

RÉPUBLIQUE ALGÉRIENNE DÉMOCRATIQUE ET POPULAIRE

Ministère de l'enseignement supérieur et de la recherche scientifique

Université A.MIRA - BÉJAIA

FACULTÉ DES SCIENCES EXACTES

DÉPARTEMENT DE MATHÉMATIQUES



MÉMOIRE DE FIN D'ÉTUDE

En vue de l'obtention du Diplôme de MASTER en Mathématiques

Option : Probabilités, Statistique et Applications

Thème

Étude de la fonction taux de hasard inversé

Réalisé par :

CHADLI Maïssa

Soutenu le 24/06/2023

devant le jury composé de :

Encadreur	Mme. Karima LAGHA	Professeur	Université de Béjaïa
Président	M. Yacine BOUMZAID	M.C.B	Université de Béjaïa
Examineur	M. Mohand BOURAINE	M.A.A	Université de Béjaïa

Année universitaire : 2022/2023

Dédicace

Je dédié ce modeste mémoire :

A mes chers parents, pour tous leurs sacrifices, leur amour, leur tendresse, leur soutien et leurs prières tout au long de mes études.

A mon fiancé, pour ses encouragements permanents, et son soutien moral.

A mes sœurs et leurs maris, a mes frères, pour leur appui et leur encouragement.

A toute ma famille pour leur soutien tout au long de mon parcours universitaire,

A tout ce qui m'ont encouragé durant la réalisation de mon mémoire

A mes très chères amies

Finalement à tous ceux qui m'ont aidé de près ou de loin.

Remerciement

Tout d'abord, je tiens à remercier le "BON DIEU" le tout puissant de m'avoir accordé patience, courage et volonté afin de réaliser et de mener à terme ce modeste travail.

Je remercie tout particulièrement mon encadreur «Mme LAGHA Karima» pour ses conseils, sa grande disponibilité et sa générosité.

Je voudrais également remercier les membres de jury, Mr.Boumzaid et Mr.Bouraine, pour avoir accepté d'examiner ce travail.

Je tiens aussi à exprimer ma reconnaissance à tous les enseignants qui ont contribué à ma formation tout au long de ma carrière universitaire.

Je remercie particulièrement Mme MEDJIDER Meriem.

Je remercie chaleureusement toute ma famille, qui m'a soutenu, encouragé et poussé durant toutes mes années d'étude.

Merci à tous

Notation et abréviation

$K(u)$	Noyau.
h_{opt}	La fenêtre optimale.
$\pi(h), \pi(h/t)$	loi a priori, loi a posteriori.
$S(t)$	Fonction de survie.
$\lambda(t)$	Taux de hasard.
$r(t)$	Taux de hasard inversé.
MLE	Maximum Likelihood Estimation (estimation du maximum de vraisemblance).
HR	Hasard Rate (Taux de hasard).
RHR	Reversed Hasard Rate (Taux de hasard inversé).
i.i.d	indépendant identiquement distribué.
p.s	presque sûre.
G	Gamma.
IG	Inverse Gaussien.
RIG	Inverse Gaussien Réciproque.
UCV	Unbiased Cross Validation (Validation croisée non-biaisé).
MISE	L'erreur quadratique moyenne intégrée.
ISE	L'erreur quadratique intégrée.
IC	Intervalle de Confiance.
KM	Kaplan-Meier.
IFR	Increasing Failure Rate.
DFR	Decreasing Failure Rate.

Table des matières

Table des figures	6
Liste des Tableaux	7
Introduction	9
1 Éléments fondamentaux sur l'analyse de survie et estimation	10
1.1 L'analyse de survie	10
1.2 Données de survie	11
1.2.1 Données complètes	12
1.2.2 Données censurées	12
1.2.3 Données tronquées	14
1.3 Distributions de la durée de survie	15
1.3.1 Fonction de survie	15
1.3.2 Fonction de répartition	16
1.3.3 Fonction de densité	16
1.3.4 Fonctions du risque et du risque cumulé	17
1.3.5 Comparaison stochastique	18
1.4 Quantités associées à la distribution de survie	19
1.4.1 Moyenne de la durée de survie	19
1.4.2 Moyenne résiduelle de la durée de survie	19
1.4.3 Variance de la durée de survie	19
1.5 Estimation paramétrique	19
1.5.1 La méthode du maximum de vraisemblance (MLE)	20
1.5.2 Estimation par intervalle de confiance	20
1.6 Estimation non paramétrique	21
1.6.1 Estimateur de Kaplan-Meier de la survie (KM)	21

1.6.2	Estimation à noyau asymétrique de la densité	23
1.6.3	Choix du paramètre de lissage	25
1.7	Distributions de la durée de vie	28
1.8	Conclusion	30
2	Taux de hasard et taux de hasard inversé	31
2.1	Le taux de hasard (HR)	31
2.1.1	Propriétés	32
2.1.2	Fonction exponentiée F^{**}	33
2.1.3	Estimation paramétrique du taux de hasard (HR)	34
2.1.4	Estimation non paramétrique du HR par la méthode de Kaplan-Meier . .	41
2.2	Taux de hasard inversé (RHR)	41
2.2.1	Propriétés du RHR	43
2.2.2	Estimation paramétrique du RHR	48
2.2.3	Comparaison entre F et F^{**}	52
2.2.4	Estimation non paramétrique du RHR	55
2.3	Conclusion	60
3	Application	61
3.1	Estimation paramétrique	61
3.1.1	Analyse d'un ensemble de données non censurées	61
3.1.2	Analyse d'un ensemble de données censurées	67
3.2	Estimation non paramétrique	74
3.2.1	Estimation par la méthode de Kaplan Meier	74
3.2.2	Estimation à noyau de taux de hasard inversé	78
3.3	conclusion	80
	Conclusion	81
	Bibliographie	84
	Annexe	85

Table des figures

1.1	Représentation de différentes dates et temps écoulé.	11
1.2	Types de censures	14
1.3	Les estimateurs de K.M de la fonction de survie des 21 malades traités par le traitement 6MP	23
2.1	Courbe en baignoire de HR.	32
2.2	Fonction d'élasticité (axe de gauche) et fonction de densité (axe de droite) pour $\alpha = 10$	47
3.1	L'intervalle de confiance de la fonction de survie (données non censurées).	64
3.2	L'intervalle de confiance de la fonction de HR (données non censurées)	65
3.3	L'intervalle de confiance de la fonction RHR (données non censurées)	67
3.4	L'intervalle de confiance de la fonction de survie (données censurées).	70
3.5	L'intervalle de confiance de RH (données censurées).	71
3.6	L'intervalle de confiance de la fonction RHR (données censurées).	73
3.7	l'estimation de la fonction de survie par KM et les intervalles de confiance à un niveau de 95%	76
3.8	Estimateurs de KM de taux de hasard et taux de hasard inversé	77
3.9	Estimateur des noyaux gamma, IG et RIG de la distribution gamma pour n=500	79
3.10	Estimateurs du RHR de la distribution Gamma(3, 1) pour n=500 et $h[UCV]$. .	79
3.11	Estimateurs du RHR de la distribution $LN(1, 1)$ pour n=500 et $h[UCV]$	79

Liste des tableaux

1.1	L'estimateur de K.M de la fonction de survie du groupe de 21 malades traités par le traitement 6MP	22
1.2	Quelques noyaux continus asymétriques.	24
3.1	Durée de vie de 50 appareils - Aarset Data.	62
3.2	Durée de survie des patients de l'essai sur le cancer de la tête et du cou : Efron(1988)	68
3.3	les données sur les défaillances ou censures du composant «Bague d'étanchéité HP»	74
3.4	Les estimateurs de K.M de la fonction de survie.	75
3.5	Estimations du taux de hasard et taux de hasard inversé	77
3.6	Quelques valeurs du ISE pour les estimateurs de la fonction de densité et la fonction RHR, pour les modèles considérés dans la simulation, en utilisant le $h[UCV]$	78

Introduction

L'analyse de la survie est un ensemble de procédures statistiques permettant de modéliser le temps restant avant la mort pour des organismes biologiques (l'espérance de vie) ou le temps restant avant l'échec ou la panne dans les systèmes artificiels, ce que l'on représente graphiquement sous la forme d'une courbe de survie. On parle aussi d'analyse de la fiabilité en ingénierie, d'analyse de la durée en économie ou d'analyse de l'histoire d'événements en sociologie. Plus généralement, l'analyse de survie implique la modélisation du facteur temps dans la probabilité d'occurrence des événements, notamment grâce à des concepts tels que le taux de défaillance instantané ou la loi de fiabilité d'un système.

L'analyse de la survie remonte au XVII^e siècle passé dans le domaine de la démographie et avait pour objectif l'estimation de la croissance d'une population à partir des registres de décès et de diverses caractéristiques. À partir du XIX^e siècle, ces analyses se sont affinées avec l'apparition de catégorisations suivant des variables exogène (sexe, nationalité,...), durant ce siècle, les statisticiens ont commencé les premières modélisations concernant la probabilité de mourir à un certain *âge*, probabilité qui sera par la suite désignée sous le terme de fonction de risque. Jusqu'en 1950, la communauté des statisticiens s'intéresse peu à 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 utilisé en analyse de survie (la loi de Weibull) [1].

En 1958, Kaplan et Meier proposent d'utiliser dans le domaine médical un estimateur non paramétrique, ils présentent des résultats importants concernant cet estimation. En 1972, un modèle statistique semi-paramétrique voit le jour grâce aux travaux de Cox, il est intervenu des variables explicatives (exogènes) dans la fonction de risque [21].

Plusieurs fonctions sont utilisées dans l'analyse de survie et sont introduites en science actuarielle, notamment, la fonction de fiabilité ou de survie, la fonction de cotes, la fonction de hasard et de hasard inversé.

La fonction de hasard a fait l'objet de plusieurs travaux, notamment en ce qui concerne l'estimation paramétrique et non paramétrique. Le principal problème de l'approche paramétrique est que les familles de distribution de probabilité classiques existantes sont limitées face à une multitude de structures de données. Une hypothèse erronée concernant le modèle de distribution sous-jacent des données peut conduire à des interprétations trompeuses. Dans de telles situations, les méthodes non paramétriques peuvent être plus appropriées. Les méthodes non paramétriques n'imposent que des hypothèses légères, telles que la régularité, sur la distribution de probabilité sous-jacente et évitent ainsi le risque de spécifier un modèle erroné pour les données. La fonction de hasard est utile dans l'ingénierie, la médecine et les assurances [7].

La fonction de taux de hasard inversé, étroitement liée à la fonction de hasard, est moins intuitive. Bien qu'initialement introduite par la recherche actuarielle, le taux de hasard inversé a jusqu'à présent été principalement appliqué à l'ingénierie de la fiabilité (Desai et al, 2011) [12]. Mais l'ingénierie n'est cependant pas le seul domaine où cet outil s'est avéré utile. Le hasard inversé a également été utilisé pour analyser des données tronquées à droite et censurées à gauche. Kalbfleisch et Lawless (1991) [20], Gross et Huber-Carol (1992) [17] l'ont fait dans le domaine de la médecine. En psychologie, Toznsend et Wenger (2004) [34] l'ont trouvé utile dans la modélisation de la capacité de traitement de l'information.

L'objectif de ce mémoire est d'analyser la fonction du taux de hasard inversé (RHR) qui est peu connue par rapport à la fonction taux de hasard (RH) et étudier la relation entre ces deux fonctions. On s'intéresse aussi à l'estimation de RHR, et on propose l'estimation paramétrique par la méthode MLE (Maximum Likelihood Estimation) et l'estimation non paramétrique par deux méthodes : la méthode de Kaplan-Meier (KM) et la méthode du noyau asymétrique.

Le travail est organisé comme suit :

- Le premier chapitre donne le cadre théorique des fonctions fondamentales d'analyse de survie et les méthodes d'estimation paramétrique et non-paramétrique utilisées dans ce mémoire.
- Le deuxième chapitre concerne l'étude de la fonction de taux de hasard et le taux de hasard inversé et la relation entre les deux.
- Le troisième chapitre présente l'estimation du taux de hasard inversé (RHR) par les méthodes MLE, Kaplan-Meier et à noyaux continus asymétriques (Gamma, Inverse-Gaussien et Inverse-Gaussien Réciproque). Des simulations sont effectuées afin d'analyser les propriétés de cet estimateur.

Chapitre 1

Éléments fondamentaux sur l'analyse de survie et estimation

Introduction

L'objectif de ce chapitre est de fixer le cadre générale de ce travail et d'introduire certaines notions et concepts élémentaires sur l'analyse de survie nécessaires et connus dans la littérature.

1.1 L'analyse de survie

L'analyse des données (durées) de survie est l'étude de délai de la survenue d'un événement. En médecine, l'analyse de survie permet d'évaluer l'efficacité d'un traitement. En démographie, sert à construire des tables de mortalités. En ingénierie, permet d'estimer la fiabilité de machines et des composants électroniques, elle est aussi utile en astrophysique.

Le terme de durée de survie désigne le temps écoulé jusqu'à la survenue d'un événement précis. L'événement étudié (communément appelé "décès") est le passage irréversible entre deux états (communément nommés "vivant" et "décès"). L'événement terminal peut s'agir de l'apparition d'une maladie (par exemple, le temps avant une rechute ou un rejet de greffe), d'une guérison (temps entre le diagnostic et la guérison), la panne d'une machine (durée de fonctionnement d'une machine, en fiabilité) ou la survenue d'un sinistre (temps entre deux sinistres, en actuariat). Cette durée est liée à trois dates particulières définies comme suit :

- **Date d'origine**

L'origine de la durée étudiée peut être la date de naissance, le début d'une exposition à un facteur de risque, la date d'une opération chirurgicale, la date de début d'une maladie

ou la date d'entrée dans l'étude.

- **Date de point**

C'est la date au-delà de laquelle on arrêtera l'étude et on ne tiendra plus compte des informations sur les sujets.

- **Date des dernières nouvelles**

C'est la date la plus récente où les informations sur un sujet ont été recueillies.

On définit également :

- **Recul**

C'est le délai écoulé entre la date d'origine et la date de point.

- **Temps de participation**

C'est le temps écoulé entre :

- la date d'origine et la date de dernières nouvelles, si cette dernière est antérieure à la date de point.
- la date d'origine et la date de point, si celle-ci est antérieure à la date de dernières nouvelles

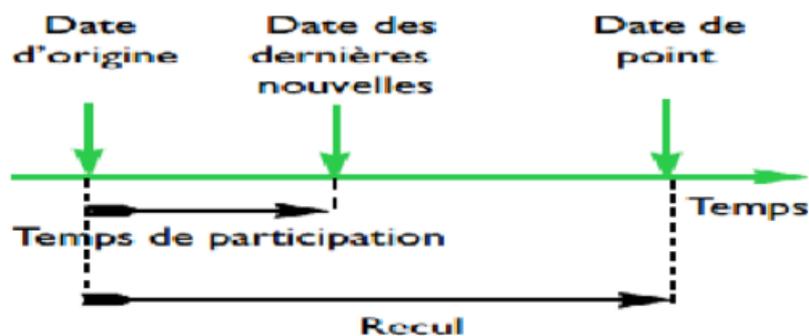


FIGURE 1.1 – Représentation de différentes dates et temps écoulé.

1.2 Données de survie

Pour analyser les données de survie, on est souvent confronté à des données complètes et des données incomplètes.

1.2.1 Données complètes

Des données complètes signifient que des données valides sont fournies sur toutes les mesures spécifiées dans le protocole d'évaluation. Il ne doit pas y avoir de lacunes ou d'informations manquantes pour que les données soient vraiment complètes.

1.2.2 Données censurées

Le phénomène de censure est lié aux événements perturbateurs qui peuvent se produire dans le laps de temps nécessaire au recueil d'une donnée. Il intervient donc fréquemment lors de mesures qui portent sur les variables modélisant le temps écoulé entre deux événements : durée de vie d'un individu, durée entre le début d'une maladie et la guérison,... etc. Ces perturbations empêchent l'observateur d'accéder à la totalité de l'information concernant le phénomène qu'il étudie et conduit à l'apparition d'observations incomplètes dites censurées.

Définition 1.1 (*Variable censure*)

La variable de censure Y est définie par la non-observation de l'événement étudié. Si au lieu d'observer X , on observe Y , et que l'on sait que $X > Y$ (respectivement $X < Y$, $Y_1 < X < Y_2$), on dit qu'il y a censure à droite (respectivement censure à gauche, censure par intervalle). Pour un individu donné i , on va considérer :

- son temps de survie X_i .
- son temps de censure Y_i .

Types de censure

La censure est le phénomène le plus couramment rencontré lors du recueil de données de survie, ce phénomène se fait selon plusieurs mécanismes tels que la censure à droite, la censure à gauche et la censure double (ou mixte).

1. Censure à droite

La variable d'intérêt est dite censurée à droite si l'individu concerné n'a aucune information sur sa dernière observation. Ainsi, en présence de censure à droite les variables d'intérêt ne sont pas toutes observées.

Un exemple typique est celui où l'événement considéré est le décès d'un patient malade et la durée d'observation est une durée totale d'hospitalisation. On trouve aussi ce genre de phénomène dans les études de fiabilité quand la panne d'un appareil ou d'un composant

électronique ne permet pas de continuer l'observation pour un autre appareil ou composant. On peut aussi trouver ces genres de phénomènes en hydrologie, en pluviométrie,.... L'expérimentateur peut fixer une date de fin d'expérience et les observations pour les individus pour lesquels on n'a pas observé l'événement d'intérêt avant cette date seront censurées à droite.

2. Censure à gauche

La censure à gauche correspond au cas où l'individu a déjà subi l'événement avant que l'individu ne soit observé. On sait uniquement que la date de l'événement est inférieure à une certaine date connue.

Par exemple, si on veut étudier en fiabilité un certain composant électronique qui est branché en parallèle avec un ou plusieurs autres composants : le système peut continuer à fonctionner, quoique de façon aberrante, jusqu'à ce que cette panne soit détectée (par exemple lors d'un contrôle ou en cas de l'arrêt du système). Ainsi donc, la durée observée pour ce composant est censurée à gauche.

Dans la vie courante il y a plusieurs phénomènes qui présentent à la fois des données censurées à droite et à gauche.

3. Censure double ou mixte

On dit qu'on a une censure double ou mixte si on a des données censurées à droite et des données censurées à gauche dans le même échantillon. Plusieurs modèles non paramétriques ont été présentés pour l'étude de la double censure, par exemple, le modèle de Turnbull [34].

Dans la littérature d'autres modèles ont été proposé notamment la censure par intervalle.

4. Censure par intervalle

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

Un avantage de ce type est qu'il permet de présenter les données censurées à droite ou à gauche par des intervalles du type $[a, \infty[$ et $[0, a]$, respectivement.

Ces quatre catégories de censure décrites ci-dessus peuvent se présenter en fonction du mode ou mécanisme de censure. Ainsi, dans la littérature on retrouve les types suivants :

- **Censure de type I (fixée)** : L'expérimentateur fixe une valeur (une date par exemple non aléatoire de fin d'expérience). Ce mécanisme de censure est fréquemment rencontré dans les applications industrielles.

- **Censure de type II (attente)** : L'expérimentateur fixe a priori le nombre d'événements à observer. La date de fin d'expérience devient alors aléatoire, le nombre d'événements étant quant à lui, non aléatoire. Ce modèle est souvent utilisé dans les études de fiabilité, d'épidémiologie.
- **Censure de type III (aléatoire)** : C'est typiquement ce modèle qui est utilisé pour les essais thérapeutiques. Dans ce type d'expériences, la date d'inclusion du patient dans l'étude est fixée, mais la date de fin d'observation est inconnue (celle-ci correspond, par exemple, à la durée d'hospitalisation du patient).

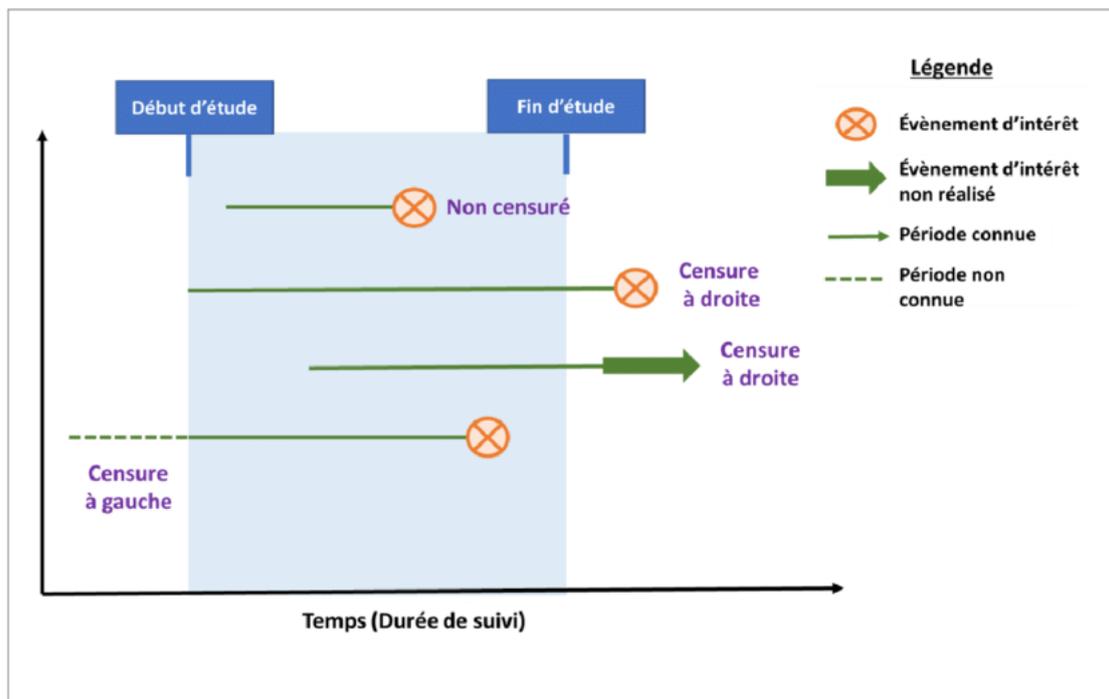


FIGURE 1.2 – Types de censures

Les données censurées ne sont pas le type unique de données incomplètes. L'autre cas classique de données incomplètes est celui des données dites tronquées.

1.2.3 Données tronquées

La troncature, quant à elle, élimine de l'étude une partie des observations X_i lors d'une étude pratique sur les durées de vie, il n'est pas rare que la variable d'intérêt X ne soit pas observable quand elle est inférieure à un seuil aléatoire Y , ce qui aura pour conséquence que l'analyse ne pourra porter que sur la loi conditionnelle de X sachant $X > Y$. Il y a trois types de troncature :

1. Troncature à gauche

Soit Y une variable aléatoire indépendante de X , on dit qu'il y a troncature à gauche lorsque X n'est observable que si $X > Y$. On observe le couple (X, Y) , avec $X > Y$.

Par exemple, si la durée de vie d'une population est étudiée à partir d'une cohorte tirée au sort dans cette population, seule la survie des sujets vivants à l'inclusion pourra être étudiée (il y a troncature à gauche car seuls les sujets ayant survécu jusqu'à la date d'inclusion dans la cohorte sont observables).

2. Troncature à droite

De même, il y a troncature à droite lorsque X n'est observable que si $X < Y$.

3. Troncature par intervalle

Quand une durée est tronquée à droite et à gauche, on dit qu'elle est tronquée par intervalle. Par exemple, on rencontre ce type de troncature lors de l'étude des patients d'un registre : les patients diagnostiqués avant la mise en place du registre ou répertoriés après la consultation du registre ne seront pas inclus dans l'étude.

1.3 Distributions de la durée de survie

La durée de survie T est une variable aléatoire (*v.a*) positive ou nulle, et absolument continue de distribution entièrement caractérisée par l'une des fonctions suivantes, en considérant l'événement d'intérêt "décès".

1.3.1 Fonction de survie

Définition 1.2 La fonction de survie notée par $S(t)$ ou $\bar{F}(t)$ est définie sur \mathbb{R}_+ pour $t > 0$, par :

$$S(t) = \bar{F}(t) = \mathbb{P}(T > t). \quad (1.1)$$

Pour t fixé c'est la probabilité de survivre jusqu'à l'instant t .

Propriété 1 La fonction de survie d'une *v.a* T est décroissante monotone continue à gauche et vérifie

$$\lim_{t \rightarrow 0} S(t) = 1 \quad \text{et} \quad \lim_{t \rightarrow \infty} S(t) = 0.$$

1.3.2 Fonction de répartition

Définition 1.3 La fonction de répartition (ou c.d.f pour "cumulative distribution function") de la v.a T est la fonction sur \mathbb{R}_+ définie par :

$$\begin{aligned} F : \mathbb{R}_+ &\longrightarrow [0, 1] \\ t &\longrightarrow \mathbb{P}(T \leq t) \end{aligned}$$

Pour t fixé, c'est la probabilité de mourir avant l'instant t .

Propriété 2 La fonction F est une fonction croissante monotone continue à droite telle que,

$$\lim_{t \rightarrow 0} F(t) = 0 \quad \text{et} \quad \lim_{t \rightarrow \infty} F(t) = 1.$$

Si la v.a est discrète, il s'agit d'une fonction en escalier. Soit \mathbb{P}_i la probabilité que T prenne la valeur t_i .

Appliquons la propriété des probabilités totales :

$$F(t) = \sum_{t_i \leq t} \mathbb{P}_i.$$

1.3.3 Fonction de densité

Définition 1.4 Si la fonction $f(t) \geq 0$ telle que pour tout $t \geq 0$,

$$F(t) = \int_0^t f(u) du,$$

et si F est dérivable sur \mathbb{R}_+ au point t , la fonction de densité de probabilité existe et définit pour tout $t \geq 0$ par

$$f(t) = \lim_{dt \rightarrow \infty} \frac{\mathbb{P}(t \leq T \leq t + dt)}{dt} = F'(t) = -S'(t). \quad (1.2)$$

Pour t fixé, la densité de probabilité caractérise la probabilité de mourir dans un petit intervalle de temps après l'instant t , relativement à son amplitude.

1.3.4 Fonctions du risque et du risque cumulé

Appelé selon les domaines d'application : "taux instantané ou cumulé de défaillance, de risque, de hasard" ou encore "quotient de mortalité".

Risque instantané λ (ou taux de hasard, ou taux de défaillance)

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) :

$$\lambda(t) = \lim_{h \rightarrow 0} \frac{\mathbb{P}(t \leq T < t + h | T \geq t)}{h} = \lim_{h \rightarrow 0} \frac{1}{h} \left[\frac{F(t+h) - F(t)}{S(t)} \right]. \quad (1.3)$$

D'où,

$$\lambda(t) = \frac{f(t)}{S(t)} = \frac{f(t)}{1 - F(t)} = -\ln(S(t))', \quad t > 0. \quad (1.4)$$

Taux de hasard cumulé Λ

Le taux de hasard cumulé est l'intégrale du risque instantané λ :

$$\Lambda(t) = \int_0^t \lambda(u) du = -\ln(S(t)), \quad t > 0. \quad (1.5)$$

On peut déduire de cette formule une expression de la fonction de survie en fonction du taux de hasard cumulé (ou du risque instantané) :

$$S(t) = \exp(-\Lambda(t)) = \exp\left(-\int_0^t \lambda(u) du\right), \quad t > 0. \quad (1.6)$$

D'où la densité et la fonction de répartition sont respectivement :

$$f(t) = \lambda(t) \exp\left(-\int_0^t \lambda(u) du\right) = \lambda(t) S(t) \quad (1.7)$$

et

$$F(t) = 1 - \exp\left(-\int_0^t \lambda(u) du\right) = 1 - \exp(-\Lambda(t)). \quad (1.8)$$

Fonctions empiriques de répartition et de survie

Soit T_1, \dots, T_n un échantillon de taille $n \geq 1$ d'une *v.a* positive T de fonction de répartition F et de fonction de survie S . Les fonctions empiriques de répartition et de survie, F_n et S_n sont respectivement définies par :

$$F_n(t) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{T_i \leq t}, \quad t \geq 0 \iff F_n(t) = \begin{cases} 0 & \text{si } t < T_{1:n} \\ \frac{i}{n} & \text{si } T_{1:n} \leq t < T_{i+1:n} \\ 1 & \text{si } t \geq T_{n:n} \end{cases} \quad (1.9)$$

et

$$S_n(t) = 1 - F_n(t) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{T_i > t}, \quad t \geq 0 \iff S_n(t) = \begin{cases} 1 & \text{si } t < T_{1:n} \\ 1 - \frac{i}{n} & \text{si } T_{1:n} \leq t \leq T_{i+1:n} \\ 0 & \text{si } t > T_{n:n} \end{cases} \quad (1.10)$$

1.3.5 Comparaison stochastique

Soient X et Y deux *v.a* positives continues avec des fonctions de répartitions $F_X(t)$ et $F_Y(t)$ et des fonctions de densité $f_X(t)$ et $f_Y(t)$ respectivement [14]. On dit alors que X est plus petit que Y en

- Ordre du rapport de vraisemblance ($X \leq_{lr} Y$) si $\frac{f_X(t)}{f_Y(t)}$ décroît en $t > 0$.
- Ordre de taux de hasard ($X \leq_{hr} Y$) si $\lambda_X(t) \geq \lambda_Y(t)$ pour tous $t > 0$.
- Ordre stochastique ($X \leq_{st} Y$) si $S_X(t) \leq S_Y(t)$ pour tous $t > 0$.
- Ordre de durée de vie résiduelle moyenne ($X \leq_{MRL} Y$) si $\mu_X(t) \leq \mu_Y(t)$ pour tous $t > 0$, où $\mu_X(t) = \frac{\int_t^\infty S(x)dx}{S(t)}$ et $\mu_Y(t) = \frac{\int_t^\infty S(y)dy}{S(t)}$.

1.4 Quantités associées à la distribution de survie

1.4.1 Moyenne de la durée de survie

Le temps moyen de survie $E(T)$ est définie par :

$$E(T) = \int_0^{+\infty} t f(t) dt = \int_0^{+\infty} S(t) dt.$$

Le moment d'ordre 2 :

$$E(T^2) = \int_0^{+\infty} t^2 f(t) dt = 2 \int_0^{+\infty} t S(t) dt. \quad (1.11)$$

1.4.2 Moyenne résiduelle de la durée de survie

La moyenne résiduelle de la durée de survie est donnée par :

$$\mu(t) = \frac{\int_t^{+\infty} S(x) dx}{S(t)}. \quad (1.12)$$

1.4.3 Variance de la durée de survie

La variance de la durée de survie $Var(T)$ est définie par :

$$Var(T) = 2 \int_0^{+\infty} t S(t) dt - (E(T))^2.$$

Remarque 1.1 *On peut déduire l'espérance et la variance à partir de n'importe laquelle des fonctions F, S, f, λ où Λ (mais pas l'inverse).*

1.5 Estimation paramétrique

Lorsque l'on souhaite ajuster un modèle paramétrique sur un échantillon de données de survie, nous devons choisir le modèle approprié. L'objectif est d'identifier ce modèle par l'estimation de ses paramètres. Pour le faire, plusieurs méthodes existent : méthode des moment, méthode des graphique, méthode du maximum du vraisemblance, méthode des moindres carrés (MMC), ... [24].

1.5.1 La méthode du maximum de vraisemblance (MLE)

Soit $f(t, \theta_1, \theta_2, \dots, \theta_p)$ la fonction de densité où $\theta_i, i = 1, \dots, p$, sont des paramètres inconnus.

La vraisemblance des observation s'écrit :

$$L(t, \theta) = \prod_{i=1}^p f(t, \theta_i), \quad \text{avec } \theta = (\theta_1, \theta_2, \dots, \theta_p) \quad (1.13)$$

Les estimateurs de maximum de vraisemblance sont ceux qui vérifient :

$$\frac{\partial L(t, \theta)}{\partial \theta_i} = 0 \quad \text{et} \quad \frac{\partial^2 \log L(t, \theta)}{\partial \theta_i^2} < 0.$$

Il s'agit de résoudre les systèmes suivants :

$$\left\{ \begin{array}{l} \frac{\partial L(t, \theta_1)}{\partial \theta_1} = 0 \\ \frac{\partial L(t, \theta_2)}{\partial \theta_2} = 0 \\ \cdot \\ \cdot \\ \frac{\partial L(t, \theta_n)}{\partial \theta_p} = 0 \end{array} \right. \quad \text{et} \quad \left\{ \begin{array}{l} \frac{\partial^2 L(t, \theta_1)}{\partial \theta_1^2} = 0 \\ \frac{\partial^2 L(t, \theta_2)}{\partial \theta_2^2} = 0 \\ \cdot \\ \cdot \\ \frac{\partial^2 L(t, \theta_n)}{\partial \theta_p^2} = 0 \end{array} \right.$$

1.5.2 Estimation par intervalle de confiance

L'estimation ponctuelle d'un paramètre θ donne une valeur numérique unique à ce paramètre, mais, n'apporte aucune information sur la précision des résultats, c-à-d qu'elle ne tient pas compte des erreurs ducs aux fluctuations d'échantillonnage, par exemple pour évaluer la confiance que l'on peut avoir en une estimation, il est nécessaire de lui associer un intervalle qui contient avec une certaine probabilité la vraie valeur du paramètre, c'est l'estimation par intervalle de confiance.

Définition 1.5 Soit (T_1, \dots, T_n) un échantillon de T . On appelle intervalle de confiance (IC), de niveau de confiance $1 - \alpha$ telles que $\alpha \in [0, 1]$ donné, un intervalle aléatoire $[\theta_1, \theta_2]$ où $\theta_1 \leq \theta_2$ sont deux statistiques, fonction de l'échantillon, telles que :

$$\mathbb{P}(\theta_1 < \theta < \theta_2) = 1 - \alpha.$$

Remarque 1.2 α est donc la probabilité que $[\theta_1, \theta_2]$ ne recouvre pas la vraie valeur du para-

mètre. C'est le seuil de confiance.

1.6 Estimation non paramétrique

1.6.1 Estimateur de Kaplan-Meier de la survie (KM)

En pratique on estime la fonction de survie à partir des données collectées, cette approximation s'appelle la courbe de Kaplan-Meier.

L'estimateur de Kaplan-Meier (Kaplan-et Meier, 1958) aussi dénommé estimateur "produit-limite", est l'estimateur du maximum de vraisemblance non-paramétrique de la fonction de survie. Il permet d'estimer la fonction de survie $S(t)$ à partir d'un échantillon de n sujets avec des durées de survie qui peuvent être censurées à droite [21].

L'estimateur de Kaplan-Meier est donné par :

$$\hat{S}(t) = \prod_{T_i \leq t} \left(1 - \frac{d_i}{n_i}\right), \quad t > 0. \quad (1.14)$$

Avec

- d_i : le nombre d'évènements à T_i
- $T_i, (i = 1, \dots, n)$: les temps d'évènements (décès et censure) distincts ordonnés.
- n_i : le nombre d'individus à risque de subir l'évènement juste avant le temps T_i

Espérance et variance de l'estimateur

- **Biais de l'estimateur de Kaplan-Meier**

De nombreux auteurs se sont intéressés au calcul du biais de l'estimateur de Kaplan-Meier. Un premier résultat dû à Gill (1980) dit qu'à moins qu'aucune observation ne soit censurée, l'estimateur de Kaplan-Meier de la fonction de survie présente un biais positif, c-à-d :

$$E(\hat{S}(t) - S(t)) \geq 0$$

- **Variance de l'estimateur de Kaplan-Meier**

L'erreur standard de l'estimateur de Kaplan-Meier est approchée selon la formule de

Greenwood, au temps de décès T_i , par :

$$\sigma(\hat{S}(T_i)) = [\hat{S}(T_i)]^2 \sum_{i=1}^n \frac{d_i}{n_i(n_i - d_i)} \quad (1.15)$$

Exemple 1.1 Calculons l'estimateur de Kaplan-Meier sur les données de Frechet (1963) "Étude des durées de rémission, exprimées en semaines, des sujets atteints de leucémie, selon qu'ils ont reçu de 6-mercaptopurine ou un placebo". Les 21 observations se présentent de la manière suivante, le signe + indiquant une donnée censurées à droite) [1] :

6-MP	6	6	6	6+	7	9+	10	10+	11+	13	16
	17+	19+	20+	22	23	25+	32+	32+	34+	35+	
Placébo	1	1	2	2	3	4	4	5	5	8	8
	8	8	11	11	12	12	15	17	22	23	

L'estimateur de Kaplan-Meier de la survie du ce groupe de 21 malades traités par le traitement 6MP se calcule par la formule (1.14) et donne le tableau (1.1)

temps	n_i	d_i	$\hat{S}_{6MP}(t)$
0	21	0	1
6	21	3	0.875
7	17	1	0.807
10	15	1	0.753
13	12	1	0.690
16	11	1	0.627
22	7	1	0.538
23	6	1	0.448

TABLE 1.1 – L'estimateur de K.M de la fonction de survie du groupe de 21 malades traités par le traitement 6MP

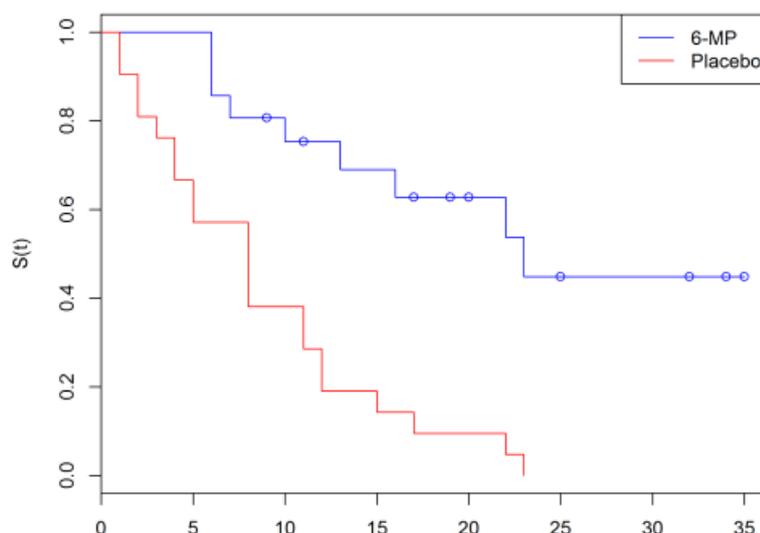


FIGURE 1.3 – Les estimateurs de K.M de la fonction de survie des 21 malades traités par le traitement 6MP

Interprétation

Le graphe dans la figure (1.3) présente les estimateurs de la fonction de survie par Kaplan-Meier pour les deux traitements. Nous remarquons que la courbe de survie 6MP est nettement au dessus de celle du placebo. D'où l'efficacité du traitement 6MP.

1.6.2 Estimation à noyau asymétrique de la densité

Les définitions suivantes présentent les notions du noyau associé, et de l'estimateur à noyau associé pour la fonction de densité f inconnue sur le support D .

Définition 1.6 Soit $t \in D$ et $h > 0$. On appelle noyau associé $K_{t,h}$ toute densité de probabilité liée à une variable aléatoire continue ou discrète $\mathcal{K}_{t,h}$ de support $D_{t,h}$ vérifiant les conditions suivantes :

- $D_{t,h} \cap D \neq \emptyset$,
- $\cup D_{t,h} \supseteq D$,
- $\lim_{h \rightarrow 0} E(\mathcal{K}_{t,h}) = t$,
- $\lim_{h \rightarrow 0} Var(\mathcal{K}_{t,h}) = 0$.

Noyau associé continu asymétrique

L'estimateur à noyau associée continu asymétrique est approprié pour estimer des densités à support compact et bornées. Soit T_1, T_2, \dots, T_n un échantillon de $v.a$ (iid) de densité de probabilité continue inconnue f à support $D = [a, b]$, avec $(a, b \in \mathbb{R})$. L'estimateur à noyau continu est de la forme :

$$\hat{f}_h(t) = \frac{1}{n} \sum_{i=1}^n K_{t,h}(T_i). \quad (1.16)$$

où h est le paramètre de lissage et $K_{t,h}$ est associé au noyau continu asymétrique qui dépend du paramètre de lissage h et de la cible t .

Noyau	$K(u)$	Espérance	Variance
$Gamma(\frac{t}{h} + 1, h)$	$\frac{u^{\frac{t}{h}} \exp(-\frac{u}{h})}{h^{\frac{t}{h}+1} \Gamma(\frac{t}{h} + 1)}$	$t + h$	$h(t + h)$
$Inv.Gam(\frac{1}{th} - 1, \frac{1}{h})$	$\frac{1}{\Gamma(\frac{1}{th} - 1)} u^{\frac{1}{th}-1} \exp(-\frac{1}{uh}) (\frac{1}{h})^{\frac{1}{th}-1}$	$\frac{1}{t} - 2$	$\frac{(\frac{1}{h})^2}{(\frac{1}{ht} - 2)^2 (\frac{1}{ht} - 3)}$
$Inv.Gauss(t, \frac{1}{h})$	$\frac{1}{\sqrt{2h\pi u^3}} \exp(-\frac{1}{2ht} (\frac{u}{t} - 2 + \frac{t}{u}))$	t	ht^3
$Recip.Inv.Gau(\frac{1}{t-h}, \frac{1}{h})$	$\frac{1}{\sqrt{2h\pi u}} \exp(-\frac{(t-h)}{2h} (\frac{u}{t-h} - 2 + \frac{(t-h)}{u}))$	t	$h(t + h)$

TABLE 1.2 – Quelques noyaux continus asymétriques.

Propriétés de l'estimateur

- **Biais :**

$$Biais(\hat{f}_h(t)) = f[E(\mathcal{K}_{t,h})] + \frac{1}{2} Var(\mathcal{K}_{t,h}) f''(t) - f(t) + o(h^2).$$

- **La variance :**

$$Var(\hat{f}_h(t)) = \frac{1}{n} \int_{\mathbb{R}_+} K_{t,h}^2 f(u) du - \frac{1}{n} (Biais[\hat{f}_h(t)] + f(t))^2.$$

- **MSE :**

$$MSE(\hat{f}_h(t)) = \left[f[E(\mathcal{K}_{t,h})] + \frac{1}{2} Var(\mathcal{K}_{t,h}) f''(t) - f(t) + o(h^2) \right]^2 + \frac{1}{n} \int_{\mathbb{R}_+} K_{t,h}^2 f(u) du - \frac{1}{n} (Biais[\hat{f}_h(t)] + f(t))^2.$$

- **MISE :**

$$\begin{aligned}
MISE(\hat{f}_h(t)) &= \int_D \left[f[E(\mathcal{K}_{t,h})] + \frac{1}{2}Var(\mathcal{K}_{t,h})f''(t) - f(t) \right]^2 dt \\
&+ \frac{1}{n} \int_D \left(\int_{\mathbb{R}_+} K_{t,h}^2 f(u) du - (Bias[\hat{f}_h(t)] + f(t))^2 \right) dt \quad (1.17)
\end{aligned}$$

Quelques résultats de convergence de l'estimateur à noyau asymétrique [9]

- Si $\frac{1}{n} \int_{\mathbb{R}_+} K_{t,h}^2(u)f(u)du \rightarrow 0$ et $h = h(n) \rightarrow 0$ quand $n \rightarrow \infty$, alors

$$\lim_{n \rightarrow \infty} MISE[\hat{f}_h(t)] = 0, \quad \forall t > 0.$$

- Si $\frac{1}{n} \int_{\mathbb{R}_+} \left[\int_{\mathbb{R}_+} K_{t,h}^2(u)f(u)du \right] dt = 0$ et $h = h(n) \rightarrow 0$ quand $n \rightarrow \infty$, alors

$$\lim_{n \rightarrow \infty} MISE[\hat{f}_h] = 0.$$

- Si f a un support dans \mathbb{R}_+ , continue et bornée sur un intervalle compact $I \subset \mathbb{R}_+$, avec \hat{f}_G son estimateur à noyau gamma

$$\begin{aligned}
\sup_{t \in I} |\hat{f}_G(t) - f(t)| &\xrightarrow{P} \quad sih + (nh^2)^{-1} \rightarrow 0, \quad n \rightarrow \infty \\
\sup_{t \in I} |\hat{f}_G(t) - f(t)| &\xrightarrow{a.s} \quad sih + \log(n)/(nh^2) \rightarrow 0, \quad n \rightarrow \infty
\end{aligned}$$

1.6.3 Choix du paramètre de lissage

- **Méthode de plug-In**

Dans le cas d'un estimateur à noyau asymétrique, le paramètre de lissage est obtenue par la minimisation du MISE asymptotique noté AMISE, i.e.,

$$h^* = \operatorname{argmin}[MISE(\hat{f}_h(t))]$$

Cependant, le paramètre de lissage obtenue dépend des fonctions inconnues f, f' et f'' , ce qui rend les calculs impossibles. Scaillet (2004) [29] a proposé une méthode empirique et a remplacé les fonctions inconnues f, f' et f'' en choisissant la distribution lognormale comme référence pour la densité f avec les paramètres μ et σ , c-à-d, $f \sim \mathcal{LN}(\mu, \sigma)$. Les paramètres de lissage optimales obtenues par Scaillet (2004) en utilisant Inverse Gaussian

(IG) et la gaussienne inverse réciproque (RIG) sont présentées respectivement comme suit

$$h_{IG} = \left[\frac{16\sigma^2 \exp(\frac{1}{8})(7\sigma^2 - 20\mu)}{12 + 68\sigma^2 + 255\sigma^4} \right]^{\frac{2}{5}} n^{-\frac{2}{5}},$$

et

$$h_{RIG} = \left[\frac{16\sigma^5 \exp(\frac{1}{8}(-17\sigma^2 + 20\mu))}{12 + 4\sigma^2 + \sigma^4} \right]^{\frac{2}{5}} n^{-\frac{2}{5}}.$$

Remarque 1.3 *On peut utiliser n'importe quel modèle de référence.*

- **Methode de UCV (Unbiased Cross Validation)**

Le principe de cette méthode consiste à estimer le ISE par la technique de validation croisée et par la suite de sélectionner le paramètre de lissage qui minimise l'estimateur, la formule de ISE suivante :

$$ISE(\hat{f}_h) = \int_0^\infty (\hat{f}(t) - f(t))^2 dt = \int_0^\infty (\hat{f}(t))^2 dt + \int_0^\infty \hat{f}(t)f(t)dt + \int_0^\infty (f(t))^2 dt$$

Comme le dernier terme ne dépend pas de h , il s'agit de minimiser UCV_f :

$$UCV_f[h] = \int_0^\infty (\hat{f}(t))^2 dt - 2 \int_0^\infty (\hat{f}(t)f(t)) dt$$

Donc un estimateur de $UCV_f(h)$:

$$UCV[h] = \int_0^\infty (\hat{f}(t))^2 dt - \frac{2}{n} \sum_{i=1}^n \hat{f}_{h,t}(t_i)$$

avec $\hat{f}_{h,t}(t_i) = \frac{1}{n-1} \sum_{\substack{j=1 \\ j \neq i}}^n K(t_j)$.

D'où le h_{UCV} optimal est donné par :

$$h_{UCV} = \underset{h>0}{\operatorname{argmin}} UCV[h].$$

- **Approche bayésienne**

Considérant T_1, T_2, \dots, T_n , n v.a i.i.d, de densité f et $t = (t_1, t_2, \dots, t_n)$, et $h \in H \subset \mathbb{R}$ le paramètre à estimer.

Cette approche considère le paramètre inconnu h comme une v.a. avec une distribution a priori $\pi(h)$, et combine à la fois l'information du paramètre h (information a priori) et l'information conduite par les données, pour fournir une information a posteriori du

paramètre h .

Dans ce cas, la loi a posteriori s'obtient par la formule de Bayes,

$$\pi(h/t) = \frac{f(t/h)\pi(h)}{\pi(t)},$$

où $\pi(h/t) = \prod_{i=1}^n f(t_i, h)$ représente la fonction du maximum de vraisemblance, et $\pi(t) = \int_H \prod_{i=1}^n f(t_i, h)\pi(h)dh$ est la loi marginale.

Dans certains cas, il est difficile d'obtenir une forme explicite de $\pi(h/t)$, donc les méthodes MCMC sont généralement utilisées pour surmonter ce problème.

– Méthodes MCMC

Les méthodes MCMC sont utilisées pour approximer la distribution a posteriori d'un paramètre d'intérêt par échantillonnage aléatoire. Le principe est de générer une chaîne de Markov, en utilisant la transition noyau et une valeur initiale arbitraire. Après un nombre d'itération, suffisamment grand, la chaîne de Markov converge vers la densité a posteriori d'intérêt.

Le but principal de cette approche est de trouver un estimateur \hat{h} du paramètre h qui minimise le coût moyen, appelé aussi risque bayésien $E(\mathbb{C}(\hat{h} - h))$, ou $\mathbb{C}(\hat{h} - h)$ représente la fonction du coût, tel que :

$$E(\mathbb{C}(\hat{h} - h)) = \int_{\mathbb{H}} \mathbb{C}(\hat{h} - h)\pi(h/t)dh.$$

L'estimateur \hat{h} qui minimise le risque bayésien en utilisant le coût quadratique moyen, donnée par

$$\hat{h} = E(h/t) = \int_{\mathbb{H}} h\pi(h/t)dh.$$

Il existe trois techniques essentielles pour la sélection du paramètre de lissage optimal de l'estimateur du noyau \hat{f}_h en utilisant l'approche bayésienne [36] :

- * Approche bayésienne globale
- * Approche bayésienne locale
- * Approche bayésienne adaptative

1.7 Distributions de la durée de vie

Nous utilisons le terme « distributions de durée de vie » pour décrire la collection de distributions de probabilités statistiques que nous utilisons dans l'ingénierie de la fiabilité et l'analyse des données de durée de vie [16].

Distribution exponentielle

On dit que T suit une loi exponentielle de paramètres $\lambda > 0$ si :

- Sa fonction de densité est définie par :

$$f(t) = \lambda \exp(-\lambda t), \quad \forall t > 0.$$

- Sa fonction de survie est donné par :

$$S(t) = \exp(-\lambda t), \quad \forall t > 0.$$

- Sa fonction de hasard est donné par :

$$\lambda(t) = \lambda, \quad \forall t > 0.$$

Remarque 1.4 Une distribution continue a un taux de hasard constant si et seulement si c'est une distribution exponentielle.

Distribution gamma

On dit que T suit une loi de gamma de paramètres $\alpha > 0$ et $\beta > 0$ si :

- Sa fonction de densité est définie par :

$$f(t) = \frac{\beta^\alpha}{\Gamma(\alpha)} t^{\alpha-1} \exp(-\beta t), \quad \forall t \geq 0; \quad \alpha, \beta > 0.$$

$$\text{ou } \Gamma(\alpha) = \int_0^\infty t^{\alpha-1} \exp(-t) dt.$$

- Sa fonction de survie est donné par :

$$S(t) = 1 - \frac{1}{\Gamma(\alpha)} \int_0^{\beta t} u^{\alpha-1} \exp(-u) du, \quad \forall t > 0.$$

- Sa fonction de hasard est donné par :

$$\lambda(t) = \frac{\frac{\beta^\alpha}{\Gamma(\alpha)} t^{\alpha-1} \exp(-\beta t)}{1 - \frac{1}{\Gamma(\alpha)} \int_0^{\beta t} u^{\alpha-1} \exp(-u) du}, \quad \forall t > 0.$$

Distribution Log-normale

T suit une loi log-normale avec des paramètres m et σ^2 , notée par $\mathcal{LN}(m, \sigma^2)$, alors,

- Sa fonction de densité est définie par :

$$f(t) = \frac{1}{t\sigma\sqrt{2\pi}} \exp\left[-\frac{(\ln(t) - m)^2}{2\sigma^2}\right], \quad \forall t > 0.$$

- Sa fonction de survie est donné par :

$$S(t) = 1 - \Phi\left(\frac{\ln(t) - m}{\sigma}\right), \quad \forall t > 0.$$

- Sa fonction de taux de hasard est donné par :

$$\lambda(t) = \frac{\phi\left[\frac{\ln(t) - m}{\sigma}\right]}{t\sigma\left[1 - \Phi\left[\frac{\ln(t) - m}{\sigma}\right]\right]}, \quad \forall t > 0.$$

Distribution Weibull

La distribution de Weibull avec les paramètres η et β notée $W(\eta, \beta)$ peut être considérée comme une généralisation de la distribution exponentielle.

- Sa fonction de densité est définie par :

$$f(t) = \frac{\beta}{\eta} \left(\frac{t}{\eta}\right)^{\beta-1} \exp\left[-\left(\frac{t}{\eta}\right)^\beta\right], \quad t > 0.$$

- Sa fonction de survie est donné par :

$$S(t) = \exp\left[-\left(\frac{t}{\eta}\right)^\beta\right], \quad t > 0.$$

- Sa fonction de hasard est donné par :

$$\lambda(t) = \frac{\beta}{\eta} \left(\frac{t}{\eta} \right)^{\beta-1}, \quad t > 0.$$

1.8 Conclusion

Dans ce chapitre, nous avons présenté les notions et éléments fondamentaux sur l'analyse de survie et de l'estimation paramétrique et non paramétrique.

Chapitre 2

Taux de hasard et taux de hasard inversé

2.1 Le taux de hasard (HR)

La notion de taux de hasard est bien connue dans la littérature. Le taux de hasard ou le taux de défaillance instantané joue un rôle crucial dans l'analyse de survie et de la fiabilité. Il est noté h en survie et λ en fiabilité, il est défini comme la probabilité conditionnelle d'une panne (décès) étant donné que le système est en état de marche (en survie) jusqu'à cette date (actuellement).

C'est une mesure de performance du système ou du composant ayant une durée de vie définie. Supposons que la durée de survie d'un système (ou composant) T est une variable aléatoire positive de fonction de répartition $F(t) = P(T \leq t), t > 0$, supposée absolument continue, alors la *v.a* T a une fonction de densité de probabilité $f(t) = \frac{dF(t)}{dt}, t > 0$.

On note $S(t) = 1 - F(t), t > 0$, la fonction de survie associée.

Le taux de hasard est définie par :

$$\lambda(t) = \lim_{dt \rightarrow 0} \frac{1}{dt} [P(t \leq T \leq t + dt | T \geq t)] = \frac{f(t)}{1 - F(t)}, \quad t > 0. \quad (2.1)$$

Le graphique de la fonction du taux de hasard est en général une "courbe en baignoire", où l'on distingue trois périodes (voir la figure (2.1)) :

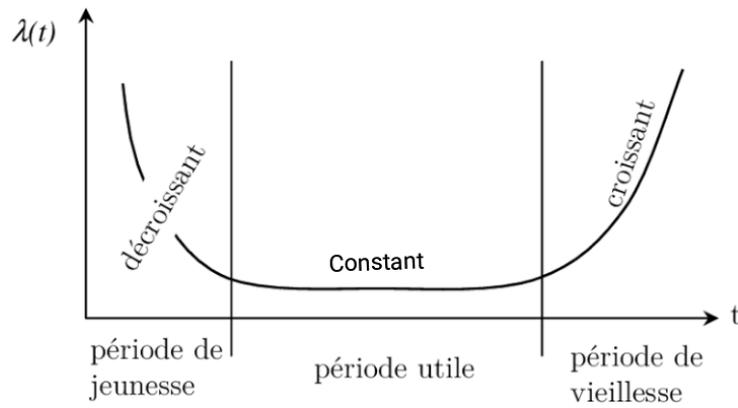


FIGURE 2.1 – Courbe en baignoire de HR.

1. **Le taux de hasard décroissant (période de jeunesse)** : Lorsque le risque de panne diminue avec le temps, on parle de «mortalité infantile» : les systèmes ayant des «défauts de jeunesse» ont des défaillances précoces, les systèmes qui «survivent» sont intrinsèquement robustes. Cela peut aussi décrire une situation de rodage. En survie, on parle de l'exposition d'une personne à une maladie particulière.
2. **Taux de hasard constant (période de vie utile)** : Lorsque le taux de hasard est constant, cela signifie que le risque de panne est totalement aléatoire, on a un système dit «sans effet de mémoire, sans usure, il n'y a pas de cumul de dommage». C'est typiquement la situation des systèmes électroniques. Dans le domaine de la survie, cette période représente le patient recevant un traitement, ce qui conduit à la stabilisation de son état.
3. **Taux de hasard croissant (période de vieillesse)** : Lorsque le risque de panne augmente avec le temps, cela indique un phénomène d'usure. C'est typiquement le cas des systèmes mécanique. En survie, à un moment donné, la personne peut être exposée à une autre maladie ou peut-être l'aggravation de la maladie.

2.1.1 Propriétés

1. $\Lambda(t) = \int_0^t \lambda(u)du < +\infty, \forall t > 0$, mais, $\int_0^{+\infty} \lambda(u)du = +\infty$.
2. $\lambda(\cdot)$ n'est pas nécessairement monotone, mais peut être localement monotone.
3. $\lambda(t) = \frac{f(t)}{S(t)} = -\frac{S'(t)}{S(t)} = -\frac{d}{dt} \ln S(t), \quad t > 0$.

2.1.2 Fonction exponentiée F^*

Soit T une v.a non négative représentant la durée de vie d'un composant ayant une fonction de distribution F avec $F(0) = 0$ et une fonction de densité f . Le taux de hasard de T est alors donné par [28]

$$\lambda(t) = \frac{f(t)}{S(t)}, \quad \text{avec, } S(t) = 1 - F(t), t > 0.$$

Soit T^* une v.a non négative, la fonction de distribution $F^*(t)$ est dite une fonction exponentiée de $F(t)$, si :

$$F^*(t) = [F(t)]^\theta, \quad t > 0, \text{ avec, } \theta > 0. \quad (2.2)$$

La densité et le taux de hasard de T^* sont donnés par :

$$f^*(t) = \frac{d}{dt}F^*(t) = \frac{d}{dt}[(F(t))^\theta] = \theta[F(t)]^{\theta-1}f(t), \quad t > 0. \quad (2.3)$$

et

$$\lambda^*(t) = \frac{f^*(t)}{S^*(t)} = \frac{f^*(t)}{1 - F^*(t)} = \frac{\theta[F(t)]^{\theta-1}f(t)}{1 - [F(t)]^\theta}, \quad t > 0. \quad (2.4)$$

D'où

$$\lambda^*(t) = \theta\lambda(t)g(t), \quad t > 0, \quad (2.5)$$

$$\text{où } g(t) = \frac{[1 - F(t)]F^{\theta-1}}{1 - F^*(t)}, \quad t > 0.$$

Critères de vieillissement [28]

Soit T une v.a positive continue représentant la durée de vie d'un composant. Soit F sa fonction de distribution et $S(t) = 1 - F(t)$ sa fonction de survie. Alors $S_t(x) = P(T > x + t | T > t)$ est la fonction de survie d'une unité d'âge t . Il est évident que toute étude du phénomène du vieillissement doit être basée sur $S_t(x)$ et les fonctions qui lui sont liées.

Définition 2.1 On dit que F est PF_2 (a la propriété du rapport de vraisemblance croissant) si $\ln(f(t))$ est concave, où $f(\cdot)$ est la densité correspondant à $F(\cdot)$, c-à-d que $\frac{f(x+t)}{f(x)}$ est décroissante en x pour tout $t > 0$.

Définition 2.2 On dit que F est IFR (DFR) si $S(x) = \frac{S(x+t)}{S(t)}$ est décroissant (croissant) en t . Si F est absolument continue avec une densité f , alors F est dans la classe IFR(DFR) si le taux de défaillance est croissant (décroissant), i.e, $\lambda_F(t) = \frac{f(t)}{S(t)}$ est croissant (décroissant).

Définition 2.3 On dit que F a un taux de défaillance moyen croissant (décroissant), notée IFRA (DFRA) si $\int_0^t r_F(x)dx/t$ est croissant (décroissant).

Définition 2.4 On dit de F qu'il est meilleur (pire) neuf qu'usagé, notée NBU(NWU) si $S_t(x) \leq (\geq)S(x)$ pour tout $x \geq 0, t \geq 0$.

Définition 2.5 On dit que F a une fonction de durée de vie résiduelle moyenne décroissante (croissante), DMRL(IMRL), si la durée de vie résiduelle moyenne $\mu_F(t) = \int_t^\infty \frac{S(x)dx}{S(t)}$ est décroissante (croissante), en supposant que la moyenne $\mu_F(0)$ existe.

Définition 2.6 On dit que F a un nouveau meilleur (pire) que celui utilisé dans l'attente, noté NBUE (NWUE) si $\mu_F(t) \leq (\geq)\mu_F(0)$ pour tout $t \geq 0$.

• **Héritage de la notion de vieillissement par F^***

- Si $\theta > (<) 1$ et $F \in IFR(DFR)$, alors, $F^* \in IFR(DFR)$.
- Si $\theta > (<) 1$ et $F \in NBU(NWU)$, alors, $F^* \in NBU(NWU)$.
- Si $\theta > (<) 1$ et $F \in PF_2$, alors, $F^* \in PF_2$.
- Si $\theta > 1$ et $\frac{d^2}{dt^2} \ln(F(t)) < \frac{1}{(1-\theta)\frac{d^2}{dt^2} \ln(f(t))}$, $\forall t > 0$, alors $F^* \in IFR$.

On utilise les modèles exponentié pour rendre le modèle plus flexible pour des fonctions croissantes ou décroissantes pour $\lambda(t)$.

2.1.3 Estimation paramétrique du taux de hasard (HR)

Considérons le modèle de weibull exponentié définie par la densité f et la fonction de répartition F de la v.a T [28],

$$f(t) = \left(\frac{\alpha}{\sigma}\right) \left(\frac{t}{\sigma}\right)^{\alpha-1} e^{-\left(\frac{t}{\sigma}\right)^\alpha}, \quad F(t) = 1 - e^{-\left(\frac{t}{\sigma}\right)^\alpha}, \quad t > 0.$$

La densité et la fonction de répartition de la v.a T^* sont données par :

$$f^*(t) = \left(\frac{\alpha\theta}{\sigma}\right) \left(\frac{t}{\sigma}\right)^{\alpha-1} [1 - e^{-\left(\frac{t}{\sigma}\right)^\alpha}]^{\theta-1} e^{-\left(\frac{t}{\sigma}\right)^\alpha}, \quad F^*(t) = [1 - e^{-\left(\frac{t}{\sigma}\right)^\alpha}]^\theta, \quad t > 0.$$

1^{er} cas : Données complètes

On considère un échantillon aléatoire $T = (T_1, T_2, \dots, T_N)$ de taille N provenant d'une distribution de Weibull exponentiée.

La fonction de vraisemblance du modèle weibull exponentié est

$$\begin{aligned}
 L^*(\alpha, \theta, \sigma) &= \prod_{i=1}^N f^*(t_i) \\
 &= \prod_{i=1}^N \left[\left(\frac{\alpha\theta}{\sigma} \right) \left(\frac{t}{\sigma} \right)^{\alpha-1} [1 - e^{-(\frac{t}{\sigma})^\alpha}]^{\theta-1} e^{-(\frac{t}{\sigma})^\alpha} \right] \\
 &= \left(\frac{\alpha\theta}{\sigma} \right)^N \prod_{i=1}^N [1 - e^{-(\frac{t_i}{\sigma})^\alpha}] e^{-\sum_{i=1}^N (\frac{t_i}{\sigma})^\alpha} \prod_{i=1}^N \left(\frac{t_i}{\sigma} \right)^{\alpha-1}
 \end{aligned} \tag{2.6}$$

• MLE des paramètres et leurs erreurs standard

$$(\hat{\alpha}, \hat{\theta}, \hat{\sigma}) = \underset{(\alpha, \theta, \sigma)}{\operatorname{argmin}} l^*(\alpha, \theta, \sigma)$$

où

$$l^* = \ln(L^*) = N \ln \left(\frac{\alpha\theta}{\sigma} \right) + (\theta - 1) \sum_{i=1}^N \ln(g(t_i)) - \sum_{i=1}^N \left(\frac{t_i}{\sigma} \right)^\alpha + (\alpha - 1) \sum_{i=1}^N \ln \left(\frac{t_i}{\sigma} \right) \tag{2.7}$$

les équations de vraisemblance sont données par le système

$$(S1) = \begin{cases} \frac{\partial l^*}{\partial \alpha} = \frac{N}{\alpha} + (\theta - 1) \sum_{i=1}^N \frac{g_\alpha(t_i)}{g(t_i)} - \sum_{i=1}^N \left(\frac{t_i}{\sigma} \right)^\alpha \ln \left(\frac{t_i}{\sigma} \right) + \sum_{i=1}^N \ln \left(\frac{t_i}{\sigma} \right) = 0, \\ \frac{\partial l^*}{\partial \theta} = \frac{N}{\theta} + \sum_{i=1}^N \ln(g(t_i)) = 0, \\ \frac{\partial l^*}{\partial \sigma} = \frac{-N\alpha}{\sigma} + (\theta - 1) \sum_{i=1}^N \frac{g_\sigma(t_i)}{g(t_i)} + \left(\frac{\alpha}{\sigma} \right) \sum_{i=1}^N \left(\frac{t_i}{\sigma} \right)^\alpha = 0, \end{cases} \tag{2.8}$$

où $g_\alpha(t_i) = \left(\frac{t_i}{\sigma} \right)^\alpha \ln \left(\frac{t_i}{\sigma} \right) e^{-(\frac{t_i}{\sigma})^\alpha}$ et $g_\sigma(t_i) = \left(\frac{t_i}{\sigma} \right)^\alpha \left(\frac{-\alpha}{\sigma} \right) e^{-(\frac{t_i}{\sigma})^\alpha}$,
tel que les dérivées secondes sont toutes négatives.

Les estimateurs des maximum de vraisemblance des paramètres sont $(\hat{\alpha}, \hat{\theta}, \hat{\sigma})$, construire

en résolvant le système (S1).

• **Estimation de la fonction de survie et son IC**

Pour le même modèle, la fonction de survie

$$S^*(t) = 1 - [1 - e^{-(\frac{t}{\hat{\sigma}})^\alpha}]^\theta, \quad t > 0,$$

estimée par

$$\hat{S}^*(t) = S^*(t)|_{\hat{\alpha}, \hat{\theta}, \hat{\sigma}} = 1 - [1 - e^{-(\frac{t}{\hat{\sigma}})^{\hat{\alpha}}}]^{\hat{\theta}}, \quad t > 0.$$

L'intervalle de confiance asymptotique à 95% pour $S(t)$ est,

$$\hat{S}^*(t) \pm 1.96 \sqrt{Var(\hat{S}^*(t))}, \quad t > 0, \quad (2.9)$$

où

$$Var[\hat{S}^*(t)] = \sum_{i=1}^3 (\hat{S}_i^*)^2 Var(A_i) + 2 \sum_{i>j} \hat{S}_i^* \hat{S}_j^* Cov(A_i, A_j), \quad (2.10)$$

obtenue à partir de l'approximation par la série de Taylor.

Avec $A_1 = \hat{\alpha}$, $A_2 = \hat{\theta}$, $A_3 = \hat{\sigma}$, et,

$$\begin{cases} \hat{S}_1^* = -\hat{\theta} \left(1 - e^{-(\frac{t}{\hat{\sigma}})^{\hat{\alpha}}}\right)^{\hat{\theta}-1} e^{-(\frac{t}{\hat{\sigma}})^{\hat{\alpha}}} \left(\frac{t}{\hat{\sigma}}\right)^{\hat{\alpha}} \ln\left(\frac{t}{\hat{\sigma}}\right), \\ \hat{S}_2^* = -\left(1 - e^{-(\frac{t}{\hat{\sigma}})^{\hat{\alpha}}}\right)^{\hat{\theta}} \ln\left(1 - e^{-(\frac{t}{\hat{\sigma}})^{\hat{\alpha}}}\right), \\ \hat{S}_3^* = \left(\frac{\hat{\alpha}\hat{\theta}}{\hat{\sigma}}\right) \left(1 - e^{-(\frac{t}{\hat{\sigma}})^{\hat{\alpha}}}\right)^{\hat{\theta}-1} e^{-(\frac{t}{\hat{\sigma}})^{\hat{\alpha}}} \left(\frac{t}{\hat{\sigma}}\right)^{\hat{\alpha}}. \end{cases} \quad (2.11)$$

• Estimation du taux de hasard et son IC

Le taux de hasard est donné par,

$$\begin{aligned}\lambda^*(t) &= \frac{f^*(t)}{S^*(t)} = \frac{\left(\frac{\alpha\theta}{\sigma}\right) \left(1 - e^{-\left(\frac{t}{\sigma}\right)^\alpha}\right)^{\theta-1} e^{-\left(\frac{t}{\sigma}\right)^\alpha} \left(\frac{t}{\sigma}\right)^{\alpha-1}}{1 - \left(1 - e^{-\left(\frac{t}{\sigma}\right)^\alpha}\right)^\theta} \\ &= \left(\frac{\alpha\theta}{\sigma}\right) g^{\theta-1}(t) \frac{g_\sigma(t)}{1 - g^\theta(t)}, \quad t > 0.\end{aligned}$$

Estimé par

$$\begin{aligned}\hat{\lambda}^*(t) &= \lambda^*|_{\hat{\alpha}, \hat{\theta}, \hat{\sigma}} = \frac{\left(\frac{\hat{\alpha}\hat{\theta}}{\hat{\sigma}}\right) \left(1 - e^{-\left(\frac{t}{\hat{\sigma}}\right)^{\hat{\alpha}}}\right)^{\hat{\theta}-1} e^{-\left(\frac{t}{\hat{\sigma}}\right)^{\hat{\alpha}}} \left(\frac{t}{\hat{\sigma}}\right)^{\hat{\alpha}-1}}{1 - \left(1 - e^{-\left(\frac{t}{\hat{\sigma}}\right)^{\hat{\alpha}}}\right)^{\hat{\theta}}} \\ &= \left(\frac{\hat{\alpha}\hat{\theta}}{\hat{\sigma}}\right) \hat{g}^{\theta-1}(t) \frac{\hat{g}_\sigma(t)}{1 - \hat{g}^\theta(t)}, \quad t > 0.\end{aligned}$$

L'intervalle de confiance asymptotique à 95% pour $\lambda(t)$ est

$$\hat{\lambda}^*(t) \pm 1.96 \sqrt{Var[\hat{\lambda}^*(t)]}, \quad t > 0, \quad (2.12)$$

où

$$Var(\hat{\lambda}^*(t)) = \sum_{i=1}^3 (\hat{\lambda}_i^*)^2 Var(A_i) + 2 \sum_{i>j} \hat{\lambda}_i^* \hat{\lambda}_j^* Cov(A_i, A_j). \quad (2.13)$$

Avec, $A_1 = \hat{\alpha}$, $A_2 = \hat{\theta}$, $A_3 = \hat{\sigma}$, et

$$\left\{ \begin{aligned} \hat{\lambda}_1^* &= \left(\hat{g}^{\theta-2}(t) \hat{g}_\sigma(t) \hat{g}_\alpha(t) (\hat{\theta} - 1 + \hat{g}^\theta(t)) + \left(\frac{\hat{g}^{\theta-1}(t)}{\hat{\sigma} \hat{\alpha}} \right) (1 - \hat{g}^\theta) \left(\hat{g}_\alpha(t) (\hat{\sigma} - \hat{\alpha}^2 - \hat{\alpha}^2 \left(\frac{t}{\hat{\sigma}}\right)^{\hat{\alpha}}) \right) \right), \\ \hat{\lambda}_2^* &= \left(\frac{\hat{\sigma} \hat{g}_\sigma(t)}{t} \right) \hat{g}^{\theta-1} \left(\frac{1 - \hat{g}^\theta + \hat{\theta} \ln(\hat{g}(t))}{(1 - \hat{g}^\theta(t))^2} \right), \\ \hat{\lambda}_3^* &= -\frac{\hat{\theta}}{t} \left(\frac{\hat{\theta} \hat{g}^{\theta-2}(t) \hat{g}_\sigma^2(t) (\hat{\theta} - 1 + \hat{g}^\theta(t)) + \hat{\alpha} \hat{g}^\theta(t) \hat{g}_\sigma(t) \left(\left(\frac{t}{\hat{\sigma}}\right)^{\hat{\alpha}} - 1 \right) (1 - \hat{g}^\theta(t))}{(1 - \hat{g}^\theta(t))^2} \right). \end{aligned} \right. \quad (2.14)$$

2^{ime} cas : Données censurées

On considère un échantillon aléatoire de taille N tels que T_1, T_2, \dots, T_k des observations non censurées et T_k, T_{k+1}, \dots, T_N des observations censurées, provenant d'une distribution weibull exponentiée.

La fonction du maximum de vraisemblance du modèle est

$$L^*(\alpha, \theta, \sigma) = \prod_{i=1}^k f^*(t_i) \prod_{i=k+1}^N (1 - F^*(t_i)) \quad (2.15)$$

$$= \left[\prod_{i=1}^k \left(\frac{\alpha\theta}{\sigma} \right) g^{\theta-1} e^{-(t_i/\sigma)^\alpha} \left(\frac{t_i}{\sigma} \right)^{\alpha-1} \right] \left[\prod_{i=k+1}^N (1 - g^\theta(t_i)) \right] \quad (2.16)$$

avec $g(t) = 1 - e^{-(\frac{t}{\sigma})^\alpha}$.

- MLE des paramètres et leurs erreurs standard

$$(\hat{\alpha}, \hat{\theta}, \hat{\sigma}) = \underset{(\alpha, \theta, \sigma)}{\operatorname{argmin}} l^*(\alpha, \theta, \sigma),$$

où

$$l^* = \ln L^* = k \ln \left(\frac{\alpha\theta}{\sigma} \right) + (\theta-1) \sum_{i=1}^k \ln(g(t_i)) - \sum_{i=1}^k \left(\frac{t_i}{\sigma} \right)^\alpha + (\alpha-1) \sum_{i=1}^k \ln \left(\frac{t_i}{\sigma} \right) + \sum_{i=k+1}^N \ln(1 - g^\theta(t_i)).$$

Les équations de vraisemblance sont données par

$$(S2) = \begin{cases} \frac{\partial l^*}{\partial \alpha} = \frac{k}{\alpha} + (\theta-1) \sum_{i=1}^k \frac{g_\alpha(t_i)}{g(t_i)} - \sum_{i=1}^k \left(\frac{t_i}{\sigma} \right)^\alpha \ln \left(\frac{t_i}{\sigma} \right) + \sum_{i=1}^k \ln \left(\frac{t_i}{\sigma} \right) \\ - \sum_{i=k+1}^N \frac{\theta g^{\theta-1}(t_i) g_\alpha(t_i)}{1 - g^\theta(t_i)} = 0, \\ \frac{\partial l^*}{\partial \theta} = \frac{k}{\theta} + \sum_{i=1}^k \ln(g(t_i)) - \sum_{i=k+1}^N \frac{g^\theta(t_i) \ln(g(t_i))}{1 - g^\theta(t_i)} = 0, \\ \frac{\partial l^*}{\partial \sigma} = \frac{-k\alpha}{\sigma} + (\theta-1) \sum_{i=1}^k \frac{g^\sigma(t_i)}{g(t_i)} + \frac{\alpha}{\sigma} \sum_{i=1}^k \left(\frac{t_i}{\sigma} \right)^\alpha - \sum_{i=k+1}^N \frac{\theta g^{\theta-1}(t_i) g_\sigma(t_i)}{1 - g^\theta(t_i)} = 0. \end{cases} \quad (2.17)$$

tel que les deuxièmes dérivées sont toutes négatives.

En résolvant le système (S2) ci-dessus, nous obtenons les estimateurs $(\hat{\alpha}, \hat{\theta}, \hat{\sigma})$ des para-

mètres (α, θ, σ) .

- **Estimation de la fonction de survie et son IC**

Pour notre modèle, la fonction de survie

$$S^*(t) = 1 - [1 - e^{-(\frac{t}{\sigma})^\alpha}]^\theta, \quad t > 0.$$

Estimée par

$$\hat{S}^*(t) = S^*(t)|_{\hat{\alpha}, \hat{\theta}, \hat{\sigma}} = 1 - [1 - e^{-(\frac{t}{\hat{\sigma}})^{\hat{\alpha}}}]^{\hat{\theta}}, \quad t > 0.$$

L'intervalle de confiance asymptotique à 95% pour $S(t)$ est

$$\hat{S}^*(t) \pm 1.96 \sqrt{Var(\hat{S}^*(t))}, \quad t > 0. \quad (2.18)$$

ou

$$Var[\hat{S}^*(t)] = \sum_{i=1}^3 (\hat{S}_i^*)^2 Var(A_i) + 2 \sum_{i>j} \hat{S}_i^* \hat{S}_j^* Cov(A_i, A_j). \quad (2.19)$$

Obtenue à partir de l'approximation par la série de Taylor.

Avec $A_1 = \hat{\alpha}$, $A_2 = \hat{\theta}$, $A_3 = \hat{\sigma}$, et,

$$\left\{ \begin{array}{l} \hat{S}_1^* = -\hat{\theta} \left(1 - e^{-(\frac{t}{\hat{\sigma}})^{\hat{\alpha}}}\right)^{\hat{\theta}-1} e^{-(\frac{t}{\hat{\sigma}})^{\hat{\alpha}}} \left(\frac{t}{\hat{\sigma}}\right)^{\hat{\alpha}} \ln\left(\frac{t}{\hat{\sigma}}\right), \\ \hat{S}_2^* = -\left(1 - e^{-(\frac{t}{\hat{\sigma}})^{\hat{\alpha}}}\right)^{\hat{\theta}} \ln\left(1 - e^{-(\frac{t}{\hat{\sigma}})^{\hat{\alpha}}}\right), \\ \hat{S}_3^* = \left(\frac{\hat{\alpha}\hat{\theta}}{\hat{\sigma}}\right) \left(1 - e^{-(\frac{t}{\hat{\sigma}})^{\hat{\alpha}}}\right)^{\hat{\theta}-1} e^{-(\frac{t}{\hat{\sigma}})^{\hat{\alpha}}} \left(\frac{t}{\hat{\sigma}}\right)^{\hat{\alpha}}. \end{array} \right.$$

• Estimation du taux de hasard et son IC

Le taux de hasard est donné par,

$$\begin{aligned}\lambda^*(t) &= \frac{f^*(t)}{S^*(t)} = \frac{\left(\frac{\alpha\theta}{\sigma}\right) \left(1 - e^{-\left(\frac{t}{\sigma}\right)^\alpha}\right)^{\theta-1} e^{-\left(\frac{t}{\sigma}\right)^\alpha} \left(\frac{t}{\sigma}\right)^{\alpha-1}}{1 - \left(1 - e^{-\left(\frac{t}{\sigma}\right)^\alpha}\right)^\theta} \\ &= \left(\frac{\alpha\theta}{\sigma}\right) g^{\theta-1}(t) \frac{g_\sigma(t)}{1 - g^\theta(t)}, \quad t > 0.\end{aligned}$$

Estimé par

$$\begin{aligned}\hat{\lambda}^*(t) &= \lambda^*|_{\hat{\alpha}, \hat{\theta}, \hat{\sigma}} = \frac{\left(\frac{\hat{\alpha}\hat{\theta}}{\hat{\sigma}}\right) \left(1 - e^{-\left(\frac{t}{\hat{\sigma}}\right)^{\hat{\alpha}}}\right)^{\hat{\theta}-1} e^{-\left(\frac{t}{\hat{\sigma}}\right)^{\hat{\alpha}}} \left(\frac{t}{\hat{\sigma}}\right)^{\hat{\alpha}-1}}{1 - \left(1 - e^{-\left(\frac{t}{\hat{\sigma}}\right)^{\hat{\alpha}}}\right)^{\hat{\theta}}} \\ &= \left(\frac{\hat{\alpha}\hat{\theta}}{\hat{\sigma}}\right) g^{\hat{\theta}-1}(t) \frac{g_{\hat{\sigma}}(t)}{1 - g^{\hat{\theta}}(t)}, \quad t > 0.\end{aligned}$$

L'intervalle de confiance asymptotique à 95% pour $\lambda(t)$ est

$$\hat{\lambda}^*(t) \pm 1.96 \sqrt{Var[\hat{\lambda}^*(t)]}, \quad t > 0, \quad (2.20)$$

où

$$Var(\hat{\lambda}^*(t)) = \sum_{i=1}^3 (\hat{\lambda}_i^*)^2 Var(A_i) + 2 \sum_{i>j} \hat{\lambda}_i^* \hat{\lambda}_j^* Cov(A_i, A_j). \quad (2.21)$$

Avec, $A_1 = \hat{\alpha}$, $A_2 = \hat{\theta}$, $A_3 = \hat{\sigma}$, et

$$\left\{ \begin{aligned} \hat{\lambda}_1^* &= \left(\hat{g}^{\theta-2}(t) \hat{g}_\sigma(t) \hat{g}_\alpha(t) (\hat{\theta} - 1 + \hat{g}^\theta(t)) + \left(\frac{\hat{g}^{\theta-1}(t)}{\hat{\sigma} \hat{\alpha}} \right) (1 - \hat{g}^\theta) \left(\hat{g}_\alpha(t) (\hat{\sigma} - \hat{\alpha}^2 - \hat{\alpha}^2 \left(\frac{t}{\hat{\sigma}}\right)^{\hat{\alpha}}) \right) \right), \\ \hat{\lambda}_2^* &= \left(\frac{\hat{\sigma} \hat{g}_\sigma(t)}{t} \right) \hat{g}^{\theta-1} \left(\frac{1 - \hat{g}^\theta + \hat{\theta} \ln(\hat{g}(t))}{(1 - \hat{g}^\theta(t))^2} \right), \\ \hat{\lambda}_3^* &= -\frac{\hat{\theta}}{t} \left(\frac{\hat{\theta} \hat{g}^{\theta-2}(t) \hat{g}_\sigma^2(t) (\hat{\theta} - 1 + \hat{g}^\theta(t)) + \hat{\alpha} \hat{g}^\theta(t) \hat{g}_\sigma(t) \left(\left(\frac{t}{\hat{\sigma}}\right)^{\hat{\alpha}} - 1 \right) (1 - \hat{g}^\theta(t))}{(1 - \hat{g}^\theta(t))^2} \right). \end{aligned} \right.$$

2.1.4 Estimation non paramétrique du HR par la méthode de Kaplan-Meier

On considère T une v.a qui suit une loi de weibull de paramètres β et η [23].

On estime

$$f(t) = \frac{\beta}{\eta} \left(\frac{t}{\eta} \right)^{\beta-1} e^{-\left(\frac{t}{\eta}\right)^\beta},$$

par

$$\hat{f}(t) = \frac{\hat{\beta}}{\hat{\eta}} \left(\frac{t}{\hat{\eta}} \right)^{\hat{\beta}-1} e^{-\left(\frac{t}{\hat{\eta}}\right)^{\hat{\beta}}},$$

avec $\hat{\beta}$ et $\hat{\eta}$ sont les estimateurs de β et η calculées par la méthode des moindres carrées [13].

- **Estimation du taux de hasard**

On estime

$$\lambda(t) = \frac{f(t)}{S(t)},$$

par

$$\hat{\lambda}(t) = \frac{\hat{f}(t)}{S^{KM}(t)} = \frac{\frac{\hat{\beta}}{\hat{\eta}} \left(\frac{t}{\hat{\eta}} \right)^{\hat{\beta}-1} e^{-\left(\frac{t}{\hat{\eta}}\right)^{\hat{\beta}}}}{\prod_{T_i \leq t} \left(1 - \frac{d_i}{n_i} \right)}.$$

Avec

- d_i : le nombre d'évènements à T_i ,
- $T_i, (i = 1, \dots, n)$: les temps d'évènements (décès et censure) distincts ordonnées,
- n_i : le nombre d'individus à risque de subir l'évènement juste avant le temps T_i .

2.2 Taux de hasard inversé (RHR)

Une fonction étroitement liée à la fonction de hasard est **la fonction de taux de hasard inversé (Reversed Hasard Rate (RHR))**, notée $r(t)$, qui au mieux de notre connaissance, a été introduite pour la première fois par Barlow et al.

La fonction de hasard inversée est une fonction moins intuitive, elle pourrait être interprétée comme la probabilité conditionnelle que le changement d'état se produise dans un intervalle infinitésimale précédent t , étant donné que le changement d'état a lieu en t ou avant t . De tous les résultats ou le changement d'état s'est produit, RHR serait la proportion de ces résultats

qui se sont produits immédiatement. Il n'est donc pas surprenant que la fonction de hasard inversée n'ait attiré l'attention des chercheurs que relativement récemment (Finkelstein, 2002) [14]. Bien qu'initialement introduit par la recherche actuarielle, le taux de hasard inversée a jusqu'à présent été principalement appliqué à l'ingénierie de la fiabilité (Desai et al, 2011) [12]. Bien qu'il s'agisse d'une fonction double du taux de hasard dans une certaine mesure, son comportement typique le rend adapté à l'évaluation du temps d'attente, des défaillances cachées, des temps d'inactivité et à l'étude des systèmes, notamment l'optimisation de la fiabilité et la probabilité de bon fonctionnement (Block et al., 1998 ; Chandra et Roy., 2001 ; Xie et al., 2002 ; Badia et Barrade, 2008 ; Poursaeed, 2010) [3][6][35][2][26].

L'ingénierie n'est cependant pas le seul domaine où cet outil s'est avéré utile. Le hasard inversé a également été utilisé pour analyser des données tronquées à droite et censurées à gauche. Kalbfleisch et Lawless (1991), Gross et Huber-Carol (1992) l'ont fait dans le domaine de la médecine [20][17].

En psychologie, Toznsend et Wenger (2004)[34] l'ont trouvé utile dans la modélisation de la capacité de traitement de l'information. De même, la comparaison stochastique des statistiques d'ordre est un autre sujet où cette fonction a trouvé une niche (Shaked et Shanthikumar, 2006)[30]. Dans cette lignée, Cheng et Zhu(1993)[8] caractérisent la meilleur stratégie d'allocation des serveurs dans un système en tandem et Finkelstein (2002) discute des applications possibles à l'ordonnancement des variables aléatoires via le taux de hasard inversé proportionnel. Enfin, Kajima (1998) [19] a utilisé le taux de hasard inversé dans l'étude des chaînes de Markov à temps continu, Gupta et al.(2004)[18], Razmkhah et al.(2012)[27] ont montré comment les mesures d'information de Fisher et d'entropie de Shannon pouvaient être calculées à l'aide de la fonction de hasard inversé.

Toutes ces applications, cependant, sont dans une certaine mesure résiduelle, le hasard inversé reste en dehors du courant dominant de la littérature statistique.

La fonction de taux de hasard inversé est définie comme [6][14] :

$$r(t) = \lim_{dt \rightarrow 0} \frac{P(t - dt \leq T | T \leq t)}{dt} = \frac{f(t)}{F(t)}, \quad t > 0. \quad (2.22)$$

Le taux de hasard inversé intégrée est alors :

$$\tilde{r}(t) = \int_t^{\infty} \frac{dF(x)}{F(x)}, \quad t > 0. \quad (2.23)$$

Étant donné les définitions de $\lambda(t)$ et de $r(t)$, il s'ensuit que :

$$r(t) = \lambda(t)w(t), \quad \text{avec,} \quad w(t) = \frac{F(t)}{S(t)}, \quad t > 0. \quad (2.24)$$

De (2.22) et (2.24), on obtient :

$$r(t) = \frac{\lambda(t)F(t)}{S(t)}, \quad t > 0. \quad (2.25)$$

2.2.1 Propriétés du RHR

Cas continu : [13]

Soit T une v.a continue $T \in D = (a, b) \subseteq \mathbb{R}$, le RHR vérifie les propriétés suivantes

1. $r(t)$ est positive, $\forall t \in (a, b)$.
2. $r(t)$ est continue en $D - \{a\}$.
3. $\lim_{t \rightarrow a} \left(\int_t^b r(u) du \right) = +\infty$ et $\lim_{t \rightarrow \infty} r(t) = 0$

4. À partir de 1-3, on a

$$F(t) = \begin{cases} 0 & \forall t \leq a \\ e^{-\int_t^b r(u) du} & \forall t \in (a, b) \\ 1 & \forall t \geq b \end{cases}$$

D'où

$$f(t) = r(t) \exp(\tilde{r}(t)), \quad \forall t \in (a, b).$$

Cas discret : [13]

Soit T une v.a discrète qui prend les valeurs $-\infty \leq t_0 \leq t_1 \leq \dots \leq t_N \leq +\infty$, où N peut être fini ou infini, le RHR est donné par (Chechile, 2011)[4] :

$$r_i = \begin{cases} 1 & \text{pour } i = 0, \\ \frac{p_i}{F_i} = \frac{F_i - F_{i-1}}{F_i} & \text{pour } i > 0, \end{cases}$$

où $P_i, i = 1, \dots, N$ est la probabilité de masse en t_i tel que $\sum_{i=1}^N P_i = 1$, et $F_i = F(t_i)$ vérifiant les propriétés suivantes :

1. $r_i > 0, \quad \forall i \geq 0, \quad \text{et } \lim_{t \rightarrow \infty} r(t) = 0.$

2. $1 - r_i = \frac{F_{i-1}}{F_i}$ pour tous $i = 1, 2, 3, \dots$

3. $1 - r(t) = 1, \quad \text{pour } t \neq t_0, t_1, t_2, \dots$

4. À partir des propriétés 2 et 3, la fonction de répartition (Voir théorème 25 de Chile(2011)[4]) on a :

- Si $(t_i)_{i \geq 0}$ est non-bornée :

$$F_i = \prod_{j>i} (1 - r_j), \quad \forall i.$$

- Si $(t_i)_{i \geq 0}$ est bornée :

$$F_i = \begin{cases} \prod_{j>i}^N (1 - r_j) & \text{si } i = 0, 1, \dots, N - 1 \\ 1 & \text{si } i = N. \end{cases}$$

La relation entre le taux de hasard inversé et l'élasticité [13]

Les difficultés d'interprétation du RHR peuvent conduire à le considérer à tort comme de simples outils théoriques avec des implications pratiques minimales. Par conséquent, la RHR peut être réévaluée sur la base de certains concepts bien établis qui permettront d'améliorer notre compréhension. Parmi eux se trouve le concept d'élasticité.

Élasticité : est l'un des concepts les plus importants de la théorie économique, qui mesure la sensibilité d'une variable de sortie aux variations d'une variable d'entrée. Elle se définit comme le rapport entre le pourcentage de variation d'une variable et le pourcentage d'une variation d'une autre variable :

1. **Cas continue** Soit T une v.a continue $T \in D = (a, b) \subseteq \mathbb{R}$, la fonction d'élasticité est définie par

$$e(t) = \frac{d(\ln F(t))}{d(\ln t)} = \frac{\frac{F'(t)}{F(t)}}{\frac{1}{|t|}} = \frac{|t|f(t)}{F(t)}. \quad (2.26)$$

Les valeurs possibles de $e(t)$ s'interprètent comme le concept économique d'élasticité lui-même. Par conséquent,

- $e(t) = 0$, implique une situation d'inélasticité parfaite : lorsque la variable a cette valeur, une variation infinitésimale de t n'augmente pratiquement pas la probabilité.
- $0 < e(t) < 1$, décrit une situation d'inélasticité, dans la mesure où une augmentation de la valeur de la variable entraîne une moindre augmentation du cumul de probabilité.
- $e(t) = 1$, exprime l'élasticité unitaire, un accroissement infinitésimal de la variable entraînant le même accroissement du cumul de probabilité.
- $e(t) > 1$, exprime une situation d'élasticité, dans laquelle une augmentation de la variable se traduit par une augmentation plus importante du cumul de probabilité.
- Lorsque $e(t)$ tend vers l'infini, on est dans une situation d'élasticité parfaite, où une augmentation infinitésimale de la valeur de la variable aléatoire conduit à une augmentation très importante (théoriquement infinie) du cumul de probabilité.

2. Le cas discret :

Ce concept pourrait également être étendu au cas discret, dans ce cas, la fonction d'élasticité est définie par :

$$e_i = \frac{|T_i|(F_{i+1} - F_i)}{F_i(T_{i+1} - T_i)}. \quad (2.27)$$

Les relations entre le RHR et la fonction d'élasticité est donné comme suit :

• Cas continue

$$\begin{aligned} e(t) &= |t|.r(t), \quad \forall t \in D - \{a\} \\ r(t) &= \frac{e(t)}{|t|}. \end{aligned} \quad (2.28)$$

• Cas discret

$$\begin{aligned} r_i &= |t_i| \cdot \frac{r_{i+1}}{1 - r_{i+1}} \\ r_{i+1} &= \frac{e_i}{e_i + |t_i|} \end{aligned} \quad (2.29)$$

Exemple 2.1 Nous considérons une situation dans laquelle le problème propose une distribution raisonnable pour modéliser la variable aléatoire « $T =$ "temps de retard pour signaler une

perte, un vol ou un clone de carte de crédit à l'émetteur"», Une variable qui ne prend que des valeurs positives et n'est théoriquement pas liée. Afin de résoudre ce problème, nous suivrons la stratégie consistant à exprimer le comportement attendu des détenteurs de cartes de crédit face à ces contingences en termes de fonction d'élasticité. Compte tenu de l'importance de la perte d'une carte pour son propriétaire, on peut supposer que le propriétaire de la carte s'en occupe généralement et signalera tout ce qui lui est arrivé dès qu'il se rendra compte que quelque chose ne va pas (sans délai). Ainsi, même si seuls quelques clients s'apercevront immédiatement qu'ils ont égaré leur carte, on peut supposer qu'ils seront nombreux par rapport au court laps de temps qui s'est écoulé après le vol ou la perte de leur carte. Cela signifie que l'augmentation de la probabilité qui se produit dans la probabilité de la distribution dans ces premiers cas sera supérieure au temps écoulé. Cela signifie que l'élasticité de T doit être pour de petites valeurs de t supérieure à l'unité, À mesure que la valeur de la variable augmente (c'-à-d à mesure que le temps nécessaire à un client pour détecter l'égaré de sa carte augmente), beaucoup plus de titulaires de cartes se rendront compte que leurs cartes ont été volées ou perdues et, par conséquent, le nombre de signalements augmentera progressivement.

Cependant, l'incorporation de nouveaux clients qui se rendent compte qu'ils ont égaré leur carte diminuera progressivement par unité de temps par rapport au nombre cumulé de lanceurs d'alerte. Autrement dit, le taux entre l'augmentation relative enregistrant le nombre de déclarations et l'augmentation relative du temps écoulé avant la déclaration continuera donc de diminuer, et donc l'élasticité diminuera. De plus, à mesure que t continue d'augmenter, il y aura un point où le nombre de rapports commencera à diminuer. A partir de là, plus la valeur du temps de signalement d'un égaré augmente, plus le nombre cumulé de clients signalant un égaré augmente de moins en moins, la plupart des personnes concernées signalant déjà leur égaré. Ainsi, la relation entre les augmentations relatives continuera de diminuer. L'élasticité de l'augmentation des nouveaux clients, par rapport à l'augmentation du temps pris, sera de plus en plus faible.

Nous pouvons même théoriquement accepter que ce nombre tende vers une inélasticité parfaite à mesure que t grandit et grandit. Ainsi, à mesure que la résilience évoluera en théorie d'une résilience parfaite à une résilience parfaite, il y aura un point d'équilibre auquel le quotient de l'inclusion de nouveaux rapports coïncidera avec l'augmentation du temps qu'il faut pour le faire. A ce moment, lorsque $T = \alpha > 0$, l'élasticité sera unitaire. Par conséquent, la situation décrite ci-dessus doit donc être modélisée à l'aide d'une fonction d'élasticité décroissante qui évolue à partir de valeurs élevées (théoriquement non limitées), atteint zéro asymptotiquement (ce qui indique une inélasticité parfaite) et présente une élasticité unitaire à l'instant α .

Une fonction possible qui correspond à ce comportement est la fonction hyperbolique rectangulaire classique dans \mathbb{R}_+ . [13]

$$e(t) = \frac{\alpha}{t}, \quad t > 0. \quad (2.30)$$

D'après , on aura la fonction de hasard inversé :

$$r(t) = \frac{\alpha}{t^2}, \quad t > 0. \quad (2.31)$$

$r(x)$ étant une fonction de hasard inversé car elle vérifie les propriétés suivantes :

- $r(t) > 0, \quad t > 0.$
- $r(t)$ est continue, $\forall t > 0.$
- $\lim_{\substack{b \rightarrow 0 \\ a \rightarrow 0}} \int_a^b \frac{\alpha}{t^2} dt = \lim_{\substack{b \rightarrow 0 \\ a \rightarrow 0}} \left[\frac{\alpha}{a} - \frac{\alpha}{b} \right] = +\infty \quad \text{et} \quad \lim_{t \rightarrow \infty} \frac{\alpha}{t^2} = 0.$

Par conséquent, en utilisant la propriété 4 du cas continu, le comportement de la v.a T peut être modélisé par sa fonction de répartition donné par

$$F(t) = \begin{cases} 0 & \forall t \leq a \\ e^{-\lim_{b \rightarrow \infty} \int_t^b \frac{\alpha}{u^2} du} = e^{-\frac{\alpha}{t}} & t > 0. \end{cases}$$

Par conséquent, la densité correspondante sera :

$$f(t) = \begin{cases} 0 & \forall t \leq a \\ \frac{\alpha}{t^2} e^{-\frac{\alpha}{t}} & t > 0. \end{cases}$$

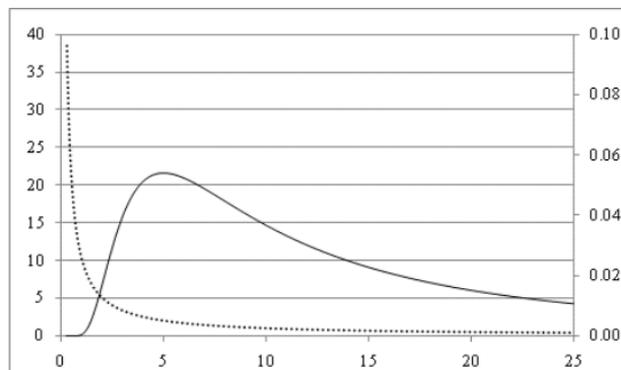


FIGURE 2.2 – Fonction d'élasticité (axe de gauche) et fonction de densité (axe de droite) pour $\alpha = 10$.

Interprétation On remarque que la fonction de densité atteint son maximum à $t = 5 = \frac{\alpha}{2}$ commençant à diminuer en ce point là. La fonction d'élasticité atteinte sa situation unitaire en $t = 10 = \alpha$.

Notons enfin que bien que de nombreuses fonctions remplissent les caractéristiques qualitatives issues de notre argument d'élasticité, permettant d'obtenir d'autres solutions, cette voie permet de proposer des fonctions de distribution difficilement imaginables ou justifiables dans d'autres cas. les fonctions de densité et de cumul obtenues dépendent de la relation particulière supposée pour les fonctions de hasard inversé.

2.2.2 Estimation paramétrique du RHR

Rappelons l'expression du taux de hasard inversé d'une *v.a* continue T ,

$$r(t) = \frac{d}{dt} \ln(F(t)) = \frac{f(t)}{F(t)}, \quad t > 0.$$

La fonction RHR de la *v.a* T^* est donnée par

$$\begin{aligned} r^*(t) &= \frac{f^*(t)}{F^*(t)} = \frac{d}{dt} \ln(F^*(t)) \\ &= \frac{\theta [F(t)]^{\theta-1} f(t)}{[f(t)]^\theta} = \theta \frac{f(t)}{F(t)}, \quad t > 0 \end{aligned}$$

D'où

$$r^*(t) = \theta r(t). \quad (2.32)$$

et $\frac{r^*(t)}{r(t)} = \theta > 0, \forall t$. Donc $r^*(t)$ est proportionnelle à $r(t)$, $\forall t$.

Les propriétés de $r^*(t)$ se déduisent de celles de $r(t)$ par proportionnalité.

1^{er} cas : Données complètes

Pour notre modèle Weibull exponentié, nous avons

$$f(t) = \left(\frac{\alpha}{\sigma}\right) \left(\frac{t}{\sigma}\right)^{\alpha-1} e^{-\left(\frac{t}{\sigma}\right)^\alpha} \quad \text{et} \quad F(t) = 1 - e^{-\left(\frac{t}{\sigma}\right)^\alpha}, \quad t > 0.$$

D'où

$$f^*(t) = \left(\frac{\alpha\theta}{\sigma}\right) \left(\frac{t}{\sigma}\right)^{\alpha-1} [1 - e^{-\left(\frac{t}{\sigma}\right)^\alpha}]^{\theta-1} e^{-\left(\frac{t}{\sigma}\right)^\alpha} \quad \text{et} \quad F^*(t) = [1 - e^{-\left(\frac{t}{\sigma}\right)^\alpha}]^\theta, \quad t > 0,$$

où α, θ, σ sont des paramètres positifs, σ est un véritable paramètre d'échelle.

Étant donné un échantillon aléatoire $T = (T_1, T_2, \dots, T_N)$ de taille N provenant d'une distribution de Weibull exponentié, ses paramètres estimés par la méthode MLE par $\hat{\alpha}, \hat{\theta}$ et $\hat{\sigma}$ solution du système (S1) (2.8).

- **Estimation de la fonction de répartition et son IC**

Pour le modèle weibull exponentialisé, la fonction de répartition est donnée par :

$$F^*(t) = [1 - e^{-(\frac{t}{\sigma})^\alpha}]^\theta, \quad t > 0,$$

estimée par

$$\hat{F}^*(t) = [1 - e^{-(\frac{t}{\hat{\sigma}})^{\hat{\alpha}}}]^{\hat{\theta}}, \quad t > 0,$$

l'intervalle de confiance asymptotique à 95% pour $F(t)$ est donnée à partir de (2.9) par

$$\hat{F}^*(t) \pm 1.96\sqrt{Var(\hat{F}^*(t))}, \quad t > 0, \quad (2.33)$$

où

$$Var(\hat{F}^*(t)) = Var(\hat{S}^*(t)) = \sum_{i=1}^3 (\hat{S}_i^*)^2 Var(A_i) + 2 \sum_{i>j} \hat{S}_i^* \hat{S}_j^* Cov(A_i, A_j). \quad (Voir(2.10))$$

- **Estimation du taux de hasard inversé et son IC**

Pour le modèle weibull exponentié, le RHR est donné par

$$\begin{aligned} r^*(t) &= \frac{f^*(t)}{F^*(t)} = \frac{\left(\frac{\alpha\theta}{\sigma}\right) \left(1 - e^{-(\frac{t}{\sigma})^\alpha}\right)^{\theta-1} e^{-(\frac{t}{\sigma})^\alpha} \left(\frac{t}{\sigma}\right)^{\alpha-1}}{\left(1 - e^{-(\frac{t}{\sigma})^\alpha}\right)^\theta}, \quad t > 0, \\ &= \left(\frac{\alpha\theta}{\sigma}\right) \frac{g_\sigma(t)}{g(t)}, \quad t > 0. \end{aligned}$$

On estime $r^*(t)$ par $\hat{r}^*(t)$

$$\begin{aligned}\hat{r}^*(t) &= \frac{f^*(t)}{F^*(t)} = \frac{\left(\frac{\hat{\alpha}\hat{\theta}}{\hat{\sigma}}\right) e^{-(\frac{t}{\hat{\sigma}})^{\hat{\alpha}}} \left(\frac{t}{\hat{\sigma}}\right)^{\hat{\alpha}-1}}{\left(1 - e^{-(\frac{t}{\hat{\sigma}})^{\hat{\alpha}}}\right)}, \\ &= \left(\frac{\hat{\alpha}\hat{\theta}}{\hat{\sigma}}\right) \frac{\hat{g}_\sigma(t)}{\hat{g}(t)}, \quad t > 0.\end{aligned}$$

L'intervalle de confiance asymptotique à 95% pour $r(t)$ est

$$\hat{r}^*(t) \pm 1.96\sqrt{Var(\hat{r}^*(t))}, \quad t > 0. \quad (2.34)$$

Obtenue à partir de l'approximation par la série de Taylor.

Où

$$Var(\hat{r}^*(t)) = \sum_{i=1}^3 (\hat{r}_i^*)^2 Var(A_i) + 2 \sum_{i>j} \hat{r}_i^* \hat{r}_j^* Cov(A_i, A_j). \quad (2.35)$$

Avec $A_1 = \hat{\alpha}$, $A_2 = \hat{\theta}$, $A_3 = \hat{\sigma}$

$$\left\{ \begin{aligned}\hat{r}_1^* &= \frac{\left(\frac{\hat{\theta}}{\hat{\sigma}}\right) e^{-(\frac{t}{\hat{\sigma}})^{\hat{\alpha}}} \left(\frac{t}{\hat{\sigma}}\right)^{\hat{\alpha}-1} \left[\left(1 - \hat{\alpha} \left(\frac{t}{\hat{\sigma}}\right)^{\hat{\alpha}} \ln\left(\frac{t}{\hat{\sigma}}\right) + (\hat{\alpha} - 1) \left(\frac{t}{\hat{\sigma}}\right)^{-1} \right) 1 - e^{-(\frac{t}{\hat{\sigma}})^{\hat{\alpha}}} - \hat{\alpha} e^{-(\frac{t}{\hat{\sigma}})^{\hat{\alpha}}} \left(\frac{t}{\hat{\sigma}}\right)^{\hat{\alpha}} \ln\left(\frac{t}{\hat{\sigma}}\right) \right]}{\left(1 - e^{-(\frac{t}{\hat{\sigma}})^{\hat{\alpha}}}\right)^2} \\ \hat{r}_2^* &= \frac{\left(\frac{\hat{\alpha}}{\hat{\sigma}}\right) e^{-(\frac{t}{\hat{\sigma}})^{\hat{\alpha}}} \left(\frac{t}{\hat{\sigma}}\right)^{\hat{\alpha}-1} \left[\left(1 - e^{-(\frac{t}{\hat{\sigma}})^{\hat{\alpha}}}\right) - \hat{\theta} \left(\frac{t}{\hat{\sigma}}\right)^{\hat{\alpha}} \ln\left(\frac{t}{\hat{\sigma}}\right) e^{-(\frac{t}{\hat{\sigma}})^{\hat{\alpha}}} \right]}{\left(1 - e^{-(\frac{t}{\hat{\sigma}})^{\hat{\alpha}}}\right)^2} \\ \hat{r}_3^* &= \frac{\left(\frac{\hat{\alpha}\hat{\theta}}{\hat{\sigma}}\right) e^{-(\frac{t}{\hat{\sigma}})^{\hat{\alpha}}} \left(\frac{t}{\hat{\sigma}}\right)^{\hat{\alpha}-1} \left[\left(1 + \left(\frac{\hat{\alpha}}{\hat{\sigma}}\right) \left(\frac{t}{\hat{\sigma}}\right)^{\hat{\alpha}} + \frac{1-\hat{\alpha}}{\hat{\sigma}} \right) \left(1 - e^{-(\frac{t}{\hat{\sigma}})^{\hat{\alpha}}}\right) - \left(\frac{\hat{\theta}}{\hat{\sigma}}\right)^{\hat{\alpha}} \ln\left(\frac{t}{\hat{\sigma}}\right) e^{-(\frac{t}{\hat{\sigma}})^{\hat{\alpha}}} \right]}{\left(1 - e^{-(\frac{t}{\hat{\sigma}})^{\hat{\alpha}}}\right)^2}\end{aligned} \right. \quad (2.36)$$

2^{eme} cas : Données censurées

Considérons un échantillon aléatoire de taille N tel que T_1, T_2, \dots, T_k des observations non censurées et T_k, T_{k+1}, \dots, T_N des observations censurées, provenant d'une distribution weibull exponentielle.

La fonction du maximum de vraisemblance du modèle est

$$L^*(\alpha, \theta, \sigma) = \prod_{i=1}^k f^*(t_i) \prod_{i=k+1}^N (1 - F^*(t_i)) \quad (2.37)$$

$$= \left[\prod_{i=1}^k \left(\frac{\alpha\theta}{\sigma} \right) g^{\theta-1} e^{-(t_i/\sigma)^\alpha} \left(\frac{t_i}{\sigma} \right)^{\alpha-1} \right] \left[\prod_{i=k+1}^N (1 - g^\theta(t_i)) \right] \quad (2.38)$$

avec $g(t) = 1 - e^{-(\frac{t}{\sigma})^\alpha}$, $t > 0$.

Les equations de vraisemblance sont donnés dans le système (S2) (2.17).

Les estimateurs du maximum de vraisemblance des paramètres sont $(\hat{\alpha}, \hat{\theta}, \hat{\sigma})$, sont obtenus à partir du système (S2)

- **Estimation de la fonction de répartition et son IC**

Pour le modèle weibull exponentié, la fonction de répartition est donnée par :

$$F(t) = [1 - e^{-(\frac{t}{\sigma})^\alpha}]^\theta, \quad t > 0,$$

estimée par

$$F^*(t) = [1 - e^{-(\frac{t}{\hat{\sigma}})^{\hat{\alpha}}}]^{\hat{\theta}}, \quad t > 0.$$

L'intervalle de confiance asymptotique à 95% pour F(t) est

$$\hat{F}^*(t) \pm 1.96 \sqrt{Var(\hat{F}^*(t))}, \quad t > 0. \quad (2.39)$$

Où

$$Var(\hat{F}(t)) = Var(\hat{S}(t)) = \sum_{i=1}^3 (\hat{S}_i)^2 Var(A_i) + 2 \sum_{i>j} \hat{S}_i \hat{S}_j Cov(A_i, A_j). \quad (Voir(2.10))$$

- **Estimation du taux de hasard inversé et son IC**

Pour le modèle weibull exponentié, le RHR est donné par

$$r^*(t) = \frac{\left(\frac{\alpha\theta}{\sigma} \right) e^{-(\frac{t}{\sigma})^\alpha} \left(\frac{t}{\sigma} \right)^{\alpha-1}}{\left(1 - e^{-(\frac{t}{\sigma})^\alpha} \right)}, \quad t > 0.$$

On estime $r^*(t)$ par $\hat{r}^*(t)$

$$\hat{r}^*(t) = \frac{\left(\frac{\hat{\alpha}\hat{\theta}}{\hat{\sigma}}\right) e^{-\left(\frac{t}{\hat{\sigma}}\right)^{\hat{\alpha}} \left(\frac{t}{\hat{\sigma}}\right)^{\hat{\alpha}-1}}}{\left(1 - e^{-\left(\frac{t}{\hat{\sigma}}\right)^{\hat{\alpha}}}\right)}, \quad t > 0.$$

L'intervalle de confiance asymptotique à 95% pour $r(t)$ est

$$\hat{r}^*(t) \pm 1.96\sqrt{Var(\hat{r}^*(t))}, \quad t > 0, \quad (2.40)$$

où

$$Var(\hat{r}^*(t)) = \sum_{i=1}^3 (\hat{r}_i^*)^2 Var(A_i) + 2 \sum_{i>j} \hat{r}_i^* \hat{r}_j^* Cov(A_i, A_j). \quad (2.41)$$

Avec $A_1 = \hat{\alpha}$, $A_2 = \hat{\theta}$, $A_3 = \hat{\sigma}$ et $\hat{r}_1, \hat{r}_2, \hat{r}_3$ sont donnés dans le système (2.25).

2.2.3 Comparaison entre F et F^*

Si Θ est une *v.a.* prenant la valeur θ . Alors dans le cas du modèle weibull exponentié on a la probabilité conditionnelle sachant θ est [28] :

$$F^*(t|\Theta = \theta) = [F(t)]^\theta, \quad t > 0.$$

La fonction de distribution conditionnelle est donc la suivante

$$F^*(t|\Theta) = \int_0^\infty F^*(t|\Theta = \theta)g(\theta)d\theta = \int_0^\infty [F(t)]^\theta g(\theta)d\theta, \quad t > 0, \quad \text{avec } \tilde{r}(t) = \int_0^t r(x)dx. \quad (2.42)$$

Où $g(\theta)$ est la densité de probabilité de Θ .

La fonction de distribution conditionnelle peut s'écrire aussi comme suit

$$F^*(t|\Theta) = \int_0^\infty e^{\theta\tilde{r}(t)}g(\theta)d\theta, \quad t > 0. \quad (2.43)$$

En effets, on a $r(t) = \frac{d}{dt} \ln F(t)$

D'où

$$\begin{aligned} \theta \int_0^t r(x)dx = \theta \ln F(x) &\implies [F(x)]^\theta = e^{\theta \int_0^t r(x)dx} \\ &\implies F^*(t) = \int_0^\infty e^{\theta \tilde{r}(t)} g(\theta) d\theta \end{aligned}$$

Ainsi, $F^*(t) = M_\Theta(T(t))$, ou $M_\Theta(s)$ est la fonction génératrice de moment de Θ calculée au point s .

On suppose que le taux de hasard inversée de T sachant $\Theta = \theta$ est $r(t|\theta)$, alors

$$\begin{aligned} r^*(t) &= \frac{f^*(t)}{F^*(t)} = \frac{\int_0^\infty f(t|\theta)g(\theta)d\theta}{F^*(t)} \\ &= \int_0^\infty \frac{f(t|\theta)}{F(t|\theta)} \frac{F(t|\theta)}{F^*(t)} g(\theta) d\theta \\ &= \int_0^\infty r(t|\theta) \frac{F(t|\theta)}{F^*(t)} g(\theta) d\theta \\ &= \int_0^\infty r(t|\theta) \psi(\theta|T < t) d\theta, \quad \text{ou } \psi(\theta|T < t) = \frac{F(t|\theta)}{F^*(t)} g(\theta) \\ &= E_{\Theta|T < t}(r(t|\theta)) \\ r^*(t) &= E_{\Theta|T < t}(\Theta r(t)) \end{aligned}$$

D'où

$$r^*(t) = r(t)E(\Theta|T < t) \tag{2.44}$$

Donc

$$\frac{r^*(t)}{r(t)} = E(\Theta|T < t) \tag{2.45}$$

En prenant la dérivée de $E(\Theta|T < t)$ par rapport à t , on peut montrer que

$$\begin{aligned} \frac{d}{dt}E(\Theta|T < t) &= \frac{d}{dt} \left[\frac{E(\Theta e^{\Theta T(t)})}{E(e^{\Theta T(t)})} \right] \\ &= r(t)Var(\Theta|T < t) > 0. \end{aligned} \tag{2.46}$$

Par conséquent, $\frac{r^*(t)}{r(t)}$ est une fonction croissante de t . De plus, $r(t)$ et $r^*(t)$ ne peuvent se croiser qu'en un seul point et le point de croisement est solution de l'équation [28],

$$\frac{d}{dt}M_\Theta[T(t)] = M_\Theta[T(t)]. \tag{2.47}$$

Remarque 2.1 Si $E(\Theta) \geq 1$, alors $r^*(t) \geq r(t)$, ce qui équivaut à $\frac{F^*(t)}{F(t)}$ est une fonction non décroissante de $t \geq 0$ [28].

Exemple d'application [14]

Tsodikov et al (1997) décrivent un modèle stochastique de cancérogenèse spontanée qui permet d'obtenir un modèle simple de cinétique de croissance des tumeurs. On suppose qu'une tumeur devient détectable lorsque sa taille atteint un certain seuil. On peut supposer que la croissance tumorale obéit aux postulats du processus de naissance et de mort avec deux états d'absorption, de sorte que le premier temps de passage par rapport à la barrière supérieure correspondra au temps de progression de la tumeur. Les hypothèses de base du modèle sont les suivantes [14]

1. L'événement initial du processus de cancérogenèse est la formation de certaines cellules, qui se produit à des moments aléatoires et leur séquence dans le temps est modélisée comme un processus homogène de Poisson.
2. Une fois qu'une première cellule maligne apparaît à la suite de la promotion tumorale, sa croissance ultérieure est irréversible et la phase de progression commence. C'est cette cellule clonogène qui donne naissance à une tumeur détectable après un certain temps, qui est considérée comme une variable aléatoire avec une fonction de distribution cumulative $F(t)$.
3. Une tumeur devient détectable lorsque sa taille atteint une certaine valeur seuil N . Un processus linéaire de naissance pure avec une barrière absorbante N est utilisé pour modéliser la dynamique de la croissance tumorale.

Le nombre critique de cellules tumorales est représenté par $N = cV$, où V est le volume d'une tumeur et c la concentration de cellules tumorales par unité de volume. La distribution conditionnelle du temps de prégression compte tenu du volume seuil V est représentée par un modèle de weibull comme

$$F(t|v) = (1 - e^{-\lambda t})^{cv}.$$

Où λ est le taux de naissance.

2.2.4 Estimation non paramétrique du RHR

Estimation de Kaplan-Meier (données censurées)

On considère T une *v.a* qui suit une loi weibull de paramètres β et η .

On estime

$$f(t) = \frac{\beta}{\eta} \left(\frac{t}{\eta}\right)^{\beta-1} e^{-\left(\frac{t}{\eta}\right)^\beta}, \quad t > 0,$$

par

$$\hat{f}(t) = \frac{\hat{\beta}}{\hat{\eta}} \left(\frac{t}{\hat{\eta}}\right)^{\hat{\beta}-1} e^{-\left(\frac{t}{\hat{\eta}}\right)^{\hat{\beta}}}, \quad t > 0,$$

avec $\hat{\beta}$ et $\hat{\eta}$ sont les estimateurs de β et η calculées par la méthode des moindres carrées [13].

- **Estimation du taux de hasard inversé**

On estime

$$r(t) = \frac{f(t)}{1 - S(t)}, \quad t > 0$$

par

$$\hat{r}(t) = \frac{\hat{f}(t)}{1 - S^{KM}(t)} = \frac{\frac{\hat{\beta}}{\hat{\eta}} \left(\frac{t}{\hat{\eta}}\right)^{\hat{\beta}-1} e^{-\left(\frac{t}{\hat{\eta}}\right)^{\hat{\beta}}}}{1 - \prod_{T_i \leq t} \left(1 - \frac{d_i}{n_i}\right)}, \quad t > 0$$

Avec

- $T_i, (i = 1, \dots, n)$: les temps d'événements (décès et censure) distincts ordonnées,
- d_i : le nombre d'évènements à T_i ,
- n_i : le nombre d'individus à risque de subir l'événement juste avant le temps T_i .

Estimation à noyau du RHR

Estimateur de RHR [3]

Rappelons l'expression du RHR de la durée de vie T :

$$r(t) = \frac{f(t)}{F(t)}, \quad t > 0.$$

où f est une densité de probabilité inconnu de T et F est sa fonction de répartition.
l'estimateur du HRH est donné par :

$$\hat{r}_h(t) = \frac{\hat{f}(t)}{\hat{F}(t)}, \quad t > 0.$$

Soit \hat{f} l'estimateur de f et \hat{F} l'estimateur de F donné comme suit :

$$\hat{f}_h(t) = \frac{1}{n} \sum_{i=1}^n K_{t,h}(T_i), \quad t > 0 \quad (2.48)$$

et

$$\hat{F}_h(t) = \int_0^t \hat{f}(x) dx = \int_0^t \frac{1}{n} \sum_{i=1}^n K_{x,h}(T_i) dx. \quad (2.49)$$

Estimateur RHR à noyau gamma

On a

$$\hat{f}_G(t) = \frac{1}{n} \sum_{i=1}^n \left(\frac{T_i^{\frac{t}{h}} e^{-\frac{T_i}{h}}}{h^{\frac{t}{h}+1} \Gamma(\frac{t}{h} + 1)} \right), \quad t > 0.$$

et

$$\hat{F}_G(t) = \frac{1}{n} \sum_{i=1}^n \int_0^t \left(\frac{T_i^{\frac{x}{h}} e^{-\frac{T_i}{h}}}{h^{\frac{x}{h}+1} \Gamma(\frac{x}{h} + 1)} \right) dx, \quad t > 0.$$

L'estimateur RHR est donné par

$$\begin{aligned} \hat{r}_{G(\frac{t}{h}+1,h)} &= \frac{\hat{f}_{G(\frac{t}{h}+1,h)}}{\hat{F}_{G(\frac{t}{h}+1,h)}} \\ &= \frac{\sum_{i=1}^n \left(\frac{T_i^{\frac{t}{h}} e^{-\frac{T_i}{h}}}{h^{\frac{t}{h}+1} \Gamma(\frac{t}{h} + 1)} \right)}{\sum_{i=1}^n \int_0^t \left(\frac{T_i^{\frac{x}{h}} e^{-\frac{T_i}{h}}}{h^{\frac{x}{h}+1} \Gamma(\frac{x}{h} + 1)} \right) dx}, \quad t > 0 \end{aligned} \quad (2.50)$$

Propriétés asymptotiques

En utilisant le fait que la fonction de répartition estimée converge *p.s* vers F quand $n \rightarrow \infty$ [29], l'espérance, le Biais et la variance de l'estimateur de RHR à noyau gamma sont donnés par

- L'espérance

$$E(\hat{r}_G(t)) = \frac{E[\hat{f}_G(t)]}{F(t)} = \frac{f(x) + hf'(x) + \frac{1}{2}f''(t) + o(h)}{F(t)} \quad (2.51)$$

- Le biais

$$\text{Biais}(\hat{r}_G(t)) = \frac{h(f'(t) + \frac{1}{2}tf''(t))}{F(t)} \quad (2.52)$$

- La variance

$$\text{Var}(\hat{r}_G(t)) = \frac{n^{-1}m^*f(t)}{[F(t)]^2} \quad (2.53)$$

ou

$$m^* = \begin{cases} \frac{1}{\sqrt{2\pi}h^{-\frac{1}{2}}t^{-\frac{1}{2}}}, & \text{si } \frac{t}{h} \rightarrow \infty \\ \frac{\Gamma(2k+2)}{2^{2k+1}(k+1)}h^{-1}, & \text{si } \frac{t}{h} \rightarrow k \end{cases}$$

- MISE

$$\text{MISE}(\hat{r}_G(t)) = \int_0^\infty \left[\frac{h(f'(t) + \frac{1}{2}tf''(t))}{F(t)} \right]^2 dt + \int_0^\infty \frac{n^{-1}m^*f(t)}{[F(t)]^2} dt \quad (2.54)$$

- Le paramètre de lissage optimal

En minimisant le MISE, on obtient

$$h_{opt} = \begin{cases} \left[\frac{\frac{1}{2\sqrt{\pi}} \int_0^\infty \frac{t^{-(\frac{1}{2})}f(t)}{[F(t)]^2} dt}{4 \left[\int_0^\infty \left[\frac{h(f'(t) + \frac{1}{2}tf''(t))}{F(t)} \right]^2 dt \right]} \right]^{\frac{2}{5}} n^{-(\frac{2}{5})} \text{ si } \frac{t}{h} \rightarrow \infty \\ \left[\frac{\frac{1}{2\sqrt{\pi}} \int_0^\infty \frac{t^{-(\frac{1}{2})}f(t)}{[F(t)]^2} dt}{\left[\frac{\int_0^\infty tf''(t)}{F(t)} dt \right]^2} \right]^{\frac{2}{5}} n^{-(\frac{2}{5})} \text{ si } \frac{t}{h} \rightarrow k \end{cases}$$

Estimateur RHR à noyau Inverse-Gaussien (IG)

On a

$$\hat{f}_{IG}(t) = \frac{1}{n} \sum_{i=1}^n \frac{1}{\sqrt{2h\pi T_i^3}} \exp\left(-\frac{1}{2ht} \left(\frac{T_i}{t} - 2 + \frac{t}{T_i}\right)\right), \quad t > 0.$$

et

$$\hat{F}_{IG}(t) = \frac{1}{n} \int_0^t \sum_{i=1}^n \frac{1}{\sqrt{2h\pi T_i^3}} \exp\left(-\frac{1}{2hx} \left(\frac{T_i}{x} - 2 + \frac{x}{T_i}\right)\right) dx, \quad t > 0.$$

L'estimateur RHR à noyau IG est donné par

$$\begin{aligned} \hat{r}_{IG(t, \frac{1}{h})}(t) &= \frac{\hat{f}_{IG(t, \frac{1}{h})}(t)}{\hat{F}_{IG(t, \frac{1}{h})}(t)} \\ &= \frac{\sum_{i=1}^n \frac{1}{\sqrt{2h\pi T_i^3}} \exp\left(-\frac{1}{2ht} \left(\frac{T_i}{t} - 2 + \frac{t}{T_i}\right)\right)}{\int_0^t \sum_{i=1}^n \frac{1}{\sqrt{2h\pi T_i^3}} \exp\left(-\frac{1}{2hx} \left(\frac{T_i}{x} - 2 + \frac{x}{T_i}\right)\right) dx}, \quad t > 0. \end{aligned} \quad (2.55)$$

Propriétés asymptotiques

L'espérance, le Biais et la variance de l'estimateur de RHR à noyau IG sont donnés par

- **L'espérance**

$$E(\hat{r}_{IG}(t)) = \frac{f(t) + \frac{1}{2}ht^3 f''(t) + o(\frac{1}{h})}{F(t)}. \quad (2.56)$$

- **Le biais**

$$Biais(\hat{r}_{IG}(t)) = \frac{\frac{1}{2}t^3 f''(t)}{F(t)}. \quad (2.57)$$

- **La variance**

$$Var(\hat{r}_{IG}(t)) = \frac{\frac{1}{2\sqrt{h\pi}} n^{-1} t^{-\frac{3}{2}} f(t) + o(n^{-1} h^{-\frac{1}{2}})}{[F(t)]^2}. \quad (2.58)$$

- **MISE**

$$MISE(\hat{r}_{IG}(t)) = \int_0^\infty \left[\frac{\frac{1}{2}t^3 f''(t)}{F(x)} \right]^2 dt + \int_0^\infty \left[\frac{\frac{1}{2\sqrt{h\pi}} n^{-1} t^{-\frac{3}{2}} f(t) + o(n^{-1} h^{-\frac{1}{2}})}{[F(t)]^2} \right] dt + o(n^{-1} h^{-1} h^{-\frac{1}{2}}). \quad (2.59)$$

- **Le paramètre de lissage optimal**

en minimisant le MISE, on obtient

$$h_{opt(IG)} = \left[\frac{\int_0^\infty \frac{t^{-3}}{2\sqrt{\pi}} \int_0^\infty \frac{f(t)}{[F(t)]^2} dt}{\int_0^\infty [f(t)]^2} \right]^{\frac{2}{5}} n^{-\frac{2}{5}}. \quad (2.60)$$

Estimateur RHR à noyau Inverse-Gaussien Réciproque (RIG)

On a

$$\hat{f}_{RIG}(t) = \frac{1}{n} \sum_{i=1}^n \frac{1}{\sqrt{2h\pi T_i}} \exp\left(-\frac{t-h}{2h} \left(\frac{T_i}{t-h} - 2 + \frac{t-h}{4}\right)\right), \quad t > 0.$$

et

$$\hat{F}_{RIG}(t) = \frac{1}{n} \int_0^t \sum_{i=1}^n \frac{1}{\sqrt{2h\pi T_i}} \exp\left(-\frac{x-h}{2h} \left(\frac{T_i}{x-h} - 2 + \frac{x-h}{4}\right)\right) dx, \quad t > 0.$$

L'estimateur RHR à noyau RIG est donné par

$$\begin{aligned} \hat{r}_{RIG(\frac{1}{t-h}, \frac{1}{h})} &= \frac{\hat{f}_{RIG}(t)}{\hat{F}_{RIG}(t)} \\ &= \frac{\sum_{i=1}^n \frac{1}{\sqrt{2h\pi T_i}} \exp\left(-\frac{t-h}{2h} \left(\frac{T_i}{t-h} - 2 + \frac{t-h}{4}\right)\right)}{\int_0^t \sum_{i=1}^n \frac{1}{\sqrt{2h\pi T_i}} \exp\left(-\frac{x-h}{2h} \left(\frac{T_i}{x-h} - 2 + \frac{x-h}{4}\right)\right) dx}, \quad t > 0. \end{aligned} \quad (2.61)$$

Propriétés asymptotiques

Le biais, la variance et le MISE de l'estimateur de RHR à noyau RIG sont donnés comme suit

- **Le biais**

$$Biais(\hat{r}_{RIG}(t)) = \frac{\frac{1}{2}htf''(t) + o(h)}{F(t)}. \quad (2.62)$$

- **La variance**

$$Var(\hat{r}_{RIG}(t)) = \frac{\frac{1}{2\sqrt{\pi}}n^{-1}h^{-\frac{1}{2}}t^{-\frac{1}{2}}f(t) + (n^{-1}h^{-1})}{[F(t)]^2}. \quad (2.63)$$

- **MISE**

$$MISE(\hat{r}_{RIG}(t)) = \int_0^\infty \left[\frac{\frac{1}{2}htf''(t) + o(h)}{F(t)} \right]^2 dt + \int_0^\infty \left[\frac{\frac{1}{2\sqrt{\pi}}n^{-1}h^{-\frac{1}{2}}t^{-\frac{1}{2}}f(t) + (n^{-1}h^{-1})}{[F(t)]^2} \right] dt + o\left(\frac{h}{n}h^{-\frac{1}{2}}\right). \quad (2.64)$$

- **Le paramètre de lissage optimal**

On minimisant le MISE, on aura

$$h_{opt(RIG)} = \left[\frac{\frac{1}{2\sqrt{\pi}} \int_0^\infty \frac{t^{-\frac{1}{2}} f''(t)}{[F(t)]^2} dt}{\int_0^\infty \left[\frac{t f''(t)}{F(t)} \right]^2 dt} \right]^{\frac{2}{5}} n^{-\frac{2}{5}}. \quad (2.65)$$

Le choix du paramètre de lissage

- **La méthode de UCV**

La méthode UCV est basée sur l'optimisation de l'erreur quadratique intégrée (ISE), qui est donnée dans notre cas par

$$ISE[\hat{r}(t)] = \int_0^\infty [\hat{r}(t)]^2 dt - 2 \int_0^\infty \hat{r}(t)r(t)dt + \int_0^\infty [r(t)]^2 dt. \quad (2.66)$$

Le dernier terme ne dépend pas de h, nous devons donc minimiser la fonction CV donné par

$$\begin{aligned} CV[h] &= \int_0^\infty [\hat{r}(t)]^2 dt - 2 \int_0^\infty \hat{r}(t)r(t)dt, \\ &= \int_0^\infty \left[\frac{\hat{f}(t)}{\hat{F}(t)} \right]^2 dt - 2E \left[\frac{\hat{f}(t)}{F(t)\hat{F}(t)} \right]. \end{aligned}$$

Et on obtient

$$UCV[h] = \int_0^\infty \left[\frac{\sum_{i=1}^n K_{t,h}(T_i)}{\int_0^t \sum_{i=1}^n K_{x,h}(T_i)dx} \right]^2 - \frac{2}{n} \sum_{i=1}^n \frac{\sum_{j=1, j \neq i}^n K_{T_i,h}(T_j)}{\int_0^t \sum_{j=1, j \neq i}^n K_{X_i,h}^2(T_j)dx}. \quad (2.67)$$

Le h optimale UCV est alors donnée par,

$$h_{UCV} = \underset{h>0}{\operatorname{argmin}} UCV[h]. \quad (2.68)$$

2.3 Conclusion

Dans ce chapitre, nous avons étudié les fonctions taux de hasard et taux de hasard inversé, donné leurs propriétés et la relation entre ces deux fonctions. Nous avons aussi donné des estimations paramétriques et non paramétriques de ces deux fonctions dans le cas des données complètes et censurées.

Une application de ces méthodes d'estimations sera développée dans le chapitre suivant.

Chapitre 3

Application

Introduction

Dans ce chapitre, nous nous sommes intéressés aux applications à des données réelles et aux simulations de la fonction de taux de hasard inversé dans des cas paramétriques et non paramétriques.

Dans le cas paramétrique, nous avons traité les cas des données complètes et censurées en utilisant la méthode MLE sur le modèle de Weibull exponentié réalisé par Remesh, Gupta et Han Wu [28] pour la fonction de survie et HR, cependant nous avons refait les calculs et les représentations graphiques afin de valider notre travail.

Dans le cas non-paramétrique, nous avons travaillé sur deux méthodes, kaplan Meier (données censurées [23]) et la méthode à noyau asymétrique.

3.1 Estimation paramétrique

3.1.1 Analyse d'un ensemble de données non censurées

Dans cette partie, nous avons analysé les données d'Aerset (1987) relatives à la durée de vie de 50 appareils et fourni les MLE et les bandes de confiance pour le taux de hasard, la fonction de survie la fonction du taux de hasard inversé.

0.1	0.2	1	1	1	1	1	2	3	6	7	11	12	18	18	18	18
18	21	32	36	40	45	46	47	50	55	60	63	63	67	67	67	67
72	75	79	82	82	83	84	84	84	85	85	85	85	85	86	86	

TABLE 3.1 – Durée de vie de 50 appareils - Aarset Data.

MLE des paramètres et leurs erreurs standard

Pour notre modèle Weibull exponentiée, nous avons

$$f(t) = \left(\frac{\alpha}{\sigma}\right) \left(\frac{t}{\sigma}\right)^{\alpha-1} e^{-\left(\frac{t}{\sigma}\right)^\alpha} \quad \text{et} \quad F(t) = 1 - e^{-\left(\frac{t}{\sigma}\right)^\alpha}, \quad t > 0.$$

D'où

$$f^*(t) = \left(\frac{\alpha\theta}{\sigma}\right) \left(\frac{t}{\sigma}\right)^{\alpha-1} [1 - e^{-\left(\frac{t}{\sigma}\right)^\alpha}]^{\theta-1} e^{-\left(\frac{t}{\sigma}\right)^\alpha} \quad \text{et} \quad F(t) = [1 - e^{-\left(\frac{t}{\sigma}\right)^\alpha}]^\theta, \quad t > 0.$$

ou α, θ, σ sont des paramètres positifs, σ est un paramètre d'échelle.

On estime les paramètres (α, θ, σ) de cette distribution par la méthode du maximum de vraisemblance et on obtient les MLE $\hat{\alpha}$, $\hat{\theta}$ et $\hat{\sigma}$, en résolvant le système (S1).

Les MLE des paramètres sont obtenus comme suit : $\hat{\alpha} = 4.956$, $\hat{\theta} = 0.139$, $\hat{\sigma} = 91.164$.

La matrice d'information observée est donnée par (Voir Annexe)

$$I_0 = \begin{bmatrix} 2.1367 & 71.0000 & 0.0492 \\ 71.0000 & 2587.8578 & 2.4333 \\ 0.0492 & 2.4333 & 0.0225 \end{bmatrix}$$

La matrice de variance-covariance est donnée par

$$I_0^{-1} = \begin{bmatrix} 5.7636 & -0.01628 & 5.0073 \\ -0.1628 & 0.0050 & -0.1880 \\ 5.0073 & -0.1880 & 53.8257 \end{bmatrix}$$

L'erreur standard des estimations $\hat{\alpha}$, $\hat{\theta}$, $\hat{\sigma}$ est donc 2.4007, 0.0707, 7.3366 respectivement.

Estimation de la fonction de survie et son IC

On estime $S(t) = 1 - [1 - e^{-\left(\frac{t}{\sigma}\right)^\alpha}]^\theta$, $t > 0$ par $\hat{S}(t) = 1 - [1 - e^{-\left(\frac{t}{91.164}\right)^{4.956}}]^{0.139}$, $t > 0$.

L'intervalle de confiance asymptotique à 95% pour $S(t)$ est

$$1 - [1 - e^{-\left(\frac{t}{91.164}\right)^{4.956}}]^{0.139} \pm \sqrt{Var(\hat{S}(t))}. \quad (3.1)$$

ou

$$\begin{aligned} Var(\hat{S}(t)) &= \sum_{i=1}^3 (\hat{S}_i)^2 Var(A_i) + 2 \sum_{i>j} \hat{S}_i \hat{S}_j Cov(A_i, A_j), \\ &= [5.7636(\hat{S}_1^2) + 0.005(\hat{S}_2^2) + 53.8257(\hat{S}_3^2)], \\ &+ 2 [(-0.01628)\hat{S}_2\hat{S}_1 + (5.0073)\hat{S}_3\hat{S}_1 + (-0.188)\hat{S}_3\hat{S}_2]. \end{aligned}$$

Avec

$$\left\{ \begin{array}{l} \hat{S}_1 = -(0.139) \left(1 - e^{-\left(\frac{t}{91.164}\right)^{4.956}}\right)^{-0.861} e^{-\left(\frac{t}{91.164}\right)^{4.956}} \left(\frac{t}{91.164}\right)^{4.956} \ln\left(\frac{t}{91.164}\right), \\ \hat{S}_2 = - \left(1 - e^{-\left(\frac{t}{91.164}\right)^{4.956}}\right)^{0.139} \ln\left(1 - e^{-\left(\frac{t}{91.164}\right)^{4.956}}\right), \\ \hat{S}_3 = (0.0075) \left(1 - e^{-\left(\frac{t}{91.164}\right)^{4.956}}\right)^{-0.861} e^{-\left(\frac{t}{91.164}\right)^{4.956}} \left(\frac{t}{91.164}\right)^{4.956}. \end{array} \right.$$

L'intervalle de Confiance de $S(t)$ des données Aarset de durée de vie de 50 appareils (Tableau (3.1)) est donné dans la figure (3.1)

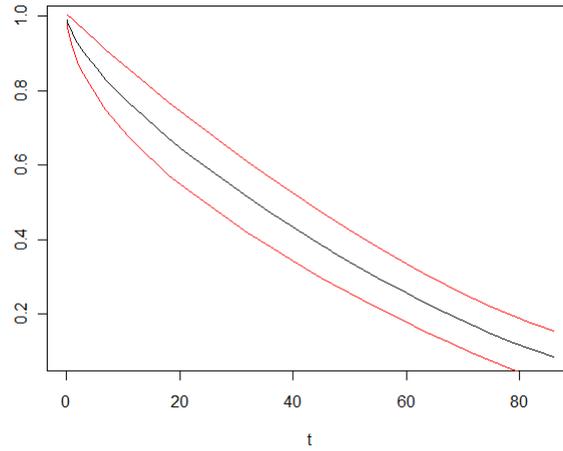


FIGURE 3.1 – L'intervalle de confiance de la fonction de survie (données non censurées).

Interprétation

On voit dans la figure (3.1) que la fonction de survie est décroissante et l'étendu de l'intervalle de confiance est petit, ce qui donne une bonne estimation ensembliste de la survie des appareils.

Estimation du taux de hasard et son IC

On estime le taux de hasard

$$\lambda(t) = \frac{f^*(t)}{S^*(t)} = \frac{\left(\frac{\alpha\theta}{\sigma}\right) \left(1 - e^{-\left(\frac{t}{\sigma}\right)^\alpha}\right)^{\theta-1} e^{-\left(\frac{t}{\sigma}\right)^\alpha} \left(\frac{t}{\sigma}\right)^{\alpha-1}}{1 - \left(1 - e^{-\left(\frac{t}{\sigma}\right)^\alpha}\right)^\theta}, \quad t > 0.$$

par

$$\hat{\lambda}(t) = \frac{(0.0075) \left(1 - e^{-\left(\frac{t}{91.164}\right)^{4.956}}\right)^{-0.861} e^{-\left(\frac{t}{91.164}\right)^{4.956}} \left(\frac{t}{91.164}\right)^{-0.861}}{1 - \left(1 - e^{-\left(\frac{t}{91.164}\right)^{4.956}}\right)^{0.139}}, \quad t > 0.$$

L'intervalle de confiance asymptotique à 95% pour $\lambda(t)$ est

$$\frac{(0.0075) \left(1 - e^{-\left(\frac{t}{91.164}\right)^{4.956}}\right)^{-0.861} e^{-\left(\frac{t}{91.164}\right)^{4.956}} \left(\frac{t}{91.164}\right)^{-0.861}}{1 - \left(1 - e^{-\left(\frac{t}{91.164}\right)^{4.956}}\right)^{0.139}} \pm 1.96 \sqrt{\text{Var}[\hat{\lambda}(t)]}. \quad (3.2)$$

où

$$\begin{aligned} \text{Var}(\hat{\lambda}(t)) &= \sum_{i=1}^3 (\hat{\lambda}_i)^2 \text{Var}(A_i) + 2 \sum_{i>j} \hat{\lambda}_i \hat{\lambda}_j \text{Cov}(A_i, A_j) \\ &= \left[5.7636(\hat{\lambda}_1^2) + 0.005(\hat{\lambda}_2^2) + 53.8257(\hat{\lambda}_3^2) \right] \\ &\quad + 2 \left[(-0.01628)\hat{\lambda}_2\hat{\lambda}_1 + (5.0073)\hat{\lambda}_3\hat{\lambda}_1 + (-0.188)\hat{\lambda}_3\hat{\lambda}_2 \right] \end{aligned}$$

Avec,

$$\left\{ \begin{aligned} \hat{\lambda}_1 &= \hat{g}^{-1.861}(t) \hat{g}_\sigma(t) \hat{g}_\alpha(t) (-0.861 + \hat{g}^\theta(t)) + \left(\frac{\hat{g}^{\theta-1}(t)}{451.8087} \right) (1 - \hat{g}^\theta) \\ &\quad (\hat{g}_\alpha(t) (66.602 - (24.561) \left(\frac{t}{91.164}\right)^{0.139})) \\ \hat{\lambda}_2 &= \left(\frac{(91.164) \hat{g}_\sigma(t)}{t} \right) \hat{g}^{\theta-1} \left(\frac{1 - \hat{g}^\theta + (0.139) \ln(\hat{g}(t))}{(1 - \hat{g}^\theta(t))^2} \right) \\ \hat{\lambda}_3 &= -\frac{0.139}{t} \left(\frac{(0.139) \hat{g}^{\theta-2}(t) \hat{g}_\sigma^2(t) (-0.861 + \hat{g}^\theta(t)) + (4.956) \hat{g}^\theta(t) \hat{g}_\sigma(t) \left(\left(\frac{t}{91.164}\right)^{4.956} - 1 \right) (1 - \hat{g}^\theta(t))}{(1 - \hat{g}^\theta(t))^2} \right) \end{aligned} \right. \quad (3.3)$$

L'intervalle de Confiance de $\lambda(t)$ de durée de vie de 50 appareils est donné dans la figure (3.2)

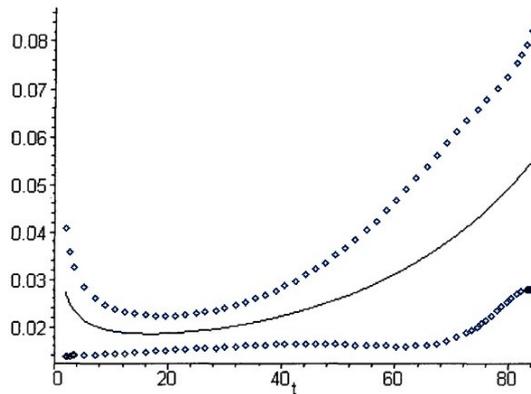


FIGURE 3.2 – L'intervalle de confiance de la fonction de HR (données non censurées)

Interprétation :

On remarque dans la figure (3.2) que l'intervalle de confiance est plus précis entre (0 et 40) et que la fonction de risque est quasi constante, et cela représente la période de bonne performance des appareils. A partir de 40, l'intervalle de confiance s'écarte et la fonction de risque augmente, ce qui représente la période de vieillissement des appareils.

Comme la fonction de hasard est croissante et d'après les propriétés du modèle weibull exponentié, on déduit que $F \in IFR$, de plus $\hat{\theta} = 0.139 < 1$, d'où $F^* \in IFR$.

Estimation du taux de hasard inversé et son IC

On estime

$$r(t) = \frac{\left(\frac{\alpha\theta}{\sigma}\right) \left(1 - e^{-\left(\frac{t}{\sigma}\right)^\alpha}\right)^{\theta-1} e^{-\left(\frac{t}{\sigma}\right)^\alpha} \left(\frac{t}{\sigma}\right)^{\alpha-1}}{\left(1 - e^{-\left(\frac{t}{\sigma}\right)^\alpha}\right)^\theta}, \quad t > 0.$$

par

$$\hat{r}(t) = \frac{(0.0075) e^{-\left(\frac{t}{91.164}\right)^{4.956}} \left(\frac{t}{91.164}\right)^{3.956}}{\left(1 - e^{-\left(\frac{t}{91.164}\right)^{4.956}}\right)}, \quad t > 0.$$

L'intervalle de confiance asymptotique à 95% pour $r(t)$ est

$$\frac{(0.0075) e^{-\left(\frac{t}{91.164}\right)^{4.956}} \left(\frac{t}{91.164}\right)^{3.956}}{\left(1 - e^{-\left(\frac{t}{91.164}\right)^{4.956}}\right)} \pm 1.96\sqrt{Var(\hat{r}(t))} \quad (3.4)$$

où

$$\begin{aligned} Var(\hat{r}(t)) &= \sum_{i=1}^3 (\hat{r}_i)^2 Var(A_i) + 2 \sum_{i>j} \hat{r}_i \hat{r}_j Cov(A_i, A_j) \\ &= [5.7636(\hat{r}_1^2) + 0.005.(\hat{r}_2^2) + 53.8257(\hat{r}_3^2)] \\ &\quad + 2[(-0.01628)\hat{r}_2\hat{r}_1 + (5.0073)\hat{r}_3\hat{r}_1 + (-0.188)\hat{r}_3\hat{r}_2] \end{aligned}$$

avec

$$\left\{ \begin{array}{l} \hat{r}_1 = (0.0015)e^{-\left(\frac{t}{91.164}\right)^{4.956}} \left(\frac{t}{91.164}\right)^{3.956} \\ \left[\frac{(1 - (4.956)\left(\frac{t}{91.164}\right)^{4.956} \ln\left(\frac{t}{91.164}\right) + (3.956)\left(\frac{t}{91.164}\right)^{-1})1 - e^{-\left(\frac{t}{91.164}\right)^{4.956}} - (4.956)e^{-\left(\frac{t}{91.164}\right)^{4.956}} \left(\frac{t}{91.164}\right)^{4.956} \ln\left(\frac{t}{91.164}\right)}{(1 - e^{-\left(\frac{t}{91.164}\right)^{4.956}})^2} \right] \\ \hat{r}_2 = \frac{(0.0015)e^{-\left(\frac{t}{91.164}\right)^{4.956}} \left(\frac{t}{91.164}\right)^{3.956} \left[(1 - e^{-\left(\frac{t}{91.164}\right)^{4.956}}) - (0.139)\left(\frac{t}{91.164}\right)^{4.956} \ln\left(\frac{t}{91.164}\right) e^{\left(\frac{t}{91.164}\right)^{4.956}} \right]}{(1 - e^{-\left(\frac{t}{91.164}\right)^{4.956}})^2} \\ \hat{r}_3 = (0.0075)e^{-\left(\frac{t}{91.164}\right)^{4.956}} \left(\frac{t}{91.164}\right)^{3.956} \\ \left[\frac{(1 + (0.054)\left(\frac{t}{91.164}\right)^{4.956} + (-0.043)(1 - e^{-\left(\frac{t}{91.164}\right)^{4.956}}) - (2.3247)\ln\left(\frac{t}{91.164}\right) e^{-\left(\frac{t}{91.164}\right)^{4.956}})}{(1 - e^{-\left(\frac{t}{91.164}\right)^{4.956}})^2} \right] \end{array} \right. \quad (3.5)$$

L'intervalle de Confiance de $r(t)$ des données Aarset est donné dans la figure (3.3)

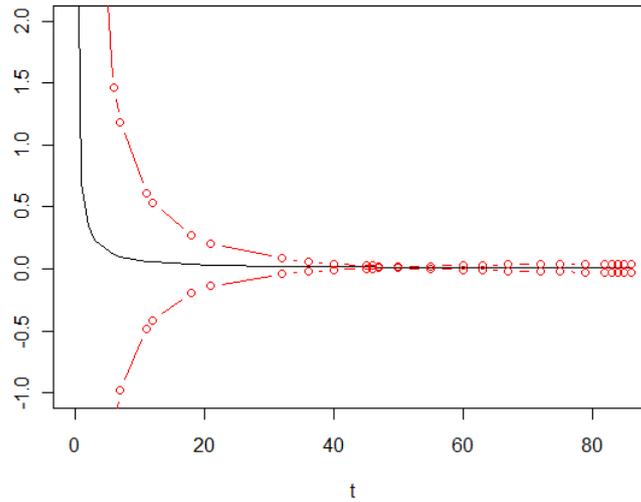


FIGURE 3.3 – L'intervalle de confiance de la fonction RHR (données non censurées)

Interprétation

D'après la figure (3.3), on remarque que l'intervalle de confiance est très éloigné de 0 à 20. Il est de plus en plus précis à partir de 20.

3.1.2 Analyse d'un ensemble de données censurées

Nous analyserons les données du tableau (3.2) suivantes donné par Efron (1988) qui représente la durée de vie des patients de l'essai sur le cancer de la tête et du cou.

7	34	42	63	64	74+	83	84	91	108	112	129
133	133	139	140	140	146	149	154	157	160	160	165
173	176	184+	218	225	241	248	273	277	279+	297	319+
405	417	420	440	523	523+	583	594	1101	1116+	1146	1225+
1349+	1412+	1417									

TABLE 3.2 – Durée de survie des patients de l’essai sur le cancer de la tête et du cou : Efron(1988)

MLE des paramètres et leurs erreurs standard

On estime les paramètres (α, θ, σ) de cette distribution par la méthode du maximum de vraisemblance comme suit

$$L(\alpha, \theta, \sigma) = \left[\prod_{i=1}^k \left(\frac{\alpha\theta}{\sigma} \right) g^{\theta-1} e^{-(t_i/\sigma)^\alpha} \left(\frac{t_i}{\sigma} \right)^{\alpha-1} \right] \left[\prod_{i=k+1}^N (1 - g^\theta(t_i)) \right]$$

$$l = \ln L = k \ln \left(\frac{\alpha\theta}{\sigma} \right) + (\theta - 1) \sum_{i=1}^k \ln(g(t_i)) - \sum_{i=1}^k \left(\frac{t_i}{\sigma} \right)^\alpha + (\alpha - 1) \sum_{i=1}^k \ln \left(\frac{t_i}{\sigma} \right) + \sum_{i=k+1}^N \ln(1 - g^\theta(t_i))$$

On obtient les MLE $\hat{\alpha}$, $\hat{\theta}$ et $\hat{\sigma}$, en résolvant le système (S2)

L’erreur standard (SE) des estimateurs $\hat{\alpha}$, $\hat{\theta}$ et $\hat{\sigma}$ peut être obtenue à partir de la matrice d’information de l’échantillon.

Les MLE des paramètres sont obtenus comme suit : $\hat{\alpha} = 0.3105$, $\hat{\theta} = 15.1850$, $\hat{\sigma} = 0.2117$.

La matrice d’information observée est donnée par

$$I_0 = \begin{bmatrix} 25873.1778 & -1.8462 & -7042.0836 \\ -1.8462 & 0.2213 & 0.5553 \\ -7042.0836 & 0.5553 & 3910.1092 \end{bmatrix}$$

La matrice de variance-covariance est donnée par

$$I_0^{-1} = \begin{bmatrix} 0.000076 & 0.00029 & 0.00014 \\ 0.00029 & 4.52147 & -0.00012 \\ 0.00014 & -0.00012 & 0.00050 \end{bmatrix}$$

L'erreur standard des estimations $\hat{\alpha}$, $\hat{\theta}$, $\hat{\sigma}$ est donc 0.0087, 2.1264, 0.0224 respectivement.

Estimation de la fonction de survie et son IC

On estime $S(t) = 1 - [1 - e^{-\left(\frac{t}{\sigma}\right)^\alpha}]^\theta$, $t > 0$ par $\hat{S}(t) = 1 - [1 - e^{-\left(\frac{t}{0.2117}\right)^{0.3105}}]^{15.185}$, $t > 0$.
l'intervalle de confiance asymptotique à 95% pour S(t) est

$$1 - [1 - e^{-\left(\frac{t}{0.2117}\right)^{0.3105}}]^{15.185} \pm \sqrt{Var(\hat{S}(t))}. \quad (3.6)$$

ou

$$\begin{aligned} Var(\hat{S}(t)) &= \sum_{i=1}^3 (\hat{S}_i)^2 Var(A_i) + 2 \sum_{i>j} \hat{S}_i \hat{S}_j Cov(A_i, A_j) \\ &= [0.000076(\hat{S}_1^2) + 4.52147.(\hat{S}_2^2) + 0.0005(\hat{S}_3^2)] \\ &\quad + 2 [(0.00029)\hat{S}_2\hat{S}_1 + (0.00014)\hat{S}_3\hat{S}_1 + (-0.00012)\hat{S}_3\hat{S}_2] \end{aligned}$$

Avec

$$\left\{ \begin{array}{l} \hat{S}_1 = -(15.185) \left(1 - e^{-\left(\frac{t}{0.2117}\right)^{0.3105}}\right)^{14.185} e^{-\left(\frac{t}{0.2117}\right)^{0.3105}} \left(\frac{t}{0.2117}\right)^{0.3105} \ln\left(\frac{t}{0.2117}\right) \\ \hat{S}_2 = -\left(1 - e^{-\left(\frac{t}{0.2117}\right)^{0.3105}}\right)^{15.185} \ln\left(1 - e^{-\left(\frac{t}{0.2117}\right)^{0.3105}}\right) \\ \hat{S}_3 = (22.2718) \left(1 - e^{-\left(\frac{t}{0.2117}\right)^{0.3105}}\right)^{14.185} e^{-\left(\frac{t}{0.2117}\right)^{0.3105}} \left(\frac{t}{0.2117}\right)^{0.3105} \end{array} \right.$$

L'intervalle de Confiance de S(t) est donné dans la figure (3.4).

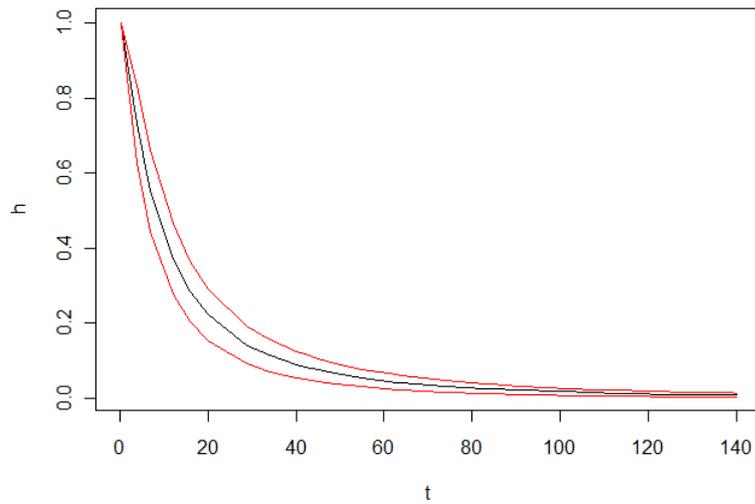


FIGURE 3.4 – L'intervalle de confiance de la fonction de survie (données censurées).

Interprétation

De la figure (3.4), on remarque que La fonction de survie est décroissante, l'entendu de l'intervalle de confiance est petit ce qui donne une bonne approximation.

Estimation du taux de hasard et son IC

On estime le taux de hasard

$$\lambda(t) = \frac{\left(\frac{\alpha\theta}{\sigma}\right) \left(1 - e^{-\left(\frac{t}{\sigma}\right)^\alpha}\right)^{\theta-1} e^{-\left(\frac{t}{\sigma}\right)^\alpha} \left(\frac{t}{\sigma}\right)^{\alpha-1}}{1 - \left(1 - e^{-\left(\frac{t}{\sigma}\right)^\alpha}\right)^\theta}, \quad t > 0,$$

par

$$\hat{\lambda}(t) = \frac{(22.2718) \left(1 - e^{-\left(\frac{t}{0.2117}\right)^{0.3105}}\right)^{14.185} e^{-\left(\frac{t}{0.2117}\right)^{0.3105}} \left(\frac{t}{0.2117}\right)^{22.2718}}{1 - \left(1 - e^{-\left(\frac{t}{0.2117}\right)^{0.3105}}\right)^{15.185}}, \quad t > 0.$$

L'intervalle de confiance asymptotique à 95% pour $\lambda(t)$ est

$$\frac{(22.2718) \left(1 - e^{-\left(\frac{t}{0.2117}\right)^{0.3105}}\right)^{14.185} e^{-\left(\frac{t}{0.2117}\right)^{0.3105}} \left(\frac{t}{0.2117}\right)^{22.2718}}{1 - \left(1 - e^{-\left(\frac{t}{0.2117}\right)^{0.3105}}\right)^{15.185}} \pm 1.96 \sqrt{Var[\hat{\lambda}(t)]}. \quad (3.7)$$

où

$$\begin{aligned} Var(\hat{\lambda}(t)) &= \sum_{i=1}^3 (\hat{\lambda}_i)^2 Var(A_i) + 2 \sum_{i>j} \hat{\lambda}_i \hat{\lambda}_j Cov(A_i, A_j) \\ &= \left[0.000076(\hat{\lambda}_1^2) + 4.52147(\hat{\lambda}_2^2) + 0.0005(\hat{\lambda}_3^2)\right] \\ &\quad + 2 \left[(0.00029)\hat{\lambda}_2 \hat{\lambda}_1 + (0.00014)\hat{\lambda}_3 \hat{\lambda}_1 + (-0.00012)\hat{\lambda}_3 \hat{\lambda}_2\right] \end{aligned}$$

Avec,

$$\left\{ \begin{aligned} \hat{\lambda}_1 &= \left(\hat{g}^{13.185}(t) \hat{g}_\sigma(t) \hat{g}_\alpha(t) (13.185 + \hat{g}^\theta(t)) + \left(\frac{\hat{g}^{\theta-1}(t)}{0.0657}\right) (1 - \hat{g}^\theta) (\hat{g}_\alpha(t) (0.1152 - (0.0964) \left(\frac{t}{0.2117}\right)^{15.185}))\right) \\ \hat{\lambda}_2 &= \left(\frac{(0.2117) \hat{g}_\sigma(t)}{t}\right) \hat{g}^{\theta-1} \left(\frac{1 - \hat{g}^\theta + (15.185) \ln(\hat{g}(t))}{(1 - \hat{g}^\theta(t))^2}\right) \\ \hat{\lambda}_3 &= -\frac{15.185}{t} \left(\frac{(15.185) \hat{g}^{theta-2}(t) \hat{g}_\sigma^2(t) (14.185 + \hat{g}^\theta(t)) + (0.3105) \hat{g}^\theta(t) \hat{g}_\sigma(t) \left(\left(\frac{t}{0.2117}\right)^{0.3105} - 1\right) (1 - \hat{g}^\theta(t))}{(1 - \hat{g}^\theta(t))^2}\right) \end{aligned} \right. \quad (3.8)$$

L'intervalle de confiance de $\lambda(t)$ est donné dans la figure (3.5).

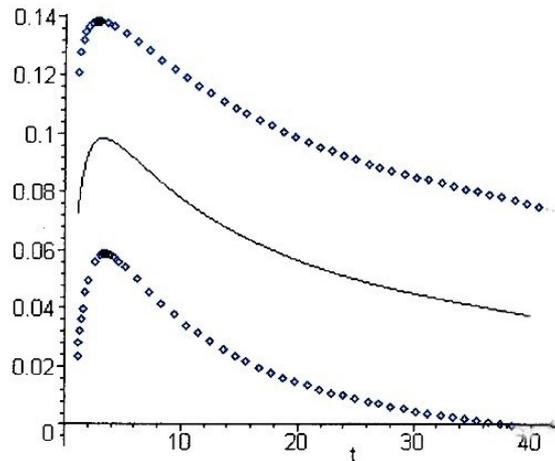


FIGURE 3.5 – L ;intervalle de confiance de RH (données censurées).

Interprétation On constate une légère augmentation de la fonction de hasard sur [0,6]. À partir de $t = 7$, la fonction de risque diminue progressivement jusqu'à atteindre zéro, et c'est ce qui représente la période de jeunesse des patients (les malades se sont rétablis).

Comme la fonction RH est décroissante alors $F \in DFR$. De plus, $\hat{\theta} = 15.185 > 1$, d'où $F^* \in DFR$

Estimation du taux de hasard inversé et son IC

On estime

$$r(t) = \frac{\left(\frac{\alpha\theta}{\sigma}\right) \left(1 - e^{-\left(\frac{t}{\sigma}\right)^\alpha}\right)^{\theta-1} e^{-\left(\frac{t}{\sigma}\right)^\alpha} \left(\frac{t}{\sigma}\right)^{\alpha-1}}{\left(1 - e^{-\left(\frac{t}{\sigma}\right)^\alpha}\right)^\theta}, \quad t > 0.$$

par

$$\hat{r}(t) = \frac{(22.2718) e^{-\left(\frac{t}{0.2117}\right)^{0.3105}} \left(\frac{t}{0.2117}\right)^{-3.6895}}{\left(1 - e^{-\left(\frac{t}{0.2117}\right)^{0.3105}}\right)}, \quad t > 0.$$

L'intervalle de confiance asymptotique à 95% pour $r(t)$ est

$$\frac{(22.2718) e^{-\left(\frac{t}{0.2117}\right)^{0.3105}} \left(\frac{t}{0.2117}\right)^{-3.6895}}{\left(1 - e^{-\left(\frac{t}{0.2117}\right)^{0.3105}}\right)} \pm 1.96 \sqrt{Var(\hat{r}(t))}, \quad (3.9)$$

où

$$\begin{aligned} Var(\hat{r}(t)) &= \sum_{i=1}^3 (\hat{r}_i)^2 Var(A_i) + 2 \sum_{i>j} \hat{r}_i \hat{S}_j Cov(A_i, A_j) \\ &= [0.000076(\hat{r}_1^2) + 4.52147.(\hat{r}_2^2) + 0.0005(\hat{r}_3^2)] \\ &\quad + 2 [(0.00029)\hat{r}_2\hat{r}_1 + (0.00014)\hat{r}_3\hat{r}_1 + (-0.00012)\hat{r}_3\hat{r}_2] \end{aligned}$$

avec

$$\left\{ \begin{array}{l} \hat{r}_1 = (71.72)e^{-\left(\frac{t}{0.2117}\right)^{0.3105}} \left(\frac{t}{0.2117}\right)^{-0.68} \\ \left[\left(1 - (0.3105)\left(\frac{t}{0.2117}\right)^{0.3105} \ln\left(\frac{t}{0.2117}\right) + (-0.68)\left(\frac{t}{0.2117}\right)^{-1}\right) 1 - e^{-\left(\frac{t}{0.2117}\right)^{0.3105}} - \hat{g}_\alpha \right] \left(1 - e^{-\left(\frac{t}{0.2117}\right)^{0.3105}}\right)^{-2}, \\ \hat{r}_2 = \frac{(1.466)e^{-\left(\frac{t}{0.2117}\right)^{0.3105}} \left(\frac{t}{0.2117}\right)^{14.185} \left[\left(1 - e^{-\left(\frac{t}{0.2117}\right)^{0.3105}}\right) - (15.185)\left(\frac{t}{0.2117}\right)^{0.3105} \ln\left(\frac{t}{0.2117}\right) e^{\left(\frac{t}{0.2117}\right)^{0.3105}} \right]}{\left(1 - e^{-\left(\frac{t}{0.2117}\right)^{0.3105}}\right)^2}, \\ \hat{r}_3 = \frac{(22.2718)e^{-\left(\frac{t}{0.2117}\right)^{0.3105}} \left(\frac{t}{0.2117}\right)^{-0.6895}}{\left[\left(1 + (1.466)\left(\frac{t}{0.2117}\right)^{0.3105} + 3.2569\right) \left(1 - e^{-\left(\frac{t}{0.2117}\right)^{0.3105}}\right) - (3.7687) \ln\left(\frac{t}{0.2117}\right) e^{-\left(\frac{t}{0.2117}\right)^{0.3105}} \right]} \\ \left(1 - e^{-\left(\frac{t}{0.2117}\right)^{0.3105}}\right)^2 \end{array} \right. \quad (3.10)$$

L'intervalle de Confiance de $r(t)$ est donné dans la figure (3.6).

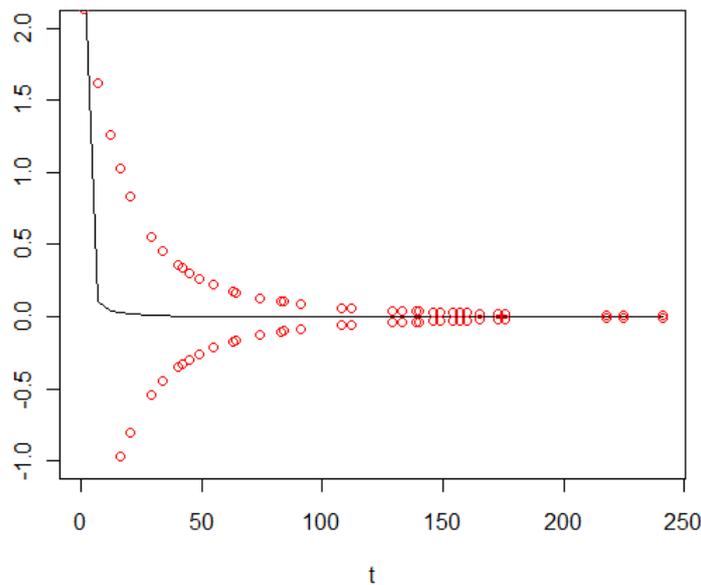


FIGURE 3.6 – L'intervalle de confiance de la fonction RHR (données censurées).

Interprétation

On constate un écart significatif de l'intervalle de confiance sur $[0, 50]$. À partir du $t=100$, l'intervalle de confiance est très précis.

3.2 Estimation non paramétrique

3.2.1 Estimation par la méthode de Kaplein Meier

Dans cette partie, nous analyserons les données du tableau (3.3) qui représentent les données sur les défaillances ou censures du composant «Bague d'étanchéité HP» de l'équipement "Compresseur du Magnaforming" récoltées lors d'une étude faite au sein de la raffinerie de pétrole de Skikda afin d'optimiser la maintenance préventive.

Numéro	Temps	Censure/Défaillance
01	60	C
02	82	D
03	115	C
04	115	C
05	121	D
06	129	D
07	141	D
08	153	D
09	162	D
10	163	D
11	193	C
12	210	D
13	227	D
14	278	D
15	287	D
16	288	C
17	312	D
18	316	D
19	334	D
20	412	D
21	430	C

TABLE 3.3 – les données sur les défaillances ou censures du composant «Bague d'étanchéité HP»

Ces données sont modélisées dans [23] par une distribution weibull,

$$f(t) = \frac{\beta}{\eta} \left(\frac{t}{\eta} \right)^{\beta-1} e^{-\left(\frac{t}{\eta}\right)^\beta}.$$

l'estimation des paramètres β et η de la loi de Weibull correspondant à ces données a été faite par la méthode des moindres carrées [23].

- MMC simple : $\hat{\beta} = 2,33, \hat{\eta} = 268, z = 0,0386.$
- MMC pondéré : $\hat{\beta} = 1,99, \hat{\eta} = 271, z = 0,0335.$

où β est le paramètre de forme, η est le paramètre d'échelle, z est le résidu.

Donc

$$\begin{aligned} \hat{f}(t) &= \frac{\hat{\beta}}{\hat{\eta}} \left(\frac{t}{\hat{\eta}} \right)^{\hat{\beta}-1} e^{-\left(\frac{t}{\hat{\eta}}\right)^{\hat{\beta}}}, \\ &= 0.007 \left(\frac{t}{271} \right)^{0.99} e^{-\left(\frac{t}{271}\right)^{1.99}}. \end{aligned}$$

- **Estimateur de la fonction de survie**

L'estimateur de Kaplan-Meier de la fonction de survie S est calculer par

$$\begin{aligned} S^{KM} &= \prod_{T(i) \leq t} \left(1 - \frac{d_i}{n_i} \right) \\ S^{KM}(t_i) &= S^{KM}(t_{i-1}) \left(1 - \frac{d_i}{n_i} \right) \end{aligned}$$

Les résultats sont donnée dans le tableau (3.4)

Temps	n_i	d_i	$S^{KM}(t_i)$
0	21	0	$(1 - \frac{0}{21}) = 1$
82	20	1	$(1 - \frac{1}{20}) = 0,950$
121	17	1	$0.950(1 - \frac{1}{17}) = 0,894$
129	16	1	$0.894(1 - \frac{1}{16}) = 0,838$
141	15	1	$0.838(1 - \frac{1}{15}) = 0,782$
153	14	1	$0.782(1 - \frac{1}{14}) = 0,726$
162	13	1	$0.726(1 - \frac{1}{13}) = 0,671$
163	12	1	$0.671(1 - \frac{1}{12}) = 0,615$
210	10	1	$0.615(1 - \frac{1}{10}) = 0,553$
227	9	1	$0.553(1 - \frac{1}{9}) = 0,492$
278	8	1	$0.492(1 - \frac{1}{8}) = 0,430$
287	7	1	$0.430(1 - \frac{1}{7}) = 0,369$
312	5	1	$0.369(1 - \frac{1}{5}) = 0,295$
316	4	1	$0.295(1 - \frac{1}{4}) = 0,221$
334	3	1	$0.221(1 - \frac{1}{3}) = 0,147$
412	2	1	$0.147(1 - \frac{1}{2}) = 0,074$

TABLE 3.4 – Les estimateurs de K.M de la fonction de survie.

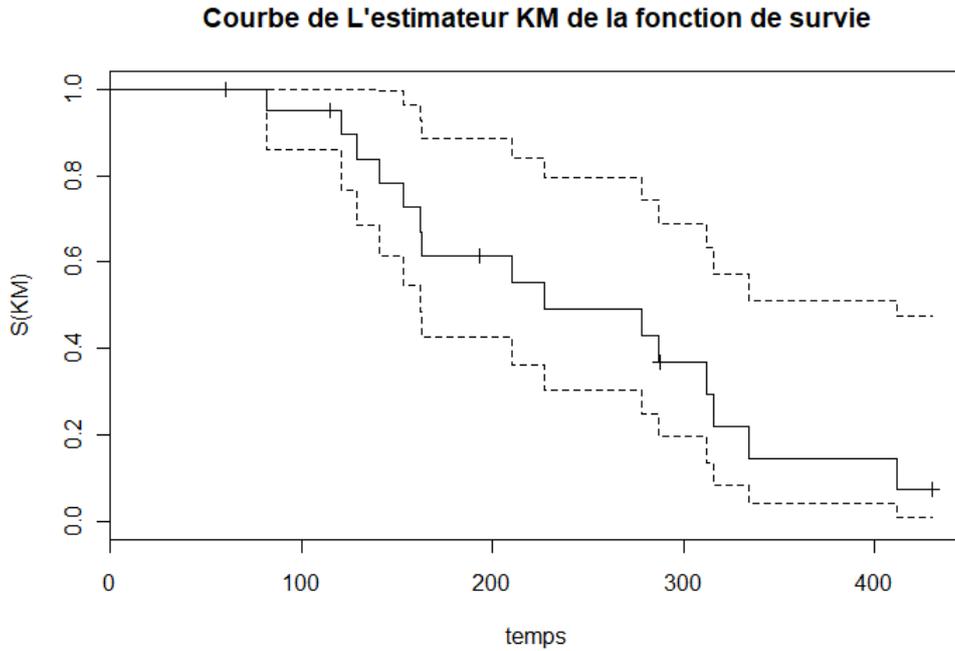


FIGURE 3.7 – l'estimation de la fonction de survie par KM et les intervalles de confiance à un niveau de 95%

Interprétation

À partir de la figure (3.7), on constate que l'intervalle de confiance est bien précis de [0, 150]. À partir de 200 la borne supérieure de l'intervalle de confiance s'éloigne progressivement.

- **Estimateur de la fonction de hasard et de hasard inversé**

On estime le taux de hasard

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

par

$$\hat{\lambda}(t) = \frac{0.007 \left(\frac{t}{271}\right)^{0.99} e^{-\left(\frac{t}{271}\right)^{0.99}}}{\prod_{T(i) \leq t} \left(1 - \frac{d_i}{n_i}\right)}$$

et le taux de hasard inversé

$$r(t) = \frac{\frac{\beta}{\eta} \left(\frac{t}{\eta}\right)^{\beta-1} e^{-\left(\frac{t}{\eta}\right)^\beta}}{1 - S(t)}$$

par

$$\hat{r}(t) = \frac{0.007 \left(\frac{t}{271}\right)^{0.99} e^{-\left(\frac{t}{271}\right)^{0.99}}}{1 - \prod_{T(i) \leq t} \left(1 - \frac{d_i}{n_i}\right)}$$

Le tableau (3.5) fournit les résultats obtenus des HR et RHR estimées par Kaplan Meier

Temps	S^{KM}	$\hat{\lambda}(t).[10^{-3}]$	$\hat{r}(t).[10^{-2}]$
82	0,950	2,156	4,099
121	0,8941	3,024	2,550
129	0,8382	3,345	1,730
141	0,7823	3,745	1,343
153	0,7264	4,162	1,108
162	0,6706	4,591	0,936
163	0,6147	5,018	0,802
210	0,5532	5,650	0,699
227	0,4918	6,201	0,601
278	0,4303	6,116	0,461
287	0,3688	6,866	0,401
312	0,2950	7,617	0,319
316	0,2213	9,953	0,282
334	0,1475	13,402	0,232
412	0,0738	15,037	0,120

TABLE 3.5 – Estimations du taux de hasard et taux de hasard inversé

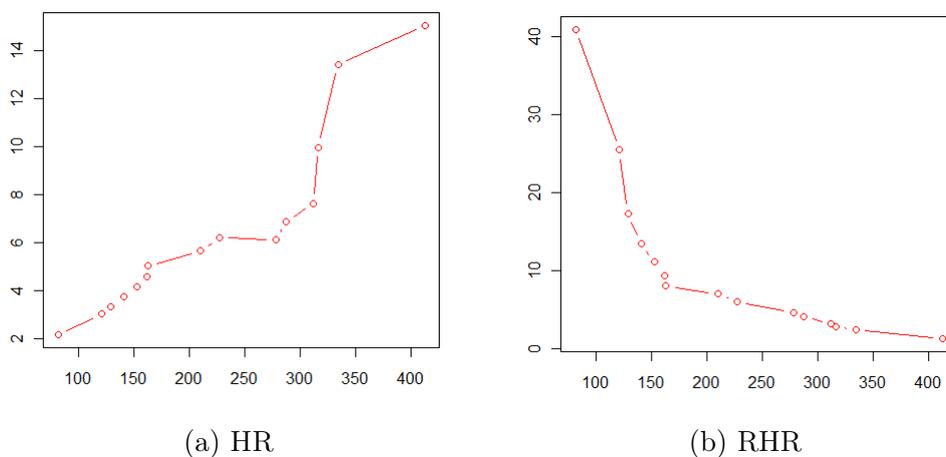


FIGURE 3.8 – Estimateurs de KM de taux de hasard et taux de hasard inversé

Interprétation

À partir de la figure (3.8), nous notons que la fonction de taux de hasard augmente ce qui représente la période de vieillissement du composant. Contrairement la fonction de hasard inversée est décroissante.

3.2.2 Estimation à noyau de taux de hasard inversé

Simulation

Dans cette partie, nous illustrons certains estimateurs à noyaux continus asymétriques à savoir le noyau Gamma et le Gaussien Inverse. Nous simulons un échantillon de taille $n = (50, 100, 500, 1000)$. Pour chaque noyau fixé, la fenêtre optimale est choisie par la méthode de validation croisée non biaisé.

Pour l'application numérique nous avons considéré les distributions suivantes :

1. Une loi gamma $G(3,1)$
2. Une loi log normal $\mathcal{LN}(1,1)$

Nous calculons le critère ISE dans le cas de méthode UCV, et les résultats de ISE moyen sont illustrés dans le tableau (3.6).

	Noyau	Gamma		LN	
		\hat{f}	\hat{r}	\hat{f}	\hat{r}
n = 50	Gamma	0.00341798	0.02096957	0.0220139	0.05991312
	IG	0.02594182	0.1198164	0.08429569	0.1029933
	RIG	0.008934334	0.01734897	0.02257169	0.06478979
n=100	Gamma	0.00172617	0.01226283	0.09514616	0.03279968
	IG	0.004401226	0.02507696	0.02651065	0.0993843
	RIG	0.003502734	0.01529820	0.002194561	0.03838449
n=500	Gamma	0.001182117	0.002781097	0.007465783	0.01339525
	IG	0.002046395	0.009323277	0.004235446	0.03511898
	RIG	0.0008877393	0.009506522	0.0008150676	0.01569476
n=1000	Gamma	0.0006962391	0.00165076	0.006140206	0.008710409
	IG	0.001247073	0.002421368	0.002345129	0.01132195
	RIG	0.0006546306	0.00284560	0.0007361259	0.009729070

TABLE 3.6 – Quelques valeurs du ISE pour les estimateurs de la fonction de densité et la fonction RHR, pour les modèles considérés dans la simulation, en utilisant le $h[UCV]$

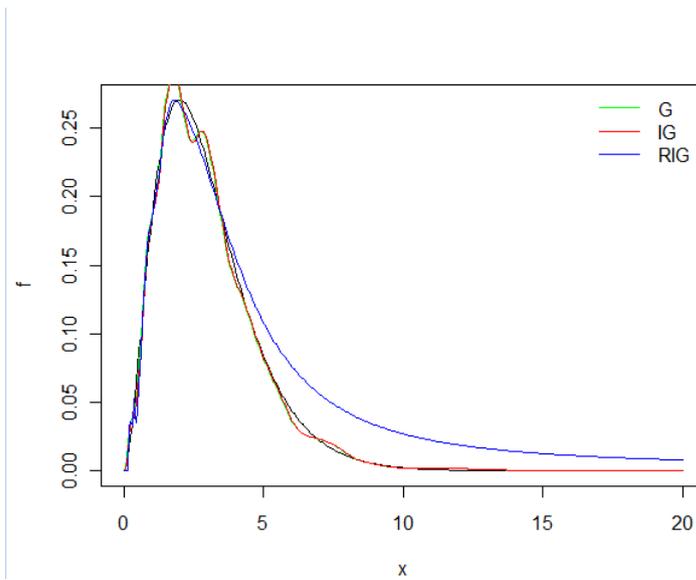
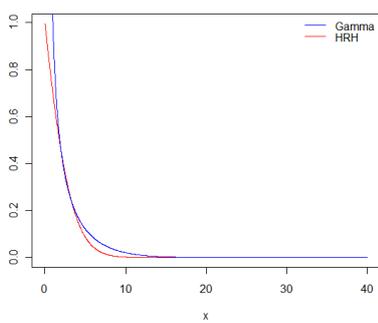
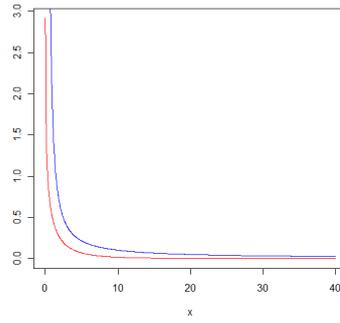


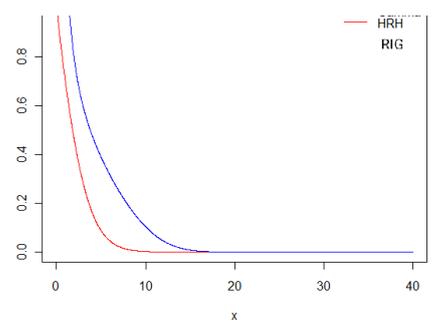
FIGURE 3.9 – Estimateur des noyaux gamma, IG et RIG de la distribution gamma pour $n=500$



(a) estimateur Gamma-RHR

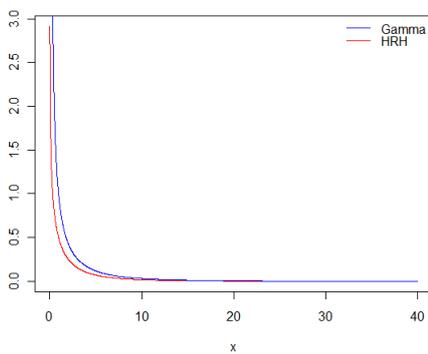


(b) estimateur IG-RHR

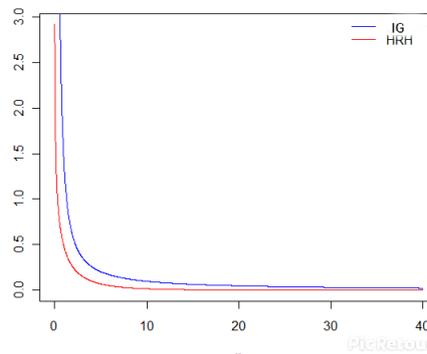


(c) estimateur RIG-RHR

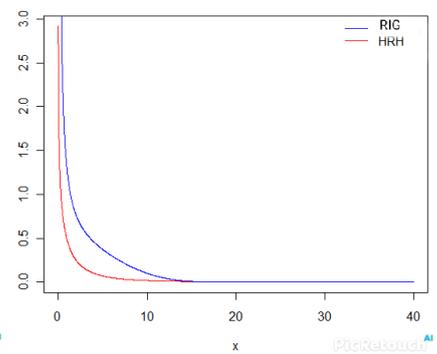
FIGURE 3.10 – Estimateurs du RHR de la distribution $\text{Gamma}(3, 1)$ pour $n=500$ et $h[UCV]$



(a) estimateur Gamma-RHR



(b) estimateur IG-RHR



(c) estimateur RIG-RHR

FIGURE 3.11 – Estimateurs du RHR de la distribution $LN(1, 1)$ pour $n=500$ et $h[UCV]$

D'après les résultats présentés dans le tableau 3.6, on constate globalement que les valeurs moyennes de ISE diminuent lorsque la taille de l'échantillon n augmente,

- Dans le cas de la distribution Gamma,
 - Pour la densité, les meilleurs valeurs du ISE sont obtenus à partir du cas de noyau Gamma pour les petites tailles n et le noyau RIG pour les échantillons de taille n moyenne et grande.
 - Pour le RHR, Contrairement au résultats de la densité les plus faibles valeurs du ISE sont obtenus dans le cas du RIG pour $n=50$, et le cas du Gamma pour les grandes tailles n .

- Dans le cas de la distribution Log normal,
 - Pour la densité, la valeur de ISE obtenue pour le noyau gamma est bonne par rapport à celle obtenus pour les noyaux IG et RIG dans le cas de $n=50$. Cependant, les meilleurs résultats sont obtenus pour le noyau RIG pour les taille $n=(100, 500, 1000)$.
 - Pour le RHR, les plus faibles valeurs du ISE sont obtenus dans le cas du noyau Gamma.

3.3 conclusion

Dans ce dernier chapitre, nous avons réalisé des applications à des données réelles et des simulations de la fonction de taux de hasard et du taux de hasard inversé dans les cas paramétriques et non paramétriques.

Conclusion générale

Dans ce mémoire on s'est intéressé à l'étude de la fonction taux de hasard inversé qui est une fonction peu connue relativement à la fonction de défaillance, et qui trouve des application dans plusieurs domaines (l'ingénierie, médecine, psychologie,...). Nous avons étudié la relation entre ces deux fonctions et donnée certains propriétés. Nous avons également considéré l'estimation de cette fonction par deux approches : paramétrique et non paramétrique, nous avons appliqué ces deux approches sur des données complètes et censurées et estimer le RHR par les méthodes MLE, KM et noyau.

Les résultats des estimations ponctuelles et ensemblistes sont remarquables, le modèle weibull exponentié est suffisamment flexible pour tenir compte des différentes phases de vie (HR monotone et non monotone). La méthode non paramétrique du KM pour des données censurées donne un HR croissant et RHR décroissant et du noyau avec une fenêtre estimée par la méthode UCV est performante, le noyau Gamma performe mieux que les noyaux IG et RIG. Comme perspective, nous pouvons citer

- Considérer d'autre fonctions exponentié que la weibull.
- Améliorer l'estimation par l'estimation de h par d'autres méthodes telles que, boyes, RT,...
- Appliquer ces méthodes dans le cas discret.
- Comparer entre les méthodes non paramétrique à noyau et KM.

Bibliographie

- [1] Alimi, F., 2022, Test d'une différence de survie entre plusieurs échantillons, Mémoire Mater, Université Mohamed Khider, Biskra.
- [2] Badia, F. G. et Berrade, M.D., 2008, On the reversed hazard rate and inactivity time of mixtures in advances in Mathematical modeling for reliability.
- [3] Block, H., Savits, T. H. and Singh, H., 1998, The reversed hazard rate function, Probability in Engineering and Informational Sciences
- [4] Bouhadjera, F., 2020, Estimation non paramétrique de la fonction de régression pour des données censurées, Thèse Doctorat, Université Badji Mokhtar, Annaba.
- [5] Burkschat, M. et Torrado, N., 2013, On the reversed hazard rate of sequential order statistics. Statistics and Probability Letters.
- [6] Chandra, N. K. and Roy, D., 2001, Some results on reversed hazard rate, Probability in the Engineering and Informational Sciences
- [7] Chekkal, S., 2021, Estimation non paramétrique du taux de défaillance par la méthode du noyau, Thèse Doctorat, Université A-Mira, Bejaia.
- [8] Cheng, Dinah, W., et Zhu, Y., 1993, Optimal order of servers in a tandem queue with general blocking, Queueing systems
- [9] Chen, S. X., 2000, Gamma kernel estimators for density functions, Annals of the Institute of Statistical Mathematics.
- [10] Cherchile, R., 2010, Properties of reverse hazard functions, Journal of Mathematical Psychology.
- [11] Cherfaoui, M., Boualem, M., Aissani, D. et Adjabi, S., 2016, Quelques propriétés des estimateurs à noyaux gamma pour des échantillons de petites tailles, Afrika Statistika.

- [12] Desai, D., Mariappan, V. and Sakhardande, M., 2011, Nature of reversed hazard rate : An investigation. International Journal of Performability Engineering.
- [13] Ferrer, E. et Pavia, J. M., 2014, On the relationship between the reversed hazard rate and elasticity, Article in Statistical Papers.
- [14] Finkelstein, M. S., 2002, On the reversed hazard rate, Reliability Engineering and System Safety.
- [15] Gangopadhyay, K. and Cheung, K. N., Bayesian approach to the choice of smoothing parameter in kernel density estimation. Journal of Nonparametric Statistics,
- [16] Girardin, V., Limnios, N., 2014, Probabilités et introduction à la statistique.
- [17] Gross, S.T. et Huber-Carol, C., 1992, Modeles de regression pour les donnees de survies tronquees, Scandinavian journal of statistics.
- [18] Gupta, Rameshwar, D., Gupta, Ramesh, C. et Sankaran P.G, 2004, Some characterization results based on factorization of the (reversed) hazard rate function, Communication in statistics.
- [19] Kajima, M., (1998), Hazard rate and reversed hazard rate monotonicities in continuous time markov chains, journal of applied probability
- [20] Kalbfleisch, J.D. et Lawless, J.F., 1991, Regression models for right truncated data with application to AIDS incubation times and reporting and reporting lags, Statistica Sinica.
- [21] Kaplan, E.L et Meier, P., 1958, Nonparametric estimation from incomplete observations, Journal of the American statistical association.
- [22] Klein, M., 2003, Survival analysis techniques for censored and truncated data, Spring.
- [23] Laggoune, R., 1998, Analyse du retour d'experience pour l'optimisation de la maintenance dans une raffinerie de petrole, These Doctorat, Universite de Boumerdes.
- [24] Lheritier, H., 2004, Theorie de l'estimation parametrique, Edp Sciences.
- [25] Oliveria, P. et Torrado, N., 2015, On proportional reversed failure rate class, Article in Statistical Papers.
- [26] Poorsaeed, M.H., 2010, A note on the mean past and the mean residual life $(n-k+1)$ out-of- n system under multi monitoring, Statistical papers.

- [27] Razmakhah, M.,Morrabi, H., et Ahmadi, J., 2012, Comparing two sampling schemes based on entropy of record statistics, Statistical papers
- [28] Remesh, C., Gupta et Han Wu, 2001, Analysing survival data by proportional reversed hasard model, International journal of reliability and application.
- [29] Scaillet, O., 2004, Density estimation using inverse and reciprocal inverse Gaussian kernels, Journal of Nonparametric Statistics. v
- [30] Shaked, Moshe et Shanthikumar,J., George, 2006, Stochastic orders, New york, Springer.
- [31] Silverman, B., 1986, Density Estimation for Statistics and Data Analysis, Monographs on statistics and spplied probability, Chapman and Hall. New York.
- [32] Srivastava, D. K., Mudholkar, G. S., Freimer, M., 1995, The exponentiated Weibull family ; a reanalysis of the bus motor failure data, Technometrics.
- [33] Tavangar, M., 2020, On the behaviour of the failure rate and reversed failure rate in engineering systems, Article in Journal of applied Probability.
- [34] Townsend, J.T. et Wenger, M.J., 2004, A theory of interactive parallel precessing : Nouvelles mesures de capacite et predictions pour une serie d'inegalites de temps de reponse, Psychological Review.
- [35] Turnbull, B.W., 1976, The Empirical Distribution Function with Arbitrary Grouped Censored and Truncated Data.
- [36] Xie,M., Gaudoin, O. et Bracquemond, C., 2002, Redefining failure rate function for discret distribution, International journal of Reliability.
- [37] Zougab, N., 2013, Approche bayésienne dans l'estimation non paramétrique de la densité de probabilité et la courbe de régression de la moyenne, Thèse Doctorat, Université A.Mira, Bejaia.

Annexe

Pour les données complètes, les dérivées secondes du système (S1) sont données par :

$$\begin{aligned}
\partial^2 l / \partial \alpha^2 &= -N/\alpha^2 + (\theta - 1) \sum_{i=1}^N g_\alpha(t_i) (g(t_i) \ln(t_i/\sigma) (1 - (t_i/\sigma)^\alpha) - g_\alpha(t_i)) / g^2(t_i) - \sum_{i=1}^N \ln^2(t_i/\sigma) (t_i/\sigma)^\alpha \\
\partial^2 l / \partial \alpha \partial \theta &= \sum_{i=1}^N g_\alpha(t_i) / g(t_i) \\
\partial^2 l / \partial \alpha \partial \sigma &= \sum_{i=1}^N ((\alpha/\sigma) (t_i/\sigma)^\alpha \ln(t_i/\sigma) + (1/\sigma) (t_i/\sigma)^\alpha) - \sum_{i=1}^N 1/\sigma + (\theta - 1) \sum_{i=1}^N (\alpha^2 (t_i/\sigma)^\alpha g_\alpha(t_i) g(t_i) - \\
&\alpha^2 g_\alpha(t_i) g(t_i) + \sigma g_\alpha(t_i) g(t_i) - \alpha \sigma g_\alpha(t_i) g_\sigma(t_i)) / \alpha \sigma g^2(t_i) \\
\partial^2 l / \partial \theta \partial \alpha &= \sum_{i=1}^N g_\alpha(t_i) / g(t_i) \\
\partial^2 l / \partial \theta^2 &= -N/\theta^2 \\
\partial^2 l / \partial \theta \partial \sigma &= \sum_{i=1}^N g_\sigma(t_i) / g(t_i) \\
\partial^2 l / \partial \sigma \partial \alpha &= -N/\sigma + (1/\sigma) \sum_{i=1}^N (t_i/\sigma)^\alpha + (\alpha/\sigma) \sum_{i=1}^N (t_i/\sigma)^\alpha \ln(t_i/\sigma) + (\theta - 1) \sum_{i=1}^N ((\alpha/\sigma) ((t_i/\sigma)^\alpha - \\
&1) g_\alpha(t_i) g(t_i) + (1/\sigma) g_\sigma(t_i) g(t_i) - g_\alpha(t_i) g_\sigma(t_i)) / g^2(t_i) \\
\partial^2 l / \partial \sigma \partial \theta &= \sum_{i=1}^N g_\sigma(t_i) / g(t_i) \\
\partial^2 l / \partial \sigma^2 &= N\alpha/\sigma^2 - (\alpha/\sigma^2 + \alpha^2/\sigma^2) \sum_{i=1}^N (t_i/\sigma)^\alpha + (\theta - 1) \sum_{i=1}^N (g(t_i) g_\sigma(t_i) ((\alpha/\sigma) (t_i/\sigma)^\alpha - 1/\sigma - \\
&\alpha/\sigma) - g_\sigma^2(t_i)) / g^2(t_i)
\end{aligned}$$

La matrice d'information observée est donnée par

$$I_0 = - \begin{bmatrix} \frac{\partial^2 l}{\partial \alpha^2} & \frac{\partial^2 l}{\partial \alpha \partial \theta} & \frac{\partial^2 l}{\partial \alpha \partial \sigma} \\ \frac{\partial^2 l}{\partial \theta \partial \alpha} & \frac{\partial^2 l}{\partial \theta^2} & \frac{\partial^2 l}{\partial \theta \partial \sigma} \\ \frac{\partial^2 l}{\partial \sigma \partial \alpha} & \frac{\partial^2 l}{\partial \sigma \partial \theta} & \frac{\partial^2 l}{\partial \sigma^2} \end{bmatrix} \Big|_{\alpha, \theta, \sigma = \hat{\alpha}, \hat{\theta}, \hat{\sigma}}$$

Résumé

Dans ce mémoire, nous avons traité le problème de l'estimation de la fonction taux de hasard inversé (RHR) par les méthodes MLE, Kaplan-Meier et méthode du noyau, dans le cas des données complètes et censurées. Des applications sont effectuées dans le cas d'un modèle weibull exponentié et par des noyaux asymétriques : Gamma, IG et RIG et une fenêtre de lissage estimée par UCV. Les résultats d'application confirment les résultats théoriques et les performance des estimateurs.

Mots clés : Estimation paramétrique, estimations non paramétrique, fonction de survie, taux de hasard, taux de hasard inversé, paramètre de lissage, noyau, Kaplan-Meier, UCV, MLE

Abstract

In this work, we have dealt with the problem of estimating the inverse hazard rate function (RHR) by the MLE, Kaplan-Meier and kernel methods, in the case of complete and censored data. Applications are made in the case of an exponential weibull model and by asymmetric kernels : Gamma, IG and RIG and a bandwidth estimated by UCV.

The application results confirm the theoretical results and the performance of the estimators.

Keywords : Parametric estimation, non-parametric estimation, survival function, hazard rate, reversed hazard rate, bandwidth, kernel, Kaplan-Meier, UCV, MLE