, on s'intéresse à un modèle de génétique des populations (voir chapitre 2) sur lequel nous confrontons les trois schémas étudiés précédemment. Il s'agit de l'approximation de la loi *a posteriori* sachant un jeu de données simulé. Nous avons réduit le nombre de paramètres du modèle que nous étudions à deux (un taux de migration et un taux de mutation) afin de pouvoir faire les calculs sur la vraisemblance en un temps raisonnable. Le modèle que nous étudions est composé de deux populations. Chacune de ces populations est de taille 10000. Le scénario démographique de ce modèle est restreint à une migration symétrique entre les deux populations. Les taux de migrations sont les mêmes dans les



deux sens (Pop1  $\rightarrow$  Pop2 et Pop2  $\rightarrow$  Pop1). Le modèle mutationnel considéré est de type SMM (*Single Mutation Model*). Le jeu de données qu'on notera  $\mathbf{x}_{\text{obs}}$  est simulé sur 5 locus indépendants. Sur chaque locus, on simule les génotypes des individus à l'aide du logiciel *IBDSim* de Leblois et al. [65]. Pour ce jeu de données, nous avons fixé le taux de mutation  $\gamma_{\text{mut}}$  à 2.3, et le taux de migration symétrique  $\gamma_{\text{mig}} = m_{12} = m_{21}$  à 0.04.

Sur ce modèle, la vraisemblance d'un jeu de données  $\ell(\mathbf{x}|\gamma_{\text{mut}}, \gamma_{\text{mig}})$  est le produit de 5 intégrales (voir chapitre 2). Chaque intégrale représente la vraisemblance du jeu de données en un locus. Dans cette étude, nous faisons appel à des approximations par échantillonnage préférentiel de ces intégrales. Ces approximations sont fournies par le logiciel *Migraine* de Rousset and Leblois [92].

Nous considérons la loi *a priori*  $\pi(\gamma_{\text{mut}}, \gamma_{\text{mig}})$  uniforme sur le support  $\Gamma = [10^{-1}, 10] \times [10^{-3}, 0.5]$ , ce qui simplifie l'expression de la densité de la loi *a posteriori* à

$$\pi(\gamma_{\text{mut}}, \gamma_{\text{mig}}|\mathbf{x}_{\text{obs}}) \propto \ell(\mathbf{x}_{\text{obs}}|\gamma_{\text{mut}}, \gamma_{\text{mig}}) \mathbf{1}_{\Gamma}(\gamma_{\text{mut}}, \gamma_{\text{mig}}). \quad (5.27)$$

Sur ce modèle et pour chacun des schémas considérés, la calibration des lois de proposition

$$q(\mathbf{x}, \hat{\boldsymbol{\theta}}_1), \dots, q(\mathbf{x}, \hat{\boldsymbol{\theta}}_T),$$

est plus simple que précédemment. Chaque densité est une gaussienne multivariée tronquée sur le support  $\Gamma$ . La calibration de la loi de proposition à une itération donnée revient à estimer le paramètre  $\boldsymbol{\theta} = \{\boldsymbol{\mu}_{\text{mut}}, \boldsymbol{\mu}_{\text{mig}}, \boldsymbol{\sigma}_{\text{mut}}^2, \boldsymbol{\sigma}_{\text{mig}}^2\}$  par maximum de vraisemblance, à partir des particules pondérées obtenues dans le passé. Pour évaluer les poids d'importance dans chaque schéma, nous avons reconstruit la loi *a posteriori* à l'aide d'un krigeage (voir figure 5.4) effectué sur une grille  $500 \times 500$  du support  $\Gamma$ . Autrement dit, pour chaque point de cette grille, nous avons approché la vraisemblance à l'aide du logiciel *Migraine*. Ainsi, notre densité *a posteriori* est le résultat de la normalisation de la fonction obtenue par krigeage sur cette grille.

Pour chacune des 100 répliques de chaque schéma, on a  $T = 45$  et  $N_t = N \times t$  avec  $N = 100$ . Soulignons que la suite  $\{N_t\}_{t \geq 1}$  est linéaire en  $t$ , ce qui ne vérifie pas les hypothèses de convergence de la section 5.4.2. Rappelons que chacun des trois schémas étudiés renvoie un système final de particules

$$(\omega_1^1, \mathbf{x}_1^1), \dots, (\omega_{N_1}^1, \mathbf{x}_{N_1}^1), \dots, (\omega_1^T, \mathbf{x}_1^T), \dots, (\omega_{N_T}^T, \mathbf{x}_{N_T}^T).$$

Pour confronter les trois schémas, on s'intéresse à la distance entre la fonction de répartition empirique

$$\hat{F}_{N_{1:T}}(\mathbf{x}) = \sum_{t=1}^T \sum_{i=1}^{N_t} \omega_i^t \mathbf{1}\{\mathbf{x}_i^t \leq \mathbf{x}\}, \quad (5.28)$$

associée au système précédent après normalisation des poids, et la fonction de répartition calculée numériquement à partir du krigeage de la densité  $\pi(\gamma_{\text{mut}}, \gamma_{\text{mig}}|\mathbf{x}_{\text{obs}})$  et qu'on notera  $F(\mathbf{x})$ .

Nous illustrons sur la figure 5.5 les résultats de la comparaison des trois schémas AMIS (boxplot (1) sur chacune des trois colonnes), AMISII (boxplot (2)) et IS (boxplot (3)). Nous

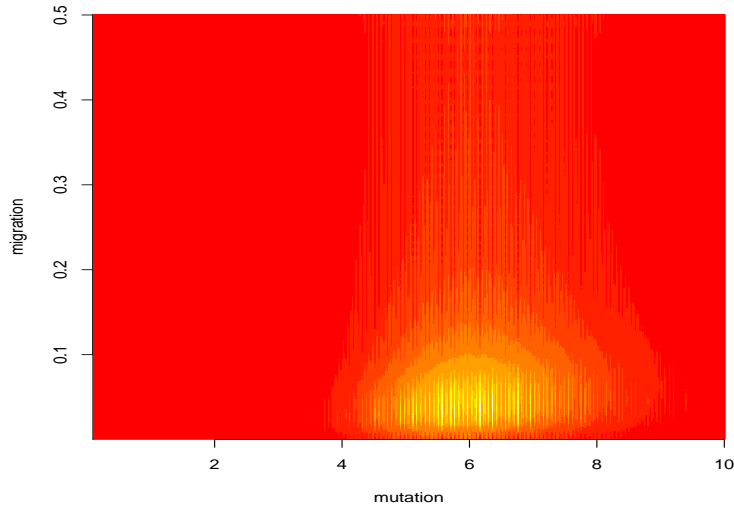


FIGURE 5.4: Représentation de la densité a posteriori  $\pi(\gamma_{mut}, \gamma_{mig} | \mathbf{x}_{obs})$  obtenue par krigeage.

comparons ces schémas via la distance de Cramer Von Mises (figure 5.5 (a)), la distance  $\mathbb{L}^2$  (figure 5.5 (b)) et la distance  $\mathbb{L}^\infty$  (figure 5.5 (c)) entre la fonction de répartition empirique  $\hat{F}_{N_{1:T}}(\mathbf{x})$  et la fonction de répartition  $F(\mathbf{x})$ . Les trois distances sont calculées numériquement sur une grille. Sur les trois distances, le schéma AMISII l’emporte en terme de précision et de stabilité (variance de la distance). Nous remarquons aussi que les distances obtenues par le schéma IS (privé de stratégie de recyclage) sont nettement plus élevées (avec plus de variabilité) que celles obtenues par les deux autres schémas.

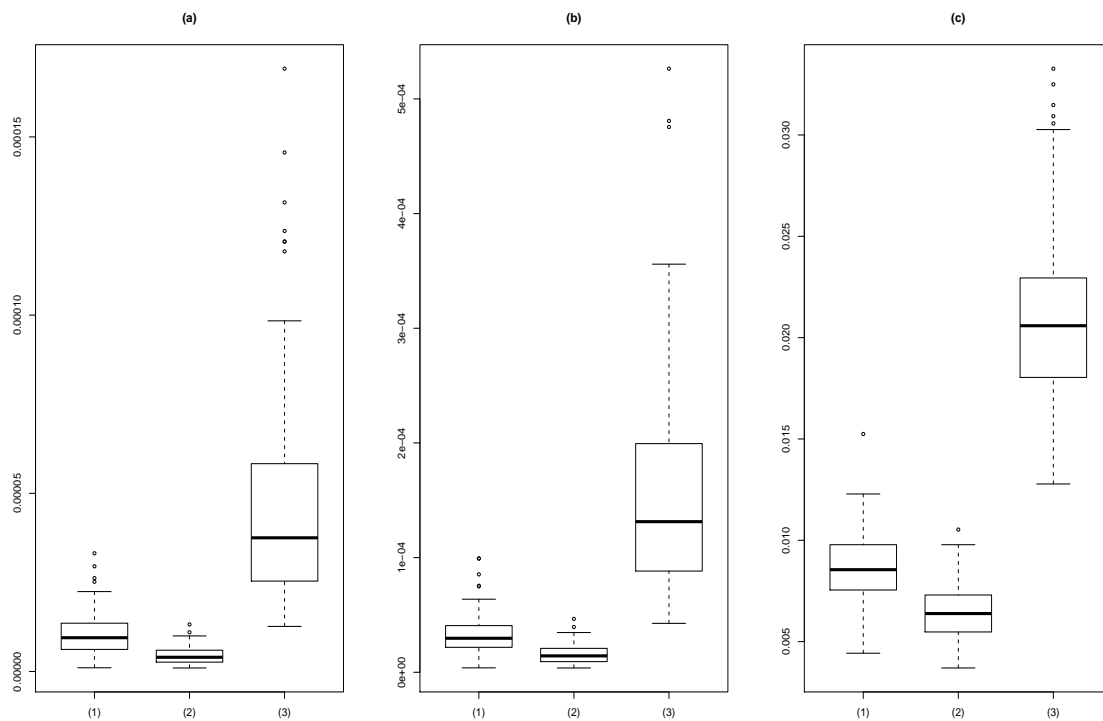


FIGURE 5.5: Comparaison des trois schémas AMIS, AMISII et IS sur les trois distances: Cramer Von-Mises(colonne (a)),  $L^2$  (colonne (b)) et  $L^\infty$  (colonne (c)).

## **Troisième partie**

### **Méthodes bayésiennes approchées**



# 6

## Méthodes bayésiennes approchées

### 6.1 Introduction

Dans un contexte d'estimation bayésienne paramétrique, on s'intéresse à la loi *a posteriori* de  $\theta \in \Theta \subset \mathbb{R}^d$ . Notons  $\pi(\theta)$  la densité *a priori* et  $\ell(\mathbf{x}|\theta)$  la vraisemblance. L'inférence bayésienne se fait en utilisant la densité *a posteriori* donnée par

$$\pi(\theta|\mathbf{x}_{\text{obs}}) \propto \pi(\theta)\ell(\mathbf{x}_{\text{obs}}|\theta),$$

où  $\mathbf{x}_{\text{obs}} \in \mathcal{D} \subset \mathbb{R}^n$  est le jeu de données observé qui peut représenter un échantillon iid dans certains cas. Lorsque le modèle d'intérêt est complexe (voir chapitre 2), le calcul de la vraisemblance  $\ell(\mathbf{x}_{\text{obs}}|\cdot)$  est numériquement très coûteux, voire impossible. Dans cette situation, l'approximation directe de  $\pi(\theta|\mathbf{x}_{\text{obs}})$  est extrêmement compliquée. Les méthodes bayésiennes approchées (dites ABC en anglais pour Approximate Bayesian Computation) se contentent de simuler suivant  $\ell(\cdot|\theta)$  pour approcher la loi *a posteriori*. Ces techniques ont été développées dans la dernière décennie par la communauté des généticiens des populations. Marin et al. [75], Beaumont [6], Csilléry et al. [24] proposent des revues détaillées des méthodes ABC ainsi que leurs utilisation en biologie. Commençons par l'algorithme ABC d'acceptation-rejet de Pritchard et al. [86]. Dans l'algorithme 11, on simule

---

**Algorithme 11** Échantillonneur ABC acceptation-rejet

---

- 1: **Pour**  $i = 1 \rightarrow N$  **faire**
  - 2:   **Répéter**
  - 3:     Simuler  $\theta$  suivant la loi *a priori*  $\pi(\cdot)$
  - 4:     Simuler  $\mathbf{x}$  suivant la vraisemblance  $\ell(\cdot|\theta)$
  - 5:     **jusqu'à**  $d(\eta(\mathbf{x}), \eta(\mathbf{x}_{\text{obs}})) \leq \varepsilon$
  - 6:      $\theta_i \leftarrow \theta$
  - 7: **Fin pour**
- 

une réalisation de  $\theta$  suivant la loi *a priori* et le jeu de données  $\mathbf{x}$  correspondant, suivant le modèle. Puis, le jeu de données simulé est comparé à  $\mathbf{x}_{\text{obs}}$  via un vecteur de statistiques résumées données par l'application  $\eta : \mathcal{D} \rightarrow \mathcal{S}$  et une distance  $d : \mathcal{S} \times \mathcal{S} \rightarrow \mathbb{R}^+$ . Si le jeu de données simulé  $\mathbf{x}$  est à une distance inférieure ou égale à  $\varepsilon$  de  $\mathbf{x}_{\text{obs}}$ , la réalisation  $\theta$  est

retenue. L'algorithme 11 produit des simulations suivant l'approximation ABC de la loi *a posteriori* de densité

$$\pi_\varepsilon(\boldsymbol{\theta}|\mathbf{x}_{\text{obs}}) \propto \int_{\{\mathbf{x} \in \mathcal{D} | d(\eta(\mathbf{x}), \eta(\mathbf{x}_{\text{obs}})) \leq \varepsilon\}} \pi(\boldsymbol{\theta}) \ell(\mathbf{x}|\boldsymbol{\theta}) d\mathbf{x}. \quad (6.1)$$

Les techniques ABC utilisent les statistiques résumées données par l'application  $\eta$  pour réduire la dimension du jeu de données en conservant le maximum d'information. L'idée de ces méthodes est de fournir une approximation précise de la densité *a posteriori*  $\pi_\varepsilon(\boldsymbol{\theta}|\mathbf{x}_{\text{obs}}) \approx \pi(\boldsymbol{\theta}|\mathbf{x}_{\text{obs}})$  en utilisant un niveau d'acceptation  $\varepsilon$  suffisamment bas [11].

Marjoram et al. [78] proposent un schéma MCMC pour échantillonner suivant la cible (6.1) qui ne nécessite aucun calcul de vraisemblance. Pour des petites valeurs du niveau d'acceptation  $\varepsilon$ , l'algorithme 11 et le schéma MCMC de Marjoram et al. [78] sont peu efficaces. Soulignons qu'en pratique, le niveau  $\varepsilon$  n'est pas fixé à l'avance et correspond à un quantile des distances  $d(\eta(\mathbf{x}), \eta(\mathbf{x}_{\text{obs}}))$ . Nous résumons dans l'algorithme 12, l'échantillonneur ABC avec rejet basé sur un quantile des distances.

---

### Algorithme 12 Échantillonneur ABC avec quantile

---

ENTRÉES : Fixer un niveau de quantile  $\alpha \in ]0, 1[$ .

- 1: **Pour**  $i = 1 \rightarrow N$  **faire**
  - 2:     Simuler  $\boldsymbol{\theta}_i$  suivant la loi *a priori*  $\pi(\cdot)$ .
  - 3:     Simuler  $\mathbf{x}_i$  suivant la vraisemblance  $\ell(\cdot | \boldsymbol{\theta}_i)$
  - 4: **Fin pour**
  - 5: Ordonner les couples  $(\boldsymbol{\theta}_1, \mathbf{x}_1), \dots, (\boldsymbol{\theta}_N, \mathbf{x}_N)$  suivant la distance à  $\mathbf{x}_{\text{obs}}$ .
  - 6: Garder seulement les  $\lfloor \alpha N \rfloor$  premiers couples ( $\lfloor a \rfloor$  la partie entière de  $a$ ).
- 

De nombreux schémas SMC (Sequential Monte Carlo en anglais) [39, 28, 68] ont été proposés comme alternatives aux deux techniques précédentes. Parmi ces schémas, on peut citer Sisson et al. [96, 95], Beaumont et al. [7], Drovandi and Pettitt [40] et Del Moral et al. [31]. Ces schémas introduisent une dimension temporelle au problème d'approximation de la loi *a posteriori*. Ils sont initialisés par une approximation ABC fournie par l'échantillonneur 12 avec un niveau d'acceptation  $\varepsilon_0$  élevé. L'idée est de décroître ce niveau à chaque itération (*i.e.*  $\varepsilon_t < \varepsilon_{t-1}$ ). Cela assure l'amélioration successive de la précision de l'approximation ABC. Soulignons que la simulation des particules suivant une approximation ABC est d'autant plus difficile que le niveau d'acceptation  $\varepsilon_t$  est bas.

Notons qu'il existe des versions ABC Parallel-Tempering [3]. Le schéma proposé dans Beaumont et al. [7] correspond à une correction de biais dans le schéma de Sisson et al. [96, 95] et se situe dans la famille des algorithmes PMC [13]. La proposition de Beaumont et al. [7] nécessite l'introduction d'une suite décroissante de seuils d'acceptation  $\varepsilon_0 > \varepsilon_1 > \dots > \varepsilon_T$ . Cette hypothèse est irréaliste en pratique. En revanche, les propositions de Del Moral et al. [31] et Drovandi and Pettitt [40] sont munies de mécanismes de calibration du niveau d'acceptation. Ces deux schémas sont des versions sans vraisemblance (qui ne font pas appel au calcul de celle-ci) de l'échantillonneur SMC de Del Moral et al. [29].

On considère le cas où les temps dûs aux calculs numériques intermédiaires sont négligeables par rapport au temps de simulation des jeux de données suivant le modèle.

C'est le cas typique des modèles complexes de génétique des populations (chapitre 2). Nous proposons d'adapter le schéma SMC de Del Moral et al. [31] en vue de minimiser le nombre de simulations suivant le modèle.

Dans la section 6.2, nous introduisons les schémas SMC proposés par Del Moral et al. [31] et Drovandi and Pettitt [40]. Puis, nous présentons notre schéma d'apprentissage efficace auto-calibré dans la section 6.3. Ces propriétés de convergence sont brièvement discutées dans cette partie ainsi que le schéma d'apprentissage et de calibration des paramètres. Nous présentons les expérimentations numériques sur un modèle complexe de génétique des populations dans la section 6.4. Nous terminons ce chapitre par des comparaisons de stratégies de parallélisation de notre schéma ABC-SMC. Nous détaillons ces comparaisons dans la section 6.5.

## 6.2 Schémas sans vraisemblance séquentiels

Rappelons que le schéma ABC basique d'acceptation-rejet décrit dans l'algorithme 11 consiste à simuler un ensemble de particules  $(\theta_i, \mathbf{x}_i)_{1 \leq i \leq N}$  et conserver les  $\theta_i$  tels que  $d(\eta(\mathbf{x}_i), \eta(\mathbf{x}_{\text{obs}})) \leq \varepsilon$ . Ainsi, on ne retient que les particules  $\{\theta_j\}_{j \in J(\varepsilon, N)}$  où l'ensemble  $J(\varepsilon, N)$  est donné par

$$J(\varepsilon, N) = \left\{ j \in \{1, \dots, N\} : \mathbf{x}_j \in \left\{ \mathbf{x} \in \mathcal{D} \mid \rho(\eta(\mathbf{x}), \eta(\mathbf{x}_{\text{obs}})) \leq \varepsilon \right\} \right\}. \quad (6.2)$$

Lorsque le niveau  $\varepsilon$  est bas, l'ensemble  $J(\varepsilon, N)$  peut être vide. Si le niveau  $\varepsilon$  est élevé, l'approximation  $\pi_\varepsilon(\theta | \mathbf{x}_{\text{obs}})$  est éloignée de la densité *a posteriori*. En pratique, comme dans l'algorithme 12, on choisit un quantile  $\alpha$  de la distance  $d(\eta(\mathbf{x}), \eta(\mathbf{x}_{\text{obs}}))$ . Ce quantile correspond à un choix du niveau  $\varepsilon$  tel que l'ensemble  $J(\varepsilon, N)$  contienne  $\lfloor \alpha N \rfloor$  indices distincts.

Que ce soit dans l'algorithme 11 ou l'algorithme 12, les  $\theta_i$  sont simulés suivant la loi *a priori*. Cette stratégie peut s'avérer très inefficace, particulièrement lorsque la loi *a priori* est non-informative. L'idée des schémas séquentiels est de construire des lois plus pertinentes que la loi *a priori* pour simuler les réalisations du paramètre  $\theta$ . Commençons par une approche simple basée sur le schéma SMC de Del Moral et al. [29].

Dans ce qui suit, on considère que l'ensemble  $\Theta$  est compact et que la loi *a priori* est uniforme sur cet ensemble. La stratégie d'initialisation de tous les schémas séquentiels consiste à générer un système de particules issu d'une loi de densité  $\pi_{\varepsilon_0}(\theta | \mathbf{x}_{\text{obs}})$ . Ce système initial est obtenu en utilisant le schéma ABC d'acceptation-rejet décrit dans l'algorithme 12 et un certain niveau de quantile. Nous supposons qu'à l'itération  $t + 1$ , nous disposons d'un système de particules de taille  $N$  noté  $\mathcal{P}^t = \left\{ (\theta_i^t, \mathbf{x}_i^t), 1 \leq i \leq N \right\}$  issu de la loi  $\pi_{\varepsilon_t}(\theta, \mathbf{x} | \mathbf{x}_{\text{obs}})$  où

$$\pi_{\varepsilon_t}(\theta, \mathbf{x} | \mathbf{x}_{\text{obs}}) \propto \pi(\theta) \ell(\mathbf{x} | \theta) \mathbf{1} \left\{ d(\eta(\mathbf{x}), \eta(\mathbf{x}_{\text{obs}})) \leq \varepsilon_t, \theta \in \Theta \right\}.$$

Une stratégie basique consiste à fixer une suite de quantiles  $\{\alpha_t\}_{t \geq 1}$  (où  $\alpha_0$  est associé



à  $\varepsilon_0$ ), une valeur cible du niveau  $\varepsilon$  et utiliser le schéma que nous décrivons dans l'algorithme 13.

---

**Algorithme 13** Échantillonneur ABC-SMC naïf
 

---

- 1: Ordonner les particules  $\left\{ (\theta_i^t, \mathbf{x}_i^t), 1 \leq i \leq N \right\}$  suivant leurs distances au jeu de données observé  $\mathbf{x}_{\text{obs}}$ , i.e
 
$$d_1 = d(\eta(\mathbf{x}_1), \eta(\mathbf{x}_{\text{obs}})) \leq d_2 = d(\eta(\mathbf{x}_2), \eta(\mathbf{x}_{\text{obs}})) \leq \dots \leq d_N = d(\eta(\mathbf{x}_N), \eta(\mathbf{x}_{\text{obs}})).$$
  - 2: ## Choix de  $\varepsilon_{t+1}$  ##
  - 3: On fixe  $\varepsilon_{t+1}$  tel que  $\varepsilon_{t+1} = d_{\lfloor \alpha_{t+1} N \rfloor}$ .
  - 4: ## Le nouveau système  $\mathcal{P}^{t+1}$  ##
  - 5: Construire un nouveau système de particule  $\mathcal{P}^*$  i.e
    - (i) Dupliquer  $\lfloor 1/\alpha_{t+1} \rfloor$  fois les  $\lfloor \alpha_{t+1} N \rfloor$  premières particules du système  $\mathcal{P}^t$ .
    - (ii) Les  $N - \lfloor \lfloor 1/\alpha_{t+1} \rfloor \alpha_{t+1} N \rfloor$  dernières particules du système  $\mathcal{P}^*$  sont tirées aléatoirement parmi  $\left\{ (\theta_i^t, \mathbf{x}_i^t), 1 \leq i \leq \lfloor \alpha_{t+1} N \rfloor \right\}$ .
  - 6: Appliquer un pas d'un noyau MCMC  $K_{t+1}$  de loi stationnaire  $\pi_{\varepsilon_{t+1}}(\boldsymbol{\theta}, \mathbf{x} | \mathbf{x}_{\text{obs}})$  sur chaque particule  $(\theta_i^*, \mathbf{x}_i^*)$  du système  $\mathcal{P}^*$  et obtenir le nouveau système de particules  $\mathcal{P}^{t+1}$ .
  - 7: **Si**  $\varepsilon_{t+1} < \varepsilon$  **alors**
  - 8: renvoyer le système  $\mathcal{P}^{t+1}$
  - 9: **Sinon**
  - 10:  $t \leftarrow t + 1$  et aller à l'étape 1.
  - 11: **Fin si**
- 

**Mouvements suivant le noyau  $K_{t+1}$**  L'étape 6 de l'algorithme 13 fait appel à un noyau Hastings-Metropolis. Éventuellement, on peut calculer la variance  $\sigma_t^2$  des  $\{\theta_i^t, 1 \leq i \leq N\}$  dans le système  $\mathcal{P}^t$ . Pour chaque particule  $(\theta_i^*, \mathbf{x}_i^*)$  du système  $\mathcal{P}^*$ , on simule alors  $\tilde{\boldsymbol{\theta}}$  suivant une gaussienne  $\mathcal{N}(\theta_i^*, 2\sigma_t^2)$  et  $\tilde{\mathbf{x}}$  suivant  $\ell(\cdot | \tilde{\boldsymbol{\theta}})$ . Puisque la loi *a priori* est uniforme sur le compact  $\boldsymbol{\Theta}$ , la probabilité d'acceptation est binaire pour tout noyau symétrique (voir [31]). Ainsi, si  $\tilde{\boldsymbol{\theta}} \in \boldsymbol{\Theta}$  et  $d(\eta(\tilde{\mathbf{x}}), \eta(\mathbf{x}_{\text{obs}})) \leq \varepsilon_{t+1}$  alors  $(\theta_i^{t+1}, \mathbf{x}_i^{t+1}) = (\tilde{\boldsymbol{\theta}}, \tilde{\mathbf{x}})$ , sinon on ne fait pas de mouvement et ainsi  $(\theta_i^{t+1}, \mathbf{x}_i^{t+1}) = (\theta_i^*, \mathbf{x}_i^*)$ .

**Sans "Rao-Blackwellisation"** On peut remplacer l'étape 5 de l'algorithme 13 par la procédure de ré-échantillonnage suivante: conditionnellement au système  $\mathcal{P}^t$ , le système  $\mathcal{P}^*$  est constitué par tirage avec remise de  $N$  particules parmi les  $\lfloor \alpha_{t+1} N \rfloor$  premières particules du système  $\mathcal{P}^t$ . Si  $N - \lfloor \lfloor 1/\alpha_{t+1} \rfloor \alpha_{t+1} N \rfloor = 0$ , l'étape 5 (i) du schéma 13 correspond à l'espérance conditionnelle de cette alternative [16, 38].

**Qualité des sorties** Les sorties des schémas séquentiels sont constituées de particules pondérées  $(\mathbf{x}_1, \omega_1), \dots, (\mathbf{x}_N, \omega_N)$  associées à une loi cible  $Q$  et telles que

$$\sum_{i=1}^N \omega_i \delta_{\mathbf{x}_i}(A) \rightarrow Q(A),$$

pour tout ensemble  $A$   $Q$ -mesurable. Une façon de mesurer la qualité d'un échantillon pondéré issue de la théorie de l'échantillonnage préférentiel est l'ESS [68, chapitre 2] donné par

$$\text{ESS}\left((\mathbf{x}_1, \omega_1), \dots, (\mathbf{x}_N, \omega_N)\right) = \left(\sum_{i=1}^N \omega_i^2\right)^{-1}. \quad (6.3)$$

Dans de nombreuses situations, le système de particules est obtenu après une étape de ré-échantillonnage suivi d'une attribution automatique d'un poids  $1/N$  pour chaque particule. Un tel système peut être composé d'un petit nombre de particules distinctes répétées un grand nombre de fois. Dans ce cas, on ne peut pas appliquer directement la formule (6.3). En effet, cette dernière donnerait un indicateur ESS égal à  $N$ . Ainsi, pour calculer l'ESS, il ne faut tenir compte que des particules distinctes avec poids agrégé. Par exemple, si une particule  $\mathbf{x}^*$  est répétée  $r$  fois avec un poids  $1/N$ , son poids après agrégation est  $r/N$ . Dans le cas où il y a des répétitions, un autre indicateur de la qualité d'un échantillon pondéré est le nombre de particules uniques. Si les particules sont répétées le même nombre de fois, il est égal à l'ESS. Cet indicateur est un majorant de l'ESS. L'écart entre ces deux mesures est d'autant plus grand que l'entropie des poids est forte.

Si aucun mouvement n'est accepté à l'étape 6 de l'algorithme 13, le système final noté  $\mathcal{P}^\top$  est composé de  $\alpha_1 \alpha_2 \dots \alpha_\top N$  particules distinctes ayant toutes le même poids. Chacune de ces particules est répétée un grand nombre de fois dans le système final. Autrement dit, c'est équivalent à une réalisation de l'algorithme 12 en considérant le niveau  $\alpha = \alpha_1 \alpha_2 \dots \alpha_\top$ . Ainsi, on obtient une valeur de l'ESS (Effective Sample Size) qui correspond au nombre de particules distinctes

$$\text{ESS}_\top = \alpha_0 \alpha_1 \alpha_2 \dots \alpha_\top N,$$

qui peut être de très mauvaise qualité. Supposons maintenant qu'à la  $t^{\text{ème}}$  itération, une proportion  $\rho_t$  de particules est acceptée à l'étape 6. Ainsi, le nombre de particules distinctes est donné par

$$\alpha_0(\alpha_1 + \rho_1)(\alpha_2 + \rho_2) \dots (\alpha_\top + \rho_\top)N.$$

Détaillons cette approximation, supposons qu'à l'itération  $t$ , le système  $\mathcal{P}^t$  contienne  $N_t$  particules distinctes et supposons que  $\alpha_t = 1/\ell_t$  où  $\ell_t$  est un entier diviseur de  $N_t$ . Supposons aussi que chaque particule  $(\tilde{\theta}, \tilde{\mathbf{x}})$  proposée à l'étape 6 est acceptée dans le système  $\mathcal{P}^{t+1}$  avec probabilité  $\rho_t$  indépendamment des autres particules. Rappelons que le système  $\mathcal{P}^*$  est composé de  $\ell_t$  copies des  $[\alpha_t N]$  premières particules de  $\mathcal{P}^t$ . Ainsi le nombre moyen de particules distinctes dans le système  $\mathcal{P}^{t+1}$  est donné par

$$\alpha_t N_t + \sum_{i=1}^{\ell_t} \alpha_t \rho_t N_t = \alpha_t N_t + \ell_t \alpha_t \rho_t N_t = (\alpha_t + \rho_t) N_t.$$

**Schémas ABC-SMC Del Moral et al. [31] et Drovandi and Pettitt [40]** Del Moral et al. [40] proposent un algorithme adaptatif où l'étape 3 du schéma naïf précédent est remplacée par une étape de calibration qui consiste à contrôler l'ESS. Ils proposent en effet de choisir la valeur  $\varepsilon_{t+1}$  telle que

$$\text{ESS}_{t+1} = \alpha \text{ESS}_t.$$

avec  $\alpha \in ]0, 1[$ . En pratique, ils regardent le nombre de particules issues de l'étape  $t$  qui respectent la condition selon laquelle la distance à l'observation est inférieure à  $\varepsilon_{t+1}$ . Ils assimilent cette quantité à l'ESS et ne tiennent pas compte des répétitions dans le système. Si  $\alpha$  est proche de 1, les cibles successives sont très similaires et le temps d'atteinte du niveau souhaité est lent. Par contre, la qualité des approximations successives est bonne. A l'inverse si  $\alpha$  est proche de 0, la stabilisation vers le niveau souhaité est très rapide mais la qualité des approximations successives est mauvaise. Dans les exemples qu'ils traitent, ils fixent la valeur de  $\alpha$  entre 0.9 et 0.99. Ainsi, le schéma n'est pas optimisé pour limiter le nombre de simulations suivant le modèle. Par ailleurs, pour une même valeur de paramètre ils proposent de simuler  $M$  jeux de données et considérer la cible ABC augmentée suivante

$$\pi_\varepsilon(\mathbf{x}_{1:M}, \boldsymbol{\theta}) \propto \pi(\boldsymbol{\theta}) \prod_{k=1}^M \ell(\boldsymbol{\theta} | \mathbf{x}_k) \frac{1}{M} \sum_{k=1}^M \mathbf{1}\left\{d(\eta(\mathbf{x}_k), \eta(\mathbf{x}_{\text{obs}})) \leq \varepsilon\right\}.$$

Minimiser le nombre d'appels au modèle n'est pas leur préoccupation. Enfin, notons qu'avant l'étape de diversification par le noyau, ils n'effectuent une étape de ré-échantillonnage que si l'ESS est inférieur à un certain seuil.

Drovandi and Pettitt [40] utilisent quant à eux le schéma naïf décrit dans l'algorithme 13 en fixant la suite des niveaux  $\alpha_t$  à une valeur donnée  $1 - \alpha$  et en modifiant seulement les  $\lfloor \alpha N \rfloor$  particules ayant les distances les plus éloignées de l'observation. Ce schéma n'est clairement pas calibré pour minimiser le nombre d'appels aux simulations suivant le modèle tout en assurant des bonnes approximations.

### 6.3 Notre proposition auto-calibrée

L'idée de notre proposition est de choisir le plus petit  $\varepsilon_{t+1}$  tel que  $(\alpha_t + \rho_t) \geq 1$ . En effet, plus  $\varepsilon$  est petit, meilleure est l'approximation ABC. Par ailleurs, comme expliqué précédemment, si l'on arrive à calibrer  $\alpha_t + \rho_t = 1$ , le nombre de particules distinctes à l'itération  $t + 1$  vaut  $N_t$ . Elles sont en moyenne toutes répétées le même nombre de fois et donc l'ESS moyen vaut  $N_t$ . Pour munir l'algorithme 13 de cette procédure, nous permutons l'étape 3 (dans laquelle nous fixons le seuil  $\varepsilon_{t+1}$ ) et l'étape 6 (durant laquelle on peut calculer  $\rho_t$ ). Le noyau  $K_t$  de l'étape 6 dépend étroitement du seuil  $\varepsilon_{t+1}$  fixé à l'étape 3. Ce qui complique la permutation des deux étapes. Comme dans le schéma SMC basique décrit dans l'algorithme 13, nous supposons que nous disposons à l'étape  $t + 1$  d'un système de particules  $\mathcal{P}^t = \{(\boldsymbol{\theta}_i^t, \mathbf{x}_i^t), 1 \leq i \leq N\}$  issues de la loi  $\pi_{\varepsilon_t}(\boldsymbol{\theta}, \mathbf{x} | \mathbf{x}_{\text{obs}})$ . L'algorithme 14 résume notre schéma ABC-SMC auto-calibré en décrivant l'itération  $t + 1$ .

**Le schéma adaptatif et le niveau d'acceptation final** Clairement,  $\rho'(\alpha')$  est une fonction croissante de  $\alpha'$ . En effet, lorsque  $\alpha'$  augmente, la condition

$$\tilde{\theta} \in \Theta \quad \text{et} \quad d(\eta(\tilde{\mathbf{x}}), \eta(\mathbf{x}_{\text{obs}})) \leq \varepsilon'(\alpha')$$

devient moins restrictive et le noyau Hastings-Metropolis diversifie de plus en plus de particules. Ainsi,  $\alpha' + \rho'(\alpha')$  augmente en fonction de  $\alpha'$ .

Abordons deux questions naturelles. Premièrement, quel est le comportement du niveau d'acceptation adapté à chaque itération quand  $N$  tend vers l'infini ? Deuxièmement, que peut-on dire du niveau d'acceptation de sortie de notre schéma ?

Quand le nombre de particules par itération  $N$  tend vers l'infini, le niveau  $\varepsilon_t$  ne décroît pas. En effet,  $\varepsilon_t$  converge vers une quantité positive dépendant seulement de la densité *a priori*  $\pi(\theta)$  et de la vraisemblance  $\ell(\mathbf{x}|\theta)$ . À chaque itération de l'algorithme 14, le triplet  $(\alpha, \rho, \varepsilon)$  vit dans une variété de dimension une. En supposant que cette variété est paramétrée par  $\varepsilon$ , la procédure de calibration choisit le seuil  $\varepsilon$  tel que  $\alpha(\varepsilon) + \rho(\varepsilon) = 1$ . Considérons l'itération  $t+1$  décrite précédemment, lorsque  $N$  tend vers l'infini la loi empirique associée au système de particules  $\mathcal{P}^t$  converge vers la loi jointe  $\pi_{\varepsilon_t}(\theta, \mathbf{x}|\mathbf{x}_{\text{obs}})$ . Ainsi,

$$\alpha(\varepsilon) = \iint \pi_{\varepsilon_t}(\theta) \ell(\mathbf{x}|\theta) \mathbf{1}\left\{d(\eta(\mathbf{x}), \eta(\mathbf{x}_{\text{obs}})) \leq \varepsilon\right\} d\mathbf{x} d\theta,$$

pour tout  $\varepsilon > 0$  et l'étape de rejet de l'algorithme 14 retient une proportion  $\alpha(\varepsilon)$  des  $\theta_i$ . La loi empirique associée converge vers

$$\pi_\varepsilon(\theta) = \alpha(\varepsilon)^{-1} \iint \pi_{\varepsilon_t}(\theta) \ell(\mathbf{x}|\theta) \mathbf{1}\left\{d(\eta(\mathbf{x}), \eta(\mathbf{x}_{\text{obs}})) \leq \varepsilon\right\} d\mathbf{x},$$

pour tout  $\varepsilon > 0$ . Asymptotiquement, le taux d'acceptation  $\rho(\varepsilon)$  du noyau Hastings-Metropolis  $K_{t+1}$  est donné par

$$\rho(\varepsilon) = \iiint \pi_\varepsilon(\theta) K_{t+1}(\theta^*|\theta) \ell(\mathbf{x}^*|\theta^*) \mathbf{1}\left\{d(\eta(\mathbf{x}^*), \eta(\mathbf{x}_{\text{obs}})) \leq \varepsilon\right\} d\mathbf{x}^* d\theta^* d\theta.$$

On en déduit que le niveau d'acceptation  $\varepsilon_{t+1}$  obtenu par notre méthode de calibration est asymptotiquement solution de l'équation

$$\begin{aligned} & \iint \pi_{\varepsilon_t}(\theta) \ell(\mathbf{x}|\theta) \mathbf{1}\left\{d(\eta(\mathbf{x}), \eta(\mathbf{x}_{\text{obs}})) \leq \varepsilon\right\} d\mathbf{x} d\theta \\ & + \iiint \pi_\varepsilon(\theta) K_{t+1}(\theta^*|\theta) \ell(\mathbf{x}^*|\theta^*) \mathbf{1}\left\{d(\eta(\mathbf{x}^*), \eta(\mathbf{x}_{\text{obs}})) \leq \varepsilon\right\} d\mathbf{x}^* d\theta^* d\theta = 1. \end{aligned}$$

Ainsi donc, les paramètres calibrés durant les différentes itérations sont des estimateurs de quantités positives qui dépendent uniquement de la loi *a priori* et de la vraisemblance. Une preuve complète et formelle consisterait à montrer la convergence de ces quantités empiriques. Les paramètres calibrés convergent quand on atteint le régime asymptotique ( $N$  grand) (voir illustration sur les figures 6.1 et 6.2). Quand on lance plusieurs instances indépendantes du schéma, l'instant d'atteinte du critère d'arrêt ne varie pas significativement. Cette convergence est illustrée sur l'exemple de Sisson et al. [96] dans la partie 6.4.1.

Ainsi, en prenant un grand nombre de particules  $N$  à chaque itération, les paramètres calibrés peuvent être considérés comme déterministes.

La deuxième question est sur la qualité du seuil  $\varepsilon$  final quand le nombre de particules  $N$  tend vers l'infini. En admettant que le niveau d'acceptation final de notre schéma ABC-SMC ne varie pas significativement d'une réalisation à une autre. L'approximation ABC donnée par le niveau  $\varepsilon_T$  de sortie (où  $T$  est l'instant d'arrêt) peut ne pas être suffisamment bas pour fournir une bonne approximation de la loi *a posteriori*. En effet, il se peut que le niveau d'acceptation final  $\varepsilon$  soit élevé (il dépend du critère d'arrêt de l'algorithme 14). Il suffit d'appliquer une étape supplémentaire d'acceptation-rejet sur le système final renvoyé par le schéma séquentiel pour atteindre le niveau souhaité. Ajouter une étape d'acceptation-rejet est numériquement moins coûteux que d'effectuer des itérations supplémentaires dans le schéma séquentiel en modifiant le critère d'arrêt. Dans une étape d'acceptation-rejet, l'ESS du système de particules dépend linéairement de la taille de celui-ci. Notons que l'approximation ABC de la loi *a posteriori* à partir du système final de particules s'obtient à l'aide des méthodes classiques (e.g estimateur à noyau, régression linéaire locale etc). Ainsi donc, les caractéristiques de la loi *a posteriori* (moyenne, médiane et différents quantiles) peuvent être obtenues par les estimateurs classiques. La qualité de ces estimations dépend étroitement de l'ESS du système de particules. L'augmentation de  $N$  augmente la qualité et la précision de ces différents estimateurs. Lorsqu'on vise une bonne qualité d'approximation ABC, on peut répartir le coût de calcul en lançant différentes instances indépendantes de notre schéma ABC-SMC en parallèle. Les différentes sorties de ces instances sont fusionnées. Le niveau d'acceptation  $\varepsilon$  de l'échantillon résultant correspond au plus grand niveau des systèmes fusionnés.

**Apprentissage efficace** Les techniques séquentielles visent à apprendre progressivement la région de l'espace paramétrique  $\Theta$  la plus pertinente pour la simulation des  $\theta$ , et cela sans introduire de biais sur l'approximation ABC de la loi *a posteriori*. La simulation des particules  $(\theta, \mathbf{x})$  suivant la loi jointe  $\pi(\theta)\ell(\mathbf{x}|\theta)$  peut se révéler inefficace. La simulation des  $\theta$  suivant la loi *a priori* génère des jeux de données éloignés de  $\mathbf{x}_{\text{obs}}$ . Dans notre schéma ABC-SMC auto-calibré, on apprend sur les particules  $(\theta, \mathbf{x})$  simulées dans le passé et on construit une nouvelle approximation ABC en faisant décroître progressivement le seuil  $\varepsilon$ . Ainsi, si la loi *a priori* est proche de la loi *a posteriori*, cet apprentissage est inutile (voir discussion sur l'exemple de Sisson et al. [96] dans la partie 6.4.1). Soulignons que l'application d'un pas d'un noyau Hastings-Metropolis peut être interprétée comme une simulation de  $\theta$  suivant un estimateur à noyau de l'approximation ABC de la loi *a posteriori*. Ainsi, le paramètre  $\theta$  d'une particule  $(\theta, \mathbf{x})$  proposée durant cette étape provient du mélange gaussien

$$\frac{1}{\alpha_t N} \sum_{i=1}^{\lfloor \alpha_t N \rfloor} \varphi\left(\mathbf{b}_t^{-1}(\theta - \theta_i^t)\right),$$

où  $\mathbf{b}_t = \sqrt{2\sigma_t^2}$  est le paramètre de lissage de cet estimateur à noyau. Ce mélange correspond à un estimateur à noyau basé sur  $\{\theta_i^t, 1 \leq i \leq \lfloor \alpha_t N \rfloor\}$  et dont la loi empirique approche  $\pi_{\varepsilon_t}$ .

**Comparaison avec l’algorithme basique d’acceptation-rejet** Notre schéma ABC-SMC présenté dans l’algorithme 14 renvoie un système final de particules noté  $\{(\theta_i^T, \mathbf{x}_i^T), 1 \leq i \leq N\}$  avec un niveau d’acceptation final  $\varepsilon_T$ . Nous considérons ce système comme un échantillon iid de taille  $ESS_T$ . Dans le contexte d’inférence en génétique des populations, le coût numérique est essentiellement dû à la simulation des jeux de données suivant le modèle. Nous comparons les seuils d’acceptation renvoyés par notre schéma et le schéma d’acceptation-rejet présenté dans l’algorithme 11. Nous effectuons cette comparaison pour le même effort de calcul et la même qualité en terme de ESS. Notons  $N_T$  le nombre total de simulations suivant le modèle nécessaires pour notre schéma. Le même coût de calcul dans l’algorithme d’acceptation-rejet revient à simuler  $N_T$  couples  $(\theta, \mathbf{x})$  à partir de loi jointe  $\pi(\theta)\ell(\mathbf{x}|\theta)$ . On peut comparer  $\varepsilon_T$  au quantile noté  $\varepsilon_{\text{acc-rej}}$  de niveau  $ESS_T/N_T$  associé à la variable aléatoire  $D = d(\eta(\mathbf{x}), \eta(\mathbf{x}_{\text{obs}}))$  où  $(\theta, \mathbf{x})$  est de densité jointe  $\pi(\theta)\ell(\mathbf{x}|\theta)$ . Pour mesurer l’efficacité de notre schéma par rapport à l’algorithme d’acceptation-rejet, nous calculons son facteur de gain. On calcule le nombre  $N_{\text{acc-rej}}$  de particules nécessaires lorsque l’on utilise l’échantillonneur ABC acceptation-rejet pour fournir une approximation de niveau  $\varepsilon_T$  en acceptant  $ESS_T$  particules. Le facteur de gain de notre schéma ABC-SMC en terme de nombre de simulations suivant le modèle est donné par

$$r_T = \frac{N_{\text{acc-rej}}}{N_T}. \quad (6.4)$$

Notre schéma gagne en terme de simulations suivant le modèle face à l’algorithme acceptation-rejet lorsque le facteur de gain  $r_T$  est supérieur à un.

**Étude de convergence** Quand la suite décroissante de seuils  $\{\varepsilon_t\}_{t \geq 1}$  est fixée, la suite de densités  $\{\pi_{\varepsilon_t}\}_{t \geq 1}$  est déterministe. Ainsi, si la suite de noyaux markovien  $\{K_t\}_{t \geq 1}$  est déterministe, il suffit d’appliquer les résultats de Del Moral et al. [29] sur les schémas SMC pour obtenir des résultats de convergence.

Si l’on considère le cas où les noyaux markoviens sont construits à partir des particules issues des itérations passées, les travaux récents de Del Moral et al. [30] permettent éventuellement de montrer des résultats de convergence. Nous pouvons également appliquer dans ce cas les théorèmes limites sur les tableaux triangulaires [37, 15]. Ces théorèmes sont énoncés dans le chapitre 4.

Notons par  $\{\xi_{N,i}^t = (\theta_i^t, \mathbf{x}_i^t), 1 \leq i \leq N\}$  le système de particules obtenu à la fin de la  $t^{\text{ème}}$  itération. Introduisons la sous-tribu  $\mathcal{F}_N^{t-1} = \sigma(\xi_{N,i}^{t-1}, 1 \leq i \leq N)$ . Conditionnellement à la tribu  $\mathcal{F}_N^t$ , les variables aléatoires  $\{\xi_{N,i}^t, 1 \leq i \leq N\}$  sont indépendantes. Ainsi, sous les hypothèses d’application des théorèmes 4.1 et 4.2, nous avons la loi des grands nombres et le TCL suivants.

- Si  $\int |\varphi(\xi)| \pi_{\varepsilon_t}(d\xi) < +\infty$ , quand  $N$  tend vers l’infini, nous avons la convergence

$$\frac{1}{N} \sum_{t=1}^N \varphi(\xi_{N,i}^t) \longrightarrow \pi_{\varepsilon_t}(\varphi)$$

en probabilité où  $\pi_{\varepsilon_t}(\varphi) = \int \varphi(\xi) \pi_{\varepsilon_t}(d\xi)$ .

- Quand  $N$  tend vers l'infini, nous avons la convergence

$$\frac{1}{\sqrt{N}} \left( \sum_{i=1}^N \varphi(\xi_{N,i}^t) - N\pi_{\varepsilon_t}(\varphi) \right) \rightarrow \mathcal{N}(0, \sigma_t^2(\varphi)),$$

en loi où  $\sigma_t^2(\varphi) = \int \varphi^2(\xi) \pi_{\varepsilon_t}(d\xi) - \pi_{\varepsilon_t}(\varphi)^2$ .

## 6.4 Expérimentations numériques

### 6.4.1 Exemple jouet

On considère ici l'exemple jouet introduit par Sisson et al. [96] et largement repris dans la littérature. C'est un mélange de deux gaussiennes de variances distinctes et une moyenne commune inconnue,

$$\begin{cases} \boldsymbol{\theta} \sim \mathcal{U}_{[-10,10]}, \\ \ell(\mathbf{x}|\boldsymbol{\theta}) = \frac{1}{2}\phi(\mathbf{x}; \boldsymbol{\theta}, 1) + \frac{1}{2}\phi(\mathbf{x}; \boldsymbol{\theta}, 1/100), \end{cases} \quad (6.5)$$

où  $\mathcal{U}_{[-10,10]}$  est la loi uniforme sur l'intervalle  $[a, b]$  et  $\phi(\cdot; \mu, \sigma^2)$  la densité de probabilité associée à une variable aléatoire gaussienne de moyenne  $\mu$  et de variance  $\sigma^2$ . Comme il est mentionné dans Beaumont et al. [7], Del Moral et al. [31], la loi *a posteriori* associée à l'observation  $\mathbf{x}_{\text{obs}} = 0$  est donnée par sa densité

$$\pi(\boldsymbol{\theta}|\mathbf{x}_{\text{obs}}) \propto \left\{ \phi(\boldsymbol{\theta}; 0, 1) + \frac{1}{2}\phi(\boldsymbol{\theta}; 0, 1/100) \right\} \mathbf{1}\left\{ -10 \leq \boldsymbol{\theta} \leq 10 \right\}.$$

Cet exemple est suffisamment simple pour que l'on puisse comparer les différents schémas séquentiels.

Le tableau 6.1 présente une comparaison de notre schéma ABC-SMC à ceux proposés par Del Moral et al. [31] et Drovandi and Pettitt [40] en terme de coût total en nombre de simulations suivant le modèle nécessaire pour obtenir le niveau d'acceptation  $\varepsilon = 0.09$ . Les schémas proposés par Del Moral et al. [31] et Drovandi and Pettitt [40] ne sont pas conçus pour optimiser le nombre de simulations suivant la vraisemblance. Nous remarquons un net avantage pour notre proposition.

TABLE 6.1: Comparaison des coûts de calcul des différents schémas SMC.

	Coût total	ESS	Niveau final $\varepsilon$
[31]	$46 \times 10^5$	29250	0.09
[40]	$109 \times 10^5$	32000	0.09
Notre schéma ABC-SMC	$23 \times 10^5$	33285	0.09

Nous avons appliqué notre schéma avec  $10^5$  particules par réalisation. Les figures 6.1 et 6.2 représentent les itérations 1 et 8 de notre schéma ABC-SMC. Elles montrent la convergence du niveau d'acceptation calibré lorsque  $N$  augmente. Nous remarquons aussi que la variance du niveau  $\varepsilon$  est proportionnelle à  $1/N$ .

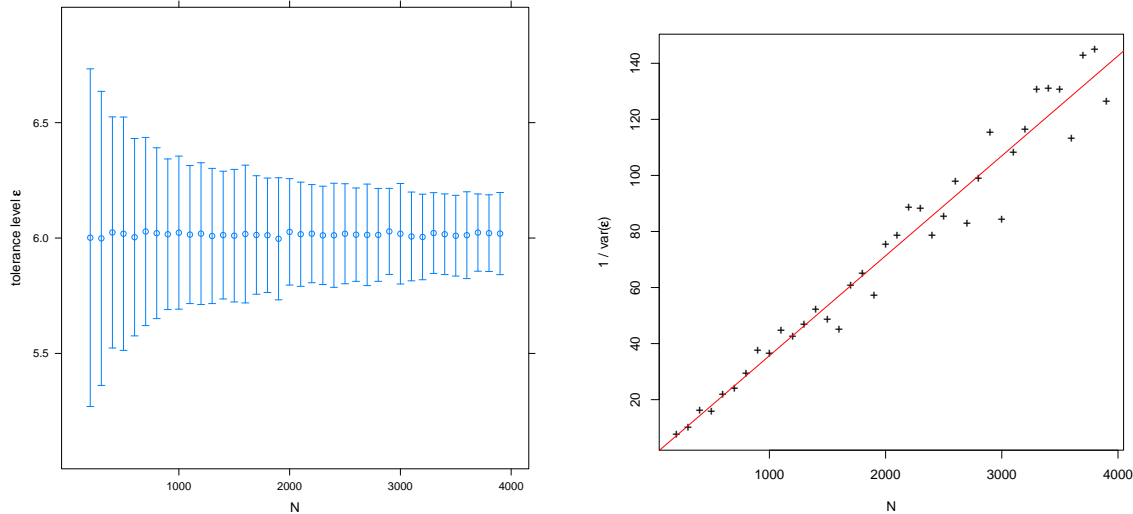
Nous avons mentionné précédemment que les schémas séquentiels sont déconseillés lorsque la loi *a priori* apporte suffisamment d'information sur le paramètre. La figure 6.3 (les deux courbes du haut) montre que lorsque le support de la densité *a priori* est étroit autour de 0, le critère d'arrêt est réalisé dès les premières itérations et le facteur de gain n'est pas favorable à l'emploi du schéma séquentiel. Les quatre courbes de cette figure présentent l'évolution du facteur de gain présenté en (6.4) sur les différentes itérations. La ligne en pointillés verticale (en bleu) correspond à l'instant d'arrêt. Sur chacune des quatre études, on initialise notre schéma avec un système de particules obtenu par acceptation-rejet avec un quantile de niveau  $\alpha_0$ . Ainsi, nous déduisons que notre schéma ABC-SMC est nettement plus efficace et économe en simulations suivant le modèle quand la densité *a priori* contient peu d'information (voir figure 6.3: courbes du bas).

Quand le support de la densité *a priori* est très large, le schéma séquentiel consacre de nombreuses itérations pour détecter la région pertinente pour la simulation des particules. Ceci peut s'avérer très inefficace et peut impacter négativement le facteur de gain. Le choix du niveau du quantile  $\alpha_0$  est important dans ce genre de situation. Un bon choix de  $\alpha_0$  et donc de  $\varepsilon_0$  permet d'éliminer les régions dans lesquelles l'apprentissage est inutile. La première approximation ABC  $\pi_{\varepsilon_0}(\theta|\mathbf{x}_{\text{obs}})$  attribue une très faible masse à ces parties du support de la densité *a priori*. Nous illustrons ce choix de  $\alpha_0$  sur la figure 6.3 (les courbes du bas). Avec un support  $[-100; 100]$  de la densité *a priori* et  $\alpha_0 = 1/4$  le facteur de gain maximal est de 14. Ce choix de  $\alpha_0$  s'obtient à partir d'un système de particules issues de la loi jointe  $\pi(\theta)\ell(\theta|\mathbf{x})$ . On choisit  $\alpha_0$  tel que la variance des  $\theta_i$  retenus après une étape d'acceptation-rejet soit significativement réduite. Plus précisément, avec le choix  $\alpha_0 = 1/4$ , on réduit d'un facteur de 4 l'écart type de  $\theta$  où  $(\theta, \mathbf{x})$  est de loi jointe  $\pi(\theta)\ell(\theta|\mathbf{x})$ .

## 6.4.2 Exemple de génétique des populations

On souhaite également étudier les performances de notre schéma sur un modèle de génétique des populations (chapitre 2). Le problème d'inférence dans ces modèles est à l'origine du développement des techniques ABC. La figure 6.4 synthétise la phylogéographie de l'abeille européenne (*Apis mellifera*) depuis son aire d'origine (Asie orientale). On suppose qu'il y a eu deux voies d'invasion : l'une contournant les Alpes par le Nord et l'autre par le Sud. Cette première division est modélisée par la divergence à la date  $t_1$ , dont l'ordre de grandeur supposée est le demi million d'années. L'abeille présente en Italie (*Apis mellifera ligustica*) est un mélange entre la lignée *Apis mellifera mellifera* présente par exemple sur la côte occidentale française et la lignée *Apis mellifera carnica* que l'on retrouve en Europe du Sud Est. Ce mélange est modélisé par une admixture à la date  $t_3$ . Les trois sous-espèces apparaissent dans nos échantillons : Pop1 a été échantillonnée dans les Landes (France) pour représenter *Apis mellifera mellifera*, Pop2 a été échantillonnée en Lombardi (Italie) pour représenter *Apis mellifera ligustica* et Pop4 en Croatie pour représenter *Apis mellifera carnica*. Enfin, la dernière population échantillonnée provient des ruches





**FIGURE 6.1:** À gauche: intervalles de confiance de niveau 0.95 du seuil d'acceptation  $\varepsilon$  calibré à la première itération avec différentes valeurs de  $N$ . À droite: l'inverse de la variance du seuil  $\varepsilon$  calibré en fonction de  $N$ . Ces variances et ces intervalles de confiance sont calculés sur 250 réalisations indépendantes.

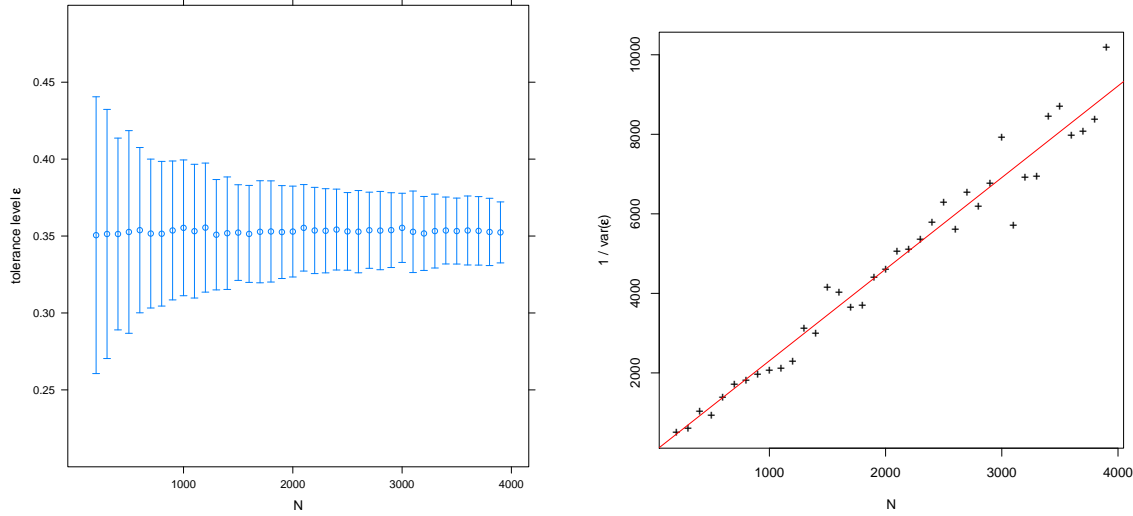
de Courmayeur (Val d'Aoste). L'abeille résidente ici est plutôt *Apis mellifera*, mais a été enrichie récemment en gènes d'*Apis mellifera ligustica*, suite à l'introduction répétée de reines du centre de l'Italie par les apiculteurs, et modélisée par l'admixture à la date  $t_5$ .

Résumons le scénario synthétisé sur la figure 6.4. Il est composé de cinq événements inter-populationnels respectivement datés en  $t_1, \dots, t_5$  du plus ancien au plus récent. Il est composé de deux types d'événements: trois divergences paramétrées par leurs dates  $t_1, t_2, t_4$  et deux admixtures paramétrées par leurs dates  $t_3, t_5$  et les taux  $r_1$  et  $r_2$ .

Ce scénario implique six populations qui ont divergées aux différentes dates d'évènements décrits précédemment. Ce scénario inclut aussi deux populations non observées. On distingue l'évolution d'une population sur la figure 6.4 par une couleur différente des autres.

Le jeu de données de cette étude est composé de différents échantillons d'individus prélevés sur les quatre populations observées dans le présent et notées  $Pop1, \dots, Pop4$  sur la figure 6.4. Chaque échantillon est de l'ordre de la cinquantaine. Nos données sont constitués par le génotypage de chaque individu en huit locus microsatellite indépendants. Nous considérons 30 statistiques résumées pour mener l'analyse ABC. Elles sont sélectionnées parmi les statistiques proposées dans le logiciel *DIYABC* de Cornuet et al. [22]. En utilisant l'approximation ABC avec acceptation-rejet et la régression locale de Beaumont et al. [4], ce logiciel permet d'analyser une large classe de scénarios. La figure 6.4 a été également obtenue à l'aide de ce logiciel.

Le modèle GSM est considéré dans l'étude. Le taux de mutation est donné par unité de temps et par individu. Pour le  $i^{\text{ème}}$  locus de l'étude, le taux de mutation total se décompose



**FIGURE 6.2:** À gauche: intervalles de confiance de niveau 0.95 du seuil d'acceptation  $\varepsilon$  calibré à la huitième itération avec différentes valeurs de  $N$ . À droite: l'inverse de la variance du seuil  $\varepsilon$  calibré en fonction de  $N$ . Ces variances et ces intervalles de confiance sont calculés sur 250 réalisations indépendantes.

en deux parties comme suit

$$\mu_{i,GLO} = \mu_i + \mu_{i,SNI},$$

où  $\mu_i$  correspond à l'augmentation ou la diminution de la séquence d'ADN par une longueur  $G \times m$  où  $G$  une réalisation d'une loi géométrique de paramètre 0.42 et  $m$  la longueur du motif (voir le chapitre 2 pour plus de détails). Le paramètre  $\mu_{i,SNI}$  correspond à l'augmentation ou la diminution de la séquence d'une seule longueur.

**Lois a priori** Notons par  $Ne_1, \dots, Ne_6$ , les tailles efficaces associées respectivement aux six populations impliquées dans le scénario. Cette taille est différente d'une population à une autre. Rappelons que les événements inter-populationnels du scénario sont datés à  $t_1, \dots, t_6$ . Ainsi, en prenant en compte les taux d'admixture  $r_1, r_2$  et les taux de mutations  $\mu_i$  et  $\mu_{i,SNI}$ , nous obtenons un vecteur de paramètres de dimension 15.

Décrivons les lois *a priori* associées aux paramètres décrits précédemment.

1. Les lois *a priori* associées aux tailles efficaces  $Ne_1, \dots, Ne_6$  sont données par

$$Ne_i \sim \text{Gamma}_{[0.1;500000]}(\overline{Ne}/2, 2),$$

où  $\text{Gamma}_{[a;b]}(p, s)$  est la loi gamma de paramètres  $p$  et  $s$  tronquée sur l'intervalle  $[a; b]$ . Notons que  $\mathbb{E}[Ne_i | \overline{Ne}] = \overline{Ne}$  où  $\overline{Ne}$  est un hyper-paramètre de loi  $\mathcal{U}_{[5;100,000]}$ .

2. Les lois *a priori* associées aux différentes dates des événements du scénario sont données par
 
$$t_5 \sim \mathcal{U}_{[0.01;800]}, t_4 - t_5 \sim \mathcal{U}_{[0;50,000]}, t_3 - t_5 \sim \mathcal{U}_{[0;50,000]}, t_2 - \max(t_3, t_4) \sim \mathcal{U}_{[100;500,000]},$$

$$t_1 - t_2 \sim \mathcal{U}_{[100;2000,000]}.$$

3. Les lois *a priori* des paramètres  $\mu_i$  et  $\mu_{i,\text{SNI}}$  sont

$$\mu_i \sim_{\text{iid}} \text{Gamma}_{[5 \times 10^{-7}; 5 \times 10^{-2}]}(\bar{\mu}/2, 2),$$

où  $\bar{\mu} = 10^{-4}$  et  $\mathbb{E}[\mu_i] = \bar{\mu}$ . Nous avons aussi

$$\mu_{i,\text{SNI}} \sim_{\text{iid}} \text{Gamma}_{[10^{-9}; 10^{-3}]}(\bar{\mu}_{\text{SNI}}/2, 2),$$

où  $\bar{\mu}_{\text{SNI}} = 10^{-6}$  et  $\mathbb{E}[\mu_{i,\text{SNI}}] = \bar{\mu}_{\text{SNI}}$ .

4. Les lois *a priori* sur les paramètres d'admixture  $r_1$  et  $r_2$  sont uniformes sur l'intervalle  $[0.01; 0.999]$ .

On présente sur la figure 6.5 l'approximation ABC des lois *a posteriori* des paramètres notés  $\theta_1, \dots, \theta_6$  où  $\theta_i = 4N\epsilon_i\mu$ . Notons que nous avons une bonne répétabilité dans les sorties des figures 6.5 et 6.6. La figure 6.6 présente les approximations des lois *a posteriori* associées aux deux paramètres d'admixture  $r_1$  et  $r_2$ .

Enfin, la figure 6.7 présente une comparaison de notre schéma ABC-SMC avec l'échantillonneur ABC basique d'acceptation-rejet. Plus précisément, nous présentons le facteur de gain en temps de calcul défini en (6.4) en fonction du numéro de l'itération du schéma séquentiel. La ligne verticale en pointillés correspond à l'itération où le critère d'arrêt est réalisé, *i.e* le premier temps  $T$  tel que  $\rho_T \leq 0.1$ . Notre schéma ABC-SMC est arrêté au bout de 14 itérations. Nous remarquons que l'instant d'arrêt donné par ce critère correspond à l'itération où le gain est maximal. Avec un tel facteur de gain, l'algorithme ABC acceptation-rejet nécessite deux fois plus de simulations suivant le modèle que notre schéma pour la même qualité d'approximation.

## 6.5 Comparaison de stratégies de parallélisation dans ABC

L'aspect parallélisation n'a pas été étudié dans les paragraphes précédents. Le critère de comparaison de l'algorithme 14 aux autres schémas est le nombre de jeux de données nécessaires pour atteindre le même niveau de précision. La parallélisation des techniques Monte-Carlo est souvent accessible et simple. Ça revient dans la plupart des cas à moyenner un grand nombre de simulations issues des différentes réalisations indépendantes d'une distribution donnée. Le coût numérique de la parallélisation provient essentiellement de la gestion de la communication entre les différents blocs impliqués. Le problème majeur rencontré dans la parallélisation des méthodes de Monte-Carlo est lié au générateur de nombres aléatoires (RNG pour Random Number Generators). Il est souvent difficile d'implémenter un RNG minimisant le nombre de communications entre les processeurs en parallèle. Le choix du RNG soulève de nombreuses questions qui constituent un domaine de recherche à part entière. De nombreux RNGs parallèles sont proposés dans la littérature. Ici, nous considérons le générateur dynamique de Matsumoto and Nishimura [80] dont la mise en oeuvre est simple et donne de bons résultats. Il engendre un ensemble de générateurs de type Mersenne-Twister indépendants avec différents paramètres internes et des graines distinctes. Chacun des RNGs engendrés par la technique

de Matsumoto and Nishimura [80] constitue un seul générateur Mersenne-Twister indépendant et ne nécessitant aucune communication avec les autres. L'algorithme 12 se parallélise de manière simple en deux manières

- a) L'effort de calcul est uniformément et entièrement réparti sur les différents coeurs. L'étape 5 de l'algorithme 12 de tri des distances est incluse dans cette parallélisation.
- b) On parallélise seulement la simulation des particules et on fusionne les sorties des différents coeurs avant d'effectuer l'étape 5 de l'algorithme 12.

La parallélisation des algorithmes séquentiels comme le schéma SMC de Del Moral et al. [29] peut s'avérer compliquée. Une itération dans ce type d'algorithme revient à simuler des particules indépendantes à partir d'une loi calibrée sur les itérations précédentes. Ici, on s'intéresse seulement aux schéma auto-calibré ABC-SMC que nous avons introduit dans l'algorithme 14. La parallélisation de ce schéma sur différents coeurs nécessite de nombreuses communications entre les différents coeurs dont chacun représente un bloc. Dans nos expérimentations, nous nous servons de l'API OpenMP pour gérer le partage de la mémoire et la communication entre les différents threads.

La parallélisation interne d'un schéma SMC est difficile mais peut être contourner en lançant plusieurs réalisations indépendantes de ce schéma. Une technique de parallélisation hybride peut être envisagée. Elle consiste à lancer plusieurs instances indépendantes du schéma SMC et chaque instance est parallélisée sur plusieurs coeurs.

Nous considérons le problème de génétique des populations étudié dans la partie 6.4.2. Nous utilisons le schéma 14 pour produire une approximation de référence de la loi *a posteriori*. Le nombre de jeux de données affectées à l'étape d'initialisation est fixé à 1,200,000 et nous avons utilisé  $N = 200,000$  particules par itération. Après 10 itérations, le schéma SMC est arrêté. Ainsi, le nombre total de jeux de données est de 3,200,000 et la valeur du ESS obtenu est de 72,000 pour le seuil  $\epsilon$  final. Pour la même qualité d'approximation (mêmes ESS et seuil  $\epsilon$ ), l'algorithme 12 nécessite deux fois plus de particules que le schéma 14 7,200,000.

Nous avons comparé les stratégies de parallélisation suivantes:

- Des instances indépendantes de l'algorithme 12 (des réalisations indépendantes sur 40 coeurs pour simuler les 7,200,000 jeux de données simulés, on simule 180,000 jeux de données sur chaque coeurs).
- Des réalisations indépendantes du schéma 14 (sur 40 coeurs et 200,000 jeux de données par itération, plus précisément, nous effectuons 40 réalisations indépendantes avec 5,000 jeux de données par itération).
- Des réalisations indépendantes du schéma 14 parallélisées sur 4 coeurs (sur 40 coeurs et avec 200,000 jeux de données par itération, nous effectuons 10 réalisations indépendantes avec 20,000 jeux de données par itération, chaque réalisation est parallélisée sur 4 coeurs).
- Des réalisations indépendantes du schéma 14 parallélisées sur 10 coeurs (sur 40 coeurs et avec 200,000 jeux de données par itération, nous effectuons 4 réalisations indépendantes avec 50,000 jeux de données par itération, chaque réalisation est parallélisée sur 10 coeurs).

Les sorties des différentes réalisations indépendantes sont fusionnées à la fin pour chacune des stratégies précédentes.

**TABLE 6.2:** Comparaison des différents stratégies de parallélisation

Nombre total de coeurs	L'algorithme ABC acceptance-rejet	Schéma ABC-SMC	Schéma ABC-SMC parallélisé sur 4 coeurs	Schéma ABC-SMC parallélisé sur 10 coeurs
20	2010	866	1240	2427
40	1046	453	662	1213
80	522	237	350	685
100	419	197	295	560

Le tableau 6.2 présente les résultats des différentes comparaisons. Il contient le temps de calcul en secondes des différentes stratégies pour la même qualité d'approximation, c'est-à-dire le même niveau d'acceptation  $\varepsilon$  et le même ESS. Nous avons répété ces expérimentations pour s'assurer que le sorties ne varient pas d'un schéma à un autre. Lorsque le nombre de processeurs est important, les résultats obtenus par le schéma ABC-SMC sont nettement plus variables (absence de répétabilité). En effet, lorsqu'on effectue le test avec le nombre maximal de coeurs (ici 100), le processus de calibration des paramètres se fait seulement sur 2000 particules à chaque itération. Citons quelques remarques sur les sorties du tableau 6.2:

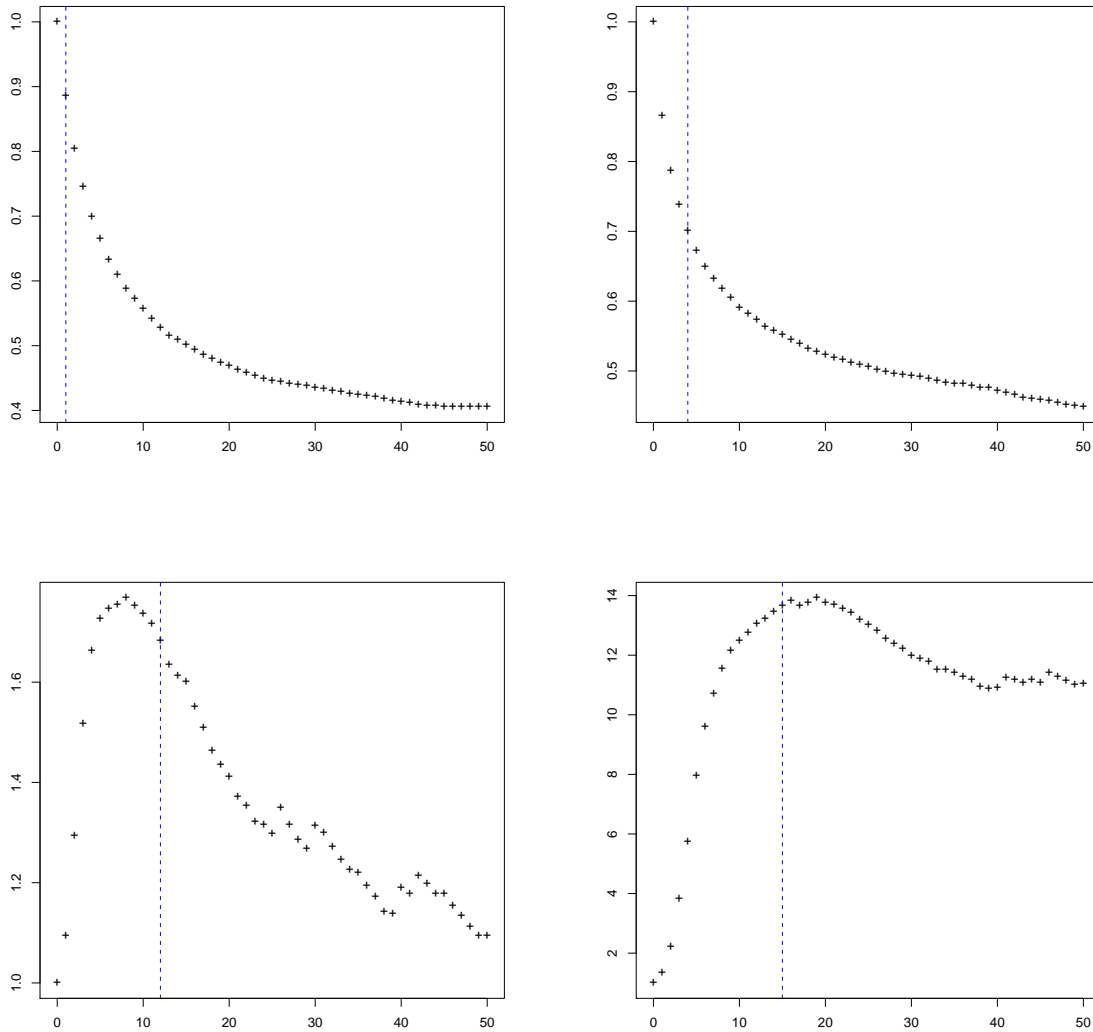
- La parallélisation avec un grand nombre de coeurs par instance du schéma ABC-SMC est coûteuse en temps.
- Il est préférable de réaliser plusieurs instances indépendantes du schéma ABC-SMC en allouant le même nombre de particules pour chaque instance. Ainsi, on assure un minimum de particules pour le processus la calibration des paramètres.
- En plus de la simplicité de la parallélisation de l'algorithme 12, il est conseillé lorsque on a un grand nombre de processeurs (à partir de 100 dans notre cas).

Pour mener nos comparaisons, nous avons considérés le problème de génétique des populations étudié dans la partie 6.4.2 et l'API OpenMP. Nous avons déduit sur cet exemple que le schéma ABC-SMC décrit dans l'algorithme 14 est inutile lorsqu'on a plus de 4 processeurs par instance du programme. L'utilisation de l'algorithme 12 est conseillée lorsqu'on possède plus d'une centaine de coeurs. Il est difficile de définir une méthode générique de choix de stratégie de parallélisation, une manière efficace pour choisir une stratégie est de lancer différentes instances du schéma 14 et d'analyser la répétabilité et la stabilité des résultats obtenus.

**Algorithme 14** Échantillonneur ABC-SMC auto-calibré

- 
- 1: Ordonner les particules  $\{(\theta_i^t, \mathbf{x}_i^t), 1 \leq i \leq N\}$  suivant leurs distances au jeu de données observé  $\mathbf{x}_{\text{obs}}$ , *i.e*

$$d_1 = d(\eta(\mathbf{x}_1), \eta(\mathbf{x}_{\text{obs}})) \leq d_2 = d(\eta(\mathbf{x}_2), \eta(\mathbf{x}_{\text{obs}})) \leq \dots \leq d_N = d(\eta(\mathbf{x}_N), \eta(\mathbf{x}_{\text{obs}})).$$
  - 2: Calcul de  $\sigma_t^2$ , la variance des  $\{\theta_i^t, 1 \leq i \leq N\}$  dans le système  $\mathcal{P}^t$ .
  - 3: **##** Calibration de  $\varepsilon_{t+1}$  **##**
  - 4:  $\alpha' \leftarrow 0.0$
  - 5: **Répéter**
  - 6:    $\alpha' \leftarrow \alpha' + 0.01$
  - 7:    $\varepsilon'(\alpha') \leftarrow d_{\lfloor \alpha' N \rfloor}$
  - 8:   **Pour**  $i = 1 \rightarrow \lfloor \alpha' N \rfloor$  **faire**
  - 9:     **Si**  $(\tilde{\theta}_i, \tilde{\mathbf{x}}_i)$  n'a pas encore été simulé **alors**
  - 10:       Simuler  $\tilde{\theta}_i$  suivant  $\mathcal{N}(\theta_i^t, 2\sigma_t^2)$  et  $\tilde{\mathbf{x}}_i$  suivant  $\ell(\cdot | \tilde{\theta}_i)$ .
  - 11:     **Fin si**
  - 12:   **Fin pour**
  - 13:   Fixer  $N'(\alpha')$  comme le nombre de particules  $(\tilde{\theta}_i, \tilde{\mathbf{x}}_i)$  parmi les  $\lfloor \alpha' N \rfloor$  premières telles que  $\tilde{\theta} \in \Theta$  et  $d(\eta(\tilde{\mathbf{x}}), \eta(\mathbf{x}_{\text{obs}})) \leq \varepsilon'(\alpha')$
  - 14:    $\rho'(\alpha') \leftarrow N'(\alpha') / \lfloor \alpha' N \rfloor$
  - 15: **jusqu'à**  $\alpha' + \rho'(\alpha') \geq 0.9$
  - 16:  $\alpha_{t+1} \leftarrow \alpha'$ ,  $\varepsilon_{t+1} \leftarrow \varepsilon'(\alpha')$  et  $\rho_{t+1} \leftarrow \rho'(\alpha')$
  - 17: **##** Le nouveau système de particules **##**
  - 18: **Pour**  $j = 1 \rightarrow \lfloor \alpha_{t+1} N \rfloor$  **faire**
  - 19:   **Si**  $\tilde{\theta}_i \in \Theta$  et  $d(\eta(\tilde{\mathbf{x}}_i), \eta(\mathbf{x}_{\text{obs}})) \leq \varepsilon_t$  **alors**
  - 20:      $(\theta_j^{t+1}, \mathbf{x}_j^{t+1}) \leftarrow (\tilde{\theta}_i, \tilde{\mathbf{x}}_i)$
  - 21:   **Sinon**
  - 22:      $(\theta_j^{t+1}, \mathbf{x}_j^{t+1}) \leftarrow (\theta_j^t, \mathbf{x}_j^t)$
  - 23:   **Fin si**
  - 24: **Fin pour**
  - 25: **Pour**  $j = \lfloor \alpha_{t+1} N \rfloor + 1 \rightarrow N$  **faire**
  - 26:   Tirer un entier  $I$  uniformément entre 1 et  $\lfloor \alpha N \rfloor$
  - 27:   Simuler  $\tilde{\theta}$  suivant  $\mathcal{N}(\theta_I^t, 2\sigma_t^2)$  et  $\tilde{\mathbf{x}}$  suivant  $\ell(\mathbf{x} | \tilde{\theta})$
  - 28:   **Si**  $\tilde{\theta} \in \Theta$  et  $d(\eta(\tilde{\mathbf{x}}), \eta(\mathbf{x}_{\text{obs}})) \leq \varepsilon_t$  **alors**
  - 29:      $(\theta_j^{t+1}, \mathbf{x}_j^{t+1}) \leftarrow (\tilde{\theta}, \tilde{\mathbf{x}})$
  - 30:   **Sinon**
  - 31:      $(\theta_j^{t+1}, \mathbf{x}_j^{t+1}) \leftarrow (\theta_I^t, \mathbf{x}_I^t)$
  - 32:   **Fin si**
  - 33: **Fin pour**
  - 34: **Si**  $\rho_{t+1} \leq 0.1$  **alors**
  - 35:   Renvoyer comme sortie le dernier système  $\mathcal{P}^{t+1}$  et le dernier seuil  $\varepsilon_{t+1}$ .
  - 36: **Sinon**
  - 37:    $t \leftarrow t + 1$  et aller à l'étape 1.
  - 38: **Fin si**
-



**FIGURE 6.3:** Étude de l'efficacité de notre schéma ABC-SMC en fonction de l'information apportée par la loi a priori. En Haut, à gauche: le support de densité a priori est  $[-0.1; 0.1]$  et  $\alpha_0 = 1/10$ . En Haut, à droite: le support de densité a priori est  $[-1; 1]$  et  $\alpha_0 = 1/10$ . En bas, à gauche: le support de densité a priori est  $[-10; 10]$  et  $\alpha_0 = 1/2$ . En bas, à droite: le support de densité a priori est  $[-100; 100]$  et  $\alpha_0 = 1/4$ .

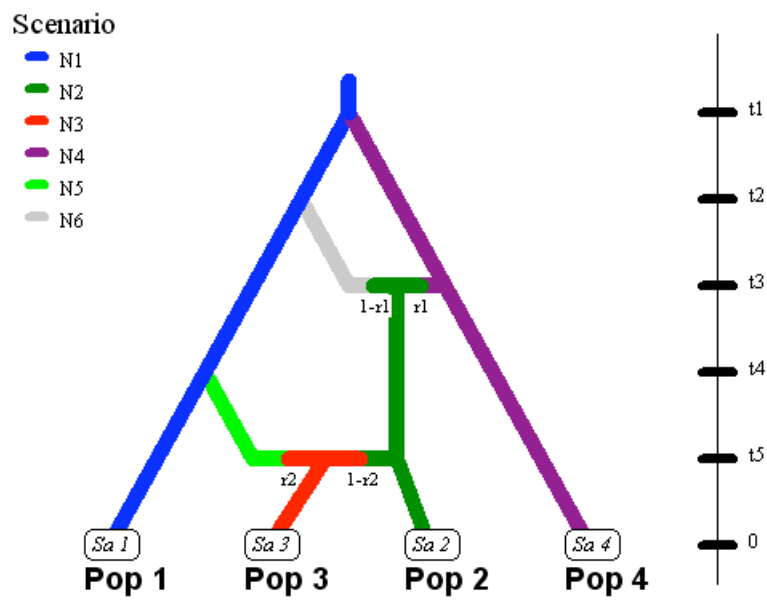


FIGURE 6.4: Scénario évolutif des abeilles en Europe.



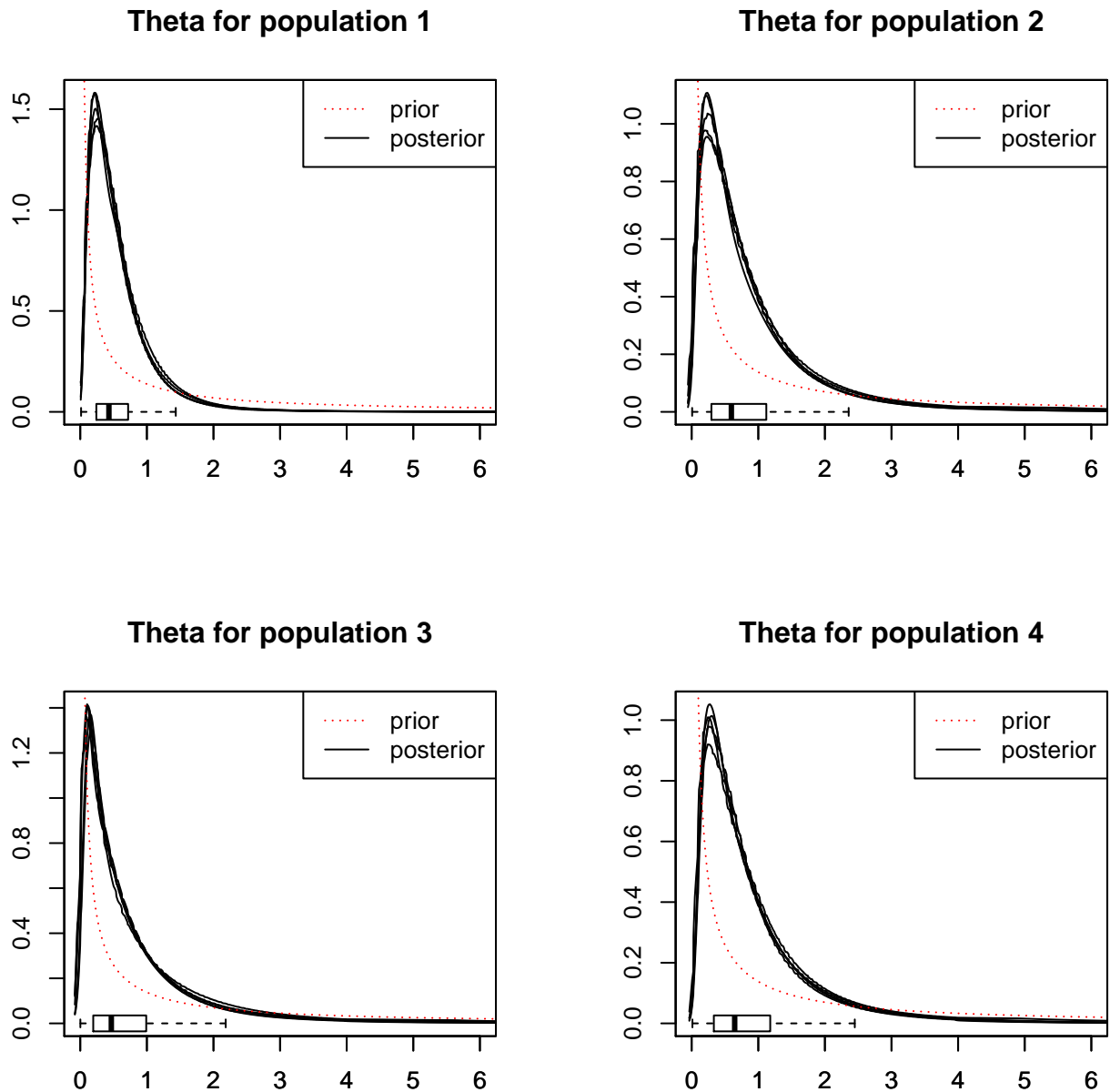
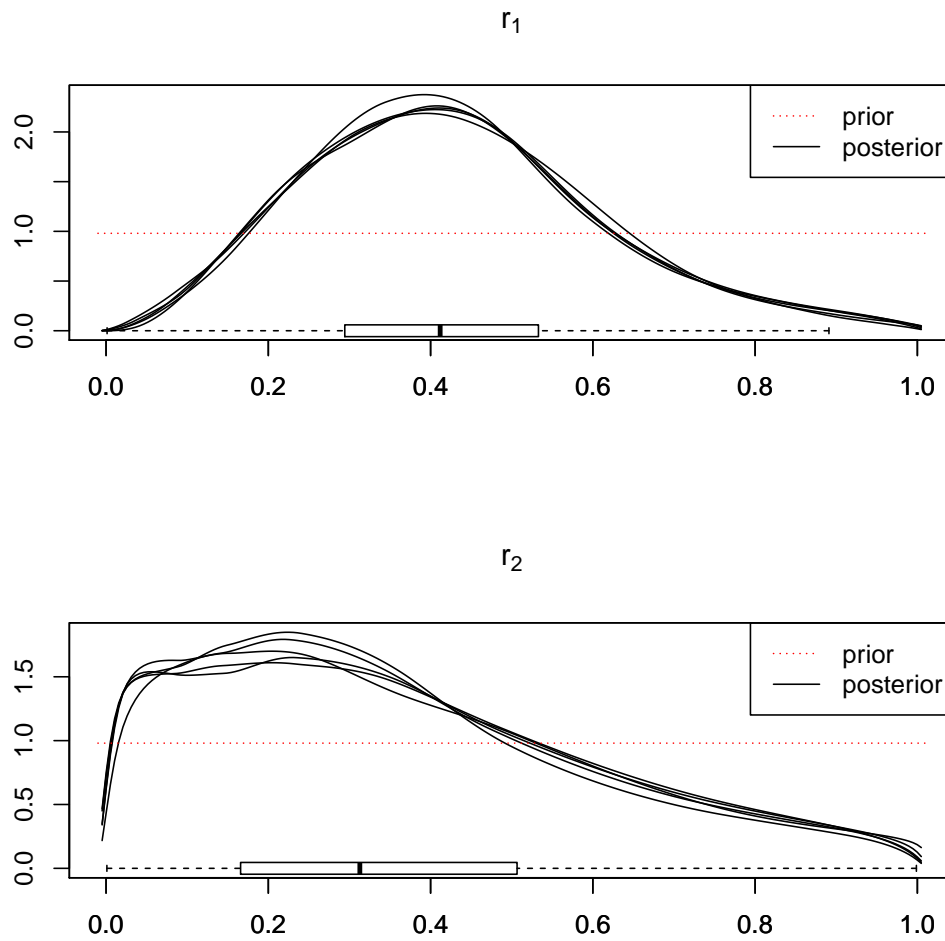


FIGURE 6.5: Cinq approximations ABC de la loi a posteriori des  $\theta_i$  données par cinq réalisations indépendantes de notre schéma ABC-SMC.



**FIGURE 6.6:** Cinq répliques indépendantes de l'approximation ABC de la loi a posteriori associées aux paramètres d'admixture  $r_1$  et  $r_2$ .

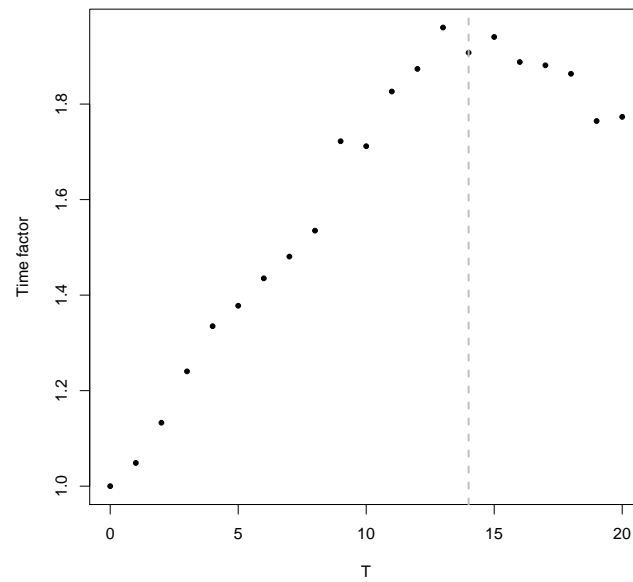


FIGURE 6.7: Évolution du facteur de gain au cours des différentes itérations de notre schéma ABC-SMC.

# 7

## Discussion et perspectives

### 7.1 Questions autour de l’algorithme AMIS

Concernant l’algorithme AMIS, nous avons obtenu un résultat de convergence des estimateurs basés sur le système final de particules et cela pour une certaine classe de fonctions. Ce résultat est sous forme d’une loi des grands nombres et la classe de fonctions est donnée par une multitude d’hypothèses de régularité sur la cible et la famille de lois d’importance. Une amélioration de ce travail consisterait à alléger ces hypothèses via d’autres outils probabilistes et obtenir un théorème central limite sur le système final de particules comme nous l’avons obtenu sur la suite de paramètres intermédiaires  $\{\hat{\theta}_t\}_{t \geq 1}$ .

Les tests présentés dans cette partie visent à montrer que nos modifications dans le schéma AMIS de Cornuet et al. [21] n’impactent pas ses qualités. Nous avons considéré un exemple simple de génétique des populations avec un nombre réduit de paramètres. On retrouve dans Sirén et al. [94], Cornuet et al. [21] des études numériques plus abouties sur des exemples réalistes de génétique des populations. Une suite intéressante à nos expérimentations numériques serait de confronter notre schéma AMIS aux problèmes d’inférence bayésienne de scénarios plus complexes.

Une question récurrente dans l’application de l’algorithme AMIS est liée à la répartition du coût de calcul total (représenté par la taille du système final de particules noté  $\Omega_T$ ) sur les différentes itérations. Une réponse à cette question consisterait à proposer et étudier des stratégies de répartition de  $\Omega_T$  sur les différentes itérations via la suite  $N_1, \dots, N_T$ .

### 7.2 Quelques perspectives autour des méthodes ABC

La prise en compte de scénarios évolutifs de plus en plus complexes nécessite le développement de techniques adaptées. Typiquement, en génétique des populations, il est important que les méthodes aient une bonne précision. D’un autre côté, on considère des modèles complexes avec un grand nombre de paramètres et d’événements. Sur l’exemple

particulier du chapitre 6, nous avons déduit qu'il est préférable d'utiliser l'algorithme ABC d'acceptation-rejet (voir algorithmes 11 et 12 du chapitre 6) quand on est en possession d'une machine avec un grand nombre de processeurs. Il serait intéressant d'étendre la comparaison des stratégies de parallélisation à différents modèles pour conforter cette conclusion.

La technique de régression locale introduite par Beaumont et al. [4] est hautement utilisée en pratique. Elle apporte des améliorations considérables en réduisant le niveau d'approximation  $\varepsilon$ . Blum et al. [10], Fearnhead and Prangle [44] proposent d'ajouter à la régression du vecteur de paramètres  $\theta$  sur les statistiques résumées une étape de sélection de variables. Cette sélection permet de conserver les statistiques résumées qui apportent le plus d'information en réduisant le bruit. Notre contribution à ce post-traitement ABC est la discussion Pudlo and Sedki [87] dont les résultats sont présentés à la fin du chapitre 1. Nous nous sommes concentrés sur le critère BIC. On retrouve dans Blum et al. [10] une analyse comparative de différents critères de sélection. Il serait intéressant d'avoir des résultats plus formels que des simples comparaisons numériques. Ces problèmes soulèvent des questions qui rentrent dans le champ de la perte d'exhaustivité en statistique.

Dans le contexte de biologie évolutive, les généticiens des populations soulèvent des questions de choix de scénarios qui font appel aux techniques de choix de modèles. Par l'approche bayésienne de cette thèse, la réponse à ces questions est donnée par les techniques ABC d'approximation des probabilités *a posteriori* des modèles en compétition. En dépit des résultats pertinents obtenus sur des exemples réels, les récents travaux de Robert et al. [90] montrent que le choix de modèles par l'approximation ABC peut être déficient. Les résultats récents de Marin et al. [74] apportent une réponse aux questions de Robert et al. [90] en montrant la convergence de la sélection de modèles par l'approximation ABC. Leurs résultats sont donnés sous des conditions sur la loi du vecteur des statistiques résumées. D'un point de vue pratique, on peut envisager de vérifier ces conditions sur les modèles de génétique des populations en étudiant les lois des statistiques résumées couramment utilisées et dont la littérature est abondante.

Que ce soit pour l'inférence des paramètres d'un modèle ou le choix de celui-ci, on peut améliorer le post-traitement des techniques ABC en utilisant des méthodes d'apprentissage sophistiquées pour inférer le vecteur des paramètres  $\theta$  (ou l'indice du modèle) en fonction des statistiques résumées.

Notre schéma ABC-SMC se stoppe lorsque le taux d'acceptation du noyau de diversification est inférieur à un seuil fixé. On pourrait envisager d'estimer le facteur de gain  $r_t$  et arrêter lorsque celui-ci est inférieur à 1. Pour estimer ce facteur de gain, il faut être capable, pour chaque itération, d'approcher la probabilité que la distance à l'observation soit inférieure au niveau  $\varepsilon_t$

$$\mathbb{P} \left( d(\eta(\mathbf{x}), \eta(\mathbf{x}_{\text{obs}})) \leq \varepsilon_t \right)$$

lorsque la loi de  $\mathbf{x}$  est la loi marginale de  $(\theta, \mathbf{x}) \sim \pi(\theta) \ell(\mathbf{x}|\theta)$ . C'est équivalent à estimer le nombre de simulations nécessaires pour l'échantillonneur ABC acceptation-rejet pour en

garder  $\text{ESS}_t$ . Nous avons,

$$\mathbb{P}\left(d(\eta(\mathbf{x}), \eta(\mathbf{x}_{\text{obs}})) \leq \varepsilon_t\right) = \mathbb{P}\left(d(\eta(\mathbf{x}), \eta(\mathbf{x}_{\text{obs}})) \leq \varepsilon_t \mid d(\eta(\mathbf{x}), \eta(\mathbf{x}_{\text{obs}})) \leq \varepsilon_{t-1}\right) \dots \\ \mathbb{P}\left(d(\eta(\mathbf{x}), \eta(\mathbf{x}_{\text{obs}})) \leq \varepsilon_1 \mid d(\eta(\mathbf{x}), \eta(\mathbf{x}_{\text{obs}})) \leq \varepsilon_0\right) \mathbb{P}\left(d(\eta(\mathbf{x}), \eta(\mathbf{x}_{\text{obs}})) \leq \varepsilon_0\right).$$

Ainsi, on peut estimer la probabilité de l'événement rare

$$\left\{d(\eta(\mathbf{x}), \eta(\mathbf{x}_{\text{obs}})) \leq \varepsilon_t\right\}$$

par le produit des niveaux des quantiles  $\alpha_t$  calibrés par notre algorithme. En effet,  $\mathbf{x}$  à l'itération  $t$  est distribué suivant la marginale de  $\pi(\boldsymbol{\theta})\ell(\mathbf{x}|\boldsymbol{\theta})$  sachant que  $d(\eta(\mathbf{x}), \eta(\mathbf{x}_{\text{obs}})) \leq \varepsilon_{t-1}$ . Il serait intéressant d'analyser les propriétés de cette technique.

## RÉSUMÉ

Cette thèse propose et étudie deux techniques d'inférence bayésienne dans les modèles où la vraisemblance possède une composante latente. Dans ce contexte, la vraisemblance d'un jeu de données observé est l'intégrale de la vraisemblance dite complète sur l'espace de la variable latente. On s'intéresse aux cas où l'espace de la variable latente est de très grande dimension et comporte des directions de différentes natures (discrètes et continues), ce qui rend cette intégrale incalculable. Le champs d'application privilégié de cette thèse est l'inférence dans les modèles de génétique des populations. Pour mener leurs études, les généticiens des populations se basent sur l'information génétique extraite des populations du présent et représente la variable observée. L'information incluant l'histoire spatiale et temporelle de l'espèce considérée est inaccessible en général et représente la composante latente. Notre première contribution dans cette thèse suppose que la vraisemblance peut être évaluée via une approximation numériquement coûteuse. Le schéma d'échantillonnage préférentiel adaptatif et multiple (AMIS pour Adaptive Multiple Importance Sampling) de Cornuet et al. [21] nécessite peu d'appels au calcul de la vraisemblance et recycle ces évaluations. Cet algorithme approche la loi a posteriori par un système de particules pondérées. Cette technique est conçue pour pouvoir recycler les simulations obtenues par le processus itératif (la construction séquentielle d'une suite de lois d'importance). Dans les nombreux tests numériques effectués sur des modèles de génétique des populations, l'algorithme AMIS a montré des performances numériques très prometteuses en terme de stabilité. Ces propriétés numériques sont particulièrement adéquates pour notre contexte. Toutefois, la question de la convergence des estimateurs obtenus par cette technique reste largement ouverte. Dans cette thèse, nous montrons des résultats de convergence d'une version légèrement modifiée de cet algorithme. Sur des simulations, nous montrons que ses qualités numériques sont identiques à celles du schéma original. Dans la deuxième contribution de cette thèse, on renonce à l'approximation de la vraisemblance et on supposera seulement que la simulation suivant le modèle (suivant la vraisemblance) est possible. Notre apport est un algorithme ABC séquentiel (Approximate Bayesian Computation). Sur les modèles de la génétique des populations, cette méthode peut se révéler lente lorsqu'on vise une approximation précise de la loi a posteriori. L'algorithme que nous proposons est une amélioration de l'algorithme ABC-SMC de Del Moral et al. [31] que nous optimisons en nombre d'appels aux simulations suivant la vraisemblance, et que nous munissons d'un mécanisme de choix de niveaux d'acceptations auto-calibré. Nous implémentons notre algorithme pour inférer les paramètres d'un scénario évolutif réel et complexe de génétique des populations. Nous montrons que pour la même qualité d'approximation, notre algorithme nécessite deux fois moins de simulations par rapport à la méthode ABC avec acceptation couramment utilisée.

## MOTS-CLÉS

Méthodes bayésiennes, méthodes de Monte-Carlo, échantillonnage préférentiel adaptatif et multiple, méthodes ABC, modèles de génétique des populations.

### ABSTRACT

This thesis consists of two parts which can be read independently. The first part is about the Adaptive Multiple Importance Sampling (AMIS) algorithm presented in Cornuet et al. [21] provides a significant improvement in stability and Effective Sample Size due to the introduction of the recycling procedure. These numerical properties are particularly adapted to the Bayesian paradigm in population genetics where the modelization involves a large number of parameters. However, the consistency of the AMIS estimator remains largely open. In this work, we provide a novel Adaptive Multiple Importance Sampling scheme corresponding to a slight modification of Cornuet et al. [21] proposition that preserves the above-mentioned improvements. Finally, using limit theorems on triangular arrays of conditionally independent random variables, we give a consistency result for the final particle system returned by our new scheme. The second part of this thesis lies in ABC paradigm. Approximate Bayesian Computation has been successfully used in population genetics models to bypass the calculation of the likelihood. These algorithms provide an accurate estimator by comparing the observed dataset to a sample of datasets simulated from the model. Although parallelization is easily achieved, computation times for assuring a suitable approximation quality of the posterior distribution are still long. To alleviate this issue, we propose a sequential algorithm adapted from Del Moral et al. [31] which runs twice as fast as traditional ABC algorithms. Its parameters are calibrated to minimize the number of simulations from the model.

### KEYWORDS

Bayesian inference, Monte Carlo methods, adaptive multiple importance sampling, ABC methods, population genetics models.





# Bibliographie

- [1] Andrieu, C., Doucet, A., and Holenstein, R. (2010). Particle Markov chain Monte Carlo (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 72(4):357–385.
- [2] Andrieu, C. and Roberts, G. O. (2009). The pseudo-marginal approach for efficient Monte Carlo computations. *The Annals of Statistics*, 37(2):697–725.
- [3] Baragatti, M., Grimaud, A., and Pommeret, D. (2012). Likelihood-free parallel tempering. *to appear in Statistics and Computing*.
- [4] Beaumont, M., Zhang, W., and Balding, D. (2002). Approximate Bayesian Computation in population genetics. *Genetics*, 162:2025–2035.
- [5] Beaumont, M. A. (2003). Estimation of Population Growth or Decline in Genetically Monitored Populations. *Genetics*, 164(3):1139–1160.
- [6] Beaumont, M. A. (2010). Approximate Bayesian Computation in Evolution and Ecology. *Annual Review of Ecology, Evolution, and Systematics*, 41(1):379–406.
- [7] Beaumont, M. A., Cornuet, J.-M., Marin, J.-M., and Robert, C. P. (2009). Adaptive approximate Bayesian computation. *Biometrika*, 96(4):983–990.
- [8] Beaumont, M. A. and Rannala, B. (2004). The Bayesian revolution in genetics. *Nature Reviews Genetics*, 5(4):251–261.
- [9] Beerli, P. and Felsenstein, J. (2001). Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach. *Proceedings of the National Academy of Sciences*, 98(8):4563–4568.
- [10] Blum, M., Nunes, M., Prangle, D., and Sisson, S. (2012). A comparative review of dimension reduction methods in approximate Bayesian computation. *Statistical Science*, *to appear*.
- [11] Blum, M. G. B. (2010). Approximate Bayesian Computation: A Nonparametric Perspective. *Journal of the American Statistical Association*, 105(491):1178–1187.
- [12] Brooks, S., Gelman, A., Jones, G., L., and Meng, X.-L. (2011). *Handbook of Markov Chain Monte Carlo (Chapman & Hall/CRC Handbooks of Modern Statistical Methods)*. Chapman and Hall/CRC, 1 edition.
- [13] Cappé, O., Guillin, A., Marin, J.-M., and Robert, C. P. (2004). Population monte carlo. *Journal of Computational and Graphical Statistics*, 13:907–929.

- [14] Cappé, O., Guillin, A., Marin, J.-M., and Robert, C. P. (2008). Adaptive importance sampling in general mixture classes. *Statistics and Computing*, 18(4):447–459.
- [15] Cappé, O., Moulines, E., and Rydén, T. (2005). *Inference in hidden Markov models*. Springer, New York.
- [16] Casella, G. and Robert, C. P. (1996). Rao-Blackwellisation of sampling schemes. *Biometrika*, 83(1):81–94.
- [17] Casella, G. and Robert, C. P. (1998). Post-processing accept-reject samples: recycling and rescaling. *Journal of Computational and Graphical Statistics*, 7(2):139–157.
- [18] Celeux, G. and Diebolt, J. (1985). The SEM algorithm: A probabilistic teacher algorithm derived from the EM algorithm for the mixture problem. *Computational Statistics Quarterly*, 2:73–82.
- [19] Chopin, N. (2004). Central Limit Theorem for Sequential Monte Carlo Methods and Its Application to Bayesian Inference. *The Annals of Statistics*, 32(6):2385–2411.
- [20] Cornuet, J. M., Beaumont, M. A., Estoup, A., and Solignac, M. (2006). Inference on microsatellite mutation processes in the invasive mite, *Varroa destructor*, using reversible jump Markov chain Monte Carlo. *Theoretical Population Biology*, 69(2):129–144.
- [21] Cornuet, J.-M., Marin, J.-M., Mira, A., and Robert, C. (2012). Adaptive Multiple Importance Sampling. *Scandinavian Journal of Statistics*.
- [22] Cornuet, J.-M., Santos, F., Beaumont, M. A., Robert, C. P., Marin, J.-M., Balding, D. J., Guillemaud, T., and Estoup, A. (2008). Inferring population history with DIYABC: a user-friendly approach to Approximate Bayesian Computation. *Bioinformatics*, 24(23):2713–2719.
- [23] Crow, J. F. and Kimura, M. (2009). *An Introduction to Population Genetics Theory*. The Blackburn Press.
- [24] Csilléry, K., Blum, M. G. B., Gaggiotti, O. E., and François, O. (2010). Approximate Bayesian Computation (ABC) in practice. *Trends in Ecology & Evolution*.
- [25] De Iorio, M. and Griffiths, R. C. (2004a). Importance sampling on coalescent histories. I. *Advances in Applied Probability*, 36(2):417–433.
- [26] De Iorio, M. and Griffiths, R. C. (2004b). Importance sampling on coalescent histories. II: Subdivided population models. *Advances in Applied Probability*, 36(2):434–454.
- [27] De Iorio, M., Griffiths, R. C., Leblois, R., and Rousset, F. (2005). Stepwise mutation likelihood computation by sequential importance sampling in subdivided population models. *Theoretical Population Biology*, 68(1):41–53.
- [28] Del Moral, P. (2004). *Feynman-Kac formulae. Genealogical and interacting particle systems with applications*. Probability and its Applications (New York). Springer-Verlag, New York.

- [29] Del Moral, P., Doucet, A., and Jasra, A. (2006). Sequential monte carlo samplers. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(3):411–436.
- [30] Del Moral, P., Doucet, A., and Jasra, A. (2011). On adaptative resampling procedures for sequential Monte Carlo methods. *Bernoulli*, 18(1):252–278.
- [31] Del Moral, P., Doucet, A., and Jasra, A. (2012). An adaptive sequential Monte Carlo method for approximate Bayesian computation. *Statistics and Computing*, 22(5).
- [32] Delyon, B., Lavielle, M., and Moulines, E. (1999). Convergence of a Stochastic Approximation Version of the EM Algorithm. *The Annals of Statistics*, 27(1):94–128.
- [33] Dempster, A., Laird, N., and Rubin, D. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 39(1):1–38.
- [34] Donnelly, P. and Tavaré, S. (1995). Coalescents and genealogical structure under neutrality. *Annu Rev Genet*, 29:401–21.
- [35] Douc, R., Guillin, A., Marin, J. M., and Robert, C. P. (2007a). Convergences of adaptive mixtures of importance sampling schemes. *Annals of Statistics*, 35(1):420–448.
- [36] Douc, R., Guillin, A., Marin, J.-M., and Robert, C. P. (2007b). Minimum variance importance sampling via Population Monte Carlo. *ESAIM: Probability and Statistics*, 11:427–447.
- [37] Douc, R. and Moulines, E. (2008). Limit theorems for weighted samples with applications to Sequential Monte Carlo Methods. *The Annals of Statistics*, 36(5):2344–2376.
- [38] Douc, R. and Robert, C. P. (2010). A vanilla Rao-Blackwellisation of Metropolis-Hastings algorithms. *to appear in The Annals of Statistics*.
- [39] Doucet, A., De Freitas, N., and Gordon, N., editors (2001). *Sequential Monte Carlo Methods in Practice (Statistics for Engineering and Information Science)*. Springer, 1 edition.
- [40] Drovandi, C. C. and Pettitt, A. N. (2011). Estimation of parameters for macroparasite population evolution using approximate Bayesian computation. *Biometrics*, 67(1):225–233.
- [41] Ethier, S. N. and Griffiths, R. C. (1987). The Infinitely-Many-Sites Model as a Measure-Valued Diffusion. *The Annals of Probability*, 15(2):515–545.
- [42] Ewens, W. (2004). *Mathematical population genetics*. Springer.
- [43] Fearnhead, P. and Donnelly, P. (2001). Estimating Recombination Rates From Population Genetic Data. *Genetics*, 159(3):1299–1318.
- [44] Fearnhead, P. and Prangle, D. (2012). Constructing Summary Statistics for Approximate Bayesian Computation: Semi-automatic approximate Bayesian computation. *Journal of Royal Statistical Society: Series B (Statistical Methodology)*, 74(3):419–474.

- [45] Gelfand, A. E. and Smith, A. F. M. (1990). Sampling-Based Approaches to Calculating Marginal Densities. *Journal of the American Statistical Association*, 85(410):398–409.
- [46] Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(6):721–741.
- [47] Geyer, C. J. and Thompson, E. A. (1995). Annealing Markov Chain Monte Carlo with Applications to Ancestral Inference. *Journal of the American Statistical Association*, 90(431):909–920.
- [48] Gilks, W. R., Richardson, S., and Spiegelhalter, D. J. (1995). *Markov Chain Monte Carlo in Practice: Interdisciplinary Statistics (Chapman & Hall/CRC Interdisciplinary Statistics)*. Chapman and Hall/CRC, 1 edition.
- [49] Griffiths, R. C. (1989). Genealogical-tree probabilities in the infinitely-many-site model. *Journal of mathematical biology*, 27(6):667–680.
- [50] Griffiths, R. and Tavaré, S. (1994). Ancestral inference in population genetics. *Statistical Science*, 9:307–319.
- [51] Haario, H., Saksman, E., and Tamminen, J. (1999). Adaptive proposal distribution for random walk Metropolis algorithm. *Computational Statistics*, 14:375–395.
- [52] Hesterberg, T. (1988). *Advances in Importance Sampling*. PhD thesis, Stanford University.
- [53] Hey, J. and Nielsen, R. (2004). Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. *Genetics*, 167(2):747–760.
- [54] Hey, J. and Nielsen, R. (2007). Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. *Proceedings of the National Academy of Sciences*, 104(8):2785–2790.
- [55] Hudson, R. R. (2001). Two-locus sampling distributions and their application. *Genetics*, 159(4):1805–1817.
- [56] Hudson, R. R. (2002). Generating samples under a Wright-Fisher neutral model of genetic variation. *Bioinformatics*, 18(2):337–338.
- [57] Kimura, M. (1968). Evolutionary rate at the molecular level. *Nature*, 217(5129):624–626.
- [58] Kimura, M. (1983). *The Neutral Theory of Molecular Evolution*. Cambridge University Press.
- [59] Kingman, J. F. C. (1982a). The coalescent. *Stochastic Processes and their Applications*, 13:235–248.

- [60] Kingman, J. F. C. (1982b). Exchangeability and the Evolution of Large Populations. In Koch, G. and Spizzichino, F., editors, *Exchangeability in Probability and Statistics*, pages 97–112. North-Holland, Amsterdam.
- [61] Kingman, J. F. C. (1982c). On the Genealogy of Large Populations. *Journal of Applied Probability*, 19:27–43.
- [62] Kuhner, M. K., Yamato, J., and Felsenstein, J. (1995). Estimating effective population size and mutation rate from sequence data using Metropolis-Hastings sampling. *Genetics*, 140(4):1421–1430.
- [63] Kuhner, M. K., Yamato, J., and Felsenstein, J. (2000). Maximum Likelihood Estimation of Recombination Rates From Population Data. *Genetics*, 156(3):1393–1401.
- [64] Leblois, R. (2004). *Estimation de paramètres de dispersion en populations structurées à partir de données génétiques*. PhD thesis, École nationale supérieure agronomique de Montpellier.
- [65] Leblois, R., Estoup, A., and Rousset, F. (2009). IBDSim: A computer program to simulate genotype data under Isolation By Distance. *Molecular Ecology Resources*, 9(1):107–109.
- [66] Liu, J. S. (1994). The Collapsed Gibbs Sampler in Bayesian Computations with Applications to a Gene Regulation Problem. *Journal of the American Statistical Association*, 89(427).
- [67] Liu, J. S. (1996). Metropolized Independent Sampling with Comparisons to Rejection Sampling and Importance Sampling. *Statistics and Computing*, 6(2):113–119.
- [68] Liu, J. S. (2008). *Monte Carlo Strategies in Scientific Computing*. Springer Series in Statistics. New York, springer edition.
- [69] Liu, J. S., Liang, F., and Wong, W. (2001). A theory of dynamic weighted in Monte Carlo computation. *Journal of the American Statistical Association*, 96(454):561–573.
- [70] Liu, J. S. and Sabatti, C. (2000). Generalized Gibbs sampler and multigrid Monte Carlo for Bayesian computation. *Biometrika*, 87(2):353–369.
- [71] Liu, J. S., Wong, W. H., and Kong, A. (1994). Covariance structure of the Gibbs sampler with applications to the comparisons of estimators and augmentation schemes. *Biometrika*, 81(1):27–40.
- [72] Lunn, D. J., Thomas, A., Best, N., and Spiegelhalter, D. (2000). WinBUGS A Bayesian modelling framework: Concepts, structure, and extensibility. *Statistics and Computing*, 10(4):325–337.
- [73] Majoram, P. and Donnelly, P. (1997). Human demography and the time since mitochondrial Eve. In Donnelly, P. and Tavaré, S., editors, *Progress in Population Genetics and Human Evolution*, volume 87. Springer.
- [74] Marin, J. M., Pillai, N., Robert, C. P., and Rousseau, J. (2011a). Relevant statistics for Bayesian model choice. Available on Arxiv : <http://fr.arXiv.org/abs/1110.4700>.

- [75] Marin, J.-M., Pudlo, P., Robert, C. P., and Ryder, R. J. (2011b). Approximate Bayesian computational methods. *Statistics and Computing*.
- [76] Marin, J.-M., Pudlo, P., and Sedki, M. (2012). Optimal parallelization of a sequential approximate bayesian computation algorithm. In Laroque, C., Himmelspach, R., Papsupathy, R., Rose, O., and M., U. A., editors, *Proceedings of the 2012 Winter Simulation Conference*.
- [77] Marin, J.-M. and Robert, C. P. (2007). *Bayesian Core: A Practical Approach to Computational Bayesian Statistics (Springer Texts in Statistics)*. Springer, 1st edition.
- [78] Marjoram, P., Molitor, J., Plagnol, V., and Tavaré, S. (2003). Markov chain Monte Carlo without likelihoods. *Proceedings of the National Academy of Sciences*, 100(26):15324–15328.
- [79] Marshall, A. W. (1956). The Use of Multistage Sampling Schemes in Monte Carlo Computations. pages 123–140. Symposium on Monte Carlo Methods, A. H. Meyer(ed.).
- [80] Matsumoto, M. and Nishimura, T. (1998). Dynamic Creation of Pseudorandom Number Generators. In *Proceedings of the Third International Conference on Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing*, pages 56–69.
- [81] Nielsen, R. and Wakeley, J. (2001). Distinguishing Migration From Isolation: A Markov Chain Monte Carlo Approach. *Genetics*, 158(2):885–896.
- [82] Ntzoufras, I. (2009). *Bayesian Modeling Using WinBUGS*. Wiley.
- [83] Owen, A. and Zhou, Y. (2000). Safe and Effective Importance Sampling. *Journal of the American Statistical Association*, 95(449):135–143.
- [84] Peel, D. and McLachlan, G. J. (2000). Robust mixture modelling using the t distribution. *Statistics and Computing*, 10(4):339–348.
- [85] Pennanen, T. and Koivu, M. (2004). An adaptive importance sampling technique. In Springer-Verlag, editor, *Monte Carlo and Quasi-Monte Carlo Methods 2004*, pages 443–455. Niederreiter, H. and Talay, D.
- [86] Pritchard, J., Seielstad, M., Perez-Lezaun, A., and Feldman, M. (1999). Population growth of human Y chromosomes: a study of Y chromosome microsatellites. *Molecular Biology and Evolution*, 16(12):1791–1798.
- [87] Pudlo, P. and Sedki, M. (2012). Discussion on the paper by fearnhead and prangle "constructing summary statistics for approximate bayesian computation: semi-automatic approximate bayesian computation". *Journal of Royal Statistical Society: Series B (Statistical Methodology)*, 74(3):466.
- [88] Ripley, B. (1987). *Stochastic Simulation*. John Wiley & Sons Inc.
- [89] Robert, C. and Casella, G. (2004). *Monte Carlo Statistical Methods*. Springer-Verlag, second edition edition.

- [90] Robert, C. P., Cornuet, J.-M., Marin, J.-M., and Pillai, N. S. (2011). Lack of confidence in approximate Bayesian computation model choice. *Proceedings of the National Academy of Sciences*, 108(37):15112–15117.
- [91] Roberts, G. O. and Rosenthal, J. S. (2006). Harris recurrence of Metropolis-within-Gibbs and trans-dimensional Markov chains. *Ann. Appl. Probab.*, 16(4):2123–2139.
- [92] Rousset, F. and Leblois, R. (2012). Likelihood-Based Inferences under Isolation by Distance: Two-Dimensional Habitats and Confidence Intervals. *Molecular Biology and Evolution*, 29(3):957–973.
- [93] Rubinstein, R. and Kroese, D. (2004). *The cross-entropy method. A unified approach to combinatorial optimization, Monte-Carlo simulation, and machine Learning*. Springer.
- [94] Sirén, J., Marttinen, P., and Corander, J. (2010). Reconstructing population histories from single-nucleotide polymorphism data. *Molecular Biology and Evolution*, 28(1):673–683.
- [95] Sisson, S., Fan, Y., and Tanaka, M. (2009). Sequential Monte Carlo without likelihoods: Errata. *Proceedings of the National Academy of Sciences*, 106(39):16889.
- [96] Sisson, S. A., Fan, Y., and Tanaka, M. M. (2007). Sequential Monte Carlo without likelihoods. *Proceedings of the National Academy of Sciences*, 104(6):1760–1765.
- [97] Stephens, M. and Donnelly, P. (2000). Inference in molecular population genetics. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 62(4):605–635.
- [98] Stumpf, M. P. and McVean, G. A. (2003). Estimating recombination rates from population-genetic data. *Nature reviews. Genetics*, 4(12):959–968.
- [99] Tanner, M. A. and Wong, W. H. (1987). The Calculation of Posterior Distributions by Data Augmentation. *Journal of the American Statistical Association*, 82(398):528–540.
- [100] Tavaré, S. (2004). *Ancestral inference in population genetics*, volume 1837, pages 1–188. Springer Verlag.
- [101] Veach, E. and Guibas, L. (1995). Optimally Combining Sampling Techniques For Monte Carlo Rendering. pages 419–428. SIGGRAPH 95 Proceeding, Addison-Wesley.
- [102] Wakeley, J. (2005). *Coalescent Theory: An Introduction*. Roberts & Company Publishers.
- [103] Wei, G. C. G. and Tanner, M. A. (1990). A Monte Carlo Implementation of the EM Algorithm and the Poor Man's Data Augmentation Algorithms. *Journal of the American Statistical Association*, 85(411):699–704.
- [104] Whittaker, J. C., Harbord, R. M., Boxall, N., Mackay, I., Dawson, G., and Sibly, R. M. (2003). Likelihood-based estimation of microsatellite mutation rates. *Genetics*, 164(2):781–787.
- [105] Woodard, D. B. and Rosenthal, J. S. (2011). Convergence Rate of Markov Chain Methods for Genomic Motif Discovery. *available at: <http://people.orie.cornell.edu/woodard/WoodRose2011.pdf>*.