

**REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE**

Ministère de l'enseignement Supérieur et de la Recherche Scientifique

**UNIVERSITE Mohamed Seddik Ben Yahia – Jijel**

**Faculté des Sciences Exactes et Informatique**

**Département de Mathématiques**



## **Mémoire**

Pour l'obtention du diplôme de : **Master**

**Spécialité : Mathématiques**

**Option : Probabilités et Statistique**

**Thème**

**La méthode d'Echantillonnage Descriptif Amélioré  
dans un contexte Bayésienne**

**Présenté par :**

**Zenifeche Abla**

**Ghibour Aicha**

**Devant le jury :**

**Président : pro Zerzayhi Tahar**

**Encadreur :Mme Ghouil Djoweyda**

**Examineur :Mme Djeridi Zahra**

**Examineur :Mme Madi Meriem**

**Promotion 2016/2017**

---

# Remerciements

Avant tout, nous remercions Dieu tout puissant de nous avoir donné la force, le courage, la persistance et nous a permis d'exploiter les moyens disponibles afin d'accomplir ce modeste travail. Merci de nous avoir éclairé le chemin de la réussite.

Nous adressons nos plus sincères remerciements à notre promotrice **Mme Ghouil Djoweyda** qui nous a encadré et dirigé ce travail avec une grande rigueur scientifique, sa disponibilité, ses précieux conseils, ses encouragements, sa patience, sa compréhension, la confiance qu'elle nous a accordé nous a permis de réaliser ce travail et sa gentillesse pendant le suivi de ce travail.

Nous offrons nos plus sincères remerciements aux membres du jury **Mme Djeridi Zohra, Mme Madi Meriem** et **prof Zerzaiyi Tahar** Pour avoir bien voulu donner de leur temps pour lire ce travail et faire partie des examinateurs aussi pour ses précisions remarques pour corriger ce travail et pour l'assistance par des conseils objectifs et éclairés.

Nous adressons nos plus sincères remerciements à Monsieur **Gharda Mabrouk** pour ses conseils.

Enfin, un très grand merci à nos enseignants et nos professeurs de département de mathématique qui ont contribué positivement façonnage de nos esprits plein de clarté et de constructions scientifiques.

Un grand merci à toutes les personnes qui, d'une quelconque manière, nous ont apporté leur amitié, leur attention, leurs encouragements, leur appui et leur assistance pour réaliser ce travail : nos familles, nos amis. Nous ne saurons citer chacun par son nom. Que tous trouvent ici l'expression de ma franche et profonde reconnaissance !

**Abla** et **Aicha**

# Table des matières

<b>Introduction</b>	<b>1</b>
<b>1 L'analyse statistique Bayésienne</b>	<b>2</b>
1.1 Introduction . . . . .	2
1.2 La loi a priori . . . . .	3
1.2.1 Le choix d'une loi a priori . . . . .	3
1.2.2 Lois a priori non informatives . . . . .	3
1.2.3 Lois a priori conjuguées . . . . .	5
1.2.4 Lois a priori subjectives . . . . .	7
1.3 La loi a posteriori . . . . .	8
1.4 Fonction de perte et risque . . . . .	8
1.5 L'estimateur bayésien . . . . .	9
<b>2 Fondements des méthodes de MC</b>	<b>13</b>
2.1 Introduction . . . . .	13
2.2 Description de la méthode . . . . .	14
2.2.1 Convergence de la méthode . . . . .	15
2.2.2 Principe de la méthode . . . . .	17
2.3 Simulation statistique . . . . .	18
2.3.1 Simulation de Monte Carlo . . . . .	18
2.3.2 Algorithme de Monte Carlo . . . . .	19
2.4 Méthode d'inverse . . . . .	20
<b>3 La méthode EDA</b>	<b>21</b>
3.1 Introduction . . . . .	21
3.2 Echantillonnage aléatoire . . . . .	21
3.2.1 Echantillonnage descriptif . . . . .	22
3.2.2 La procédure d'échantillonnage . . . . .	23
3.2.3 Echantillonnage descriptif amélioré (EDA) . . . . .	24
3.3 Application . . . . .	27
3.3.1 Interprétation des résultats . . . . .	28

# Introduction générale

Depuis quelques années, la modélisation et l'inférence numérique " bayésiennes " sont de plus en plus présentes en écologie et plus généralement en sciences de la vie. Ceci est dû au développement des méthodes numériques qu'à l'évolution de l'inférence bayésienne elle-même. Ces méthodes sont enrichies de techniques de Monte Carlo pertinentes (MCMC et EDA). L'approche hiérarchique permet de construire des modèles adaptés à ces méthodes numériques. Les techniques de simulation de Monte Carlo sont utilisées pour simuler des systèmes déterministes avec des paramètres ou des entrées stochastiques. Le nom de Monte Carlo a été proposé par les scientifiques du projet Manhattan lors de la deuxième guerre mondiale et fait allusion aux jeux de hasard pratiqués à Monaco.

Les méthodes MC sont aujourd'hui utilisées pour simuler des phénomènes physiques complexes dans plusieurs domaines scientifiques et appliqués : radioactivité, physique des hautes énergies, réseaux, économétrie, logistique. Elles s'appuient sur l'échantillonnage des distributions des quantités incertaines. Elle est l'une des principales applications mathématiques et statistiques utilisant les nombres aléatoires. Cette méthode a pour objet de fournir une approximation numérique de la valeur d'une intégrale (quelque soit sa dimension) d'une fonction donnée pourvu qu'elle soit intégrable.

L'échantillonnage aléatoire est une technique numérique largement utilisée permettant de solutionner des problèmes généralement trop complexes pour qu'une solution analytique soit disponible. De nouvelles méthodes, toujours de même principe que Monte Carlo, sont venues pour combler ses défauts, à savoir l'échantillonnage Descriptif et l'échantillonnage Descriptif Amélioré. Le processus de simulation est généralement très coûteux en terme de temps et d'espace mémoire, ce qui motive l'introduction du parallélisme pour faire face à une demande sans cesse réitérée : traiter vite des problèmes plus grands.

Dans ce mémoire, nous visons à appliquer la méthode EDA dans un sens bayésien, ainsi ce travail est organisé comme suit : Dans le premier chapitre nous avons présenter le bagage nécessaire pour comprendre l'analyse bayésienne, ensuite dans le deuxième chapitre les fondements de la méthode de Monte Carlo sont abordés, et enfin le dernier chapitre aborde une explication détaillée de la méthode EDA avec une application sur un modèle de poisson dans un contexte bayésien.

# Chapitre 1

## L'analyse statistique Bayésienne

### 1.1 Introduction

La statistique bayésienne est une approche statistique fondée sur l'inférence bayésienne, on utilise le terme de "bayésienne" pour la différencier de la statistique fréquentiste (ou statistique classique) qui (où elle donne les mêmes résultats que l'approche bayésienne par des procédés moins coûteux en calcul). La statistique bayésienne est surtout utilisée lorsque l'on n'a que de petits échantillons, typiquement quand chaque observation est elle-même très coûteuse, contrairement à la statistique classique, elle n'exige pas au départ qu'on se fixe une hypothèse précise à confirmer ou infirmer.

Les méthodes bayésiennes constituent une manière différente de construire la connaissance par rapport aux statistiques conventionnelles. Elles offrent la possibilité de mettre à jour ce que nous savons déjà à la lumière de ce que nous apportent nos expérimentations et observations, le tout dans un calcul très intuitif. Elles peuvent résoudre un grand nombre de problèmes rencontrés en statistique classique (multicolinéarité, petits échantillons, données manquantes, modèles très complexes...). On s'attend à ce que ces méthodes remplacent les outils classiques d'inférence dans les années à venir.

Dans ce premier chapitre, nous ferons une brève introduction à l'approche bayésienne ; nous présenterons les notions et les outils sur lesquels se fonde une analyse bayésienne et dont nous aurons besoin pour établir les prochains chapitres de ce mémoire à savoir : les densités a priori et a posteriori, le choix de la loi a priori et l'estimateur bayésien.

On fait l'hypothèse que les variables aléatoires  $X_i$  sont indépendantes et identiquement distribuées (i.i.d). La densité de  $X_i$  dépend d'un paramètre  $\theta$  (inconnu).

## 1.2 La loi a priori

Ce qui fait toute la différence entre la statistique bayésienne et la statistique classique est qu'en statistique bayésienne on traite le paramètre  $\theta$  comme une variable aléatoire, on doit donc lui attribuer une distribution de probabilité, cette distribution est appelée la distribution a priori.

La loi a priori est une transformation, sous une distribution de probabilité, du savoir de l'expert qu'on appelle encore l'expertise déjà connu sur le problème en main en dehors des informations apportées par les résultats expérimentaux. Dans la pratique, l'information a priori peut être codée selon une des façons suivantes :

1. Prendre une loi a priori vague, c'est-à-dire non informative
2. Choisir une loi a priori conjuguée à la vraisemblance
3. Déterminer une loi a priori subjectivement.

### 1.2.1 Le choix d'une loi a priori

Le choix de loi a priori est une étape fondamentale dans l'analyse bayésienne, ce choix peut avoir différentes motivations. Les stratégies sont diverses, elle peuvent se baser sur des expériences du passé, sur une intuition ou sur une idée que le praticien a sur le phénomène aléatoire qu'il est en train de suivre. Elles peuvent être, également, motivées par des aspects calculables.

Enfin, ces stratégies peuvent également tenir compte du fait qu'on ne sait rien par le truchement des lois non informatives.

### 1.2.2 Lois a priori non informatives

Les lois a priori non informatives représentent une ignorance sur le problème en main, mais ne signifient pas que l'on sache absolument rien sur la distribution statistique du paramètre. En effet, on connaît au moins son domaine de variation, c'est-à-dire l'ensemble des états de la nature,  $\Theta$ , et le rôle de chaque composante du paramètre sur les observables.

Ces lois doivent être donc, particulièrement, construites à partir de la distribution de l'échantillonnage, puisque, c'est le seul moyen disponible pour avoir des informations sur le paramètre  $\theta$ . À cet égard, les lois a priori non informatives peuvent être considérées comme des lois de références, aux quelles chacun pourrait avoir recours quand toute information a priori sur  $\theta$  est absente.

Nous décrirons dans ce qui suit, quelques-unes des techniques les plus populaires dans la construction des lois a priori non informatives.

### Lois a priori invariantes

Le fait de formaliser l'absence d'information a priori par une propriété d'invariance est naturelle, au sens où seuls les paramètres de la distribution de  $\theta$  changent lorsqu'on effectue une transformation de  $\theta$ . Par exemple, les distributions de  $\theta$  et de  $(\theta - \theta_0)$ , en réalité, ne sont pas les mêmes, mais dire qu'elles sont les mêmes, c'est-à-dire ;  $\pi(\theta) = \pi(\theta - \theta_0)$ , pour tout  $\theta_0$ , exprime certainement l'ignorance sur la valeur de  $\theta$ . On dit, dans ce cas, que la loi a priori  $\pi$  est invariante par translation et  $\pi(\theta) = c$ , la loi Uniforme sur  $\Theta$ .

Cette technique de construction des lois non informatives n'est que partiellement satisfaisante, car elle implique la référence à une structure d'invariance, qui peut être parfois choisie de plusieurs manières, ne pas exister, ou être sans intérêt pour le décideur.

### Lois a priori de Jeffreys

Jeffreys(1946, 1961) propose une approche intrinsèque qui évite effectivement le besoin de prendre en compte une structure d'invariance potentielle, tout en étant souvent compatible lorsque cette structure existe. Les lois a priori non informatives de Jeffreys sont fondées sur l'information de Fisher, donnée par

$$I(\theta) = -E_{\theta} \left[ \frac{\partial^2 \log f(x|\theta)}{\partial \theta^2} \right]$$

d'où la loi a priori de Jeffreys est

$$\pi(\theta) = I^{1/2}(\theta)$$

Dans le cas où le paramètre  $\theta$  est multidimensionnel, on définit la matrice d'information de Fisher par généralisation s'obtient pour  $\theta \in \mathbb{R}^k$ ,  $I(\theta)$  a les éléments suivants :

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

et la loi non informative de Jeffreys est alors définie par

$$\pi(\theta) = [\det(I(\theta))]^{1/2}$$

La technique de Jeffreys fournit une des meilleures techniques pour construire une loi a priori non informative ; et elle permet bien souvent de retrouver les estimateurs classiques surtout dans des cas unidimensionnels, mais de sa part, elle a été critiquée par certains Bayésiens comme étant un outil sans justification subjective en terme d'information a priori.

### Lois a priori de référence

Une loi a priori de référence est tout simplement une loi a priori non informative construite d'une manière particulière. Mais d'une certaine sorte, toutes les lois a priori non informatives sont des lois de référence du fait que chaque loi a priori non informative peut être considérée comme un point de référence auquel chacun pourrait avoir recours quand toute information sur  $\theta$  est absente.

Cette approche est une modification de l'approche de Jeffreys qui a été proposé par Bernardo (1979), elle repose sur le principe de faire la distinction entre l'importance des paramètres c'est-à-dire, entre les paramètres de nuisance et les paramètres d'intérêt.

Considérons tout d'abord le cas d'un paramètre à deux composantes,  $\theta = (\theta_1, \theta_2)$ , où  $\theta_1$  est le paramètre d'intérêt (de plus importance) et  $\theta_2$  est le paramètre de nuisance, et soit  $x \sim f(x|\theta)$ .

La stratégie introduite par Bernardo est la suivante : pour  $\theta_1$  fixé, on détermine tout d'abord la densité conditionnelle  $\pi(\theta_2|\theta_1)$  comme la loi de Jeffreys associée à  $f(x|\theta)$ , puis on calcule  $\pi(\theta_1)$  qui est la loi de Jeffreys associée à la loi marginale :

$$\tilde{f}(x|\theta_1) = \int f(x|\theta_1, \theta_2) \pi(\theta_2|\theta_1) d\theta_2$$

La loi de référence de  $\theta$  est le produit des deux lois, c'est-à-dire :

$$\pi(\theta_1, \theta_2) = \pi(\theta_2|\theta_1)\pi(\theta_1)$$

Cette manière de faire peut se généraliser si  $\theta = (\theta_1, \dots, \theta_n)$  et si l'on a ordonné sans perte de généralité les  $\theta_i$  par intérêt croissant. Il est clair que ce raisonnement n'est pas purement objectif parce que donner plus d'importance à un paramètre qu'à un autre relève une fois encore d'un choix.

### 1.2.3 Lois a priori conjuguées

Une loi a priori  $\pi(\theta)$  est conjuguée si  $f(x|\theta)$  et  $\pi(\theta)$  appartiennent à la même famille de lois. Ce type de lois a priori est utilisé quand l'information a priori disponible sur le modèle est trop vague ou peu faible. Dans ce cas l'analyste regarde la forme de la fonction de vraisemblance et choisit une famille de lois qui se marie bien avec elle.

Rappelons ici qu'une famille  $\mathcal{F}$  de distributions de probabilité sur  $\Theta$  est dite conjuguée (ou fermée par échantillonnage) par une vraisemblance  $f(x|\theta)$  si pour toute loi a priori  $\pi \in \mathcal{F}$ , la distribution a posteriori  $\pi(\cdot|x)$  appartient également à  $\mathcal{F}$ .

L'avantage des familles conjuguées est avant tout la simplicité des calculs. Avant l'essor du calcul numérique, ces familles étaient pratiquement les seules qui permettaient de faire

aboutir des calculs. L'intérêt principal du caractère conjugué se manifeste quand  $\mathcal{F}$  est paramétrée. Effectivement le passage de la distribution a priori à la distribution a posteriori n'est dans ce cas qu'une mise à jour des paramètres correspondants, ce que nous pouvons le constater dans l'exemple ci-dessus. Et par conséquent, les distributions a posteriori sont toujours calculables dans ce cas.

**Définition 1.2.1. Familles exponentielles**[11]

La famille exponentielle regroupe les lois de probabilité qui admettent une densité de la forme :

$$f(x|\theta) = h(x) e^{\alpha(\theta) T(x) - \psi(\theta)}, \quad \theta \in \Theta$$

$T$  est une statistique exhaustive.

Une telle famille est dite régulière si  $\Theta$  est un ouvert tel que :

$$\Theta = \left\{ \theta \mid \int h(x) e^{\alpha(\theta) T(x)} d\mu(x) < \infty \right\}$$

En outre, on appelle paramétrisation canonique, l'écriture :

$$f(x|\theta) = h(x) e^{\theta T(x) - \psi(\theta)}$$

et famille naturelle l'expression :

$$f(x|\theta) = h(x) e^{\theta T(x)}$$

**Théorème 1.2.1. Famille exponentielles**[11]

Si  $x \sim f(x|\theta) = h(x)e^{\theta T(x) - \psi(\theta)}$ , alors la famille de lois a priori

$$\{\pi_{\lambda, \mu}(\theta) \propto h(x) e^{\theta \mu - \lambda \psi(\theta)}, \lambda, \mu\}$$

est conjuguée. On note que  $\pi_{\lambda, \mu}$  est une densité de probabilité si et seulement si  $\lambda > 0$  et  $\mu/\lambda \in \Theta$ . La loi a posteriori correspondante est  $\pi(\theta|\lambda + 1, \mu + T(x))$ .

En effet,

$$\begin{aligned} \pi_{\lambda, \mu}(\theta|x) &\propto h(x) e^{\theta T(x) - \psi(\theta)} e^{\theta \mu - \lambda \psi(\theta)} \\ &\propto h(x) e^{\theta(T(x) + \mu) - (\lambda + 1)\psi(\theta)} \\ &= \pi_{\lambda + 1, \mu + T(x)} \end{aligned}$$

Le tableau ci-dessous présente quelques lois a priori conjuguées pour quelques familles exponentielles usuelles.

TAB. 1.1 – lois a priori conjuguées pour quelques familles exponentielles usuelles

$f(x \theta)$	$\pi(\theta)$	$\pi(\theta x)$
Normale $\mathcal{N}(\theta, \sigma^2)$	Normale $\mathcal{N}(\mu, \tau^2)$	$\mathcal{N}(\varrho(\sigma^2\mu + \tau^2x), \varrho\sigma^2\tau^2)$ $\varrho = 1/(\sigma^2 + \tau^2)$
Poisson $\mathcal{P}(\theta)$	Gamma $\mathcal{G}(\alpha, \beta)$	$\mathcal{G}(\alpha + x, \beta + 1)$
Gamma $\mathcal{G}(\nu, \theta)$	Gamma $\mathcal{G}(\alpha, \beta)$	$\mathcal{G}(\alpha + \nu, \beta + x)$
Binomiale $\mathcal{B}(n, \theta)$	Bêta $\mathcal{B}e(\alpha, \beta)$	$\mathcal{B}e(\alpha + x, \beta + n - x)$
Binomiale Négative $\mathcal{N}eg(m, \theta)$	Bêta $\mathcal{B}e(\alpha, \beta)$	$\mathcal{B}e(\alpha + m, \beta + x)$
Normale $\mathcal{N}(\mu, 1/\theta)$	Gamma $\mathcal{G}(\alpha, \beta)$	$\mathcal{G}(\alpha + 0.5, \beta + (\mu - x)^2/2)$

### 1.2.4 Lois a priori subjectives

Précisons tout d'abord que cette démarche n'est pas forcément facile dans la pratique. L'idée est d'utiliser les données antérieures. Par exemple, dans un cadre paramétrique, cela revient à choisir une valeur particulière du paramètre.

Dans un cas concret, il peut être judicieux de baser son raisonnement sur les avis des experts, notamment à l'aide de questionnaires. Il est, alors, nécessaire de veiller à ce que les questions soient compréhensibles « Quelle est la loi a priori ? », par exemple en prenant comme base les quantiles plutôt que les moments. Pour plusieurs experts, il peut être utile de pondérer leurs réponses et d'utiliser des modèles hiérarchiques. Ainsi, la difficulté ici n'est pas mathématique mais plus psychométrique pour réduire les biais sur les réponses fournies. Nous allons nous concentrer sur le second aspect de la détermination.

### 1.3 La loi a posteriori

Supposons désormais qu'en plus d'une densité d'échantillonnage,  $f(x|\theta)$ , une densité a priori sur  $\Theta$ ,  $\pi(\theta)$  soit disponible, c'est à dire que nous disposons d'un modèle complètement bayésien. Une fois données ces deux densités nous pouvons en construire plusieurs autres a savoir :

1. La densité jointe de  $(\theta, x)$  :

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

2. La densité marginale de  $x$  :

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

3. La densité a posteriori de  $\theta$ , obtenue par la formule de bayes :

$$\pi(\theta|x) = \frac{f(x|\theta)\pi(\theta)}{m(x)} \propto f(x|\theta)\pi(\theta)$$

Autrement dit, la densité a posteriori représente une actualisation de l'information a priori au vu de l'information apportée par les observation. La loi a posteriori d'un paramètre est une mise à jour de la loi a priori lorsque les observations du processus sont prises en compte.

### 1.4 Fonction de perte et risque

#### Définition 1.4.1.

Soit  $\delta \in \mathcal{D}$  une règle de décision.

Une fonction de perte (de coût) est une fonction mesurable de  $(\mathcal{D} \times \Theta)$  à valeurs dans  $\mathbb{R}_+$  notée  $l(\delta, \theta)$  et définie telle que,

1.  $\forall (\delta, \theta), l(\delta, \theta) > 0$
2.  $\forall \theta, \exists \delta^*$ , tels que :  $l(\delta^*(x), \theta) = 0$

#### Remarque 1.4.1.

1. Une règle de décision  $\delta_1$  est dite meilleure que  $\delta_2$  si son risque associé est moins que celui associé à  $\delta_2$
2. Une règle de décision  $\delta$  est la meilleure de toutes les décisions si et seulement si son risque est le plus petit.
3. Une règle de décision  $\delta$  est dite admissible s'elle est la meilleure des autres.

4. Une règle de décision qui n'est pas admissible est dite inadmissible.

### Définition 1.4.2. Le risque fréquentiste

Pour une fonction de perte donnée  $l(\theta, \delta)$ , la fonction de risque associée est

$$\begin{aligned} R(\delta, \theta) &= E_{\theta}[l(\theta, \delta(x))] \\ &= \int_X l(\theta, \delta(x)) f(x|\theta) d\mu(x) \end{aligned}$$

c'est une fonction de  $\theta$  et ne définit pas un ordre total sur  $\mathcal{D}$  et ne permet donc pas de comparer toutes décisions et estimateurs.

### Définition 1.4.3. Le risque a posteriori

Pour un bayésien,  $\theta$  est une variable aléatoire dite a posteriori et une fois que les données seront disponibles, la distribution pertinente de  $\theta$  le risque a posteriori  $\pi(\theta|x)$  et le risque pertinent sera le risque a posteriori ou bien le risque bayésien défini par

$$\begin{aligned} \rho(\pi, \delta|x) &= E^{\pi}(l(\theta, \delta(x))|x) \\ &= \int_{\Theta} l(\theta, \delta(x)) \pi(\theta|x) d\theta \end{aligned}$$

### Définition 1.4.4. Le risque intégré

Pour une fonction de perte donnée, le risque intégré est défini par

$$\begin{aligned} r(\pi, \delta) &= E(R(\theta, \delta)|x) \\ &= \int_{\Theta} R(\theta, \delta) \pi(\theta) d\theta \end{aligned}$$

## 1.5 L'estimateur bayésien

### Définition 1.5.1. L'estimateur Bayésien

Un estimateur bayésien est la règle de décision  $\delta^{\pi}$  qui minimise  $r(\pi, \delta)$ , c'est-à-dire qui vérifie

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

Pour obtenir la valeur de l'infimum, il faut en théorie minimiser une intégrale double. En effet, nous cherchons à minimiser la fonction de risque bayésienne  $r(\theta, \delta)$ , nous pouvons écrire :

$$\begin{aligned}
 r(\theta, \delta) &= \int_{\theta} R(\delta, \theta) \pi(\theta) d\theta \\
 &= \int_{\theta} \int_x l(\delta, \theta) f(x|\theta) dx \pi(\theta) d\theta \\
 &= \int_{\theta} \int_x l(\delta, \theta) \frac{f(x|\theta) \pi(\theta)}{m_{\pi}(x)} m_{\pi}(x) dx d\theta \\
 &= \int_x \int_{\theta} l(\delta, \theta) \pi(\theta|x) m_{\pi}(x) d\theta dx \\
 &= \int_x \left\{ \int_{\theta} l(\delta, \theta) \pi(\theta|x) d\theta \right\} m_{\pi}(x) dx \\
 &= \int_x \rho(\pi, \delta|x) m_{\pi}(x) dx
 \end{aligned}$$

et minimiser  $r(\pi, \delta)$  pour toute valeur de  $x$ , sera donc équivalent à minimiser la fonction de risque a posteriori

$$\rho(\pi, \delta|x) = \int_{\theta} l(\delta, \theta) \pi(\theta|x) d\theta$$

### Exemple 1. La perte quadratique

Une fonction de perte quadratique est une fonction  $l : (\Theta \times \mathcal{D}) \longrightarrow \mathbb{R}_+$  donnée par

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

Ainsi, soit

$$\begin{aligned}
 f(\delta, x) &= \rho(\pi, \delta|x) = E(l(\delta, \theta)) \\
 &= \int_{\Theta} (\theta - \delta)^2 \pi(\theta|x) d\theta \\
 &= \int_{\Theta} \theta^2 \pi(\theta|x) d\theta - 2\delta \int_{\Theta} \theta \pi(\theta|x) d\theta + \delta^2 \int_{\Theta} \pi(\theta|x) d\theta \\
 &= E(\theta^2|x) - 2\delta E(\theta|x) + \delta^2
 \end{aligned}$$

La décision  $\delta$  qui minimise  $f(\delta, x)$  est celle qui vérifie

$$\frac{d}{d\delta} f(\delta, x) = 0$$

ce qui donne,  $-2E(\theta|x) + 2\delta = 0$  et donc,  $\delta = E(\theta|x)$

Donc pour la perte quadratique, l'estimateur de Bayes est la moyenne de la loi a posteriori.

### Définition 1.5.2. Estimateur admissible

On dit que  $\delta \in \mathcal{D}$  est inadmissible si et seulement si :

$$\exists \delta_0 \in \mathcal{D}, \forall \theta \in \Theta : R(\theta, \delta) \geq R(\theta, \delta_0) \text{ et } \exists \theta_0 \in \Theta : R(\theta_0, \delta) > R(\theta_0, \delta_0).$$

de ce fait,  $\delta$  est dite admissible si elle n'est pas inadmissible et par conséquent, un estimateur est dit admissible si et seulement s'il n'est pas inadmissible.

Un estimateur  $\delta_1$  est inadmissible s'il existe un estimateur  $\delta_2$  qui domine  $\delta_1$ , c'est-à-dire tel que pour tout  $\theta$ ,

$$R(\theta, \delta_1) \geq R(\theta, \delta_2)$$

et pour au moins une valeur  $\theta_0$  du paramètre,

$$R(\theta_0, \delta_1) > R(\theta_0, \delta_2)$$

sinon  $\delta_1$  est dit admissible.

### Théorème 1.5.1. Estimateur Bayésien admissible

*Si l'estimateur bayésien  $\delta^\pi$  associé à une fonction de perte  $l$  et une loi a priori  $\pi$  est unique, alors il est admissible.*

### Théorème 1.5.2.

*Si  $\pi > 0$  sur  $\Theta$ ,  $r(\pi) < \infty$  pour une fonction de perte  $l$  donnée, si  $\delta^\pi$  l'estimateur bayésien correspondant existe et si  $\theta \mapsto R(\theta, \delta)$  est continue. Alors  $\delta^\pi$  est admissible.*

### L'estimateur MAP

On appelle estimateur MAP (estimateur de maximum a posteriori) tout estimateur  $\delta^\pi(x)$  qui maximise l'information sur  $\theta$  représentée par sa loi a posteriori, c'est-à-dire tout estimateur  $\delta^\pi(x)$  tel que  $\delta^\pi(x) \in \underset{\theta}{\text{Argmax}} \pi(\theta|x)$ .  $\delta^\pi(x)$  doit donc être le mode de la distribution a posteriori.

L'estimateur MAP est le pendant bayésien de l'estimateur de maximum de vraisemblance, de ce fait ils partagent les mêmes inconvénients comme : la non unicité, l'instabilité (dus aux calculs d'optimisation) et la dépendance vis-à-vis de la mesure de référence (dominant  $\Theta$ ), seulement l'estimateur MAP ne vérifie pas la non invariance par reparamétrisation qui peut apparaître importante intuitivement.

### Exemple 2.

Supposons qu'on a pas d'information a priori sur  $\theta$ . Pour trouver la loi a priori on choisit la loi de Jeffreys.

La fonction de densité s'écrit :

$$f(x|\theta) = \theta^x(1 - \theta)^{n-x}$$

L'information de Fisher est donnée comme suit

$$I(\theta) = \frac{n}{\theta(1 - \theta)}$$

La loi a priori de Jeffreys s'écrit

$$\pi(\theta) = I^{1/2}(\theta) \propto \sqrt{\frac{n}{\theta(1 - \theta)}}$$

On a :

$$\pi(\theta) = \theta^{-1/2}(1 - \theta)^{-1/2}$$

c-à-d :  $\theta \sim \mathcal{Be}(1/2, 1/2)$

$$\pi(\theta|x) \propto \theta^{x-1/2}(1 - \theta)^{n-x-1/2}$$

Donc :

$$\theta(x) \sim \mathcal{Be}(x + 1/2, n - x + 1/2)$$

Alors

$$\hat{\theta} = E(\theta|x) = \frac{x + 1/2}{n + 1}$$

# Chapitre 2

## Fondements des méthodes de MC

### 2.1 Introduction

Le terme méthode de Monte-Carlo désigne une famille de méthodes algorithmiques visant à calculer une valeur numérique approchée en utilisant des procédés aléatoires, c'est-à-dire des techniques probabilistes. Le nom de ces méthodes, qui fait allusion aux jeux de hasard pratiqués à Monte-Carlo, a été inventé en 1947 par Nicholas Metropolis et publié pour la première fois en 1949 dans un article coécrit avec Stanislaw Ulam.

Les méthodes de Monte-Carlo sont particulièrement utilisées pour calculer des intégrales en dimensions plus grandes que 1. Elles sont également couramment utilisées en physique des particules, où des simulations probabilistes permettent d'estimer la forme d'un signal ou la sensibilité d'un détecteur. La comparaison des données mesurées à ces simulations peut permettre de mettre en évidence des caractéristiques inattendues.

La méthode de simulation de Monte-Carlo permet aussi d'introduire une approche statistique du risque dans une décision financière. Elle consiste à isoler un certain nombre de variables-clés du projet, tels que le chiffre d'affaires ou la marge et à leur affecter une distribution de probabilité. Pour chacun de ces facteurs, un grand nombre de tirages aléatoires est effectué dans les distributions de probabilité déterminées précédemment, afin de trouver la probabilité d'occurrence de chacun des résultats.

Dans ce chapitre, nous ferons une petite introduction à la méthode de Monte Carlo, son principe et la simulation statistique.

## 2.2 Description de la méthode

Pour utiliser une méthode de Monte-Carlo, on doit tout d'abord mettre sous la forme d'une espérance la quantité que l'on cherche à calculer. C'est souvent simple (calcul d'intégrale par exemple) mais peut-être plus compliqué (équations aux dérivées partielles par exemple).

Une application classique des méthodes de Monte-Carlo est le calcul des quantités du type :

$$I = E[g(X)] = \int g(x)f(x)dx$$

où  $g : \mathbb{R}^d \rightarrow \mathbb{R}$  est une fonction donnée et  $X$  un vecteur aléatoire de densité  $f$  suivant laquelle on sait simuler. Dans ce contexte, l'estimateur de Monte Carlo est défini par

$$\hat{I}_n = \frac{1}{n} \sum_{i=1}^n g(X_i),$$

où les  $X_i$  sont générées de façon *i.i.d.* selon  $f$ .

### Théorème 2.2.1.

Soit  $(U_i)_{i \in \mathbb{N}}$  une suite de variables aléatoires indépendantes et de loi Uniforme sur  $[0, 1]$  alors

$$\frac{1}{n} \sum_{i=1}^n g(U_i) \rightarrow E[g(U)] \text{ p.s.}$$

En d'autres termes, si  $u_1, u_2, \dots, u_n$  sont des nombres tirés au hasard dans  $[0, 1]$

$$\frac{1}{n} [g(u_1) + g(u_2) + \dots + g(u_n)]$$

est une approximation de l'intégrale  $\int_0^1 g(x)dx$ .

Plus généralement prenons l'exemple d'une intégrale du type :  $\int_{\mathbb{R}} g(x)f(x)dx$  où  $f(x) \geq 0$  et  $\int_{\mathbb{R}} f(x)dx = 1$ , alors  $I = E[g(X)]$ .

Toujours par la loi des grands nombres, si  $(X_i)_{i \in \mathbb{N}}$  est une suite de variables aléatoires indépendantes sur  $\mathbb{R}$  de loi de densité  $f$ ,

$$\frac{1}{n} \sum_{i=1}^n g(X_i) \rightarrow E[g(X)] \text{ p.s.}$$

et donc si  $(x_1, x_2, \dots, x_n)$  est une réalisation de  $(X_1, X_2, \dots, X_n)$ ,

$$\frac{1}{n} [g(x_1) + g(x_2) + \dots + g(x_n)]$$

est une approximation de  $I$ .

### 2.2.1 Convergence de la méthode

C'est la loi forte des grands nombres qui permet de justifier la convergence de la méthode de Monte Carlo et le théorème de la limite central qui précise la vitesse de convergence.

#### **Théorème 2.2.2. Lois des grands nombres**[13]

*La méthode de Monte Carlo trouve son fondement mathématique dans la loi des grands nombres.*

*Soit  $(X_i)_{i \geq 1}$  une suite de réalisations de la variable aléatoire  $X$ .*

*Si  $E|g(X)| < \infty$ , alors*

$$\hat{I}_n = \frac{1}{n} \sum_{i=1}^n g(X_i) \xrightarrow[n \rightarrow \infty]{p.s.} I$$

Il est clair que l'estimateur  $\hat{I}_n$  est sans biais, c'est-à-dire que  $E[\hat{I}_n] = I$ . La loi forte des grands nombres assure qu'il est convergent.

#### **Exemple 3. Estimateur de $(\pi)$ .**

Supposons que  $(X, Y)$  suit la loi Uniforme sur le carré  $C = [0, 1] \times [0, 1]$  et que

$$\varphi(x, y) = \mathbf{1}_{x^2+y^2 \leq 1}$$

En notant  $D = \{(x, y) \in \mathbb{R}_+^2, x^2 + y^2 \leq 1\}$ , le quart de disque unité, on a donc

$$I = \int \int_C \mathbf{1}_D(x, y) dx dy = \lambda(D) = \frac{\pi}{4}$$

Simuler des points  $(X_i, Y_i)$  uniformément dans  $C$  est facile la propriété précédente assure donc que

$$\hat{I}_n = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_D(x_i, y_i) \xrightarrow[n \rightarrow \infty]{p.s.} \frac{\pi}{4} \iff 4 \times \hat{I}_n \xrightarrow[n \rightarrow \infty]{p.s.} \pi$$

On dispose donc d'un estimateur Monte-Carlo pour la constante  $\pi$ .

#### **Remarque 2.2.1.**

La méthode de Monte Carlo ne peut donc s'utiliser que des variables aléatoires intégrables.

Pour avoir une idée de l'intérêt de la méthode, il faut pouvoir évaluer l'erreur commise.

#### **Définition 2.2.1.**

L'erreur est définie par  $\epsilon_n = E(X) - \frac{1}{n}(X_1 + \dots + X_n)$ .

Le théorème de la limite centrale donne un asymptotique de l'erreur  $\epsilon_n$ , mais de nature aléatoire. Il dit que la loi de l'erreur finit par ressembler à une loi Gaussienne Centrée.

**Théorème 2.2.3. Théorème Central Limite**[13]

Soit  $X_i, i \geq 1$  une suite de réalisation de la variable aléatoire  $X$ .

On suppose que  $E(X^2) < +\infty$  on note  $\sigma^2$  la variance de  $X$  ( $\sigma^2 = E(X^2) - (E(X))^2$ ), alors  $\frac{\sqrt{n}}{\sigma}\epsilon_n$  converge en loi vers une gaussienne centrée réduite  $G$ .

Cela signifie que si  $G$  est une variable aléatoire de loi

$$\int \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$$

et si  $f$  est une fonction continue bornée;  $E(f(\frac{\sqrt{n}}{\sigma}\epsilon_n))$  converge vers

$$E(f(G)) = \int_{-\infty}^{+\infty} f(x) \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$$

**Remarque 2.2.2.**

On peut aussi montrer que pour tout  $C_1$  et  $C_2$

$$\lim_{n \rightarrow \infty} P\left(\frac{\sigma}{\sqrt{n}}C_1 \leq \epsilon_n \leq \frac{\sigma}{\sqrt{n}}C_2\right) = \int_{C_1}^{C_2} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$$

Dans les applications, on oublie le passage à la limite et on remplace  $\epsilon_n$  par une gaussienne centrée de variance  $\frac{\sigma^2}{n}$ .

**Exemple 4. Estimateur de ( $\pi$ )**

Dans ce cas, la variance vaut tout simplement

$$\sigma^2 = I - I^2 = \frac{\pi}{4} \left(1 - \frac{\pi}{4}\right)$$

laquelle est donc estimée par

$$\hat{\sigma}_n^2 = \hat{I}_n - \hat{I}_n^2$$

## 2.2.2 Principe de la méthode

### Dans un sens classique

À la base de la méthode de Monte Carlo est que d'après la loi des grands nombres, on peut approximer une espérance mathématique d'une variable  $X$  par la moyenne arithmétique de  $n$  réalisations indépendantes de cette variable. Cette espérance mathématique s'exprime comme une intégrale d'une certaine fonction  $g$ .

Soit  $I = E(g(X))$  où

$X$  est une variable aléatoire de densité  $f$ . Il s'en suit en vertu de la loi des grands nombres qu'une approximation de l'intégrale  $I$  est donnée par la moyenne arithmétique de  $n$  réalisations  $x_i$  de v.a  $g(X)$

$$I \approx \frac{\sum_{i=1}^n g(x_i)}{n}.$$

où

$x_i$  sont  $n$  réalisations d'une variable aléatoire  $X$  ayant  $f$  pour densité de probabilité.

Il convient cependant de noter que dans les applications, on dispose rarement d'un échantillon de  $n$  réalisations de variable aléatoire  $X$ , par application de la méthode, on peut disposer de  $n$  valeurs simulées de  $X$ . Ces valeurs peuvent alors remplacer les vraies réalisations dans la formule d'approximation, l'erreur commise par ce remplacement ne devrait pas être a priori importante surtout lorsque le générateur utilisé est de bonne qualité.

On a approximer l'intégrale  $I$  par la quantité suivant :

$$\hat{I} = \frac{\sum_{i=1}^n g(x_i)}{n}$$

où

Les  $x_i$  sont  $n$  valeurs simulées d'une variable aléatoire  $X$  ayant  $f$  pour densité de probabilité.

La quantité  $\hat{I}$  est appelée l'approximation de  $I$  par la méthode de Monte Carlo.

### Dans un sens bayésien

Dans un contexte bayésien le principe des méthodes Monte Carlo reste le même, il se base sur la loi des grands nombres, seulement dans un sens bayésien on simule selon la loi a posteriori au lieu de la loi de la variable.

On suppose que l'on souhaite calculer  $E[g(\theta)|x]$  (la moyenne a posteriori) où  $g(\theta)$  est une fonction réel et  $x$  désigne l'ensemble des observations.

On suppose que l'on sait simuler selon la loi a posteriori, le principe de ces méthodes est de simuler une suite de variables aléatoires  $\theta_i$ ,  $i = \overline{1, n}$  (i.i.d) selon la loi a posteriori, puis d'approcher  $E[g(\theta)|x]$  à l'aide de la loi des grands nombres :

$$\frac{1}{n} \sum_{i=1}^n g(\theta_i) \xrightarrow{n \rightarrow \infty} E[g(\theta)|x]$$

## 2.3 Simulation statistique

La simulation est une technique de modélisation du monde réel, elle permet de représenter le fonctionnement d'un système composé de différents centres d'activités, de mettre en évidence les caractéristiques de ceux-ci et les interactions entre eux, de décrire la circulation des différents objets traités par ces processus et enfin d'observer le comportement du système dans son ensemble et son évolution dans le temps. La formalisation du phénomène peut s'effectuer aux moyens de modèles mathématiques qui présentent l'avantage de fournir des résultats reflétant une évolution stable du système dans le temps. Les étapes de la simulation sont exécutées dans l'ordre suivant : Formulation du problème, modélisation, validation du modèle, programmation et vérification du problème.

### 2.3.1 Simulation de Monte Carlo

Notre intérêt est porté sur la simulation stochastique appelée aussi simulation de Monte Carlo. Dans cette dernière, des modèles de simulation contenant une ou plusieurs variables aléatoires pour les quelles des nombres aléatoires, suivant une loi Uniforme doivent être générés afin d'obtenir des échantillons des distributions correspondantes. Le but d'une telle simulation est d'estimer les paramètres inconnus du système étudié les modèles de gestion de stocks et les files d'attentes.

La simulation de Monte Carlo est une expérience d'échantillonnage, c'est à dire que les nombres aléatoires sont utilisés pour générer seulement quelques réponses parmi toutes les réponses possibles qui correspondent à toutes les permutations de nombres aléatoires. Dans la simulation, la procédure d'échantillonnage doit extraire autant d'information possible des variables aléatoires d'entrée. Ceci s'opère en représentant une variable stochastique à travers un échantillon qui imite complètement son comportement aléatoire.

### 2.3.2 Algorithme de Monte Carlo

On peut calculer l'approximation de Monte Carlo en utilisant l'algorithme suivant :

- Générer en utilisant un générateur de bonne qualité  $n$  nombres aléatoires (valeurs simulées d'une variable aléatoire suivant la loi Uniforme continue sur  $[0, 1]$  ) :  $U_1, U_2, \dots, U_n$ .
- Choisir une densité  $f$  pour simuler à partir des  $U_i$ ,  $n$  valeurs d'une variable  $X$ .
- Déterminer les quantités :  $g(X_1), g(X_2), \dots, g(X_i), \dots, g(X_n)$ .
- Calculer la moyenne arithmétique de ces quantités donnant  $\hat{I}$

tq :

$$\hat{I} = \frac{\sum_{i=1}^n g(x_i)}{n}$$

#### Exemple 5.

On approxime par la méthode de Monte Carlo l'intégrale suivante :

$I = \int_0^2 e^{-x^2} dx$  , il convient au préalable d'écrire  $I$  comme une espérance mathématique en

se donnant une densité de probabilité  $f$  sur  $[0, 2]$  facile à simuler . On choisit pour cet exemple la densité de la loi Uniforme continue sur  $[0, 2]$  qui est la plus simple.

Soit,  $f(x) = \frac{1}{2} \mathbf{1}_{[0,2]}$  d'où, on déduit directement :

$$I = \int_0^2 2e^{-x^2} \frac{1}{2} dx$$

Soit,  $I = E(2e^{-x^2})$  où  $X \longrightarrow U(0, 2)$  et donc :

$$\hat{I} = \frac{\sum_{i=1}^n 2e^{-x_i^2}}{n}$$

(Les  $X_i$  étant des valeurs simulées de  $X \longrightarrow U(0, 2)$ ).

Définir l'approximation de  $I$  par la méthode de Monte Carlo, en conséquence pour trouver cette approximation, on peut procéder comme suit ( $n$  étant donnée) :

- Générer  $n$  valeurs indépendantes  $U_1, U_2, \dots, U_i, \dots, U_n$  en utilisant un bon générateur.
- Calculer  $X_1, X_2, \dots, X_i, \dots, X_n$  par la méthode d'inversion, soit  $X_i = 2U_i$
- Calculer  $g(X_i) = 2e^{-x_i^2}$  pour  $i = 1$  à  $n$  et leur moyenne arithmétique simple pour trouver  $\hat{I}$ .

## 2.4 Méthode d'inverse

### Lemme 2.4.1. Fonction d'inverse

Soit  $X$  une variable aléatoire réelle de fonction de répartition  $F$ , posons pour  $0 \leq t \leq 1$ ,

$$F^{-1}(u) = \inf(x, F(x) \geq u)$$

Alors, si  $U \sim \mathcal{U}([0, 1])$ , alors  $F^{-1}(U)$  a même loi que  $X$ .

### Exemple 6.

Nous avons :

$$F(x) = \begin{cases} \frac{x}{3} & \text{si } x \in [0, 1[ \\ \frac{x}{3} + \frac{1}{3} & \text{si } x \in [1, 2] \\ 1 & \text{si } x \geq 1 \\ 0 & \text{si } x \leq 0 \end{cases}$$

$$F^{-1}(u) = \begin{cases} 3u & \text{si } u \in [0, \frac{1}{3}] \\ 1 & \text{si } u \in [\frac{1}{3}, \frac{2}{3}] \\ 3u - 1 & \text{si } u \in [\frac{2}{3}, 1] \end{cases}$$

### Exemple 7. Loi exponentielle

Si  $U$  suit la loi Uniforme sur  $]0, 1[$

$$Y = -\frac{\ln u}{\lambda}$$

suit la loi exponentielle de paramètre  $\lambda$ . En fait ici  $F(x) = 1 - \exp(-\lambda x)$  s'inverse en

$$F^{-1}(u) = -\frac{\ln(1-u)}{\lambda}$$

mais on exploite le fait que  $1 - U$  est de même loi que  $U$ .

### Exemple 8. Loi de Weibull.

La loi de Weibull est très utilisée en fiabilité. La loi  $\text{Weib}(a, b, c)$  de paramètres  $a > 0$ ,  $b \geq 0$  et  $c > 0$  est caractérisée par sa fonction de survie  $G(x) = 1 - F(x)$  donnée par

$$G_{(a,b,c)}(x) = \exp\left(-\left(\frac{x-b}{c}\right)^a\right), \text{ pour } x \geq b$$

clairement  $b$  est un paramètre de localisation et  $c$  un paramètre d'échelle de sorte que  $\text{Weib}(a, b, c)$  se déduit par translation et changement d'échelle de la loi  $\text{Weib}(a) = \text{Weib}(a, 0, 1)$  de fonction de survie

$$G_a(x) = \exp(-x^a), \text{ pour } x \geq 0$$

La simulation de la loi  $\text{Weib}(a, b, c)$  se ramène ainsi à celle de  $\text{Weib}(a)$ . En exploitant à nouveau le fait que  $U$  et  $1 - U$  ont même loi, on voit immédiatement que  $Y = (-\ln U)^{1/a}$  suit la loi  $\text{Weib}(a)$ .

# Chapitre 3

## La méthode EDA

### 3.1 Introduction

Lorsque l'on souhaite produire des estimations sur une population, l'échantillonnage permet de réduire le volume de données à traiter. Utilisé dans les enquêtes, il permet de sélectionner aléatoirement un échantillon d'unités dans une base de sondage, unités qui seront ensuite enquêtées pour inférer sur la population d'étude. La méthode d'échantillonnage utilisée est généralement un compromis entre la recherche d'estimations précises, et la nécessité de respecter un budget imposé. L'échantillonnage peut également être utilisé en situation de données volumineuses, afin de se restreindre à un volume de données à traiter compatible avec ses ressources informatiques. Dans ce cas, le choix de la méthode d'échantillonnage repose également sur des contraintes techniques (algorithme séquentiel).

Dans ce chapitre, nous proposons une introduction aux méthodes d'échantillonnage en population finie. Après nous enons quelques rappels sur la méthode d'échantillonnage descriptif amélioré. Nous présenterons également l'algorithme d'EDA. La méthode EDA sera illustrée par une application.

### 3.2 Echantillonnage aléatoire

#### Définition 3.2.1.

Un échantillon aléatoire de taille  $n$  de la variable aléatoire  $X$  est un vecteur aléatoire. où est le vecteur  $x_i$  ;  $i = 1, \dots, n$  sont des variables aléatoires indépendantes et de même loi que la variable aléatoire  $X$ . En utilisant l'échantillonnage aléatoire, considérons le cas d'un problème de simulation avec une variable aléatoire d'entrée et définissons les estimateurs de la simulation dans les deux cas suivants :

**Premier cas**

Une variable de sortie possédant  $k$  paramètres à estimer,  $\theta_j$ ,  $j = 1, 2, \dots, k$ .  
 Dans ce cas, les estimateurs de la simulation obtenus sont définis par :

$$Y_j = F_j(u_1, u_2, \dots, u_n), j = 1, 2, \dots, k$$

où

-  $Y_j$  est l'estimateur du paramètre inconnu  $\theta_j$ ,  $j = 1, 2, \dots, k$ .

-  $F_j$ ,  $j = 1, 2, \dots, k$  est la fonction de simulation, généralement définie par un programme informatique, reliant les valeurs de la variable d'entrée et chaque estimateur.

-  $u_1, u_2, \dots, u_n$  sont des nombres aléatoires indépendants uniformément distribués entre 0 et 1.

En simulation, une histoire crée un estimateur pour chaque paramètre étudié. Chaque histoire crée donc  $k$  estimateurs car nous avons  $k$  paramètres à estimer.

**Deuxième cas**

$m$  variables de sortie dont chacune possède  $k$  paramètres à estimer suivants :  $\theta_{j1}, \theta_{j2}, \dots, \theta_{jk}$ ,  $j = 1, 2, \dots, m$ . Dans ce cas, les estimateurs de la simulation obtenus sont définis par :

$$Y_{1l} = F_{1l}(u_1, u_2, \dots, u_n)$$

$$\vdots$$

$$Y_{jl} = F_{jl}(u_1, u_2, \dots, u_n)$$

$$\vdots$$

$$Y_{ml} = F_{ml}(u_1, u_2, \dots, u_n), l = 1, 2, \dots, k.$$

où  $Y_{jl}$ ,  $j = \overline{1..m}$  sont les estimateurs des  $\theta_{jl}$  et chaque histoire crée  $km$  estimateurs.

**3.2.1 Echantillonnage descriptif****Définition 3.2.2.**

L'échantillonnage descriptif est une méthode qui contrôle complètement l'ensemble des valeurs de l'échantillon. Elle est basée sur un choix régulier de l'ensemble des valeurs de l'échantillon et sur leurs permutations aléatoires. Cette méthode génère deux types de problèmes : le biais des estimateurs des paramètres des variables de sortie et la connaissance a priori de la taille de l'échantillon.

### 3.2.2 La procédure d'échantillonnage

Supposons que le problème de simulation en étude contient une variable aléatoire d'entrée  $X$ . Soit  $H$  sa fonction de répartition et  $H^{-1}$  son inverse telle que :  $X = H^{-1}(R)$  et  $R$  suit la loi Uniforme entre 0 et 1.

La procédure de génération d'échantillons descriptifs est la suivante :

- On subdivise l'intervalle  $[0, 1]$  en  $n$  sous-intervalles équiprobables.

Soit  $\{r_i, i = 1, 2, \dots, n\}$  l'ensemble de  $n$  points réguliers

où

$r_i$  est le milieu du  $i^{eme}$  sous intervalles tels que :

$$r_i = \frac{i - 0.5}{n}, \quad i = 1, 2, \dots, n$$

- On génère les valeurs de l'échantillon en utilisant l'échantillonnage sans remise.

$$x_{di} = H^{-1}(r_i), \quad i = 1, 2, \dots, n$$

- On permute aléatoirement l'ensemble des valeurs  $\{x_{di}; i = 1, 2, \dots, n\}$

- On garde en mémoire les valeurs de l'échantillon afin de les utiliser à la demande de la simulation.

Dans une histoire, les estimateurs obtenus par la simulation, en utilisant l'ED sont donnés par :

$$(Yr)_j = F_j(r_1, r_2, \dots, r_n), \quad j = 1, 2, \dots, k$$

où

$(Yr)_j$  est l'estimateur du paramètre inconnu  $\theta_j, j = 1, 2, \dots, k$ .

$r_1, r_2, \dots, r_n$  sont des nombres réguliers dépendants uniformément distribués entre 0 et 1.

Etant donné que l'échantillonnage descriptif utilise des nombres réguliers, on parle alors de la surface des réponses régulière. Sa définition est identique à celle de la surface des réponses sauf qu'au lieu de considérer l'ensemble  $\{u_1, u_2, \dots, u_n\}$ , nous utilisons l'ensemble  $\{r_1, r_2, \dots, r_n\}$ .

### 3.2.3 Echantillonnage descriptif amélioré (EDA)

#### Définition 3.2.3.

L'Echantillonnage Descriptif Amélioré (EDA) est une méthode d'échantillonnage qui peut être utilisée pour produire des valeurs d'entrée pour l'estimation de l'espérance de fonctions de variables de sortie. Un estimateur d'EDA est défini et a été démontré qu'il est sans biais et efficace par rapport à l'échantillonnage aléatoire simple selon le critère de la variance pour une classe d'estimateurs.

C'est une méthode proposée par  $M^{me}$  M.Tari, cette méthode est basée sur le principe de la méthode définie précédemment (ED), elle élimine les inconvénients cités plus haut et analyse les conditions où le biais se produit. Cette méthode considère un bloc d'ensembles de nombres réguliers  $r_i$  de cardinal des nombres premiers  $p_i$  et distribue les observations  $X_i$  à la demande de la simulation.

On arrête le processus de génération des nombres lorsque la simulation se termine. L'échantillon généré est obtenu par  $X_i = F^{-1}(r_i)$

où

$$r_i = \frac{i - 0.5}{p}, \quad i = 1, \dots, p.$$

#### Remarque 3.2.1.

Si  $M$  histoires répliquées sont nécessaires, alors  $M$  bloc de  $m_1, m_2, \dots, m_M$  ensembles de nombres réguliers seront nécessaires à considérés.

Pour réduire le risque du biais de la méthode précédente, une approche a été proposée, qui consiste en un bloc qui doit être situé à l'intérieur d'un générateur visant à distribuer aléatoirement des sous-ensembles réguliers de tailles de nombre premiers  $p_q, p_{q+1} \dots$  pour tout  $q$  dans un ordre aléatoire à la demande de la simulation. Nous arrêtons de produire ces sous suites lorsque la simulation s'arrête (le critère d'arrêt est la durée de la simulation dans une exécution).

Dans cette approche, chaque exécution est déterminée par un bloc de différents nombres premiers. Si on a besoin de  $M$  exécutions répétées, il est nécessaire de considérer  $M$  blocs de sous-ensembles réguliers  $m_1, m_2, \dots, m_M$ , les nombres premiers et les valeurs de la variable aléatoire des sous-ensembles ne sont pas les mêmes pour toutes les exécutions répliquées.

**Principe :**

- Générer un nombre premier  $p_i$  puis calculer les sous suites de nombres

$$r_i = \frac{i - 0.5}{p_1}, i = \overline{1, p_1}$$

- Mélanger les  $r_i$ .
- Générer un deuxième nombre premier  $p_2$  et calculer

$$r_i = \frac{i - 0.5}{p_2}, i = \overline{1, p_2}$$

**Les avantages :**

Elle garde les avantages de la méthode précédente de plus cette approche élimine la nécessité de déterminer à l'avance la taille de l'échantillon. Cette nouvelle approche réduit le biais pouvant être provoqué par l'échantillonnage descriptif (Les estimateurs sont sans biais). De plus, elle évite la connaissance a priori de la taille de l'échantillon.

**Algorithme EDA**

- (a) Initialisation de l'expérience

- 1- Avant chaque début de l'histoire, générer un nombre premier aléatoirement.
- 2- Générer les nombres réguliers  $r_i$ ;  $i = 1, 2, \dots, p$  et les sauvegardés dans un vecteur  $R$ .

- (b) Initialisation de la sous histoire

Au début de chaque sous histoire, poser  $i_p = 1$ .

- (c) Echantillonnage sans remise durant la sous histoire

- 3- Si  $i_p > P$  aller à (d).
- 4- Générer aléatoirement un entier naturel  $iaux [i_p, p]$
- 5- Permuter  $r(i_p)$  avec  $r(iaux)$ .
- 6- Générer une observation  $X_i$  correspondante. Si aucune valeur de l'échantillon descriptive n'est demandée, arrêter et collecter les résultats finaux du dernier nombre premier utilisé et aller à (e)
- 7- Sinon, soit  $i_p = i_p + 1$ , aller à (3).

- (d) Collecter les résultats après chaque sous-histoire et aller à (1)
- (e) Collecter les résultats après chaque histoire.

### 3.3 Application

Nous visons dans cette section d'appliquer la méthode EDA abordée dans ce chapitre pour estimer les paramètres d'un modèle de défaillances de dix pompe dans une centrale nucléaire. Le modèle fait l'hypothèse que les défaillances de la pompe  $i$  suivent un processus de Poisson de paramètre  $\lambda_i$ ,  $i = \overline{1,10}$  pour un temps d'observation  $t_i$ . Les données figurent dans le tableau (3.1).

En 2012, Benbait Soumia et Hank Nawal avec Mme Djeridi ont considéré une estimation bayésienne pour le même modèle, mais en utilisant les méthodes MCMC. Elles ont supposés dans un premier cas que les  $\lambda_i$  suivent une loi a priori conjuguée  $(\alpha, \beta)$  avec  $\alpha = 1.8$  et  $\beta$  inconnu, donc, elles ont proposé une loi a priori pour  $\beta$  qui est  $(\gamma, \sigma)$  avec  $\gamma = 0.01$  et  $\sigma = 1$ .

Après calcul de la loi a posteriori, elles ont appliqué les méthodes MCMC pour estimer les  $\lambda_i$ ,  $i = \overline{1,10}$  et  $\beta$  (La référence [5]).

Les résultats de la simulation figurent dans le tableau (3.2).

En gardant la même loi a priori pour :

$$\lambda_i \sim \mathcal{G}(\alpha, \beta), \quad i = \overline{1,10}$$

avec  $\alpha = 1.8$ , mais  $\beta$  constant et égale à 1, dans notre cas, nous allons appliquer la méthode EDA pour estimer les paramètres  $\lambda_i$ ,  $i = \overline{1,10}$ .

Nous avons

$$X \sim \mathcal{P}(\lambda_i t_i), \quad i = \overline{1,10}$$

c-à-d

$$\mathcal{P}(X|\lambda_i) = \frac{e^{-\lambda_i t_i} (\lambda_i t_i)^{x_i}}{x_i!}$$

$$\lambda_i \sim \mathcal{G}(\alpha, \beta), \quad i = \overline{1,10} \quad \alpha = 1.8 \quad \text{et} \quad \beta = 1$$

c-à-d

$$\pi(\lambda_i) = \frac{1}{\Gamma(\alpha)} \lambda_i^{\alpha-1} e^{-\lambda_i}$$

En applique le théorème de Bays, la distribution a posteriori est donné par :

$$\begin{aligned}\pi(\lambda_1, \dots, \lambda_{10}, \beta | x_1, \dots, x_{10}, t_1, \dots, t_{10}) &\propto \mathcal{P}(X | \lambda_i) \pi(\lambda_i) \\ &\propto \left( \prod_{i=1}^{10} \frac{e^{-\lambda_i t_i} (\lambda_i t_i)^{x_i}}{x_i!} \frac{1}{\Gamma(\alpha)} \lambda_i^{\alpha-1} e^{-\lambda_i} \right) \\ &\propto \prod_{i=1}^{10} e^{-\lambda_i t_i} (\lambda_i t_i)^{x_i} \lambda_i^{\alpha-1} e^{-\lambda_i} \\ &\propto \prod_{i=1}^{10} e^{-\lambda_i (1+t_i)} (\lambda_i t_i)^{x_i} \lambda_i^{\alpha-1}\end{aligned}$$

qui est une *Gamma*( $x_i + \alpha, t_i + 1$ ).

Comme nous l'avons indiqué précédemment, la méthode EDA consiste à générer des nombres premiers  $p_1, \dots, p_m$  et pour chaque nombre premier on effectue la sous histoire suivante :

1. On calcule les nombres :  $r_i = \frac{i-0.5}{p_i}$ ,  $i = \overline{1.p_i}$  et  $x_i = P(r_i)$   
où  $P$  est la loi a posteriori.
2. On permute les  $x_i$  pour chaque  $p_i$ , on obtient des valeurs  $x'_i$ .
3. Pour chaque sous histoire l'estimateur EDA du paramètre  $\lambda_i$  est

$$\hat{\lambda}_{ij} = \frac{1}{p_i} \sum_{i=1}^n x'_i, \quad i = \overline{1.10} \text{ et } j = \overline{1.m}$$

On fait ces étapes pour chaque paramètre  $\lambda_i$ ,  $i = \overline{1.10}$ .

Et nous avons appliqué cette procédure pour 3 nombres premiers  $p = 5$ ,  $p = 7$  et  $p = 11$ . Nous avons utilisé le langage R pour effectuer les calculs, les résultats sont stockés dans le tableau (3.3).

### 3.3.1 Interprétation des résultats

On remarque d'après le tableau (3.3) que pour chaque  $p_i$ ,  $i = \overline{1.3}$  les estimateurs  $\hat{\lambda}_i$  sont proches entre eux.

Et si on compare ses résultats avec ceux obtenus par Mme Djeridi donnés dans le tableau (3.2), on constate que les estimateurs obtenus sont proches des estimateurs obtenus par MCMC.

Alors, on peut constater d'après cette étude que les méthode MCMC et la méthode EDA donnent les même résultats, donc on peut dire qu'elles sont de même importance.

TAB. 3.1 – Nombre de défaillances et temps d'observation pour dix pompes d'une centrale nucléaire

(Source : Graver and O'Muircheartaigh 1987)

Pompe	1	2	3	4	5	6	7	8	9	10
Défaillances	5	1	5	14	3	19	1	1	4	22
Temps	94	16	63	126	5	31	105	105	2	10

TAB. 3.2 – Résultats de simulation de l'échantillonnage de Gibbs

Paramètres	moyenne
$\lambda_1$	0.0700731
$\lambda_2$	0.1590098
$\lambda_3$	0.1048920
$\lambda_4$	0.1241629
$\lambda_5$	2.1848445
$\lambda_6$	0.6307686
$\lambda_7$	1.0208779
$\lambda_8$	1.0380995
$\lambda_9$	1.5512630
$\lambda_{10}$	2.0278068
moy	0.8911798

TAB. 3.3 – Résultats de la simulation par EDA

$\lambda_i$	$p = 5$	$p = 7$	$p = 11$
$\lambda_1$	0.07016496	0.07037402	0.07057604
$\lambda_2$	0.1595352	0.1500842	0.1579864
$\lambda_3$	0.1042441	0.1042441	0.1042441
$\lambda_4$	0.1241936	0.1247882	0.1247882
$\lambda_5$	1.354923	1.354923	1.354923
$\lambda_6$	0.6368817	0.6321082	0.6394699
$\lambda_7$	0.04941772	0.03591173	0.04020137
$\lambda_8$	0.04017198	0.04038216	0.04323801
$\lambda_9$	1.565845	1.56577	1.543762
$\lambda_{10}$	2.018532	2.028043	2.026159
moy	0.6123909	0.6106629	0.6105348

# conclusion

Dans ce travail, nous avons étudié une approche bayésienne en employant la méthode EDA pour calculer l'estimateur d'un modèle de poisson hiérarchique.

D'après ce mémoire, on peut constater que la méthode EDA a une même importance que les méthodes MCMC dans la statistique bayésienne. Aussi on peut remarquer la valeur qui ont ajouté ces méthodes à la statistique bayésienne qui est devenue de plus en plus très flexible est demandée dans divers domaines.

# Bibliographie

- [1] **Aissani.A.** Modélisation et Simulation, 2007.
- [2] **Baranger.c, Mathiand.J.** Méthode de Monte Carlo. ENSTA, (2012/2013).
- [3] **Bontempi.G.** Modélisation et simulation. Département d'Informatique, INFO-F-305, 2008.
- [4] **Boreux.J, Parent.E, Bernier.J.** Pratique du Calcul Bayésien.Springer verlag. France, Paris, 2010.
- [5] **Benbait.S et Hank.N.** Choix de La loi a priori dans la Statistique Bayésienne : Mémoire de Master, 2012.
- [6] **Dimov.** Monte Carlo methods for applied scientists. World Scientific Publishing Co, 2008.
- [7] **Guyader.A.** Simulations Monte-Carlo. Université Pierre et Marie curie. Master Mathématiques et Applications spécialité Statistique, 2016/2017.
- [8] **Mathlouthi.H.** Cours de méthodes de simulation, (2014/2015)
- [9] **Millet.A.** méthode de Monte Carlo. Université Paris, 2011.
- [10] **Pardoux.E.** La méthode de Monte Carlo. Marseille, 13/9/2006.
- [11] **Robert.C.P.** Le choix Bayésien : Principe et pratique. Springer-verlag France, paris, 2006.
- [12] **Rousseau.J.** Statistique Bayésienne Notes de cours. (2009/2010)

- [13] **Rubenthaler.S.** Méthodes de Monte Carlo (cours et exercices). Université nice Sophia Antipolis, (2015-2016).
- [14] **Tari.M, Dahmani.A.** Refined descriptive Sampling : A better approach to Monte Carlo simulation. *Journal Simulation Modelling*, 1 april 2005.

## Résumé

Dans ce travail, nous présentons une nouvelle méthode pour faire une estimation des paramètres d'un modèle dans un contexte bayésien. Dans un premier temps, nous donnons des notions sur la statistique bayésienne et la méthode de Monte Carlo. En outre, nous s'adressons à la méthode d'EDA et nous avons appliqué cette méthode à un modèle de Poisson hiérarchique à l'aide de langage R.

## Mots clés

Méthode Monte-Carlo, Méthode ED, Méthode EDA, Statistique Bayésienne.

## Abstract

In this work, we present a new method for estimating parameters of a model in a bayesian context. In first step, we will define the bayesian statistic and the method of Monte Carlo. In a second step, we address the EDA method and we have applied this method on hierarchical the Poisson model using R language.

## Keywords

Monte-Carlo method, ED method, EDA method, Bayesian statistic.

## ملخص

في هذا العمل نقدم طريقة جديدة لتقدير الوسائط لنموذج في سياق البايز. في البداية قدمنا بعض التعاريف في منهجية نظرية بايز و طريقة مونتي كارلو.

ثانيا تطرقنا الى طريقة اختيار العينة الوصفية المحسنة و قمنا بتنفيذها على نموذج بواسون.

## كلمات المفاتيح

منهجية النظرية الافتراضية, طريقة العينة الوصفية, طريقة العينة الوصفية المحسنة, طريقة مونتي كارلو