

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE  
MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE  
SCIENTIFIQUE  
UNIVERSITE MOHAMED SEDDIK  
BEN YAHIA - JIJEL



FACULTE DES SCIENCES EXACTES ET INFORMATIQUE  
DEPARTEMENT DE MATHEMATIQUES

Série : .....

**Mémoire présenté pour obtenir le diplôme de  
Master en Mathématique**

**Spécialité : probabilités et statistique**

**Par**

**BOUZERAA Sarra**

**Intitulé**

**Estimation bayésienne sous des données progressivement  
censurées**

Soutenue le : .... /10 /2020

Devant les jurys :

Président :	CHERAITIA Hassen	M.C.B	Univ. de Jijel
Rapporteur :	BOUDJERDA Khawla	M.C.B	Univ. de Jijel
Examineurs:	ABDI Zineb	M.A.A	Univ. de Jijel

Année université 2019/2020

## ✱ *Remerciements* ✱

Avant tout nous remercions ALLAH tout puissant qui nous a donné la force et la volonté pour pouvoir finir ce mémoire.

Nous tenons à exprimer nos profond respect à ma encadreur de mémoire, Melle ”**Boudjerda Khawla**”, pour ces conseils précieux et sa critique constructive qui ont été très positifs.

Nous remercions sincèrement les membres du jury :

- CHERAITIA Hassen : Président
- ABDI Zineb : Examineur.

Il est important de remercier mes famille qui ont toujours été une source d’encouragement.

Merci à tout les enseignants d’université de Mohammed Seddik Ben Yahia -JIJEL

Sans oublier toutes les personnes qui ont contribue de près ou de loin.

✱ *Dédicace* ✱

J'ai le grand plaisir de dédier ce modeste mémoire :

♡ A mon très chère père Maseoud, " le premier et le dernier homme de ma vie, source d'amour, d'affection, de générosité et de sacrifices, Tu étais toujours là près de moi pour me soutenir, et me guider avec les précieux conseils. Aucune dédicace ne saurait exprimer l'amour que je porte au grand homme que vous êtes. Que dieu me le garde".

♡ A ma très chère mère zohra,  
"source de ma vie, d'amour et de tendresse qui n'a pas cessé de m'encourager et de prier pour moi, puisse Dieu le tout puissant t'accorder meilleure santé, longue vie et bonheur".

♡ A mes chère frères "Samir et Walid" et mes belles Soeurs "Nessrin et Lamiss".

Puis Dieu vous donne santé, bonheur, courage et surtout réussite.

♡ A tout mes ami(e)s et mes collègues

♡ A tout ceux qui m'ont aidé de près ou de loin.

✱ *Sarra* ✱

# Résumé

Dans ce travail, on s'est intéressé essentiellement à un modèle de survie : le modèle de Lindley.

L'étude des estimateurs de Bayes de paramètre, de la fonction de fiabilité et de la fonction de taux de panne sous différentes fonctions de perte avec des données progressivement censurées. La loi a priori utilisée dans ce travail est la loi a priori conjuguée naturelle.

Deux approches ont été utilisées, l'approche classique du maximum de vraisemblance puis une approche bayésienne et les risques a posteriori sont obtenus en utilisant des fonctions de perte symétriques (la fonction de perte quadratique) puis des fonctions de perte asymétriques (la fonction de perte Linex et entropie).

Mot clé : Modèle de Lindley, Estimateurs de Bayes, les lois a priori, les méthodes de Monte-Carlo

# Table des matières

<b>1 Outils mathématiques</b>	<b>3</b>
1.1 Définitions . . . . .	3
1.1.1 Statistique . . . . .	3
1.1.2 Variable aléatoire . . . . .	3
1.1.3 Loi de probabilité . . . . .	4
1.2 Distribution de la durée de fiabilité . . . . .	4
1.2.1 Analyse de fiabilité . . . . .	4
1.2.2 La fonction de fiabilité . . . . .	4
1.2.3 La fonction de répartition . . . . .	5
1.2.4 La fonction de densité de probabilité . . . . .	5
1.2.5 La fonction du taux de hasard . . . . .	5
1.2.6 La fonction du taux de hasard cumulé . . . . .	6
1.3 Quelques lois usuelles . . . . .	6
1.3.1 Loi uniforme . . . . .	6
1.3.2 Loi exponentielle . . . . .	7
1.3.3 La loi Weibull . . . . .	8
1.3.4 Loi Gamma . . . . .	9
1.4 Estimation . . . . .	10
1.5 Qualité d'un estimateur . . . . .	10

1.5.1	Le biais d'un estimateur . . . . .	11
1.5.2	Convergence d'un estimateur . . . . .	11
1.6	Efficacité et Risque d'un estimateur . . . . .	11
1.7	Information de Fisher . . . . .	13
1.7.1	Information de Fisher d'un échantillon( $\theta$ réel) . . . . .	13
1.7.2	Information de Fisher d'un échantillon (cas multidimensionnel) . . . . .	14
1.8	Estimateur efficace . . . . .	15
1.9	Estimation par intervalle de confiance . . . . .	15
1.10	Méthodes de construction d'un estimateur . . . . .	16
1.10.1	La méthode du maximum de vraisemblance . . . . .	16
1.10.2	La méthode des moments . . . . .	17
<b>2</b>	<b>La statistique bayésienne</b>	<b>19</b>
2.1	Introduction . . . . .	19
2.2	Le principe bayésien . . . . .	21
2.3	Risque fréquentiste . . . . .	22
2.4	Risque de Bayes . . . . .	22
2.5	Distribution a posteriori . . . . .	23
2.6	Choix de la distribution a priori . . . . .	25
2.6.1	La distribution a priori conjuguée . . . . .	26
2.6.2	Distribution a priori impropre . . . . .	27
2.6.3	Distribution a priori de Jeffrey . . . . .	28
2.7	Fonctions de perte . . . . .	30
2.7.1	Fonction de perte quadratique . . . . .	30
2.7.2	La Fonction de perte Linex . . . . .	30
2.7.3	Fonction de perte de DeGroot . . . . .	32

2.7.4	Fonction de perte d'entropie . . . . .	32
2.8	Les Méthodes numériques stochastique . . . . .	33
2.8.1	Méthodes MCMC . . . . .	33
2.8.2	Algorithme Hasting-Metropolis . . . . .	33
2.8.3	Algorithme de type Gibbs . . . . .	34
<b>3</b>	<b>La loi de Lindley à un paramètre</b>	<b>35</b>
3.1	Introduction . . . . .	35
3.2	Le modèle . . . . .	38
3.3	Estimation par la méthode du maximum de vraisemblance . . . . .	41
3.3.1	Estimation du paramètre $\theta$ . . . . .	41
3.3.2	Estimation de la fonction de fiabilité . . . . .	42
3.4	Estimation bayésienne sous différentes fonctions de perte . . . . .	42
3.4.1	Les distributions a posteriori et a priori . . . . .	42
3.4.2	Fonctions de perte . . . . .	43
	<b>Bibliography</b>	<b>50</b>

# Introduction général

L'approche bayésienne de la statistique connaît à l'heure actuelle un essor considérable notamment grâce aux progrès de l'informatique et des méthodes numérique de type MCMC. Lorsque l'on réalise une étude, on a souvent des informations a priori provenant soit l'études antérieures soit l'avis d'expert.

La statistique bayésienne permet d'utiliser ces connaissances a priori et les combine avec l'information apportées par les données pour obtenir une information a posteriori. La statistique bayésienne est également très utilisés dans les analyses, c'est-à-dire les analyses qui mettent ensemble plusieurs études réalisées dans des conditions par fois différentes pour en extraire de l'information avec une meilleur précision. Au cours de la formations nous efforcerons de comparer les avantages et les inconvénients de l'approche bayésienne par rapport à l'approche classique (ou fréquentiste).

Dans ce mémoire, on s'est intéressé au modèle de Lindley, les propriétés mathématiques telles que la fonction de fiabilité, les différentes moments ont été traités dans l'article (H.Krishna et K.Kumar (2011)).

On se propose d'étudier le problème de l'estimation de paramètre et de la fonction de fiabilité à l'aide de l'approche bayésienne, dans un plan d'expérience progressivement censurées. Pour le choix de la loi a priori, on choisi une loi conjuguée naturelle sur le paramètre. Les estimateurs de Bayes dépendant du choix de la fonction de perte qui souvent est une fonction de perte quadratique de forme symétrique d'expression des estimateurs



de notre modèle n'a pas une forme analytique simple, aussi bien par l'approche classique que par l'approche bayésienne (solution d'un système non linéaire pour l'une et forme intégrale pour l'autre).

Le manuscrit est organisé de la manière suivante :

Dans le premier chapitre nous nous limitons à un rappel sur des notions et définitions de base de la théorie de fiabilité que nous estimons utiles pour la suite de notre travail. Le chapitre deux consacré à l'analyse bayésienne.

Au chapitre trois, on étudié le modèle de Lindley, on donne les estimateurs de son paramètre, sa fonction de fiabilité et sa fonction de taux de panne à l'aide de d'une approche classique de maximum de vraisemblance, puis à l'aide d'une approche bayésienne, dans un plan de données progressivement censurés.

# Chapitre 1

## Outils mathématiques

Dans ce chapitre introductif, nous énonçons quelques définitions et propriétés dont nous servirons par la suite.

### 1.1 Définitions

#### 1.1.1 Statistique

La statistique est un ensemble des méthodes qui consiste à réunir les données sur un phénomène, puis à analyser, à commenter et à critiquer ces données. On peut définir la statistique descriptive comme l'instrument statistique qui permet de donner un sens, une expression à l'information recueillie. Elle donne une image concise et simplifiée de la réalité.

#### 1.1.2 Variable aléatoire

Dans une expérience dont les résultats ne dépendent que du hasard, on n'est pas nécessairement intéressé par le résultat de l'expérience, mais bien souvent par une certaine fonction de ce résultat.

**Définition 1.1** *Une variable aléatoire  $X$  sur un ensemble fondamental  $\Omega$  est une fonc-*

tion de  $\Omega$  dans  $\mathbb{R}$  telle que pour tout intervalle  $[a, b] \subset \mathbb{R}$ , l'ensemble  $\{X \in [a, b]\} = \{\omega \in \Omega : X(\omega) \in [a, b]\}$  est un événement.

### 1.1.3 Loi de probabilité

En théorie des probabilités et en statistique, une loi de probabilité est une mesure qui décrit le comportement aléatoire d'un phénomène au hasard. Autrement dit, la loi de probabilité décrit le comportement d'un variable aléatoire discrète ou continue dépendant d'une expérience aléatoire.

Il y a plusieurs moyens de caractériser la loi de probabilité d'un variable aléatoire, la plus simple et couramment utilisée est la fonction de répartition.

## 1.2 Distribution de la durée de fiabilité

### 1.2.1 Analyse de fiabilité

Soit  $X$  une variable aléatoire positive ou nulle et absolument continue qui représente le temps de fiabilité, et  $t$  la valeur précise de la variable aléatoire  $X$ , alors sa loi de probabilité peut être définie par l'une des cinq fonctions équivalentes suivantes (chacune des fonctions ci-dessous peut être obtenue à partir de l'une des autres fonctions).

### 1.2.2 La fonction de fiabilité

La fonction de fiabilité est la probabilité que la variable aléatoire  $X$  dépasse le temps spécifié  $t$ .

La fonction de fiabilité est définie par :

$$S(t) = P(X > t), \quad t \geq 0$$

### 1.2.3 La fonction de répartition

La fonction de répartition (où c.d.f pour "cumulative distribution function") représente, pour  $t$  fixé, la probabilité de mourir avant l'instant  $t$ , c'est-à-dire :

$$F(t) = P(X \leq t)$$

$$= 1 - P(X > t)$$

$$F(t) = 1 - S(t), \quad t \geq 0$$

### 1.2.4 La fonction de densité de probabilité

Si la fonction de répartition  $F$  admet une dérivée au point  $t$  alors,

$$f(t) = \lim_{h \rightarrow 0} \frac{P(t \leq X < t + h)}{h} = F'(t) = -S'(t)$$

### 1.2.5 La fonction du taux de hasard

Le taux de hasard est défini comme :

$$h(t) = \lim_{r \rightarrow 0} \frac{P(t \leq X < t + r | X \geq t)}{r} = \frac{f(t)}{S(t)}$$

pour  $t$  fixé, il caractérise la probabilité de mourir dans un petit intervalle de temps après l'instant  $t$ , conditionnellement au fait d'avoir survécu jusqu'à l'instant  $t$ , Aussi cela signifie-t-il le risque de mort instantané pour ceux qui ont survécu.

## 1.2.6 La fonction du taux de hasard cumulé

La fonction du taux de hasard cumulé est définie par :

$$H(t) = \int_0^t h(x)dx = -\ln(S(t))$$

## 1.3 Quelques lois usuelles

### 1.3.1 Loi uniforme

La loi uniforme continue est une généralisation de la fonction rectangle à cause de la forme de sa fonction de densité de probabilité. Elle est paramétrée par les plus petites et les plus grandes valeurs  $a$  et  $b$  que la variable aléatoire uniforme peut prendre. Cette loi continue est souvent notée  $U[a, b]$ .

La distribution uniforme est utile pour l'échantillonnage à partir des distribution arbitraires. La distribution uniforme est souvent utilisée dans la simulation des échantillons, de nombreux langages de programmation ont la capacité de générer des nombres pseudo-aléatoires qui sont distribués efficacement selon la distribution uniforme standard.

La fonction de densité est :

$$f(t) = \frac{1}{b-a} \mathbf{1}_{[0,+\infty[}(t), \quad 0 < t < \infty$$

La fonction de répartition est :

$$F(t) = \frac{t}{b-a}$$

La fonction de survie est :

$$S(t) = 1 - F(t) = 1 - \frac{t}{b-a}$$

La fonction de hasard est :

$$h(t) = \frac{f(t)}{S(t)} = ((b-a) - t)^{-1}$$

La fonction de hasard cumulé est :

$$H(t) = -\ln(S(t)) = -\ln\left(1 - \frac{t}{(b-a)}\right)$$

### 1.3.2 Loi exponentielle

Cette distribution modélise la durée de vie d'un phénomène sans mémoire, sans vieillissement ou sans usure : la probabilité que le phénomène dure au moins  $s+t$  heures sachant qu'il a déjà duré  $t$  heures sera la même que la probabilité de durer  $s$  heures à partir de sa mise en fonction initiale. En d'autres termes, le fait que le phénomène ait duré pendant  $t$  heures ne change rien à son espérance de vie à partir du temps  $t$ .

La fonction de densité est :

$$f(t) = \lambda e^{-\lambda t} \mathbf{1}_{[0,+\infty[}(t)$$

La fonction de répartition est :

$$F(t) = 1 - e^{-\lambda t}$$

La fonction de survie est :

$$S(t) = e^{-\lambda t}$$

La fonction de hasard est :

$$h(t) = \lambda$$

La fonction de hasard cumulé est :

$$H(t) = \lambda t$$

### 1.3.3 La loi Weibull

Nommée d'après Waloddi Weibull en 1951, est une loi de probabilité continue. La loi de Weibull est un cas spécial de loi d'extremum généralisée au même titre que la loi de Gumbel ou la loi de Fréchet. Les différentes type de fonctions de cette distribution sont les suivantes :

La fonction de densité est :

$$f(t) = \frac{\beta}{\lambda} \left( \frac{t}{\lambda} \right)^{\beta-1} e^{-(t/\lambda)^\beta} \mathbf{1}_{[0,+\infty[}(t);$$

La fonction de répartition est :

$$F(t) = 1 - e^{-(t/\lambda)^\beta}$$

La fonction de survie est :

$$S(t) = e^{-(t/\lambda)^\beta}$$

La fonction de hasard est :

$$h(t) = \frac{\beta}{\lambda} \left( \frac{t}{\lambda} \right)^{\beta-1}$$

La fonction de hasard cumulé est :

$$H(t) = \left(\frac{t}{\lambda}\right)^\beta$$

### 1.3.4 Loi Gamma

Une distribution Gamma est caractérisée par deux paramètres qui affectent respectivement la forme et l'échelle de sa représentation graphique, les distribution Gamma sont utilisées pour modéliser une grande variété de phénomène, et tout particulièrement les phénomènes se déroulant au cours du temps où par essence, le temps écoulé est une grandeur réelle positive ; c'est le cas par exemple dans l'analyse de fiabilité.

La fonction de densité est :

$$f(t) = \frac{t^{k-1}e^{-t/\theta}}{\Gamma(k)\theta^k} \mathbf{1}_{[0,+\infty[}(t)$$

La fonction de répartition est :

$$F(t) = \frac{\gamma(k, t/\theta)}{\Gamma(k)}$$

La fonction de fiabilité est :

$$S(t) = 1 - \frac{\gamma(k, t/\theta)}{\Gamma(k)}$$

La fonction de hasard cumulé est :

$$H(t) = -\ln\left(1 - \frac{\gamma(k, t/\theta)}{\Gamma(k)}\right)$$



## 1.4 Estimation

En mathématiques, un estimateur est une statistique permettant d'évaluer un paramètre inconnu relatif à une loi de probabilité. Un estimateur est toujours établi à partir d'une statistique d'échantillon, il doit refléter au mieux le vrai paramètre de la population qui restera inconnu. On souhaite calculer à partir d'un échantillon statistique une valeur pour ce paramètre aussi proche que possible de sa vraie valeur, et on souhaite aussi déterminer un intervalle autour de cette valeur, qui caractérise la précision de la valeur estimée.

La qualité des estimateurs s'exprime par leur convergence, leur biais et leur efficacité. Diverses méthodes permettent d'obtenir des estimateurs de qualités différentes.

**Définition 1.2** *Soit un échantillon de taille  $n$  fini, issu d'une population mère et soient  $x_1, \dots, x_n$  des réalisations indépendantes de la variable aléatoire  $X$  dont la densité  $f(x, \theta)$  dépend d'un paramètre inconnu  $\theta$ . On appelle estimateur de  $\theta$ , une statistique notée  $T_n$ , toute fonction qui dépend uniquement du  $n$ -échantillon. Sa valeur est notée par :*

$$\hat{\theta} = T_n(x_1, \dots, x_n)$$

## 1.5 Qualité d'un estimateur

Un estimateur  $T_n$  de  $\theta$  sera bon s'il est suffisamment proche de  $\theta$ . On définit donc une mesure de l'écart type entre  $\theta$  et  $T_n$ , qu'on appellera le risque de l'estimateur. Plus ce risque est petit mieux sera notre estimateur.

### 1.5.1 Le biais d'un estimateur

**Définition 1.3** On appelle le biais de l'estimateur  $T_n$  de  $\theta$ , l'écart type entre sa moyenne et la vraie valeur du paramètre  $\theta$ .

$$b_\theta(T_n) = E(T_n) - \theta$$

#### *cas particulier*

Un estimateur  $T_n$ , d'un paramètre  $\theta$  est dit sans biais si son espérance mathématique est la vraie valeur du paramètre :

$$E(T_n) = \theta$$

Un estimateur  $T_n$ , est dit asymptotiquement sans biais si :

$$\lim_{n \rightarrow \infty} E(T_n) = \theta$$

### 1.5.2 Convergence d'un estimateur

**Définition 1.4** Un estimateur  $T_n$ , d'un paramètre  $\theta$  est dit convergent si :

$$P(|T_n - \theta| < \varepsilon) = 0$$

et, on écrit :

$$T_n \xrightarrow{\text{prob}} \theta$$

**Théorème 1.1** Un estimateur  $T_n$ , de  $\theta$  dont l'espérance mathématique tend vers  $\theta$  et dont la variance tend vers 0 lorsque  $n \rightarrow \infty$ , est un estimateur convergent :

$$E(T_n) = \theta \text{ et } V(T_n) = 0 \text{ quand } n \rightarrow \infty$$

**Remarque 1.1** Si deux estimateurs sans biais et convergents du paramètre  $\theta$ , alors le plus précis donc le meilleur est celui qui possède la variance (l'écart quadratique) la plus faible.

## 1.6 Efficacité et Risque d'un estimateur

Pour étudier l'efficacité et calculer le risque d'un estimateur on utilise le suivant :

**Théorème 1.2** *Le risque quadratique d'un estimateur  $\hat{\theta}$  d'un paramètre  $\theta$  est défini par :*

$$EQM(\hat{\theta}) = R(\hat{\theta}, \theta) = E(\hat{\theta} - \theta)^2$$

*On l'appelle aussi l'erreur quadratique moyenne de  $\hat{\theta}$  "EQM( $\hat{\theta}$ )"*

– **Cas d'un estimateur sans biais :**

$$\begin{aligned} R(\hat{\theta}, \theta) &= E(\hat{\theta} - \theta)^2 = E(\hat{\theta}^2 - 2\hat{\theta}\theta + \theta^2) = E(\hat{\theta}^2) - 2E(\hat{\theta}\theta) + E(\theta^2) \\ &= E(\hat{\theta}^2) - 2\theta E(\hat{\theta}) + E(\theta)^2 = E(\hat{\theta})^2 - 2\theta^2 + \theta^2 \\ &= E(\hat{\theta})^2 - \theta^2 \\ &= \text{Var}(\hat{\theta}) \end{aligned}$$

– **Cas d'un estimateur quelconque :**

$$\begin{aligned} R(\hat{\theta}, \theta) &= E(\hat{\theta} - \theta)^2 = E(\hat{\theta} - E(\hat{\theta}) + E(\hat{\theta}) - \theta)^2 \\ &= E(\hat{\theta} - E(\hat{\theta}))^2 - 2E((\hat{\theta} - E(\hat{\theta}))(E(\hat{\theta}) - \theta)) + E((E(\hat{\theta}) - \theta))^2 \\ &= \text{Var}(\hat{\theta}) - E(B)^2 \end{aligned}$$

**Définition 1.5** *Un estimateur est dit efficace si son risque est fini et qu'il n'existe aucun autre estimateur ayant un risque plus faible.*

*Soient  $\hat{\theta}_1$  et  $\hat{\theta}_2$  deux estimateur d'un même paramètre  $\theta$ , on appelle efficacité relative de  $\hat{\theta}_2$  par rapport à  $\hat{\theta}_1$ , le rapport*

$$E f . r e l a t i v e = \frac{RQM(\hat{\theta}_2)}{RQM(\hat{\theta}_1)} = \frac{R(\hat{\theta}_2, \theta)}{R(\hat{\theta}_1, \theta)}$$

**Remarque 1.2** *Si ce rapport  $\leq 1$  alors  $\hat{\theta}_2$ , est meilleur (plus efficace) que  $\hat{\theta}_1$ ,*

$$RQM(\hat{\theta}_2) \leq RQM(\hat{\theta}_1)$$

*Si  $\hat{\theta}_1$  et  $\hat{\theta}_2$  sont des estimateurs sans biais de  $\theta$ , alors l'efficacité relative est donné par le rapport :*

$$\frac{\text{Var}(\hat{\theta}_2)}{\text{Var}(\hat{\theta}_1)}$$

*Si deux estimateurs sans biais et convergents du paramètre  $\theta$ , alors le plus précis donc le meilleur est celui qui possède la variance (l'écart quadratique) la plus faible.*

## 1.7 Information de Fisher

Considérons une variable aléatoire  $X$  dont la loi dépend de  $s$  paramètres notés comme un vecteur  $\theta \in T \subset \mathbb{R}^s$ . Nous considérons soit ses probabilités  $f(x, \theta) = P(X = x, \theta)$  (**cas discret**), soit sa densité  $f(x, \theta)$  (**cas continu**). Nous utiliserons dans les énoncés la fonction de densité, mais le cas discret est tout à fait analogue. Nous étudions d'abord le cas  $s = 1$ , puis le cas général. La notion suivante a été introduite par Ronald Aylmer Fisher.

### 1.7.1 Information de Fisher d'un échantillon( $\theta$ réel)

**Score de l'échant** : Si  $f(x, \theta)$  est différentiable en  $\theta$ ,  $l$  est une fonction dérivable de  $\theta$  qui représente le log de la vraisemblance :  $l(x, \theta) = \ln(L(x, \theta))$ , et le score est sa dérivée :

$$S_n(\theta) = \frac{\partial}{\partial \theta} l(X_1, \dots, X_n; \theta) = \frac{1}{L(x_1, \dots, X_n; \theta)} \frac{\partial}{\partial \theta} L(X_1, \dots, X_n; \theta)$$

- Pour tout  $\theta$ , le score est une variable aléatoire.
- Le score s'annule à un optimum en  $\theta$  de la fonction de vraisemblance.
- Pour une réalisation donnée  $(x_1, x_2, \dots, x_n)$  de l'échantillon aléatoire (un sondage particulier), la valeur du score est une fonction de  $\theta$ , réalisation de la variable aléatoire définie ci-dessus sur ce type de données :

$$S_n(\theta) = \frac{\partial}{\partial \theta} l(X_1, \dots, X_n; \theta)$$

- Le score est centré :  $E(S_n(\theta)) = 0$
- La variance du score (si elle existe) s'appelle l'information de Fisher apportée par l'échantillon sur  $\theta$ ,

$$I_n(\theta) = E((S_n(\theta))^2)$$

Cette quantité mesure l'information apportée par un échantillon sur le paramètre.

**propriétés de l'information de Fisher :**

- Autre formulation : si le domaine de définition de  $X$  ne dépend pas de  $\theta$  et que cette quantité existe,

$$I_n(\theta) = -E\left(\frac{\partial^2 l(X_1, \dots, X_n; \theta)}{\partial \theta^2}\right) = -E\left(\frac{\partial S_n(\theta)}{\partial \theta}\right)$$

- Additivité : si le domaine de définition de  $X$  ne dépend pas de  $\theta$ , chaque observation apporte la même information,

$$I_n(\theta) = nI_1(\theta)$$

## 1.7.2 Information de Fisher d'un échantillon (cas multidimensionnel)

### Extension au cas multidimensionnel

Le score est un vecteur aléatoire de dimension  $s$  :  $\theta = (\theta_1, \dots, \theta_s) \in \mathbb{R}^s$

$$S_n(\theta) = \left( \frac{\partial l(X_1, \dots, X_n; \theta)}{\partial \theta_1}, \dots, \frac{\partial l(X_1, \dots, X_n; \theta)}{\partial \theta_s} \right)$$

Il est caractérisé par son vecteur espérance (= 0) et sa matrice de variance covariance appelée matrice d'information de Fisher,

$$I_n(\theta) = (I_{i,j}^n)_{1 \leq i,j \leq s}$$

Définie positive de terme général

$$I_{i,j}^n = cov\left(\frac{\partial l(X_1, \dots, X_n; \theta)}{\partial \theta_i}, \dots, \frac{\partial l(X_1, \dots, X_n; \theta)}{\partial \theta_j}\right)$$

## 1.8 Estimateur efficace

Une estimateur efficace de  $\theta$  est un estimateur sans biais telle que :

$$V(\hat{\theta}) = \frac{1}{I_n(\theta)}$$

### proposition

- Si un estimateur efficace de  $\theta$  existe, il est unique p.s.
- Si un estimateur efficace de  $\theta$  existe, il est égal p.s. à l'estimateur sans biais de variance minimale de  $\theta$ .

## 1.9 Estimation par intervalle de confiance

Les estimateur ponctuelles n'apportent pas d'information sur la précision des résultat, c'est-à-dire qu'elles ne tiennent pas compte des erreurs dues aux fluctuations d'échantillonnage.

Pour évaluer la confiance que l'on peut avoir en une valeur, il est nécessaire de déterminer un intervalle contenant avec une certaine probabilité fixée au préalable, la vraie valeur du paramètre : c'est l'estimation par intervalle de confiance. Le principe de cette méthode d'estimation est de proposer un encadrement d'un paramètre inconnu d'une population dont la loi est connue.

Soit  $\alpha \in ]0, 1[$ . S'il existe des variables aléatoires réelles  $\theta_{min}(X_1, \dots, X_n)$  et  $\theta_{max}(X_1, \dots, X_n)$  telles que :

$$P(\theta \in [\theta_{min}(X_1, \dots, X_n), \theta_{max}(X_1, \dots, X_n)]) = 1 - \alpha$$

Où :  $\theta_{min}(X_1, \dots, X_n)$  et  $\theta_{max}(X_1, \dots, X_n)$  sont les limites de confiance.

On dit alors que  $[\theta_{min}(X_1, \dots, X_n), \theta_{max}(X_1, \dots, X_n)]$  est un intervalle de confiance pour  $\theta$  avec une probabilité  $1 - \alpha$  dite risque d'erreur. On le note  $IC_{1-\alpha}(\theta)$ .

## 1.10 Méthodes de construction d'un estimateur

Dans certain situations il n'est y a pas d'estimateur évident donc on est amené à recourir à une méthode de construction d'un estimateur.

### 1.10.1 La méthode du maximum de vraisemblance

L'estimation du maximum de vraisemblance est une méthode statistique courante utilisée pour inférer les paramètres de la distribution de probabilité d'un échantillon donné. Cette méthode consiste à maximiser une fonction appelée fonction de vraisemblance, contenant le paramètre que l'on souhaite estimer.

Soit  $X$  une variable aléatoire réelle de loi de probabilité, dont on veut estimer le paramètre  $\theta$ . Alors on définit une fonction  $f$  telle que ;

Si  $X$  est une variable aléatoire continue de densité de probabilité  $f$ , alors

$$f(x, \theta) = f_{\theta}(x)$$

Si  $X$  est une variable aléatoire discrète de probabilité ponctuelle  $P$ , alors

$$f(x, \theta) = P_{\theta}(X = x)$$

On appelle fonction de vraisemblance de  $\theta$  pour une réalisation  $(x_1, \dots, x_n)$  de n-échantillon  $X_1, \dots, X_n$  i.i.d, d'un variable aléatoire  $X$ , si la loi de probabilité de  $X$  dépend d'un paramètre  $\theta$ , on définit alors la fonction de vraisemblance de  $\theta$  comme suit :

$$L(x_1, \dots, x_n; \theta) = f(x_1, \dots, x_n; \theta) = \prod_{i=1}^n f(x_i; \theta)$$

Lorsque  $\theta$  n'est pas observable, la méthode du maximum de vraisemblance consiste à estimer  $\theta$  par la valeur qui maximise  $L$  (vraisemblance), et on obtient ;

$$\hat{\theta} = \left\{ \theta / L(\hat{\theta}) = \max_{\theta} L(x_1, \dots, x_n; \theta) \right\}$$

La vraisemblance étant positive et le logarithme népérien est une fonction strictement positive; il est équivalent et souvent plus simple de maximiser le logarithme népérien de la vraisemblance.

Si  $L(x_1, \dots, x_n; \theta)$  est la vraisemblance du paramètre  $\theta$ , alors l'estimateur du maximum de vraisemblance vérifie :

$$\begin{cases} \frac{\partial}{\partial \theta} \log L(x_1, \dots, x_n; \theta) = 0 \\ \frac{\partial^2}{\partial \theta^2} \log L(x_1, \dots, x_n; \theta) < 0 \end{cases}$$

### 1.10.2 La méthode des moments

L'estimation par la méthode des moments c'est la méthode la plus naturelle. L'idée de base est d'estimer une espérance mathématique par une moyenne empirique, une variance par une variance empirique, etc...

On suppose que l'échantillon  $X_1, \dots, X_n$  i.i.d, d'une variable aléatoire  $X$  et la loi de probabilité de  $X$  dépend d'un paramètre  $\theta$

Si le paramètre à estimer est l'espérance de la loi des  $X_i (i = 1, \dots, n)$ , alors on peut l'estimer par la moyenne empirique de l'échantillon. Autrement dit, si  $\theta = E[X]$ , alors l'estimateur de  $\theta$  par la méthode des moment est :

$$\hat{\theta}_n = \bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$$

Plus généralement, pour  $\theta \in \mathbb{R}$ , si  $E(X) = \varphi(\theta)$ , où  $\varphi$  est une fonction inversible, alors l'estimateur de  $\theta$  par la méthode des moments est :



$$\hat{\theta}_n = \varphi^{-1}(\bar{X}_n)$$

De la même manière, on estime la variance de la loi des  $X_i$  ; par la variance empirique de l'échantillon :

$$S_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$$

Ce principe peut naturellement se généraliser aux moments de tous ordres, centrés ou non centrés :

$$E[(X - E(X))^k] \text{ et } E(X^k), k \geq 1$$

# Chapitre 2

## La statistique bayésienne

### 2.1 Introduction

Dans de nombreuses situations d'expériences aléatoires, il semble raisonnable d'imaginer que le praticien a une certaine idée du phénomène aléatoire qu'il est en train d'observer. Or, la démarche statistique classique repose essentiellement sur un principe de vraisemblance qui consiste à considérer que ce qui a été observé rend compte de manière exhaustive du phénomène. Mais l'observation ne fournit qu'une image et celle-ci peut être mauvaise. Certes cet inconvénient est en générale gommé par les considérations asymptotiques et un certain nombre de théorèmes permettent d'évaluer la bonne qualité des estimateurs si le nombre d'observations est suffisant.

L'analyse bayésienne des problèmes statistiques propose d'introduire dans la démarche d'inférence, l'information dont on dispose a priori le praticien. Dans le cadre de la statistique paramétrique, ceci se traduira par le choix d'une loi sur le paramètre d'intérêt. Dans l'approche classique, le modèle paramétrique  $(\mathcal{X}, \mathcal{A}, P_\theta, \theta \in \Theta)$ . Ayant un a priori sur le paramètre, modélisé par une densité de probabilité que nous noterons  $\pi(\theta)$ , on "réactualise" cet a priori au vu de l'observation en calculant la densité a posteriori  $\pi(\theta|x)$ , et c'est à

partir de cette loi que l'on mène l'inférence.

On peut alors, par exemple, de manière intuitive pour le moment retenir l'espérance mathématique ou encore le mode de cette densité a posteriori comme l'estimateur de  $\theta$ .

Le paramètre  $\theta$  devient donc en quelque sorte une variable aléatoire, à laquelle on associe une loi de probabilité dite loi a priori.

On sent bien d'emblée que les estimateurs bayésiens sont très dépendants du choix de la loi a priori.

Différentes méthodes existent pour déterminer ces lois a priori. On peut se référer à des techniques bayésiennes empiriques, où l'on construit la loi a priori sur la base d'une expérience passée, usant de méthodes fréquentistes, pour obtenir formes et valeurs des paramètres pour cette loi. Nous verrons que l'on peut aussi modéliser l'absence d'information sur le paramètre au moyen des lois dites non informative.

L'approche bayésienne se différencie donc de l'approche classique dans le sens où le paramètre  $\theta$  n'est plus considéré comme étant totalement inconnu, il est devenu une variable aléatoire dont le comportement est supposé connu. On fait intervenir dans l'analyse statistique une distribution associée à ce paramètre. On appelle modèle statistique bayésien ; la donnée d'un modèle statistique paramétrique  $(\mathcal{X}, \mathcal{A}, P_\theta, \theta \in \Theta)$  avec  $f(x; \theta)$  densité de  $P_\theta$  et d'une loi  $\pi(\theta)$  sur le paramètre.

La démarche de l'analyse bayésienne conduit au calcul d'une loi a posteriori  $\pi(\theta|x)$  ; actualisation de la loi a priori  $\pi(\theta)$  au vu de l'observation.

Ce calcul repose sur la version continue du théorème de Bayes

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

$f(x; \theta)$  désignant la loi de l'observation ou vraisemblance et  $f(x)$  la loi marginale ou prédictive

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

## 2.2 Le principe bayésien

$\Theta$  : l'espace des paramètres dans le cas d'un problème statistique.

$\mathcal{A}$  : l'espace des actions ou décisions, dont les évènements sont des images de l'observation par une application  $\delta$  appelée règle de décision.

$\mathcal{D}$  : l'ensemble des règles de décisions  $\delta$ , applications de  $\chi$  dans  $\mathcal{A}$  (les estimateurs possibles).

On note  $a$  une action. On a :  $a = \delta(x)$ .

L'inférence consiste à choisir une règle de décision  $\delta \in \mathcal{D}$  concernant  $\theta \in \Theta$  sur la base d'une observation  $x \in \chi$ ,  $x$  et  $\theta$  étant liés par la loi  $f(x; \theta)$ .

En statistique, la règle de décision est un estimateur, l'action est une estimation (valeur de l'estimateur au point d'observation).

Pour choisir une décision, construit une relation de préférence en considérant une mesure du coût ou perte en courue lorsqu'on prend la décision  $\delta(x)$  et que l'état de la nature est  $\theta$ .

Pour faire on introduit la fonction  $L$ , appelée fonction de coût (ou de perte) définie de la manière suivante : On appelle fonction de coût, tout fonction  $L$  de  $\Theta \times \mathcal{D}$  dans  $\mathbb{R}$ .  $L(\theta, a)$  évalue le coût d'une décision  $a$  quand le paramètre vaut  $\theta$ .

Elle permet donc, en quelque sorte, de quantifier la perte encourue par une mauvaise décision, une mauvaise évaluation de  $\theta$ . Il s'agit d'un fonction de  $\theta$ . Un coût négatif correspond à un gain.

## 2.3 Risque fréquentiste

On dira qu'une décision est une bonne décision si elle conduit à un coût nul.

Autrement dit, une bonne décision est solution de l'équation

$$L(\theta, \delta(x)) = 0$$

$\theta$  étant inconnu, on ne peut évidemment pas résoudre cette équation. Classer les décisions par la seule considération du coût est donc impossible. Celui-ci ne prend pas compte l'information apportée par le modèle  $f(x; \theta)$ . Ces remarques conduisent à considérer la moyenne de la perte, c'est le risque fréquentiste.

**définition** : On appelle risque fréquentiste le coût moyen (l'espérance mathématique) du coût d'une règle de décision :

$$R(\theta, \delta) = E_{\theta}(L(\theta, \delta)) = \int L(\theta, \delta) dP_{\theta}(x)$$

On dira que  $\delta_1$  est préférable à  $\delta_2$  et on note  $\delta_1 < \delta_2$  si :

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

cette définition permet d'établir un préordre sur l'ensemble  $\mathcal{D}$  des décisions. Cependant, ce préordre est partiel puisqu'il ne permet pas de comparer deux règles de décision telles que :

$$R(\theta_1, \delta_1) < R(\theta_1, \delta_2) \text{ et } R(\theta_2, \delta_1) > R(\theta_2, \delta_2)$$

## 2.4 Risque de Bayes

Puisque l'approche bayésienne met à la disposition du statisticien une loi a priori  $\pi(\theta)$ , on peut considérer la moyenne du risque fréquentiste i.e la moyenne du coût moyen sui-

vant la loi a priori :  $E^\pi(R(\theta, \delta(x)))$

Il s'agit du risque bayésien ou risque de Bayes que l'on note  $r(\pi, \delta)$ . On a :

$$\begin{aligned} r(\pi, \delta) &= E^\pi(R(\theta, \delta)) \\ &= \int_{\Theta} R(\theta, \delta) \pi(\theta) d\theta \\ &= \int_{\Theta} \int_{\mathcal{X}} L(\theta, \delta(x)) f(x; \theta) dx \pi(\theta) d\theta \\ &= \int_{\Theta} \int_{\mathcal{X}} L(\theta, \delta(x)) \pi(\theta, x) f(x) dx d\theta \end{aligned}$$

On définit alors le coût a posteriori  $\rho(\pi, \delta(x))$  comme étant la moyenne du coût par rapport à la loi a posteriori

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

Il s'agit d'une fonction de  $x$ .

On a le résultat suivant : Le risque de Bayes  $r(\pi, \delta)$  est la moyenne de coût a posteriori  $\rho(\pi, \delta(x))$  suivant la loi marginale  $f(x)$ .

**Preuve**

$$r(\pi, \delta) = \int_{\Theta} \int_{\mathcal{X}} L(\theta, \delta(x)) f(x; \theta) \pi(\theta) dx d\theta$$

Où :  $f(x; \theta) \pi(\theta) = \pi(\theta|x) f(x)$ , on a donc ;

$$r(\pi, \delta) = \int_{\Theta} \int_{\mathcal{X}} L(\theta, \delta(x)) \pi(\theta|x) d\theta f(x) dx = \int_{\mathcal{X}} \rho(\pi, \delta(x)) f(x) dx$$

## 2.5 Distribution a posteriori

La distribution a posteriori est la quantité la plus importante dans l'inférence bayésienne.

Il contient toutes les informations disponibles sur le paramètre  $\theta$  inconnu après avoir ob-

servé les données  $X = x$ .

### distribution a posteriori

Soit  $X = x$  désigne la réalisation observée d'une (éventuellement multi-variée) variable aléatoire  $X$  avec la fonction de densité  $\pi(x|\theta)$ .

Soit  $\pi(\theta)$  est la densité a priori qui nous permet de calculer la fonction de densité  $\pi(\theta|x)$  de la distribution a posteriori en utilisant le théorème de Bayes :

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

Le terme  $f(x; \theta)$  est la fonction de vraisemblance. Puisque  $\theta$  est aléatoire, nous conditionnons explicitement sur une valeur spécifique  $\theta$  et on écrit  $L(\theta) = f(x; \theta)$ .

Le dénominateur peut être écrit comme :

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

qui souligne que cela ne dépend pas de  $\theta$ . La quantité  $f(x)$  est connue comme la probabilité marginale et est important pour le choix du modèle bayésien.

La densité de la distribution a posteriori est donc proportionnelle au produit de la vraisemblance est la densité a posteriori de la distribution avec une constante de proportionnalité  $\frac{1}{f(x)}$ . Ceci est habituellement noté

$$\pi(\theta|x) \propto f(x, \theta)\pi(\theta).$$

Où  $\propto$  représente "est proportionnel à" et implique que  $\frac{1}{\int L(\theta)\pi(\theta)d\theta}$  est la constante de normalisation pour assurer que :  $\int \pi(\theta|x)d\theta = 1$ , telle que  $\pi(\theta|x)$  est une fonction de densité valide.

L'inférence statistique à propos de  $\theta$  est basée uniquement sur la distribution a posteriori. L'estimation ponctuelle appropriés sont les paramètres de la localisation, telles que la moyenne, médiane ou le mode de la distribution a posteriori.

La moyenne a posteriori  $E(\theta|x)$  est l'espérance de la distribution a posteriori

$$E(\theta|x) = \int \pi(\theta|x)dx$$

Le mode a posteriori  $Mod(\theta|x)$  est le mode de la distribution a posteriori

$$Mod(\theta|x) = argmax_{\theta}\pi(\theta|x)$$

La médiane a posteriori  $Med(\theta|x)$  est la médiane de la distribution a posteriori est le nombre  $\alpha$  qui vérifié.

$$\int_{-\infty}^{\alpha} \pi(\theta|x)d\theta = 0,5 \text{ et } \int_{\alpha}^{+\infty} \pi(\theta|x)d\theta = 0,5$$

Implicitement, on suppose souvent que la moyenne a posteriori est finie, auquel cas il est également unique. Cependant, le mode a posteriori et la médiane a posteriori ne sont pas nécessairement unique. En effet, une distribution a posteriori peut avoir plusieurs modes et donc est appelée multimodale. L'estimation bayésienne par intervalle est également dérivée de la distribution a posteriori. Pour les distinguer des intervalles de confiance, qui ont une interprétation différent, ils sont appelés intervalles de crédibilité. Ici la définition est pour un paramètre scalaire  $\theta$ .

**Définition 2.1 Intervalle de crédibilité**

Pour un  $\gamma \in [0, 1]$  fixé, un  $\gamma - 100$  intervalle de crédibilité est définie par deux nombres réels  $t_l$  et  $t_u$  qui remplissent :

$$\int_{t_l}^{t_u} f(\theta; x)d\theta = \gamma$$

La quantité  $\gamma$  est appelée le niveau crédible de l'intervalle de crédibilité  $[t_l, t_u]$ .

## 2.6 Choix de la distribution a priori

L'inférence bayésienne permet la spécification probabiliste des croyances antérieures par le biais d'une distribution préalable.



Il est souvent utile et justifié de restreindre l'éventail des possibles distributions a priori a une famille spécifique avec un ou deux paramètres. Le choix de cette famille peut être basée sur le type de fonction de vraisemblance rencontré.

### 2.6.1 La distribution a priori conjuguée

Une approche pragmatique de choisir une distribution a priori est de sélectionner un membre d'une famille spécifique de distributions telles que la distribution a posteriori appartient à la même famille. Elle est appelée distribution a posteriori conjuguée.

#### La distribution a priori conjuguée

Soit  $L(x, \theta) = f(x|\theta)$  est la fonction de maximum de vraisemblance basée sur l'observation  $X = x$ . La classe  $G$  des distributions est appelée conjuguée par rapport à  $L(\theta)$ , si la distribution a posteriori  $\pi(\theta|x)$  est dans  $G$  pour tout  $x$  chaque fois que la distribution  $\pi(\theta)$  est dans  $G$ .

La famille  $G = \{toutes\ les\ distributions\}$  est conjuguée trivialement par rapport à toutes fonction de vraisemblance. Dans la pratique, on essaie de trouver des petits ensembles  $G$  qui sont spécifiques à la probabilité  $L(x, \theta)$ .

Avant d'étudier les distributions a priori conjuguées, on note qu'il suffit d'étudier conjugaison pour un membre  $X_i$  d'un échantillon aléatoire  $X_{1:n}$ . En effet, si l'a priori est conjuguée, la partie a posteriori après avoir observé la première observation, est par définition, du même type de sert de nouvelle distribution a posteriori, incorporant désormais la deuxième observation, n'est à niveau dans une classe conjuguée,...etc.

Seulement les paramètres de la distribution changeront dans un tel traitement séquentiel

des données.

Le tableau suivant donne quelques exemples des distributions a priori avec la fonction de vraisemblance correspondante :

vraisemblance	distr a priori	distr a posteriori
$X \theta \sim Bin(n, \theta)$	$\theta \sim Be(\alpha, \beta)$	$\theta x \sim Be(\alpha + x, \beta + n - x)$
$X \theta \sim Geom(\theta)$	$\theta \sim Be(\alpha, \beta)$	$\theta x \sim Be(\alpha + 1, \beta + x - 1)$
$X \pi \sim Po(e\lambda)$	$\lambda \propto G(\alpha, \beta)$	$\lambda x \sim G(\alpha + x, \beta + e)$
$X \pi \sim exp(\lambda)$	$\lambda \sim G(\alpha, \beta)$	$\lambda x \sim G(\alpha + 1, \beta + x)$
$X \mu \sim N(\mu, \sigma^2 \text{connu})$	$\mu \sim N(v, \tau^2)$	$\mu X \sim N((\frac{1}{\sigma^2} + \frac{1}{\tau^2})^{-1}(\frac{x}{\sigma^2} + \frac{v}{\tau^2}), (\frac{1}{\sigma^2} + \frac{1}{\tau^2})^{-1})$
$X \sigma^2 \sim N(\mu \text{connu}, \sigma^2)$	$\sigma^2 \sim IG(\alpha, \beta)$	$\sigma^2 \sim IG(\alpha + \frac{1}{2}, \beta + \frac{1}{2}(x - \mu)^2)$

Tableau.2.1- Représente quelques distributions a priori pour différentes fonctions de vraisemblance

### 2.6.2 Distribution a priori impropre

La distribution a priori a un influence sur la distribution a posteriori. Si on veut minimiser l'influence de la distribution a priori, il est courant de spécifier une a priori "vague", par exemple avec une très grande variance.

Dans la limite cela peut conduire à une distribution préalable mauvaise avec une fonction de densité qui n'intègre pas. En raison de la constante de normalisation manquante, ces fonctions de densité sont généralement spécifiées en utilisant le signe de proportionnalité "∝". Si on utilise l'a priori impropre, alors il est nécessaire de vérifier que au moins la distribution a priori est propre. Si ça le cas, alors l'a priori impropre peuvent être utilisées dans l'analyse bayésienne.

**Distribution a priori impropre :** Une distribution a priori avec une fonction de densité

$\pi(\theta) \geq 0$  est appelée impropre si,

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

Où,  $\sum_{\theta \in \Theta} \pi(\theta) = \infty$ , pour un paramètre  $\theta$  continu ou discret, respectivement.

### 2.6.3 Distribution a priori de Jeffrey

Dans certaines situations, il peut être utile de choisir une distribution a priori qui ne donne pas beaucoup d'informations sur le paramètre en raison de la connaissance préalable faible ou manquante. Un premier choix naïf est un a priori uniforme  $\pi_{\theta}(\theta) \propto 1$ , dans ce cas, l'a posteriori est proportionnelle à la fonction de vraisemblance. Noter qu'un a priori uniforme localement sera incorrecte si l'espace de paramètre est pas borné. Cependant, il existe des problèmes liés à cette approche. On suppose que  $\phi = h(\theta)$  est une transformation différentiable de  $\theta$ , qui a une loi a priori localement uniforme avec une densité  $\pi_{\theta}(\theta) \propto 1$ . On utilise un changement de variable, on obtient l'a priori correspondant à  $\phi$ .

$$\begin{aligned} \pi_{\phi}(\phi) &= \pi_{\theta} \{ \theta h^{-1}(\phi) \} \left| \frac{dh^{-1}(\phi)}{d\phi} \right| \\ &\propto \left| \frac{dh^{-1}(\phi)}{d\phi} \right| \end{aligned}$$

On note que ce terme est nécessairement constant. En effet,  $\pi_{\phi}(\phi)$  sera indépendante de  $\phi$  seulement si  $h$  est linéaire. Si  $h$  est non linéaire, la densité a posteriori  $\pi_{\phi}(\phi)$  dépend de  $\phi$  et ne sera pas (localement) uniforme. Cependant, si nous avons choisi une para-métrisation avec  $\phi$  dès le départ, nous avons choisi une a priori uniforme (localement)  $\pi_{\phi}(\phi) \propto 1$ .

#### La distribution a priori de Jeffrey

Soit  $X$  une variable aléatoire avec une fonction de maximum de vraisemblance  $L(x; \theta)$  où  $\theta$  est le paramètre inconnu.

L'a priori de Jeffrey est définie comme suit

$$\pi(\theta) \propto \sqrt{J(\theta)}$$

Où  $J(\theta)$  est l'**information de Fisher**. Cette dernière équation est nommée : la règle de Jeffrey.

L'a priori de Jeffrey est proportionnelle à la racine carrée de l'information de Fisher, qui donne une distribution a priori impropre.

**Résultat** (Invariance de la loi a priori de Jeffrey)

L'a priori de Jeffrey est invariant sous la para-métrisation de  $\theta$  si :

$$\pi_\theta(\theta) \propto \sqrt{J_\theta(\theta)}$$

Donc, la fonction de densité de  $\phi = h(\theta)$  est :

$$\pi_\phi(\phi) \propto \sqrt{J_\phi(\phi)}$$

Où  $J_\phi(\phi)$  est l'information de Fisher de  $\phi$ .

**Preuve**

On a  $\pi_\theta \propto \sqrt{J_\theta(\theta)}$  et on fait un changement de variable, on a

$$\begin{aligned} \pi_\phi(\phi) &= \pi_\theta(\theta) \left| \frac{dh^{-1}(\phi)}{d\phi} \right| \\ &\propto \sqrt{J_\theta(\theta)} \left| \frac{dh^{-1}(\phi)}{d\phi} \right| = \sqrt{J_\theta(\theta) \left\{ \frac{dh^{-1}(\phi)}{d\phi} \right\}^2} \\ &= \sqrt{J_\theta(\theta)} \end{aligned}$$

## 2.7 Fonctions de perte

### 2.7.1 Fonction de perte quadratique

La fonction de perte quadratique est la fonction définie par

$$L(\theta, d) = (\theta - d)^2$$

Une variante de cette fonction de perte est une fonction de perte quadratique pondérée (fonction de perte quadratique généralisée) de la forme

$$L(\theta, d) = \omega(\theta)(\theta - d)^2$$

Sous l'hypothèse d'un coût quadratique, l'estimateur de Bayes  $\delta^\pi(x)$  de  $\theta$  associé à la loi a priori  $\pi$  est la moyenne a posteriori de  $\theta$

$$\delta^\pi(x) = E_{\pi(\cdot|x)}(\theta) = \int_{\theta \in \mathbb{O}} L(\theta, \delta(x)) \pi(\theta|x) d\theta$$

#### Preuve

Par définition, l'estimateur de Bayes minimise le coût a posteriori i.e  $\rho(\pi, \delta) = E^{\pi(\cdot|x)}(L(\theta, \delta(x)))$ .

Sous l'hypothèse d'un coût quadratique, on a :

$$\begin{aligned} \rho(\pi, \delta) &= E^{\pi(\cdot|x)}((\theta - \delta(x))^2) \\ &= E^{\pi(\cdot|x)}(\theta^2) - 2\delta(x)E^{\pi(\cdot|x)}(\theta) + \delta^2(x). \end{aligned}$$

Il s'agit d'un polynôme du second degré en  $\delta(x)$ . Il sera minimum en  $E^{\pi(\cdot|x)}(\theta)$ .

### 2.7.2 La Fonction de perte Linex

Une fonction de perte asymétrique très pratique est la fonction de perte Linex (Linear Exponential). Elle a été introduite par Varian (1975). Cette fonction de perte presque

exponentiellement d'un coté de zéro sous l'hypothèse que la perte minimale est obtenu pour  $\hat{\delta}(x) = \theta$ , la fonction de perte Linex pour  $\theta$  :

$$L(\Delta) \propto e^{a\Delta} - a\Delta - 1, \quad a \neq 0$$

Où :  $\Delta = (\delta(x) - \theta)$  où  $\delta(x)$  est un estimateur du  $\theta$ . Le signe de  $a$  représentant respectivement la direction et le degré de symétrie ( $a > 0$  : la surestimation est plus grave que la sous-estimation et vice versa). Pour  $a$  proche de zéro, la perte Linex est approximativement la fonction de perte quadratique :

$$E_{\theta}(L(\delta(x) - \theta)) \propto e^{a\delta(x)} E_{\theta}(e^{-a\theta}) - a(\delta(x) - E_{\theta}(\theta) - 1) \dots (*)$$

Où  $E_{\theta}(\cdot)$  représente l'espérance a posteriori relative à la densité a posteriori de  $\theta$ .

L'estimateur de Bayes  $\delta_{\pi}(x)$  qui minimise (\*). Pour trouver l'estimateur, nous dérivons l'équation (\*) par rapport à  $\delta(x)$ , nous obtenons

$$\frac{d}{d\delta(x)} (E_{\theta}(L(\delta(x) - \theta))) = ae^{a\delta(x)} E_{\theta}(e^{-a\theta}) - a$$

En égalant cette expression à zéro, nous obtenons

$$e^{-a\delta(x)} E_{\theta}(e^{-a\theta}) = a$$

Alors, l'estimateur de Bayes  $\hat{\delta}_L(x)$  sous la fonction de perte Linex est :

$$\delta(x) = -\frac{1}{a} \log(E_{\theta}(e^{-a\theta}))$$

étant donné que  $E_{\theta}(e^{-a\theta})$  existe et est finie.

### 2.7.3 Fonction de perte de DeGroot

DeGroot (1970) a introduit plusieurs types des fonctions de perte et est obtient les estimateurs de Bayes sous cette fonction de perte. Un exemple d'une fonction de perte symétrique est la fonction de perte de DeGroot définie par :

$$L(\theta, \delta(x)) = \left(\frac{\theta - \delta(x)}{\delta(x)}\right)^2$$

sous cette fonction de perte, l'estimateur de Bayes est

$$\delta_\pi(x) = \frac{E^\pi(\theta^2|x)}{E^\pi(\theta|x)}$$

### 2.7.4 Fonction de perte d'entropie

Galabria et Pulcini (1994) ont proposé une fonction de perte qui découle de la fonction de perte Linex appelée la fonction de perte entropie est définie par :

$$L_E(\theta, d) \propto \left(\frac{d}{\theta}\right)^p - p \ln\left(\frac{d}{\theta}\right) - 1$$

qui a minimum lorsque  $d = \theta$ .

L'estimateur de Bayes de paramètre  $\theta$  sous cette fonction de perte est :

$$\delta(x) = (E_\theta(\theta^{-p}))^{-\frac{1}{p}}$$

- Lorsque  $p = 1$ , l'estimateur de Bayes coïncide avec l'estimateur de Bayes sous la fonction de perte quadratique pondéré  $\frac{(d-\theta)^2}{\theta}$
- Lorsque  $p = -1$ , l'estimateur de Bayes coïncide avec l'estimateur de Bayes sous la fonction de perte quadratique.

## 2.8 Les Méthodes numériques stochastique

### 2.8.1 Méthodes MCMC

Le but des méthodes MCMC est de simuler selon  $\pi$  et l'idée de base est de construire une chaîne de Markov ergodique de loi stationnaire  $\pi$ .

### 2.8.2 Algorithme Hasting-Metropolis

Pour  $\theta^{(0)}$  est une valeur initiale, on définit par récurrence les valeurs de  $\theta^{(t)}$ .

A l'étape  $t$ , à partir de  $\theta^{(t-1)}$ ,  $\theta^{(t)}$  est construit en tirant un  $\theta'$  à l'aide d'une distribution de probabilité instrumentale :  $\theta' \sim q(\cdot | \theta^{(t-1)})$ .  $\theta^{(t)}$  est alors donné par :

$$\theta^{(t)} = \begin{cases} \theta', & \text{avec une probabilité } \alpha(\theta', \theta^{(t-1)}) \\ \theta^{(t-1)}, & \text{avec une probabilité } 1 - \alpha(\theta', \theta^{(t-1)}) \end{cases}$$

Où  $\alpha(\theta', \theta^{(t-1)}) = \min\left(\frac{\pi(\theta')}{\pi(\theta^{(t-1)})} \frac{q(\theta^{(t-1)} | \theta')}{q(\theta' | \theta^{(t-1)})}, 1\right)$ . Notons qu'il est possible suivant cette construction de rester au même endroit après une itération. On peut alors montrer en écrivant la condition de balance, que pour ce choix de  $\alpha$ , on obtient une chaîne de Markov de loi stationnaire  $\pi$ .

Cette chaîne de Markov est ergodique si et seulement si  $(\theta^t)_t$  est irréductible et apériodique.

**Exemple 2.1** *Marche aléatoire (ou symétrique)*

$q(\theta | \theta') = q(\theta' | \theta)$  et  $\theta' = \theta^{t-1} + \varepsilon \sigma$  Où  $\varepsilon$  a une loi symétrique par rapport à 0.

Le choix de  $\sigma$  est ici primordial. Comme  $\alpha(\theta', \theta^{(t-1)}) = \min\left(\frac{\pi(\theta')}{\pi(\theta^{(t-1)})}, 1\right)$ , un  $\sigma$  trop faible ne permet pas d'explorer tout le support puisque les  $\theta^{(t)}$  sont proches les uns des autres ; si  $\sigma$  est grand, alors on refuse souvent  $\theta'$  et donc on ne profite pas bien du nombre de simulations. Empiriquement, le taux d'acceptation optimal est entre 0,1 et 0,6, il faut donc choisir  $\sigma$  en fonction.



### 2.8.3 Algorithme de type Gibbs

Pour  $\theta = (\theta_1, \dots, \theta_p)$ , on veut simuler  $\pi(\theta)$  à partir de  $\pi_i(\theta_i|\theta_{(-i)}) = \pi_i(\theta_i|\theta_j, j \neq i)$  pour tout  $i$ . On initialise avec  $\theta^{(0)}$  et à l'instant  $t$ , on écrit :

$$\begin{aligned} (\theta_1^{(t)}|\theta^{(t-1)}) &\sim \pi_1(\theta_1^{(t)}|\theta_{(-1)}^{(t-1)}) \\ (\theta_2^{(t)}|\theta^{(t-1)}, \theta_1^{(t)}) &\sim \pi_2(\theta_2|\theta_1^{(t)}, \theta_3^{(t-1)}, \dots, \theta_p^{(t-1)}) \\ &\vdots \sim \vdots \\ (\theta_p^{(t)}|\theta^{(t-1)}, \theta_{(-p)}^{(t)}) &\sim \pi_p(\theta_p^{(t)}|\theta_{(-p)}^{(t)}) \end{aligned}$$

Dans le cas où une telle loi  $\pi$  existe,  $\theta^{(t)}$  issu de cet algorithme est une chaîne de Markov ergodique de loi stationnaire  $\pi$ .

En pratique, on combine souvent les deux grandes approches précédentes comme le montre l'exemple qui suit.

**Exemple 2.8.1** *Lois normales*

$(X_{ij}|\mu_j, \sigma_j^2) \sim N(\mu_j, \sigma_j^2)$  avec  $\mu_j \sim N(\mu_0, \tau_0^2)$  et  $\sigma_j^2 \sim IG(a, b)$  on peut récupérer  $\tau_0, \mu_0, a, b$  à partir des lois conjuguées mais les calculs ne sont pas si simples. On considère plutôt  $\pi(\mu_0|X_i, \tau_0, a, b) \propto \tilde{\pi}(\tau_0, \mu_0, a, b)$  et  $(\mu_0^t|\tau_0^t, a^t, b^t) \sim q(\cdot|\mu_0^{t-1}, \tau_0^t, a^t, b^t)$ ; c'est alors une chaîne de Markov de loi stationnaire  $\pi$ .

# Chapitre 3

## La loi de Lindley à un paramètre

### 3.1 Introduction

Dans la théorie de la fiabilité, on s'intéresse à la durée de vie d'un système ou d'un élément. On considère des systèmes réparables justables à une distribution de durée de survie de Lindley généralisée, dont la distribution de Lindley est un cas particulier. On traite le problème de l'estimation paramétrique des paramètres et de quelques caractéristiques de fiabilité.

Noter que la distribution de Lindley est un mélange des distributions  $exp(\theta)$  et  $Gamma(2, \theta)$  avec les proportions de mélange sont respectivement  $p = \frac{\theta}{1+\theta}$  et  $(1 - p) = \frac{1}{1+\theta}$ .

Cette distribution fournit un meilleur ajustement à l'ensemble empirique des données considérées que les distributions binomiale et Hermite négatifs. Récemment, une bonne partie de l'attention a été accordée à cette fonction de densité de probabilité (pdf) dans la littérature statistique. Par exemple, Ghitany et al. (2008) et Ghitany et Al-Mutairi (2009) ont étudié certaines propriétés de la distribution de Poisson-Lindley discrète proposée par Sankaran (1970). Ce travail a été prolongé par Mahmoudi et Zakerzadeh (2010). Shanker et. Al. Ont introduit une distribution Lindley à deux paramètres. Zakerzadeh et

al. Ont proposé une nouvelle distribution de durée de vie à deux paramètres : modèle et propriétés. Ghitany et al. Ont travaillé sur l'estimation de la fiabilité d'un système de contraintes de la distribution de puissance de Lindley. Elbatal et al. Ont proposé une nouvelle distribution généralisée de Lindley.

### Censure

La censure est le phénomène le plus couramment rencontré lors du recueil des données de survie :

Pour l'individu  $i$ , considérons :

- Son temps de survie  $X_i$ .
- Son temps de censure  $C_i$ .
- La durée réellement observée  $T_i$ .

### Censure à droite

La durée de vie est dite censurée à droite si l'individu n'a pas subi l'évènement à sa dernière observation. En présence de censure à droite, les durées de vie ne sont pas toutes observées ; pour certaines d'entre elles, on sait seulement qu'elles sont supérieures à une certaine valeur connue **Censure à gauche** La censure à gauche correspond au cas où l'individu a déjà subi l'évènement avant que l'individu soit observé. On sait uniquement la date de l'évènement inférieure à une certaine date connue. Pour chaque individu, on peut associer un couple des variables aléatoire  $(T, \delta)$  :

$$T = X \vee C = \max(X, C)$$

$$\delta = \mathbf{1}_{X \geq C}$$

### Censure par intervalle

Une date est censurée par intervalle, si au lieu d'observer avec certitude le temps de l'évènement, la seule information disponible est qu'il a eu lieu entre deux dates connues

### Données progressivement censurées et l'algorithme de Balakrishnan

Les données progressivement censurées peuvent être décrites par la méthode suivante :

On suppose que  $n$  unités indépendantes sont mis dans un test et le censure  $R = (r_1, r_2, \dots, r_m)$  est déjà fixé. Lors du premier échec, dit  $X_{1:m:n}$ ,  $r_1$  unités sont éliminées au hasard de  $n - 1$  unités restantes. Dans le deuxième échec, dit  $X_{2:m:n}$ ,  $r_2$  unités sont éliminés au hasard de  $n - r_1 - 2$  unités restantes. On continue la procédure jusqu'au le  $m^{ime}$  échec, dit  $X_{m:m:n}$ , tous le reste  $r_m = n - m - r_1 - \dots - r_{m-1}$  sont éliminés. Puis,  $X_{1:m:n} < X_{2:m:n} < \dots < X_{m:m:n}$  sont appelés les statistiques d'ordre progressivement censurées.

**Remarque** Pour générer les données progressivement censurées d'une distribution connue, on utilise l'algorithme de Balakrishnan et Sandhu avec les étapes suivantes :

- 1) Générer  $m$  variables identiquement indépendantes distribuées  $(u_1, u_2, \dots, u_m)$  de la loi uniforme  $U(0, 1)$ .
- 2) Soit  $z_i = -\log(1 - u_i)$ ,  $z_i$  sont identiquement indépendantes distribuées de la distribution exponentielle standard.
- 3) En donnant les censures  $R(R_1, R_2, \dots, R_m)$ , soit  $y_1 = \frac{z_1}{m}$ , et pour  $i = 1, \dots, m$

$$y_i = y_{i-1} + \frac{z_i}{n - \sum_{j=1}^{i-1} R_j - i - 1}$$

Donc,  $(y_1, y_2, \dots, y_m)$  sont des données progressivement censurées d'un échantillon  $U(0, 1)$ .

- 4) Soit  $w_i = 1 - \exp(-y_i)$ .
- 5) Soit  $x_i = F^{-1}(w_i)$  i.e  $w_i = F(x_i)$ , donc,  $(x_1, x_2, \dots, x_m)$  sont des données progressivement censurées de la distribution qu'on veut générer.

## 3.2 Le modèle

On considère une variable aléatoire de loi de Lindley à un paramètre  $\theta > 0$ , la fonction de densité est donnée par :

$$f(x; \theta) = \frac{\theta^2(1+x)e^{-\theta x}}{1+\theta}, x > 0, \theta > 0$$

Noter que la distribution de Lindley est un mélange des distributions  $exp(\theta)$  et  $Gamma(2, \theta)$  avec les proportions de mélange sont respectivement  $\frac{\theta}{1+\theta}$  et  $\frac{1}{1+\theta}$  La fonction de répartition cumulative correspondante est donnée par :

$$F(x; \theta) = 1 - \left(1 + \frac{\theta x}{1+\theta}\right)e^{-\theta x}, x > 0, \theta > 0$$

La fonction de fiabilité correspondante est définie comme suit :

$$S(x; \theta) = 1 - F(x; \theta) = \left(1 + \frac{\theta x}{1+\theta}\right)e^{-\theta x}, x > 0, \theta > 0$$

La fonction de taux de hasard est :

$$h(x; \theta) = \frac{f(x; \theta)}{S(x; \theta)} = \frac{\theta(1+x)}{1+\theta(1+x)}$$

Le moment d'ordre  $r$  de la distribution de Lindley est :

$$\mu'_r = E(X^r) = \frac{r!(\theta + r + 1)}{\theta^r(\theta + 1)}; r = 1, 2, \dots$$

pour que :

$$\mu'_1 = \text{moyenne} = \frac{(\theta + 2)}{\theta(\theta + 1)}$$

et la variance est :

$$Var(X) = \frac{(\theta^2 + 4\theta + 2)}{\theta^2(\theta + 1)^2}$$

La médiane (Md) de la distribution de Lindley est la solution de l'équation non linéaire suivante :

$$F(Md) = 0.5$$

$$0.5 - \frac{e^{-\theta Md}(1 + \theta + \theta Md)}{(1 + \theta)} = 0$$

La mode est la valeur de  $x$  pour laquelle  $f(x)$  est maximum. Par méthode de différenciation, on obtient le mode (Mo) comme :

$$Mo = \begin{cases} \frac{1-\theta}{\theta}, & \text{si } 0 < \theta < 1; \\ 0, & \text{sinon} \end{cases}$$

**Exemple 3.2.1** *La moyenne, médiane, mode et variance de la loi de Lindley à un paramètre*

$\theta$	0,1	0,2	0,5	0,75	1	1,5	2	5	10	20
<i>moyenne</i>	19,091	9,167	3,333	2,095	1,5	0,779	0,667	0,233	0,109	0,052
<i>Médian</i>	15,858	7,532	2,654	1,631	1,146	0,574	0,487	0,164	0,076	0,036
<i>Mode</i>	9	4	1	0,333	0	0	0	0	0	0
<i>Variance</i>	199,174	49,306	7,556	3,229	1,75	0,521	0,389	0,052	0,012	0,003

*Tableau.2- Représente la moyenne, médiane, mode et variance de la loi de Lindley à un paramètre pour différentes valeurs de  $\theta$*

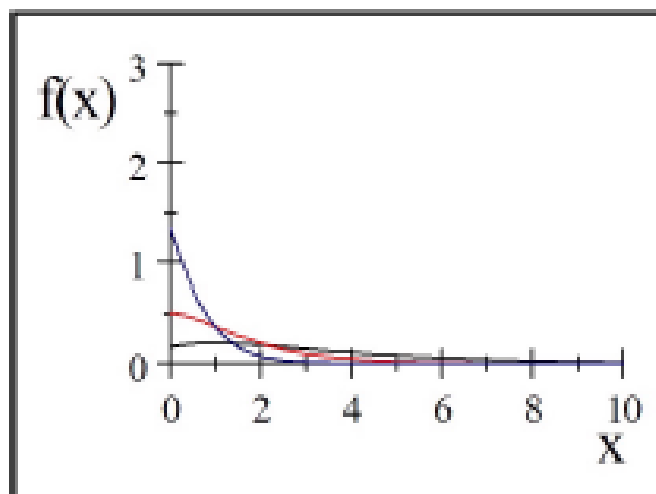


FIGURE 3.1 – Présentation graphique de la fonction de densité pour quelques valeurs de  $\theta$ , noir( $\theta = 0.5$ ); rouge ( $\theta = 1$ ); bleu( $\theta = 2$ )

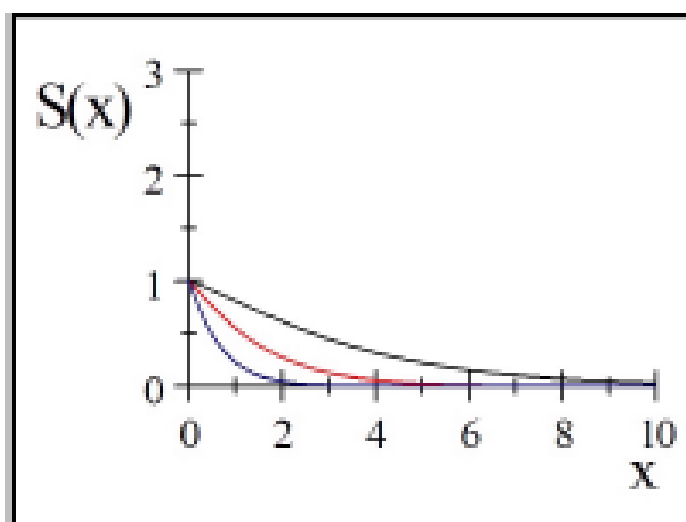


FIGURE 3.2 – Présentation graphique de la fonction de fiabilité pour quelques valeurs de  $\theta$ , noir( $\theta = 0.5$ ); rouge ( $\theta = 1$ ); bleu( $\theta = 2$ )

### 3.3 Estimation par la méthode du maximum de vraisemblance

#### 3.3.1 Estimation du paramètre $\theta$

Soit un n-échantillon  $(x_1, x_2, \dots, x_n)$  de la loi de Lindley à un paramètre  $\theta$ , on suppose les données sont progressivement censurée à l'ordre  $m$  avec :  $\underline{R} = (R_1, R_2, \dots, R_m)$ .

La fonction du vraisemblance est :

$$\begin{aligned} L(\theta, \underline{x}) &= A \prod_{i=1}^m f(x_i) [1 - F(x_i)]^{R_i} \\ &= A \prod_{i=1}^m \frac{\theta(1+x_i)}{\theta+1} e^{-\theta x_i} \left[1 - \left(1 - \left(1 + \frac{\theta x_i}{1+\theta}\right) e^{-\theta x_i}\right)\right]^{R_i} \end{aligned}$$

avec

$$A = n(n - R_1)(n - 2 - R_1 - R_2) \dots \left(n - \sum_{i=1}^{m-1} (R_i - 1)\right)$$

On prend le logarithme de la vraisemblance on obtient :

$$\begin{aligned} l(\theta, \underline{x}) &= \ln L(\theta, \underline{x}) = \ln \left[ A \left( \frac{\theta^2}{\theta+1} \right)^m \prod_{i=1}^m (1+x_i) e^{-\theta \sum_{i=1}^m x_i} \prod_{i=1}^m \left[ \left(1 + \frac{\theta x_i}{1+\theta}\right) e^{-\theta x_i} \right]^{R_i} \right] \\ &= \ln A + m \ln \left( \frac{\theta^2}{\theta+1} \right) + \sum_{i=1}^m \ln(1+x_i) + \ln(e^{-\theta \sum_{i=1}^m x_i}) + \sum_{i=1}^m \ln \left[ \left(1 + \frac{\theta x_i}{1+\theta}\right) e^{-\theta x_i} \right]^{R_i} \\ &= \ln A + m(\ln(\theta^2) - \ln(\theta+1)) + \sum_{i=1}^m \ln(1+x_i) - \theta \sum_{i=1}^m x_i + \sum_{i=1}^m R_i \left[ \ln\left(1 + \frac{\theta x_i}{1+\theta}\right) + \ln(e^{-\theta x_i}) \right] \\ &= \ln A + m(\ln(\theta^2) - \ln(\theta+1)) + \sum_{i=1}^m \ln(1+x_i) - \theta \sum_{i=1}^m x_i + \sum_{i=1}^m R_i \left[ \ln\left(1 + \frac{\theta x_i}{1+\theta}\right) - \theta x_i \right] \\ &= \ln A + m(\ln(\theta^2) - \ln(\theta+1)) + \sum_{i=1}^m \ln(1+x_i) - \theta \sum_{i=1}^m x_i + \sum_{i=1}^m R_i \ln\left(1 + \frac{\theta x_i}{1+\theta}\right) - \theta \sum_{i=1}^m R_i x_i \end{aligned}$$

L'estimateur du maximum de vraisemblance de paramètre  $\theta$  est la solution de l'équation

$$l(\theta, \underline{x}) = 0$$



On dérive par rapport à  $\theta$  on obtient :

$$\begin{aligned} \frac{\partial l(\theta)}{\partial \theta} &= m \left( \frac{2\theta}{\theta^2} - \frac{1}{\theta+1} \right) - \sum_{i=1}^m x_i + \sum_{i=1}^m R_i \left( \frac{x_i}{(1+\theta)(1+\theta+\theta x_i)} \right) - \sum_{i=1}^m R_i x_i \\ &= \frac{2m}{\theta} - \frac{m}{\theta+1} - \sum_{i=1}^m x_i + \sum_{i=1}^m R_i \left( \frac{x_i}{(1+\theta)(1+\theta+\theta x_i)} \right) - \sum_{i=1}^m R_i x_i \end{aligned}$$

L'équation est non linéaire donc il n'y a pas de solution analytique. On utilise les méthodes numériques pour obtenir la valeur approximée au estimateur du maximum de vraisemblance  $\hat{\theta}_{MV}$  de paramètre  $\theta$ . Dans ce chapitre, on utilise le package du R (le package BB) qui a une capacité pour résoudre l'équation non linéaire

### 3.3.2 Estimation de la fonction de fiabilité

L'estimateur de la fonction de fiabilité  $S(\theta, x)$  est obtenu, en remplaçant  $\theta$  par  $\hat{\theta}_{MV}$  dans l'expression de la fonction de fiabilité, donc :

$$\hat{S}_{MV}(\theta, x) = \left[ 1 + \frac{\hat{\theta}_{MV} x}{1 + \hat{\theta}_{MV}} \right] e^{-\hat{\theta}_{MV} x}, x > 0$$

L'erreur quadratique de l'estimateur de la fonction de fiabilité est :

$$EQ(\hat{S}_{MV}(\theta, x)) = (\hat{S}_{MV}(\theta, x) - S(\theta, x))^2$$

## 3.4 Estimation bayésienne sous différentes fonctions de perte

### 3.4.1 Les distributions a posteriori et a priori

Pour le modèle de Lindley, plusieurs auteurs se sont intéressés à l'estimation du paramètre par différentes méthodes. Dans le cadre d'une approche bayésienne, nous proposons en premier lieu, une loi a priori conjuguée naturelle

$$\pi(\theta) = \frac{b^a}{\Gamma(a)} \theta^{a-1} e^{-b\theta}, a, b > 0$$

La distribution a postérieure du paramètre  $\theta$  est :

$$\begin{aligned} \pi(\theta|\underline{x}) &= \frac{\pi(\theta)L(\theta|\underline{x})}{\int_{\theta} \pi(\theta)L(\theta|\underline{x})d\theta}, \theta > 0 \\ &= \frac{\frac{b^a}{\Gamma(a)} \theta^{a-1} e^{-b\theta} A \left( \frac{\theta^2}{1+\theta} \right)^m \prod_{i=1}^m (1+x_i) e^{-\theta \sum_{i=1}^m x_i} \prod_{i=1}^m \left[ \left(1 + \frac{\theta x_i}{\theta+1}\right) e^{-\theta x_i} \right]^{R_i}}{\int_{\theta} \frac{b^a}{\Gamma(a)} \theta^{a-1} e^{-b\theta} A \left( \frac{\theta^2}{1+\theta} \right)^m \prod_{i=1}^m (1+x_i) e^{-\theta \sum_{i=1}^m x_i} \prod_{i=1}^m \left[ \left(1 + \frac{\theta x_i}{\theta+1}\right) e^{-\theta x_i} \right]^{R_i} d\theta} \\ &= \frac{\theta^{a-1} e^{-b\theta} \left( \frac{\theta^2}{1+\theta} \right)^m e^{-\theta \sum_{i=1}^m x_i} \prod_{i=1}^m \left[ \left(1 + \frac{\theta x_i}{\theta+1}\right) e^{-\theta x_i} \right]^{R_i}}{\int_{\theta} \theta^{a-1} e^{-b\theta} \left( \frac{\theta^2}{1+\theta} \right)^m e^{-\theta \sum_{i=1}^m x_i} \prod_{i=1}^m \left[ \left(1 + \frac{\theta x_i}{\theta+1}\right) e^{-\theta x_i} \right]^{R_i} d\theta} \\ &= \frac{\theta^{a-1+2m} (\theta+1)^{-m} e^{-b\theta-\theta \sum_{i=1}^m x_i} \prod_{i=1}^m \left[ \left(1 + \frac{\theta x_i}{\theta+1}\right) e^{-\theta x_i} \right]^{R_i}}{\int_{\theta} \theta^{a-1+2m} e^{-b\theta-\theta \sum_{i=1}^m x_i} \prod_{i=1}^m \left[ \left(1 + \frac{\theta x_i}{\theta+1}\right) e^{-\theta x_i} \right]^{R_i} d\theta} \\ &= k^{-1} \theta^{a-1+2m} e^{-b\theta-\theta \sum_{i=1}^m x_i} \prod_{i=1}^m \left[ \left(1 + \frac{\theta x_i}{\theta+1}\right) e^{-\theta x_i} \right]^{R_i} \\ \text{Où, } k &= \int_{\theta} \theta^{a-1+2m} e^{-b\theta-\theta \sum_{i=1}^m x_i} \prod_{i=1}^m \left[ \left(1 + \frac{\theta x_i}{\theta+1}\right) e^{-\theta x_i} \right]^{R_i} d\theta \end{aligned}$$

### 3.4.2 Fonctions de perte

On considère la fonction de perte quadratique, Linex et entropie, le tableau suivant présente les trois fonctions de perte et l'expression de l'estimateur bayésien avec le risque a posteriori correspondant.

Fonction de perte	expression	Estimateur de Bayes	Risque a posteriori
quadratique	$L(\theta, \delta) = (\theta - \delta)^2$	$\hat{\theta}_Q = E_{\pi}(\theta)$	$E_{\pi}(\theta - \hat{\theta}_Q)^2$
Entropie	$L(\theta, \delta) = \left(\frac{\delta}{\theta}\right)^p - p \ln\left(\frac{\delta}{\theta}\right) - 1$	$\hat{\theta}_E = [E_{\pi}(\theta)^{-p}]^{-1/p}$	$p[E_{\pi}(\ln(\theta) - \ln(\hat{\theta}_E))]$
Linex	$L(\theta, \delta) = e^{r(\delta-\theta)} - r(\delta - \theta) - 1$	$\hat{\theta}_L = -\frac{1}{r} \ln(E_{\pi}(e^{-r\theta}))$	$r(\hat{\theta}_Q - \hat{\theta}_L)$

Tableau.3-les fonctions de perte et les estimations bayésiennes (avec les risques a posteriori) correspondantes

**Estimateur bayésienne de paramètre  $\theta$  sous la fonction de perte quadratique**

Sous la fonction de perte quadratique, l'estimateur de Bayes  $\hat{\theta}_Q$  de paramètre  $\theta$  est la moyenne a posteriori :

$$\begin{aligned}\hat{\theta}_Q &= E_\pi(\theta) = \frac{\int_\theta \theta \pi(\theta|x) d\theta}{\int_\theta \pi(\theta|x) d\theta} \\ &= \frac{\int_\theta \theta \theta^{a-1+2m} (\theta+1)^{-m} e^{-b\theta-\theta \sum_{i=1}^m x_i} \prod_{i=1}^m \left[ \left(1 + \frac{\theta x_i}{\theta+1}\right) e^{-\theta x_i} \right]^{R_i} d\theta}{\int_\theta \theta^{a-1+2m} e^{-b\theta-\theta \sum_{i=1}^m x_i} \prod_{i=1}^m \left[ \left(1 + \frac{\theta x_i}{\theta+1}\right) e^{-\theta x_i} \right]^{R_i} d\theta} \\ &= \frac{\int_\theta \theta^{a+2m} (\theta+1)^{-m} e^{-b\theta-\theta \sum_{i=1}^m x_i} \prod_{i=1}^m \left[ \left(1 + \frac{\theta x_i}{\theta+1}\right) e^{-\theta x_i} \right]^{R_i} d\theta}{\int_\theta \theta^{a-1+2m} e^{-b\theta-\theta \sum_{i=1}^m x_i} \prod_{i=1}^m \left[ \left(1 + \frac{\theta x_i}{\theta+1}\right) e^{-\theta x_i} \right]^{R_i} d\theta}\end{aligned}$$

Le risque a posteriori est :

$$\begin{aligned}PR(\hat{\theta}_Q) &= E_\pi(\theta - \hat{\theta}_Q)^2 \\ &= E_\pi(\theta^2 + \hat{\theta}_Q^2 - 2\theta\hat{\theta}_Q) \\ &= E_\pi(\theta^2) + E_\pi(\hat{\theta}_Q^2) - 2E_\pi(\theta\hat{\theta}_Q)\end{aligned}$$

**Estimateur bayésienne de paramètre  $\theta$  sous la fonction de perte entropie**

Sous la fonction de perte entropie, l'estimateur de Bayes  $\hat{\theta}_E$  de paramètre  $\theta$  est :

$$\begin{aligned}\hat{\theta}_E &= [E_\pi(\theta^{-p})]^{-1/p} \\ &= \left[ \frac{\int_\theta \theta^{-p} \pi(\theta|\underline{x}) d\theta}{\int_\theta \pi(\theta|\underline{x}) d\theta} \right]^{-1/p} \\ &= \left[ \frac{\int_\theta \theta^{-p} \theta^{a-1+2m} (\theta+1)^{-m} e^{-b\theta-\theta \sum_{i=1}^m x_i} \prod_{i=1}^m \left[ \left(1 + \frac{\theta x_i}{\theta+1}\right) e^{-\theta x_i} \right]^{R_i} d\theta}{\int_\theta \theta^{a-1+2m} e^{-b\theta-\theta \sum_{i=1}^m x_i} \prod_{i=1}^m \left[ \left(1 + \frac{\theta x_i}{\theta+1}\right) e^{-\theta x_i} \right]^{R_i} d\theta} \right]^{-1/p} \\ &= \left[ \frac{\int_\theta \theta^{a-1+2m-p} (\theta+1)^{-m} e^{-b\theta-\theta \sum_{i=1}^m x_i} \prod_{i=1}^m \left[ \left(1 + \frac{\theta x_i}{\theta+1}\right) e^{-\theta x_i} \right]^{R_i} d\theta}{\int_\theta \theta^{a-1+2m} e^{-b\theta-\theta \sum_{i=1}^m x_i} \prod_{i=1}^m \left[ \left(1 + \frac{\theta x_i}{\theta+1}\right) e^{-\theta x_i} \right]^{R_i} d\theta} \right]^{-1/p}\end{aligned}$$

Le risque a posteriori :

$$PR(\hat{\theta}_E) = pE_\pi(\ln(\theta) - \ln(\hat{\theta}_E))$$

**Estimateur bayésienne de paramètre  $\theta$  sous la fonction de perte Linex**

Sous la fonction de perte Linex, l'estimateur de Bayes  $\hat{\theta}_L$  de paramètre  $\theta$  est :

$$\begin{aligned}\hat{\theta}_L &= -\frac{1}{r} \ln [E_\pi(e^{-r\theta})] \\ &= -\frac{1}{r} \ln \left[ \int_{\theta} e^{-r\theta} \pi(\theta | \underline{x}) d\theta \right] \\ &= -\frac{1}{r} \ln \left[ \int_{\theta} e^{-r\theta} \theta^{a-1+2m} (\theta + 1)^{-m} e^{-b\theta - \theta \sum_{i=1}^m x_i} \prod_{i=1}^m \left[ \left(1 + \frac{\theta x_i}{\theta + 1}\right) e^{-\theta x_i} \right]^{R_i} d\theta \right] \\ &= -\frac{1}{r} \ln \left[ \int_{\theta} \theta^{a-1+2m} (\theta + 1)^{-m} e^{-b\theta - r\theta - \theta \sum_{i=1}^m x_i} \prod_{i=1}^m \left[ \left(1 + \frac{\theta x_i}{\theta + 1}\right) e^{-\theta x_i} \right]^{R_i} d\theta \right]\end{aligned}$$

Le risque a posteriori est :

$$PR(\hat{\theta}_L) = r(\hat{\theta}_Q - \hat{\theta}_L)$$

**Exemple 3.4.1** (*H.Krishna et K.Kumar (2011)*) Dans cette section, une étude de simulation est menée pour obtenir les estimations du paramètre et des caractéristiques de fiabilité de Lindley ( $\theta$ ) développé dans les sections précédentes. Les estimations du maximum de vraisemblance et de Bayes sont obtenues pour échantillons progressivement censurés à droite de type II générés pour différentes valeurs de  $\theta$ ,  $\alpha$  et  $\beta$ . Tous les calculs sont effectués sous R. Nous avons pris trois tailles d'échantillon petite taille d'échantillon  $n = 20$ , taille d'échantillon modérée  $n = 30$ , grande taille de l'échantillon  $n = 50$ . En outre, chaque taille d'échantillon a 10 schémas de censure. L'étude comprend les étapes suivantes :

- 1) Choisir les valeurs des hyper-paramètres  $\alpha$  et  $\beta$ , le temps de mission  $t$ , la taille de l'échantillon  $n$ , le nombre de pannes  $m$  et le système de censure  $\underline{R} = (R_1, R_2, R_3, \dots, R_m)$ .
- 2) Prenons  $\theta = \beta/\alpha$ , la moyenne de la distribution a priori de  $\theta$ .
- 3) Calculer les valeurs réelles de  $R(t)$  et  $h(t)$ .
- 4) Générer un échantillon progressivement censuré à droite de type II de taille  $n$  avec  $m$  échecs en utilisant l'algorithme de Balakrishnan

*Les résultats sont obtenus dans le tableau suivant :*

$n$	$m$	$R_i$	$\hat{\theta}_Q$	$PR(\hat{\theta}_Q)$	$\hat{R}_Q(t)$	$PR(\hat{R}_Q(t))$	$\hat{h}_Q(t)$	$PR(\hat{h}_Q(t))$
20	8	(12, 0 * 7)	2.0750	0.1497	0.5650	0.0053	0.7316	0.5519
20	8	(4 * 3, 0 * 5)	2.0898	0.1547	0.5622	0.0053	0.7331	0.5497
20	8	(0 * 5, 4 * 3)	2.0948	0.1640	0.5612	0.0056	0.7336	0.5490
20	8	(0 * 7, 12)	2.0702	0.1706	0.5663	0.0061	0.7308	0.5533
20	10	(10, 0 * 9)	2.0879	0.1672	0.5620	0.0058	0.7335	0.5491
20	10	(5 * 2, 0 * 8)	2.0872	0.1500	0.5616	0.0053	0.7339	0.5484
20	10	(0 * 8, 5 * 2)	2.0537	0.1377	0.5678	0.0050	0.7313	0.5524
20	10	(0 * 9, 10)	2.0813	0.1513	0.5626	0.0054	0.7336	0.5489
20	16	(4, 0 * 15)	2.0548	0.1107	0.5653	0.0042	0.7339	0.5483
20	16	(2 * 2, 0 * 14)	2.0603	0.1144	0.5643	0.0042	0.7344	0.5475
30	12	(18*, 0 * 11)	2.0441	0.1363	0.5692	0.0051	0.7309	0.5529
30	12	(6 * 3, 0 * 9)	2.0675	0.1429	0.5646	0.0051	0.7331	0.5496
30	12	(0 * 9, 6 * 3)	2.0817	0.1364	0.5613	0.0048	0.7351	0.5466
30	12	(0 * 11, 18)	2.0875	0.1378	0.5602	0.0049	0.7355	0.5460
30	15	(15, 0 * 14)	2.0594	0.1268	0.5650	0.0045	0.7338	0.5485
30	15	(5 * 3, 0 * 12)	2.0544	0.1120	0.5656	0.0042	0.7336	0.5486
30	15	(0 * 12, 5 * 3)	2.0574	0.1142	0.5649	0.0043	0.7340	0.5481
30	15	(0 * 14, 15)	2.0588	0.1197	0.5647	0.0044	0.7341	0.5479
30	24	(6, 0 * 23)	2.0378	0.0739	0.5665	0.0029	0.7347	0.5467
30	24	(2 * 3, 0 * 21)	2.0324	0.0915	0.5681	0.0036	0.7336	0.5485
50	20	(30, 0 * 19)	2.0558	0.1026	0.5641	0.0038	0.7351	0.5463
50	20	(5 * 6, 0 * 14)	2.0557	0.0983	0.5640	0.0037	0.7353	0.5461
50	20	(0 * 14, 5 * 6)	2.0515	0.0946	0.5646	0.0036	0.7351	0.5462
50	20	(0 * 19, 30)	2.0515	0.1082	0.5650	0.0040	0.7348	0.5469
50	25	(25, 0 * 24)	2.0524	0.0895	0.5639	0.0034	0.7358	0.5452
50	25	(5 * 5, 0 * 20)	2.0373	0.0856	0.5668	0.0033	0.7354	0.5471
50	25	(0 * 20, 5 * 5)	2.0404	0.0791	0.5659	0.0030	0.7352	0.5461
50	25	(0 * 24, 25)	2.0453	0.0766	0.5648	0.0029	0.7357	0.5452
50	40	(10, 0 * 39)	2.0428	0.0597	0.5640	0.0023	0.7369	0.5433
50	40	(5 * 2, 0 * 38)	2.0314	0.0552	0.5662	0.0022	0.7359	0.5447

Tableau.4-Estimation bayésienne de paramètre, de fonction de fiabilité et la fonction de taux de panne avec des données progressivement censurée

**Exemple 3.4.2** (*H.Krishna et K.Kumar (2011)*)

*Exemple de données réelles : Dans cette section, pour illustrer l'utilisation de la distribution de Lindley comme modèle de fiabilité, nous utilisons l'ensemble de données des temps d'attente (en minutes) avant le service de 100 clients bancaires comme discuté par Ghitany et al. Les temps d'attente (en minutes) sont les suivants :*

0.8, 0.8, 1.3, 1.5, 1.8, 1.9, 1.9, 2.1, 2.6, 2.7, 2.9, 3.1, 3.2, 3.3, 3.5, 3.6, 4.0, 4.1, 4.2, 4.2, 4.3, 4.3, 4.4, 4.4, 4.6, 4.7, 4.7, 4.8, 4.9, 4.9, 5.0, 5.3, 5.5, 5.7, 5.7, 6.1, 6.2, 6.2, 6.2, 6.3, 6.7, 6.9, 7.1, 7.1, 7.1, 7.1, 7.4, 7.6, 7.7, 8.0, 8.2, 8.6, 8.6, 8.6, 8.8, 8.8, 8.9, 8.9, 9.5, 9.6, 9.7, 9.8, 10.7, 10.9, 11.0, 11.0, 11.1, 11.2, 11.2, 11.5, 11.9, 12.4, 12.5, 12.9, 13.0, 13.1, 13.3, 13.6, 13.7, 13.9, 14.1, 15.4, 15.4, 17.3, 17.3, 18.1, 18.2, 18.4, 18.9, 19.0, 19.9, 20.6, 21.3, 21.4, 21.9, 23.0, 27.0, 31.6, 33.1, 38.5.

*Les résultats sont obtenus dans le tableau suivant :*

$n$	$m$	$R_i$	$\hat{\theta}$	$\hat{R}(t = 9)$	$\hat{h}(t = 9)$
100	40	(1 * 39, 21)	0.1218	0.6607	0.0669
	40	(2 * 15, 0 * 10, 2 * 15)	0.1040	0.7248	0.0530
	40	((3, 0) * 20)	0.1054	0.7197	0.0541
	40	((0 * 39), 60)	0.1845	0.4563	0.1197
	50	(1 * 50)	0.1189	0.6710	0.0646
	50	((2, 0) * 25)	0.1191	0.6703	0.0647
	50	((3, 0, 0) * 16, 2, 0)	0.1195	0.6690	0.0650
	50	((0 * 49), 50)	0.1834	0.4595	0.1187
	80	(1 * 20, 0 * 60)	0.1572	0.5402	0.0960
	80	((1, 0, 0, 0) * 20)	0.1755	0.4831	0.1118
	80	((2, 0, 0, 0, 0, 0, 0) * 10)	0.1709	0.4969	0.1078
	80	(0 * 79, 20)	0.1901	0.4404	0.1246
	100	(0 * 100)	0.1866	0.4505	0.1215

*Tableau.5-Estimation bayésienne de paramètre, de fonction de fiabilité et la fonction de taux de panne avec des données réelles*

## Conclusion

Dans ce travail, nous nous sommes intéressés à la distribution de Lindley à un seul paramètre pour la modélisation des données progressivement censurées. L'estimation du paramètre a été effectuée par la méthode du maximum de vraisemblance et par une approche bayésienne sous les trois fonctions de perte (quadratique, entropie et Linex) avec un échantillon progressivement censuré.

En perspectives, ce travail peut être élargi pour le modèle de Lindley par l'étude de prédiction de statistique d'ordre et des valeurs extrêmes.



# Bibliographie

- [1] Aouf Fairouz, Estimation Bayésienne dans un modèle de Lindley généralisé, THESE de Doctorat en Mathématique, Année : 2016-2017.
- [2] Bolshev, L. N, Smirnov, N. V. Tables of Mathematical Statistics. Science, Moscow, [in Russian], (1983).
- [3] Chen, Z, A new two-parameter lifetime distribution with bathtub shape or increasing failure rate function, Statistics and Probability Letters, 49 ; 155-161.(2000).
- [4] Deniz, E. G. and Calderin-Ojeda, E. The discrete Lindley distribution : properties and applications, Journal of Statistical Computation and Simulation. 81(11), 1405 - 1416, 2011.
- [5] FELFLI ASMA, La distribution Extension de Weibull(EW), Mémoire de Master en Mathématique, Année : 2019-2020.
- [6] Ghitany, M .E, Atieh, B, Nadarajah, Lindley distribution and its applications, Math. Comput. Simul. 78, 493 - 506, 2008.
- [7] Hougaard, P. Survival models for heterogeneous populations derived from stable distributions, Biometrika 73 ; 387-396. (1986).
- [8] Howlader HA, HossainA, On Bayesian estimation and prediction from realibility based on type II censored data, communications in statistics-Theory and methods 24, 2249 ?2259 (1995)
- [9] Jeong, J. H, A new parametric family for modelling cumulative incidence

function : application to breast cancer data, Journal of the Royal Statistical Society : Series A, 169(2); 289-303. (2006)

[10] Judith ROUSSEAU, STATISTIQUE BAYESIENNE-Notes de cours, Année : 2009-2010

[11] Krishna, H. and Kumar, K. Reliability estimation in Lindley distribution with progressively type-II right censored sample, Mathematics and Computers in Simulation. 82(2), 281 - 294, 2011.

[12] Lindley, D. V. Fiducial distributions and bayes theorem. Journal of the Royal Society, series B, 20, 102-107. (1958).

[13] NEDJAR SIHEM, Poisson Pseudo Lindley Distributions et leurs applications en assurance-vie, THESE de Doctorat en sciences, Année : 2017.

[14] Robert. C, L'analyse statistique Bayesienne, Economica (1992).