

وزارة التعليم العالي والبحث العلمي

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

BADJI MOKHTAR-ANNABA
UNIVERSITY
UNIVERSITE BADJI MOKHTAR
ANNABA



جامعة باجي مختار
- عنابة -

Faculté des Sciences

Département de Mathématiques

Année : 2024/2025



THÈSE

Présentée en vue de l'obtention du diplôme de Doctorat

Le Titre

Modélisation Statistique

Filière

Mathématiques Appliquées

Spécialité

Actuariat

Par

BOUAZIZ Asma

DIRECTEUR DE THÈSE : SEDDIK AMEUR Nacira

Prof. U.B.M. Annaba

CO-DIRECTEUR DE THÈSE : CHOUIA Sana

M.C.A. U.B.M. Annaba

Devant le jury

PRESIDENT : Chadli Assia

Prof. U.B.M. Annaba

EXAMINATEUR : Zeghdoudi Halim

Prof. U.B.M. Annaba

EXAMINATEUR : Remita Med Riad

Prof. ENSIA - Alger

EXAMINATEUR : Merzougui Mouna

Prof. U.B.M. Annaba

Remerciements

Avant toute chose, je rends grâce à **Allah**, le Tout-Puissant, qui m'a soutenue dans les moments de faiblesse, réconfortée dans l'incertitude et insufflé l'espoir dans les instants de doute. Sa précieuse guidance m'a accompagnée tout au long de mon parcours et m'a donné la force et la détermination nécessaires pour mener à bien ce travail de recherche.

[كما قال تعالى : [وَإِذْ تَأَذَّنَ رَبُّكُمْ لَئِن شَكَرْتُمْ لَأَزِيدَنَّكُمْ وَلَئِن كَفَرْتُمْ إِنَّ عَذَابِي لَشَدِيدٌ]

سورة ابراهيم (٧)

Je tiens à exprimer ma profonde gratitude à Madame le Professeur **SEDDIK-AMEUR Nacira**, ma directrice de thèse, dont l'engagement sans faille, la générosité intellectuelle et la bienveillance ont été des piliers essentiels dans l'élaboration de ce travail. Son savoir, son expérience et ses conseils avisés ont enrichi ma réflexion et guidé la rédaction de cette thèse. Que Dieu la récompense pleinement.

Je suis également profondément reconnaissante envers le Docteur **Chouia Sana**, mon co-encadrant.

Mes respectueux remerciements vont à Madame le Professeur **Chadli Assia**, qui m'a fait l'honneur de présider le jury, ainsi qu'à Monsieur le Professeur **Zeghdoudi Halim**, Monsieur le Professeur **Remita Mohamed Riad** et Madame le Professeur **Merzougui Mouna**, pour avoir accepté de faire partie du jury et pour le temps précieux qu'ils y ont consacré.

Un merci tout particulier à **mes très chers parents**, dont l'amour inconditionnel, le soutien indéfectible et les encouragements constants ont été une source inestimable de force et de persévérance.

Un immense merci à mon mari, bien-aimé, **Hamdi**, dont l'amour, la présence attentive, le soutien indéfectible et les nombreux sacrifices ont rendu possible la réalisation de cette œuvre.

Enfin, et surtout, merci à Allah de m'avoir comblée des plus doux des anges : mes enfants **Nizar, Maria** et **Noursine**, qui illuminent ma vie par leur simple regard.

Je remercie mes amis, Oh ! combien nombreux, je leurs témoigne toute ma reconnaissance. Merci à eux pour leur sympathie et leur bonne humeur. en particulier **Imene Ouchen**, ainsi qu'à toutes les personnes qui, de près ou de loin, ont contribué à faire éclore cette thèse.

Dédicace

Avec une profonde gratitude et un amour incommensurable, je dédie cette thèse à celles et ceux qui ont illuminé mon chemin par leur soutien et leur bienveillance.

À mes parents, **Rabah** et **Nassima**, Tout au long de mon cursus, ils m'ont toujours soutenue, encouragée et aidée. Ils ont su me donner toutes les chances pour réussir. Qu'ils trouvent, dans la réalisation de ce travail, l'aboutissement de leurs efforts ainsi que l'expression de ma plus affectueuse gratitude.

À mon mari, **Hamdi**, dont le soutien inébranlable, l'écoute attentive et la présence rassurante ont été essentiels tout au long de ces années.

À mes précieux enfants, **Nizar**, **Maria** et **Noursine**, mes anges, source inépuisable de motivation et d'inspiration, dont l'amour emplit ma vie de bonheur.

À mes très chères sœurs **Amira** et **Abir** et ses enfants, **Meriem**, **Bissan** et **Tamim**, ainsi qu'à mes frères bien-aimés **Hamdi** et **Raed**, je vous adresse toute ma reconnaissance et mon amour profond. Merci pour votre affection sincère, votre bienveillance constante et cette complicité unique qui illumine chacun de mes jours.

À mes beaux-parents **Nacer** et **Heddi**.

À toute ma famille paternelle et maternelle.

Merci à vous tous pour votre amour, votre patience et votre confiance. Cette réussite est aussi la vôtre.

Modélisation Statistique

Résumé

Cette thèse est dédiée d'abord à l'étude de modèles qui traitent le cas où les durées de vie sont sujets à plusieurs causes de panne (modèles à risques concurrents) comme le modèle bi-Weibull et un modèle de l'extension exponentielle généralisée par la méthode alpha-power transformation. Ensuite on propose une nouvelle distribution généralisant celle de Chen qui peut décrire des données de différents domaines comme la fiabilité, la médecine et les assurances. Pour vérifier l'ajustement de chacune de ces distributions aux données réelles, on a développé de nouveaux critères de tests puissants, pratiques et efficaces.

Mots-Clés : Analyse de survie, fiabilité, maximum de vraisemblance estimateurs, modèle à risques concurrents, tests d'ajustement, tests du khi-deux modifié.

Statistical Modeling

Abstract

This thesis is dedicated first to the study of models that deal with the case where lifetimes are subject to several causes of failure (competing risk models) such as the bi-Weibull model and a generalization of the extension exponential distribution by the alpha-power transformation method. Then we propose a new model generalizing the Chen distribution that can describe data from various fields such as reliability, medicine and insurance. To verify the fit of each of these distributions to real data, we have developed new powerful, practical and efficient test criteria.

Keywords : Competing risks models, maximum likelihood estimates, modified chi-square statistics, goodness-of-fit tests, reliability, survival analysis.

النمذجة الإحصائية

ملخص

تركز هذه الأطروحة في البداية على دراسة النماذج التي تتناول الحالات التي تكون فيها أعمار الأفراد عرضة لأسباب متعددة للفشل (نماذج المخاطر المتنافسة)، مثل نموذج ثنائي-وييل ونموذج التوسع الأسي المعمم باستخدام طريقة تحويل القوة ألفا. ثم نقترح توزيعًا جديدًا يعمم توزيع Chen ، والذي يمكنه وصف البيانات من مجالات مختلفة مثل الموثوقية والطب والتأمين. وللتحقق من ملاءمة كل من هذه التوزيعات للبيانات الواقعية، قمنا بتطوير معايير اختبار جديدة قوية وعملية وفعالة.

الكلمات المفتاحية: تحليل البقاء، الموثوقية، مقدرات أقصى احتمالية، نموذج المخاطر المتنافسة، اختبارات جودة الملاءمة، اختبارات مربع كاي المعدل.

Table des matières

Introduction générale	1
1 Préliminaires	4
1.1 Objectifs et intérêts de l'analyse de survie	5
1.2 Concepts généraux de l'analyse de survie	5
1.2.1 Distributions de la durée de survie	5
1.2.2 Estimation par la méthode du maximum de vraisemblance	9
1.3 Tests d'ajustement	10
1.3.1 Tests non paramétriques	10
1.3.2 Test de Nikulin, Rao et Robson (NRR)	12
2 Modèle des risques concurrents Bi-weibull	14
2.1 Présentation du modèle BW	15
2.2 Estimation	18
2.2.1 Estimation des paramètres inconnus	18
2.2.2 La matrice d'information de Fisher	20
2.3 Test d'ajustement pour le modèle BW	21
2.3.1 Tests non paramétriques	22
2.3.2 Construction de la statistique NRR pour la distribution BW	23
2.4 Simulations et Applications réelles	27
2.4.1 Simulations	27
2.4.2 Applications	30

3	Le modèle alpha-power transformé de l'extension exponentielle (APTEE)	34
3.1	Présentation du modèle	35
3.2	Estimation	37
3.2.1	Estimation du maximum vraisemblance	37
3.2.2	La matrice d'information de Fisher	38
3.3	Tests d'ajustement pour le modèle APTEE	39
3.3.1	Tests non paramétriques	39
3.3.2	Construction du test statistique NRR pour la distribution APTEE .	41
3.4	Simulations et Applications réelles	43
3.4.1	Simulations	44
3.4.2	Applications	47
4	Modèle alpha-power transformé de Chen (APTC)	50
4.1	Présentation du modèle APTC	51
4.2	Estimation	54
4.2.1	Estimation des paramètres inconnus	54
4.2.2	La Matrice d'information de Fisher	55
4.2.3	Estimateurs classiques	56
4.3	Test d'ajustement pour le modèle APTC	57
4.4	Simulations et Applications	60
4.4.1	Simulations	61
4.4.2	Application	64
	Conclusion	66

Introduction générale

La modélisation statistique est un outil principal dans la description des durées de vie qui représentent, par exemple, le temps restant avant la panne dans les systèmes artificiels, le temps avant l'occurrence d'un événement en médecine comme le décès ou l'apparition d'une réaction à une thérapie, la durée de vie des produits industriels en fiabilité, la durée de chômage, la durée du temps d'attente entre deux sinistres en assurance, etc... La méthode actuarielle a été la première méthode de l'analyse de survie, ensuite elle a été élargie à tous les domaines d'applications. Modéliser des durées de vie revient à les ajuster à un modèle mathématique ou une loi de probabilité afin de comprendre le phénomène étudié. Vu les développements dans le domaine industriel, le domaine médical, économique et les assurances, les modèles classiques ne suffisent plus à représenter toutes les données recueillies pour l'analyse, d'où l'intérêt des chercheurs à produire de nouveaux modèles statistiques qui seront plus adaptés.

Pour modéliser les taux de défaillances dans les essais de vie, en théorie de fiabilité, dans la modélisation des montants des sinistres en réassurance, en ingénierie des systèmes de communication, dans les systèmes de radars et beaucoup d'autres domaines, diverses lois sont proposées dans la littérature statistique. Le premier modèle est celui de Weibull (1951) à un seul paramètre, qui a été ensuite généralisé à différentes formes. On cite les distributions Weibull à deux et trois paramètres, la Weibull modifiée, l'inverse Weibull, l'inverse Weibull généralisée (*De Gusmao et al, 2011*) et la Weibull exponentielle (*Mudholkar et al, 1993*). Ensuite, de nouvelles formes de généralisations sont apparues utilisant des générateurs à partir d'une distribution de base. Parmi celles-ci, on trouve la Weibull généralisée basée sur la loi beta appelée beta-Weibull-G distributions (*Cordeiro*

et al., 2014), la famille des distributions Weibull-X (Alzaatreh et al., 2016), la famille Weibull-G (Bourguignon et al., 2014) et la Weibull-X exponentiée (Alzaghal et al., 2013).

Dans ce travail, on s'est intéressé à une distribution maniable basée sur la Weibull à deux paramètres. Ce modèle, noté Bi-Weibull, obtenu en prenant la loi du minimum entre deux distributions de Weibull à deux paramètres chacune, représente des durées de survie où le taux de panne est soumis à différentes causes, c'est ce qu'on appelle un modèle à risques concurrents. Ces modèles sont très présents en fiabilité où le taux de panne dépend des facteurs extérieurs d'utilisation, en médecine où le patient peut mourir d'une autre cause que sa maladie, en assurance où les sinistres peuvent se produire d'une cause extérieure, etc

D'autres lois que celle de Weibull ont été aussi généralisées pour les rendre plus flexibles. Parmi celles-ci, une nouvelle famille de générateurs de distributions (Mahdavi and Kundu, 2017) a été introduite récemment, appelée la transformée à puissance α (α -power transformed distributions). Ce générateur peut inclure plusieurs modèles selon la distribution de base choisie et comme dans leur cas c'est le modèle *APTE* (α -power transformed exponential) avec l'exponentielle comme baseline. Afify et al., (2020) ont appliqué cette distribution (*APTE*) en actuariat pour évaluer le degré de risque d'exposition des entreprises à des aspects particuliers du risque; tandis que Alghamedi et al., (2019) ont considéré le cas de la distribution exponentielle généralisée (*EE*) de Nadarajah et Haghghi (2011), pour obtenir le modèle *APTEE*. Les auteurs ont passé en revue ses propriétés statistiques et les estimateurs du maximum de vraisemblance mais les autres méthodes d'estimation n'ont pas été considérées et la validation du modèle n'a pas été étudiée, ce que nous avons réalisée dans ce modeste travail.

En utilisant cette approche, on a construit un nouveau modèle dont la distribution de base est la loi de Chen (2000) et qu'on appellera α -power transformed Chen distribution, notée *APTC*. Après l'étude des caractéristiques, les paramètres inconnus sont évalués selon différentes méthodes (maximum de vraisemblance, Anderson-Darling, Cramer Von-Mises, Kolmogorov-Smirnov), un critère de test du type du khi-deux modifié a aussi été fourni permettant de valider ce modèle de manière efficace.

Cette thèse est scindée en quatre chapitres.

Le premier chapitre présente quelques notions de base de la théorie de l'analyse de survie.

Le chapitre deux est dédié à l'étude du modèle bi-Weibull à risques concurrents. Différentes méthodes d'estimation sont données pour évaluer les paramètres inconnus, ensuite en utilisant une approche proposée par Nikulin, Rao et Robson, on a construit un test d'ajustement efficace pour ce modèle.

Le chapitre trois concerne la distribution *APTEE*. Différents estimateurs ont été proposés pour déterminer les paramètres du modèle. Une statistique du type du khi-deux modifié a été fournie ainsi que quelques tests basés sur les EDF (empirical distribution functions) pour valider ce modèle.

Le dernier chapitre est réservé à l'introduction d'un nouveau modèle α -power transformed Chen (*APTC*). Les caractéristiques statistiques sont présentées et les paramètres du modèle sont calculés de plusieurs manières. Enfin la mise en oeuvre d'une statistique de test, capable d'ajuster cette distribution sans avoir à évaluer les distributions alternatives possibles, a été élaborée.

Des simulations et des exemples de données réelles de la fiabilité, de l'analyse de survie et des assurances ont été utilisées pour illustrer la flexibilité des modèles proposés et la praticabilité des tests construits pour chacun d'eux.

Au cours des dernières années, l'analyse de survie a suscité un intérêt croissant, tant pour ses applications essentielles dans les domaines industriels que pour son rôle central en modélisation statistique. Son développement théorique s'est considérablement enrichi, notamment en raison de son utilité dans l'étude des processus stochastiques.

Ce chapitre est consacré à l'exploration approfondie de la théorie de l'analyse de survie. Il définit les concepts fondamentaux de cette discipline et les grandeurs qui la caractérisent, tout en mettant en évidence les méthodes d'estimation des paramètres associés. L'un des défis majeurs en modélisation statistique réside dans le choix de la distribution de probabilité la plus adaptée à l'analyse des données. Pour cela, divers critères paramétriques et non paramétriques sont mobilisés, parmi lesquels les tests du khi-deux occupent une place prépondérante.

L'analyse de survie permet d'évaluer la capacité d'un système ou d'un matériel à remplir une fonction spécifique dans des conditions données et sur une période déterminée. Elle constitue ainsi un outil essentiel pour la fiabilité des systèmes et l'optimisation des performances dans divers domaines scientifiques et industriels.

D'après les normes de l'AFNOR X60-500, l'analyse de survie se définit comme la capacité d'un système à exécuter une fonction spécifique pendant une durée déterminée et dans des conditions opérationnelles précises. Les conditions se réfèrent à toutes les contraintes externes, qu'elles soient d'origine humaine, climatique, physique, etc.

1.1 Objectifs et intérêts de l'analyse de survie

L'évaluation de la durée de fonctionnement (fiabilité) est une étape essentielle dans toute analyse de la sécurité des systèmes. Initialement, l'étude de la survie était principalement appliquée aux systèmes de haute technologie tels que les centrales nucléaires et le domaine aérospatial. De nos jours, l'analyse de survie est devenue un élément essentiel pour évaluer la qualité et faciliter la prise de décision dans l'étude de divers domaines tels que le transport, l'énergie, les bâtiments, les composants électroniques, les composants informatiques, etc.

De nombreux acteurs de l'industrie s'investissent dans l'évaluation et l'amélioration de l'analyse de la durée de vie de leurs produits tout au long de leur cycle de développement, depuis la phase de conception jusqu'à la mise en service (incluant la conception, la fabrication et l'exploitation). Cette démarche vise à approfondir leur compréhension du rapport coût/analyse de la durée de vie et à identifier et contrôler les sources de défaillance.

1.2 Concepts généraux de l'analyse de survie

Le néologisme "analyse de survie", qui correspond à la traduction française du terme anglais "Reliability", reflète la confiance de l'utilisateur envers l'appareil qu'il utilise ou auquel il est confronté. C'est seulement le 9 avril 1962 que ce terme nouveau a été officiellement reconnu par l'Académie des sciences, qui l'a défini comme suit : "Une grandeur décrivant la fiabilité de fonctionnement, ou l'évaluation de la probabilité de fonctionnement d'un équipement selon des normes spécifiques".

Cette définition englobe divers concepts tels que la probabilité de bon fonctionnement, les performances requises, les conditions opérationnelles et la durée de vie.

1.2.1 Distributions de la durée de survie

Supposons que la durée de survie T soit une variable aléatoire positive ou nulle, et absolument continue, alors sa loi de probabilité peut être définie par l'une des cinq fonctions

équivalentes suivantes (chacune des fonctions ci-dessous peut être obtenues à partir de l'une des autres fonctions) :

Fonction de survie S (ou fonction de fiabilité R)

La fonction de survie est, pour t fixé, la probabilité de survivre un être vivant jusqu'à l'instant t (dans le domaine medical), c'est-à-dire:

$$S(t) = \mathbf{P}(T > t), \quad t \geq 0.$$

Considérons un matériel dont on étudie la fiabilité. Alors, on a la fonction de fiabilité R définie pour tout $t \geq 0$ par:

$$R(t) = S(t) = 1 - F(t).$$

Fonction de répartition F

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

$$F(t) = \mathbf{P}(T \leq t) = 1 - S(t) = 1 - R(t).$$

On appelle fonction défaillance la fonction F définie pour tout $t \geq 0$, où elle représente la probabilité qu'un dispositif choisi au hasard ait une défaillance avant l'instant t .

- **Remarque:** *Il est arbitraire de décider que $S(t) = P(T \geq t)$ ou $S(t) = P(T > t)$. Cela n'a aucune importance quand la loi de T est continue car $P(T \geq t) = P(T > t)$. Dans les cas où F a des sauts (quand le temps est discret, par exemple, compté en mois ou semaine), on utilise les notations suivantes:*

$$F^-(t) = P(T < t) \quad \text{et} \quad F^+(t) = P(T \leq t)$$

où F^- est la limite à gauche et F^+ la limite à droite de F (définitions et notations sont identiques pour la fonction S). Remarquons que $F^- \leq F^+$ et $S^- \geq S^+$.

Densité de probabilité f

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

$$F(t) = \int_0^t f(s) ds.$$

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

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

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

Risque instantané h

Le risque instantané, aussi souvent appelé "taux de hasard" (taux d'incidence), pour t fixé caractérise la probabilité de mourir dans un petit intervalle de temps après l'instant t , conditionnellement au fait d'avoir survécu jusqu'à l'instant t . Aussi cela signifie-t-il le risque de mort instantané pour ceux qui ont survécu :

$$h(t) = \lim_{h \rightarrow 0} \frac{\mathbf{P}(t \leq T < t + h / X \geq t)}{h} = \frac{f(t)}{S(t)} = -\ln(S(t))' = -\ln(R(t))'.$$

Soit T une variable aléatoire qui à chaque matériel associe son temps de bon fonctionnement. On choisit un de ces matériels au hasard.

Soit les événements A : «Le matériel est en état de bon fonctionnement à l'instant t ».

et B : «Le matériel est défaillant à l'instant $t + \Delta t$ » on a alors:

$$\mathbf{P}(A) = \mathbf{P}(T > t) \text{ et } \mathbf{P}(B) = \mathbf{P}(T \leq t + \Delta t)$$

$$\text{Donc } \mathbf{P}(A \cap B) = \mathbf{P}(t < T < t + \Delta t)$$

$$= F(t + \Delta t) - F(t)$$

$$= (1 - R(t + \Delta t)) - (1 - R(t))$$

$$= R(t) - R(t + \Delta t).$$

$$\text{On en déduit que: } \mathbf{P}(B/A) = \frac{\mathbf{P}(A \cap B)}{\mathbf{P}(A)} = \frac{R(t) - R(t + \Delta t)}{R(t)}$$

Alors:

$$h(t) = \lim_{\Delta t \rightarrow 0} \left(\frac{1}{\Delta t} \cdot \frac{R(t) - R(t + \Delta t)}{R(t)} \right).$$

• Importance de la fonction de risque

La fonction de risque $h(t)$ fournit la description la plus concrète d'une distribution de survie.

La distribution la plus simple est celle où le risque instantané est constant. Le risque instantané de décès ne dépend pas du temps: on dit qu'il n'y a pas de vieillissement.

Ce n'est à peu près vrai, dans le cas d'une population humaine occidentale, que pour la mortalité entre 5 et 15 ans qui est surtout liée aux accidents. On considère encore que ce modèle peut convenir pour des adultes s'ils sont observés pendant une courte période.

En général, dans une population adulte, le risque instantané de décès augmente avec l'âge: on dit qu'il y a vieillissement, $h(t)$ est une fonction croissante.

Le risque instantané de décès pour les enfants de moins d'un an, diminue avec l'âge entre zéro et un an, $h(t)$ est décroissant.

Taux de hasard cumulé λ

Le taux de hasard cumulé est l'intégrale du taux de hasard h (risque instantané) :

$$\lambda(t) = \int_0^t h(u) du = -\ln[S(t)].$$

On peut déduire la fonction de survie du taux de hasard cumulé grâce à la relation:

$$S(t) = \exp[-\lambda(t)] = \exp\left[-\int_0^t h(u) du\right].$$

On peut définir la distribution de probabilité de la variable T à partir de l'une des fonctions: $f(t)$, $h(t)$, $F(t)$, $S(t)$ ou $R(t)$.

Tableau -Relations entre $F(t)$, $R(t)$, $f(t)$, et $h(t)$				
Fonction	$F(t)$	$R(t)$	$f(t)$	$h(t)$
$F(t)$		$1 - R(t)$	$\int_0^t f(u) du$	$1 - \exp\left(-\int_0^t h(u) du\right)$
$R(t)$	$1 - F(t)$		$\int_0^\infty f(u) du$	$\exp\left(-\int_0^t h(u) du\right)$
$f(t)$	$\frac{dF(t)}{dt}$	$-\frac{dR(t)}{dt}$		$h(t) \exp\left(-\int_0^t h(u) du\right)$
$h(t)$	$\frac{\frac{dF(t)}{dt}}{1-F(t)}$	$-\frac{R'(t)}{R(t)}$	$\frac{f(t)}{\int_0^{+\infty} f(u) du}$	

1.2.2 Estimation par la méthode du maximum de vraisemblance

soit T une variable aléatoire réelle de loi paramétrique (discrète ou continue), dont veut estimer le paramètre θ . Alors, on définit une fonction f telle que

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

si T est une variable aléatoire continue de densité f_{θ} et

$$f(t, \theta) = P_{\theta}(T = t)$$

Si X est une variable aléatoire discrète.

- **Définition:**

On appelle fonction de vraisemblance de θ pour une réalisation (t_1, \dots, t_n) d'un échantillon, la fonction de θ :

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

La méthode qui consiste à estimer θ par la valeur qui maximise $L_{(t_1, \dots, t_n)}$ (la vraisemblance) s'appelle "méthode du maximum de vraisemblance".

On appelle $\hat{\theta}$ l'estimateur associé:

$$\hat{\theta} = \left\{ \theta : L_{(t_1, \dots, t_n)}(\hat{\theta}) = \sup_{\theta \in \Theta} L_{(t_1, \dots, t_n)}(\theta) \right\}$$

Ceci est un problème d'optimisation. On utilise généralement le fait que si $L_{(t_1, \dots, t_n)}$ est dérivable et si $L_{(t_1, \dots, t_n)}$ admet un maximum global en une valeur, alors la dérivée première de $L_{(t_1, \dots, t_n)}$ s'y annule et sa dérivée seconde y est négative.

Réciproquement, si la dérivée première de $L_{(t_1, \dots, t_n)}$ s'annule en $\theta_0 = \theta^*$, et si sa dérivée seconde est négative strictement en θ^* , alors θ^* est un point où $L_{(t_1, \dots, t_n)}$ admet un maximum global.

La vraisemblance étant positive et le logarithme népérien, \log étant une fonction croissante, il est équivalent et souvent plus simple de maximiser le logarithme népérien de la

vraisemblance (le produit se transforme en somme, ce qui est plus simple à dériver) et plus facile à calculer numériquement.

1.3 Tests d'ajustement

L'évaluation de l'adéquation d'un modèle statistique à un ensemble de données empiriques repose sur des tests d'ajustement. Ces tests permettent de vérifier si une distribution théorique représente correctement les observations et jouent un rôle crucial en analyse statistique. Ils sont particulièrement utiles pour la validation des modèles probabilistes et la sélection de distributions optimales.

1.3.1 Tests non paramétriques

Les tests non paramétriques jouent un rôle fondamental dans l'évaluation statistique, notamment lorsque l'on cherche à comparer une distribution théorique avec un ensemble de données empiriques, ou à déterminer si deux échantillons suivent la même loi de probabilité. Contrairement aux méthodes paramétriques, ces tests ne nécessitent pas de supposer une forme particulière de la distribution sous-jacente, ce qui les rend particulièrement utiles pour l'analyse de données en conditions incertaines.

Parmi les tests les plus couramment utilisés, on retrouve :

Test de Kolmogorov -Smirnov

Le test de Kolmogorov-Smirnov (KS), introduit par Kolmogorov (1933) et amélioré par Smirnov (1948), est un outil statistique fondamental permettant d'évaluer l'adéquation entre une distribution théorique et un échantillon empirique, ou de comparer deux échantillons indépendants pour déterminer s'ils suivent la même loi de probabilité. Ce test repose sur la fonction de distribution empirique (EDF) et se base sur la plus grande différence verticale entre la distribution observée et celle attendue sous l'hypothèse nulle. Conover (1999) a précisé que la statistique KS est définie à partir des n observations ordonnées $t_1 < t_2 < \dots < t_n$ et permet d'obtenir une mesure robuste de la divergence entre

les distributions. Grâce à sa simplicité et son efficacité, ce test est largement utilisé en analyse des données et en validation des modèles probabilistes, notamment lorsqu'il s'agit de tester la normalité d'un échantillon ou de comparer des distributions sous diverses conditions expérimentales.

$$D_n = \max [D^+; D^-]$$

$$D^+ = \max_{1 \leq i \leq n} \left[\binom{i}{n} - F(t_{(i)}) \right]$$

et

$$D^- = \max_{1 \leq i \leq n} \left[F(t_{(i)}) - \binom{i-1}{n} \right]$$

Test de Cramer-Von-Mises

Conover (1999) a déclaré que le test Cramer-Von-Mises avait été développé par Cramer (1928), Von Mises et Smirnov (1936). La statistique CVM utilisait la fonction de pondération $\psi = 1$, de sorte que la statistique AD de l'équation $D = \max_x |F^*(t) - S_n(t)|$ (test de Lilliefors) devient (Thadewald and buning 2007):

$$W_n^2 = \sum_{i=1}^n \left[F(t_i) - \frac{2i-1}{2n} \right]^2 + \frac{1}{12n}$$

Test d'Andersson-Darling

Le test d'Anderson-Darling (AD), introduit par Anderson et Darling en 1954, est une extension du test de Cramér-von Mises (CVM) qui accorde une importance accrue aux extrémités de la distribution. Contrairement au test CVM, qui mesure la somme des écarts quadratiques entre la fonction de distribution empirique $F_n(t)$ et la distribution théorique $F^*(t)$, le test AD pondère ces écarts en fonction de la position des observations, mettant ainsi davantage l'accent sur les queues de la distribution (Farrel et Stewart, 2006).

Selon Arshad et al. (2003), le test AD est considéré comme l'un des plus puissants parmi les tests basés sur la fonction de distribution empirique (EDF). Il appartient à la

classe des statistiques quadratiques EDF et repose sur la mesure de la différence au carré entre la distribution empirique et la distribution théorique :

$$A_n^2 = -n - \frac{1}{n} \left\{ \sum_{i=1}^n (2i-1) [\ln F(t_{(i)}) + \ln(1 - F(t_{(n+1-i)}))] \right\}$$

1.3.2 Test de Nikulin, Rao et Robson (NRR)

L'un des problèmes importants de la modélisation statistique est le choix de la distribution de probabilité utilisée dans l'analyse. En plus des critères basés sur les fonctions de répartition empiriques (EDF) cités plus haut, on a le fameux test du khi-deux de Pearson X^2 . Seulement, celui-ci n'est pas applicable quand les paramètres sont inconnus et sa distribution n'est plus un khi-deux, ce qui a poussé les chercheurs à le modifier. Dans ce travail, on s'intéresse à la statistique proposée par Nikulin (1973) et Rao et Robson (1974) qui a montré son efficacité.

Soit un échantillon T_1, T_2, \dots, T_n de T suivant une distribution paramétrique $F_0(t, \theta)$, et soit $v = (v_1, v_2, \dots, v_r)^T$ le vecteur des fréquences obtenu en regroupant les T_i dans des intervalles I_j (r classes). Pour tester l'hypothèse nulle H_0 , Pearson (1900) a proposé le premier test basé sur la statistique

$$\chi_n^2(\theta) = X_n^T(\theta) X_n(\theta) = \sum_{j=1}^r \frac{[v_j - nP_j(\theta)]^2}{nP_j(\theta)}$$

où

$$X_n(\theta) = \left(\frac{v_1 - nP_1(\theta)}{\sqrt{nP_1(\theta)}}, \frac{v_2 - nP_2(\theta)}{\sqrt{nP_2(\theta)}}, \dots, \frac{v_r - nP_r(\theta)}{\sqrt{nP_r(\theta)}} \right)^T$$

et que

$$P_j(\theta) = \int_{a_{j-1}}^{a_j} f(t) dt = F(a_j) - F(a_{j-1}), \quad i = 1, 2, \dots, r.$$

avec

$$I_j =]a_{j-1}, a_j]; \quad I_i \cap I_j = \emptyset \quad i \neq j; \quad \bigcup_{j=1}^r I_j = R^1.$$

et

$$-\infty = a_0 \prec a_1 \prec \dots \prec a_{r-1} \prec a_r = +\infty$$

Les auteurs ont modifié le test de Pearson pour en déduire la statistique NRR, notée Y^2 , qui sera distribuée selon un khi-deux. Pour le calcul on utilise souvent la formule quadratique suivante qui dépendra du modèle étudié:

$$Y_n^2(\hat{\theta}_n) = X_n^2(\hat{\theta}_n) + Q_n(\hat{\theta}_n) = X_n^2(\hat{\theta}_n) + \frac{1}{n} L^T(\hat{\theta}_n) \left(I(\hat{\theta}_n) - J(\hat{\theta}_n) \right)^{-1} L(\hat{\theta}_n)$$

où $\hat{\theta}_n$ est l'estimateur du maximum de vraisemblance des paramètres inconnus et I, J sont les matrices d'information sur les données non-groupées et groupées respectivement.

Modèle des risques concurrents Bi-weibull

En théorie des probabilités, la loi de Weibull, du nom du mathématicien Waloddi Weibull (1951), constitue une loi de probabilité continue. Il s'agit de la distribution la plus couramment employée pour modéliser les durées de vie et les temps de défaillance dans les essais de vie, de survie et la théorie de la fiabilité. Sa pertinence réside dans sa flexibilité et ses multiples domaines d'application, notamment l'analyse de survie, l'étude de la fiabilité, la théorie des valeurs extrêmes, l'économie, ainsi que la modélisation des montants des sinistres en réassurance et en assurance générale. Elle est également pertinente en génie électrique, ingénierie industrielle, prévisions météorologiques, dispersion des signaux dans les systèmes radars, industrie éolienne pour caractériser les distributions de vitesse du vent, et ingénierie des systèmes de communication.

Dans les études de fiabilité et en analyse de survie, les données observées présentent fréquemment un comportement non monotone, contrairement à la distribution Weibull classique, ce qui a conduit à la proposition de nouvelles généralisations. Parmi celles-ci, nous avons examiné une distribution à quatre (4) paramètres, dénommée distribution Bi-Weibull (BW).

Les modèles à risques concurrents sont particulièrement pertinents lorsque plusieurs facteurs de défaillance peuvent affecter un système ou un individu, chacun ayant un impact différencié sur la probabilité de survie. La distribution Bi-Weibull, caractérisée

par deux composants Weibull distincts, permet de capturer la complexité des processus de défaillance en intégrant plusieurs mécanismes sous-jacents.

Ce chapitre débute par la présentation de la distribution BW, suivie de l'établissement de ses caractéristiques et d'une représentation graphique des diverses formes du taux de hasard. Par la suite, les statistiques de test d'ajustement sont élaborées pour valider le choix de ce modèle. Des simulations numériques et des applications à des données empiriques sont employées pour valider les résultats obtenus.

2.1 Présentation du modèle BW

Dans cette section, nous supposons que la variable aléatoire T_j suit la distribution de *Weibull* avec les paramètres α_j et β_j , pour $j = 1, 2, \dots, J$. Le cas de cette distribution a été discuté dans plusieurs travaux sur les risques concurrents.

La distribution de *Weibull*, est une fonction caractérisée par deux paramètres, β facteur d'échelle ou paramètre caractéristique de vie et α le facteur de forme. Ces paramètres permettent à la distribution de modéliser un large champ de comportements qui dépendent du temps. La loi de *Weibull* possède un risque instantané monotone.

La notation $T \rightsquigarrow \mathcal{W}(\alpha, \beta)$ est généralement utilisée pour indiquer que la durée de vie T a une distribution de *Weibull* avec les paramètres α et β ($\alpha, \beta > 0$). La durée de vie T admet alors ces fonctions.

Fonction de survie $S(t; \alpha, \beta)$

Spécifiquement, la fonction de survie est donnée par:

$$S(t; \alpha, \beta) = e^{-\left(\frac{t}{\beta}\right)^\alpha} \quad t \geq 0$$

Densité de probabilité $f(t; \alpha, \beta)$

De même, la densité de probabilité est donnée par :

$$f(t; \alpha, \beta) = \frac{d}{dt} F(t; \alpha, \beta) = \begin{cases} \frac{\alpha}{\beta} \left(\frac{t}{\beta}\right)^{\alpha-1} e^{-\left(\frac{t}{\beta}\right)^\alpha} & t \geq 0 \\ 0 & \text{ailleurs} \end{cases}$$

Taux de hasard $\lambda(t; \alpha, \beta)$

La fonction de risque est la suivante:

$$\lambda(t; \alpha, \beta) = -\frac{\frac{d}{dt}e^{-\left(\frac{t}{\beta}\right)^\alpha}}{e^{-\left(\frac{t}{\beta}\right)^\alpha}} = \frac{\frac{\alpha}{\beta}t^{\alpha-1}e^{-\left(\frac{t}{\beta}\right)^\alpha}}{e^{-\left(\frac{t}{\beta}\right)^\alpha}} = \frac{\alpha}{\beta} \left(\frac{t}{\beta}\right)^{\alpha-1} \quad t \geq 0$$

Taux de hasard cumulé $h(t; \alpha, \beta)$

Par conséquent, la fonction de hasard cumulé est

$$h(t; \alpha, \beta) = -\ln(S(t; \alpha, \beta)) = -\ln\left(e^{-\left(\frac{t}{\beta}\right)^\alpha}\right) = \left(\frac{t}{\beta}\right)^\alpha \quad t \geq 0$$

Fonction de répartition $F(t; \alpha, \beta)$

Une variable aléatoire T suit la distribution de Weibull à deux paramètres si sa fonction de distribution cumulative $P(T \leq t) = F(t; \alpha, \beta)$ prend la forme:

$$F(t; \alpha, \beta) = 1 - \exp\left\{-\left(\frac{t}{\beta}\right)^\alpha\right\}$$

où $\beta > 0$ est le paramètre d'échelle et $\alpha > 0$ est le paramètre de forme de Weibull.

Distribution bi-Weibull

Soit $T^{(0)}$ et $T^{(1)}$ des variables aléatoires indépendantes distribuées selon des lois de Weibull de paramètres distincts α_0, β_0 et α_1, β_1 et une fonction de distribution cumulative $F_i(t), i = 0, 1$. Le minimum $T = \min\{T^{(0)}, T^{(1)}\}$ est dit avoir une distribution bi-Weibull, notée *BW*, dont la fonction de répartition et la densité de probabilité sont données respectivement par:

$$F(t; \alpha_0, \alpha_1, \beta_0, \beta_1) = 1 - \exp\left\{-\left[\left(\frac{t}{\beta_0}\right)^{\alpha_0} + \left(\frac{t}{\beta_1}\right)^{\alpha_1}\right]\right\} \quad t \geq 0$$

$$f(t) = \left[\frac{\alpha_1}{\beta_1} \left(\frac{t}{\beta_1}\right)^{\alpha_1-1} + \frac{\alpha_0}{\beta_0} \left(\frac{t}{\beta_0}\right)^{\alpha_0-1}\right] \exp\left\{-\left[\left(\frac{t}{\beta_0}\right)^{\alpha_0} + \left(\frac{t}{\beta_1}\right)^{\alpha_1}\right]\right\}$$

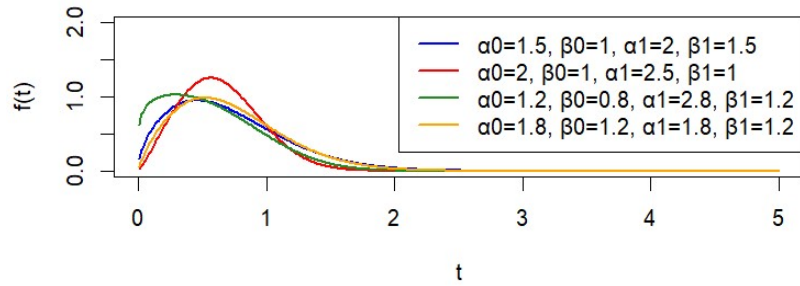


Figure 2.1: Graphe de PDF du modèle BW

Les fonctions de survie et du taux de hasard

$$S(t; \alpha_0, \alpha_1, \beta_0, \beta_1) = \exp \left\{ - \left[\left(\frac{t}{\beta_0} \right)^{\alpha_0} + \left(\frac{t}{\beta_1} \right)^{\alpha_1} \right] \right\} \quad t \geq 0$$

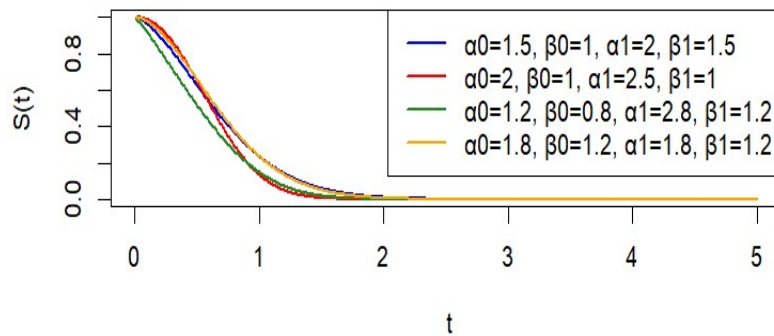


Figure 2.2: Graphe de la fonction de survie du modèle BW

$$\lambda(t) = \frac{\alpha_1}{\beta_1} \left(\frac{t}{\beta_1} \right)^{\alpha_1-1} + \frac{\alpha_0}{\beta_0} \left(\frac{t}{\beta_0} \right)^{\alpha_0-1}$$

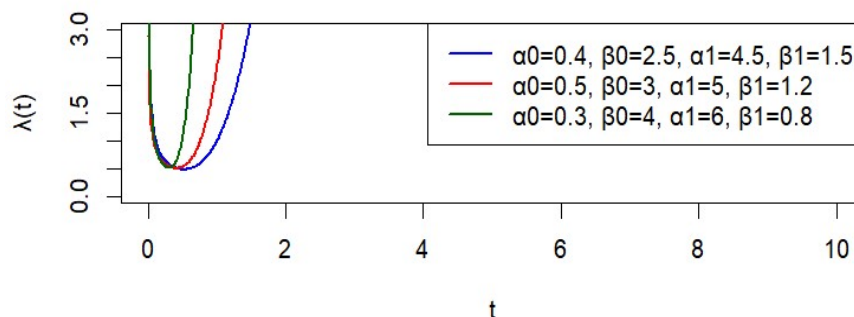


Figure 2.3: Graphe de la fonction de taux de hasard du modèle BW

Comme on le voit sur ces courbes, la fonction du taux de hasard est en forme de baignoire ce qui lui permet de décrire tous types de pannes, aussi bien les pannes de jeunesse (accidentelles) que les pannes de vieillesse ou d'usure.

2.2 Estimation

L'estimation des paramètres inconnus est une étape cruciale dans l'analyse statistique, permettant de déduire les caractéristiques d'une population à partir d'un échantillon de données observées. Dans ce contexte, la méthode du maximum de vraisemblance (MLE) est largement utilisée en raison de ses propriétés de convergence et de normalité asymptotique.

2.2.1 Estimation des paramètres inconnus

Pour calculer les valeurs des paramètres de cette distribution, on utilise la méthode du maximum de vraisemblance (MLE).

Prenons T_1, \dots, T_n comme un échantillon de variables aléatoires complètes indépendantes et identiquement distribuées (i.i.d.) suivant la distribution BW, avec un vecteur

de paramètres $\theta = (\alpha_0, \alpha_1, \beta_0, \beta_1)$. Dans ce contexte, la fonction de vraisemblance de T s'exprime comme suit :

$$\begin{aligned}
 L(t, i) &= \prod_{j=1}^n f(t_j, i_j) \\
 &= \prod_{j=1}^n \left[\frac{\alpha_0}{\beta_0} \left(\frac{t_j}{\beta_0} \right)^{\alpha_0-1} \right]^{1-i_j} \left[\frac{\alpha_1}{\beta_1} \left(\frac{t_j}{\beta_1} \right)^{\alpha_1-1} \right]^{i_j} \left[\exp \left\{ - \left(\frac{t_j}{\beta_0} \right)^{\alpha_0} \right\} \right] \left[\exp \left\{ - \left(\frac{t_j}{\beta_1} \right)^{\alpha_1} \right\} \right] \\
 &= \prod_{j=1}^n \left[\frac{\alpha_0}{\beta_0} \left(\frac{t_j}{\beta_0} \right)^{\alpha_0-1} \right]^{1-i_j} \left[\frac{\alpha_1}{\beta_1} \left(\frac{t_j}{\beta_1} \right)^{\alpha_1-1} \right]^{i_j} \exp \left\{ - \left[\left(\frac{t_j}{\beta_0} \right)^{\alpha_0} + \left(\frac{t_j}{\beta_1} \right)^{\alpha_1} \right] \right\}
 \end{aligned}$$

L'expression de la fonction de log-vraisemblance peut être formulée de la manière suivante :

$$\begin{aligned}
 \log L(t, i) &= \sum_{j=1}^n (1 - i_j) \log \left[\frac{\alpha_0}{\beta_0} \left(\frac{t_j}{\beta_0} \right)^{\alpha_0-1} \right] + \sum_{j=1}^n i_j \log \left[\frac{\alpha_1}{\beta_1} \left(\frac{t_j}{\beta_1} \right)^{\alpha_1-1} \right] \\
 &\quad + \sum_{j=1}^n \log \left[e^{-\left(\frac{t_j}{\beta_0}\right)^{\alpha_0}} \right] + \sum_{j=1}^n \log \left[e^{-\left(\frac{t_j}{\beta_1}\right)^{\alpha_1}} \right] \\
 &= \sum_{j=1}^n (1 - i_j) \left[\log \left(\frac{\alpha_0}{\beta_0} \right) + (\alpha_0 - 1) \log \left(\frac{t_j}{\beta_0} \right) \right] \\
 &\quad + \sum_{j=1}^n i_j \left[\log \left(\frac{\alpha_1}{\beta_1} \right) + (\alpha_1 - 1) \log \left(\frac{t_j}{\beta_1} \right) \right] - \sum_{j=1}^n \left(\frac{t_j}{\beta_0} \right)^{\alpha_0} - \sum_{j=1}^n \left(\frac{t_j}{\beta_1} \right)^{\alpha_1}
 \end{aligned}$$

Les fonctions de score

Les fonctions de score pour les paramètres $\alpha_0, \alpha_1, \beta_0, \beta_1$ sont données par:

$$\frac{\partial \log L(t, i)}{\partial \beta_0} = -\frac{\alpha_0}{\beta_0} \sum_{j=1}^n (1 - i_j) + \frac{\alpha_0}{\beta_0} \sum_{j=1}^n \left(\frac{t_j}{\beta_0} \right)^{\alpha_0}$$

$$\frac{\partial \log L(t, i)}{\partial \beta_1} = -\frac{\alpha_1}{\beta_1} \sum_{j=1}^n i_j + \frac{\alpha_1}{\beta_1} \sum_{j=1}^n \left(\frac{t_j}{\beta_1} \right)^{\alpha_1}$$

$$\frac{\partial \log L(t, i)}{\partial \alpha_0} = \frac{1}{\alpha_0} \sum_{j=1}^n (1 - i_j) + \sum_{j=1}^n \log \frac{t_j}{\beta_0} \left(1 - \left(\frac{t_j}{\beta_0} \right)^{\alpha_0} \right)$$

$$\frac{\partial \log L(t, i)}{\partial \alpha_1} = \frac{1}{\alpha_1} \sum_{j=1}^n i_j + \sum_{j=1}^n \log \frac{t_j}{\beta_1} \left(1 - \left(\frac{t_j}{\beta_1} \right)^{\alpha_1} \right)$$

En annulant ces équations, on obtient les estimateurs dits estimateurs du maximum de vraisemblance $\hat{\alpha}_0, \hat{\alpha}_1, \hat{\beta}_0, \hat{\beta}_1$. On fait appel aux méthodes numériques pour les calculer.

2.2.2 La matrice d'information de Fisher

Les composantes de la matrice d'information de Fisher $\hat{I} = [i_{ss}]_{l \times l'} = -E \left(\frac{\partial^2 \log L(t, i)}{\partial \theta^2} \right)$ dérivées du second ordre de la fonction de log-vraisemblance par rapport aux paramètres sont obtenues comme suit:

$$I_{11} = \frac{\partial^2 \log L(t, i)}{\partial \alpha_0^2} = -\frac{\hat{\alpha}_0}{\hat{\beta}_0^2} \left[\sum_{j=1}^n (1 - i_j) + (\hat{\alpha}_0 + 1) \sum_{j=1}^n \left(\frac{t_j}{\hat{\beta}_0} \right)^{\hat{\alpha}_0} \right]$$

$$I_{14} = \frac{\partial^2 \log L(t, i)}{\partial \alpha_0 \partial \beta_1} = 0$$

$$I_{41} = \frac{\partial^2 \log L(t, i)}{\partial \beta_1 \partial \alpha_0} = 0$$

$$I_{22} = \frac{\partial^2 \log L(t, i)}{\partial \beta_0^2} = -\frac{\hat{\alpha}_1}{\hat{\beta}_1^2} \left[\sum_{j=1}^n i_j - (\hat{\alpha}_1 + 1) \sum_{j=1}^n \left(\frac{t_j}{\hat{\beta}_1} \right)^{\hat{\alpha}_1} \right]$$

et

$$I_{23} = \frac{\partial^2 \log L(t, i)}{\partial \beta_0 \partial \alpha_1} = 0$$

$$I_{32} = \frac{\partial^2 \log L(t, i)}{\partial \alpha_1 \partial \beta_0} = 0$$

$$I_{33} = \frac{\partial^2 \log L(t, i)}{\partial \alpha_1^2} = \frac{1}{\hat{\alpha}_0^2} \sum_{j=1}^n (1 - i_j) + \sum_{j=1}^n \left(\frac{t_j}{\hat{\beta}_0} \right)^{\hat{\alpha}_0} \left[\ln \left(\frac{t_j}{\hat{\beta}_0} \right) \right]^2$$

$$I_{44} = \frac{\partial^2 \log L(t, i)}{\partial \beta_1^2} = \frac{1}{\hat{\alpha}_1^2} \sum_{j=1}^n i_j - \sum_{j=1}^n \left(\frac{t_j}{\hat{\beta}_1} \right)^{\hat{\alpha}_1} \left[\ln \left(\frac{t_j}{\hat{\beta}_1} \right) \right]^2$$

$$I_{13} = I_{31} = \frac{\partial^2 \log L(t, i)}{\partial \alpha_0 \partial \alpha_1} = \frac{1}{\widehat{\beta}_0} \left[\sum_{j=1}^n (1 - i_j) + \sum_{j=1}^n \left(\frac{t_j}{\widehat{\beta}_0} \right)^{\widehat{\alpha}_0} \left(1 - \widehat{\alpha}_0 \ln \left(\frac{t_j}{\widehat{\beta}_0} \right) \right) \right]$$

$$I_{12} = \frac{\partial^2 \log L(t, i)}{\partial \alpha_0 \partial \beta_0} = 0$$

$$I_{21} = \frac{\partial^2 \log L(t, i)}{\partial \beta_0 \partial \alpha_0} = 0$$

$$I_{24} = I_{42} = \frac{\partial^2 \log L(t, i)}{\partial \beta_0 \partial \beta_1} = \frac{1}{\widehat{\beta}_0} \left[\sum_{j=1}^n i_j + \sum_{j=1}^n \left(\frac{t_j}{\widehat{\beta}_1} \right)^{\widehat{\alpha}_1} \left(1 - \widehat{\alpha}_1 \ln \left(\frac{t_j}{\widehat{\beta}_1} \right) \right) \right]$$

et

$$I_{34} = \frac{\partial^2 \log L(t, i)}{\partial \alpha_1 \partial \beta_1} = 0$$

$$I_{43} = \frac{\partial^2 \log L(t, i)}{\partial \beta_1 \partial \alpha_1} = 0$$

Cette matrice est indispensable au calcul des intervalles de confiance des paramètres inconnus, ainsi qu'à la réalisation du test d'ajustement proposé.

2.3 Test d'ajustement pour le modèle BW

La validation d'un modèle statistique repose sur plusieurs tests d'adéquation, tels que les statistiques d'Anderson-Darling, de Kolmogorov-Smirnov, de Cramér-von Mises et d'autres outils incontournables, notamment le test du khi-deux. Cependant, l'application de ce dernier devient problématique lorsque les paramètres du modèle sont inconnus. Pour pallier cette limitation, Nikulin (1973) a proposé une version modifiée du test du khi-deux adaptée aux distributions continues dont les paramètres de forme et d'échelle ne sont pas préalablement définis, tout en tenant compte de données non groupées. Dans la même lignée, Rao et Robson (1974) ont obtenu un résultat similaire pour les familles de

distributions exponentielles. Depuis, cette approche est désignée sous le nom de statistique NRR (Nikulin-Rao-Robson), et constitue une référence en analyse statistique pour tester l'ajustement des modèles sous des conditions plus générales. Dans cette section, on expose d'abord les tests d'ajustement EDF ensuite et basé sur l'approche de la NRR, on construit un critère de ce type pour vérifier si un ensemble d'observations est modélisé par la distribution Bi-Weibull.

2.3.1 Tests non paramétriques

Kolmogorov-Smirnov

La statistique du test de Kolmogorov-Smirnov D_n est le test d'ajustement le plus ancien et le plus utilisé. Il est défini par

$$\begin{aligned} D_n &= \text{Max} [D^+; D^-] \\ D^+ &= \text{Max}_{1 \leq i \leq n} \left[\left(\frac{i}{n} \right) - 1 + \exp \left(- \left(\left(\frac{t_i}{\beta_0} \right)^{\alpha_0} + \left(\frac{t_i}{\beta_1} \right)^{\alpha_1} \right) \right) \right] \\ D^- &= \text{Max}_{1 \leq i \leq n} \left[1 - \exp \left(- \left(\left(\frac{t_i}{\beta_0} \right)^{\alpha_0} + \left(\frac{t_i}{\beta_1} \right)^{\alpha_1} \right) \right) - \left(\frac{i-1}{n} \right) \right] \end{aligned}$$

Cramer- Von- Mises

Le deuxième test populaire est celui de Cramer-Von-Mises dont la forme est la suivante

$$\begin{aligned} W_n^2 &= \sum_{i=1}^n \left[F(t_i) - \frac{2i-1}{2n} \right]^2 + \frac{1}{12n} \\ &= \sum_{i=1}^n \left[1 - \exp \left(- \left(\left(\frac{t_i}{\beta_0} \right)^{\alpha_0} + \left(\frac{t_i}{\beta_1} \right)^{\alpha_1} \right) \right) - \frac{2i-1}{2n} \right]^2 + \frac{1}{12n} \end{aligned}$$

Anderson- Darling

Le test d'Anderson-Darling A_n^2 est une alternative aux tests de Kolmogorov-Smirnov et de Cramer-Von-mises. Il a la forme suivante

$$A_n^2 = -n - \frac{1}{n} \left[\sum_{i=1}^n (2i-1) \left[\ln \left(1 - \exp \left(- \left(\left(\frac{t_i}{\beta_0} \right)^{\alpha_0} + \left(\frac{t_i}{\beta_1} \right)^{\alpha_1} \right) \right) \right) \right] - \left(\left(\frac{t_{n+1-i}}{\beta_0} \right)^{\alpha_0} + \left(\frac{t_{n+1-i}}{\beta_1} \right)^{\alpha_1} \right) \right]$$

2.3.2 Construction de la statistique NRR pour la distribution BW

Soit T_1, \dots, T_n n variables aléatoires i.i.d., pour tester l'hypothèse nulle H_0 telle que :

$$P(T_j \leq t \mid H_0) = F_{BW}(t; \theta), t \geq 0, \quad \theta = (\alpha_0, \alpha_1, \beta_0, \beta_1)^T,$$

où $F_{BW}(t; \theta)$ est la fonction de répartition de la distribution de Bi-Weibull et $f_{BW}(t; \theta)$ sa densité de probabilité; nous allons adapter la statistique de Nikulin-Rao-Robson, basée sur les estimateurs du maximum de vraisemblance sur les données initiales non groupées.

Pour cela, on partage l'ensemble \mathbb{R}^+ en r intervalles I_1, I_2, \dots, I_r mutuellement disjoints, par les points :

$$0 = K_0 < K_1 < \dots < K_{r-1} < K_r = +\infty$$

où

$$I_j =]K_{j-1}, K_j]; \quad I_i \cap I_j = \emptyset \quad i \neq j; \quad \bigcup_{j=1}^r I_j = \mathbb{R}^+.$$

Soit v le vecteur de fréquence $v = (v_1, v_2, \dots, v_r)^T$ obtenu comme résultat du regroupement des T_i dans les intervalles I_j .

Dans le cas d'équibrosabilité (i.e. $P_1 = P_2 = \dots = P_r$), les K_j sont définies telles que:

$$\begin{aligned} K_j &= F_{BW}^{-1}(P_1 + P_2 + \dots + P_j) = F_{BW}^{-1}\left(\frac{j}{r}\right); \quad j = 1, \dots, r-1. \\ &= \left(-\left(\frac{t}{\beta_0}\right)^{\alpha_0} - \left(\frac{t}{\beta_1}\right)^{\alpha_1} \right) - \log\left(1 - \frac{j}{r}\right) \end{aligned}$$

Soit: $P(\theta) = (P_1(\theta), P_2(\theta), \dots, P_r(\theta))^T$

$$P_j(\theta) = \int_{I_j} f_{BW}(t; \theta) dt = F_{BW}(K_{j-1}) - F_{BW}(K_j), \quad j = 1, \dots, r.$$

Soit le vecteur $X_n(\theta)$:

$$X_n(\theta) = \left(\frac{v_1 - np_1(\theta)}{\sqrt{np_1(\theta)}}, \frac{v_2 - np_2(\theta)}{\sqrt{np_2(\theta)}}, \dots, \frac{v_r - np_r(\theta)}{\sqrt{np_r(\theta)}} \right)^T$$

Pour calculer la statistique NRR, on utilise souvent la formule quadratique suivante:

$$Y_n^2(\hat{\theta}_{BW}) = X_n^2(\hat{\theta}_{BW}) + \frac{1}{n} L^T(\hat{\theta}_{BW}) \left(I(\hat{\theta}_{BW}) - J(\hat{\theta}_{BW}) \right)^{-1} L(\hat{\theta}_{BW})$$

Où $\hat{\theta}_{BW}$ est l'estimateur du maximum de vraisemblance de θ .

Pour la distribution de Bi-Weibull, on obtient

$$P_j(\hat{\theta}) = -\exp\left(-\left(\frac{K_j}{\hat{\beta}_0}\right)^{\hat{\alpha}_0} - \left(\frac{K_j}{\hat{\beta}_1}\right)^{\hat{\alpha}_1}\right) + \exp\left(-\left(\frac{K_{j-1}}{\hat{\beta}_0}\right)^{\hat{\alpha}_0} - \left(\frac{K_{j-1}}{\hat{\beta}_1}\right)^{\hat{\alpha}_1}\right)$$

les dérivées partielles de $P_j(\theta)$ s'obtiennent comme suit :

$$\begin{aligned} \frac{\partial P_j(\hat{\theta})}{\partial \hat{\beta}_0} &= -\left(\frac{\hat{\alpha}_0}{\hat{\beta}_0}\right) \left(\frac{K_j}{\hat{\beta}_0}\right)^{\hat{\alpha}_0} \exp\left(-\left(\frac{K_j}{\hat{\beta}_0}\right)^{\hat{\alpha}_0} - \left(\frac{K_j}{\hat{\beta}_1}\right)^{\hat{\alpha}_1}\right) \\ &\quad + \left(\frac{\hat{\alpha}_0}{\hat{\beta}_0}\right) \left(\frac{K_{j-1}}{\hat{\beta}_0}\right)^{\hat{\alpha}_0} \exp\left(-\left(\frac{K_{j-1}}{\hat{\beta}_0}\right)^{\hat{\alpha}_0} - \left(\frac{K_{j-1}}{\hat{\beta}_1}\right)^{\hat{\alpha}_1}\right) \end{aligned}$$

$$\begin{aligned} \frac{\partial P_j(\hat{\theta})}{\partial \hat{\beta}_1} &= -\left(\frac{\hat{\alpha}_1}{\hat{\beta}_1}\right) \left(\frac{K_j}{\hat{\beta}_1}\right)^{\hat{\alpha}_1} \exp\left(-\left(\frac{K_j}{\hat{\beta}_0}\right)^{\hat{\alpha}_0} - \left(\frac{K_j}{\hat{\beta}_1}\right)^{\hat{\alpha}_1}\right) \\ &\quad + \left(\frac{\hat{\alpha}_1}{\hat{\beta}_1}\right) \left(\frac{K_{j-1}}{\hat{\beta}_1}\right)^{\hat{\alpha}_1} \exp\left(-\left(\frac{K_{j-1}}{\hat{\beta}_0}\right)^{\hat{\alpha}_0} - \left(\frac{K_{j-1}}{\hat{\beta}_1}\right)^{\hat{\alpha}_1}\right) \end{aligned}$$

$$\begin{aligned} \frac{\partial P_j(\hat{\theta})}{\partial \hat{\alpha}_0} &= \left(\frac{K_j}{\hat{\beta}_0}\right)^{\hat{\alpha}_0} \ln\left(\frac{K_j}{\hat{\beta}_0}\right) \exp\left(-\left(\frac{K_j}{\hat{\beta}_0}\right)^{\hat{\alpha}_0} - \left(\frac{K_j}{\hat{\beta}_1}\right)^{\hat{\alpha}_1}\right) \\ &\quad - \left(\frac{K_{j-1}}{\hat{\beta}_0}\right)^{\hat{\alpha}_0} \ln\left(\frac{K_{j-1}}{\hat{\beta}_0}\right) \exp\left(-\left(\frac{K_{j-1}}{\hat{\beta}_0}\right)^{\hat{\alpha}_0} - \left(\frac{K_{j-1}}{\hat{\beta}_1}\right)^{\hat{\alpha}_1}\right) \end{aligned}$$

$$\begin{aligned} \frac{\partial P_j(\hat{\theta})}{\partial \hat{\alpha}_1} &= \left(\frac{K_j}{\hat{\beta}_1}\right)^{\hat{\alpha}_1} \ln\left(\frac{K_j}{\hat{\beta}_1}\right) \exp\left(-\left(\frac{K_j}{\hat{\beta}_0}\right)^{\hat{\alpha}_0} - \left(\frac{K_j}{\hat{\beta}_1}\right)^{\hat{\alpha}_1}\right) \\ &\quad - \left(\frac{K_{j-1}}{\hat{\beta}_1}\right)^{\hat{\alpha}_1} \ln\left(\frac{K_{j-1}}{\hat{\beta}_1}\right) \exp\left(-\left(\frac{K_{j-1}}{\hat{\beta}_0}\right)^{\hat{\alpha}_0} - \left(\frac{K_{j-1}}{\hat{\beta}_1}\right)^{\hat{\alpha}_1}\right) \end{aligned}$$

On doit d'abord estimer les paramètres par la méthode du maximum de vraisemblance, puis calculer la matrice d'information de Fisher pour les données non groupées $I = \left(\widehat{i}_{ll'} \right)_{4 \times 4}$ (voir sous-section 2.2.2) et enfin la matrice d'information pour des données groupées $J(\theta)$, où

$$J(\theta) = B(\theta)^T \cdot B(\theta), \quad B(\theta) = [b_{ll'}]_{4 \times 4}$$

$$b_{j1}(\theta) = \frac{1}{\sqrt{P_j}} \cdot \frac{\partial P_j(\widehat{\theta})}{\partial \widehat{\alpha}_0}; \quad b_{j2}(\theta) = \frac{1}{\sqrt{P_j}} \cdot \frac{\partial P_j(\widehat{\theta})}{\partial \widehat{\alpha}_1};$$

$$b_{j3}(\theta) = \frac{1}{\sqrt{P_j}} \cdot \frac{\partial P_j(\widehat{\theta})}{\partial \widehat{\beta}_0}; \quad ; b_{j4}(\theta) = \frac{1}{\sqrt{P_j}} \cdot \frac{\partial P_j(\widehat{\theta})}{\partial \widehat{\beta}_1}$$

$$j = 1, \dots, r$$

dont les valeurs sont obtenues par :

$$b_{j1}(\theta) = \frac{1}{\sqrt{P_j}} \cdot \left[\left(\left(\frac{K_j}{\widehat{\beta}_0} \right)^{\widehat{\alpha}_0} \ln \left(\frac{K_j}{\widehat{\beta}_0} \right) \exp \left(- \left(\frac{K_j}{\widehat{\beta}_0} \right)^{\widehat{\alpha}_0} - \left(\frac{K_j}{\widehat{\beta}_1} \right)^{\widehat{\alpha}_1} \right) \right) \right]$$

$$- \frac{1}{\sqrt{P_j}} \cdot \left[\left(\left(\frac{K_{j-1}}{\widehat{\beta}_0} \right)^{\widehat{\alpha}_0} \ln \left(\frac{K_{j-1}}{\widehat{\beta}_0} \right) \exp \left(- \left(\frac{K_{j-1}}{\widehat{\beta}_0} \right)^{\widehat{\alpha}_0} - \left(\frac{K_{j-1}}{\widehat{\beta}_1} \right)^{\widehat{\alpha}_1} \right) \right) \right]$$

$$b_{j2}(\theta) = \frac{1}{\sqrt{P_j}} \cdot \left[\left(\frac{K_j}{\widehat{\beta}_1} \right)^{\widehat{\alpha}_1} \ln \left(\frac{K_j}{\widehat{\beta}_1} \right) \exp \left(- \left(\frac{K_j}{\widehat{\beta}_0} \right)^{\widehat{\alpha}_0} - \left(\frac{K_j}{\widehat{\beta}_1} \right)^{\widehat{\alpha}_1} \right) \right]$$

$$- \frac{1}{\sqrt{P_j}} \cdot \left[\left(\frac{K_{j-1}}{\widehat{\beta}_1} \right)^{\widehat{\alpha}_1} \ln \left(\frac{K_{j-1}}{\widehat{\beta}_1} \right) \exp \left(- \left(\frac{K_{j-1}}{\widehat{\beta}_0} \right)^{\widehat{\alpha}_0} - \left(\frac{K_{j-1}}{\widehat{\beta}_1} \right)^{\widehat{\alpha}_1} \right) \right]$$

$$b_{j3}(\theta) = \frac{1}{\sqrt{P_j}} \cdot \left[- \left(\frac{\widehat{\alpha}_0}{\widehat{\beta}_0} \right) \left(\frac{K_j}{\widehat{\beta}_0} \right)^{\widehat{\alpha}_0} \exp \left(- \left(\frac{K_j}{\widehat{\beta}_0} \right)^{\widehat{\alpha}_0} - \left(\frac{K_j}{\widehat{\beta}_1} \right)^{\widehat{\alpha}_1} \right) \right]$$

$$+ \frac{1}{\sqrt{P_j}} \cdot \left[\left(\frac{\widehat{\alpha}_0}{\widehat{\beta}_0} \right) \left(\frac{K_{j-1}}{\widehat{\beta}_0} \right)^{\widehat{\alpha}_0} \exp \left(- \left(\frac{K_{j-1}}{\widehat{\beta}_0} \right)^{\widehat{\alpha}_0} - \left(\frac{K_{j-1}}{\widehat{\beta}_1} \right)^{\widehat{\alpha}_1} \right) \right]$$

$$\begin{aligned}
 b_{j4}(\theta) &= \frac{1}{\sqrt{P_j}} \left[- \left(\frac{\hat{\alpha}_1}{\hat{\beta}_1} \right) \left(\frac{K_j}{\hat{\beta}_1} \right)^{\hat{\alpha}_1} \exp \left(- \left(\frac{K_j}{\hat{\beta}_0} \right)^{\hat{\alpha}_0} - \left(\frac{K_j}{\hat{\beta}_1} \right)^{\hat{\alpha}_1} \right) \right] \\
 &\quad + \frac{1}{\sqrt{P_j}} \left[\left(\frac{\hat{\alpha}_1}{\hat{\beta}_1} \right) \left(\frac{K_{j-1}}{\hat{\beta}_1} \right)^{\hat{\alpha}_1} \exp \left(- \left(\frac{K_{j-1}}{\hat{\beta}_0} \right)^{\hat{\alpha}_0} - \left(\frac{K_{j-1}}{\hat{\beta}_1} \right)^{\hat{\alpha}_1} \right) \right]
 \end{aligned}$$

les éléments du vecteur L sont calculés comme suit :

$$L(\theta) = (L_1, L_2, L_3, L_4)^T$$

$$\begin{aligned}
 L_1(\theta) &= \sum_{j=1}^r \frac{v_j}{P_j} \frac{\partial P_j(\hat{\theta})}{\partial \hat{\alpha}_0} \\
 &= \sum_{j=1}^r \frac{v_j}{P_j} \left[\left(\frac{K_j}{\hat{\beta}_0} \right)^{\hat{\alpha}_0} \ln \left(\frac{K_j}{\hat{\beta}_0} \right) \exp \left(- \left(\frac{K_j}{\hat{\beta}_0} \right)^{\hat{\alpha}_0} - \left(\frac{K_j}{\hat{\beta}_1} \right)^{\hat{\alpha}_1} \right) \right] \\
 &\quad - \sum_{j=1}^r \frac{v_j}{P_j} \left[\left(\frac{K_{j-1}}{\hat{\beta}_0} \right)^{\hat{\alpha}_0} \ln \left(\frac{K_{j-1}}{\hat{\beta}_0} \right) \exp \left(- \left(\frac{K_{j-1}}{\hat{\beta}_0} \right)^{\hat{\alpha}_0} - \left(\frac{K_{j-1}}{\hat{\beta}_1} \right)^{\hat{\alpha}_1} \right) \right]
 \end{aligned}$$

$$\begin{aligned}
 L_2(\theta) &= \sum_{j=1}^r \frac{v_j}{P_j} \frac{\partial P_j(\hat{\theta})}{\partial \hat{\alpha}_1} \\
 &= \sum_{j=1}^r \frac{v_j}{P_j} \left[\left(\frac{K_j}{\hat{\beta}_1} \right)^{\hat{\alpha}_1} \ln \left(\frac{K_j}{\hat{\beta}_1} \right) \exp \left(- \left(\frac{K_j}{\hat{\beta}_0} \right)^{\hat{\alpha}_0} - \left(\frac{K_j}{\hat{\beta}_1} \right)^{\hat{\alpha}_1} \right) \right] \\
 &\quad - \sum_{j=1}^r \frac{v_j}{P_j} \left[\left(\frac{K_{j-1}}{\hat{\beta}_1} \right)^{\hat{\alpha}_1} \ln \left(\frac{K_{j-1}}{\hat{\beta}_1} \right) \exp \left(- \left(\frac{K_{j-1}}{\hat{\beta}_0} \right)^{\hat{\alpha}_0} - \left(\frac{K_{j-1}}{\hat{\beta}_1} \right)^{\hat{\alpha}_1} \right) \right]
 \end{aligned}$$

$$\begin{aligned}
 L_3(\theta) &= \sum_{i=1}^r \frac{v_j}{P_j} \frac{\partial P_j(\hat{\theta})}{\partial \hat{\beta}_0} \\
 &= \sum_{i=1}^r \frac{v_j}{P_j} \left[- \left(\frac{\hat{\alpha}_0}{\hat{\beta}_0} \right) \left(\frac{K_j}{\hat{\beta}_0} \right)^{\hat{\alpha}_0} \exp \left(- \left(\frac{K_j}{\hat{\beta}_0} \right)^{\hat{\alpha}_0} - \left(\frac{K_j}{\hat{\beta}_1} \right)^{\hat{\alpha}_1} \right) \right] \\
 &\quad + \sum_{i=1}^r \frac{v_j}{P_j} \left[\left(\frac{\hat{\alpha}_0}{\hat{\beta}_0} \right) \left(\frac{K_{j-1}}{\hat{\beta}_0} \right)^{\hat{\alpha}_0} \exp \left(- \left(\frac{K_{j-1}}{\hat{\beta}_0} \right)^{\hat{\alpha}_0} - \left(\frac{K_{j-1}}{\hat{\beta}_1} \right)^{\hat{\alpha}_1} \right) \right]
 \end{aligned}$$

et

$$\begin{aligned}
 L_4(\theta) &= \sum_{j=1}^r \frac{v_j}{P_j} \frac{\partial P_j(\hat{\theta})}{\partial \hat{\beta}_1} \\
 &= \sum_{j=1}^r \frac{v_j}{P_j} \left[- \left(\frac{\hat{\alpha}_1}{\hat{\beta}_1} \right) \left(\frac{K_j}{\hat{\beta}_1} \right)^{\hat{\alpha}_1} \exp \left(- \left(\frac{K_j}{\hat{\beta}_0} \right)^{\hat{\alpha}_0} - \left(\frac{K_j}{\hat{\beta}_1} \right)^{\hat{\alpha}_1} \right) \right] \\
 &\quad + \sum_{j=1}^r \frac{v_j}{P_j} \left[\left(\frac{\hat{\alpha}_1}{\hat{\beta}_1} \right) \left(\frac{K_{j-1}}{\hat{\beta}_1} \right)^{\hat{\alpha}_1} \exp \left(- \left(\frac{K_{j-1}}{\hat{\beta}_0} \right)^{\hat{\alpha}_0} - \left(\frac{K_{j-1}}{\hat{\beta}_1} \right)^{\hat{\alpha}_1} \right) \right]
 \end{aligned}$$

Ainsi toutes les composantes de la statistique Y^2 sont calculées, on en déduit alors la valeur Y^2 pour la comparer à une distribution du khi-deux à $(r - 1)$ degrés de liberté, où r est le nombre de classe choisi.

2.4 Simulations et Applications réelles

Nous avons mené une importante étude par simulations numériques pour montrer la maniabilité et l'efficacité des tests proposés dans ce travail. Ensuite, nous avons appliqué ces tests à des données réelles issues de la fiabilité et de l'analyse de survie.

2.4.1 Simulations

Estimation des paramètres du modèle BW

Nous avons considéré un modèle de risques concurrents Bi-Weibull. Nous avons effectué 10000 simulations, avec des tailles d'échantillon $n = 15, 50, 100$ et 200 , avec les valeurs des paramètres sont $\beta_0 = 0.5$, $\beta_1 = 1$, $\alpha_0 = 1.5$ et $\alpha_1 = 2$.

En utilisant le logiciel R, nous avons calculé les estimateurs des paramètres inconnus par différentes méthodes, ainsi que leurs erreurs quadratiques moyennes (SME), les résultats sont résumés dans les tableaux 1, 2, 3, 4 respectivement pour les méthodes maximum vraisemblance; Kolmogorov-Smirnov; Cramer-Von-Mises et Anderson-Darling :

$M = 10000$		$\hat{\alpha}_0$	$\hat{\beta}_0$	$\hat{\alpha}_1$	$\hat{\beta}_1$
$n = 15$	Mean	1.42172	0.32922	1.80297	0.70655
	SME	0.01032	0.08924	0.03489	0.09123
$n = 50$	Mean	1.48036	0.45988	1.84011	0.76410
	SME	0.00343	0.02263	0.02263	0.03811
$n = 100$	Mean	1.49068	0.47950	1.92406	0.89660
	SME	0.00244	0.00850	0.00819	0.03083
$n = 200$	Mean	1.50730	0.50945	1.98636	0.98772
	SME	0.00039	0.00436	0.00435	0.00955

Tableau 1: ML estimateurs et leurs erreurs quadratiques moyennes

$M = 10000$		$\hat{\alpha}_0$	$\hat{\beta}_0$	$\hat{\alpha}_1$	$\hat{\beta}_1$
$n = 15$	Mean	1.42135	0.48970	1.87589	0.86034
	SME	0.00630	0.07554	0.03280	0.07637
$n = 50$	Mean	1.45627	0.33257	1.80787	0.89540
	SME	0.00720	0.03272	0.04266	0.04848
$n = 100$	Mean	1.47742	0.54131	1.93220	0.90235
	SME	0.00381	0.00960	0.02312	0.03244
$n = 200$	Mean	1.50325	0.53493	1.97310	0.93285
	SME	0.00107	0.00401	0.00687	0.00982

Tableau 2: Estimateurs KS des paramètres et leurs erreurs quadratiques moyennes

$M = 10000$		$\hat{\alpha}_0$	$\hat{\beta}_0$	$\hat{\alpha}_1$	$\hat{\beta}_1$
$n = 15$	Mean	1.43752	0.38650	1.66156	0.86928
	SME	0.00489	0.02367	0.03638	0.02823
$n = 50$	Mean	11.39721	0.53427	1.77915	0.84509
	SME	0.00376	0.02239	0.03067	0.02573
$n = 100$	Mean	1.42639	0.42487	1.87606	0.87307
	SME	0.00226	0.00650	0.02165	0.02035
$n = 200$	Mean	1.47239	0.44975	1.90849	0.98152
	SME	0.00050	0.00323	0.00663	0.00508

Tableau 3: Estimateurs CVM des paramètres et leurs erreurs quadratiques moyennes

$M = 10000$		$\hat{\alpha}_0$	$\hat{\beta}_0$	$\hat{\alpha}_1$	$\hat{\beta}_1$
$n = 15$	Mean	1.45930	0.35685	1.92943	0.81630
	SME	0.03238	0.08264	0.06279	0.07526
$n = 50$	Mean	1.42024	0.42163	1.82683	0.89053
	SME	0.00751	0.03485	0.03400	0.05339
$n = 100$	Mean	1.49315	0.42759	1.96315	0.93049
	SME	0.00408	0.00897	0.05421	0.02738
$n = 200$	Mean	1.49511	0.49382	1.98239	0.97203
	SME	0.00099	0.00591	0.00964	0.00866

Tableau 4: Estimateurs AD des paramètres et leurs erreurs quadratiques moyennes

D'après les expériences numériques, on observe que les valeurs des estimateurs sont proches des vraies valeurs des paramètres. De plus, leurs erreurs quadratiques moyennes (SME) diminuent à mesure que la taille de l'échantillon augmente, et ce, pour toutes les méthodes.

Cela montre que l'augmentation de la taille de l'échantillon améliore la précision des estimations, réduisant ainsi l'erreur d'estimation et renforçant la fiabilité des résultats obtenus.

Critères de Test

Pour démontrer la praticabilité du test statistique proposé dans cette étude, nous vérifions l'hypothèse nulle H_0 selon laquelle les échantillons sont issus de la distribution bi-Weibull.

Le nombre de rejets de l'hypothèse nulle H_0 est calculé pour des milliers d'échantillons de différentes tailles ($n = 15, 50, 100, 200$) et comparé aux valeurs théoriques correspondantes pour les seuils de signification ($\xi = 0.01, \xi = 0.1$ et $\xi = 0.05$)

Cette approche permet d'évaluer la validité et la performance du test statistique en vérifiant sa capacité à détecter les écarts par rapport à la distribution supposée.

n	$\xi = 0.01$	$\xi = 0.1$	$\xi = 0.05$
15	0.01697	0.1230	0.03879
50	0.01513	0.1080	0.04918
100	0.01222	0.1022	0.05204
200	0.01124	0.1004	0.05073

Tableau 5: Valeurs critiques du khi-deux et leurs valeurs empiriques correspondantes

D'après le tableau 5, les valeurs des niveaux de signification empiriques de Y^2 sont relativement proches des niveaux théoriques correspondants de la distribution du Khi-deux χ_{r-1}^2 . Cela nous permet de conclure que le test statistique Y^2 développé dans cette étude est bien adapté pour ajuster la distribution BW.

2.4.2 Applications

Pour illustrer l'applicabilité et la pertinence de cette nouvelle distribution Bi-Weibull, nous analyserons des ensembles de données réelles, issus de divers domaines d'application

Données médicales

Dans cette section, nous avons appliqué les résultats obtenus à un ensemble de données réelles prises sur le site: <https://www.kaggle.com/datasets/marshalpatel3558/diabetes-prediction-dataset-legit-dataset>. L'ensemble de données semble lié au diabète et

contient diverses mesures biomédicales et caractéristiques des patients. Dans le cadre de nos travaux, nous avons sélectionné les données suivantes : taux de cholestérol sanguin (probablement mesuré en mg/dL ou mmol/L). Cela fait généralement référence au cholestérol total (Chol(4,2, 3,7, 4,2, 4,2, 0,9, 2,9, 3,6, 2,9, 3,8, 3,8, 3,6, 4,0, 4,9, 4,2, 4,0, 3,6, 5,3, 4,9, 0,5, 4,4, 0,5, 6,2, 4,2, 4,8, 4,6, 4,9, 0,8, 2,8, 3,7, 5,6, 3,2, 4,6, 3,6, 0,2, 6,5, 4,0, 3,0, 3,7, 3,7, 4,4, 3,8, 9,5, 3,7, 4,7, 9,5, 0,7, 4,2, 5,2, 0,3, 4,7)) et le taux de triglycérides dans le sang (probablement mesuré en mg/dL ou mmol/L). Les triglycérides sont un type de graisse dans le sang (TG(0,9, 1,4, 0,9, 0,9, 1,0, 1,0, 1,3, 0,8, 0,9, 2,0, 0,7, 1,1, 1,3, 1,7, 1,5, 1,1,0,8, 1,3, 1,3, 0,9, 1,9, 1,0, 1,5, 1,0, 1,3, 1,3, 1,8, 2,9, 1,3, 1,4, 1,8, 0,8,0,7, 1,0, 1,5, 0,6, 1,2, 1,3, 1,5, 1,6, 0,7, 1,7, 1,3, 1,9, 1,4, 1,1, 2,1, 1,8, 1,7, 1,6)).

On suppose que ces données sont modélisées par la distribution Bi-Weibull et on considère le modèle à risques concurrents de Bertholon comme alternative.

$$Betholon \quad f(t/\lambda, \eta, \beta) = \lambda \exp(-\lambda t) + \frac{\beta}{\eta} \left(\frac{t-T}{\eta}\right)^{\beta-1} \exp\left(-\left(\frac{t-T}{\eta}\right)^\beta\right) \quad \beta = 2, T = 0.5$$

Dans plusieurs études statistiques et applications en fiabilité des systèmes, $\beta = 2$ est utilisé pour modéliser des phénomènes où le risque augmente avec le temps.

Critères classiques de sélection de modèles

Model	LL	AIC	BIC	K_S - pvalue
Bi - Weibull	-29.90045	63.80091	67.62495	0.712648
Bertholon	-90.68764	18537528	189.1993	0.035918

Tableau 6: Comparaison des critères de sélection des modèles Bi-Weibull et Bertholon

D'après ces résultats, la distribution Bi-Weibull est clairement meilleure pour modéliser les données:

- La log-vraisemblance (LL) comparée à Bertholon indique une meilleure adéquation aux données.

- Les valeurs AIC et BIC plus faibles pour Bi-Weibull, comparées à Betholon, signifient que Bi-Weibull minimise la perte d'information et évite les paramètres inutiles.

-Test de Kolmogorov-Smirnov (K-S) : une valeur plus élevée pour Bi-Weibull (0,7126) signifie que Bi-Weibull s'adapte mieux aux données que Betholon (0,0359, très faible).

Test de Nukilun-Rao-Robson

D'autre part, nous supposons l'hypothèse nulle que l'ensemble de données I appartient à la distribution BW. Si nous choisissons $r = 8$ classes de regroupement, les calculs intermédiaires du test Y^2 sont obtenues comme suit :

$$L(0.001849, -0.0053089, 0.00293371, 0.0013059)$$

$$I_n(\hat{\theta}) = \begin{pmatrix} 1.66416 & 0 & 0.3747132 & 0 \\ 0 & 2.53475 & 0 & 0.88591 \\ 0.3747132 & 0 & 0.267186 & 0 \\ 0 & 0.88591 & 0 & 0.103789 \end{pmatrix}$$

$$J(\hat{\theta}) = \begin{pmatrix} 0.722107 & 0.735629 & 0.9193473 & 0.203987 \\ 0.735629 & 0.058435 & 0.958696 & 0.283743 \\ 0.9193473 & 0.958696 & 0.5153546 & 0.305858 \\ 0.203987 & 0.283743 & 0.305858 & 0.879862 \end{pmatrix}$$

Nous obtenons $Y^2 = 3.8935$, donc pour le niveau de signification $\xi = 0.05$, la valeur critique du Khi-deux est $X_{r-1}^2 = 14,06714$. Comme $Y^2 < X_{r-1}^2$, l'hypothèse nulle H_0 ne peut pas être rejetée, on en déduit que la distribution Bi-Weibull est la mieux adaptée pour modéliser cet ensemble de données sur le diabète.

Données d'assurance

Dans cet exemple, nous avons considéré à un ensemble de données réelles du site : <https://www.kaggle.com/datasets/devansodariya/road-accident-united-kingdom-uk-dataset>. Le gouvernement britannique a collecté des données sur la circulation entre 2000 et 2018, enregistrant plus de 1,8 million d'accidents, ce qui en fait l'un des ensembles de données les plus complets du marché. C'est un aperçu complet d'un pays en pleine mutation.

(nombre de victimes

(1,1,1,1,1,1,2,2,5,1,1,1,1,2,1,1,1,1,1,1,1,1,1,1,1,2,1,1,2,1,1,2,1,1,1,1,1,1,1,1,1,1,1,1,2,1))

et le nombre de véhicules

(1,1,2,1,1,2,2,1,2,2,1,2,1,2,1,1,2,2,2,1,1,2,2,2,2,1,4,2,2,2,2,2,2,2,2,2,2,2,2,1,2,1,2,2,2,2,1,2,2)).

On suppose l'hypothèse nulle que cet ensemble de données suit la distribution BW. Si nous choisissons $r = 5$ classes de regroupement, les calculs intermédiaires du test Y^2 sont:

$$L(-1.216817; 0.361550; 1.886662; 0.751409)$$

$$I_n(\hat{\theta}) = \begin{pmatrix} 0.03767 & 0 & -0.29617 & 0 \\ 0 & 2.72875 & 0 & -1.22389 \\ -0.29617 & 0 & 1.60026 & 0 \\ 0 & -1.22389 & 0 & 0.06067 \end{pmatrix}$$

$$J(\hat{\theta}) = \begin{pmatrix} 0.03767 & 1.180405 & -0.296178 & 0.039840 \\ 1.180405 & 2.728752 & 0.028484 & -1.22389 \\ -0.296178 & -0.028484 & -1.600264 & -1.332989 \\ 0.039804 & -1.22389 & -1.332989 & 0.0606761 \end{pmatrix}$$

Nous obtenons $Y^2 = 4.3910$, donc pour le niveau de signification $\xi = 0.05$, la valeur critique du khi-deux est $X_{r-1}^2 = 11.0705$. H_0 ne peut pas être rejetée, d'où l'on peut dire que la distribution Bi-Weibull modélise convenablement cet ensemble de données sur l'assurance.

Conclusion

L'ensemble des résultats obtenus démontre que la distribution Bi-Weibull constitue un outil statistique puissant pour la modélisation des phénomènes d'assurance, de survie et de fiabilité. Sa flexibilité lui permet de s'adapter à divers contextes.

Le modèle alpha-power transformé de l'extension exponentielle (APTEE)

Dans l'analyse des phénomènes de durée de vie et de fiabilité, il est essentiel de disposer de modèles statistiques adaptés aux diverses formes de risques. La distribution Alpha-Power Transformed Extended Exponential (APTEE) (el Ghamedi et al.2019) constitue une avancée significative dans la modélisation des systèmes à risques concurrents, offrant une flexibilité accrue pour représenter des données réelles. Ce modèle peut être considéré comme une mixture de deux distributions.

La distribution APTEE, introduite comme une extension de la distribution extension exponentielle de Nadarajah et Haghghi (2011), intègre un paramètre de transformation de puissance Alpha qui lui permet de modéliser divers types de courbes de taux de hasard.

L'intérêt de la distribution APTEE réside dans sa capacité à englober plusieurs distributions connues, notamment les distributions exponentielle, extension exponentielle et exponentielle transformée par puissance (Mahdavi et Kundu, 2017). Cela permet d'améliorer les ajustements statistiques et d'offrir une modélisation plus réaliste des phénomènes à risques concurrents.

Son utilisation dans des applications pratiques, allant de la fiabilité industrielle à la biomédecine, démontre sa pertinence dans l'analyse des défaillances complexes où une simple distribution exponentielle ne suffit pas. Les résultats obtenus, notamment à partir de simulations et d'applications sur des données réelles, mettent en évidence la supériorité de ce modèle pour représenter des processus de dégradation et d'échec dans plusieurs domaines scientifiques et techniques.

Ce chapitre explorera en détails les propriétés mathématiques et statistiques de la distribution APTEE, les méthodes d'estimation, les tests d'ajustement, ainsi que ses applications réelles .

3.1 Présentation du modèle

Nous présentons le modèle APTEE qui peut être spécifié par la fonction de répartition (CDF) et de densité de probabilité suivantes (PDF) :

$$F(t) = \frac{\alpha^{1-\exp(1-(1+\lambda t)^\beta)} - 1}{\alpha - 1}, \quad x > 0$$

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

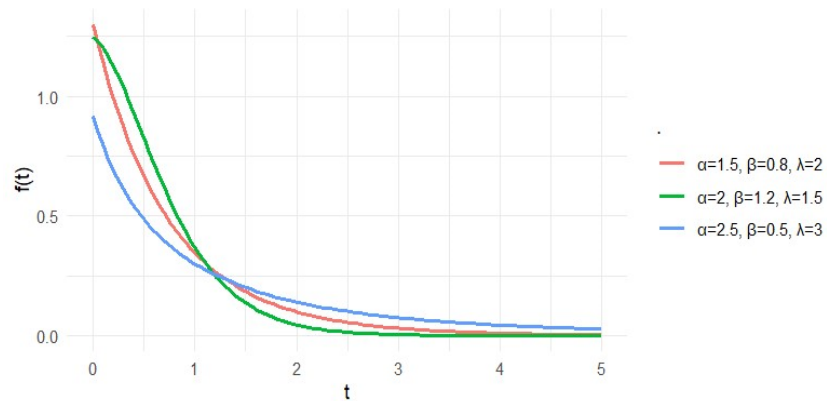


Figure 3.1: Graphe de la fonction de la densité du modèle APTEE

où $\lambda > 0$ est un paramètre d'échelle et $\alpha > 0, \alpha \neq 1$ et $\beta > 0$ sont des paramètres de forme. En posant $\beta = 1$, l'APTEE se réduit à la distribution exponentielle de puissance

alpha (Mahdavi et Kundu, 2017), la distribution exponentielle exponentielle (Gupta et Kundi 2001) découle de la distribution APTE avec $\alpha = 1$, tandis que l'exponentielle est déduite pour $\alpha = 1$ et $\beta = 1$.

La fonction de survie et la fonction de taux de risque de la distribution APTEE ont les formes

$$h(t) = \frac{\beta\lambda \log(\alpha) (1 + \lambda t)^{\beta-1} \exp(1 - (1 + \lambda t)^\beta) \alpha^{(1-(1+\lambda t)^\beta)}}{\alpha^{(1-(1+\lambda t)^\beta)} - 1}$$

$$S(t) = \frac{\alpha - \alpha^{1-\exp(1-(1+\lambda t)^\beta)}}{\alpha - 1}$$

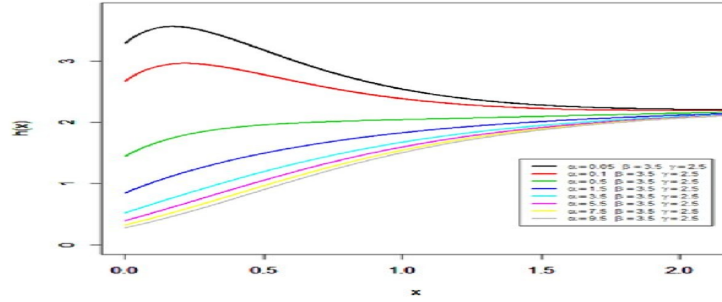


Figure 3.2: Graphe de la fonction du taux de hasard du modèle APTEE

La fonction quantile de la distribution APTEE est obtenue par :

$$Q(p) = \left[\log \left(1 - \left[\frac{\log(\alpha - 1)p + 1}{\log(\alpha)} \right]^{1/\beta} \right) \right]^{-1/\lambda}, \left\{ \begin{array}{l} 0 < p < 1 \\ \text{Let } p \sim \text{unif}(0, 1) \end{array} \right\}$$

les premier, deuxième et troisième quartiles de la distribution APTEE sont obtenus en fixant $p = 0, 25; 0, 5; 0, 75$ respectivement.

La représentation en mélange de la distribution APTEE, selon laquelle la variable aléatoire suit la loi $f(T_1)$ avec une probabilité $(\frac{\log \alpha}{\alpha - 1})$ et la loi $f(T_2)$ avec la probabilité $(1 - \frac{\log \alpha}{\alpha - 1})$, offre une interprétation naturelle en tant que modèle à risques concurrents. En effet, chaque composante du mélange peut être associée à un mécanisme de défaillance

distinct, permettant ainsi de modéliser des phénomènes dans lesquels plusieurs causes peuvent conduire à l'événement d'intérêt.

où T_1 et T_2 ont les PDF suivants :

$$f(T_1) = \beta\lambda(1 + \lambda t)^{\beta-1} \exp\left\{1 - (1 + \lambda t)^\beta\right\}$$

$$f(T_2) = \frac{\log \alpha}{(\alpha - 1 - \log \alpha)} \beta\lambda(1 + \lambda t)^{\beta-1} \exp\left\{1 - (1 + \lambda t)^\beta\right\} \left[\alpha^{1 - \exp(1 - (1 + \lambda t)^\beta)} - 1\right]$$

respectivement. Il est clair que lorsque α approche 1, T se comporte comme une distribution extension exponentielle, et lorsque α augmente, il se comporte comme T_2 .

3.2 Estimation

Dans cette section, nous déterminons les estimateurs des paramètres inconnus du modèle APTEE en appliquant la méthode du maximum de vraisemblance, qui permet d'obtenir les valeurs optimales des paramètres en maximisant la probabilité des observations disponibles.

3.2.1 Estimation du maximum vraisemblance

Dans cette section, nous abordons l'estimation des paramètres de la distribution APTEE et obtenons un estimateur du maximum de vraisemblance.

Soit T un échantillon aléatoire de taille n issu de la distribution APTEE, la fonction de la log-vraisemblance se réduit alors à :

$$l(\alpha, \beta, \lambda | T) = \prod_{i=1}^n f_T(t_i) = \prod_{i=1}^n \frac{\log \alpha}{\alpha - 1} \beta\lambda(1 + \lambda t_i)^{\beta-1} \exp\left(1 - (1 + \lambda t_i)^\beta\right) \alpha^{1 - \exp(1 - (1 + \lambda t_i)^\beta)}$$

En différenciant par rapport à α, β et λ , respectivement, et en égalant à zéro, nous obtenons :

$$\frac{\partial \log l}{\partial \alpha} = \frac{n}{\alpha \log(\alpha)} - \frac{n}{\alpha - 1} + \frac{\sum_{i=1}^n (1 - \exp(1 - (1 + \lambda t_i)^\beta))}{\alpha} = 0$$

$$\begin{aligned} \frac{\partial \log l}{\partial \beta} &= \frac{n}{\beta} + \sum_{i=1}^n \log(1 + \lambda t_i) - \sum_{i=1}^n (1 + \lambda t_i)^\beta \log(1 + \lambda t_i) \\ &\quad + \log(\alpha) \sum_{i=1}^n (1 + \lambda t_i)^\beta \log(1 + \lambda t_i) \exp(1 - (1 + \lambda t_i)^\beta) = 0 \end{aligned}$$

et

$$\begin{aligned} \frac{\partial \log l}{\partial \lambda} &= \frac{n}{\lambda} + (\beta - 1) \sum_{i=1}^n \frac{t_i}{(1 + \lambda t_i)} - \beta \sum_{i=1}^n t_i (1 + \lambda t_i)^{\beta-1} \\ &\quad + \log(\alpha) \sum_{i=1}^n \beta t_i (1 + \lambda t_i)^{\beta-1} \exp(1 - (1 + \lambda t_i)^\beta) = 0 \end{aligned}$$

En résolvant les équations précédentes, nous obtenons des estimateurs des paramètres du modèle APTEE par la méthode MLE.

3.2.2 La matrice d'information de Fisher

Les éléments constitutifs de la matrice d'information de Fisher sont alors obtenus comme suit:

$$I_{11} = -\frac{n(\log(\alpha) + 1)}{(\alpha \log(\alpha))^2} - \frac{n}{(\alpha - 1)^2} + \frac{\sum_{i=1}^n (1 - \exp(1 - (1 + \lambda t_i)^\beta))}{\alpha^2}$$

$$\begin{aligned} I_{22} &= -\frac{n}{\beta^2} - \sum_{i=1}^n (1 + \lambda t_i)^\beta [\log(1 + \lambda t_i)]^2 \\ &\quad + \log(\alpha) \sum_{i=1}^n \lambda t_i [\log(1 + \lambda t_i)]^2 \exp(1 - (1 + \lambda t_i)^\beta) (1 + \lambda t_i) \end{aligned}$$

$$\begin{aligned} I_{33} &= -\frac{n}{\lambda^2} + (\beta - 1) \sum_{i=1}^n \frac{t_i^2}{(1 + \lambda t_i)^2} - \beta(\beta - 1) \sum_{i=1}^n t_i^2 (1 + \lambda t_i)^{\beta-2} \\ &\quad - \beta \log(\alpha) \sum_{i=1}^n t_i^2 \exp(1 - (1 + \lambda t_i)^\beta) (1 + \lambda t_i)^{\beta-2} [\beta(1 + \lambda t_i) - (\beta - 1)] \end{aligned}$$

et

$$I_{12} = \frac{1}{\alpha} \sum_{i=1}^n \exp(1 - (1 + \lambda t_i)^\beta) (1 + \lambda t_i)^\beta \log(1 + \lambda t_i)$$

$$I_{13} = \frac{\beta}{\alpha} \sum_{i=1}^n t_i \exp(1 - (1 + \lambda t_i)^\beta) (1 + \lambda t_i)^{\beta-1}$$

$$\begin{aligned} I_{23} &= \sum_{i=1}^n \frac{t_i}{(1 + \lambda t_i)} - \sum_{i=1}^n t_i (1 + \lambda t_i)^{\beta-1} (\beta \log(1 + \lambda t_i) + 1) \\ &\quad - \log(\alpha) \sum_{i=1}^n \beta t_i^2 \exp(1 - (1 + \lambda t_i)^\beta) (1 + \lambda t_i)^{\beta-1} \beta t_i (1 + \lambda t_i)^\beta \log(1 + \lambda t_i) \\ &\quad + \log(\alpha) \sum_{i=1}^n \beta t_i \exp(1 - (1 + \lambda t_i)^\beta) (1 + \lambda t_i)^{\beta-1} \log(1 + \lambda t_i) - 1 \end{aligned}$$

3.3 Tests d'ajustement pour le modèle APTEE

Dans le cas du modèle APTEE, il est essentiel de vérifier sa pertinence pour décrire fidèlement les observations empiriques. Dans cette section, nous présenterons les tests d'ajustement cités plus haut et proposerons une construction d'un test statistique adapté sur la statistique Nikulin-Rao-Robson, afin d'évaluer l'ajustement du modèle APTEE. En intégrant des données empiriques, nous illustrerons l'applicabilité et la robustesse des méthodes développées, contribuant ainsi à une meilleure compréhension du comportement du modèle dans des contextes variés.

3.3.1 Tests non paramétriques

Nous donnons les formules des différents tests d'ajustement adaptés à la loi APTEE.

Kolmogorov-Smirnov

La statistique du test de Kolmogorov-Smirnov D_n pour notre modèle est donnée par :

$$D_n = \max [D^+; D^-]$$

où

$$D^+ = \max_{1 \leq i \leq n} \left[\left(\frac{i}{n} \right) - F(t_{(i)}) \right]$$

$$D^+ = \text{Max}_{1 \leq i \leq n} \left[\left(\frac{i}{n} \right) - \frac{\alpha^{1 - \exp(1 - (1 + \lambda t_i)^\beta)} - 1}{\alpha - 1} \right]$$

et

$$D^- = \max_{1 \leq i \leq n} \left[F(t_{(i)}) - \left(\frac{i-1}{n} \right) \right]$$

$$D^- = \text{Max}_{1 \leq i \leq n} \left[\frac{\alpha^{1 - \exp(1 - (1 + \lambda t_i)^\beta)} - 1}{\alpha - 1} - \left(\frac{i-1}{n} \right) \right]$$

Cramer-Von Mises Le test de Cramer-Von-Mises a la forme suivante

$$W_n^2 = \sum_{i=1}^n \left[F(t_i) - \frac{2i-1}{2n} \right]^2 + \frac{1}{12n}$$

Donc pour la distribution APTEE(α, β, λ), on obtient

$$W_n^2 = \sum_{i=1}^n \left[\frac{\alpha^{1 - \exp(1 - (1 + \lambda t_i)^\beta)} - 1}{\alpha - 1} - \frac{2i-1}{2n} \right]^2 + \frac{1}{12n}$$

Anderson-Darling La statistique d'Anderson-Darling A_n^2 égale à

$$A_n^2 = -n - \frac{1}{n} \left\{ \sum_{i=1}^n (2i-1) [\ln F(t_{(i)}) + \ln(1 - F(t_{(n+1-i)}))] \right\}$$

prend la formule suivante pour le modèle APTEE

$$A_n^2 = -n - \frac{1}{n} \left[\sum_{i=1}^n (2i-1) \left[\ln \left(\frac{\alpha^{1 - \exp(1 - (1 + \lambda t_i)^\beta)} - 1}{\alpha - 1} \right) \right] + \ln \left(1 - \frac{\alpha^{1 - \exp(1 - (1 + \lambda t_{n+1-i})^\beta)} - 1}{\alpha - 1} \right) \right]$$

3.3.2 Construction du test statistique NRR pour la distribution APTEE

Soit $T = (t_1, \dots, t_n)^T$ n variables aléatoires indépendantes et de distribution identique, pour tester l'hypothèse nulle H_0 , telles que :

$$P(T_j \leq t \mid H_0) = F_{APTEE}(t; \theta), t \geq 0, \quad \theta = (\alpha, \beta, \lambda)^T,$$

Où $F_{APTEE}(t; \theta)$ est la fonction de distribution APTEE et $f_{APTEE}(t; \theta)$ sa densité de probabilité; nous construisons la statistique de Nikulin-Rao-Robson relative au modèle APTEE. On procède comme suit, Soient les limites K_j des intervalles de groupement des observations

$$0 = K_0 < K_1 < \dots < K_{r-1} < K_r = +\infty$$

Où

$$I_j =]K_{j-1}, K_j]; \quad I_i \cap I_j = \emptyset \quad i \neq j; \quad \bigcup_{j=1}^r I_j = R^+.$$

Dans le cas d'équiprobabilité ($P_1 = P_2 = \dots = P_r$), les K_j sont obtenus par:

$$\begin{aligned} K_j &= F_{APTEE}^{-1}(P_1 + P_2 + \dots + P_j) = F_{APTEE}^{-1}\left(\frac{j}{r}\right); \quad j = 1, \dots, r-1. \\ &= \left[\log \left(1 - \left\{ \frac{\log [(\alpha - 1)] \left(\frac{j}{r}\right) + 1}{\log(\alpha)} \right\}^{\frac{1}{\beta}} \right) \right]^{-\frac{1}{\lambda}} \end{aligned}$$

Les fonctions P_j de la distribution APTEE sont alors:

$$P_j(\hat{\theta}) = \left(\frac{\alpha^{1 - \exp(1 - (1 + \lambda K_{j-1})^\beta)} - 1}{\alpha - 1} \right) - \left(\frac{\alpha^{1 - \exp(1 - (1 + \lambda K_j)^\beta)} - 1}{\alpha - 1} \right) \quad \text{avec } j = 1, \dots, r$$

La matrice symétrique estimée $J(\hat{\theta})$, pour les données groupées, est obtenue à partir des formules suivantes :

$$J_{11}(\hat{\theta}) = \sum_j \frac{1}{P_j} \left(\frac{\partial P_j}{\partial \hat{\alpha}} \right)^2 ; J_{22}(\hat{\theta}) = \sum_j \frac{1}{P_j} \left(\frac{\partial P_j}{\partial \hat{\beta}} \right)^2$$

$$J_{33}(\hat{\theta}) = \sum_j \frac{1}{P_j} \left(\frac{\partial P_j}{\partial \hat{\lambda}} \right)^2$$

$$J_{12} = J_{21} = \sum_j \frac{1}{P_j} \left(\frac{\partial P_j}{\partial \hat{\alpha}} \right) \left(\frac{\partial P_j}{\partial \hat{\beta}} \right) ; J_{13} = J_{31} = \sum_j \frac{1}{P_j} \left(\frac{\partial P_j}{\partial \hat{\alpha}} \right) \left(\frac{\partial P_j}{\partial \hat{\lambda}} \right)$$

$$J_{23} = J_{32} = \sum_j \frac{1}{P_j} \left(\frac{\partial P_j}{\partial \hat{\beta}} \right) \left(\frac{\partial P_j}{\partial \hat{\lambda}} \right)$$

Les dérivées partielles de $P_j(\hat{\theta})$ sont alors :

$$\frac{\partial P_j(\hat{\theta})}{\partial \hat{\alpha}} = \frac{(\exp(1 - (1 + \lambda K_j)^\beta)) \times (\alpha^{1 - \exp(1 - (1 + \lambda K_j)^\beta)} - 1)}{(\alpha - 1)^2} - \frac{(\exp(1 - (1 + \lambda K_{j-1})^\beta)) \times (\alpha^{1 - \exp(1 - (1 + \lambda K_{j-1})^\beta)} - 1)}{(\alpha - 1)^2}$$

$$\frac{\partial P_j(\hat{\theta})}{\partial \hat{\beta}} = \frac{\alpha^{1 - \exp(1 - (1 + \lambda K_j)^\beta)} \times \exp(1 - (1 + \lambda K_j)^\beta) \times ((1 + \lambda K_j)^\beta) \ln(1 + \lambda K_j)}{(\alpha - 1)^2} - \frac{\alpha^{1 - \exp(1 - (1 + \lambda K_{j-1})^\beta)} \times \exp(1 - (1 + \lambda K_{j-1})^\beta) \times ((1 + \lambda K_{j-1})^\beta) \ln(1 + \lambda K_{j-1})}{(\alpha - 1)^2}$$

$$\frac{\partial P_j(\hat{\theta})}{\partial \hat{\lambda}} = \frac{\alpha^{1 - \exp(1 - (1 + \lambda K_j)^\beta)} \times (\exp(1 - (1 + \lambda K_j)^\beta)) \times (\beta K_j (1 + \lambda K_j)^{\beta-1})}{(\alpha - 1)^2} - \frac{\alpha^{1 - \exp(1 - (1 + \lambda K_{j-1})^\beta)} \times (\exp(1 - (1 + \lambda K_{j-1})^\beta)) \times (\beta K_{j-1} (1 + \lambda K_{j-1})^{\beta-1})}{(\alpha - 1)^2}$$

on peut aussi calculer le vecteur $L(\hat{\theta}) = (L_1, L_2, L_3)^T$. Nous avons :

$$\begin{aligned}
 L_1(\widehat{\theta}) &= \sum_{j=1}^r \frac{v_j}{P_j} \frac{\partial P_j}{\partial \widehat{\alpha}}(\widehat{\theta}) \\
 &= \sum_{j=1}^r \frac{v_j}{P_j} \left(\frac{(\exp(1 - (1 + \lambda K_j)^\beta)) \times (\alpha^{1 - \exp(1 - (1 + \lambda K_j)^\beta)} - 1)}{(\alpha - 1)^2} \right) \\
 &\quad - \sum_{j=1}^r \frac{v_j}{P_j} \left(\frac{(\exp(1 - (1 + \lambda K_{j-1})^\beta)) \times (\alpha^{1 - \exp(1 - (1 + \lambda K_{j-1})^\beta)} - 1)}{(\alpha - 1)^2} \right)
 \end{aligned}$$

$$\begin{aligned}
 L_2(\widehat{\theta}) &= \sum_{j=1}^r \frac{v_j}{P_j} \frac{\partial P_j}{\partial \widehat{\beta}}(\widehat{\theta}) \\
 &= \sum_{j=1}^r \frac{v_j}{P_j} \left(\frac{\alpha^{1 - \exp(1 - (1 + \lambda K_j)^\beta)} \times \exp(1 - (1 + \lambda K_j)^\beta) \times ((1 + \lambda K_j)^\beta) \ln(1 + \lambda K_j)}{(\alpha - 1)^2} \right) \\
 &\quad - \sum_{j=1}^r \frac{v_j}{P_j} \left(\frac{\alpha^{1 - \exp(1 - (1 + \lambda K_{j-1})^\beta)} \times \exp(1 - (1 + \lambda K_{j-1})^\beta) \times ((1 + \lambda K_{j-1})^\beta) \ln(1 + \lambda K_{j-1})}{(\alpha - 1)^2} \right)
 \end{aligned}$$

$$\begin{aligned}
 L_3(\widehat{\theta}) &= \sum_{j=1}^r \frac{v_j}{P_j} \frac{\partial P_j}{\partial \widehat{\lambda}}(\widehat{\theta}) \\
 &= \sum_{j=1}^r \frac{v_j}{P_j} \left(\frac{\alpha^{1 - \exp(1 - (1 + \lambda K_j)^\beta)} \times (\exp(1 - (1 + \lambda K_j)^\beta)) \times (\beta K_j (1 + \lambda K_j)^{\beta-1})}{(\alpha - 1)^2} \right) \\
 &\quad - \sum_{j=1}^r \frac{v_j}{P_j} \left(\frac{\alpha^{1 - \exp(1 - (1 + \lambda K_{j-1})^\beta)} \times (\exp(1 - (1 + \lambda K_{j-1})^\beta)) \times (\beta K_{j-1} (1 + \lambda K_{j-1})^{\beta-1})}{(\alpha - 1)^2} \right)
 \end{aligned}$$

On peut déduire la statistique Y^2 pour valider le modèle APTEE.

3.4 Simulations et Applications réelles

Cette section est consacrée à l'application des résultats théoriques obtenus sur des dizaines de milliers d'échantillons générés à partir de la distribution APTEE.

3.4.1 Simulations

Estimation des paramètres du modèle APTEE

Nous avons généré $M = 10000$ échantillons de différentes tailles ($n = 15, n = 50, n = 100, n = 200$) à partir de la distribution APTEE. En utilisant différentes méthodes, nous avons calculé les estimateurs des paramètres inconnus et leurs erreurs quadratiques moyennes. Les valeurs des paramètres sont fixées à $\alpha = 2, \beta = 0.5, \lambda = 1.5$. Les résultats sont présentés dans les tableaux 7, 8, 9 et 10.

$M = 10000$		$\hat{\alpha}$	$\hat{\beta}$	$\hat{\lambda}$
$n = 15$	Mean	1.88769	0.45578	1.47863
	SME	0.10873	0.00783	0.01812
$n = 50$	Mean	1.91835	0.48869	1.48625
	SME	0.06724	0.01561	0.01561
$n = 100$	Mean	1.95092	0.50262	1.48459
	SME	0.02619	0.00354	0.01356
$n = 200$	Mean	1.98548	0.50578	1.49379
	SME	0.00672	0.00261	0.00946

Tableau 7: Estimateurs du maximum de vraisemblance et leurs erreurs quadratiques moyennes

$M = 10000$		$\hat{\alpha}$	$\hat{\beta}$	$\hat{\lambda}$
$n = 15$	Mean	2.06512	0.51193	1.48082
	SME	0.06473	0.01685	0.00902
$n = 50$	Mean	2.04507	0.50330	1.48920
	SME	0.03163	0.00934	0.00853
$n = 100$	Mean	2.01887	0.50216	1.49272
	SME	0.00848	0.00462	0.00792
$n = 200$	Mean	2.00143	0.50187	1.49972
	SME	0.00609	0.00313	0.00423

Tableau 8: Estimateurs KS et leurs erreurs quadratiques moyennes

$M = 10000$		$\hat{\alpha}$	$\hat{\beta}$	$\hat{\lambda}$
$n = 15$	Mean	1.93732	0.53416	1.51731
	SME	0.12935	0.01808	0.00912
$n = 50$	Mean	1.97703	0.51809	1.50451
	SME	0.08514	0.00911	0.00853
$n = 100$	Mean	1.97802	0.51472	1.50208
	SME	0.06088	0.00551	0.00825
$n = 200$	Mean	1.99525	0.51014	1.50127
	SME	0.03753	0.00343	0.00721

Tableau 9: Estimateurs CVM et leurs erreurs quadratiques moyennes

$M = 10000$		$\hat{\alpha}$	$\hat{\beta}$	$\hat{\lambda}$
$n = 15$	Mean	1.85056	0.51393	1.45588
	SME	0.13439	0.01578	0.00916
$n = 50$	Mean	1.92429	0.51354	1.48599
	SME	0.06237	0.00754	0.00828
$n = 100$	Mean	1.95514	0.50690	1.49510
	SME	0.02522	0.00468	0.00796
$n = 200$	Mean	1.98087	0.50548	1.50116
	SME	0.00925	0.00279	0.00681

Tableau 10: Estimateurs AD et leurs erreurs quadratiques moyennes

Les valeurs des estimateurs obtenus sont très proches des vraies valeurs, et leurs erreurs quadratiques moyennes (SME) diminuent à mesure que la taille de l'échantillon augmente, et ce pour toutes les méthodes. Ce qui confirme la convergence des méthodes utilisées.

Critères de Test

Pour vérifier l'hypothèse nulle H_0 selon laquelle les échantillons sont tirés de la distribution APTEE, le nombre de rejets de l'hypothèse nulle H_0 pour des milliers d'échantillons de tailles différentes ($n = 15, 50, 100, 200$) est calculé et comparé à leurs valeurs théoriques pour les niveaux de signification $\xi = 0.01, \xi = 0.1$ et $\xi = 0.05$

n	$\xi = 0.01$	$\xi = 0.1$	$\xi = 0.05$
15	0.0138	0.1140	0.0462
50	0.0113	0.1090	0.0492
100	0.0092	0.1055	0.0498
200	0.0108	0.1003	0.5009

Tableau 11: khi-deux critique et leurs valeurs empiriques correspondantes

Comme prévu, les valeurs empiriques obtenues sont très proches de leurs valeurs théoriques correspondantes, ce qui confirme que le test d'ajustement proposé est bien adapté pour valider la distribution APTEE.

3.4.2 Applications

L'analyse porte sur deux types de données (survie et fiabilité)

Données de durée de survie

On s'intéresse à un ensemble de données réelles (ensemble de données I) représentant la durée de survie (en jours) de 72 cobayes infectés par des bacilles tuberculeux virulents, selon Bjerkedal (1960). Les données sont les suivantes :

0.1, 0.33, 0.44, 0.56, 0.59, 0.72, 0.74, 0.77, 0.92, 0.93, 0.96, 1, 1, 1.02, 1.05, 1.07, 07, 1.08, 1.08,1.08, 1.09, 1.12, 1.13, 1.15, 1.16, 1.2, 1.21, 1.22, 1.22, 1.24, 1.3, 1.34, 1.36, 1.39, 1.44, 1.46, 1.53,1.59, 1.6, 1.63, 1.63, 1.68, 1.71, 1.72, 1.76, 1.83, 1.95, 1.96, 1.97, 2.02, 2.13, 2.15, 2.16, 2.22, 2.3,2.31, 2.4, 2.45, 2.51, 2.53, 2.54, 2.54, 2.78, 2.93, 3.27, 3.42, 3.47, 3.61, 4.02, 4.32, 4.58, 5.55.

les estimateurs des paramètres sont calculés (Tableau 12) :

<i>Methode</i>	α	β	λ
<i>AD</i>	1.29811	0.05858	1.71231
<i>CVM</i>	1.60684	0.86685	0.72275
<i>KS</i>	1.64178	0.87502	1.60815
<i>MLE</i>	0.04398	1.57090	0.05153

Tableau 12: Valeurs estimées des paramètres inconnus de la distribution APTEE

Maintenant on va vérifier l'hypothèse nulle que l'ensemble de données I appartient à la distribution APTEE. Si nous choisissons $r = 8$ classes de regroupement, le calcul intermédiaire du critère de test Y^2 nous donne :

$$L = (L_1 = -1.906953, L_2 = -3.778612, L_3 = -7.907812)$$

$$I_n(\hat{\theta}) = \begin{pmatrix} 0.2194303 & 0.1699780 & 0.1493361 \\ 0.1699780 & 0.1316705 & 0.1156806 \\ 0.1493361 & 0.1156806 & 0.1016325 \end{pmatrix}$$

$$J(\hat{\theta}) = \begin{pmatrix} 0.8662394 & 1.113803 & 3.322811 \\ 1.1138027 & 1.492910 & 4.30159 \\ 3.3281101 & 4.301590 & 12.79500 \end{pmatrix}$$

Nous obtenons $Y^2 = 6.50414$, donc pour le niveau de signification $\xi = 0.05$, la valeur critique du khi-deux est $X_{r-1}^2 = 14.06$, l'hypothèse nulle H_0 est confirmée. Ces données peuvent être décrites la distribution APTEE.

Exemple de fiabilité

Pour illustrer l'utilité du modèle APTEE dans les études de fiabilité, nous analysons l'ensemble de données réelles tiré de Linhart et Zucchini [18], qui représente les temps de défaillance du système de climatisation d'un avion. les donnés sont: 23, 261, 87, 7, 120, 14, 62, 47, 225, 71, 246, 21, 42, 20, 5, 12, 120, 11, 3, 14, 71, 11, 14, 11, 16, 90, 1, 16, 52, 95

Supposons que ces données suivent la distribution APTEE. Si nous choisissons $r = 5$ classes de regroupement, le calcul intermédiaire du critère de test Y^2 est le suivant :

$$L = (L_1 = -0.0048667, L_2 = 0.00148015, L_3 = -0.0045631)$$

$$I_n(\hat{\theta}) = \begin{pmatrix} 0.05570600 & -0.00169401 & -0.0052243 \\ -0.00169401 & 0.00515147 & 0.00158813 \\ -0.0052243 & 0.1156806 & 0.00489603 \end{pmatrix}$$

$$J(\hat{\theta}) = \begin{pmatrix} 0.04619088 & -0.0140465 & -0.043303 \\ -0.0140465 & -0.00427155 & 0.0131686 \\ -0.043303 & 0.013168681 & 0.0459746 \end{pmatrix}$$

Nous obtenons $Y^2 = 4.8999$, donc pour le niveau de signification $\xi = 0.05$, la valeur critique du khi-deux est $X_{r-1}^2 = 9.487729$ ce qui confirme l'hypothèse nulle.

À travers ces études empiriques, nous avons mis en œuvre les tests statistiques afin de vérifier l'adéquation du modèle APTEE aux données observées. Les résultats obtenus permettent non seulement de valider l'efficacité du modèle, mais aussi d'explorer son potentiel pour des applications futures dans divers domaines scientifiques et industriels.

4

Modèle alpha-power transformé de Chen (APTC)

La distribution de Chen, introduite en 2000, s'est révélée particulièrement efficace pour modéliser des variables aléatoires de divers domaines tels que l'ingénierie, les assurances, la physique et les télécommunications, elle a démontré une flexibilité importante pour représenter les processus de défaillance. Ces dernières années, cette distribution a attiré l'attention de plusieurs auteurs qui ont proposé diverses généralisations, on cite Abdulzeid Anafo et al., (2022) pour l'extension de la distribution de Chen modifiée avec application aux données pluviométriques, Méndez-González et al., (2023) pour une autre distribution de Chen obtenue par additivité, nommée additive Chen distribution. Tandis que Tarvirdizade et Ahmadpour (2019) ont introduit une nouvelle distribution en composant la distribution de Chen et la distribution de Weibull nommée Weibull–Chen (W-C) distribution. Abbas et al., (2022) se sont intéressés à l'addition de ces distributions pour obtenir la Chen-Weibull (ACW) distribution. Sousa-Ferreira et al., (2023) ont obtenu la distribution Extended Chen-Poisson en composant la loi de Poisson et celle de Chen, Abidemi and Abiodun (2023) ont quant à eux construit une distribution de Chen générée exponentiellement, Exponentially Generated Modified Chen Distribution.

En utilisant la généralisation par la méthode de transformation alpha-power introduite par Mahdavi et Kundu (2017), nous proposons de construire un nouveau modèle basé sur la distribution de Chen. Pour accroître la capacité de modélisation et mieux ajuster les données empiriques, une extension de ce modèle basée sur la transformation Alpha-Power (APT) a été introduite donnant naissance à la distribution Alpha-Power Transformed Chen (APTC). Ce modèle permet d'incorporer un paramètre supplémentaire α , offrant ainsi une plus grande flexibilité dans la représentation des dynamiques de défaillance et dans l'analyse des risques concurrents.

L'approche APT appliquée à la distribution de Chen permet d'obtenir une forme générale qui englobe plusieurs distributions connues, tout en introduisant de nouveaux paramètres de contrôle sur la structure des probabilités.

L'objectif principal de ce chapitre est d'examiner les propriétés fondamentales du modèle APTC, notamment ses fonctions de densité de probabilité, distribution cumulative, taux de hasard. Nous explorerons également les méthodes d'estimation des paramètres inconnus à l'aide du maximum de vraisemblance et nous calculons aussi la matrice d'information de Fisher. Enfin, des tests statistiques d'ajustement seront développés pour vérifier l'adéquation du modèle aux données réelles et en démontrer l'applicabilité pratique.

4.1 Présentation du modèle APTC

La distribution de Chen (Chen, 2000) de paramètres λ et β est définie par une fonction de distribution cumulative et une fonction de densité de probabilité données par

$$F(t) = 1 - \exp\left(\lambda\left(1 - e^{t^\beta}\right)\right) \quad t \geq 0, \lambda > 0, \beta > 0$$

et

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

Mahdavi et Kundu (2017) ont proposé une généralisation de distributions par une transformation de la distribution de base par une puissance, appelée alpha-power trans-

formation en ajoutant un nouveau paramètre α . Si $F(t)$ est la fonction de densité cumulée d'une distribution quelconque, alors $F_{APT}(t)$ est obtenue comme suit:

$$F_{APT}(t) = \begin{cases} \frac{\alpha^{F(t)} - 1}{\alpha - 1} & \text{Si } \alpha > 0, \alpha \neq 1 \\ F(t) & \text{Si } \alpha = 1 \end{cases}$$

Cela signifie que pour $\alpha = 1$, la transformation ne modifie pas la fonction $F(t)$.

La densité de probabilité associée à cette transformation (*PDF*) est donnée par:

$$f_{APT}(t) = \begin{cases} \frac{\log(\alpha)}{\alpha - 1} f(t) \alpha^{F(t)} & \text{Si } \alpha > 0, \alpha \neq 1 \\ f(t) & \text{Si } \alpha = 1 \end{cases}$$

En se basant sur cette approche, on construit la distribution Alpha-power transformée Chen (APTC) en substituant la fonction cumulative de la distribution Chen dans la transformation alpha-power $F_{APT}(t)$.

Soit $\theta = (\alpha, \lambda, \beta)$, où α est le paramètre supplémentaire, la fonction de répartition cumulative (CDF) de la distribution APTC est obtenue par

$$F_{APTC}(t) = \begin{cases} \frac{\alpha^{1 - \exp(\lambda(1 - e^{t^\beta}))} - 1}{\alpha - 1} & \text{Si } \alpha > 0, \alpha \neq 1 \\ F(t) & \text{Si } \alpha = 1 \end{cases}$$

sa fonction de densité de probabilité (PDF) est alors donnée par

$$f_{APTC}(t) = \begin{cases} \frac{\log(\alpha)}{\alpha - 1} \lambda \beta e^{t^{\beta-1}} e^t \exp(\lambda(1 - e^{t^\beta})) \alpha^{1 - \exp(\lambda(1 - e^{t^\beta}))} & \text{Si } \alpha > 0, \alpha \neq 1 \\ f(t) & \text{Si } \alpha = 1 \end{cases}$$

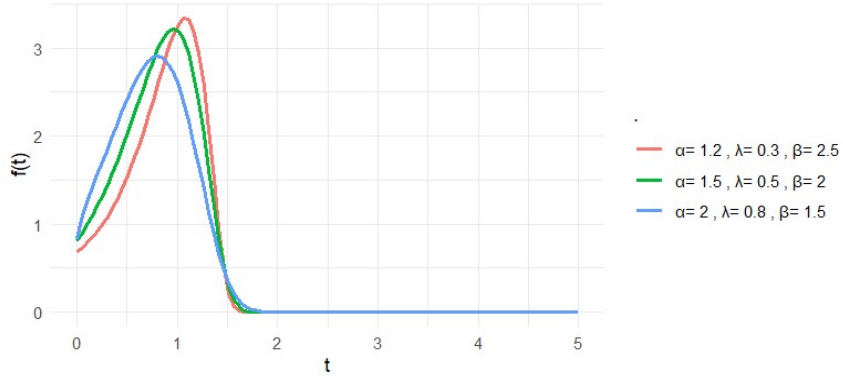


Figure 4.1: Graphe de la fonction de densité du modèle APTC

La fonction de survie et la fonction de taux de risque de la distribution APTC sont définies par

$$S_{APTC}(t) = \begin{cases} \frac{\alpha - \alpha^{1 - \exp(\lambda(1 - e^{t^\beta}))}}{\alpha - 1} & \text{Si } \alpha > 0, \alpha \neq 1 \\ \exp(\lambda(1 - e^{t^\beta})) & \text{Si } \alpha = 1 \end{cases}$$

$$h_{APTC}(t) = \begin{cases} \frac{\lambda \beta e^{t^{\beta-1}} \alpha^{1 - \exp(\lambda(1 - e^{t^\beta}))} \exp(\lambda(1 - e^{t^\beta}))}{\alpha - \alpha^{1 - \exp(\lambda(1 - e^{t^\beta}))}} & \text{Si } \alpha > 0, \alpha \neq 1 \\ \lambda \beta \exp(t)^{\beta-1} & \text{Si } \alpha = 1 \end{cases}$$

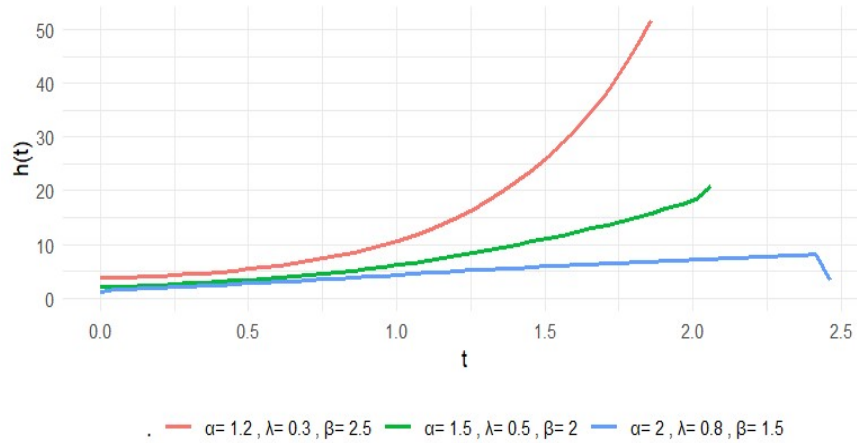


Figure 4.2: Graphe de la fonction de taux de hasard du modèle APTC

La distribution $APTC(\alpha, \beta, \lambda)$ peut être représentée par un modèle mélange pour $\alpha > 1$. $\left(\frac{\log \alpha}{\alpha-1}\right)$ est une fonction décroissante de 1 à 0, lorsque α varie de 1 à ∞ . Si $T \sim APTC(\alpha, \beta, \lambda)$, elle peut être représentée comme suit :

$$T = \begin{cases} T_1 & \text{avec probabilité } \left(\frac{\log \alpha}{\alpha-1}\right) \\ T_2 & \text{avec probabilité } 1 - \left(\frac{\log \alpha}{\alpha-1}\right) \end{cases}$$

où T_1 et T_2 ont les PDF suivants :

$$f(T_1) = \lambda \beta e^{t^{\beta-1}} e^t \exp\left(\lambda(1 - e^{t^\beta})\right)$$

$$f(T_2) = \frac{\log \alpha}{(\alpha - 1 - \log \alpha)} \lambda \beta e^{t^{\beta-1}} e^t \exp\left(\lambda(1 - e^{t^\beta})\right) \left[\alpha^{1 - \exp(\lambda(1 - e^{t^\beta}))} - 1 \right]$$

respectivement. Il est clair que lorsque α approche 1, T se comporte comme une distribution Chen, et lorsque α augmente, il se comporte comme T_2 .

4.2 Estimation

Considérons un échantillon aléatoire (t_1, t_2, \dots, t_n) de la distribution $APTC(\alpha, \beta, \lambda)$ dont les paramètres sont inconnus. Dans cette section, nous proposons différentes méthodes d'estimation pour évaluer leurs valeurs.

4.2.1 Estimation des paramètres inconnus

En raison de leurs propriétés intéressantes, à savoir la convergence et la normalité asymptotique, les estimateurs du maximum de vraisemblance sont généralement requis. La fonction de vraisemblance est :

$$l(\alpha, \beta, \lambda | t) = \prod_{i=1}^n f(t_i) = \prod_{i=1}^n \frac{\log \alpha}{\alpha - 1} \left(\exp\left(\lambda(1 - e^{t_i^\beta})\right) \lambda \beta e^{t_i^{\beta-1}} e^{t_i} \alpha^{1 - \exp(\lambda(1 - e^{t_i^\beta}))} \right)$$

en prenant le logarithme, nous obtenons :

$$\begin{aligned} \log l(\alpha, \beta, \lambda \setminus t) &= \sum_{i=1}^n \log(\log(\alpha)) - \log(\alpha - 1) + \lambda \left(1 - e^{t_i^\beta}\right) + \log(\beta) + (\beta - 1) t_i + t_i \\ &\quad + \log(\alpha) + \log\left(1 - \exp\left(\lambda \left(1 - e^{t_i^\beta}\right)\right)\right) \end{aligned}$$

En différenciant l'équation par rapport à α, β et λ , respectivement, et en égalant à zéro, nous obtenons :

$$\frac{\partial \log l(\alpha, \beta, \lambda \setminus t)}{\partial \alpha} = \frac{n}{\alpha \log(\alpha)} - \frac{n}{\alpha - 1} + \frac{n}{\alpha} = 0$$

$$\frac{\partial \log l(\alpha, \beta, \lambda \setminus t)}{\partial \beta} = \frac{n}{\beta} + \sum_{i=1}^n t_i - \sum_{i=1}^n \lambda t_i e^{t_i^\beta} \exp\left(\lambda \left(1 - e^{t_i^\beta}\right)\right) = 0$$

et

$$\frac{\partial \log l(\alpha, \beta, \lambda \setminus t)}{\partial \lambda} = \sum_{i=1}^n \left[\left(1 - e^{t_i^\beta}\right) - \frac{\exp\left(\lambda \left(1 - e^{t_i^\beta}\right)\right)}{1 - \exp\left(\lambda \left(1 - e^{t_i^\beta}\right)\right)} \right] = 0$$

Les estimateurs du maximum de vraisemblance des paramètres du modèle sont déterminés en résolvant numériquement les équations non linéaires $\frac{\partial \log l(\alpha, \beta, \lambda \setminus t)}{\partial \alpha} = 0$, $\frac{\partial \log l(\alpha, \beta, \lambda \setminus t)}{\partial \beta} = 0$, $\frac{\partial \log l(\alpha, \beta, \lambda \setminus t)}{\partial \lambda} = 0$ de manière simultanée. On utilise le logiciel R de calcul mathématique ou une autre méthode itérative.

4.2.2 La Matrice d'information de Fisher

Les éléments de la matrice d'information de Fisher nécessaires pour la construction du test d'ajustement sur les données initiales, $I(\hat{\theta})_{3 \times 3}$ pour la distribution APTC, sont obtenus comme suit :

$$I_{11} = -\frac{n(\log(\alpha) + 1)}{(\alpha \log(\alpha))^2} - \frac{n}{(\alpha - 1)^2} - \frac{n}{\alpha^2}$$

$$I_{22} = -\frac{n}{\beta^2} - \sum_{i=1}^n \lambda t_i^2 e^{t_i^\beta} \exp\left(\lambda \left(1 - e^{t_i^\beta}\right)\right)$$

$$I_{33} = - \sum_{i=1}^n \frac{\exp\left(\lambda\left(1 - e^{t_i^\beta}\right)\right)}{1 - \exp\left(\lambda\left(1 - e^{t_i^\beta}\right)\right)}$$

et

$$I_{13} = -\frac{n}{\alpha\lambda} + \sum_{i=1}^n \frac{\exp\left(\lambda\left(1 - e^{t_i^\beta}\right)\right)}{\left(\alpha^{1-\exp\left(\lambda\left(1 - e^{t_i^\beta}\right)\right)}\right)^2}$$

$$I_{23} = - \sum_{i=1}^n t_i e^{t_i^\beta} \exp\left(\lambda\left(1 - e^{t_i^\beta}\right)\right)$$

4.2.3 Estimateurs classiques

Ces méthodes sont basées sur les statistiques classiques, bien connues, pour déterminer les estimateurs des paramètres inconnus du modèle. Les valeurs sont obtenues en minimisant ces statistiques par rapport aux paramètres.

Estimateurs de Kolmogorov-Smirnov

La statistique de Kolmogorov-Smirnov D_n relative à la distribution APTC est donnée comme suit:

$$D_n = \max [D^+; D^-]$$

$$D^+ = \max_{1 \leq i \leq n} \left[\left(\frac{i}{n} \right) - F(t_{(i)}) \right] \quad ; \quad D^- = \max_{1 \leq i \leq n} \left[F(t_{(i)}) - \left(\frac{i-1}{n} \right) \right]$$

$$D^+ = \text{Max}_{1 \leq i \leq n} \left[\left(\frac{i}{n} \right) - \frac{\alpha^{1-\exp\left(\lambda\left(1 - e^{t_i^\beta}\right)\right)} - 1}{\alpha - 1} \right]$$

et

$$D^- = \text{Max}_{1 \leq i \leq n} \left[\frac{\alpha^{1-\exp\left(\lambda\left(1 - e^{t_i^\beta}\right)\right)} - 1}{\alpha - 1} - \left(\frac{i-1}{n} \right) \right]$$

Estimateurs de Cramer-Von-Mises

$$W_n^2 = \sum_{i=1}^n \left[F(t_i) - \frac{2i-1}{2n} \right]^2 + \frac{1}{12n}$$

Donc pour la distribution APTC (α, β, λ) , on obtient

$$W_n^2 = \sum_{i=1}^n \left[\frac{\alpha^{1-\exp(\lambda(1-e^{t_i^\beta}))} - 1}{\alpha - 1} - \frac{2i-1}{2n} \right]^2 + \frac{1}{12n}$$

Statistique d'Anderson-Darling

La statistique d'Anderson-Darling A_n^2

$$A_n^2 = -n - \frac{1}{n} \left\{ \sum_{i=1}^n (2i-1) [\ln F(t_{(i)}) + \ln(1 - F(t_{(n+1-i)}))] \right\}$$

prend la formule suivante pour le modèle APTC

$$A_n^2 = -n - \frac{1}{n} \left[\sum_{i=1}^n (2i-1) \left[\ln \left(\frac{\alpha^{1-\exp(\lambda(1-e^{t_{(i)}^\beta}))} - 1}{\alpha - 1} \right) \right] + \ln \left(1 - \frac{\alpha^{1-\exp(\lambda(1-e^{t_{(n+1-i)}^\beta}))} - 1}{\alpha - 1} \right) \right]$$

4.3 Test d'ajustement pour le modèle APTC

La validation du modèle choisi pour une analyse statistique doit impérativement être prouvée afin d'avoir des résultats fiables.

Dans cette section, nous proposons de construire un nouveau test statistique du khi-deux Y^2 pour ajuster la distribution APTC en utilisant l'approche de Nikulin, Rao et Robson. Cette statistique récupère l'information perdue lors du regroupement des données. Elle permet aux utilisateurs de vérifier la validité de cette distribution sans tenir compte des concurrents potentiels.

Pour vérifier si l'échantillon $T = (t_1, \dots, t_n)^T$ appartient à la distribution APTC, $F_{APTC}(t; \theta)$:

$$H_0 : P(T_j \leq t | H_0) = F_{APTC}(t; \theta), \quad t \geq 0, \quad \text{avec} \quad \theta = (\alpha, \beta, \lambda)^T$$

Premièrement, nous calculons les estimateurs ML des paramètres inconnus. Pour cette statistique, les r classes groupées Δ_j doivent être équiprobables, ce qui signifie que $P_1 = P_2 = \dots = P_r = \frac{1}{r}$. Dans ce cas, les limites K_j des classes $\Delta_j = [K_{j-1}, K_j]$ sont obtenues par :

$$K_j = F_{APTC}^{-1} \left(\frac{j}{r} \right) = \left(\frac{\ln \left(1 - \frac{\ln \left(1 + \frac{j}{r} (\alpha - 1) \right)}{\lambda \ln(\alpha)} \right)}{\beta} \right) \quad j = 1, \dots, r - 1$$

et

$$P_j(\hat{\theta}) = \left(\frac{\alpha^{1 - \exp(\lambda(1 - e^{K_j^\beta}))} - 1}{\alpha - 1} \right) - \left(\frac{\alpha^{1 - \exp(\lambda(1 - e^{K_j^\beta}))} - 1}{\alpha - 1} \right) \quad \text{avec } j = 1, \dots, r$$

Pour fournir la statistique de critères Y^2 pour ajuster la distribution $F_{APTC}(t; \theta)$, nous avons besoin des deux matrices d'estimation d'informations estimées $J(\hat{\theta})$ et $I(\hat{\theta})$ pour les données groupées et non groupées.

Matrice d'information estimée $J(\hat{\theta})$

Les composantes de la matrice symétrique estimée $J(\hat{\theta})$ pour les données groupées définies par :

$$J_{11}(\hat{\theta}) = \sum_j \frac{1}{P_j} \left(\frac{\partial P_j}{\partial \hat{\alpha}} \right)^2 ; J_{22}(\hat{\theta}) = \sum_j \frac{1}{P_j} \left(\frac{\partial P_j}{\partial \hat{\beta}} \right)^2$$

$$J_{33}(\hat{\theta}) = \sum_j \frac{1}{P_j} \left(\frac{\partial P_j}{\partial \hat{\lambda}} \right)^2$$

$$J_{12} = J_{21} = \sum_j \frac{1}{P_j} \left(\frac{\partial P_j}{\partial \hat{\alpha}} \right) \left(\frac{\partial P_j}{\partial \hat{\beta}} \right) ; J_{13} = J_{31} = \sum_j \frac{1}{P_j} \left(\frac{\partial P_j}{\partial \hat{\alpha}} \right) \left(\frac{\partial P_j}{\partial \hat{\lambda}} \right)$$

$$J_{23} = J_{32} = \sum_j \frac{1}{P_j} \left(\frac{\partial P_j}{\partial \hat{\beta}} \right) \left(\frac{\partial P_j}{\partial \hat{\lambda}} \right)$$

sont obtenues à partir des dérivées partielles suivantes:

$$\frac{\partial P_j(\hat{\theta})}{\partial \hat{\alpha}} = \left(\frac{\left(1 - \exp\left(\lambda\left(1 - e^{K_{j-1}^\beta}\right)\right)\right) \alpha^{1 - \exp\left(\lambda\left(1 - e^{K_{j-1}^\beta}\right)\right)}}{(\hat{\alpha} - 1)^2} \right) - \left(\frac{\left(1 - \exp\left(\lambda\left(1 - e^{K_j^\beta}\right)\right)\right) \alpha^{1 - \exp\left(\lambda\left(1 - e^{K_j^\beta}\right)\right)}}{(\hat{\alpha} - 1)^2} \right)$$

$$\frac{\partial P_j(\hat{\theta})}{\partial \hat{\beta}} = \left(\frac{\left(-\exp\left(\lambda\left(1 - e^{K_{j-1}^\beta}\right)\right)\right) \lambda K_{j-1}^{\beta-1} e^{K_{j-1}^\beta} \alpha^{1 - \exp\left(\lambda\left(1 - e^{K_{j-1}^\beta}\right)\right)}}{(\hat{\alpha} - 1)^2} \right) - \left(\frac{\left(-\exp\left(\lambda\left(1 - e^{K_j^\beta}\right)\right)\right) \lambda K_j^{\beta-1} e^{K_j^\beta} \alpha^{1 - \exp\left(\lambda\left(1 - e^{K_j^\beta}\right)\right)}}{(\hat{\alpha} - 1)^2} \right)$$

$$\frac{\partial P_j(\hat{\theta})}{\partial \hat{\lambda}} = \left(\frac{\alpha^{1 - \exp\left(\lambda\left(1 - e^{K_{j-1}^\beta}\right)\right)} \left(-\exp\left(\lambda\left(1 - e^{K_{j-1}^\beta}\right)\right)\right) \left(1 - e^{K_{j-1}^\beta}\right)}{(\hat{\alpha} - 1)^2} \right) - \left(\frac{\alpha^{1 - \exp\left(\lambda\left(1 - e^{K_j^\beta}\right)\right)} \left(-\exp\left(\lambda\left(1 - e^{K_j^\beta}\right)\right)\right) \left(1 - e^{K_j^\beta}\right)}{(\hat{\alpha} - 1)^2} \right)$$

On déduit également les éléments du vecteur $L(\hat{\theta}) = (L_1, L_2, L_3)^T$.

$$\begin{aligned} L_1(\hat{\theta}) &= \sum_{j=1}^r \frac{v_j}{P_j} \frac{\partial P_j}{\partial \hat{\alpha}}(\hat{\theta}) \\ &= \sum_{j=1}^r \frac{v_j}{P_j} \left(\frac{\left(1 - \exp\left(\lambda\left(1 - e^{K_{j-1}^\beta}\right)\right)\right) \alpha^{1 - \exp\left(\lambda\left(1 - e^{K_{j-1}^\beta}\right)\right)}}{(\hat{\alpha} - 1)^2} \right) \\ &\quad - \sum_{j=1}^r \frac{v_j}{P_j} \left(\frac{\left(1 - \exp\left(\lambda\left(1 - e^{K_j^\beta}\right)\right)\right) \alpha^{1 - \exp\left(\lambda\left(1 - e^{K_j^\beta}\right)\right)}}{(\hat{\alpha} - 1)^2} \right) \end{aligned}$$

$$\begin{aligned}
 L_2(\widehat{\theta}) &= \sum_{j=1}^r \frac{v_j}{P_j} \frac{\partial P_j}{\partial \widehat{\beta}}(\widehat{\theta}) \\
 &= \sum_{j=1}^r \frac{v_j}{P_j} \left(\frac{\left(-\exp\left(\lambda\left(1 - e^{K_{j-1}^\beta}\right)\right)\right) \lambda K_{j-1}^{\beta-1} e^{K_{j-1}^\beta} \alpha^{1-\exp\left(\lambda\left(1 - e^{K_{j-1}^\beta}\right)\right)}}{(\widehat{\alpha} - 1)^2} \right) \\
 &\quad - \sum_{j=1}^r \frac{v_j}{P_j} \left(\frac{\left(-\exp\left(\lambda\left(1 - e^{K_j^\beta}\right)\right)\right) \lambda K_j^{\beta-1} e^{K_j^\beta} \alpha^{1-\exp\left(\lambda\left(1 - e^{K_j^\beta}\right)\right)}}{(\widehat{\alpha} - 1)^2} \right)
 \end{aligned}$$

$$\begin{aligned}
 L_3(\widehat{\theta}) &= \sum_{j=1}^r \frac{v_j}{P_j} \frac{\partial P_j}{\partial \widehat{\lambda}}(\widehat{\theta}) \\
 &= \sum_{j=1}^r \frac{v_j}{P_j} \left(\frac{\alpha^{1-\exp\left(\lambda\left(1 - e^{K_{j-1}^\beta}\right)\right)} \left(-\exp\left(\lambda\left(1 - e^{K_{j-1}^\beta}\right)\right)\right) \left(1 - e^{K_{j-1}^\beta}\right)}{(\widehat{\alpha} - 1)^2} \right) \\
 &\quad - \sum_{j=1}^r \frac{v_j}{P_j} \left(\frac{\alpha^{1-\exp\left(\lambda\left(1 - e^{K_j^\beta}\right)\right)} \left(-\exp\left(\lambda\left(1 - e^{K_j^\beta}\right)\right)\right) \left(1 - e^{K_j^\beta}\right)}{(\widehat{\alpha} - 1)^2} \right)
 \end{aligned}$$

Ainsi, ayant calculé I , J et L , nous pouvons en déduire la valeur de la statistique de test Y^2 . Pour le niveau de signification α , H_0 est acceptée si $Y_n^2(\widehat{\theta}) \leq \chi_{\alpha, r-1}^2$.

4.4 Simulations et Applications

L'évaluation empirique d'un modèle statistique repose sur deux étapes essentielles : les simulations et les applications aux données réelles. Les simulations permettent d'étudier la performance des méthodes utilisées en générant un grand nombre d'échantillons de différentes tailles, tandis que l'application à des données réelles valide la pertinence du modèle dans des contextes pratiques.

Dans cette section, nous examinons la convergence des estimateurs obtenus à travers des simulations numériques, puis nous testons l'efficacité du modèle APTC sur des ensembles de données réelles. Cette approche nous permet de vérifier la faisabilité du test

statistique proposé et de démontrer son adéquation pour modéliser des observations de différents phénomènes.

4.4.1 Simulations

Nous avons généré $M = 10000$ échantillons de différentes tailles ($n = 15, n = 50, n = 100, n = 200$) avec les valeurs des paramètres $\alpha = 1.2, \beta = 0.3, \lambda = 2.5$, à partir de la distribution APTC. À l'aide du logiciel R, nous avons calculé les valeurs des estimateurs du maximum de vraisemblance, de Kolmogorov-Smirnov, d'Anderson-Darling, de Cramer-von-Mises et leurs erreurs quadratiques moyennes correspondantes (Tableaux 13, 14, 15, 16 respectivement).

$M = 10000$		$\hat{\alpha}$	$\hat{\beta}$	$\hat{\lambda}$
$n = 15$	Mean	0.95503	2.00483	0.13740
	SME	0.09899	0.07240	0.06496
$n = 50$	Mean	1.05683	2.08705	0.17080
	SME	0.04910	0.04970	0.03173
$n = 100$	Mean	1.15707	2.17808	0.28680
	SME	0.01966	0.00979	0.01293
$n = 200$	Mean	1.19719	2.20074	0.30740
	SME	0.00494	0.00556	0.00814

Tableau 13: Estimateurs ML et leurs erreurs quadratiques moyennes

$M = 10000$		$\hat{\alpha}$	$\hat{\beta}$	$\hat{\lambda}$
$n = 15$	Mean	1.17470	2.3899	0.1992
	SME	0.06046	0.0536	0.0357
$n = 50$	Mean	1.18986	2.4051	0.2034
	SME	0.02645	0.0380	0.00504
$n = 100$	Mean	1.190739	2.4666	0.3036
	SME	0.0138	0.02535	0.00336
$n = 200$	Mean	1.1986	2.4994	0.3094
	SME	0.00235	0.00443	0.00049

Tableau 14: Estimateurs KS pour les paramètres et leurs erreurs quadratiques moyennes

$M = 10000$		$\hat{\alpha}$	$\hat{\beta}$	$\hat{\lambda}$
$n = 15$	Mean	0.9424	2.1489	0.14546
	SME	0.4913	0.04846	0.00870
$n = 50$	Mean	0.9970	2.19105	0.25776
	SME	0.09978	0.03868	0.00780
$n = 100$	Mean	1.05562	2.39660	0.28148
	SME	0.00979	0.00782	0.00740
$n = 200$	Mean	1.19792	2.49949	0.30249
	SME	0.00556	0.00174	0.00480

Tableau 15: Estimateurs CVM pour les paramètres et leurs erreurs quadratiques moyennes

$M = 10000$		$\hat{\alpha}$	$\hat{\beta}$	$\hat{\lambda}$
$n = 15$	Mean	1.05503	2.281460	0.1489
	SME	0.09899	0.07860	0.1003
$n = 50$	Mean	1.15683	2.29325	0.19105
	SME	0.0766	0.0468	0.0390
$n = 100$	Mean	1.19707	2.31733	0.2666
	SME	0.0494	0.0157	0.01495
$n = 200$	Mean	1.2075	2.49675	0.29994
	SME	0.01941	0.00405	0.00596

Tableau 16: Estimateurs AD pour les paramètres et leurs erreurs quadratiques moyennes

Les résultats numériques montrent que les estimateurs obtenus sont convergents et que les valeurs moyennes des estimateurs sont très proches des vraies valeurs et les erreurs quadratiques moyennes MSE diminuent à mesure que la taille de l'échantillon augmente, ce qui confirme la fiabilité des résultats.

Critère de test

Pour démontrer la faisabilité du test statistique fourni dans ce travail, nous testons l'hypothèse nulle H_0 selon laquelle les échantillons sont tirés de la distribution APTC. À cette fin, les niveaux de signification théoriques $\xi = (0.01, 0.05, 0.1)$ sont comparés à ceux correspondants aux niveaux de signification empiriques pour 10000 échantillons simulés de différentes tailles ($n = 15, 50, 100, 200$). Les résultats sont résumés dans le tableau 17.

n	$\xi = 0.01$	$\xi = 0.1$	$\xi = 0.05$
15	0.01708	0.1341	0.03623
50	0.01624	0.1191	0.04095
100	0.01235	0.1133	0.04801
200	0.01198	0.1042	0.05620

Tableau 17: Khi-deux critique et leurs valeurs empiriques correspondantes

Conformément aux attentes, les valeurs empiriques obtenues présentent une forte concordance avec leurs équivalents théoriques, attestant ainsi de la pertinence et de la faisabilité du test d'ajustement proposé pour la validation de la distribution APTC.

4.4.2 Application

Données d'assurance

Un exemple d'ensembles de données réelles, extraits du site

(<https://www.kaggle.com/datasets/denkuznetz/traffic-accident-prediction>)

sont utilisées pour confirmer les résultats théoriques obtenus dans ce travail. Cet ensemble contient des données conçues pour prédire la fréquence et la gravité des accidents de la route en fonction de divers facteurs affectant l'état des routes, le comportement des conducteurs et les conditions de circulation. Nous avons sélectionné les données suivantes concernant Limitation de vitesse (100,120,60,60,195,120,60,60,60,30,60,100,60,30,60,60,100,60,50,60,100,30,30,50,100,60,100,80,80,80,30,100,60,60,50,120,100,50,60,60,50,60,60,30,30,60,50)

Pour tester l'hypothèse H_0 selon laquelle ces données suivent la distribution APTC, les utilisateurs recourent généralement aux critères classiques de sélection pour choisir le modèle approprié parmi les alternatives possibles.

Dans notre cas, on applique la statistique de test fournie ci-dessus pour vérifier si les données citées peuvent être modélisées par la distribution APTC. Nous avons regroupé les observations en $r = 5$ intervalles. Les valeurs qui composent de la statistique de test Y^2 sont obtenues comme suit :

$$L(-0.54094; 0.161064; -0.022981)$$

$$I_n(\hat{\theta}) = \begin{pmatrix} 0.84974 & 0 & 0 \\ 0 & 0.080735 & -0.208612 \\ 0 & -0.208612 & 0.065196 \end{pmatrix}$$

$$J(\hat{\theta}) = \begin{pmatrix} -1.17682 & 0.12412 & 0.03218 \\ 0.12412 & 0.83851 & -0.02321 \\ 0.032187 & -0.02321 & 0.89853 \end{pmatrix}$$

Nous obtenons $Y^2 = 3.2219$, donc pour un seuil de signification $\xi = 0.05$ la valeur critique du khi-deux est $X_{r-1}^2 = 9.4877$. On peut conclure que la distribution *APTC* décrit convenablement ces observations.

Conclusion

L'étude menée à travers des simulations numériques et l'application aux données réelles a permis de valider les résultats théoriques obtenus dans ce travail et la flexibilité de la distribution *APTC* dans la modélisation des phénomènes de durée de vie dans divers domaines.

Conclusion et Perspectives

Cette étude a examiné diverses techniques de généralisation des distributions exponentielle et Weibull, ainsi que les approches statistiques existantes en la matière. Nous avons concentré notre attention sur les distributions de risques concurrents, qui revêtent une importance cruciale dans différents domaines tels que la fiabilité, l'analyse de survie et les assurances, Une nouvelle proposition de modèle a été avancée, à savoir le modèle Alpha-Power-Transformé Chen, et ses déclinaisons liées aux temps de vie .

L'examen statistique approfondi réalisé pour chaque modèle a permis d'évaluer leur adéquation aux données empiriques, assurant ainsi leur pertinence et leur efficacité dans la modélisation de phénomènes complexes. Afin d'accomplir cela, des tests de validation spécifiques ont été conçus et les paramètres inconnus ont été estimés en utilisant différentes méthodes. Des simulations numériques approfondies ont été effectuées pour évaluer l'efficacité des estimateurs et la résistance des critères de test suggérés.

Cette étude peut être étendue aux cas où les observations sont censurées, ce qui correspond à des situations fréquentes en analyse de survie. Par ailleurs, d'autres modèles peuvent être construits à l'aide des nouvelles méthodes de génération de distributions, permettant d'enrichir la flexibilité de l'ajustement aux données réelles.

Bibliographie

- [1] Alzaghal, A., Famoye, F., & Lee, C. (2013). Exponentiated TX family of distributions with some applications. *International Journal of Statistics and probability*, 2(3), 31.
- [2] Abbas, U. F., Ahmed, A., & Mukhtar, U. (2022). Bayesian estimation of four parameters additive chen-weibull distribution. *Fudma Journal Of Sciences*, 6(1), 181-190.
- [3] Afify, A. Z., Gemeay, A. M., & Ibrahim, N. A. (2020). The heavy-tailed exponential distribution: risk measures, estimation, and application to actuarial data. *Mathematics*, 8(8), 1276.
- [4] Alzaatreh, A., Lee, C., & Famoye, F. (2013). A new method for generating families of continuous distributions. *Metron*, 71(1), 63-79.
- [5] Ali, R. A. H. (2012). Fitting a two parameters of weibull distribution using goodness of fit tests. *Al-Mustansiriyah J. Sci*, 23(6), 137-148.
- [6] Anderson, T. W., & Darling, D. A. (1954). A test of goodness of fit. *Journal of the American statistical association*, 49(268), 765-769.
- [7] An, K. (1933). Sulla determinazione empirica di una legge didistribuzione. *Giorn Dell'inst Ital Degli Att*, 4, 89-91.
- [8] Alghamedi, A., Dey, S., Kumar, D., & Dobbah, S. A. (2020). A new extension of extended exponential distribution with applications. *Annals of Data Science*, 7(1), 139-162.

-
- [9] Berger, J. O., & Sun, D. (1993). Bayesian analysis for the poly-Weibull distribution. *Journal of the American Statistical Association*, 88(424), 1412-1418.
- [10] Bousquet, N., Bertholon, H., & Celeux, G. (2006). An alternative competing risk model to the Weibull distribution for modelling aging in lifetime data analysis. *Life-time Data Analysis*, 12, 481-504
- [11] Bourguignon, M., Silva, R. B., & Cordeiro, G. M. (2014). The Weibull-G family of probability distributions. *Journal of data science*, 12(1), 53-68.
- [12] Chen, Z. (2000). A new two-parameter lifetime distribution with bathtub shape or increasing failure rate function. *Statistics & Probability Letters*, 49(2), 155-161.
- [13] Cordeiro, G. M., Ortega, E. M. M., & Silva, G. (2012). The beta extended Weibull family. *Journal of Probability and Statistical Science*, 10(10), 15-40.
- [14] Darling, D. A. (1957). The Kolmogorov-smirnov, Cramer-von Mises tests. *The annals of mathematical statistics*, 823-838.
- [15] De Gusmao, F. R., Ortega, E. M., & Cordeiro, G. M. (2011). The generalized inverse Weibull distribution. *Statistical Papers*, 52, 591-619.
- [16] Evans, J. W. (1989). Two-and three-parameter Weibull goodness-of-fit tests (Vol. 493). US Department of Agriculture, Forest Service, Forest Products Laboratory.
- [17] Freels, J. K., Timme, D. A., Pignatiello, J. J., Warr, R. L., & Hill, R. R. (2019). Maximum likelihood estimation for the poly-Weibull distribution. *Quality Engineering*, 31(4), 545-552.
- [18] H. Linhart, W. Zucchini, *Model Selection*, John Wiley & Sons, New York, (1986). 5
- [19] <https://www.kaggle.com/datasets/marshalpatel3558/diabetes-prediction-dataset-legit-dataset>.
- [20] <https://www.kaggle.com/datasets/devansodariya/road-accident-united-kingdom-uk-dataset>

- [21] <https://www.kaggle.com/datasets/denkuznetz/traffic-accident-prediction>
- [22] Kabiru Abidemi, A., & Abiodun, A. A. (2023). Exponentially Generated Modified Chen Distribution with Applications to Lifetime Dataset. *Journal of Probability and Statistics*, 2023(1), 4458562.
- [23] Kovács, Á. M., & Taylor, H. M. (2016, January). COMPETING RISKS WEIBULL MODEL: PARAMETER ESTIMATES AND THEIR ACCURACY. In *Annales Universitatis Scientiarum Budapestinensis de Rolando Eötvös Nominatae. Sectio Computatorica* (Vol. 45).
- [24] Mahdavi, A., & Kundu, D. (2017). A new method for generating distributions with an application to exponential distribution. *Communications in Statistics-Theory and Methods*, 46(13), 6543-6557.
- [25] Méndez-González, L. C., Rodríguez-Picón, L. A., Pérez-Olguín, I. J. C., & VidalPortilla, L. R. (2023). An additive Chen distribution with applications to lifetime data. *Axioms*, 12(2), 118.
- [26] Mudholkar, G. S., & Srivastava, D. K. (1993). Exponentiated Weibull family for analyzing bathtub failure-rate data. *IEEE transactions on reliability*, 42(2), 299-302.
- [27] Mudholkar, G. S., & Hutson, A. D. (1996). The exponentiated Weibull family: some properties and a flood data application. *Communications in Statistics-Theory and Methods*, 25(12), 3059-3083.
- [28] Murthy, D. P., Xie, M., & Jiang, R. (2004). *Weibull models*. John Wiley & Sons.
- [29] Nadarajah, S., & Haghghi, F. (2011). An extension of the exponential distribution. *Statistics*, 45(6), 543-558.
- [30] Nonaka, Y., & Ishioka, T. (1991). Maximum likelihood estimation of Weibull parameters for two independent competing risk. *IEEE Transactions on Reliability*, 40, 71-74.

- [31] Nikulin, M. S. (1974). Chi-square test for continuous distributions with shift and scale parameters. *Theory of Probability & Its Applications*, 18(3), 559-568.
- [32] Rao, K. C., & Robson, B. S. (1974). A chi-square statistic for goodness-of-fit tests within the exponential family. *Communications in Statistics-Theory and Methods*, 3(12), 1139-1153.
- [33] Sarhan, A. M., Hamilton, D. C., & Smith, B. (2010). Statistical analysis of competing risks models. *Reliability Engineering & System Safety*, 95(9), 953-962.
- [34] Sousa-Ferreira, I., Abreu, A. M., & Rocha, C. (2023). The extended Chen-Poisson lifetime distribution. *REVSTAT-Statistical Journal*, 21(2), 173-196.
- [35] Tarvirdizade B. and Ahmadpour M., A new extension of chen distribution with applications to lifetime data, *Communications in Mathematics and Statistics*. (2019) 9, no. 1, 23–38, <https://doi.org/10.1007/s40304-019-00185-4>, 2-s2.0-85069693397.
- [36] Treidi, W., & Seddik-Ameur, N. (2016). NRR statistic for the extension Weibull distribution. *Global Journal of Pure and Applied Mathematics*, 12(4), 2809-18.
- [37] Xie, M., & Lai, C. D. (1996). Reliability analysis using an additive Weibull model with bathtub-shaped failure rate function. *Reliability Engineering & System Safety*, 52(1), 87-93.
- [38] Xie, M., Tang, Y., & Goh, T. N. (2002). A modified Weibull extension with bathtub-shaped failure rate function. *Reliability Engineering & System Safety*, 76(3), 279-285.