الجمهورية الجزائرية الديمقراطية الشعبية

Ministère de L'Enseignement Supérieur et de la Recherche Scientifique

UNIVERSITÉ SAAD DAHLAB DE BLIDA
Faculté des sciences

Département de Mathématiques

Mémoire de Master

Spécialité : Modélisation Stochastique et Statistique

Thème :

Inférence statistique sur l'indice de performance de la loi de Lindley à partir de données censurées progressivement de type II

Réalisé par :

Benaouda Soumia & Elhattah Aya

Devant le jury composé de :

Omar TAMI	Président	MCB.	Univ.Blida 1
Redouane FRIHI	Examinateur	MCB.	Univ.Blida 1
Abdelaziz RASSOUL	Promoteur	Prof.	ENSH, Blida

Juin 2025

REMERCIEMENTS

Nous exprimons notre profonde reconnaissance et notre gratitude infinie à Dieu Tout-Puissant, qui, par Sa miséricorde et Sa bonté infinies, nous a doté de la force, de la patience et de la persévérance indispensables pour mener à bien ce travail avec succès et dévouement.

Nous adressons nos plus sincères remerciements et notre profonde gratitude à notre éminent professeur et superviseur honorable, le Dr Abdelaziz Rassoul, pour ses orientations scientifiques précises, ses précieux conseils, ainsi que son suivi continu tout au long de la préparation de ce travail. Il a été un soutien exemplaire et un guide éclairé ; que Dieu le récompense amplement et le guide vers tout ce qui est bien et couronné de succès.

Nous exprimons également notre sincère reconnaissance aux membres respectés du jury d'examen, qui nous ont honorés en acceptant d'évaluer ce travail et l'ont enrichi de leurs observations constructives et de leurs précieux conseils.

Nous ne saurions oublier d'exprimer notre profonde gratitude et notre estime à l'ensemble des professeurs du département de mathématiques de l'Université Saad Dahlab – Blida 1, qui ont joué un rôle majeur dans notre formation scientifique tout au long de nos années d'études. Chacun d'eux a laissé une marque distinctive et un impact inoubliable.

Nous adressons aussi nos vifs remerciements et notre respect à l'équipe administrative du département et de la faculté, pour leurs efforts constants, leur souci permanent de créer des conditions favorables à l'apprentissage, ainsi que pour leur attitude respectueuse et leur soutien tout au long de notre parcours.

À vous tous, nous adressons nos plus profondes expressions de reconnaissance, accompagnées de nos vœux sincères de réussite et de succès. Que Dieu vous récompense tous abondamment.

Dédicace

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

♡ **À moi-même**

«Pour avoir persévéré malgré les doutes, Pour avoir cru en mes capacités, même lorsque le chemin semblait flou. Ce mémoire est le fruit de mes efforts, de mes nuits blanches, et de mon désir constant d'apprendre et d'évoluer. Je me dédie ce travail avec fierté et gratitude.»

♡ **À mon très cher père Abd elkader** ,

«Pour l'homme qui m'a appris la valeur de l'effort et la noblesse du cœur .
Merci papa, pour ta présence discrète mais essentielle à chaque étape de ma vie»

♡ **À ma très chère mère Rachida**,

«Pour celle qui a fait de moi ce que je suis, avec tendresse et dévouement.
Pour toi, maman chaque mot de ce travail est un hommage à ton amour.»

♡ **À mes belles sœurs « Firdaous et Ritadj »**.

« mes précieuses sœurs, complices de mes secrets,
témoins de mes rêves, et les premiers visages de l'amitié véritable.»

♡ **À ma très chère amie Maria Ayadi**

«Pour mon amie de cœur, présence fidèle dans chaque saison de ma vie.
Merci d'être cette amie rare,
ton amitié est l'un des plus beaux chapitres de mon histoire.»

♡ **À Elhattah Aya**

«qui fut pour moi un soutien précieux
et une partenaire de ce parcours.»

♡ **À lui**

«À celui qui a cru en moi quand je doutais de tout,
À celui dont la foi en moi a été ma lanterne dans l'obscurité.
Tu n'as peut-être pas de nom sur ces pages, Mais ton empreinte y est gravée à jamais.
Merci, pour ta présence discrète mais essentielle. »

♡ **À tout les membres de ma famille**

♡ **À tous mes ami(e)s et mes collègues**

Benaouda Soumia

Dédicace

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

♡ **À mon très cher père Abd elkrim**,

« Source d'amour, d'affection, de générosité et de sacrifices.
Tu étais toujours là près de moi pour me soutenir et me guider
avec tes précieux conseils. Aucune dédicace ne saurait exprimer
l'amour que je porte au grand homme que vous êtes.
Que Dieu me le garde. »

♡ **À ma très chère mère Yamina**,

« 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. »

♡ **À mon cher frère** «Walid »
et mes belles sœurs « Riyane,Iméne et Rania ».

Puisse Dieu vous donner santé, bonheur, courage
et surtout réussite.

♡ **À Benaouda Soumia**,
qui fut pour moi un soutien précieux
et une partenaire de ce parcours.

♡ **À tous mes ami(e)s et mes collègues**

♡ **À tous ceux qui m'ont aidé de près ou de loin**

Elhattah Aya

TABLE DES MATIÈRES

TABLE DES FIGURES

LISTE DES TABLEAUX

ملخص :

يتمحور هذا العمل حول دراسة مؤشر الأداء في مجال الموثوقية، وذلك من خلال نمذجة أزمنة الحياة اعتمادًا على توزيع ليندلي بصِيغه المختلفة. يهدف البحث إلى تقدير هذا المؤشر باستخدام طريقة الإمكان الأعظمي، من خلال بيانات خاضعة للرقابة التصاعدية من النوع الثاني، وهي الحالة التي تمثّل بشكل أكثر واقعية العديد من التطبيقات العملية.

في هذا الإطار، تم اعتماد توزيع ليندلي بصفته نموذجًا مرنًا وفعالًا، حيث تم تناول نسخه ذات المعلمة الواحدة، المعلمتين، مع دراسة خصائصه الإحصائية الأساسية، مثل دالة الكثافة، ودالة التوزيع، والتوقع، والتباين.

الكلمات المفتاحية : دالة الكثافة الاحتمالية، دالة التوزيع التراكمي، توزيع ليندلي، توزيع ويبل، توزيع غاما، مؤشر الأداء، التباين، طريقة الإمكان الأعظمي، التقدير، ملاءمة النموذج، الرقابة، بيانات مراقبة من النوع الثاني، بيانات مراقبة تصاعديا من النوع الثاني.

Résumé :

Ce travail porte sur l'étude de l'indice de performance dans le domaine de la fiabilité, à travers la modélisation des temps de vie en s'appuyant sur les différentes formes de la distribution de Lindley. L'objectif de cette recherche est d'estimer cet indice en utilisant la méthode du maximum de vraisemblance, à partir de données soumises à une censure progressive de type II, un cadre qui reflète de manière plus réaliste de nombreuses applications pratiques.

Dans ce contexte, la distribution de Lindley a été adoptée comme un modèle flexible et efficace. Ses versions à un et deux paramètres ont été examinées, avec une étude de ses propriétés statistiques fondamentales, telles que la fonction de densité, la fonction de répartition, l'espérance et la variance.

Mots-clés : fonction de densité de probabilité, fonction de répartition cumulative, distribution de Lindley, distribution de Weibull, distribution Gamma, indice de performance, variance, méthode du maximum de vraisemblance, estimation, ajustement du modèle, censure, données censurées de type II, données censurées progressivement de type II.

Abstract :

This work focuses on the study of the performance index in the field of reliability, through the modeling of lifetime data based on the different forms of the Lindley distribution. The main objective of this research is to estimate this index using the maximum likelihood method, based on data subject to Type II progressive censoring, a framework that more realistically reflects many practical applications.

In this context, the Lindley distribution has been adopted as a flexible and efficient model. Its one-parameter and two-parameter versions have been examined, along with a study of its fundamental statistical properties, such as the probability density func-

tion, cumulative distribution function, mean, and variance.

Keywords : probability density function, cumulative distribution function, Lindley distribution, Weibull distribution, Gamma distribution, performance index, variance, maximum likelihood method, estimation, model fitting, censoring, Type II censored data, progressively Type II censored data.

Abréviations et Notations

Abréviations

MTTF	Le temps moyen avant défaillance.
IPDV	L'indice de performance de la durée de vie.
fdp	La fonction de densité de probabilité.
fdc	La fonction de distribution cumulative.
CvM	Cramér-Von Mises statistic.
BIC	Bayesian Information Criterion
AD	Anderson-Darling statistic.
EMV	Estimateur du Maximum de Vraisemblance.
AIC	Akaike Information Criterion
KS	Kolmogorov-Smirnov statistic
i.i.d	identiquement et indépendants distribués

Notations

$R(t)$	La fiabilité.
$\lambda(t)$	Le taux de défaillance.
C_{LX}	L'indice de performance de durée de vie.
σ_X	L'écart-type de X.
$E(X)$	L'espérance de X.
$Var(X)$	La variance de X.
$F_X(x)$	La fonction de répartition.
$f_X(x)$	La fonction de densité.
$(X_1, ..., X_n)$	Échantillon de taille n de X.
$(X_{1:m:n}, X_{2:m:n}, ..., X_{m:m:n})$	Échantillon progressivement censurés de type II.

INTRODUCTION GÉNÉRALE

La fiabilité constitue un domaine central en statistique appliquée, consacré à l'analyse de l'efficacité et de la durabilité des systèmes ou des composantes dans diverses conditions d'utilisation. Parmi les indicateurs majeurs de ce domaine, l'indice de performance, c'est un outil indispensable pour examiner les caractéristiques de défaillance et de mesurer l'efficacité d'un système à partir des données relatives aux durées de vie ou aux échecs.

Ce mémoire a pour objectif d'étudier cet indice à partir de différentes sortes de données, qu'elles soient complètes ou soumises à différents types de censure, en mettant particulièrement l'accent sur les méthodes d'estimation statistique, notamment la méthode du maximum de vraisemblance (Maximum Likelihood Estimation).

Le premier chapitre est consacré à la présentation des concepts fondamentaux de la fiabilité, avec une définition complète de l'indice de performance et une étude de ses propriétés théoriques.

Le deuxième chapitre se concentre sur l'inférence statistique de l'indice de performance à partir de données complètes, en présentant différentes méthodes d'estimation, avec un accent particulier sur la méthode du maximum de vraisemblance. Nous abordons ensuite les cas impliquant des données censurées, en introduisant d'abord la notion de censure à droite (censure à droite), avant de détailler les types de censure à droite de type 1 et de type 2, pour enfin traiter de la censure progressive de type II (censure progressivement de type II). Dans ce cadre, nous expliquons l'application de la méthode du maximum de vraisemblance à ces types de données.

Le troisième chapitre se focalise sur une étude plus spécialisée du distributif de Lindley, dans ses formes à un et deux deux paramètres. Nous dérivons les principales

caractéristiques de chaque modèle, telles que la fonction de densité, la fonction de distribution, l'espérance, la variance, ainsi que l'indice de performance, tout en appliquant la méthode du maximum de vraisemblance à partir de données censurées progressivement de type II. Ce chapitre est également accompagné d'une application en R permettant de mettre en œuvre les estimations numériques pour ces modèles.

En conclusion, un partie appliquée est consacrée à l'utilisation du distributif de Lindley pour estimer l'indice de performance à partir de données censurées progressivement de type II, illustrant ainsi le côté pratique des modèles théoriques étudiés.

Cette étude contribue à enrichir la compréhension théorique et pratique des méthodes d'estimation de l'indice de performance, en mettant en lumière l'efficacité du modèle de Lindley pour modéliser des données censurées, en particulier dans le cadre de la censure progressive.

CHAPITRE 1

NOTIONS DE BASE DE LA FIABILITÉ

Introduction

La fiabilité est une discipline essentielle dans l'analyse des systèmes et des processus, visant à quantifier la durée de vie et les performances des composants sous diverses conditions d'utilisation. Elle permet de garantir le bon fonctionnement d'un système pendant une période donnée et aide à minimiser les risques de défaillance.

L'indice de performance est un indicateur clé utilisé pour évaluer l'efficacité d'un système en tenant compte de divers paramètres tels que la durée de vie, la probabilité de défaillance et la robustesse du modèle statistique utilisé.

L'analyse de la fiabilité combinée à l'étude de l'indice de performance permet d'optimiser les décisions liées à la maintenance, à la gestion des risques et à la qualité des produits. Une approche statistique rigoureuse, basée sur des modèles de censure et des distributions spécifiques comme celle de Lindley, permet d'améliorer la précision des estimations et d'affiner les stratégies de gestion des systèmes techniques et industriels.

1.1 Notions fondamentales de fiabilité

1.1.1 Définition de la Fiabilité

La fiabilité est une mesure essentielle en ingénierie et en statistiques qui évalue la capacité d'un système ou d'un composant à fonctionner sans défaillance pendant une période donnée. Elle est particulièrement utilisée dans les domaines industriels, militaires et médicaux pour analyser et améliorer les performances des équipements.

Elle est exprimée par

$$R(t) = P(T \geq t) = 1 - F(t) \tag{1.1}$$

Où :

T est une variable aléatoire représentant le temps avant défaillance.

t est un instant donné

$R(t)$ est la probabilité que le système fonctionne correctement pendant $[0, t]$.

1.1.2 Fonction de Densité de Probabilité et Fonction de Répartition

La fonction de densité de probabilité (f.d.p), qui décrit la distribution du temps de défaillance, est donnée par :

$$f(t) = -\frac{dR(t)}{dt}. \tag{1.2}$$

La fonction de distribution cumulative (f.d.c) est définie par :

$$F(t) = P(T \leq t) = 1 - R(t) \tag{1.3}$$

elle représente la probabilité qu'un système ait déjà subi une défaillance avant le temps t.

1.1.3 Taux de Défaillance et Temps Moyen Avant Défaillance (MTTF)

$$\lambda(t) = \frac{f(t)}{R(t)}. \tag{1.4}$$

Ce taux mesure la probabilité instantanée de défaillance à un instant donné, conditionnellement à la survie jusqu'à cet instant.

Interprétation du taux de défaillance :

Si $\lambda(t)$ est constant, cela correspond à une distribution exponentielle.

Si $\lambda(t)$ augmente avec le temps, cela signifie que le système vieillit et que le risque de défaillance augmente

Si $\lambda(t)$ diminue, cela peut signifier que le système devient plus fiable avec le temps, souvent après une période de rodage.

Le temps moyen avant défaillance (MTTF) : est un indicateur clé pour évaluer la fiabilité d'un système :

$$MTTF = \int_0^\infty t f(t) dt \tag{1.5}$$

Ce paramètre est particulièrement utile pour les systèmes non réparables.

1.2 Durée de Vie et Modélisation

La durée de vie moyenne est souvent calculée à l'aide de distributions statistiques adaptées aux processus de défaillance.

1.2.1 Cas de la Distribution Exponentielle

Si X suit une loi exponentielle de paramètre λ : $X \sim Exp(\lambda)$, sa fonction de fiabilité est donnée par :

$$R(t) = e^{-\lambda t} \tag{1.6}$$

Le temps moyen avant défaillance (MTTF) est :

$$MTTF = \frac{1}{\lambda} \tag{1.7}$$

Le taux de défaillance est constant :

$$\lambda(t) = \lambda \tag{1.8}$$

1.2.2 Cas de la Distribution Gamma

Si X suit une loi Gamma de paramètres k et λ : $X \sim \gamma(k, \lambda)$, alors :
La fonction de fiabilité est donnée par :

$$R(t) = 1 - \frac{\gamma(k, \lambda t)}{\Gamma(k)} \tag{1.9}$$

où :

$\gamma(k, \lambda t)$ est la fonction gamma incomplète définie par :

$$\gamma(k, \lambda t) = \int_0^{\lambda.t} \mu^{k-1} e^{-\mu} d\mu \tag{1.10}$$

$\Gamma(k)$ est la fonction gamma complète :

$$\Gamma(k) = \int_0^{\infty} \mu^{k-1} e^{-\mu} d\mu \tag{1.11}$$

Le taux de défaillance est constant :

$$\lambda(t) = \frac{\lambda^k t^{k-1} e^{-\lambda t}}{(k-1)! R(t)} \tag{1.12}$$

Le temps moyen avant défaillance (MTTF) est :

$$MTTF = \frac{k}{\lambda} \tag{1.13}$$

1.2.3 Cas de la Distribution Weibull

Si X suit une loi de Weibull de paramètres θ et β : $X \sim Weibull(\theta, \beta)$, alors :
La fonction de fiabilité est :

$$R(t) = e^{-(\frac{t-\sigma}{\theta})^{\beta}} \tag{1.14}$$

où :

σ : durée de vie minimale.

θ : durée d'échelle.

β : paramètres de forme .

Le taux de défaillance est :

$$\lambda(t) = \frac{\beta(t-\sigma)^{\beta-1}}{\theta^{\beta}} \tag{1.15}$$

1.3 L'indice de Performance de Durée de Vie

1.3.1 Définition

L'indice de Performance de la Durée de vie (IPDV) est une mesure, utilisée pour évaluer la capacité d'un produit ou d'un système à maintenir ses fonctions de manière fiable et efficace tout ou long de sa durée de vie prévue. Il permet de quantifier la qualité, la fiabilité et la durabilité des produits, en prenant en compte divers facteurs tels que les conditions d'utilisation et la maintenance.

1.3.2 Importance

L'IPDV est un indicateur crucial dans divers domaines, notamment l'industrie, la santé et la technologie. Son importance réside dans sa capacité à fournir des informations précieuses sur la fiabilité et la longévité des produits, des systèmes ou même des individus.

On peut résumer son importance par les points suivants :

— L'IPDV permet aux fabricants d'évaluer la qualité de leurs produits et de s'assurer qu'ils répondent aux normes de fiabilité attendues. Par exemple, dans l'industrie automobile, l'IPDV peut mesurer la durée de vie moyenne des composants d'un moteur.

— Dans le secteur technologique, l'IPDV est essentiel pour évaluer la durée de vie des appareils électroniques, il permet de garantir la continuité des services et de réduire les coûts de maintenance.

— En maintenance prédictive,l'IPDV est utilisé pour anticiper les défaillances des équipements et planifier les interventions de maintenance de maniére proactive. Cela permet de réduire les temps d'arrêt, d'optimiser les Coûts de maintenance et d'améliorer la disponibilité des équipements.

— En médecine, il contribue à la recherche médicale en fournissant des données sur l'évolution des maladies et l'impact des interventions médicales.

En résumé : l'indice de performance de la durée de vie est un outil essentiel pour évaluer la fiabilité, la qualité et la longévité dans divers domaines. Son utilisation permet d'améliorer les produits, les services et la qualité de vie.

Aprés avoir examiné l'importance de l'indicateur de performance de la durée de vie et son rôle dans l'évaluation de la fiabilité, nous aborderons maintenant les méthodes utilisées pour estimer cet indicateur et explorerons les facteurs qui influencent sa précision.

1.4 Formulation de L'indice de Performance

Un indice de capacité de processus C_{LX} a été développé par [1] pour mesurer la caractéristique de qualité "plus c'est grand, mieux c'est " il ne fait aucun doute qu'une durée de vie plus longue implique une meillleur qualité.

Définition 1.1 *L'indice de performance de durée de vie C_{LX} est alors défini pour évaluer la performance de l'article comme suit : si X est la durée de vie des produits*

$$C_{LX} = \frac{\mu - L}{\sigma}, \tag{1.16}$$

où L, μ et σ sont respectivement la limite inférieur, la moyenne du processus et l'écart-type du processus.

En général, les durées de vie des différents produits ne sont pas identique. Dans la suite nous présentons l'indice de performance pour quelques distributions usuelles, telles que la distribution Exponontielle, la distribution de Weibull, la distribution Gamma, et autres.

1.4.1 Indice de Performance pour la Distribution de Weibull

Nous supposons maintenant que la variable aléatoire X des produits suit une distribution de Weibull à deux paramètres avec les paramètres de forme et d'échelle β et θ, respectivement. Par conséquent, la fonction densité de probabilité (f.d.p) et la fonction de distribution cumulative (f.d.c) de la durée de vie X des produits sont respectivement les suivantes :

$$f_X(x,\theta,\beta) = \frac{\beta}{\theta}\left(\frac{x}{\theta}\right)^{\beta-1} \exp\left\{-\left(\frac{x}{\theta}\right)^{\beta}\right\}, \quad x>0, \quad \theta>0, \quad \beta>0, \tag{1.17}$$

et

$$F_X(x,\theta,\beta) = 1 - \exp\left\{-\left(\frac{x}{\theta}\right)^{\beta}\right\}, \quad x>0, \quad \theta>0, \quad \beta>0, \tag{1.18}$$

Étant donné que la moyenne et l'écart-type du processus sont respectivement donnés par

$$\mu_X = E(X) = \theta\Gamma(1+1/\beta)$$

et

$$\sigma_X = \sqrt{\text{var}(X)} = \theta\sqrt{\Gamma(1+2/\beta) - (\Gamma(1+1/\beta))^2}$$

d'après 1.16, l'indice de performance de la durée de vie est donné par

$$C_{LX} = \frac{1}{A}\left[\Gamma\left(1+\frac{1}{\beta}\right) - \frac{L_X}{\theta}\right], \quad -\infty < C_{LX} < \frac{1}{A}\Gamma\left(1+\frac{1}{\beta}\right) \tag{1.19}$$

où

$$A = \left[\Gamma\left(1+\frac{2}{\beta}\right) - \left(\Gamma\left(1+\frac{1}{\beta}\right)\right)^2\right]^{1/2} \tag{1.20}$$

et $\Gamma(.)$ est la fonction gamma complète .

Évidemment, pour $\theta\Gamma\left(1+1/\beta\right) > L_X$, nous avons $C_{LX} > 0$, de plus, la fonction de taux de défaillance $\lambda(x,\theta,\beta)$ est facilement obtenue commme suit

$$\lambda(x,\theta,\beta) = \frac{f_X(x,\theta,\beta)}{1-F_X(x,\theta,\beta)} = \frac{\beta}{\theta}\left(\frac{x}{\theta}\right)^{\beta-1}, \quad x>0, \quad \theta>0, \quad \beta>0 \tag{1.21}$$

D'aprés 1.19 et 1.21, nous constatons que la fonction de taux de défaillance est décroissante en on θ, tan dis que l'indice de performance sur la durée de vie C_{LX}, est croissant en θ . Par conséquent, C_{LX} est approprié pour représenter la performance des produits pendant leur durée de vie .

Si $X > (<) L_X$, le produit est appelé produit conforme (non conforme). Par conséquent, le ratio des produits conforme est connu sous le nom de taux de conformité, qui est défini comme suit :[2]

$$p_r = p(X \geq L_X) = \exp\left\{-\left(\Gamma\left(1+\frac{1}{\beta}\right) - AC_{LX}\right)^{\beta}\right\}, \quad -\infty < C_{LX} < \frac{1}{A}\Gamma\left(1+\frac{1}{\beta}\right) \tag{1.22}$$

1.4.2 Indice de performance pour la distribution exponontielle

Supposons que X la durée de vie d'un tel produit puisse être modélisée par une distribution exponentielle à une seul paramètre. Par conséquent, la fonction de densité de probabilité (f.d.p) et la fonction de distribution cumulative (f.d.c) sont respectivement les suivants

$$f_X(x,\lambda) = \lambda\exp(-\lambda x), \quad x>0, \quad \lambda>0 \tag{1.23}$$

et

$$F_X(x,\lambda) = 1-\exp(-\lambda x), \quad x>0, \quad \lambda>0 \tag{1.24}$$

voir aussi (Johnson et al.[3]).

L'indice de performance de durée de vie C_{LX} dans ce cas est donné par

$$C_{LX} = \frac{\mu - L}{\sigma} = \frac{1/\lambda - L}{1/\lambda} = 1-\lambda L, \quad C_{LX} < 1 \tag{1.25}$$

Où la moyenne du processus $\mu = E(X) = 1/\lambda$, l'écart-type du processus $\sigma = \sqrt{\text{var(X)}} = 1/\lambda$, et L est la limite inférieure.

La fonction de taux de défaillance $\lambda(x)$ est définie par

$$\lambda(x) = \frac{f_X(x)}{1 - F_X(x)} = \frac{\lambda \exp(-\lambda x)}{1 - [1 - \exp(-\lambda x)]} = \lambda, \quad \lambda > 0 \tag{1.26}$$

Lorsque la durée de vie moyenne des produits $1/\lambda (> L)$, l'IPDV $C_{LX} > 0$. Les équations 1.25 et 1.26 montrent que plus la moyenne $1/\lambda$ est élevée, plus le taux de défaillance est faible et plus l'indice de performance sur la durée de vie C_{LX} est élevé.

Si la durée de vie d'un produit (ou d'un article) X dépasse la limite inférieure L ($X > L$), le produit est étiqueté comme étant un produit conforme. Dans le cas contraire, le produit est étiqueté comme étant un produit non conforme, le taux de conformité est donné par [4]

$$p_r = p(X \geq L) = \int_L^{\infty} \lambda \exp(-\lambda x) dx = \exp(-\lambda L) = \exp(C_{LX} - 1), \quad -\infty < C_{LX} < 1 \tag{1.27}$$

1.4.3 Indice de performance pour la distribution Burr XII

La distribution de Burr XII a été appliquee dans le domaine du controle de la qualite, des etudes de faibilité et de la modelisation des temps de défaillance (voir soliman [5]). La fonction de densite de probabilite (f.d.p) et la fonction de distribution cumulative (f.d.c) de la distrubution de Burr XII sont données respectivement par

$$f_X(x, c, k) = ckx^{c-1}(1 + x^c)^{-(k+1)}, \quad x > 0, \quad c > 0, \quad k > 0 \tag{1.28}$$

et

$$F_X(x, c, k) = 1 - (1 + x^c)^{-k}, \quad x > 0, \quad c > 0, \quad k > 0 \tag{1.29}$$

où $\theta = (c, k)^T$ le vecteur paramètres, c et k sont des paramètres de forme.

Si X (la durée de vie d'un produit) provient de la distribution Burr XII. Dans ce cas l'IPDV est définie par

$$C_{LX} = \frac{kB(k - 1/c, 1 + 1/c) - L}{\sqrt{kB(k - 2/c, 1 + 2/c) - k^2B^2(k - 1/c, 1 + 1/c)}} \tag{1.30}$$

$$= \frac{1}{M}\left[kB(k - 1/c, 1 + 1/c) - L\right], \quad -\infty < C_{LX} < \frac{kB(k - 1/c, 1 + 1/c)}{M}, \tag{1.31}$$

Où

$$M = \sqrt{kB(k - 2/c, 1 + 2/c) - k^2B^2(k - 1/c, 1 + 1/c)}, ck > 0, B(a, b)$$

désigne la fonction bêta, $\mu = kB(k - 1/c, 1 + 1/c)$ est la moyenne du processus, $\sigma = M$ l'écart-type du processus, et L est la limite inférieure. La fonction de taux de défaillance

$\lambda(x)$ est définie par

$$\lambda(x) = \frac{f_X(x/c,k)}{1-F_X(x/c,k)} = \frac{ckx^{c-1}}{1+x^c}, \quad x>0, \quad c>0, \quad k>0, \tag{1.32}$$

Si la durée de vie d'un produit X dépasse la limite inférieur de spécification L, le produit est considéré comme conforme. Le pourcentage des produits conformes est désigné sous l'appellation de taux de conformité identifié par

$$p_r = p(X \geq L) = \left\{1 + \left[kB(k-1/c, 1+1/c) - M.C_{LX}\right]^c\right\}^{-k}, \quad -\infty < C_{LX} < \frac{kB(k-1/c,1+1/c)}{M} \tag{1.33}$$

Où

$$M = \sqrt{kB(k-2/c,1+2/c) - k^2B^2(k-1/c,1+1/c)}, ck > 2$$

et c, k sont les paramètre de forme. [6]

1.4.4 Indice de performance pour la distribution Gamma

Supposons que la durée de vie d'un produit puisse être modélisée par une distribution gamma à deux paramètres de forme (α connu) et d'échelle (λ inconnu). Par conséquent, la fonction de densité de probabilité (f.d.p) et la fonction de distribution cumulative (f.d.c) sont respectivement les suivantes

$$f_X(x,\alpha,\lambda) = \frac{\lambda e^{-\lambda x}(\lambda x)^{\alpha-1}}{\Gamma(\alpha)}, \quad x \geq 0, \quad \alpha > 0, \quad \lambda > 0 \tag{1.34}$$

et

$$F_X(x,\alpha,\lambda) = \frac{1}{\Gamma(\alpha)} \int_0^{\lambda x} e^{-\mu}\mu^{\alpha-1} d\mu, \quad x \geq 0, \quad \alpha > 0 \tag{1.35}$$

Où $\mu = E(X) = \alpha/\lambda$ la moyenne du processus,

$$\sigma = \sqrt{\text{var(X)}} = \sqrt{\alpha/\lambda^2} = \sqrt{\alpha}/\lambda$$

l'écart-type du processus, α est donné et L la limite inférieure de spécification. (Voir également Johnson et al. [3])

L'indice de performance de durée de vie d'un produit de durée de vie X provient de la distribution gamma peut être réécrit comme suit

$$C_{LX} = \frac{\mu - L}{\alpha} = \frac{(\alpha/\lambda) - L}{\sqrt{\alpha}/\lambda} = \frac{\alpha - \lambda L}{\sqrt{\alpha}} \tag{1.36}$$

La fonction de taux de défaillance $\lambda(x)$ est définie par

$$\lambda(x) = \frac{f_X(x)}{1 - F_X(x)} = \left[\lambda e^{-\lambda x}(\lambda x)^{\alpha-1}\right] / \Gamma(\alpha)\left(\frac{1}{\Gamma(\alpha)}\int_0^{\lambda X} e^{-\mu}\mu^{\alpha-1}d\mu\right)^{-1}, \quad x \geq 0, \quad \alpha > 0, \quad \lambda > 0 \tag{1.37}$$

Lorsque la moyenne α/λ(>L), L'IPDV $C_{LX} > 0$. Les équations 1.36 et 1.37 montrent que pour $x < \alpha/\lambda$ et $\alpha > 0$, plus λ est petit, plus le taux de défaillance $\lambda(x)$ est petit et plus l'IPDV C_{LX} est grand. Par conséquent, l'IPDV représente raisonnablement et précisément les performances des nouveaux produits

En outre, si la durée de vie d'un produit X dépasse la limite inférieure L, le produit est défini comme un produit conforme. Dans ce cas le taux de conformité est défini comme suit [7]

$$p_r = P(X \geq L) = 1 - \frac{1}{\Gamma(\alpha)}\int_0^{\sqrt{\alpha}(\sqrt{\alpha}-C_{LX})} e^{-\mu}\mu^{\alpha-1}d\mu, \quad -\alpha < C_{LX} < \sqrt{\alpha} \tag{1.38}$$

CHAPITRE 2

INFÉRENCE STATISTIQUE POUR L'INDICE DE PERFORMANCE POUR DES DONNÉES COMPLÈTES

2.1 Définition générale de l'inférence statistique

L'inférence statistique est une branche fondamentale des statistiques qui permet de tirer des conclusions sur une population à partir d'un échantillon de données observées. Elle vise à généraliser les résultats obtenus à une population plus large tout en tenant compte de l'incertitude inhérente aux données échantillonnées. Contrairement à la statistique descriptive, qui se limite à résumer et organiser les données collectées, l'inférence statistique cherche à extrapoler ces résultats en utilisant des outils probabilistes et des modèles mathématiques.

L'inférence statistique repose sur le principe que, bien qu'il soit souvent impossible ou trop coûteux d'examiner l'ensemble d'une population, il est possible d'obtenir des informations pertinentes en analysant un échantillon représentatif. Cependant, étant donné que les données observées varient d'un échantillon à l'autre en raison du hasard, les résultats obtenus ne sont pas exacts mais sont associés à une certaine marge d'erreur. C'est pourquoi l'inférence statistique intègre des concepts probabilistes pour quantifier cette incertitude et garantir la fiabilité des conclusions tirées.

L'objectif principal de l'inférence statistique est donc d'estimer des paramètres inconnus (comme la moyenne, la variance ou un indice de performance) et de tester des hypothèses statistiques pour valider ou invalider des affirmations sur une population donnée. Cette approche permet d'exploiter un échantillon limité pour prendre des décisions informées, en mesurant rigoureusement l'incertitude associée aux estimations.

D'un point de vue méthodologique, l'inférence statistique se base sur des théories probabilistes, notamment la loi des grands nombres et le théorème central limite, qui justifient l'utilisation des distributions d'échantillonnage pour estimer des paramètres et construire des tests d'hypothèses fiables. Elle s'applique à un large éventail

de disciplines, notamment en sciences économiques, en biostatistique, en ingénierie, en intelligence artificielle et en physique expérimentale, où la prise de décision basée sur des données échantillonnées est essentielle.

Ainsi, l'inférence statistique constitue un outil indispensable pour analyser des phénomènes aléatoires, prédire des tendances et appuyer des décisions stratégiques en s'appuyant sur une méthodologie rigoureuse et quantifiable.

2.2 Types d'Estimation

L'inférence statistique repose sur des méthodes d'estimation qui permettent d'évaluer un paramètre inconnu d'une population à partir d'un échantillon observé. Dans le cadre de l'indice de performance, ces méthodes sont essentielles pour obtenir une approximation fiable de cet indicateur et mesurer l'incertitude associée aux résultats. L'estimation statistique peut être abordée sous plusieurs formes, notamment l'estimation ponctuelle, l'estimation par intervalle de confiance, et les tests d'hypothèses.

2.2.1 Estimation

L'estimation consiste à attribuer une valeur approximative à un paramètre inconnu d'une population en utilisant les données d'un échantillon. Puisque la vraie valeur du paramètre est généralement inaccessible, on cherche à la déduire à partir des observations disponibles tout en minimisant les erreurs possibles.

Les méthodes d'estimation visent à garantir que l'estimation obtenue est cohérente, non biaisée et efficace, ce qui signifie que l'estimateur doit se rapprocher de la valeur réelle du paramètre lorsque la taille de l'échantillon augmente.

2.2.1.1 Estimation Ponctuelle

L'estimation ponctuelle consiste à attribuer une valeur unique comme estimation d'un paramètre inconnu. Cette valeur est obtenue à l'aide d'un estimateur, qui est une fonction des données observées.

Exemples courants d'estimateurs :

— La moyenne empirique $\bar{X}$ est un estimateur ponctuel de la moyenne μ d'une population.

— La variance empirique S^2 est un estimateur ponctuel de la variance σ^2.

Un bon estimateur ponctuel doit être non biaisé (sa moyenne doit être égale à la valeur réelle du paramètre) et efficace (avoir la plus petite variance possible).

Cependant, une estimation ponctuelle ne fournit aucune information sur l'incertitude associée à l'estimation, d'où l'importance de l'estimation par intervalle.

2.2.1.2 Estimation par Intervalle de Confiance

L'estimation par intervalle de confiance permet de compléter l'estimation ponctuelle en fournissant une plage de valeurs dans laquelle le paramètre inconnu a une forte probabilité de se situer. Un intervalle de confiance est défini par deux bornes (IC_{inf} et IC_{sup}) et un niveau de confiance $(1-\alpha)$, qui représente la probabilité que l'intervalle contienne le vrai paramètre.

Exemple d'un intervalle de confiance pour la moyenne :

Si la variable X suit une loi normale de moyenne μ et de variance σ^2, alors un intervalle de confiance à un niveau de confiance $1-\alpha$ pour la moyenne μ est donné par ;

$$\left[\bar{X} - z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}, \bar{X} + z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right] \tag{2.1}$$

où $z_{1-\frac{\alpha}{2}}$ est la valeur critique de la loi normale standard.

Remarque

L'estimation par intervalle est plus informative qu'une estimation ponctuelle car elle fournit une mesure d'incertitude, indiquant à quel point l'estimation est fiable.

2.2.2 Tests d'Hypothèses

Un test d'hypothèse est une méthode statistique permettant de décider si une affirmation sur un paramètre inconnu d'une population est acceptable ou non, en se basant sur un échantillon de données.

Étapes principales d'un test d'hypothèse

— Formuler les hypothèses :
  H_0 : l'hypothèse nulle (ce qu'on suppose vrai par défaut).
  H_1 : l'hypothèse alternative (ce qu'on veut tester) .
— Choisir une statistique de test : C'est une fonction des données permettant de mesurer l'écart entre les données observées et ce qu'on attend sous H_0 .
— Déterminer la loi de la statistique sous H_0.
— Fixer le seuil de signification α : Souvent, $\alpha = 0.05$ C'est le risque d'erreur de rejeter H_0.
— Calculer la p−valeur :La p−valeur est la probabilité, sous H_0 , d'obtenir une statistique au moins aussi extrême que celle observée.
— Prendre une décision :
  Si p−valeur $\leq \alpha$ on rejette $H_0 \rightarrow$ résultat significatif.
  Si p−valeur $> \alpha$ on ne rejette pas $H_0 \rightarrow$ pas assez de preuve contre l'hypothèse.

2.3 Méthodes d'estimation

L'inférence statistique, plusieurs méthodes permettent d'estimer les paramètres inconnus d'une distribution à partir d'un échantillon de données observées. Ces mé-

thodes sont essentielles pour extraire des informations pertinentes sur une population lorsque l'observation complète de celle-ci est impossible. Chaque approche repose sur des principes mathématiques spécifiques et offre des performances variées en fonction du contexte et des hypothèses sous-jacentes.

Parmi les méthodes les plus courantes, on distingue :

1.Méthode des Moments (MOM)

2.Méthode du Maximum de Vraisemblance (EMV)

3.Estimation Bayésienne

2.3.1 Méthode des Moments (MOM)

Dans certains cas, il est facile de décider comment estimer un paramètre, et souvent l'intuition seule peut nous conduire à de très bons estimateurs. Par exemple, l'estimation d'un paramètre à l'aide de son analogue dans l'échantillon est généralement raisonnable. En particulier, la moyenne de l'échantillon est une bonne estimation de la moyenne de la population. Dans les modèles plus compliqués, qui se présentent souvent dans la pratique, nous avons besoin d'une méthode plus méthodique pour estimer les paramètres. Dans cette section, nous détaillons quatre méthodes pour trouver des estimateurs.

La méthode des moments est peut-être la plus ancienne méthode de recherche d'estimateurs ponctuels, remontant au moins à Karl Pearson à la fin des années 1800. Elle a l'avantage d'être très simple à utiliser et permet presque toujours d'obtenir une sorte d'estimation. Malheureusement, dans de nombreux cas, cette méthode produit des estimateurs qui peuvent être améliorés. Cependant, elle constitue un bon point de départ lorsque d'autres méthodes s'avèrent impossibles à mettre en œuvre.

Principe de la méthode

L'estimation par la méthode des moments c'est la méthode la plus naturelle. L'idee de base est d'estimer une esperance mathematique par une moyenne empirique, une variance par une variance empirique,etc...

On suppose que l'echantillon $X_1,...,X_n$ i.i.d, d'une variable aleatoire X et la loi de probabilite de X depend d'un parametre θ, Si le parametre a estimer est l'esperance de la loi des $X_i (i = 1,...,n)$, alors on peut l'estimer par la moyenne empirique de l'echantillon. Autrement dit, si $\theta = E[X]$, alors l'estimateur deθ par la methode des moment est :

$$\hat{\theta} = \bar{X_n} = \frac{1}{n} \sum_{i=1}^{n} X_i \tag{2.2}$$

Plus generalement, pour$\theta \in R$, si $E(X) = \varphi(\theta)$, où φ est une fonction inversible, alors

l'estimateur de θ par la methode des moments est :

$$\hat{\theta}_n = \varphi^{-1}(\bar{X_n}) \tag{2.3}$$

De la meme maniere, on estime la variance de la loi des X_i, par la variance empirique de l'echantillon :

$$S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 \tag{2.4}$$

Ce principe peut naturellement se généraliser aux moments de tout ordre, centres ou non centres :

$$E[(X - E(X))^k] \quad et \quad E(X^K), k \geq 1. \tag{2.5}$$

La méthode des moments (MOM) consiste à estimer les paramètres inconnus d'une distribution en égalant les moments empiriques aux moments théoriques.

2.3.2 Méthode du Maximum de Vraisemblance (EMV)

Parmi les approches les plus puissantes et les plus utilisées en inférence statistique, la méthode du maximum de vraisemblance (ou EMV, de l'anglais Maximum Likelihood Estimation) occupe une place centrale. Elle repose sur un principe intuitif : estimer les paramètres d'un modèle statistique de manière à ce que la probabilité d'observer les données disponibles soit la plus élevée possible. Autrement dit, elle cherche les valeurs des paramètres qui rendent les données observées les plus "plausibles".

Cette méthode consiste à construire une fonction de vraisemblance, qui mesure la probabilité d'observer l'échantillon en fonction des paramètres du modèle, puis à maximiser cette fonction pour obtenir les estimateurs des paramètres inconnus. Lorsque les paramètres maximisent cette vraisemblance, on considère qu'ils décrivent le mieux les données observées selon le modèle choisi.

La méthode du maximum de vraisemblance est particulièrement appréciée pour sa cohérence, sa précision asymptotique, et sa large applicabilité, aussi bien pour les distributions discrètes que continues. Elle constitue donc un outil incontournable dans les applications statistiques, notamment lorsqu'on cherche à évaluer la performance d'un système à partir d'un modèle de probabilité

Définition

La méthode du maximum de vraisemblance, proposée initialement par Ronald Fisher en 1922, est l'une des méthodes les plus utilisées pour estimer les paramètres inconnus d'un modèle statistique. Elle repose sur le principe d'adopter les valeurs des paramètres qui rendent les données observées les plus probables, selon le modèle

choisi.

Principe de la méthode

La méthode du maximum de vraisemblance est, de loin, la technique la plus populaire pour dériver des estimateurs. Rappelons que si $X_1,...,X_n$, est un échantillon i.i.d d'une population avec fdp $f(x|\theta_1,,...,\theta_k)$, la fonction de vraisemblance est définie par :

$$L(\theta|x) = L(\theta_1,...,\theta_k|x_1,.....,x_n) = \prod_{i=1}^{n} f(x_i|\theta_1,...,\theta_k). \tag{2.6}$$

Dans la pratique, on travaille souvent avec la log-vraisemblance :

$$\ell(\theta|x) = \log L(\theta|x) \tag{2.7}$$

On calcule la dérivée (ou les dérivées partielles si θ est un vecteur) de la fonction de log-vraisemblance par rapport aux paramètres.

$$\frac{\partial}{\partial\theta_i}\ell(\theta|x) = 0, \quad i = 1,...,k \tag{2.8}$$

En résolvant ce système 2.8, on obtient les estimateurs du maximum de vraisemblance $\hat{\theta_{EMV}}$ des paramètres inconnus.

L'orsque la résolution analytique est difficile, des méthodes numériques peuvent être utilisées pour approcher la solution.

2.3.3 Estimation Bayésienne

L'approche bayésienne de la statistique est fondamentalement différente de l'approche classique que nous avons adoptée. Néanmoins, certains aspects de l'approche bayésienne peuvent être très utiles pour d'autres approches statistiques. Avant d'aborder les méthodes permettant de trouver des estimateurs de Bayes, nous commencerons par discuter de l'approche bayésienne des statistiques.

Dans l'approche classique, le paramètre θ est considéré comme une quantité inconnue, mais fixée. Un échantillon aléatoire $X_1,...,X_n$ est tiré d'une population indexée par θ et, sur la base des valeurs observées dans l'échantillon, on obtient une connaissance de la valeur de θ. Dans l'approche bayésienne, θ est considérée comme une quantité dont la variation peut être décrite par une distribution de probabilité (appelée distribution préalable). Il s'agit d'une distribution subjective, basée sur la croyance de l'expérimentateur, qui est formulée avant que les données ne soient vues (d'où le nom de distribution préalable). Un échantillon est ensuite prélevé dans une population indexée par et la distribution préalable est mise à jour à l'aide de l'information sur l'échantillon. La distribution a priori mise à jour est appelée distribution a posteriori. Cette mise à jour est effectuée à l'aide de la règle de Bayes , d'où le nom de statistique bayésienne.

Si nous désignons la distribution préalable par $\pi(\theta)$ et la distribution d'échantillonnage par $f(x|\theta)$, la distribution postérieure, la distribution conditionnelle de θ étant donné l'échantillon x, est la suivant :

$$\pi(\theta|x) = \frac{f(x|\theta)\pi(\theta)}{m(x)}, \quad avec \quad (f(x|\theta)\pi(\theta) = f(x,\theta). \tag{2.9}$$

Où $m(x)$ est la distribution marginale de X, c'est-à-dire :

$$m(x) = \int f(x|\theta)\pi(\theta)d\theta. \tag{2.10}$$

Après avoir exploré les différentes méthodes d'estimation, il est essentiel de se concentrer sur l'une des approches les plus couramment utilisées en statistique, à savoir la méthode du maximum de vraisemblance, qui permet d'estimer de manière optimale les paramètres des distributions. Dans cette section, nous appliquerons cette méthode à certaines distributions spécifiques.

2.4 Censure

La censure se produit dans les ensembles de données sur la durée de vie lorsque seule une limite supérieure ou inférieure de la durée de vie est connue. La censure est fréquente dans les ensembles de données sur la durée de vie parce qu'il est souvent impossible ou peu pratique d'observer la durée de vie de tous les éléments testés. Un ensemble de données pour lequel tous les temps de défaillance sont connus est appelé ensemble de données complet. La figure 2.1 illustre un ensemble de données complet de $n = 5$ éléments mis à l'essai simultanément à l'instant $t = 0$, où les $\times$ indiquent les temps de défaillance. Considérons les deux extrémités de chacun des segments horizontaux. Il est essentiel de fournir une définition sans ambiguïté de l'origine temporelle (par exemple, le moment où un produit est acheté ou le moment où un cancer est diagnostiqué). De même, la défaillance doit être définie sans ambiguïté. Il est plus facile de la définir pour une ampoule ou un fusible que pour un roulement à billes ou une chaussette. En dehors d'un contexte de fiabilité, un ensemble de données sur les durées de vie est souvent appelé de manière générique données temps-événement, correspondant au temps écoulé entre l'origine temporelle et l'événement d'intérêt. Une observation censurée se produit lorsque seule une limite est connue pour le moment de la défaillance. Si un ensemble de données contient une ou plusieurs observations censurées, il s'agit d'un ensemble de données censurées.

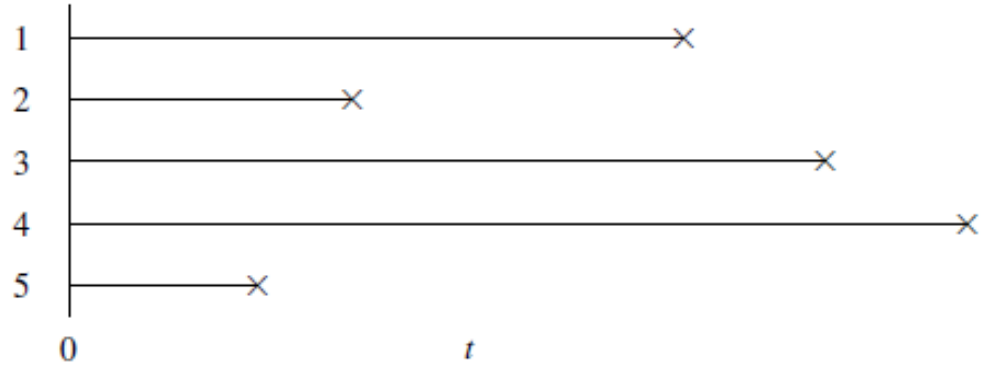

FIGURE 2.1 – Ensemble de données complet avec n = 5.

2.5 Censure à droite

Le type de censure le plus courant est connu sous le nom de censure à droite. Dans un ensemble de données censuré à droite, un ou plusieurs éléments n'ont qu'une limite inférieure connue pour leur durée de vie. Le terme « taille de l'échantillon » est désormais vague. À partir de maintenant, nous utilisons n pour désigner le nombre d'éléments testés et r pour désigner le nombre de défaillances observées. Dans une situation de test de durée de vie industriel, par exemple, $n = 12$ téléphones portables sont soumis à un test de durée de vie continu et rigoureux le 1er janvier, et $r = 3$ des téléphones portables sont tombés en panne le 31 décembre. Ces téléphones portables défectueux sont mis au rebut dès qu'ils tombent en panne. Les $n - r = 12 - 3 = 9$ téléphones cellulaires restants qui fonctionnent encore le 31 décembre ont une durée de vie supérieure à 365 jours et sont donc des observations censurées à droite. La censure à droite ne se limite pas aux applications de fiabilité. Dans une étude médicale où T est le temps de survie après le diagnostic d'un type particulier de cancer, par exemple, un patient peut soit
(a) être encore en vie à la fin de l'étude, soit
(b) mourir d'une cause autre que le type particulier de cancer, ce qui constitue une observation censurée à droite, soit
(c) perdre tout contact avec l'étude (par exemple, s'il quitte la ville), ce qui constitue une observation censurée à droite.
Dans le cadre de la censure à droite, on distingue généralement deux types principaux de censure : la censure à droite de type I et la censure à droite de type II, chacun ayant ses propres caractéristiques et applications en analyse statistique.

2.5.1 Censure à droite de type II

La censure de type II ou statistique d'ordre. Comme le montre la figure 2.2, cela correspond à l'arrêt d'une étude sur l'un des échecs ordonnés. Le diagramme correspond à un ensemble de $n = 5$ éléments placés simultanément sur un test à l'instant

$t = 0$. Le test est terminé lorsque $r = 3$ échecs sont observés. Le temps avance de gauche à droite dans la figure 2.2 et la défaillance du premier élément (correspondant à la troisième défaillance ordonnée observée) met fin au test. Les durées de vie des troisième et quatrième éléments sont censurées à droite. Les temps de défaillance observés sont indiqués par un × et les temps de censure à droite sont indiqués par un ∘. Dans la censure de type II, le temps nécessaire pour terminer le test est aléatoire.

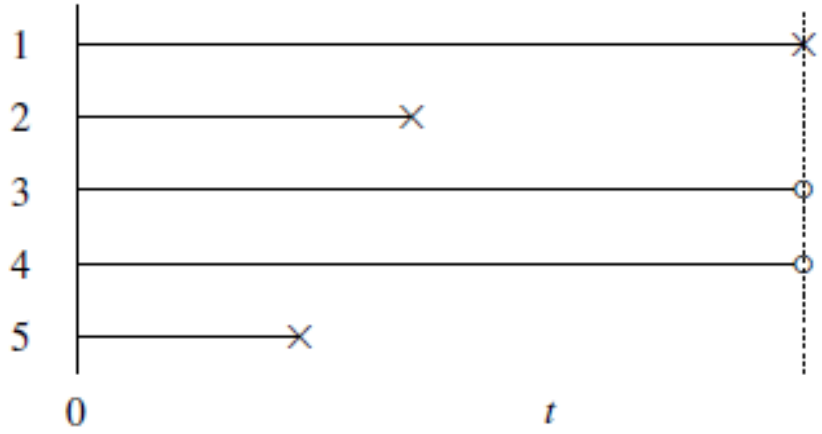

FIGURE 2.2 – Ensemble de données de type II censurées à droite avec n = 5 et r = 3.

2.5.2 Censure à droite de type I

Le deuxième cas particulier est le type I ou la censure temporelle. Comme le montre la figure 2.3, cela correspond à l'arrêt de l'étude à un moment donné. Le diagramme montre un ensemble de $n = 5$ items placés simultanément sur un test à $t = 0$ qui est terminé au moment indiqué par la ligne verticale en pointillés. Pour la réalisation illustrée à la figure 2.3, il y a $r = 4$ échecs observés. Dans la censure de type I, le nombre d'échecs r est aléatoire.

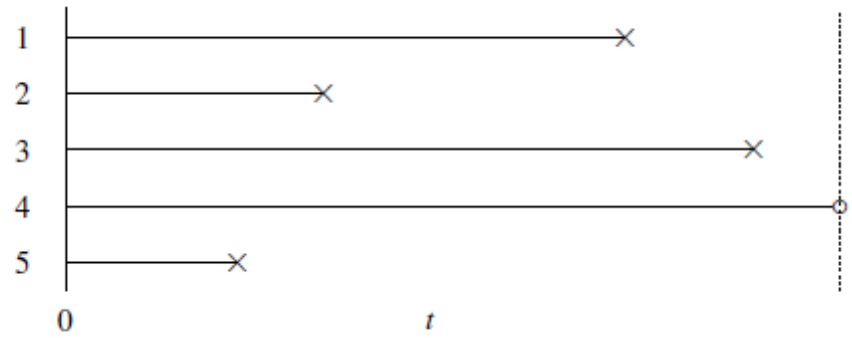

FIGURE 2.3 – Ensemble de données censurées à droite de type II avec $n = 5$ et $r = 4$.

Bien qu'il existe d'autres types de censure, tels que la censure à gauche et la censure par intervalle, ce chapitre se concentrera sur la censure à droite, car il s'agit du type de censure le plus courant. Dans le cas de la censure à droite, le ratio r/n est la fraction des éléments dont on observe l'échec. Lorsque r/n est proche de un, l'ensemble de données est qualifié de légèrement censuré , lorsque r/n est proche de zéro, l'ensemble de données est qualifié de fortement censuré. Dans le domaine de la fiabilité, de nombreux ensembles de données sont fortement censurés parce que les éléments ont une longue durée de vie. Dans le domaine biomédical, certains cancers ont de longues périodes de rémission, ce qui se traduit par des ensembles de données fortement censurés.

Il existe de nombreux scénarios dans les essais de durée de vie et les expériences de fiabilité dans lesquels des unités sont perdues ou retirées de l'essai avant la défaillance. La perte peut se produire involontairement ou être prévue dans l'étude. Dans de nombreuses situations, cependant, le retrait des unités avant la défaillance est planifié à l'avance afin d'économiser le temps et l'argent associés aux essais. Les deux schémas de censure les plus courants sont appelés censure de type I et censure de type II ; mais l'utilisation de ces types de censure ne permet pas de retirer des unités à des points autres que le point final de l'expérience. Des schémas de censure plus généraux sont donc nécessaires.

2.5.3 Censure progressive de type II

Un test de durée de vie de type II avec censure progressive est réalisé de la manière suivante. Nous concevons une expérience dans laquelle n unités sont placées dans un test de durée de vie commençant au même moment, et le test peut être interrompu au moment de tout échec. En outre, une ou plusieurs unités survivantes peuvent être retirées du test (censurées) au moment de chaque échec survenant avant la fin de l'expérience ; en d'autres termes, au moment du premier échec $X_{1,n}, r_1$, des unités sont retirées au hasard des $(n-1)$ unités survivantes restantes.

Au deuxième échec $X_{2,n}, r_2$, des unités parmi les $n-2-r_1$ unités restantes sont retirées au hasard. Le test se poursuit jusqu'à la m^{ime} défaillance. À ce moment, toutes les unités $r_m = n-m-r_1-r_2-...-r_{m-1}$ restantes sont retirées. Dans ce système de censure, r_i et m sont fixés à l'avance. Les valeurs ordonnées obtenues à la suite de ce type de censure sont appelées à juste titre statistiques d'ordre progressivement censurées de type II. Il convient de noter que si $r_1 = r_2 = ... = r_{m-1} = 0$, de sorte que $r_m = n-m$, ce schéma se réduit au schéma conventionnel de censure à droite en une étape de type II. Notez également que si $r_1 = r_2 = ... = r_m = 0$, de sorte que $m = n$, le schéma de censure progressive de type II se réduit au cas où il n'y a pas de censure (cas de l'échantillon complet)

Fonction de vraisemblance sur échantillon progressivement censuré

Supposons que $x = (X_{1,m,n}, X_{2,m,n}, ..., X_{m,m,n})$ soit un échantillon progressivement censuré de type II d'un test de durée de vie portant sur n éléments et que $r_1, r_2, ..., r_m$ désignent les nombres correspondants d'unités retirées (retirées) du test. La fonction de vraisemblance basée sur l'échantillon progressivement censuré de type II ([8]) est donnée par

$$L(\theta) = C \prod_{i=1}^{m} f(x_i) \Big[R(x_i) \Big]^{r_i} \tag{2.11}$$

où $C = n(n-r_1-1)(n-r_1-r_2-2)...(n-\sum_{j=1}^{m-1}(r_j+1))$: constante combinatoire.

$f(x_i)$: densité de probabilité .

$R(x_i) = 1 - F(x_i)$: fonction de survie.

θ : paramètre(s) du modèle .

Fonction log-vraisemblance sur échantillon progressivement censuré

$$\ell(\theta) = \sum_{i=1}^{m} \ln\{f(x_i)\} + \sum_{i=1}^{m} r_i \ln\{R(x_i)\} \tag{2.12}$$

2.6 Méthode de Newton-Raphson pour Données Progressivement Censurées

La méthode de Newton-Raphson est un algorithme itératif puissant pour résoudre des équations non linéaires, largement utilisé en statistique pour l'estimation par maximum de vraisemblance (EMV). Elle est particulièrement utile pour estimer les paramètres de distributions complexes, surtout lorsque les solutions analytiques sont introuvables.

2.6.1 Adaptation de Newton-Raphson

Calcul du gradien $\nabla\ell(\theta)$:

$$\frac{\partial\ell(\theta)}{\partial\theta} = \sum_{i=1}^{m} \frac{\partial \ln\{f(x_i)\}}{\partial\theta} + \sum_{i=1}^{m} r_i \frac{\partial \ln\{R(x_i)\}}{\partial\theta} \tag{2.13}$$

Calcul de la Hessienne $\nabla^2\ell(\theta)$:

$$\frac{\partial^2\ell(\theta)}{\partial\theta^2} = \sum_{i=1}^{m} \frac{\partial^2 \ln\{f(x_i)\}}{\partial\theta^2} + \sum_{i=1}^{m} r_i \frac{\partial^2 \ln\{R(x_i)\}}{\partial\theta^2} \tag{2.14}$$

Initialisation : choisir θ_0 (ex : $\theta_0 = 1.5$)
Itération de Newton-Raphson : pour $k = 1, 2, ...$

$$\theta^{(k+1)} = \theta^{(k)} - \left[\nabla^2\ell(\theta^{(k)})\right]^{-1} \nabla\ell(\theta^{(k)}) \tag{2.15}$$

Critère d'Arrêt : $\left|\theta_{k+1} - \theta_k\right| < \epsilon$ (Ex : $\epsilon = 10^{-1}$)

2.7 Estimation du maximum de vraisemblance pour certaines distributions

2.7.1 Distribution exponentielle

Dans les expériences de test de durée de vie des produits, l'expérimentateur n'est pas toujours en mesure d'observer la durée de vie de tous les produits testés en raison d'un manque de temps et/ou d'autres restrictions (telles que le manque de fonds, le manque de ressources matérielles, les difficultés mécaniques ou expérimentales, etc. La censure progressive est très utile dans de nombreuses situations pratiques où il existe des contraintes budgétaires ou une demande d'essais rapides.

Soit X la durée de vie d'un tel produit et X a une distribution exponentielle à un paramètre avec la (f.d.p) comme 1.23. Considérons que $X_{1,n} \leq X_{2,n} \leq ... \leq X_{m,n}$ est l'échantillon censuré à droite progressivement de type II correspondant, avec un schéma de censure $r = (r_1, r_2, ..., r_m)$. La fonction de vraisemblance est donnée par

$$\begin{aligned} L(\lambda, x) &= C \prod_{i=1}^{m} f_X(x_i)[1 - F_X(x_i)]^{r_i} \\ &= C\lambda^m exp\left\{ -\lambda \sum_{i=1}^{m} (1 + r_i)x_i \right\}. \end{aligned} \tag{2.16}$$

Où $f_X(x)$ et $F_X(x)$ sont les (f.d.p) et (f.d.c) de X comme 1.23 et 1.24, respectivement et

$C = n(n-r_1-1).....(n-r_1-r_2-...-r_{m-1}-r_{m+1})$

On prend le logarithme de la Vraisemblance on obtient :

$$\ell(\lambda,x) = logL(\lambda,x) = logC + mlog\lambda - \lambda\sum_{i=1}^{m}(1+r_i)x_i \tag{2.17}$$

On dérive par rapport à λ on obtient

$$\frac{\partial\ell}{\partial\lambda} = \frac{m}{\lambda} - \sum_{i=1}^{m}(1+r_i)x_i$$

L'estimateur du maximum de vraisemblance de paramètre λ est la solution de l'équation

$$\ell(\lambda,x) = 0$$

$$\frac{m}{\lambda} - \sum_{i=1}^{m}(1+r_i)x_i = 0$$

Par conséquent, l'estimateur du maximum de vraisemblance (EMV) $\hat{\lambda}$ de λ est donnée par

$$\hat{\lambda} = \frac{m}{\sum_{i=1}^{m}(1+r_i)X_{i,n}} \tag{2.18}$$

Où r_i et m correspondent à la définition ci-dessus (voir également Blakrishnan et Aggarwala [8]. En utilisant l'invariance de la (EMV)(voir Zehna [9]), la (EMV) de C_{LX} peut être écrite comme suit

$$\begin{aligned}\hat{C}_{LX} &= 1 - \hat{\lambda}L \\ &= 1 - \frac{mL}{\sum_{i=1}^{m}(1+r_i)X_{i,n}}.\end{aligned} \tag{2.19}$$

2.7.2 Distribution de Burr XII

2.7.2.1 EMV de l'indice de performance de la durée de vie

Supposons que $X_{1:m:n}, X_{2:m:n}, ..., X_{m:m:n}$ soient les échantillons progressifs de type II censurés à droite d'un test de durée de vie de n produits (ou articles) dont les durées de vie suivent la distribution de Burr XII avec la f.d.p de X donnée par l'équation 1.28, et que $r = (r_1, r_2, ..., r_m)$ représente les nombres correspondants de produits (ou articles) retirés du test de durée de vie. Ensuite, la f.d.p conjointe de toutes les m statistiques d'ordre progressivement censurées à droite de type II (voir Soliman [10]) est donnée par

$$C\prod_{i=1}^{m}\{f_X(x_{i:m:n},c,k)[1-F_X(x_{i:m:n},c,k)]^{r_i}\} \tag{2.20}$$

où $C = n(n-r_1-1)...(n-r_1-r_2-...-r_{m-1}-m+1)$, $f_X(x_{i:m:n},c,k)$ et $F_X(x_{i:m:n},c,k)$ sont respectivement la f.d.p et la f.d.c de X donnée par les équations 1.28 et 1.29. En substituant les équations 1.28 et 1.29 à l'équation 2.20, on obtient la fonction de vraisemblance de $X_{1:m:n}, X_{2:m:n}, ..., X_{m:m:n}$ est donné comme suit :

$$\begin{aligned} L(c,k,x) &= C\prod_{i=1}^{m}\left\{f_X(x_{i:m:n},c,k)\left[1-F_X(x_{i:m:n},c,k)\right]^{r_i}\right\} \\ &= C\prod_{i=1}^{m}\left\{\left[ckx_{i:m:n}^{c-1}(1+x_{i:m:n}^{c})^{-(k+1)}\right]\left[1-(1-(1+x_{i:m:n}^{c})^{-k})\right]^{r_i}\right\} \\ &= C(ck)^m\prod_{i=1}^{m}x_{i:m:n}^{c-1}(1+x_{i,m,n}^{c})^{(-k(r_i+1)+1)}, \end{aligned} \tag{2.21}$$

Le logarithme naturel de la fonction de vraisemblance peut alors s'écrire comme suit :

$$\ell(c,k,x) = \log L(c,k,x) = m\ln(ck)+(c-1)\sum_{i=1}^{m}\ln(x_{i:m:n})-\sum_{i=1}^{m}(k(r_i+1)+1)\ln(1+x_{i:m:n}^{c}) \tag{2.22}$$

L'(EMV) $\hat{c}$, $\hat{k}$ de c et k peut être obtenu en fixant les premières dérivées partielles de l'équation 2.22 à zéro par rapport à c et k. Ces équations de vraisemblance pour les paramètres c et k sont données par

$$\frac{\partial\ell(c,k,x)}{\partial c} = \frac{m}{c}+\sum_{i=1}^{m}\ln(x_{i:m:n})-\sum_{i=1}^{m}(k(r_i+1)+1)\frac{x_{i:m:n}^{c}\ln(x_{i:m:n})}{(1+x_{i:m:n}^{c})} = 0 \tag{2.23}$$

et

$$\frac{\partial\ell(c,k,x)}{\partial k} = \frac{m}{k}-\sum_{i=1}^{m}(r_i+1)\ln(1+x_{i:m:n}^{c}) = 0 \tag{2.24}$$

L'équation 2.24 donne le EMV de k comme étant donné par

$$\hat{k} = \frac{m}{\sum_{i=1}^{m}(r_i+1)\ln(1+x_{i:m:n}^{\hat{c}})} \tag{2.25}$$

En utilisant la EMV de K donnée par l'équation 2.25, léquation 2.23 peut être réduite à

$$\begin{aligned} &\frac{m}{\hat{c}}+\sum_{i=1}^{m}\ln(x_{i:m:n})-\left[\frac{m}{\sum_{i=1}^{m}(1+r_i)\ln(1+x_{i:m:n}^{\hat{c}})}\right] \\ &\times\sum_{i=1}^{m}(r_i+1)\frac{x_{i:m:n}^{\hat{c}}\ln(x_{i:m:n})}{(1+x_{i:m:n}^{\hat{c}})}-\sum_{i=1}^{m}\frac{x_{i:m:n}^{\hat{c}}\ln(x_{i:m:n})}{(1+x_{i:m:n})} = 0 \end{aligned} \tag{2.26}$$

L'equation est non linéaire donc il n'y a pas de solution analytique. On utilise le package maxLik du R qui a une capacité pour résoudre l'équation non linéaire . Donc l'EMV de C_{LX}, peut s'écrire comme suit[6]

$$\hat{C}_{LX} = \frac{\hat{k}B(\hat{k}-1/\hat{c},1+1/\hat{c})-L}{\sqrt{\hat{k}B(\hat{k}-2/\hat{c},1+2/\hat{c})-\hat{k}^2B^2(\hat{k}-1/\hat{c},1+1/\hat{c})}} \tag{2.27}$$

2.7.3 Distribution de Weibull

2.7.3.1 EMV de l'indice de performance de la durée de vie

Soit $X_{1:m:n:k} < X_{2:m:n:k} < ... < X_{m:m:n:k}$ l'échantillon progressif censuré au premier échec d'une population continue avec f.d.p et f.d.c donnée par les équations 1.17 et 1.18, respectivement, où λ est un vecteur de paramètres. Suivant [11], la fonction de vraisemblance associée aux données observées $x = (x_1, x_2, ..., x_m)$ présente comme suit

$$L(\lambda, x) = Ck^m \prod_{i=1}^{m} f_X(x_i, \lambda)[1 - F_X(x_i, \lambda)]^{k(r_i+1)-1}, \tag{2.28}$$

où $0 < x_1 < ... < x_m < \infty$ et

$$C = n(n - r_1 - 1)(n - r_1 - r_2 - 2)...\left(n - \sum_{i=1}^{m-1} r_i - m + 1\right).$$

En substituant les équations 1.17 et 1.18 à 2.28, la fonction de vraisemblance est la suivante

$$L(\theta, \beta, x) = C\left(\frac{k\beta}{\theta^\beta}\right)^m \prod_{i=1}^{m} x_i^{\beta-1} \exp\left(-k\sum_{i=1}^{m}(r_i+1)\left(\frac{x_i}{\theta}\right)^\beta\right) \tag{2.29}$$

On prend le logarithme de la Vraisemblance on obtient :

$$\ell(\theta,\beta,x) = logL(\theta,\beta,x) = logC + mlog(k\beta) - m\beta log(\theta) + (\beta-1)\sum_{i=1}^{m} log(x_i) - k\sum_{i=1}^{m}(r_i+1)\left(\frac{x_i}{\theta}\right)^\beta$$

On dérive par rapport à θ on obtient

$$\frac{\partial \ell}{\partial \theta} = -\frac{m\beta}{\theta} + k\beta\sum_{i=1}^{m}(r_i+1)\cdot\frac{x_i^\beta}{\theta^{\beta+1}}$$

On dérive par rapport à β on obtient

$$\frac{\partial \ell}{\partial \beta} = \frac{m}{\beta} - m\log\theta + \sum_{i=1}^{m}\log x_i - k\sum_{i=1}^{m}(r_i+1)\left(\frac{x_i}{\theta}\right)^\beta \log\left(\frac{x_i}{\theta}\right)$$

L'estimateur du maximum de vraisemblance des paramètres θ et β est la solution desl'équations

$$\frac{\partial \ell(\theta,\beta,x)}{\partial \theta} = 0 \quad et \quad \frac{\partial \ell(\theta,\beta,x)}{\partial \beta} = 0$$

Les équations sont non linéaires, il n'existe donc pas de solution analytique. Nous utilisons le package maxLik de R, qui permet de résoudre ce type d'équations par des méthodes numériques.

Donc l'EMV de C_{LX} [2], peut s'écrire comme suit

$$\hat{C_{LX}} = \frac{1}{A}\left[\Gamma\left(1+\frac{1}{\beta}\right) - \frac{L_X}{\hat{\theta}}\right] = \frac{1}{A}\left[\Gamma\left(1+\frac{1}{\beta}\right) - L_X\left(\frac{m}{W}\right)^{\frac{1}{\beta}}\right] \tag{2.30}$$

où $W = k\sum_{i=1}^{m}(r_i+1)x_{i:m:n:k}^{\beta}$ et $A = \left[\Gamma\left(1+\frac{2}{\beta}\right) - \left(\Gamma\left(1+\frac{1}{\beta}\right)\right)^2\right]^{1/2}$.

CHAPITRE 3

INFÉRENCE STATISTIQUE POUR L'INDICE DE PERFORMANCE POUR DES DONNÉES PROGRESSIVEMENT CENSURÉES DE TYPE II

3.1 Présentation de la Distribution Lindley

Les distributions de probabilité jouent un rôle fondamental en statistique, permettant de modéliser des phénomènes aléatoires dans divers domaines tels que l'ingénierie, la finance, la médecine et les sciences sociales. Parmi ces distributions, la loi exponentielle a longtemps été utilisée pour modéliser des temps de vie ou des durées entre événements. Cependant, sa rigidité (notamment son incapacité à capturer des phénomènes de surdispersion ou d'asymétrie marquée) a motivé le développement d'alternatives plus flexibles.

C'est dans ce contexte que la distribution Lindley a émergé comme une solution élégante, combinant simplicité mathématique et capacité à s'adapter à des données réelles plus complexes. Proposée initialement sous une forme à un seul paramètre, elle a connu plusieurs généralisations pour améliorer son pouvoir de modélisation, aboutissant aujourd'hui à des versions à deux, trois paramètres, voire davantage.

L'objectif de ce mémoire est de retracer l'évolution historique de la distribution Lindley, d'en présenter les propriétés mathématiques clés, et d'illustrer son utilité pratique à travers des applications concrètes.

C'est en 1958, dans un article intitulé "Fiducial Distributions and Bayes' Theorem", publié dans le Journal of the Royal Statistical Society, Series B, que Lindley a proposé la distribution qui porte aujourd'hui son nom.

Dans ce travail, il s'intéressait à des situations d'attente et d'incertitude dans un cadre bayésien, et a proposé une loi continue dérivée comme combinaison de la loi exponentielle et d'une loi gamma, pour mieux modéliser certaines variables de durée.

Pourquoi cette loi est importante ?

— La loi de Lindley a l'avantage d'être simple mathématiquement mais plus flexible que la loi exponentielle.
— Elle est adaptée à la modélisation des durées de vie, des temps d'attente, et des phénomènes de fiabilité.
— Elle a été étendue depuis dans de nombreuses variantes, notamment :
— Lindley généralisée à 2 ou 3 paramètres,
— Lindley modifiée,
— Lindley exponentielle, etc.

3.2 Distribution Lindley à 1 Paramètre θ

3.2.1 Fonction de Densité de probabilité (fdp)

La distribution Lindley est introduite par Lindley en 1958 comme une distribution à un seul paramétre $\theta > 0$. Sa fonction de densité de probabilité (fdp) est donnée par :

$$f(x,\theta) = \frac{\theta^2}{1+\theta}(1+x).e^{-\theta.x}, \quad \theta > 0 \tag{3.1}$$

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 $\frac{1}{1+\theta}$ [12].

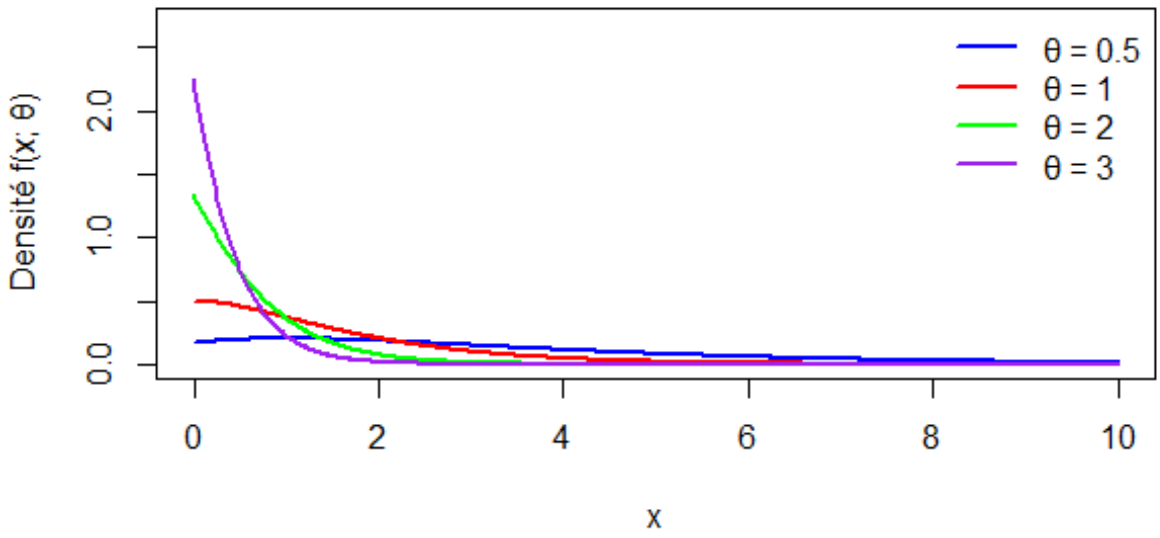

FIGURE 3.1 – Présentation graphique de la fonction de densité (Lindley 1 paramètre) pour quelques valeurs de θ

3.2.2 Fonction de répartition (fdc)

La fonction de répartition cumulative correspondante est donnée par :

$$F(x,\theta) = 1 - \left(1 + \frac{\theta x}{\theta + 1}\right) e^{-\theta.x}, \quad x > 0, \quad \theta > 0 \tag{3.2}$$

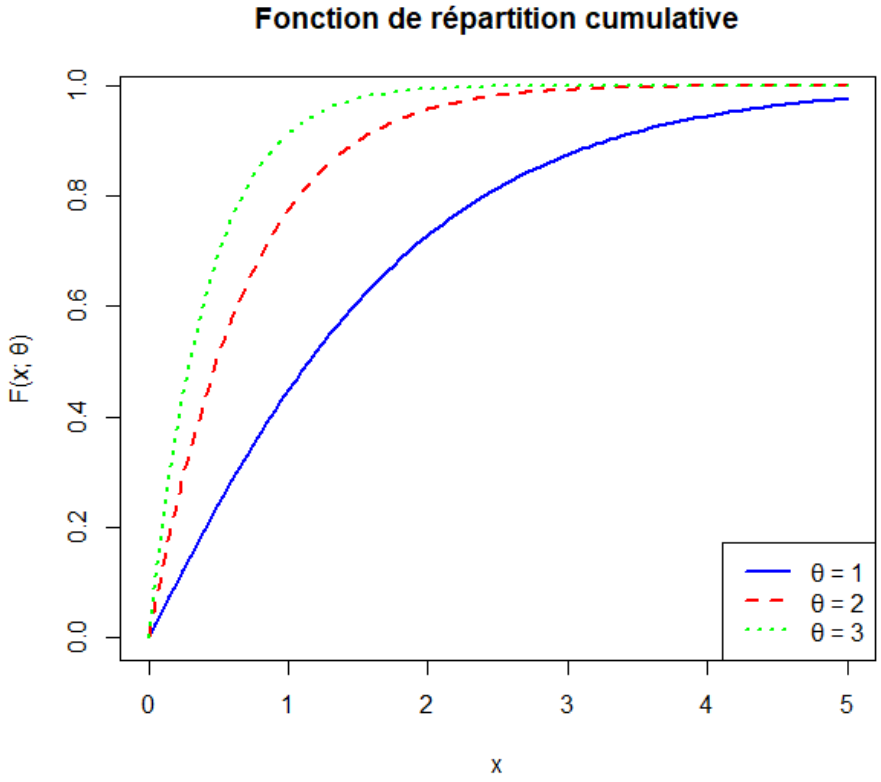

FIGURE 3.2 – Présentation graphique de la fonction de répartition (Lindley 1 paramètre) pour quelques valeurs de θ

3.2.3 Fonction de fiabilité (Survie)

La fonction de Fabilité correspondante est définie comme suit :

$$R(x,\theta) = 1 - F(x,\theta) = \left(1 + \frac{\theta x}{\theta + 1}\right) e^{-\theta.x}, \quad x > 0, \quad \theta > 0 \tag{3.3}$$

3.2.4 Moyenne (Espérance)

$$E[X] = \mu = \frac{\theta + 2}{\theta(\theta + 1)}, \quad \theta > 0 \tag{3.4}$$

3.2.5 Variance

[12]

$$Var(X) = \sigma^2 = \frac{\theta^2 + 4\theta + 2}{\theta^2(\theta + 1)^2}, \quad \theta > 0 \tag{3.5}$$

3.2.6 Indice de performance

$$C_{LX} = \frac{\mu - L}{\sigma} = \frac{(2+\theta) - L\theta(1+\theta)}{\sqrt{\theta^2 + 4\theta + 2}} \tag{3.6}$$

où L est la limite inférieur .

3.2.7 Estimation par la méthode du maximum de vraisemblance

Estimation du paramètre θ

Soit un n-échantillon $(X_1, X_2, ..., X_n)$ de la loi de Lindley a un paramètre θ, on suppose les données sont progressivement censurée de type II à l'ordre m avec : $r = (r_1, r_2, ..., r_m)$.

La fonction du vraisemblance est :

$$\begin{aligned} L(\theta, x) &= C \prod_{i=1}^{m} f(x_i, \theta)[1 - F(x_i, \theta)]^{r_i} \\ &= C\left(\frac{\theta^2}{1+\theta}\right)^m \prod_{i=1}^{m}\left((1+x_i)e^{-\theta\sum_{i=1}^{m} x_i}\right)\prod_{i=1}^{m}\left(\left(1 + \frac{\theta x_i}{1+\theta}\right)e^{-\theta x_i}\right)^{r_i} \end{aligned} \tag{3.7}$$

Avec :

- n unités en test.
- m premières défaillances observées $X_1 < X_2 < ... < X_m$
- r_i unités retirées à la i-ème défaillance (schéma de censure progressive).
- Contrainte : $\sum_{i=1}^{m} r_i = n - m$
- C : Constante de normalisation (dépend du schéma de censure).

  $C = n(n - r_1 - 1)(n - 2 - r_1 - r_2)...(n - \sum_{i=1}^{m}(r_i - 1))$

On prend le logarithme de la Vraisemblance on obtient :

$$\begin{aligned} \ell(\theta, x) = lnL(\theta, x) &= ln\left[C\left(\frac{\theta^2}{\theta+1}\right)^m \prod_{i=1}^{m}(1+x_i)e^{-\theta\sum_{i=1}^{m} x_i}\prod_{i=1}^{m}[(1+\frac{\theta x_i}{1+\theta})e^{-\theta x_i}]^{r_i}\right] \\ &= lnC + mln\left(\frac{\theta^2}{1+\theta}\right) + \sum_{i=1}^{m} ln(1+x_i) + ln(e^{-\theta\sum_{i=1}^{m} x_i}) + \sum_{i=1}^{m} ln\left[(1+\frac{\theta x_i}{1+\theta})e^{-\theta x_i}\right]^{r_i} \\ &= lnC + m(ln(\theta^2) - ln(\theta+1)) + \sum_{i=1}^{m} ln(1+x_i) - \theta\sum_{i=1}^{m} x_i + \sum_{i=1}^{m} r_i\left[ln(1+\frac{\theta x_i}{1+\theta}) + ln(e^{-\theta x_i})\right] \\ &= lnC + m(ln(\theta^2) - ln(\theta+1)) + \sum_{i=1}^{m} ln(1+x_i) - \theta\sum_{i=1}^{m} x_i + \sum_{i=1}^{m} r_i\left[ln(1+\frac{\theta x_i}{1+\theta}) - \theta x_i\right] \\ &= lnC + m(ln(\theta^2) - ln(\theta+1)) + \sum_{i=1}^{m} ln(1+x_i) - \theta\sum_{i=1}^{m} x_i + \sum_{i=1}^{m} r_i ln(1+\frac{\theta x_i}{1+\theta}) - \theta\sum_{i=1}^{m} r_i x_i \end{aligned} \tag{3.8}$$

L'estimateur du maximum de vraisemblance de paramètre θ est la solution de l'équation

$$\ell(\theta, x) = 0$$

On dérive par rapport à θ on obtient :

$$\begin{aligned}\frac{\partial \ell(\theta)}{\partial \theta} &= m\left(\frac{2\theta}{\theta^2} - \frac{1}{\theta+1}\right) - \sum_{i=1}^{m} x_i + \sum_{i=1}^{m} r_i \left(\frac{x_i}{(1+\theta)(1+\theta+\theta x_i)}\right) - \sum_{i=1}^{m} r_i x_i \\ &= \frac{2m}{\theta} - \frac{m}{\theta+1} - \sum_{i=1}^{m} x_i + \sum_{i=1}^{m} r_i \left(\frac{x_i}{(1+\theta)(1+\theta+\theta x_i)}\right) - \sum_{i=1}^{m} r_i x_i \end{aligned} \tag{3.9}$$

L'equation est non linéaire donc il n'y a pas de solution analytique. On utilise les méthodes numériques pour obtenir la valeur approximée de l'estimateur du maximum de vraisemblance $\hat{\theta}$ de paramètre θ. Dans ce chapitre, on utilise le package du **maxLik** qui a une capacité pour résoudre l'équation non linéaire .
Le tableau suivant présente les résultats de l'estimation du paramètre de la loi de Lindley, obtenus par la méthode du maximum de vraisemblance sous une censure progressive de type II, pour différentes valeurs de n et m , avec $\theta = 2$ et $L = 0.5$.

TABLE 3.1 – les résultats de l'estimation du paramètre de la loi de Lindley

n	m	$\hat{\theta}$	Biais	Erreur quadratique	C_{LX}	$\hat{C_{LX}}$
20	5	1.9347	-0.0652	0.0784	0.6681	0.6719
	8	2.3858	0.3858	0.4135	0.6681	0.6486
	15	2.3743	0.3743	0.4382	0.6681	6491
50	8	2.3925	0.3925	0.4207	0.6681	0.6483
	15	2.3609	0.3609	0.4263	0.6681	0.6497
	20	2.1127	0.1127	0.1127	0.6681	0.6619
100	8	2.3789	0.3789	0.4095	0.6681	0.6489
	15	2.3735	0.3735	0.4346	0.6681	0.6491
	20	2.1127	0.1127	0.1127	0.6681	0.6619

Ce tableau montre que plus la taille de l'échantillon n et le nombre de données observées m augmentent, plus l'estimation du paramètre θ devient précise (biais et erreur quadratique plus faibles). L'indice de performance estimé $\hat{C_{LX}}$ reste proche de sa valeur théorique (0.6681), ce qui confirme la fiabilité de l'estimation, surtout pour de grandes valeurs de m.

3.3 Distribution Lindley à 2 Paramètres θ,β

3.3.1 Fonction de Densité de probabilité (fdp)

La fonction de densité de probabilité (fdp) est donnée par :

$$f(x;\theta,\beta) = \frac{\theta^2}{\beta+\theta}(1+\beta.x)e^{-\theta.x}, \quad x>0, \quad \theta>0, \quad \beta>-\theta \tag{3.10}$$

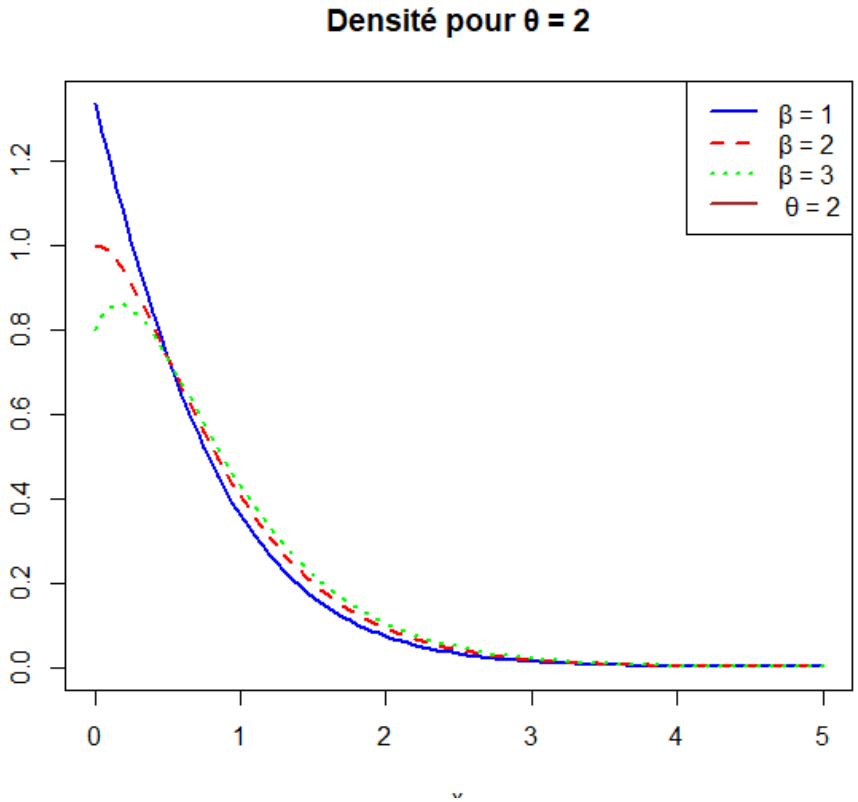

FIGURE 3.3 – Présentation graphique de la fonction de densité (Lindley 2 paramètres) pour $\theta = 2$ et quelques valeurs de β

3.3.2 Fonction de répartition (fdc)

La fonction de répartition cumulative correspondante[12] est donnée par :

$$F(x;\theta,\beta) = 1 - \frac{\theta+\beta+\beta\theta x}{\beta+\theta}e^{-\theta x}, \quad x>0, \quad \theta>0 \quad ,\beta>-\theta \tag{3.11}$$

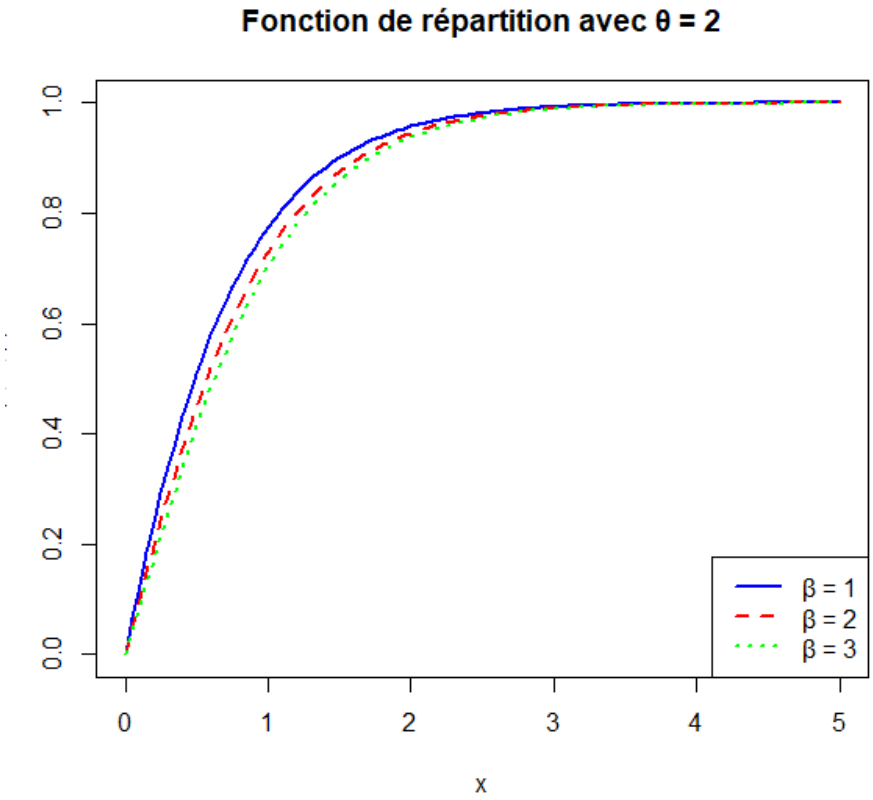

FIGURE 3.4 – Présentation graphique de la fonction de répartition (Lindley 2 paramètres) pour $\theta = 2$ et quelques valeurs de β

3.3.3 Fonction de fiabilité (Survie)

La fonction de Fabilité correspondante est définie comme suit :

$$R(x;\theta,\beta) = 1 - F(x;\theta;\beta) = \left(\frac{\theta+\beta+\beta\theta x}{\beta+\theta}\right)e^{-\theta x}, \quad x>0, \quad \theta>0, \quad \beta>-\theta \tag{3.12}$$

3.3.4 Moyenne (Espérance)

$$E[X] = \mu = \frac{\theta+2\beta}{\theta(\theta+\beta)}, \quad x>0, \quad \theta,\beta>0 \tag{3.13}$$

3.3.5 Variance

$$Var(X) = \frac{\theta^2+2\beta^2+4\theta\beta}{\theta^4+2\theta^3\beta+\theta^2\beta^2} \tag{3.14}$$

3.3.6 Indice de performance

$$C_{LX} = \frac{\mu - L}{\sigma} = \frac{\frac{\theta+2\beta}{\theta(\theta+\beta)} - L}{\sqrt{\frac{\theta^2+2\beta^2-2\theta\beta}{\theta^2(\theta+\beta)^2}}} \tag{3.15}$$
$$C_{LX} = \frac{\theta + 2\beta - L\theta(\theta+\beta)}{\sqrt{\theta^2 + 2\beta^2 + 4\theta\beta}}$$

où L est la limite inférieure.

3.3.7 Estimation par la méthode du maximum de vraisemblance(les données progressivement censurée de type II)

Estimation des paramètres θ et β

Soit un n-échantillon $(X_1, X_2, ..., X_n)$ de la loi de Lindley a deux paramètre θ et β, on suppose les données sont progressivement censuréede de type II à l'ordre m avec : $r = (r_1, r_2, ..., r_m)$.

La fonction du vraisemblance est :

$$\begin{aligned} L(\theta,\beta,x) &= C.\prod_{i=1}^{m} f(x_i,\theta,\beta).[1-F(x_i,\theta,\beta)]^{r_i} \\ &= C\left(\frac{\theta^2}{\beta+\theta}\right)^m \prod_{i=1}^{m}(1+\beta x_i)e^{-\theta\sum_{i=1}^{m}x_i}\prod_{i=1}^{m}\left(\left(\frac{\theta+\beta+\beta\theta x_i}{\beta+\theta}\right)e^{-\theta x_i}\right)^{r_i}. \end{aligned} \tag{3.16}$$

Avec : $C = n(n-r_1-1)(n-2-r_1-r_2)...(n-\sum_{i=1}^{m}(r_i-1))$

On prend le logarithme de la Vraisemblance on obtient :

$$\begin{aligned} \ell(\theta,\beta,x) = lnL(\theta,\beta,x) &= ln\left[C\left(\frac{\theta^2}{\beta+\theta}\right)^m \prod_{i=1}^{m}(1+\beta x_i)e^{-\theta\sum_{i=1}^{m}x_i}\prod_{i=1}^{m}\left(\left(\frac{\theta+\beta+\beta\theta x_i}{\beta+\theta}\right)e^{-\theta x_i}\right)^{r_i}\right]. \\ &= lnc + 2mln(\theta) - mln(\beta+\theta) + \sum_{i=1}^{m} ln(1+\beta x_i) - \theta\sum_{i=1}^{m}x_i + \sum_{i=1}^{m} r_i ln\left(\frac{\theta+\beta+\theta\beta x_i}{\beta+\theta}\right) - \theta\sum_{i=1}^{m} r_i x_i \end{aligned} \tag{3.17}$$

On dérive par rapport à θ on obtient :

$$\frac{\partial\ell(\theta,\beta,x)}{\partial\theta} = \frac{2m}{\theta} - \frac{m}{\beta+\theta} - \sum_{i=1}^{m}x_i - \sum_{i=1}^{m} r_i x_i + \sum_{i=1}^{m} r_i\left(\frac{\beta^2 x_i}{(\beta+\theta)(\theta+\beta+\theta\beta x_i)}\right)$$

On dérive par rapport à β on obtient :

$$\frac{\partial \ell(\theta,\beta,x)}{\partial \beta} = \frac{-m}{\beta+\theta} + \sum_{i=1}^{m} \frac{x_i}{1+\beta x_i} - \sum_{i=1}^{m} r_i \frac{\theta^2 x_i}{(\theta+\beta)(\theta+\beta+\theta\beta x_i)}$$

L'estimateur du maximum de vraisemblance des paramètres θ et β est la solution de système suivant :

$$\begin{cases} \dfrac{\partial \ell(\theta,\beta)}{\partial \theta} = \dfrac{2m}{\theta} - \dfrac{m}{\beta+\theta} - \displaystyle\sum_{i=1}^{m} x_i - \sum_{i=1}^{m} r_i x_i + \sum_{i=1}^{m} r_i \left(\dfrac{\beta^2 x_i}{(\beta+\theta)(\theta+\beta+\theta\beta x_i)} \right) = 0 \\ \dfrac{\partial \ell(\theta,\beta)}{\partial \beta} = \dfrac{-m}{\beta+\theta} + \displaystyle\sum_{i=1}^{m} \frac{x_i}{1+\beta x_i} - \sum_{i=1}^{m} r_i \frac{\theta^2 x_i}{(\theta+\beta)(\theta+\beta+\theta\beta x_i)} = 0 \end{cases}$$

Ce système n'a pas de solution explicite. On utilise les méthodes numériques pour obtenir la valeur approximée au estimateur du maximum de vraisemblance $\hat{\theta}$ de paramètre θ et $\hat{\beta}$ de β. Dans ce chapitre, on utilise le package du **maxLik** qui a une capacité pour résoudre le système .

3.3.8 Estimation par la méthode du maximum de vraisemblance(les données censurée de type II)

Estimation des paramètres θ et β

Soit un n-échantillon $(X_1, X_2, ..., X_n)$ de la loi de Lindley a deux paramètre θ et β, on suppose les données sont censurée de type II à l'ordre m avec :$r = (0, 0, ..., n-m)$.

La fonction du vraisemblance est :

$$\begin{aligned} L(\theta,\beta,x) &= C.\prod_{i=1}^{m} f(x_i,\theta,\beta).[1-F(x_i,\theta,\beta)]^{n-m} \\ &= C\left(\frac{\theta^2}{\beta+\theta}\right)^m \prod_{i=1}^{m}(1+\beta x_i) e^{-\theta\sum_{i=1}^{m} x_i} \prod_{i=1}^{m}\left(\left(\frac{\theta+\beta+\beta\theta x_i}{\beta+\theta}\right)e^{-\theta x_i}\right)^{n-m}. \end{aligned} \tag{3.18}$$

Avec : $C = n(n-r_1-1)(n-2-r_1-r_2)...(n-\sum_{i=1}^{m}(r_i-1))$

On prend le logarithme de la Vraisemblance on obtient :

$$\begin{aligned} \ell(\theta,\beta,x) = logL(\theta,\beta,x) &= \log C + 2mlog(\theta) - mlog(\theta+\beta) + \sum_{i=1}^{m}\log(1+\beta x_i) - \theta\sum_{i=1}^{m} x_i \\ &+ (n-m)\sum_{i=1}^{m}\left[\log\left(\frac{\theta+\beta+\beta\theta x_i}{\beta+\theta}\right) - \theta x_i\right] \end{aligned}$$

On dérive par rapport à θ en suit par rapport à β on obtient :

$$\frac{\partial \ell}{\partial \theta} = \frac{2m}{\theta} - \frac{m}{\beta+\theta} - \sum_{i=1}^{m} x_i + (n-m)\sum_{i=1}^{m}\left[\frac{1+\beta x_i}{\theta+\beta+\beta\theta x_i} - \frac{1}{\beta+\theta} - x_i\right]$$

$$\frac{\partial \ell}{\partial \beta} = -\frac{m}{\beta+\theta} + \sum_{i=1}^{m}\frac{x_i}{1+\beta x_i} + (n-m)\sum_{i=1}^{m}\left[\frac{1+\theta x_i}{\theta+\beta+\beta\theta x_i} - \frac{1}{\beta+\theta}\right]$$

L'estimateur du maximum de vraisemblance des paramètres θ et β est la solution de système suivant :

$$\begin{cases} \dfrac{\partial \ell(\theta,\beta)}{\partial \theta} = \dfrac{2m}{\theta} - \dfrac{m}{\beta+\theta} - \displaystyle\sum_{i=1}^{m} x_i - \sum_{i=1}^{m} r_i x_i + \sum_{i=1}^{m} r_i\left(\frac{\beta^2 x_i}{(\beta+\theta)(\theta+\beta+\theta\beta x_i)}\right) = 0 \\ \dfrac{\partial \ell(\theta,\beta)}{\partial \beta} = \dfrac{-m}{\beta+\theta} + \displaystyle\sum_{i=1}^{m}\frac{x_i}{1+\beta x_i} - \sum_{i=1}^{m} r_i \frac{\theta^2 x_i}{(\theta+\beta)(\theta+\beta+\theta\beta x_i)} = 0 \end{cases}$$

Ce système n'a pas de solution explicite. On utilise les méthodes numériques pour obtenir la valeur approximée au estimateur du maximum de vraisemblance $\hat{\theta}$ de paramètre θ et $\hat{\beta}$ de β. Dans ce chapitre, on utilise le package du **maxLik** qui a une capacité pour résoudre le système .

Le tableau suivant présente les résultats de l'estimation du paramètre de la loi Lindley à deux paramètres, obtenus par la méthode du maximum de vraisemblance sous une censure de type II, pour différentes valeurs de n et m , avec $\theta = 1$, $\beta = 0.5$ et $L = 0.1$.

TABLE 3.2 – Les résultats de l'estimation des paramètres de la loi de Lindley

n	m	$\hat{\theta}$	Biais1	Erreur quadratique1	$\hat{\beta}$	Biais2	quadratique2	C_{LX}	$\hat{C_{LX}}$
20	5	1.4817	0.4817	0.9211	0.4664	-0.0335	1.0340	1.6734	0.4903
	8	1.2993	0.2993	0.5364	0.6016	0.1016	0.8605	1.6734	0.8316
	15	1.1220	0.1220	0.2328	0.8207	0.3207	0.7228	1.6734	1.2007
50	8	1.3924	0.3924	0.6272	0.7675	0.2675	0.9493	1.6734	0.6794
	15	1.2588	0.2588	0.4071	0.9616	0.4616	0.8050	1.6734	0.8533
	20	1.2569	0.2569	0.3541	0.9785	0.4785	0.8008	1.6734	0.8480
	25	1.2011	0.2011	0.2817	0.9828	0.4828	0.7464	1.6734	0.9385
100	20	1.3187	0.3187	0.4385	1.0761	0.5761	0.8584	1.6734	0.7100
	25	1.2682	0.2682	0.3663	1.1119	0.6119	0.8070	1.6734	0.7613
	30	1.2639	0.2639	0.3476	1.1096	0.6096	0.7977	1.6734	0.7684
	40	1.2366	0.2366	0.2956	1.1755	0.6755	0.7681	1.6734	0.7682

Ce tableau montre que l'augmentation de la taille d'échantillon n et du nombre d'observations m permet une meilleure estimation des deux paramètres θ et β, avec une réduction progressive des biais et des erreurs quadratiques. Toutefois, l'estimateur de β reste légèrement biaisé positivement, surtout pour les petites tailles. L'indice de performance estimé $\hat{C_{LX}}$ s'améliore également avec m, se rapprochant de la valeur théorique C_{LX}, ce qui confirme la consistance des estimateurs.

3.4 Distribution Lindley à 3 Paramètre α,λ,β

3.4.1 Fonction de Densité de probabilité (fdp)

La fonction de densité de probabilité (fdp) est donnée par :

$$f(x) = \frac{\alpha.\lambda^2(\beta + x^\alpha)x^{\alpha-1}e^{-\lambda.x^\alpha}}{1+\lambda.\beta}, \quad x > 0, \quad \alpha,\beta,\lambda > 0 \tag{3.19}$$

3.4.2 Fonction de répartition (fdc)

La fonction de répartition cumulative correspondante [13] est s'obtient par intégration de la (fdp) :

$$F(x) = 1 - \frac{(1+\lambda.\beta+\lambda.x^\alpha)e^{-\lambda.x^\alpha}}{1+\lambda.\beta}, \quad x > 0, \quad \alpha,\beta,\lambda > 0 \tag{3.20}$$

3.4.3 Fonction de fiabilité (Survie)

La fonction de fiabilité est :

$$R(x) = 1 - F(x) = 1 - \left(1 - \frac{(1+\lambda.\beta+\lambda.x^\alpha)e^{-\lambda.x^\alpha}}{1+\lambda.\beta}\right) \tag{3.21}$$

$$R(x) = \frac{(1+\lambda.\beta+\lambda.x^\alpha)e^{-\lambda.x^\alpha}}{1+\lambda.\beta}, \quad x > 0, \quad \alpha,\beta,\lambda > 0$$

3.4.4 Moyenne (Espérance)

$$E[X] = \mu = \frac{[\alpha(\lambda.\beta+1)+1]\Gamma(\frac{1}{\alpha})}{\alpha^2\lambda^{\frac{1}{\alpha}}(1+\lambda.\beta)}, \quad x > 0, \quad \alpha,\beta,\lambda > 0$$

Avec : $\Gamma(\frac{1}{\alpha}) = \int_0^\infty t^{\frac{1}{\alpha}-1}e^{-t}dt$

3.4.5 Variance

$$Var(X) = E[X^2] - (E[X])^2$$

Avec :

$$E[X^2] = \frac{2[\alpha(\lambda\beta+1)+2]\Gamma\left(\frac{2}{\alpha}\right)}{\alpha^2\lambda^{\frac{2}{\alpha}}(1+\lambda\beta)} \tag{3.22}$$

Donc :

$$Var(X) = \frac{2[\alpha(\lambda\beta+1)+2]\Gamma\left(\frac{2}{\alpha}\right)}{\alpha^2\lambda^{\frac{2}{\alpha}}(1+\lambda\beta)} - \left(\frac{[\alpha(\lambda\beta+1)+1]\Gamma\left(\frac{1}{\alpha}\right)}{\alpha\lambda^{\frac{1}{\alpha}}(1+\lambda\beta)}\right)^2 \tag{3.23}$$

$$Var(X) = \sigma^2 = \frac{1}{(1+\lambda\beta)}\left(\frac{2[\alpha(\lambda\beta+1)+2]\Gamma\left(\frac{2}{\alpha}\right)}{\alpha^2\lambda^{\frac{2}{\alpha}}} - \frac{[\alpha(\lambda\beta+1)+1]^2\Gamma\left(\frac{1}{\alpha}\right)^2}{\alpha^2\lambda^{\frac{2}{\alpha}}(1+\lambda\beta)}\right)$$

3.4.6 Indice de performance

$$C_{LX} = \frac{\mu - L}{\sigma} = \frac{\frac{[\alpha(\lambda.\beta+1)+1]\Gamma(\frac{1}{\alpha})}{\alpha^2\lambda^{\frac{1}{\alpha}}(1+\lambda.\beta)} - L}{\sqrt{\frac{1}{(1+\lambda\beta)}\left(\frac{2[\alpha(\lambda\beta+1)+2]\Gamma\left(\frac{2}{\alpha}\right)}{\alpha^2\lambda^{\frac{2}{\alpha}}} - \frac{[\alpha(\lambda\beta+1)+1]^2\Gamma\left(\frac{1}{\alpha}\right)^2}{\alpha^2\lambda^{\frac{2}{\alpha}}(1+\lambda\beta)}\right)}}$$

3.4.7 Estimation par la méthode du maximum de vraisemblance(les données progressivement censurée de type II)

Estimation du paramètre α, β, λ

Soit un n-échantillon $(X_1, X_2, ..., X_n)$ de la loi de Lindley a troi paramètres α, β et λ, on suppose les données sont progressivement censurée de type II à l'ordre m avec : $r = (r_1, r_2, ..., r_m)$. La fonction du vraisemblance est :

$$L(\alpha,\beta,\lambda) = C.\prod_{i=1}^{m} f(x).(R(x))^{r_i} \tag{3.24}$$

$$L(\alpha,\beta,\lambda) = C.\prod_{i=1}^{m} \frac{\alpha.\lambda^2(\beta+x^\alpha)x^{\alpha-1}e^{-\lambda.x^\alpha}}{1+\lambda.\beta}.\prod_{i=1}^{m}\left(\frac{(1+\lambda.\beta+\lambda.x^\alpha)e^{-\lambda.x^\alpha}}{1+\lambda.\beta}\right)^{r_i}$$

On prend le logarithme de la Vraisemblance on obtient :

$$\ell(\alpha,\beta,\lambda) = \ln L(\alpha,\beta,\lambda) = logC + m\ln(\alpha) + 2m\ln(\lambda) \tag{3.25}$$

$$- m\ln(1+\lambda\beta) + \sum_{i=1}^{m}\ln(\beta+x_i^\alpha) + (\alpha-1)\sum_{i=1}^{m}\ln x_i \tag{3.26}$$

$$- \lambda\sum_{i=1}^{m} x_i^\alpha(1+r_i) + \sum_{i=1}^{m} r_i\ln(1+\lambda\beta+\lambda x_i^\alpha) - \sum_{i=1}^{m} r_i\ln(1+\lambda\beta) \tag{3.27}$$

On dérive ℓ par rapport à α, β, λ :

Dérivée par rapport à α :

$$\frac{\partial \ell}{\partial \alpha} = \frac{m}{\alpha} + \sum_{i=1}^{m} \frac{x_i^\alpha \ln x_i}{\beta + x_i^\alpha} + \sum_{i=1}^{m} \ln x_i - \lambda \sum_{i=1}^{m} (1 + r_i) x_i^\alpha \ln x_i + \sum_{i=1}^{m} r_i \frac{\lambda x_i^\alpha \ln x_i}{1 + \lambda\beta + \lambda x_i^\alpha} \tag{3.28}$$

Dérivée par rapport à β :

$$\frac{\partial \ell}{\partial \beta} = -\frac{m\lambda}{1 + \lambda\beta} + \sum_{i=1}^{m} \frac{1}{\beta + x_i^\alpha} + \sum_{i=1}^{m} r_i \frac{\lambda}{1 + \lambda\beta + \lambda x_i^\alpha} - \sum_{i=1}^{m} r_i \frac{\lambda}{1 + \lambda\beta} \tag{3.29}$$

Dérivée par rapport à λ :

$$\frac{\partial \ell}{\partial \lambda} = \frac{2m}{\lambda} - \frac{m\beta}{1 + \lambda\beta} - \sum_{i=1}^{m} x_i^\alpha (1 + r_i) + \sum_{i=1}^{m} r_i \frac{\beta + x_i^\alpha}{1 + \lambda\beta + \lambda x_i^\alpha} \tag{3.30}$$

Pour obtenir les estimateurs $\hat{\alpha}, \hat{\beta}, \hat{\lambda}$, on résout numériquement le système d'équations non linéaires obtenu :

$$\frac{\partial \ell(\alpha, \beta, \lambda)}{\partial \alpha} = 0, \quad \frac{\partial \ell(\alpha, \beta, \lambda)}{\partial \beta} = 0, \quad \frac{\partial \ell(\alpha, \beta, \lambda)}{\partial \lambda} = 0$$

Ces équations sont complexes et n'admettent pas de solutions analytiques. On utilise alors une méthode numérique (comme le package `maxLik` de R) pour les résoudre.

3.5 Application sur des données réelles

Dans cette étude, nous analysons des données réelles de temps d'attente (en minutes) pour 100 clients dans une banque, observés et rapportés par Ghitany [14]. Nous comparons l'ajustement de la loi de Lindley avec d'autres lois classiques comme l'exponentielle, log-normale, Weibull et gamma. Pour cela, nous utilisons les critères AIC et BIC, ainsi que les tests de Kolmogorov-Smirnov (KS), Cramér-von Mises (CvM) et Anderson-Darling (AD).

L'ensemble des données est présenté dans le tableau 3.3

TABLE 3.3 – Temps d'attente (en minutes) de 100 clients dans une agence bancaire

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

3.5.1 Description des données

Les statistiques suivantes ont été obtenues à partir de résumé **summary(x)**

TABLE 3.4 – Résumé statistique des temps d'attente

Statistique	Valeur
Minimum	0.800
1er Quartile	4.675
Médiane	8.100
Moyenne	9.877
3e Quartile	13.025
Maximum	38.500

Ces valeurs montront que les temps d'attente varient beaucoup, de 0,8 à 38,5 minutes. La moitié des clients attendent entre 5 et 13 minutes environ. Quelques clients ont des attentes très longues (plus de 30 minutes), ce qui fait que le temps moyen (9,9 minutes) est plus élevé que le temps médian (8,1 minutes). Cela montre que la plupart des clients sont servis relativement vite, mais certains doivent attendre beaucoup plus longtemps.

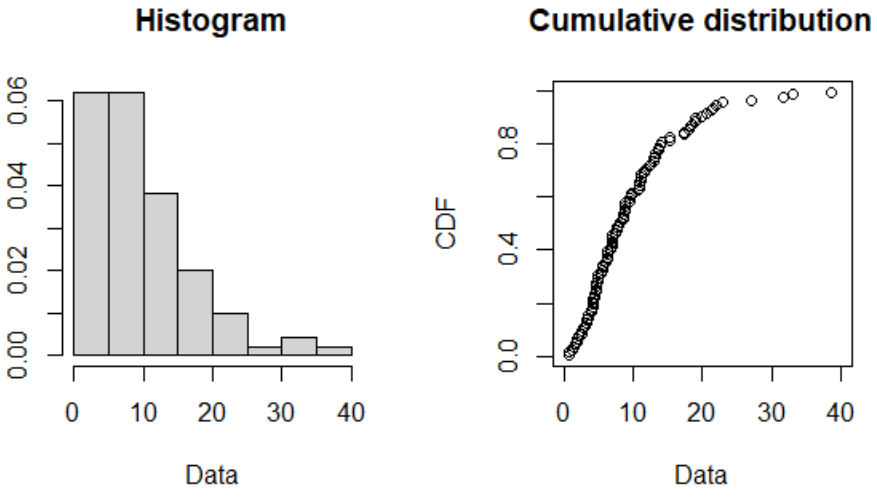

FIGURE 3.5 – Histogramme et fonction de répartition cumulée des données

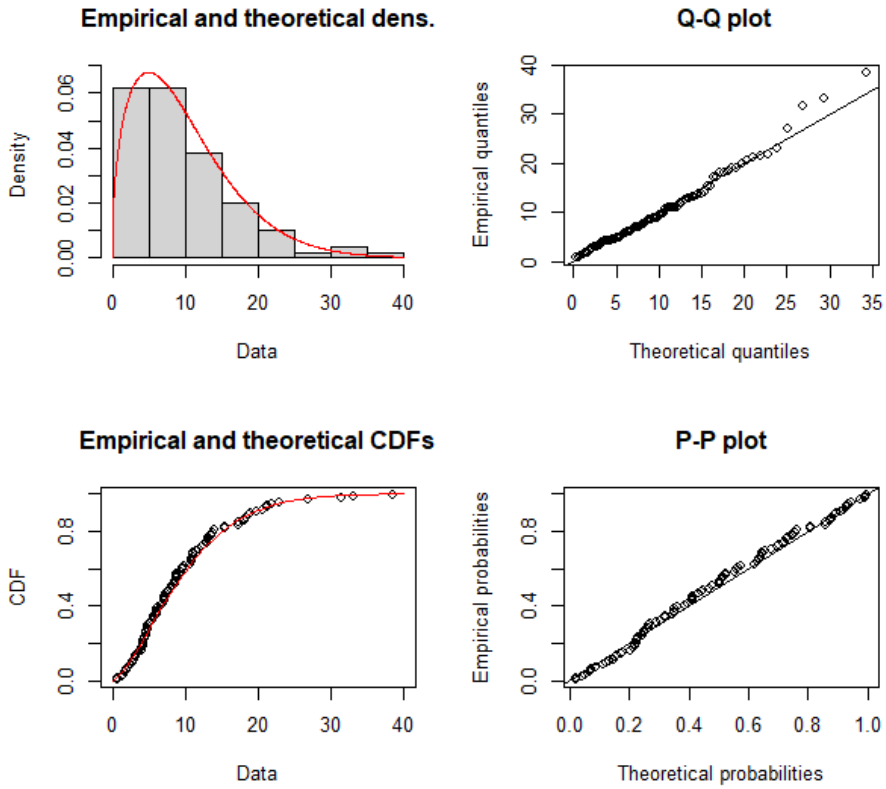

FIGURE 3.6 – Densités estimées, Densités cumulées estimées, Graphiques P-P et Q-Q pour Lindley à un paramètre ajusté l'ensemble de données

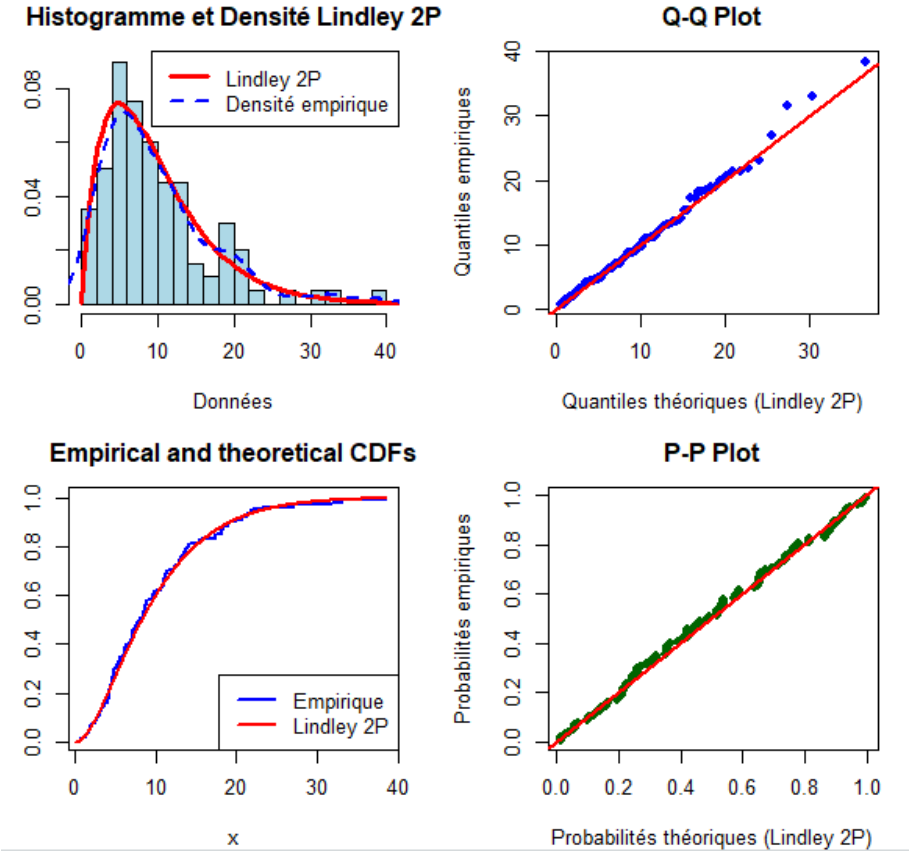

FIGURE 3.7 – Densités estimées, Densités cumulées estimées, Graphiques P-P et Q-Q pour Lindley à deux paramètre ajusté l'ensemble de données

Afin de comparer les trois modèles de distribution, nous considérons des critères comme, AIC (Akaike critère d'information), BIC (critère d'information bayésien) et KS (test de kolmogorov-Smirnov), CvM(Cramer-von Mises) et AD(Anderson-Darling) pour le jeu de données. La meilleure distribution correspond aux petits, AIC, BIC, KS, CvM, et Valeurs AD

$$AIC = 2K - 2l,$$

$$BIC = K * log(n) - 2l, \quad et \quad KS = sup_x|F_n(x) - F(x)|$$

$$V = \sqrt{\frac{x^2/n}{min(I,J)-1}}, \quad et \quad A^2 = n\int_{-\infty}^{+\infty} \frac{(F_n(x)-F(x))^2}{F(x)(1-F(x))} dF(x)$$

où

$F_n(x) = \frac{1}{n}\sum_{i=1}^{m} I_{x_i \leq x}$ est la fonction de distribution empirique .

K : le nombre de paramètres dans le modèle statistique.

$F(x)$: la fonction de distribution cumulative.

n :la taille de l'échantillon .

l :la valeur qui maximisée la fonction de log-vraisemblance sous le modèle considéré

3.5.2 Ajustement des lois avec la loi de Lindley à un paramètre

TABLE 3.5 – Comparaison des ajustements statistiques des distributions selon les critères AIC, BIC, KS, CvM, AD

Modèle	AIC	BIC	KS	AD	CvM
Lindley-1P	**637.600**	**641.680**	**0.0415**	**0.1755**	**0.0277**
Gamma	638.600	643.811	0.0425	0.1855	0.0287
Weibull	641.461	646.672	0.0578	0.4058	0.0611
Log-Normal	642.348	647.558	0.0565	0.4083	0.0541

Interprétation des résultats

Le tableau ci-dessus présente la comparaison de plusieurs modèles de distribution selon les critères statistiques d'ajustement : AIC, BIC, Kolmogorov-Smirnov (KS), Cramér-von Mises (CvM) et Anderson-Darling (AD).

Il en ressort que le modèle **Lindley-1P** offre le meilleur ajustement aux données, avec les plus faibles valeurs pour tous les critères : AIC = 637.600, BIC = 641.680, KS = 0.0415, CvM = 0.0277 et AD = 0.1755. Ces résultats ont été obtenus à l'aide du package `fitdistrplus` sous `R`, qui permet l'ajustement de distributions théoriques aux données empiriques de manière efficace.

Ces résultats indiquent que, malgré sa simplicité, ce modèle est compétitif par rapport aux distributions classiques comme Gamma, Weibull et Log-Normale.

TABLE 3.6 – estimations EMV pour le Modèle Lindely(1 paramètre) evec censure progressive de type II pour L=0.5 (n = 100)

m	$\hat{\theta}$	$\hat{C_{LX}}$
15	0.2623	2.17495
20	0.2619	2.115427
30	0.2488	2.188153
40	0.2490	2.06054

3.5.3 Ajustement des lois avec la loi de Lindley à deux paramètres

TABLE 3.7 – Comparaison des ajustements statistiques des distributions selon les critères AIC, BIC, KS, CvM, AD

Modèle	AIC	BIC	KS	AD	CvM
Lindley-2P	**633.610**	**639.321**	**0.02532**	**0.09175**	**0.0141**
Gamma	638.600	643.811	0.0425	0.1855	0.0287
Weibull	641.461	646.672	0.0578	0.4058	0.0611
Log-Normal	642.348	647.558	0.0565	0.4083	0.0541

Interprétation des résultats

Afin d'évaluer la qualité d'ajustement des différents modèles statistiques sur les données observées, plusieurs critères ont été utilisés : l'AIC , le BIC , ainsi que trois statistiques de test de bonne adéquation, à savoir KS , CvM et AD .

Le tableau montre clairement que le modèle **Lindley-2P** présente les plus faibles valeurs pour l'ensemble des critères considérés, ce qui indique une excellente adéquation aux données. En particulier, les valeurs minimales des critères AIC (633.610) et BIC (639.321), ainsi que les statistiques de tests KS (0.02532), AD (0.09175) et CvM (0.0141), confirment que ce modèle surpasse les distributions classiques telles que Gamma, Weibull ou Log-Normale.

Ces résultats suggèrent que la loi de Lindley à deux paramètres est particulièrement appropriée pour modéliser les temps d'attente dans ce contexte spécifique.

3.5.4 Comparaison entre les modèles Lindley à un paramètre et à deux paramètres selon les critères d'ajustement

TABLE 3.8 – Comparaison entre le modèle Lindley à un paramètre et celui à deux paramètres selon différents critères d'ajustement

Modèle	AIC	BIC	KS	AD	CvM
Lindley-2P	**633.610**	**639.321**	**0.02532**	**0.09175**	**0.0141**
Lindley-1P	637.600	641.680	0.04150	0.17550	0.0277

On observe, à partir du tableau ci-dessus, que le modèle **Lindley à deux paramètres** surpasse clairement le modèle à un seul paramètre selon tous les critères d'ajustement considérés.

Les valeurs plus faibles de l'AIC, du BIC ainsi que des statistiques KS, AD et CvM indiquent une meilleure adéquation du modèle à deux paramètres aux données observées. Cela s'explique par la plus grande flexibilité du modèle à deux paramètres, qui lui permet de mieux capturer la structure des données.

Ces résultats confirment l'intérêt d'utiliser le modèle généralisé de Lindley lorsqu'une meilleure précision de modélisation est recherchée.

TABLE 3.9 – estimations EMV pour le Modèle Lindely(2 paramètre) evec censure de type II pour L=0.5 (n = 100)

m	$\hat{\theta}$	$\hat{\beta}$	$\hat{C_{LX}}$
15	0.0845	0.0820	2.9427
20	0.0949	0.0909	2.9443
30	0.1146	0.1067	2.9471
40	0.1237	0.1135	2.9488

CONCLUSION GÉNÉRALE

Ce mémoire a étudié l'indice de performance dans le domaine de la fiabilité, en particulier son estimation à partir de données censurées progressivement de type II. Après avoir présenté les notions fondamentales de la fiabilité et de l'inférence statistique, nous avons focalisé notre attention sur la distribution de Lindley sous ses formes à un et deux paramètres. À l'aide de la méthode du maximum de vraisemblance, nous avons estimé les paramètres de cette distribution pour différents types de données. Les résultats obtenus, notamment sur des données réelles, montrent que le modèle Lindley à deux paramètres offre un ajustement supérieur. Ce travail ouvre des perspectives vers des généralisations plus poussées et l'utilisation de méthodes bayésiennes pour une modélisation encore plus précise.

BIBLIOGRAPHIE

[1] Douglas C Montgomery. *Introduction to statistical quality control*. John wiley & sons, 2020.

[2] Mohammad Vali Ahmadi, Mahdi Doostparast, and Jafar Ahmadi. Estimating the lifetime performance index with weibull distribution based on progressive first-failure censoring scheme. *Journal of Computational and Applied Mathematics*, 239 :93–102, 2013.

[3] Norman L Johnson, Samuel Kotz, and Narayanaswamy Balakrishnan. *Continuous univariate distributions, volume 2*, volume 2. John wiley & sons, 1995.

[4] Wen-Chuan Lee, Jong-Wuu Wu, and Ching-Wen Hong. Assessing the lifetime performance index of products with the exponential distribution under progressively type ii right censored samples. *Journal of computational and applied mathematics*, 231(2) :648–656, 2009.

[5] Ahmed A Soliman. Reliability estimation in a generalized life-model with application to the burr-xii. *IEEE Transactions on Reliability*, 51(3) :337–343, 2002.

[6] Jong-Wuu Wu, Wen-Chuan Lee, Ching-Wen Hong, and Sung-Yu Yeh. Implementing lifetime performance index of burr xii products with progressively type ii right censored sample. *International Journal of Innovative Computing, Information and Control*, 10(2) :671–693, 2014.

[7] Zhi-Sheng Ye and Nan Chen. Closed-form estimators for the gamma distribution derived from likelihood equations. *The American Statistician*, 71(2) :177–181, 2017.

[8] Narayanaswamy Balakrishnan and Rita Aggarwala. *Progressive censoring : theory, methods, and applications*. Springer Science & Business Media, 2000.

[9] Peter W Zehna et al. Invariance of maximum likelihood estimators. *Annals of Mathematical Statistics*, 37(3) :744, 1966.

[10] Ahmed A Soliman. Estimation of parameters of life from progressively censored data using burr-xii model. *IEEE Transactions on Reliability*, 54(1) :34–42, 2005.

[11] Shuo-Jye Wu and Coşkun Kuş. On estimation based on progressive first-failure-censored sampling. *Computational Statistics & Data Analysis*, 53(10) :3659–3670, 2009.

[12] Rama Shanker, Shambhu Sharma, and Ravi Shanker. A two-parameter lindley distribution for modeling waiting and survival times data. *Applied Mathematics*, 4(2) :363–368, 2013.

[13] Nosakhare Ekhosuehi and Festus Opone. A three parameter generalized lindley distribution : properties and application. *Statistica*, 78(3) :233–249, 2018.

[14] Iman Makhdoom, Shahram Yaghoobzadeh Shahrastani, and FGhazalnaz Sharifonnasabi. Bayesian inference for the lindley distribution under type-ii censoring with fuzzy data. *Journal of Data Science and Modeling*, pages 245–265, 2025.