République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
Université AMO de Bouira
Faculté des Sciences et des Sciences Appliquées
Département du Mathématiques

Mémoire de Master

Filière : Mathématiques

Spécialité : Recherche Opérationnelle

Thème

Sur les modéles d'analyse de la survie dans le cadre paramétriques

---

Présenté par :
- Rouam Hayet
- Bellatrache Djahida

Devant le jury composé de :

Président	M^r Nacer Demouche	MCB	U. A/M/O Bouira.
Promoteur	M^r Said Beddek	MAA	U. A/M/O Bouira.
Examinateur	M^r Karim Hamid	MAA	U. A/M/O Bouira.
Examinatrice	M^{me} Khadidja Boudane	MAA	U. A/M/O Bouira.

2018/2019

Dédicaces

Je dédie ce travail de longues années d'étude à :
La lumiére de ma vie, au coeur le plus tendre et le plus doux, à celle qui s'est tellement sacrifiée pour me voir toujours meilleure : ma trés chére mére.
A lêtre le plus cher à mon coeur, à celui qui m'a toujours guidée par ses conseils et qui m'a encouragée poursuivre mes études : Mon pére qu'allah l'accueille en son vaste paradis.
A l'esprit de mon chére amie Ogbi Ratiba.
Mes soeurs et Mes fréres et tout ma famille Rouam et la famille Bellatrache.
Mon fiancé Zaknoun Said et sa famille.
Tous ceux qui m'ont aidé de prés ou de loin.
A tous mes amis sans exception et spécialement à :Kahina,Sarah,Ahlam,Fatiha,Lamia,Widade

Hayet

Je dédie ce travail à :
A ceux qui m'ont tout donnée sans rien en retour A ceux qui m'ont encouragée et soutenue dans les moments les plus diffciles A vous mes chers parents Le plus beau cadeau que Dieu puissent faire à un enfant, pour leur amour et leur support continu. Que ce travail soit le temoignagée sincére et affectueux de ma profonde reconnaissance pour tout ce que vous avez fait pour moi.
A mes chéres fréres et à mes chéres Soeurs
A Les deux familles :Bellatrache et Rouam .
J'adresse aussi mes dédicaces à mes amies avec qui j'ai passée des moments agréables.

Djahida

Résumé

On s'interesse ici *à* la construction des nouveaux modéles d'analyse de la survie basée sur deux distributions continues et *à* étudié certaines propriétées statistique de ces modéles proposé, et on s'interesse aussi *à* leur application et *à* l'estimation de ses paramétres en utilisant la méthode du maximum de vraissemblance.

Mots clés :analyse de la survie,distribution, méthode du maximum de vraissemblance.

Abstract

We are interested here by the construction of two models of survival analysis based on two continuous distribution and studied some statistical properties of the proposed models ,and we also interested by the application of these models and by the maximum likelihood estimates of parameters.

Key words : Survival analysis,distribution,maximum likelihood.

Remerciements

REMERCIEMENTS

Nous remercions Dieu pour le courage, la patience et la volonté qui nous ont été utiles tout au long de notre parcours.

Nous tenons *à* remercier **M BEDDEK SAID** pour la proposition du théme, l'encadrement de ce travail, pour ses précieux conseils et orientations.

Nous remercions également les membres du jury : **M HAMID,M DEMOUCHE et Mme Boudane** pour avoir accepté d'examiner et d'évaluer notre travail.

Nos remerciements vont *à* tous ceux qui ont contribué d'une quelconque maniére *à* l'aboutissement de ce travail.

Nos sincéres remerciements s'adressent enfin *à* tous ceux qui nous ont soutenu de prés ou de loin.

Table des matières

Introduction générale

L'analyse de la survie remonte au XVII siècle passé dans le domaine de la démographie et avait pour objectif l'estimation de la croissance d'une population $\grave{a}$ partir des registres de décès et de divers caractéristiques. $\grave{a}$ partir du XIX^e siècle,Ces analyses ne sont affinées avec l'apparition de catégorisations suivant des variables exogène (sexe, nationalité,....),durant ce siècle ,les statisticiens commencé les premiers modélisations concernant la probabilité de mourir $\grave{a}$ un certain $\hat{a}ge$, probabilité qui sera par la suite désignée sous le terme de fonction de risque.
jusqu'en 1950,la communauté des statisticiens s'intéresse peu $\grave{a}$ l'analyse des données de survie, la principale contribution étant celle de Greenwood (1926) qui propose une formule pour l'erreur standard d'une table de survie. En 1951,W. Weibull propose un modèle paramétrique pour calculer la fiabilité d'un système non réparable qui sera par la suite utilisée en analyse de la survie (la loi de Weibull) [8].
en 1958 Kaplan et Meier proposent d'utiliser dans le domaine médical un estimateur non paramétrique,ils présentent des résultats importants concernant cet estimation .
l'année 1972, un modèle statistique semi-paramétrique voit le jour,gr$\grave{a}$ce aux travaux de Cox ,il est intervenu des variables explicatives(exogènes) dans la fonction de risque [8].
L'analyse de survie a pour objectifs de fournir les outils statistiques pour apprécier la rapidité avec laquelle un évènement survient, tester l'association statistique entre une exposition d'intérêt et la rapidité de survenue de l'événement, et quantifier cette association statistique.
L'analyse de survie doit être conduite sur des données issues d'un essai clinique ou dune étude de cohorte. Ne pas utiliser les outils statistiques issus de l'analyse de survie lorsque l'on travaille sur des données issues d'un essai clinique ou d'une étude de cohorte peut conduire $\grave{a}$ des biais dans les estimations des paramètres.
Les modéles de durée constituent un outil utilisé dans de nombreux domaines :
_ Fiabilité : durée de la vie d'un matériel, durée entre deux pannes d'un matériel réparable
_ Démographie, médecine : durée de la vie humaine, durée entre le déclenchement

d'une maladie et la guérison, durée séparant deux naissances
_ Économie, assurance : durée d'un épisode de chômage, durée de vie d'une entreprise, durée séparant deux sinistres, instant d'un défaut de paiement, durée avant la ruine
_ Dans tous les domaines oú l'on cherche $\grave{a}$ mesurer l'instant d'arrivée d'un événement aléatoire (panne, mort, maladie, chômage,)
Le domaine d'application de ces modèles est donc large.
L'analyse des données de survie a pour première particularité de ne concerner que des variables aléatoires positives (modélisant les durées de vie). Une conséquence de cette particularité est que la loi normale ne sera plus ici la référence en matière de distribution. Le plus souvent, toute autre loi issue de la famille exponentielle, et $\grave{a}$ support dans R^+, lui sera préférée.
Dans ce travail,nous allons nous intéresser $\grave{a}$ la construction de nouveaux modèles pour l'analyse des données de survie dans un cadre paramétrique en faisant appel $\grave{a}$ un degré de sophistication supérieur $\grave{a}$ la simple modélisation d'un échantillon I ID de loi paramétrique fixé $\grave{a}$ priori .
la premier modèle est basé sur un mélange proportionnel de lois de probabilité ,qu'on le définit par la formule :

$$G(x) = [p_1F_1(x) + p_2F_2(x)]^\alpha$$

avec : $F_1(x)$ et $F_2(x)$ sont deux fonctions de répartitions et α un nombre réel positif
et $0 \leq p_1 \leq 1$, $0 \leq p_2 \leq 1$.
le deuxième modèle est inspiré de l'approche de Kolmogrov-Lehman ,on le définit par la formule :

$$G(x) = \frac{p_1F_1(x) + p_2F_2(x)}{2 - F_1(x)}$$

avec : $F_1(x)$ et $F_2(x)$ sont des fonctions de répartitions
dans les deux modèles,nous avons pris deux lois de Weibull de paramètres différents .
pour cela nous avons structuré notre mémoire en trois chapitres ,dans le premier chapitre nous exposons les notions préliminaires d'analyse de survie que nous utiliserons par la suite,dans le deuxième chapitre,nous définissons un modèle qui s'appelle l'alternative de Lehman et nous présentons les deux modèles que nous le proposons,enfin,dans le troisième chapitre nous faisons une petite simulation et nous établissons une comparaison entre les résultats de modèle de lehman et les deux modèles proposé.

Chapitre 1

Généralités sur l'analyse de la survie

1.1 Introduction

l'objectif de ce chapitre est de fixer le cadre Générale de ce travail et d'introduire certaines notions et concepts élémentaires nécessaires et connus dans la littérature

1.2 Notions préliminaires

1.2.1 Variable aléatoire réelle

Définition 1.1. [7]
on appelle variable aléatoire (v.a) toute application mesurable T d'un espace de probabilité (Ω, F, P) dans un espace mesurable (B,$\mathbb{B}$).

1.2.2 La fonction de répartition

Définition 1.2. [10]
La fonction de répartition (distribution function) est une notion clé de la théorie des probabilités. Elle indique, pour la valeur donnée prise par une variable aléatoire (v.a), un cumul de probabilités.
Soit une fonction F associée à une v.a T. F est une fonction de répartition si :

$$F(T) = P(T \leq t).$$

Cette fonction est croissante sur R ,continue $\grave{a}$ droite. Elle varie de 0 à 1, autrement dit :

$$\lim_{t \longrightarrow -\infty} F(t) = 0, \lim_{t \longrightarrow +\infty} F(t) = 1$$

Si la v.a est discrète, il s'agit d'une fonction en escaliers. Soit p_i la probabilité que T prenne la valeur t_i.
Appliquons la propriété des probabilités totales :

$$F(t) = \sum_{t_i \leq t} P_i$$

Si la v.a est absolument continue

$$F(t) = \int_{-\infty}^{t} f(u)du$$

dans le domaine de l'analyse de la survie La fonction de répartition (ou c.d.f. pour "cumulative distribution function") représente, pour t fixé, la probabilité de mourir avant l'instant t, c'est-à -dire

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

1.2.3 Densité de probabilité f

Définition 1.3. C'est la fonction $f(t) \geq 0$ telle que pour tout $t \geq 0$

$$F(t) = \int_{0}^{t} f(u)du$$

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

$$f(t) = \lim_{h \longrightarrow 0} \frac{p(t < T < t + h)}{h} = F'(t)$$

Pour t fixé, la densité de probabilité représente la probabilité de mourir dans un petit intervalle de temps après l'instant t.

1.2.4 La fonction caractéristique

Définition 1.4. soit X une variable aléatoire,on appelle fonction caractéristique de x la fonction ϕ_x définie sur R à valeurs complexes, par :

$\phi_x(t) = E(\exp itx)$
en pratique,suivant le cas, la formule est différent :

— si X est une variable discrète finie :

$$\phi_x(t) = \sum_{j=1}^{n} \exp(itx_j)P(X = x_j)$$

— si X est une variable discrète dénombrable :

$$\phi_x(t) = \sum_{j=1}^{+\infty} \exp(itx_j)P(X = x_j)$$

— si X est une variable réelle :

$$\phi_x(t) = \int_{-\infty}^{+\infty} \exp(itx)f(x)dx$$

1.2.5 La durée de vie

Définition 1.5. On appelle durée de vie une variable aléatoire T positive, généralement, la durée s'écoulant entre deux événements.
Ex d'événements. : mort, panne, sinistre, entrée en chômage, maladie.

Modèles de durée de vie

sous-ensemble de méthodes statistiques adaptées à l'étude des durées de vie.

1.2.6 Fonction de survie S(t)

Définition 1.6. [10, 5]
La fonction de survie est égale à la probabilité que le décés intervienne après un temps t donné (appelée également courbe de survie).

$$S(t) = P(t < T)$$

avec S(t) est une fonction monotone décroissante
T est une variable aléatoire symbolisant le moment du décés, et P est la fonction

probabilité.
$S(0) = 1$ et $\lim_{t \longrightarrow +\infty} S(t) = 0$
Remarquons que :

$$S(t) = P(T \geq t) = 1 - P(T \leq t) = 1 - F(t)$$

1.2.7 Risque instantané λ (ou taux de hasard)

Définition 1.7. [10, 5]

Le risque instantané (ou taux d'incidence), pour t fixé caractérise la probabilité de mourir dans un petit intervalle de temps après t, conditionnellement au fait d'avoir survécu jusqu'au temps t (c'est *à* -dire le risque de mort instantané pour ceux qui ont survécu) :

$$\lambda(t) = \lim_{h \longrightarrow 0} \frac{p(t < T < t + h/t \leq T)}{h} = \frac{f(t)}{S(t)} = \frac{f(t)}{1 - F(t)}$$

Propriétés du taux de hasard :

Quand on analyse des durées de vie, les 5 formes les plus usuelles de taux de hasard sont :

— Un taux de hasard décroissant est caractéristique d'un système qui s'améliore (phase de déverminage ou de jeunesse, « burn in », mortalité infantile)

h décroit$\leftrightarrow$ Loi de T DFR (decreasing failure rate)

— Un taux de hasard constant est caractéristique d'un système qui ne vieillit pas (phase d'exploitation ou de vie utile)

h constant $\leftrightarrow$ Loi de T exponentielle

— Un taux de hasard croissant est caractéristique d'un système qui se détériore (phase d'obsolescence ou de vieillissement)

h croit $\leftrightarrow$ Loi de T IFR (decreasing failure rate)

— En cloche

— En forme de baignoire

1.2.8 Quantités associées *à* la distribution de survie

Soit X une variable aléatoire réelle définie sur I, de fonction de répartition F_x et de loi de probabilité P.

Moment ordinaire

Définition 1.8. [13]
Le moment (ou moment ordinaire) d'ordre $r \in N$ de X est défini, s'il existe, par :

$$\mu_r = \sum_{k \in I} k^r P_k$$

si X est discrète

$$\mu_r = \int_{x \in I} x^r f_X(x) dx$$

si X est absolument continue

Moment centré

Définition 1.9. [13]
Le moment centré d'ordre $r \in N$ de X est défini, s'il existe, par :

$$\mu_r = E([x - E(x)]^r) = \sum_{k \in I} [k - E(x)]^r P_k$$

si X est discrète

$$\mu_r = \int_{x \in I} [x - E(x)]^r f(x) dx$$

si X absolument est continue

1.3 Estimation

Un estimateur du paramètre inconnu θ d'un modèle ou loi de probabilité est une fonction qui fait correspondre à une suite d'observations $t_1 .....t_n$ issues du modèle ou de la loi de probabilité, la valeur θ que l'on nomme estimateur ou estimation $\hat{\theta} = f(t_1.....t_n)$

1.3.1 L'estimation par la Méthode du maximum de vraisemblance

La Méthode du maximum de vraisemblance est une Méthode d'estimation paramétrique qui doit sa popularité à :

- la simplicité de son approche
- l'aspect numérique accessible grâce à l'application de Méthode d'optimisation connues.

Elle permet de :

- construire des estimateurs performants
- construire des intervalles de confiances précis
- mettre en ouvre des tests statistiques "puissants"

Fonction de vraisemblance

Définition 1.10. [1]
on appelle fonction de vraisemblance pour $(t_1, ..., t_n)$ la fonction de θ :

$$L(t_1, ..., t_n, \theta) = \prod_{i=1}^{n} f(t_i\theta)$$

avec :

$$f(t_i;\theta) = \begin{cases} p_\theta(T = t) \; si \; T \; est \; une \; var \; discrte \\ f_\theta(T) \; si \; T \; est \; une \; var \; de \; densit \; f_\theta \end{cases}$$

Estimateurs du maximum de vraisemblance

On appelle estimateur du maximum de vraisemblance de θ(EMV) pour $(t_1, ..., t_n)$ un réel θ^* qui maximise la fonction de vraisemblance $L_n(t_1, ..., t_n, \theta)$ i.e pour tout θ :

$$L_n(t_1, ..., t_n, \theta) \leq L_n(t_1, ..., t_n, \theta^*)$$

Une expression alternative est :

$$\theta^* \in \arg\max L_n(t_1, ..., t_n, \theta)$$

ou $\arg\max$ désigne l'argument du maximum qui est l'ensemble des points en lesquels une expression atteint sa valeur maximale.
Puisqu'il dépend de $(t_1, ..., t_n)$, θ^* est une estimation ponctuelle de θ

Fonction de Log-vraisemblance

On appelle fonction de log-vraisemblance pour $(t_1, ..., t_n)$ la fonction de θ définie par :

$$l_n(t_1, ..., t_n, \theta) = \ln L_n(t_1, ..., t_n, \theta)$$

Elle n'a de sens que si θ vérfie $L_n(t_1, ..., t_n, \theta) > 0$
La fonction logarithme népérien étant croissante, l'emv θ^* de θ pour $(t_1, ..., t_n)$ vérfie :

$$\theta^* \in \arg\max L_n(t_1, ..., t_n, \theta) = \arg\max l_n(t_1, ..., t_n, \theta)$$

Équation de vraisemblance

On appelle équation de vraisemblance l'équation en θ :

$$\frac{\partial l_n(t_1, ..., t_n, \theta)}{\partial \theta} = 0$$

Expression analytique de l'emv

Pour envisager d'avoir une expression analytique de l'emv θ^* de θ pour $(t_1, ..., t_n)$ une idée est d'exprimer $L_n(t_1, ..., t_n, \theta)$ en fonction de produits de termes exponentiels/puissances, puis de considérer la fonction de log-vraisemblance $\ln L_n(t_1, ..., t_n, \theta)$ Si cette derniére est dérivable en θ, une condition nécessaire que doit vérifier θ^* est d'être solution de l'équation de vraisemblance. Il faut ensuite vérifier que θ^* est bien un maximum pour $L_n(t_1, ..., t_n, \theta)$

1.4 Les lois usuelles

Le choix du modéle détermine en particulier la forme de la fonction de hasard ; on distinguera notamment les modéles à fonction de hasard monotone des modéles permettant d'obtenir des fonctions de hasard " en cloche " ou en " U " ; ces derniers modéles sont peu usités en assurance, la situation de référence étant un taux de hasard croissant (au sens large) avec le temps.

1.4.1 La loi exponentielle

[5] La spécification la plus simple consiste à poser $h(t) = \lambda$, avec $\lambda > 0$. On en déduit immédiatement que :

$$S(t) = \exp(-\lambda t)$$

Le modèle exponentiel est caractérisé par le fait que les fonctions de survie conditionnelles$S_u, u > 0$ sont exponentielles de même paramètre $\lambda > 0$. Cela signifie que le comportement de la variable aléatoire T après l'instant u ne dépend pas de ce qui est survenu jusqu'en u. Il est également caractérisé par le fait que la fonction de survie est multiplicative, au sens oú :

$$S(u + t) = S(u)S(t)$$

.

On vérifie aisément par un calcul direct que :

$$E(T) = \frac{1}{\lambda}$$

et

$$V(T) = \frac{1}{\lambda^2}$$

. L'estimation du paramètre λ est classique, à partir de l'expression :

$$L(\lambda) = \lambda^n \exp(-\lambda \sum_{i=1}^{n} T_i)$$

qui conduit facilement à :

$$\hat{\lambda} = \frac{n}{\sum_{i=1}^{n} T_i}$$

1.4.2 Loi log-normale :

En théorie des probabilité et statistique , une variable aléatoire T est dite suivre une loi log-normale de paramètre μ et σ^2 si la variable $Y = \ln T$ suit une loi normale d'espérance μ et de variance σ^2.La loi log-normale est aussi appelée loi de Galton la densité de Test alors :[2, 6]

$$f(t) = \frac{1}{\sigma\sqrt{2\pi}}\frac{1}{t}\exp(-\frac{1}{2}(\frac{\log t - \mu}{\sigma})^2)$$

O$\grave{u}$ μ est le paramètre d'échelle et σ est le paramètre de forme.
T admet alors une espérance mathématique et une variance sont :

$$E(T) = \exp(\mu + \frac{\sigma^2}{2})$$

et

$$V(T) = \exp(2\mu + \sigma^2)[\exp(\sigma^2) - 1]$$

1.4.3 Loi log-logistique

T suit une Loi log-logistique de paramètre α et β si sa densité est définie par :[6]

$$f(t, \alpha, \beta) = \frac{\exp(-\frac{t-\alpha}{\beta})}{\beta(1 + \exp(-\frac{t-\alpha}{\beta}))^2}$$

et sa fonction de répartition définit comme suit :

$$F(t, \alpha, \beta) = \frac{1}{1 + \exp(-\frac{t-\alpha}{\beta})}$$

Son espérance et sa variance sont données par les formules suivantes :

$$E(T) = \alpha$$
$$V(T) = \frac{\beta^2\pi^2}{3}$$

1.5 conclusion

nous avons présenté dans ce chapitres des généralités et les différents modéles paramétriques de l'analyse de survie,dans le chapitre suivant nous allons introduire l'alternative de Lehman et les deux modèles d'analyse de la survie que nous proposons.

Chapitre 2

Les modèles d'analyse de la survie

2.1 Alternative de lehman

2.1.1 Modèle

Le modéle est définit par :[12]

$$\tilde{F}(t) = [F(t)]^{\alpha} \tag{2.1}$$

oú : F est la fonction de répartition de la distribution de probabilité de base et α un nombre réel positif.

2.1.2 quelques propriétées

- la fonction de densitée est définit par :[12]

$$\tilde{f}(t) = \tilde{F}^{'}(t) = \alpha[F(t)]^{\alpha-1} f(t) \tag{2.2}$$

avec :$f(t) = F^{'}(t)$
- la fonction de survie :[12]

$$\tilde{S}(t) = 1 - \tilde{F}(t) \tag{2.3}$$

$$\tilde{S}(t) = 1 - [F(t)]^{\alpha} \tag{2.4}$$

- la fonction de hasard :[12]

$$\tilde{h}(t) = \frac{\tilde{f}(t)}{1 - \tilde{F}(t)} \tag{2.5}$$

$$\tilde{h}(t) = \frac{\alpha[F(t)]^{\alpha-1} f(t)}{1-[F(t)]^{\alpha}} \tag{2.6}$$

$$\tilde{h}(t) = \alpha h(t) g(t) \tag{2.7}$$

avec :$g(t) = \dfrac{[F(t)]^{\alpha-1} - [F(t)]^{\alpha}}{1-[F(t)]^{\alpha}}$

2.1.3 Cas particulier

pour $\alpha = 1$ on retrouve les modèles classiques :

Le modèle de Weibull

Loi modélisant un système série[2, 5] (le système est défaillant dès lors qu'un de ses composants l'est), dans ses trois phases de vie

$$T \sim W(k,\lambda), k(forme) > 0, \lambda(\text{échelle}) > 0 \leftrightarrow \begin{cases} f(t) = \frac{k}{\lambda}(\frac{t}{\lambda})^{k-1} \exp(-\frac{t}{\lambda})^k \\ F(t) = 1 - \exp(-\frac{t}{\lambda})^k \\ h(t) = \frac{k}{\lambda}(\frac{t}{\lambda})^{k-1} \\ E(T) = \lambda\Gamma(1+\frac{1}{k}) \end{cases}$$

• $0 < k < 1 \leftrightarrow F$ DFR $\lim_{h\to 0} h(t) = +\infty \, ; \lim_{h\to\infty} h(t) = 0$

• $k = 1 \leftrightarrow$ La loi de T est exponentielle $\frac{1}{\alpha}$

• $k > 1 \leftrightarrow F$ F IFR $\lim_{h\to 0} h(t) = 0 \, ; \lim_{h\to\infty} h(t) = +\infty$

• La loi de Weibull est une généralisation simple du modèle exponentiel, permettant d'obtenir des fonctions de hasard monotone.

• Lorsque $\lambda = 1$ et $k = 2$ ce modèle porte le nom de « modèle de RAYLEIGH» ; il est utilisé en physique pour modéliser la durée de vie de certaines particules. La loi de Weibull apparaît naturellement dans l'étude de la distribution limite du minimum d'un échantillon I.I.D

- **L'estimation des paramètres :**

Soit $f(t) = \frac{k}{\lambda}(\frac{t}{\lambda})^{k-1} \exp -(\frac{t}{\lambda})^k$

on pose : $\theta = \frac{1}{\lambda^k}$

on aurra donc :
$f(t) = \theta k t^{k-1} \exp -\theta t^k$
la fonction de vraisemblance associé a cette fonction de densité est :
$L(t, \theta, k) = \prod_{i=1}^n f(t_i)$
$= \prod_{i=1}^n \theta k t_i^{k-1} \exp -\theta t_i^k$
$= \theta^n k^n \sum_{i=1}^n t_i^{k-1} \exp(-\theta \sum_{i=1}^n x_i^k)$
$\ln L(t, \theta, k) = n \ln \theta + n \ln k + (k-1) \sum_{i=1}^n \ln t_i - \theta \sum_{i=1}^n t_i^k$
on dérive $\ln L(.)$ par-rapport a θ et k

$$\begin{cases} \frac{\partial \ln L}{\partial \theta} = \dfrac{n}{\theta} - \sum_{i=1}^n t_i^k = 0 \\ \frac{\partial \ln L}{\partial k} = \dfrac{n}{k} + \sum_{i=1}^n \ln t_i - \theta \sum_{i=1}^n t_i^k \ln t_i = 0 \end{cases}$$

la résolution de ce système d'équations donne :

$$\begin{cases} \hat{\theta} = \dfrac{n}{\sum_{i=1}^n t_i^k} \\ \hat{k} = \dfrac{n}{\theta \sum_{i=1}^n t_i^k \ln t_i - \sum_{i=1}^n \ln t_i} \end{cases}$$

Loi gamma

Le modèle Gamma est une autre généralisation naturelle du modèle exponentiel :[2, 5, 4] supposons que la durée T,soit la durée d'attente de la réalisation d'un service dans une file d'attente et que la file d'attente soit composée de r serveurs indépendants et identiques qui traitent chacun une partie du service (ils sont donc montés en série). On fait l'hypothèse que la durée de réalisation du traitement de chacun des serveurs est une loi exponentielle de paramètre $\lambda > 0$. Alors la durée globale de service est la somme de r variables exponentielles de même paramètre ; on en déduit que la durée de service est distribuée selon une loi Gamma de paramètre (r, λ) :

$$T \square \Gamma(\lambda, r), \lambda(forme) > 0, r(\text{échelle}) > 0 \leftrightarrow \begin{cases} f(t) = \dfrac{\lambda^r}{\Gamma(r)} t^{r-1} \exp(-\lambda t) \\ F(t) = \dfrac{1}{\Gamma(r)} \int_0^{\lambda t} u^{r-1} \exp(-u) du \\ h(t) = \dfrac{f(t)}{1 - F(t)} \\ E(T) = \dfrac{r}{\lambda} \\ V(T) = \dfrac{r}{\lambda^2} \\ \quad avec \Gamma(r) = \int_0^{+\infty} u^{r-1} \exp(-u) du \end{cases}$$

$$0 < r < 1 \leftrightarrow F \quad \text{DFR} \quad \lim_{h\to 0} h(t) = +\infty\,;\lim_{h\to\infty} h(t) = \frac{1}{\lambda}$$

$r = 1 \leftrightarrow T$ de loi exponentielle de paramètre θ

$$r > 1 \leftrightarrow F \quad \text{IFR} \quad \lim_{h\to 0} h(t) = 0\,;\lim_{h\to\infty} h(t) = \lambda$$

•La loi gamma est une généralisation de la loi exponentielle (qui correspond à $r = 1$)
• Lorsque r est entier la loi gamma s'appelle Loi d'Erlang

2.2 Alternative de lehman généralisé

2.2.1 Modèle

Soient $F_1(x)$ et $F_2(x)$ deux fonctions de répartition de deux lois de probabilités de bases connues, le modèle est définit par la fonction G suivante :

$$G(x) = [p_1F_1(x) + p_2F_2(x)]^{\alpha} \tag{2.8}$$

avec :α un nombre réel positif, $0 \leq p_1 \leq 1$, $0 \leq p_2 \leq 1$ et $p_1 + p_2 = 1$. dans la suite on pose $\alpha = 1$

2.2.2 Propriétés

A• **Propriété 1**

$G(x)$ est une fonction de répartition

$A - 1$• **Preuve**

on montre que (2.8) vérifie les propriétés d'une fonction de répartition :

— $G(x)$ est continue car$F_1(x)$ et $F_2(x)$ sont deux fonctions continues
— $0 \leq G(x) \leq 1$ car :
  $F_1(x) \leq 1$,donc :

  $p_1F_1(x) \leq p_1 \leq 1$
  et : $F_2(x) \leq 1$, donc
  $p_2F_2(x) \leq p_2 \leq 1$
  ce qui implique :

$$p_1F_1(x) + p_2F_2(x) \leq p_1 + p_2 \tag{2.9}$$

et on a aussi : $p_2 = 1 - p_1$ donc :
$p_1 F_1(x) + p_2 F_2(x) \leq 1$

$$G(x) \leq 1 \tag{2.10}$$

$F_1(x) \geq 0$ et $F_2(x) \geq 0$ alors :
$p_1 F_1(x) \geq 0$ et $p_2 F_2(x) \geq 0$ donc :
$p_1 F_1(x) + p_2 F_2(x) \geq 0$ ce qui implique :

$$G(x) \geq 0 \tag{2.11}$$

de (2.10) et (2.11) on aurra :

$$0 \leq G(x) \leq 1 \tag{2.12}$$

— $G'(x) = p_1 f_1(x) + p_2 f_2(x) \geq 0$ car $f_1(x) \geq 0$ et $f_2(x) \geq 0$ (des fonction de densité et $0 \leq p_1 \leq 1$ et $0 \leq p_2 \leq 1$.)
— $G(+\infty) = p_1 F_1(+\infty) + p_2 F_2(+\infty) = 1$ (car$F_1(+\infty) = 1$ et $F_2(+\infty) = 1$ et $p_1 + p_2 = 1$)

— $G(-\infty) = p_1 F_1(-\infty) + p_2 F_2(-\infty) = 0$ (car$F_1(-\infty) = 0$ et $F_2(-\infty) = 0$)

$B\bullet$ Propriété 2 (la fonction de hasard)

$h(x) = \dfrac{g(x)}{1 - G(x)}$

$$h(x) = \frac{p_1 f_1(x) + p_2 f_2(x)}{1 - p_1 F_1(x) - p_2 F_2(x)} \tag{2.13}$$

et puisque on a $p_2 = 1 - p_1$ alors :

$$h(x) = \frac{p_1(f_1(x) - f_2(x)) + f_2(x)}{1 - p_1(F_1(x) - F_2(x)) - F_2(x)} \tag{2.14}$$

$C\bullet$ Propriété 3 (l'ordre statistique)

la forme générale de l'ordre statistique :

$$g_x(r) = \frac{n!}{(r-1)!(n-r)!} g(x_r) G(x_r)^{r-1} (1 - G(x_r))^{n-r}$$

1 ére ordre statistique : r=1

$$g_1(x_1) = n g(x_1)(1 - G(x_1))^{n-1} = n(p_1 f_1(x) + p_2 f_2(x))[1 - p_1 F_1(x) - p_2 F_2(x)]^{n-1}$$

n éme ordre statistique : r=n

$$g_n(x_n) = ng(x_n)(G(x_n))^{n-1} = n(p_1 f_1(x_n) + p_2 f_2(x_n))[p_1 F_1(x_n) + p_2 F_2(x_n)]^{n-1}$$

$D\bullet$ **Propriété 4 (la fonction caractéristique)**
généralement la fonction caractéristique est définit par :

$$\phi_x(t) = \int_0^{+\infty} g(x)\exp(itx)dx$$

dans notre cas :
$\phi_x(t) = p_1(\frac{k_1}{\lambda_1}) \int_0^{+\infty}(\frac{x_i}{\lambda_1})^{k_1-1}\exp(itx - (\frac{x_i}{\lambda_1})^{k_1}) + p_2(\frac{k_2}{\lambda_2})$
$\int_0^{+\infty}(\frac{x_i}{\lambda_2})^{k_2-1}\exp(itx - (\frac{x_i}{\lambda_2})^{k_2})dx$

Exemple

soit $f_1(x), f_2(x)$ deux fonctions de densités de deux loi de Weibull de paramètres différents
et soit $F_1(x), F_2(x)$ leurs fonctions de répartition respectives :
$f_1(x) = (\frac{k_1}{\lambda_1})(\frac{x}{\lambda_1})^{k_1-1}\exp -(\frac{x}{\lambda_1})^{k_1}$
$F_1(x) = 1 - \exp -(\frac{x}{\lambda_1})^{k_1}$
$f_2(x) = (\frac{k_2}{\lambda_2})(\frac{x}{\lambda_2})^{k_2-1}\exp -(\frac{x}{\lambda_2})^{k_2}$
$F_2(x) = 1 - \exp -(\frac{x}{\lambda_2})^{k_2}$
on remplace dans$G(x)$ et $g(x)$ et on trouve :

$$G(x) = p_1(1 - \exp(-\frac{x}{\lambda_1})^{k_1}) + p_2(1 - \exp(-\frac{x}{\lambda_2})^{k_2} \tag{2.15}$$

$$g(x) = p_1(\frac{k_1}{\lambda_1})(\frac{x}{\lambda_1})^{k_1-1}\exp -(\frac{x}{\lambda_1})^{k_1} + p_2(\frac{k_2}{\lambda_2})(\frac{x}{\lambda_2})^{k_2-1}\exp -(\frac{x}{\lambda_2})^{k_2} \tag{2.16}$$

et puisque :$p_2 = 1 - p_1$ donc on peut écrire :

$$G(x) = p_1(\exp -(\frac{x}{\lambda_2})^{k_2} - \exp(-\frac{x}{\lambda_1})^{k_1}) + 1 - \exp -(\frac{x}{\lambda_2})^{k_2} \tag{2.17}$$

la fonction de hasard correspondante est :

$$h(x)=\frac{g(x)}{1-G(x)}$$

alors :

$$h(x)=\frac{p_1(\frac{k_1}{\lambda_1})(\frac{x}{\lambda_1})^{k_1-1}\exp-(\frac{x}{\lambda_1})^{k_1}+p_2(\frac{k_2}{\lambda_2})(\frac{x}{\lambda_2})^{k_2-1}\exp-(\frac{x}{\lambda_2})^{k_2}}{1-p_1(1-\exp-(\frac{x}{\lambda_1})^{k_1})-p_2(1-\exp-(\frac{x}{\lambda_2})^{k_2}}$$

l'estimation des paramètres par la méthode du maximum de vraisemblance

Soit $L(x_i,\lambda_1,\lambda_2,k_1,k_2)$ la fonction de vraisemblance,avec :

$$L(x_i,\lambda_1,\lambda_2,k_1,k_2)=\prod_{i=1}^{n}g(x_i,\lambda_1,\lambda_2,k_1,k_2)$$

$\prod_{i=1}^{n}g(x_i,\lambda_1,\lambda_2,k_1,k_2)=\prod_{i=1}^{n}[p_1f_1(x_i,\lambda_1,k_1)+p_2f_2(x_i,\lambda_2,k_2)]$

$\ln\prod_{i=1}^{n}g(x_i,\lambda_1,\lambda_2,k_1,k_2)=\ln\prod_{i=1}^{n}[p_1f_1(x_i,\lambda_1,k_1)+p_2f_2(x_i,\lambda_2,k_2)]$

$=\sum_{i=1}^{n}\ln[p_1f_1(x_i,\lambda_1,k_1)+p_2f_2(x_i,\lambda_2,k_2)]$

Maintenant on dérive la fonction de vraisemblance par-rapport a ses paramètres :

$\frac{\partial\sum_{i=1}^{n}ln[p_1f_1(x_i,\lambda_1,k_1)+p_2f_2(x_i,\lambda_2,k_2)]}{\partial\lambda_1}$

$=\sum_{i=1}^{n}[\frac{\partial ln[p_1f_1(x_i,\lambda_1,k_1)+p_2f_2(x_i,\lambda_2,k_2)]}{\partial\lambda_1}]$

$=\sum_{i=1}^{n}[\frac{\partial ln[p_1f_1(x_i,\lambda_1,k_1)]}{\partial\lambda_1}]$

$=\sum_{i=1}^{n}[\frac{\partial[p_1f_1(x_i,\lambda_1,k_1)+p_2f_2(x_i,\lambda_2,k_2)]}{\partial\lambda_1}]\dfrac{1}{p_1f_1(x_i,\lambda_1,k_1)+p_2f_2(x_i,\lambda_2,k_2)}$

$=\sum_{i=1}^{n}(p_1\frac{\partial(p_1f_1(x_i,\lambda_1,k_1)}{\partial\lambda_1})\dfrac{1}{p_1f_1(x_i,\lambda_1,k_1)+p_2f_2(x_i,\lambda_2,k_2)}$

$$f_1(x_i,\lambda_1,k_1) = \frac{k_1 x_i^{k_1-1}}{{\lambda_1}^{k_1}} \exp -\frac{x_i^{k_1}}{\lambda_1^{k_1}}$$
$$= \exp[\ln(k_1 x_i^{k_1-1}) - k_1 \ln\lambda_1 - \frac{x_i^{k_1}}{\lambda_1^{k_1}}]$$
$$\tfrac{\partial f_1(x_i,\lambda_1,k_1)}{\partial\lambda_1} = -\frac{k_1}{\lambda_1} - x_i k_1(\frac{-k_1}{\lambda_1^{k_1+1}}) \exp[\ln(k_1 x_i^{k_1-1}) - k_1 \ln\lambda_1 - \frac{x_i^{k_1}}{\lambda_1^{k_1}}]$$
$$= -\frac{k_1}{\lambda_1} + \frac{x_i k_1^2}{\lambda_1^{k_1+1}} \exp[\ln(k_1 x_i^{k_1-1}) - k_1 \ln\lambda_1 - \frac{x_i^{k_1}}{\lambda_1^{k_1}}]$$
$$= \frac{-k_1\lambda_1^{k_1} + x_i k_1^2}{\lambda_1^{k_1+1}} \exp[\ln(k_1 x_i^{k_1-1}) - k_1 \ln\lambda_1 - \frac{x_i^{k_1}}{\lambda_1^{k_1}}]$$
$$= \frac{-k_1^2 x_i^{k_1-1}\lambda_1^{k_1} + k_1^3 x_i^{k_1}}{\lambda_1^{k_1+1}} \exp -\frac{x_i^{k_1}}{\lambda_1^{k_1}}$$

donc :

$$\tfrac{\partial \sum_{i=1}^n \ln g(x_i,\lambda_1,\lambda_2,k_1,k_2)}{\partial\lambda_1} = \textstyle\sum_{i=1}^n [p_1(\frac{-k_1^2 x_i^{k_1-1}\lambda_1^{k_1} + k_1^3 x_i^{k_1}}{\lambda_1^{k_1+1}} \exp -\frac{x_i^{k_1}}{\lambda_1^{k_1}})$$
$$*\,(\frac{1}{p_1(\frac{k_1}{\lambda_1})(\frac{x_i}{\lambda_1})^{k_1-1} \exp -(\frac{x_i}{\lambda_1})^{k_1} + p_2(\frac{k_2}{\lambda_2})(\frac{x_i}{\lambda_2})^{k_2-1} \exp -(\frac{x_i}{\lambda_2})^{k_2}})]$$

on dérive par-rapport a k_1

$$f_1(x_i,\lambda_1,k_1) = \frac{k_1 x_i^{k_1-1}}{{\lambda_1}^{k_1}} \exp -\frac{x_i^{k_1}}{\lambda_1^{k_1}}$$
$$= \exp[\ln k_1 + (k_1-1) ln x_i - k_1 \ln\lambda_1 - (\frac{x_i}{\lambda_1})^{k_1}]$$

$$\frac{\partial f_1(x_i,\lambda_1,k_1)}{\partial k_1} = [\frac{1}{k_1} + \ln x_i - ln\lambda_1 - \frac{\partial \exp(k_1 \ln\frac{x_i}{\lambda_1})}{\partial k_1}] \exp[\ln k_1 + (k_1-1) ln x_i - k_1 \ln\lambda_1 - (\frac{x_i}{\lambda_1})^{k_1}]$$

$$= [\frac{1}{k_1} + \ln(\frac{x_i}{\lambda_1}) - \ln(\frac{x_i}{\lambda_1}) \exp(k_1 \ln(\frac{x_i}{\lambda_1}))] \exp[\ln k_1 + (k_1-1) ln x_i - k_1 \ln\lambda_1 - (\frac{x_i}{\lambda_1})^{k_1}]$$
$$= [\frac{1}{k_1} + \ln(\frac{x_i}{\lambda_1}) - \ln(\frac{x_i}{\lambda_1})(\frac{x_i}{\lambda_1})^{k_1}] \exp[\ln k_1 + (k_1-1) ln x_i - k_1 \ln\lambda_1 - (\frac{x_i}{\lambda_1})^{k_1}]$$
$$= [\frac{1}{k_1} + \ln(\frac{x_i}{\lambda_1}) - \ln(\frac{x_i}{\lambda_1})(\frac{x_i}{\lambda_1})^{k_1}] \frac{k_1 x_i^{k_1-1}}{{\lambda_1}^{k_1}} \exp -\frac{x_i^{k_1}}{\lambda_1^{k_1}}$$

$$= \frac{x_i^{k_1-1}}{{\lambda_1}^{k_1}} + \ln(\frac{x_i}{\lambda_1})\frac{k_1 x_i^{k_1-1}}{{\lambda_1}^{k_1}} - \ln(\frac{x_i}{\lambda_1}) k_1 x_i^{-1} (\frac{x_i}{\lambda_1})^{2k_1} \exp -(\frac{x_i}{\lambda_1})^{k_1}$$

donc :

$$\frac{\partial \sum_{i=1}^{n} [\ln g(x_i,\lambda_1,\lambda_2,k_1,k_2)}{\partial k_1} = \sum_{i=1}^{n} [p_1(\frac{x_i^{k_1-1}}{{\lambda_1}^{k_1}} + \ln(\frac{x_i}{\lambda_1})\frac{k_1 x_i^{k_1-1}}{{\lambda_1}^{k_1}} - \ln(\frac{x_i}{\lambda_1}) k_1 x_i^{-1} (\frac{x_i}{\lambda_1})^{2k_1} \exp -(\frac{x_i}{\lambda_1})^{k_1})$$

$$* (\frac{1}{p_1(\frac{k_1}{\lambda_1})(\frac{x_i}{\lambda_1})^{k_1-1} \exp -(\frac{x_i}{\lambda_1})^{k_1} + p_2(\frac{k_2}{\lambda_2})(\frac{x_i}{\lambda_2})^{k_2-1} \exp -(\frac{x_i}{\lambda_2})^{k_2}})]$$

et puisque on a utilisé deux loi de Weibull de paramètres différents,donc de la même manière on trouve :

$$\frac{\partial \sum_{i=1}^{n} \ln g(x_i,\lambda_1,\lambda_2,k_1,k_2)}{\partial \lambda_2} = \sum_{i=1}^{n} [p_2(\frac{-k_2^2 x_i^{k_2-1} \lambda_2^{k_2} + k_2^3 x_i^{k_2}}{\lambda_2^{k_2+1}} \exp -\frac{x_i^{k_2}}{\lambda_2^{k_2}})$$

$$* (\frac{1}{p_1(\frac{k_1}{\lambda_1})(\frac{x_i}{\lambda_1})^{k_1-1} \exp -(\frac{x_i}{\lambda_1})^{k_1} + p_2(\frac{k_2}{\lambda_2})(\frac{x_i}{\lambda_2})^{k_2-1} \exp -(\frac{x_i}{\lambda_2})^{k_2}})]$$

et :

$$\frac{\partial \sum_{i=1}^{n} [\ln g(x_i,\lambda_1,\lambda_2,k_1,k_2)}{\partial k_2} = \sum_{i=1}^{n} [p_2(\frac{x_i^{k_2-1}}{{\lambda_2}^{k_2}} + \ln(\frac{x_i}{\lambda_2})\frac{k_2 x_i^{k_2-1}}{{\lambda_2}^{k_2}} - \ln(\frac{x_i}{\lambda_2}) k_2 x_i^{-1} (\frac{x_i}{\lambda_2})^{2k_2} \exp -(\frac{x_i}{\lambda_2})^{k_2})$$

$$* (\frac{1}{p_1(\frac{k_1}{\lambda_1})(\frac{x_i}{\lambda_1})^{k_1-1} \exp -(\frac{x_i}{\lambda_1})^{k_1} + p_2(\frac{k_2}{\lambda_2})(\frac{x_i}{\lambda_2})^{k_2-1} \exp -(\frac{x_i}{\lambda_2})^{k_2}})]$$

on aurra donc le système d'équation suivant :

$$(p)\begin{cases}
\sum_{i=1}^{n}[p_1(\frac{-k_1^2 x_i^{k_1-1}\lambda_1^{k_1}+k_1^3 x_i^{k_1}}{\lambda_1^{k_1+1}}\exp-\frac{x_i^{k_1}}{\lambda_1^{k_1}}) \\
*(\frac{1}{p_1(\frac{k_1}{\lambda_1})(\frac{x_i}{\lambda_1})^{k_1-1}\exp-(\frac{x_i}{\lambda_1})^{k_1}+p_2(\frac{k_2}{\lambda_2})(\frac{x_i}{\lambda_2})^{k_2-1}\exp-(\frac{x_i}{\lambda_2})^{k_2}})]=0 \\
\sum_{i=1}^{n}[p_1(\frac{x_i^{k_1-1}}{{\lambda_1}^{k_1}}+\ln(\frac{x_i}{\lambda_1})\frac{k_1 x_i^{k_1-1}}{{\lambda_1}^{k_1}}-\ln(\frac{x_i}{\lambda_1})k_1 x_i^{-1}(\frac{x_i}{\lambda_1})^{2k_1}\exp-(\frac{x_i}{\lambda_1})^{k_1}) \\
*(\frac{1}{p_1(\frac{k_1}{\lambda_1})(\frac{x_i}{\lambda_1})^{k_1-1}\exp-(\frac{x_i}{\lambda_1})^{k_1}+p_2(\frac{k_2}{\lambda_2})(\frac{x_i}{\lambda_2})^{k_2-1}\exp-(\frac{x_i}{\lambda_2})^{k_2}})]=0 \\
\sum_{i=1}^{n}[p_2(\frac{-k_2^2 x_i^{k_2-1}\lambda_2^{k_2}+k_2^3 x_i^{k_2}}{\lambda_2^{k_2+1}}\exp-\frac{x_i^{k_2}}{\lambda_2^{k_2}}) \\
*(\frac{1}{p_1(\frac{k_1}{\lambda_1})(\frac{x_i}{\lambda_1})^{k_1-1}\exp-(\frac{x_i}{\lambda_1})^{k_1}+p_2(\frac{k_2}{\lambda_2})(\frac{x_i}{\lambda_2})^{k_2-1}\exp-(\frac{x_i}{\lambda_2})^{k_2}})]=0 \\
\sum_{i=1}^{n}[p_2(\frac{x_i^{k_2-1}}{{\lambda_2}^{k_2}}+\ln(\frac{x_i}{\lambda_2})\frac{k_2 x_i^{k_2-1}}{{\lambda_2}^{k_2}}-\ln(\frac{x_i}{\lambda_2})k_2 x_i^{-1}(\frac{x_i}{\lambda_2})^{2k_2}\exp-(\frac{x_i}{\lambda_2})^{k_2}) \\
*(\frac{1}{p_1(\frac{k_1}{\lambda_1})(\frac{x_i}{\lambda_1})^{k_1-1}\exp-(\frac{x_i}{\lambda_1})^{k_1}+p_2(\frac{k_2}{\lambda_2})(\frac{x_i}{\lambda_2})^{k_2-1}\exp-(\frac{x_i}{\lambda_2})^{k_2}})]=0
\end{cases}$$

Ce système est compliqué,on peut pas le résoudre directement car les équations sont non linéaire,pour cela nous allons faire appelle à des méthodes numériques pour la résolution de ce système.

2.3 Modèle de Mélange

2.3.1 Modèle

Soit $F_1(x)$ et $F_2(x)$ deux fonctions de répartition,le modèle est définit par la fonction G suivante :

$$G(x) = \frac{p_1F_1(x) + p_2F_2(x)}{2 - F_1(x)} \tag{2.18}$$

avec : $0 \leq p_1 \leq 1$, $0 \leq p_2 \leq 1$ et $p_1 + p_2 = 1$.

2.3.2 Propriétés

$A\bullet$ **Propriété 1**

$G(x)$ est une fonction de répartition en effet :

1) G(x) continue car$F_1(x)$ et$F_2(x)$ sont continues et $2 - F_1(x)$ est continue donc G(x) est continue

2) $0 \leq G(x) \leq 1$ car :

$0 \leq F_1(x) \leq 1$ et $0 \leq F_2(x) \leq 1$ $\forall x$

$\Rightarrow$

$$0 \leq p_1F_1(x) \leq p_1 \tag{2.19}$$

$$0 \leq p_2F_2(x) \leq p_2 \tag{2.20}$$

et on a :$0 \leq p_1 \leq 1$, $0 \leq p_2 \leq 1$ donc :

$$0 \leq p_1F_1(x) \leq p_1 \leq 1 \tag{2.21}$$

$$0 \leq p_2F_2(x) \leq p_2 \leq 1 \tag{2.22}$$

donc :(2.21) + (2.22) $\Rightarrow 0 \leq P_1F_1(x) + p_2F_2 \leq p_1 + p_2 = 1$

$\Rightarrow$

$$0 \leq p_1F_1(x) + P_2F_2(x) \leq 1 \tag{2.23}$$

on a :$0 \leq F_1(x) \leq 1$ donc :

$-1 \leq -F_1(x) \leq 0$

$\Rightarrow 1 \leq 2 - F_1(x) \leq 2$

$$\Rightarrow \frac{1}{2} \leq \frac{1}{2 - F_1(x)} \leq 1 \tag{2.24}$$

on multiple (2.23) et (2.24) on obtient :

$0 \leq \dfrac{p1F_1(x) + p_2F_2(x)}{2 - F_1} \leq 1$
$\Rightarrow 0 \leq G(x) \leq 1$
3) $G\prime(x) = \dfrac{p_1f_1(x) + p_2f_2[2 - F_1(x)] - (-f_1)[p_1F_1(x) + p_2F_2(x)]}{(2 - F_1(x))^2}$
$G\prime(x) = \dfrac{2p_1f1(x) + f1(x)p2F2(x) + 2p2f2(x) - p2f2(x)F1(x)}{(2 - F1(x))^2}$
$= \dfrac{f_1(x)[2p_1 - p_2F_2(x) - 2p_1F_1(x)] + f_2(x)[2p_2 - p_2F_1(x)]}{(2 - F_1(x))^2} \geq 0$
$\Rightarrow G\prime(x) \geq 0$
4-A) $G(-\infty) = \dfrac{p_1F_1(-\infty) + p_2F_2(-\infty)}{2 - F_1(-\infty)} = 0$ (car$F_1(-\infty) = 0$ et $F_2(-\infty) = 0$)

4-B) $G(+\infty) = \dfrac{p_1F_1(+\infty) + p_2F_2(+\infty)}{2 - F_1(+\infty)} = \dfrac{p_1 + p_2}{2 - 1} = 1$ (car$F_1(+\infty) = 1$ et $F_2(+\infty) = 1$ et $p_1 + p_2 = 1$)

d'ou G(x) est une fonction de répartition

$B\bullet$ **Propriété 2 (La fonction de densité)**
on pose $F_1(x) \neq F_2(x)$ donc :
La fonction de densité correspondante $\grave{a}$ la formule est :
$g(x) = \dfrac{(p_1f_1(x) + p_2f_2)[2 - F_1(x)] - (-f_1)[p_1F_1(x) + p_2F_2(x)]}{(2 - F_1(x))^2}$

$$g(x) = \frac{f1[2p_1 - p_2F_2(x) - 2p_1F_1] + f_2(x)[2p_2 - p_2F_1(x)]}{(2 - F_1(x))^2} \tag{2.25}$$

$C\bullet$ **Propriété 3 (La fonction de hasard)**
la fonction hasard est définit par :

$h(x) = \dfrac{f_1[2p_1 - p_2F_2 - 2p_1F_1] + f_2[2p_2 - p_2F_1](2 - F_1)}{(2 - F_1)^2[2 - p_1F_1 - p_2F_2]}$

$$h(x) = \frac{f_1[2p_1 - p_2F_2 - 2p_1F_1] + f_2[2p_2 - F_1P_2]}{(2 - F_1)[2 - F_1(1 + p_1) - p_2F_2]} \tag{2.26}$$

2.3.3 Cas particulier

Si $F_1(x)$ =$F_2(x)$ alors :
1) la fonction de répartition est :

$$G(x) = \frac{F(x)}{2 - F(x)} \tag{2.27}$$

2) La fonction de densité correspondante est donné par :

$$g(x) = \frac{2f(x)}{(2 - F(x))^2} \tag{2.28}$$

3) la fonction de hasard correspondante est :
$h(x) = \dfrac{g(x)}{1 - G(x)}$
$= \dfrac{2f(x)}{(2 - F(x))^2} \dfrac{2 - F(x)}{2(1 - F(x))}$
donc :

$$h(x) = \frac{f(x)}{(2 - F(x))(1 - F(x))} \tag{2.29}$$

L'ordre statistique

on sait que la forme générale de l'ordre statistique est définit par :

$$g(x_r) = \frac{n!}{(r-1)!(n-r)!} g(x_r)[G(x_r)]^{r-1}[1 - G(x_r)]^{n-r}$$

dans notre cas :

$$g(x_r) = \frac{n!}{(r-1)!(n-r)!} \frac{2f(x_r)}{[2 - F(x_r)]^2} [\frac{F(x_r)}{2 - F(x_r)}]^{r-1}[1 - \frac{F(x_r)}{2 - F(x_r)}]^{n-r}$$

le premier ordre x_1 donné par

$$g_1(x_1) = ng_1(x_1)[1 - G_1(x_1)]^{n-1}$$

$$g_1(x_1) = n\frac{2f(x_1)}{(2 - F(x_1))^2}[1 - \frac{F(x_1)}{2 - F(x_1)}]^{n-1} \tag{2.30}$$

le n'iéme ordre donné par : $g_n(x_n) = ng(x_n)[G(x_n)]^{n-1}$

$$g_n(x_n) = n\frac{2f(x_n)}{(2 - F(x_n))^2}[\frac{F(x_n)}{2 - F(x_n)}]^2 \tag{2.31}$$

Exemple

soit la distribution de Wiebull définit par :

$$\begin{cases} f(x) = \frac{k}{\lambda}(\frac{x}{\lambda})^{k-1} \exp(-\frac{x}{\lambda})^{k}, x > 0 \\ F(x) = 1 - \exp(-\frac{x}{\lambda})^{k} \end{cases}$$

avec :f(x) est la fonction de densité et F(x) est la fonction de répartition
on remplace dans(2.27) et(2.28) on obtient :

$$G(x) = \frac{1 - \exp(-\frac{x}{\lambda})^{k}}{2 - (1 - \exp(-\frac{x}{\lambda})^{k})}$$

alors :

$$G(x) = \frac{1 - \exp(-\frac{x}{\lambda})^{k}}{1 + \exp(-\frac{x}{\lambda})^{k}} \tag{2.32}$$

et :

$$g(x) = \frac{2(\frac{k}{\lambda}(\frac{x}{\lambda})^{k-1} \exp(-\frac{x}{\lambda})^{k})}{(1 + \exp(-\frac{x}{\lambda})^{k})^{2}}$$

la fonction de hasard est :

$$h(x) = \frac{\frac{k}{\lambda}(\frac{x}{\lambda})^{k-1} \exp(-\frac{x}{\lambda})^{k}}{2 + (1 - \exp(-\frac{x}{\lambda})^{k})^{2} - 3(1 - \exp(-\frac{x}{\lambda})^{k})} \tag{2.33}$$

Le moment ordinaire :

le moment ordinaire est définit par :

$$\mu\prime = \int_{0}^{\infty} x^{r} g(x) dx$$

$$\mu\prime = \int_{0}^{\infty} x^{r} \frac{2(\frac{k}{\lambda}(\frac{x}{\lambda})^{k-1} \exp(-\frac{x}{\lambda})^{k})}{(1 + \exp(-\frac{x}{\lambda})^{k})^{2}} dx$$

$$\mu\prime = 2\frac{k}{\lambda^k}\int_0^\infty \frac{x^{k+r-1}\exp(-\dfrac{x}{\lambda})^k}{(1+\exp(-\dfrac{x}{\lambda})^k)^2}dx \tag{2.34}$$

La fonction caractéristique :

$\phi_x(t) = \int_{-\infty}^{+\infty}\exp(itx)f(x)dx = \int_0^{+\infty}\exp(itx)f(x)dx$

$$\phi_x(t) = \frac{k}{\lambda^k}\int_0^{+\infty}\frac{x^{k-1}\exp(itx-(\dfrac{x}{\lambda})^k)}{(1+\exp(-\dfrac{x}{\lambda})^k)^2}dx \tag{2.35}$$

l'estimation des paramètres par la méthode du Maximum de vraisemblance

on a :

$$g(x) = \frac{2\dfrac{k}{\lambda}(\dfrac{x}{\lambda})^{k-1}\exp(-\dfrac{x}{\lambda})^k}{(1+\exp(-\dfrac{x}{\lambda})^k)^2}$$

et

$$f(x) = \frac{k}{\lambda}(\frac{x}{\lambda})^{k-1}\exp(-\frac{x}{\lambda})$$

soit $l(x,k,\lambda)$ la fonction de vraisemblance associé a notre modèle :

$$l(x,k,\lambda) = \prod_{i=1}^{n} g(x_i)$$

$\prod_{i=1}^n g(x_i) = \prod_{i=1}^n \dfrac{2f(x_i)}{(2-F(x_i))^2}$

$\log\prod_1^n g(xi) = \log\prod_{i=1}^n \dfrac{2f(xi)}{(2-F(x_i))^2}$

$= \sum_{i=1}^n \log\dfrac{2f(x_i)}{(2-F(x_i))^2}$

$$= \sum_{i=1}^{n} \log[\frac{2\frac{k}{\lambda^k} x_i^{k-1} * \exp^(-\frac{x_i}{\lambda})^k}{(1+\exp(-\frac{x_i}{\lambda})^k)^2}]$$

$$= \sum_{i=1}^{n}[\log(2\frac{k}{\lambda^k} x_i^{k-1} \exp^(-\frac{x_i}{\lambda})^k) - 2\log(1+\exp(-\frac{x_i}{\lambda})^k)]$$

$$\bullet \frac{\partial \log L(x)}{\partial \lambda} = \sum_{i=1}^{n}[\frac{\frac{-k^2 x_i^{k-3}\lambda^k + k^3 x_i^k \exp(-\frac{x_i}{\lambda})^k}{\lambda^{k+1}}}{2\frac{k}{\lambda^k} x_i^{k-1}\exp(-\frac{x_i}{\lambda})^k} - \frac{2k\lambda^{k-1}x_i^k}{1+\exp(-\frac{x_i}{\lambda})^k}]$$

$$= \sum_{i=1}^{n}[\frac{(-k^2 x_i^{k-1}\lambda^k + k^3 x_i^k)\exp(-\frac{x_i}{\lambda})^k}{\lambda^{k+1} 2\frac{k}{\lambda^k} x_i^{k-1}\exp(-\frac{x_i}{\lambda})^k} - \frac{2k\lambda^{k-1}x_i^k}{1+\exp(-\frac{x_i}{\lambda})^k}]$$

$$= \sum_{i=1}^{n}[\frac{(-k^2 x_i^{k-1}\lambda^k + k^3 x_i^k)\exp(-\frac{x_i}{\lambda})^k(1+\exp(-\frac{x_i}{\lambda})^k)}{(2k\lambda x^{k-1}\exp(-\frac{x_i}{\lambda})^k)(1+\exp(-\frac{x_i}{\lambda})^k)}]$$

$$-\frac{2k\lambda^{k-1}x_i^k(2k\lambda x^{k-1}\exp(-\frac{x}{\lambda})^k)}{(2k\lambda x^{k-1}\exp(-\frac{x_i}{\lambda})^k)(1+\exp(-\frac{x_i}{\lambda})^k)}$$

$$= \sum_{i=1}^{n}[\frac{(-k^2 x_i^{k-1}\lambda^k + k^3 x_i^k)\exp(-\frac{x_i}{\lambda})^k + (-k^2 x_i^{k-1}\lambda^k + k^3 x_i^k)\exp(-\frac{x_i}{\lambda})^k}{2k\lambda x_i^{k-1}\exp(-\frac{x_i}{\lambda})^k(1+\exp(-\frac{x_i}{\lambda})^k)}]$$

$$-\frac{4k^2\lambda^{2k-1}x_i^{2k-1}\exp(-\frac{x_i}{\lambda})^k(2k\lambda x_i^{k-1}\exp(-\frac{x_i}{\lambda})^k)}{2k\lambda x_i^{k-1}\exp(-\frac{x_i}{\lambda})^k(1+\exp(-\frac{x}{\lambda})^k))}$$

$$= \sum_{i=1}^{n}[\frac{(-k^2 x_i^{k-1}\lambda^k + k^3 x_i^k)\exp(-\frac{x_i}{\lambda})^k}{2k\lambda x_i^{k-1}\exp(-\frac{x_i}{\lambda})^k(1+\exp(-\frac{x_i}{\lambda})^k)}$$

$$-\frac{\exp(-2\frac{x_i}{\lambda})^k(-k^2 x_i^{k-1}\lambda^k + k^3 x_i^k - 8k^3\lambda^{2k+1}x_i^{3k+1})}{2k\lambda x_i^{k-1}\exp(-\frac{x_i}{\lambda})^k(1+\exp(-\frac{x_i}{\lambda})^k)}] = 0$$

$$\bullet \frac{\partial \log L(x)}{\partial k} = \sum_{i=1}^{n}[(\frac{1}{k} + \ln(\frac{x_i}{\lambda}) - \ln(\frac{x_i}{\lambda})(\frac{x_i}{\lambda})^k(\frac{kx_i^{k-1}}{\lambda^k}))\exp(-\frac{x_i}{\lambda})^k$$

$$-\frac{2(\ln(\frac{x_i}{\lambda})\exp(k) * \ln(\frac{x_i}{\lambda})\exp(-\frac{x_i}{\lambda})^k)}{1+\exp(-\frac{x_i}{\lambda})^k}] = 0$$

2.3.4 cas générale :

On pose $F_1(x) \neq F_2(x)$ alors :
1) la fonction de répartition est :

$$G(x) = \frac{p_1F_1(x) + p_2F_2(x)}{2 - F_1(x)} \tag{2.36}$$

2) La fonction de densité correspondante est :

$$g(x) = \frac{(p_1f_1(x) + p_2f_2)[2 - F_1(x)] - (-f_1)[p_1F_1(x) + p_2F_2(x)]}{(2 - F_1(x))^2}$$

$$g(x) = \frac{f1[2p_1 - p_2F_2(x) - 2p_1F_1] + f_2(x)[2p_2 - p_2F_1(x)]}{(2 - F_1(x))^2} \tag{2.37}$$

3) la fonction de hasard est :

$$h(x) = \frac{f_1(x)[2p_1 - p_2F_2(x) - 2p_1F_1(x)] + f_2(x)[2p_2 - p_2F_1(x)](2 - F_1(x))}{(2 - F_1(x))^2[2 - F_1(x) - p_1F_1(x) - p_2F_2(x)]}$$

$$h(x) = \frac{f_1(x)[2p_1 - p_2F_2(x) - 2p_1F_1(x)] + f_2(x)[2p_2 - F_1(x)P_2]}{(2 - F_1(x))[2 - F_1(x)(p_1 + 1) - p_2F_2(x)]}$$

L'ordre statistique de la fonction de densité

le premier ordre statistique :

$$g_1(x_1) = ng_1(x_1)[1 - G_1(x_1)]^{n-1}$$

$$g_1(x_1) = n\frac{f_1(x_1)[2p_1 - p_2F_2(x_1) - 2p_1F_1(x_1)] + f_2(x_1)[2p_2 - p_2F_1(x_1)]}{(2 - F_1(x_1))^2}$$

$$[1 - \frac{p_1 * F_1(x_1) + p_2 * F_2(x_1)}{2 - F_1(x_1)}]^{n-1}$$

le n'iéme ordre donné par :

$$g_n(x_n) = ng(x_n)[G(x_n)]^{n-1}$$

$$g_n(x_n) = n\frac{f_1(x_n)[2p_1 - p_2F_2(x_n) - 2p_1F_1(x_n)] + f_2(x_n)[2p_2 - p_2F_1(x_n)]}{(2 - F_1(x_n))^2}$$

$$*[\frac{p_1F_1(x_n) + p_2F_2(x_n)}{2 - F_1(x_n)}]^{n-1}$$

car la forme générale de l'ordre statistique est :

$$g(x_r) = \frac{n!}{(r-1)!(n-r)!}g(x_r)[G(x_r)]^{r-1}[1 - G(x_r)]^{n-r}$$
$$= \frac{n!}{(r-1)!(n-r)!}\frac{f_1(x_r)[2p_1 - p_2F_2(x_r) - 2p_1F_1(x_r)] + f_2(x_r)[2p_2 - p_2F_1(x_r)]}{(2 - F_1(x_n))^2}$$
$$[\frac{p_1F_1(x_r) + p_2F_2(x_r)}{2 - F_1(x_r)}]^{r-1} *[1 - \frac{p_1F_1(x_r) + p_2F_2(x_r)}{2 - F_1(x_r)}]^{n-r}$$

Le moment ordinaire

$$\mu\prime = \int_0^{\infty} x^r g(x)dx$$

$$\mu\prime = \int_0^{\infty} x^r \frac{2(\frac{k}{\lambda}(\frac{x}{\lambda})^{k-1}\exp(-\frac{x}{\lambda})^k)}{(1 + \exp(-\frac{x}{\lambda})^k)^2}dx$$

$$\mu\prime = 2\frac{k}{\lambda^k}\int_0^{\infty} \frac{x^{k+r-1}\exp(-\frac{x}{\lambda})^k}{(1 + \exp(-\frac{x}{\lambda})^k)^2}dx \tag{2.38}$$

Exemple

on prend deux loi de Weibull de paramètres différents (le même exemple avec le modèle de l'alternative de Lehman généralisé)

on remplace dans G(x) on trouve :

$$G(x)=\frac{p_1(1-\exp(-\dfrac{x}{\lambda_1})^{k_1})+p_2(1-\exp(-\dfrac{x}{\lambda_2})^{k_2})}{2-(1-\exp(-\dfrac{x}{\lambda_1})^{k_1})}$$

$$G(x)=\frac{p_1(1-\exp(-\dfrac{x}{\lambda_1})^{k_1})+p_2(1-\exp(-\dfrac{x}{\lambda_2})^{k_2})}{1+\exp(-\dfrac{x}{\lambda 1})^{k1}} \tag{2.39}$$

et la fonction de densité correspondante $\grave{a}$ la formule précédante est :

$$g(x)=\frac{\dfrac{k_1}{\lambda_1}(\dfrac{x}{\lambda_1})^{k_1-1}\exp(-\dfrac{x}{\lambda_1})^{k_1}[2p_1-p_2(1-\exp(-\dfrac{x}{\lambda_2})^{k_2})-2p_1(1-\exp(-\dfrac{x}{\lambda_1})^{k_1})]}{(1+\exp(-\dfrac{x}{\lambda_1})^{k_1})^2}$$

$$+\frac{\dfrac{k_2}{\lambda_2}(\dfrac{x}{\lambda_2})^{k_2-1}\exp(-\dfrac{x}{\lambda_2})^{k_2}[2p_2-p_2(1-\exp(-\dfrac{x}{\lambda_1})^{k_1})]}{(1+\exp(-\dfrac{x}{\lambda_1})^{k_1})^2}$$

La fonction de hasard :

$$h(x)=\frac{g(x)}{1-G(x)}$$

on a :

$$1-G(x)=\frac{1+\exp(-\dfrac{x}{\lambda_1})^{k_1}-p_1(1-\exp(-\dfrac{x}{\lambda_1})^{k_1})-p_2(1-\exp(-\dfrac{x}{\lambda_2})^{k_2})}{1+\exp(-\dfrac{x}{\lambda_1})^{k_1}}$$

alors :

$$h(x)=\frac{\dfrac{k_1}{\lambda_1}(\dfrac{x}{\lambda_1})^{k_1-1}\exp(-\dfrac{x}{\lambda_1})^{k_1}[2p_1-p_2(1-\exp(-\dfrac{x}{\lambda_2})^{k_2})-2p_1(1-\exp(-\dfrac{x}{\lambda_1})^{k_1})]}{[1+\exp(-\dfrac{x}{\lambda_1})^{k_1}-p_1(1-\exp(-\dfrac{x}{\lambda_1})^{k_1})-p_2(1-\exp(-\dfrac{x}{\lambda_2})^{k_2})](1+\exp(-\dfrac{x}{\lambda_1})^{k_1})}$$

$$+\frac{\dfrac{k_2}{\lambda_2}(\dfrac{x}{\lambda_2})^{k_2-1}\exp(-\dfrac{x}{\lambda_2})^{k_2}[2p_2-p_2(1-\exp(-\dfrac{x}{\lambda_1})^{k_1})]}{[1+\exp(-\dfrac{x}{\lambda_1})^{k_1}-p_1(1-\exp(-\dfrac{x}{\lambda_1})^{k_1})-p_2(1-\exp(-\dfrac{x}{\lambda_2})^{k_2})](1+\exp(-\dfrac{x}{\lambda_1})^{k_1})}$$

Le moment ordinaire

$\mu\prime = \int_0^\infty x^r g(x)dx$

$$\mu\prime = \int_0^\infty x^r[\frac{\frac{k_1}{\lambda_1}(\frac{x}{\lambda_1})^{k_1-1}\exp(-\frac{x}{\lambda_1})^{k_1}[2p_1 - p_2(1-\exp(-\frac{x}{\lambda_2})^{k_2}) - 2p_1(1-\exp(-\frac{x}{\lambda_1})^{k_1})]}{(1+\exp(-\frac{x}{\lambda_1})^{k_1})^2}$$

$$+\frac{\frac{k_2}{\lambda_2}(\frac{x}{\lambda_2})^{k_2-1}\exp(-\frac{x}{\lambda_2})^{k_2}[2p_2 - p_2(1-\exp(-\frac{x}{\lambda_1})^{k_1})]}{(1+\exp(-\frac{x}{\lambda_1})^{k_1})^2}]dx$$

L'estimation des paramètres par la méthode du Maximum de vraisemblance

$L(x) = \prod_{i=1}^n g(x_i)$

$$L(x) = \prod_{i=1}^n \frac{f_1(x_i)[2p_1 - p_2F_2(x_i) - 2p_1F_1(x_i)] + f_2(x_i)[2p_2 - p_2F_1(x_i)]}{(2-F_1(x_i))^2}$$

$$\ln L(x) = \ln \prod_{i=1}^n \frac{f_1(x_i)[2p_1 - p_2F_2(x_i) - 2p_1F_1(x_i)] + f_2(x_i)[2p_2 - p_2F_1(x_i)]}{(2-F_1(x_i))^2}$$

$$\ln L(x) = \sum_{i=1}^n \ln[\frac{f_1(x_i)[2p_1 - p_2F_2(x_i) - 2p_1F_1(x_i)] + f_2(x_i)[2p_1 - p_2F_1(x_i)]}{(2-F_1(x_i))^2}]$$

$$= \sum_{i=1}^n \ln[(f_1(x_i)[2p_1 - p_2F_2(x_i) - 2p_1F_1(x_i)] + f_2(x_i)[2p_2 - p_2F_1(x_i)]$$
$$- 2\ln(2-F_1(x_i))]$$

les équations de la vraisemblance sont résumé dans le système d'équation suivant :

$$(p)\begin{cases}\dfrac{\partial \ln L(x,\lambda_1,\lambda_2,k_1,k_2)}{\partial\lambda_1} = \sum_{i=1}^{n}[\dfrac{\dfrac{\partial f_1(x_i)}{\partial\lambda_1}(2p_1-p_2F_2(x_i)-2p_1F_1(x_i))}{B} \\ -\dfrac{\dfrac{\partial F_1(x_i)}{\partial\lambda_1}2p_1f_1(x_i)}{B} - \dfrac{\dfrac{\partial F_1(x_i)}{\partial\lambda_1}p_2f_2(x_i)}{B} - \dfrac{2\dfrac{\partial F_1(x_i)}{\partial\lambda_1}}{2-F_1(x_i)}] \\ \dfrac{\partial \ln L(x,\lambda_1,\lambda_2,k_1,k_2)}{\partial k_1} = \sum_{i=1}^{n}[\dfrac{\dfrac{\partial f_1(x_i)}{\partial k_1}(2p_1-p_2F_2(x_i)-2p_1F_1(x_i))}{B} \\ +\dfrac{\dfrac{\partial F_1(x_i)}{\partial k_1}2p_1f_1(x_i)}{B} - \dfrac{\dfrac{\partial F_1(x_i)}{\partial k_1}p_2f_2(x_i)}{B} - \dfrac{2\dfrac{\partial F_1(x_i)}{\partial k_1}}{2-F_1(x_i)}] \\ \dfrac{\partial \log L(x,\lambda_1,\lambda_2,k_1,k_2)}{\partial\lambda_2} = \sum_{i=1}^{n}[\dfrac{-f_1(x_i)p_2\dfrac{\partial F_2(x_i)}{\partial\lambda_2} - \dfrac{\partial f_2(x_i)}{\partial\lambda_2}[2p_2-p_2F_1(x_i)]}{B}] \\ \dfrac{\partial \log L(x,\lambda_1,\lambda_2,k_1,k_2)}{\partial k_2} = \sum_{i=1}^{n}[\dfrac{-f_1(x_i)p_2\dfrac{\partial F_2(x_i)}{\partial k_2} + \dfrac{\partial f_2(x_i)}{\partial k_2}[2p_2-p_2F_1(x_i)]}{B}] \end{cases}$$

$$\bullet\frac{\partial \log L(x,\lambda_1,\lambda_2,k_1,k_2)}{\partial\lambda_1} = \sum_{i=1}^{n}[\frac{-k_1^2x_i^{k_1-1}\lambda_1^{k_1}+k_1^3x_i^{k_1}}{\lambda_1^{k_1+1}}\exp(-\frac{x_i^{k_1}}{\lambda_1^{k_1}})$$

$$\frac{(2p_1-p_2(1-\exp(-\frac{x_i}{\lambda_2})^{k_2})-2p_1(1-\exp(-\frac{x_i}{\lambda_1})^{k_1})}{A}$$

$$-\frac{x_i^{k_1}k_1\lambda_1^{3k_1-1}\exp-\frac{x_i^{k_1}}{\lambda_1^{k_1}}p_2(\frac{k_2}{\lambda_2}(\frac{x_i}{\lambda_2})^{k_2-1}\exp(-\frac{x}{\lambda_2})^{k_2})}{A}$$

$$+\frac{x_i^{k_1}k_1\lambda_1^{3k_1-1}\exp-\frac{x_i^{k_1}}{\lambda_1^{k_1}}2p_1((\frac{k_1}{\lambda_1})(\frac{x}{\lambda_1})^{k_1-1}\exp-(\frac{x}{\lambda_1})^{k_1})}{A}$$

$$-2\frac{x_i^{k_1}k_1\lambda_1^{3k_1-1}\exp(-\frac{x_i^{k_1}}{\lambda_1^{k_1}})}{1+\exp(-\frac{x_i}{\lambda_1})^{k_1}}]=0$$

$$\bullet\frac{\partial \log L(x,\lambda_1,\lambda_2,k_1,k_2)}{\partial k_1} = \sum_{i=1}^{n}[\frac{x_i^{k_1-1}}{{\lambda_1}^{k_1}}+\ln(\frac{x_i}{\lambda_1})\frac{k_1x_i^{k_1-1}}{{\lambda_1}^{k_1}}-\ln(\frac{x_i}{\lambda_1})k_1x_i^{-1}(\frac{x_i}{\lambda_1})^{2k_1}\exp-(\frac{x_i}{\lambda_1})^{k_1}$$

$$\frac{(2p_1 - p_2(1-\exp(-\frac{x_i}{\lambda_2})^{k_2}) - 2p_1(1-\exp(-\frac{x}{\lambda_1})^{k_1})}{A}$$
$$+\frac{p_2(\frac{k_2}{\lambda_2}(\frac{x_i}{\lambda_2})^{k_2-1}\exp(-\frac{x_i}{\lambda_2})^{k_2})}{A} * (-\ln(\frac{x_i}{\lambda_1})\exp(k_1\ln(-\frac{x_i}{\lambda_1})))$$
$$-\frac{2p_1(\frac{k_1}{\lambda_1}(\frac{x_i}{\lambda_1})^{k_1-1}\exp(-\frac{x_i}{\lambda_1})^{k_1})}{A}(-\ln(\frac{x_i}{\lambda_1})\exp(k_1\ln(-\frac{x_i}{\lambda_1})))$$
$$+2\frac{(-\ln(\frac{x_i}{\lambda_1})\exp(k_1\ln(-\frac{x_i}{\lambda_1})))}{1+\exp(-\frac{x_i}{\lambda_1})^{k_1}}] = 0$$

$$\bullet\frac{\partial \log L(x,\lambda_1,\lambda_2,k_1,k_2)}{\partial\lambda_2} = \sum_{i=1}^{n}[\frac{-p_2(\frac{k_1}{\lambda_1}(\frac{x_i}{\lambda_1})^{k_1-1}\exp(-\frac{x_i}{\lambda_1})^{k_1})}{A}$$
$$x_i^{k_2}k_2\lambda_2^{3k_2-1}\exp-\frac{x_i^{k_2}}{\lambda_2^{k_2}} + \frac{-k_2^2x_i^{k_2-1}\lambda_2^{k_2}+k_2^3x_i^{k_2}}{\lambda_2^{k_2+1}}\exp-\frac{x_i^{k_2}}{\lambda_2^{k_2}}$$
$$\frac{[2p2 - p_2(1-\exp(-\frac{x_i}{\lambda_1})^{k_1})]}{A}] = 0$$

$$\bullet\frac{\partial \log L(x,\lambda_1,\lambda_2,k_1,k_2)}{\partial k_2} = \sum_{i=1}^{n}[\frac{-p_2(\frac{k_1}{\lambda_1}(\frac{x_i}{\lambda_1})^{k_1-1}\exp(-\frac{x_i}{\lambda_1})^{k_1})}{A}$$
$$(\frac{x_i}{\lambda_1})^{k_1-1}\exp(-\frac{x_i}{\lambda_1})^{k_1}(-\ln(\frac{x_i}{\lambda_2})\exp(k_2\ln(-\frac{x_i}{\lambda_2})))$$
$$+\frac{x_i^{k_2-1}}{{\lambda_2}^{k_2}} + \ln(\frac{x_i}{\lambda_2})\frac{k_2x_i^{k_2-1}}{{\lambda_2}^{k_2}} - \ln(\frac{x_i}{\lambda_2})k_2x_i^{-1}(\frac{x_i}{\lambda_2})^{2k_2}\exp-(\frac{x_i}{\lambda_2})^{k_2}$$
$$\frac{[2p2 - p_2(1-\exp(-\frac{x_i}{\lambda_1})^{k_1})]}{A}] = 0$$

avec :

$$A = \frac{k_1}{\lambda_1}(\frac{x}{\lambda_1})^{k_1-1}\exp(\frac{x}{-\lambda_1})^{k_1}(2p_1 - p_2(1-\exp(-\frac{x}{\lambda_1})^{k_1} - 2p_1(1-\exp(-\frac{x}{\lambda_1})^{k_1}))$$
$$+\frac{k_2}{\lambda_2}(\frac{x}{\lambda_2})^{k_2-1}\exp(\frac{x}{-\lambda_2})^{k_2}(2p_2 - p_2(1-\exp(\frac{x}{{}_\lambda 1})^{k1})$$

et

$$B = f_1(x_i)[2p_1 - p_2F_2(x_i) - 2p_1F_1(x_i)] + f_2(x_i)[2p_2 - p_2F_1(x_i)]$$

Chapitre 3

Simulation et résultats numériques

3.1 Introduction

Dans ce chapitre, nous allons présenter le travail de simulation sur des données réelles de survie effectué pour illustrer et étayer les différents aspects théoriques abordés dans les chapitres précédents. Cette étude portera essentiellement sur :
• La comparaison des différents modèles paramétriques des données de survie proposés dans ce mémoire :
1. Alternative de LEHMAN ;
2. Alternative de LEHMAN généralisée ;
3. Modèle de mélange.
• L'étude de la performance de ces modèles ;
• L'étude de l'influence de la taille de l'échantillon sur ces différents modèles.
Pour ceci nous avons considérer deux ensembles de données réelles :

I) Le premier cas de données est issu d'une étude du temps de défaillance du système de climatisation d'un avion présentée dans Linhart and Zucchini [3, 11] :

23, 261, 87, 7, 120, 14, 62, 47, 225, 71, 246, 21, 42, 20, 5, 12, 120, 11, 3, 14, 71, 11, 14, 11, 16, 90, 1, 16, 52, 95.

Ces données ont également utilisées dans Gupta and Kundu et dans Singh et al.
II) Le deuxième cas de données décrivent une étude de rémission (en mois) d'un échantillon aléatoire de 128 patients atteints de cancer de la vessie rapportée dans Lee and Wang [14] :

0.08,2.09, 3.48, 4.87, 6.94, 8.66, 13.11, 23.63, 0.20, 2.23, 3.52, 4.98, 6.97, 9.02, 13.29, 0.40, 2.26,3.57, 5.06, 7.09, 9.22, 13.80, 25.74, 0.50, 2.46, 3.64, 5.09, 7.26, 9.47, 14.24, 25.82, 0.51, 2.54,3.70, 5.17, 7.28, 9.74, 14.76, 26.31, 0.81, 2.62, 3.82, 5.32, 7.32, 10.06, 14.77,32.15, 2.64, 3.88,5.32, 7.39, 10.34, 14.83, 34.26, 0.90, 2.69, 4.18, 5.34,7.59, 10.66, 15.96, 36.66, 1.05, 2.69, 4.23,5.41, 7.62, 10.75, 16.62, 43.01, 1.19, 2.75, 4.26, 5.41, 7.63, 17.12, 46.12, 1.26,2.83, 4.33, 5.49,7.66,11.25, 17.14, 79.05, 1.35, 2.87, 5.62, 7.87, 11.64, 17.36, 1.40, 3.02, 4.34, 5.71, 7.93, 11.79,18.10, 1.46, 4.40, 5.85, 8.26, 11.98, 19.13, 1.76, 3.25, 4.50, 6.25, 8.37, 12.02, 2.02, 3.31, 4.51, 6.54, 8.53, 12.03, 20.28, 2.02, 3.36, 6.76, 12.07, 21.73, 2.07, 3.36, 6.93, 8.65, 12.63, 22.69.

Ces données ont été déja utilisées dans Singh, S. K., Singh, U. and Kummar, M. [, 2014].

3.2 Plan de simulation

Pour notre plan de simulation, nous allons extraire aléatoirement $\grave{a}$ partir de l'ensemble des données réelles un échantillon de taille $p \in \{3, 5, 10, 15, 20, 25\}$ pour le premier cas de données et de taille $p \in \{5, 10, 15, 20, 25, 50\}$ pour le deuxiéme cas. Ces échantillons seront utilisés pour calculer les paramètres du modèle considéré et qui seront $\grave{a}$ leurs tours réinjectés afin de calculer le logarithme de la vraisemblance (log-likelihood creterion) qui nous sert de critère de comparaison pour les différents modèles. Nous réaliserons 40 répétitions d'expérience pour chaque taille d'échantillon et nous considérons la moyenne des résultats obtenus.

Nous allons suivre les étapes suivantes pour la simulation :
- Extraire un échantillon de taille n pour les données considérées.
- Calculer les paramètres modèle :
  — Alternative de LEHMAN (AL).
  — Alternative de LEHMAN généralisée (ALG).
  — Modèle de mélange (MM).
- Calculer le critère de comparaison (LL) pour ces différents modèles.

Notons que pour le premier modèle (Alternative de Lehman), nous avons considéré les deux cas particulier suivant :
— $\alpha = 1$ et $F(x) \sim Weibul$,
— $\alpha = 1$ et $F(x) \sim Gamma$.

Les simulations et les graphiques ont été réalisés $\grave{a}$ l'aide du logiciel R (package ks). Nous avons utilisé la version 2.8.0 pour la programmation. R est un système d'analyse statistique et graphique crée par **Ross Ihaka** et **Robert Gentlelman**. Il est $\grave{a}$ la fois un langage et un logiciel qui comporte de nombreuses fonctions pour les analyses statistiques et les graphiques.

3.3 Résultats de la simulation

Les résultats de la simulation sont données sous forme de tableaux. Les tableaux (1.1), (1.2), (1.3), (1.4), (1.5), (1.6), (1.7), (1.8), (1.9), (1.10), (1.11), (1.12), donnent les valeurs moyennes de LL des 40 expériences réalisées pour les deux ensembles de données réelles considérées.

n	Weibul	Gamma	ALG	MM
3	-148.336	-153.721	-148.219	-148.301
5	-149.258	-153.440	-148.308	-148.216
10	-149.813	-155.127	-148.301	-149.229
15	-150.724	-155.875	-149.000	-149.001
25	-151.029	-156.166	-149.825	-149.655

TABLE 3.1 – La moyenne de LL pour le premier cas de données avec $p_1 = 1/2$, $p_2 = 1/2$, $F_1 \sim$ Weibul et $F_2 \sim$ Weibul

n	Weibul	Gamma	ALG	MM
3	-147.190	-153.338	-144.771	-144.183
5	-147.861	-152.019	-145.468	-146.225
10	-149.007	-152.971	-145.333	-146.796
15	-150.116	-153.311	-146.267	-148.155
25	-150.746	-154.000	-147.952	-149.010

TABLE 3.2 – La moyenne de LL pour le premier cas de données avec $p_1 = 1/3$, $p_2 = 2/3$, $F_1 \sim$ Weibul et $F_2 \sim$ Weibul

n	Weibul	Gamma	ALG	MM
3	-148.336	-153.721	-147.012	-148.000
5	-149.258	-153.440	-148.671	-148.241
10	-149.813	-155.127	-148.810	-149.334
15	-150.721	-155.875	-148.997	-149.526
25	-151.029	-156.166	-149.023	-150.701

TABLE 3.3 – La moyenne de LL pour le premier cas de données avec $p_1 = 1/2$, $p_2 = 1/2$, $F_1 \sim$ Gamma et $F_2 \sim$ Gamma

n	Weibul	Gamma	ALG	MM
3	-147.190	-153.338	-146.333	-147.200
5	-147.861	-152.019	-146.502	-147.118
10	-149.007	-152.971	-146.991	-148.151
15	-150.116	-153.311	-149.997	-150.000
25	-150.746	-154.000	-150.002	-150.407

TABLE 3.4 – La moyenne de LL pour le premier cas de données avec $p_1 = 1/3$, $p_2 = 2/3$, $F_1 \sim$ Gamma et $F_2 \sim$ Gamma

n	Weibul	Gamma	ALG	MM
3	-148.336	-153.721	-142.076	-143.971
5	-149.258	-153.440	-142.611	-144.433
10	-149.813	-155.127	-143.401	-144.892
15	-150.721	-155.875	-144.127	-145.017
25	-151.029	-156.166	-145.581	-146.872

TABLE 3.5 – La moyenne de LL pour le premier cas de données avec $p_1 = 1/2$, $p_2 = 1/2$, $F_1 \sim$ Weibul et $F_2 \sim$ Gamma

n	Weibul	Gamma	ALG	MM
3	-147.190	-153.338	-140.129	-143.541
5	-147.861	-152.019	-140.551	-143.819
10	-149.007	-152.971	-140.905	-144.174
15	-150.116	-153.311	-141.630	-144.667
25	-150.746	-154.000	-142.013	-145.386

TABLE 3.6 – La moyenne de LL pour le premier cas de données avec $p_1 = 1/3$, $p_2 = 2/3$, $F_1 \sim$ Weibul et $F_2 \sim$ Gamma

n	Weibul	Gamma	ALG	MM
3	-526.241	-527.013	-501.388	-501.957
5	-529.102	-531.861	-504.700	-504.402
10	-538.549	-542.118	-506.899	-508.139
15	-541.134	-545.233	-507.126	-508.200
25	-550.020	-555.991	-511.615	-513.698

TABLE 3.7 – La moyenne de LL pour le deuxième cas de données avec $p_1 = 1/2$, $p_2 = 1/2$, $F_1 \sim$ Weibul et $F_2 \sim$ Weibul

n	Weibul	Gamma	ALG	MM
3	-526.241	-527.013	-446.178	-448.701
5	-529.102	-531.861	-446.801	-449.552
10	-538.549	-542.118	-448.691	-452.837
15	-541.134	-545.233	-453.007	-457.294
25	-550.020	-555.991	-458.521	-463.096

TABLE 3.8 – La moyenne de LL pour le deuxième cas de données avec $p_1 = 1/3$, $p_2 = 2/3$, $F_1 \sim$ Weibul et $F_2 \sim$ Weibul

n	Weibul	Gamma	ALG	MM
3	-526.241	-527.013	-439.810	-439.901
5	-529.102	-531.861	-440.057	-439.873
10	-538.549	-542.118	-449.151	-451.227
15	-541.134	-545.233	-453.762	-456.382
25	-550.020	-555.991	-460.643	-461.109

TABLE 3.9 – La moyenne de LL pour le deuxième cas de données avec $p_1 = 1/2$, $p_2 = 1/2$, $F_1 \sim$ Gamma et $F_2 \sim$ Gamma

n	Weibul	Gamma	ALG	MM
3	-526.241	-527.013	-429.713	-432.109
5	-529.102	-531.861	-430.926	-432.455
10	-538.549	-542.118	-435.528	-436.365
15	-541.134	-545.233	-438.069	-439.691
25	-550.020	-555.991	-452.576	-456.048

TABLE 3.10 – La moyenne de LL pour le deuxième cas de données avec $p_1 = 1/3$, $p_2 = 2/3$, $F_1 \sim$ Gamma et $F_2 \sim$ Gamma

n	Weibul	Gamma	ALG	MM
3	-526.241	-527.013	-422.739	-451.037
5	-529.102	-531.861	-424.336	-451.660
10	-538.549	-542.118	-430.892	-458.407
15	-541.134	-545.233	-433.195	-461.953
25	-550.020	-555.991	-439.601	-469.377

TABLE 3.11 – La moyenne de LL pour le deuxième cas de données avec $p_1 = 1/2$, $p_2 = 1/2$, $F_1 \sim$ Weibul et $F_2 \sim$ Gamma

n	Weibul	Gamma	ALG	MM
3	-526.241	-527.013	-407.541	-416.831
5	-529.102	-531.861	-409.113	-421.659
10	-538.549	-542.118	-414.839	-428.217
15	-541.134	-545.233	-417.591	-429.503
25	-550.020	-555.991	-428.380	-452.175

TABLE 3.12 – La moyenne de LL pour le deuxième cas de données avec $p_1 = 1/3$, $p_2 = 2/3$, $F_1 \sim$ Weibul et $F_2 \sim$ Gamma

3.4 Performance des estimateurs

La lecture des résultats obtenus dans les tableaux (1.1),...,(1.8), (1.9), (1.10), (1.11), (1.12), nous permet d'affirmer que :

Pour le premier ensemble de données :

Les résultats obtenus dans les tableaux (1.1), (1.2), (1.3), (1.4), montrent clairement que la loi de Weibul est nettement meilleures que celle de Gamma pour le premier modèle (Alternative de Lehman). Par contre, les résultats sont presque identiques entre les deux modèles (ALG) et (MM) avec un léger avantage pour la méthode (ALG) en particulier pour la taille de l'échantillon n = 25. Ces résultats montrent aussi que les deux modèles proposés dans ce mémoire sont plus performants que les modèles classiques, Weibul et Gamma. On remarque par exemple dans le tableau (1.2) : pour la taille d'échantillon n=25, nous avons LL = - 147.952 pour le modèle (ALG) alors que LL= -150.746 pour Weibull et LL = -154.000 pour Gamma.

Les tableaux (1.5), (1.6) confirment les résultas déjà obtenus dans les tableaux précédents (1.1), (1.2), (1.3), (1.4), pour le premier modèle (Alternative de Lehman) *à* savoir que la loi de Weibul est nettement meilleure que la loi de Gamma. Mais pour les deux modèles restants (ALG et MM), le mélange de lois ($F_1 \sim Weibul$ et $F_2 \sim Gamma$) améliorent nettement leurs performances avec un avantage claire pour le premier modèle (ALG). Par exemple, dans le tableau (1.6), nous avons pour la taille d'échantillon n=25, LL = - 150.746 (Weibull), LL = -154.000 (Gamma), LL = -142.013 (ALG), LL = -145.386 (MM).

Les résultats obtenus montrent aussi que la taille de l'échantillon n'influe pas sur la qualité de l'estimation. On remarque pour les résultats sont stable en général pour les différentes tailles de l'échantillon.

Pour le deuxième ensemble de données :

Le même constat est valable pour les résultats obtenus dans les tableaux (1.7), (1.8), (1.9), (1.10), (1.11), (1.12) avec un avantage significatif pour le modéle (ALG) exception faite pour les résultats du tableau (1.9) avec la taille d'échantillon n=5 et n=3, les résultats de LL sont presque identique pour les deux modèles proposés *à* savoir (ALG) et (MM). Ici également, nous remarquons que la taille de l'échantillon n'influe pas significativement sur les résultats obtenus. Ainsi les résultats restent en général stables pour toutes les tailles.

3.5 Conclusion

Notre travail de simulation aborde un problème crucial, bien connu dans l'analyse de la survie en particulier et en statistiques en général. Il s'agit du problème de choix d'un modèle approprié dans le cadre paramétrique pour l'analyse des données de survie. Ainsi, Les résultats obtenus pour les deux ensembles de données réelles (étude du temps de défaillance du système de climatisation d'un avion et étude de rémission, en mois, d'un échantillon aléatoire de 128 patients atteints de cancer de la vessie) confirment que les deux alternatives que nous avons proposés dans ce travail sont plus éfficace que les modèle paramétriques classiques dans la modélisation et l'analyse des données de survie avec un avantage significatif pour le premier modèle, alternative de Lehman généralisée. Notons, en outre, que les deux modèles sont stable et ne sont pas influencé significativement par la taille de l'échantillon.

Conclusion générale

Ce mémoire est une contribution à la modélisation et à l'analyse des données de survie dans un cadre paramétrique. Dans une première étape, nous avons fixé le cadre générale de ce domaine de recherche à travers la présentation d'un certain nombre de notions préliminaires bien connues dans la littérature. Dans la deuxième partie, nous avons introduit trois modèles paramétriques pour les données de survie. Il s'agit en premier lieu du modèle de LEHMAN qui généralise les modèles paramétriques classiques et proposé dans [12]. En deuxième lieu, nous avons construit deux autres modèles alternatifs, l'alternative de LEHMAN généralisé et le modèle du mélange. Dans ce chapitre, les différentes propriétés statistiques et asymptotiques des deux modèles proposés ont été présentées. Enfin et afin de tester l'efficacité des modéles introduits, une étude de simulation sur des données réelles a été réalisée. Les résultats numériques obtenus indiquent que nos deux modèles sont plus performant que les modèles classiques.

Bibliographie

[1] C.Chesneau
Sur l'estimateur de maximum de vraisemblance(env) (cour), Licence. France. 2017.

[2] C. Vassoilles
Propostion d'une nouvelle méthode de classification à base de copules , Mémoire l'université du Québec 2014

[3] D.Kumar,U.Singh,S.Kumar et S.Mukherjee
The new probability distribution :an aspect to a life time distribution , journal of mathematical science letters vol 6 $N^0$1,p 35-42 (2017)

[4] E.Doucet
Estimateurs de noyau et théorie des valeurs extrêmes , mémoire Université de Québec à Montréal 2014

[5] F. Planchet
Modèles de durées (cour), 2019-2020

[6] G.Colletaz
Économétrie des durées de survies(cour) , 9 septembre 2019

[7] J.Cristophe Breton
Fondement de probabilités de (Ω, F, P) *au conséquences de la LG et et de TCL (cour)* , 2014

[8] L.Tristan
modéles statistiques pour des données de survie corrélées, thèse de doctorat l'institut national agronomique Pari-Grignon.

[9] M.Genin
Variables aléatoires (cour), Épidémiologie et Qualité des soins octobre 2015

[10] P.Saint Pierre
Introduction à l'analyse des durées de survie (cour), Février 2015

[11] R.D.Gupta et D.Kundu
Exponentiated Exponential Family : An Alternative to Gamma and Weibull Distributions , Biometrical Journal 43 (2001) 1, 117–130

[12] R.C.Gupta ,P.L.Gupta et R.D.Gupta
Modeling failure time data by lehman alternatives , Communications in Statistics - Theory and Methods (1998)

[13] S.Nedjar
Poisson pseudo lindly distribution et leurs applications en assurances vie , thèse de doctorat université Annaba 2017

[14] Singh, S. K., Singh, U. and Kumar
Bayesian inference for exponentiated Pareto model with application to bladder cancer remission time, Statistics in Transition (2014)