

Ministère de l'enseignement supérieur et la recherche scientifique
وزارة التعليم العالي والبحث العلمي

Université Badji Mokhtar
– Annaba –
Badji Mokhtar University
– Annaba –



جامعة باجي مختار
– عنابة –
Année Universitaire : 2016 / 2017

Faculté des Sciences
Département de Mathématiques
Laboratoire LaPS



THÈSE

Présentée en vue de l'obtention du diplôme de
DOCTORAT EN MATHÉMATIQUES
Option : Modélisation Mathématiques – Probabilités Statistiques

Etude de modèles de durées de vie à temps accéléré

Présentée par
Aidi Khaoula

DIRECTEUR DE THÈSE : SEDDIK AMEUR Nacira PROF., U.B.M. ANNABA

Devant le jury

PRÉSIDENT:	Remita Mohamed Riad	PROF. U.B.M. ANNABA
EXAMINATRICE:	Chadli Assia	PROF. U.B.M. ANNABA
EXAMINATRICE:	Nemouchi Nahima	PROF. U. Mentouri CONSTANTINE
EXAMINATRICE:	Messaci Fatiha	PROF. U. Mentouri CONSTANTINE

TABLE DES MATIÈRES

1	Introduction générale	8
2	Les généralisations du modèle exponentiel	14
2.1	Modèle exponentiel généralisé <i>GEE</i>	15
2.1.1	Présentation du modèle	16
2.1.2	Estimation des paramètres dans le cas de censure droite	19
2.2	Modèle exponentiel linéaire généralisé <i>LE</i>	21
2.2.1	Présentation du modèle	22
2.2.2	Estimation des paramètres dans le cas de censure droite	24
2.3	Modèle Kumaraswamy généralisé exponentiel exponentié <i>Kw – GEE</i>	26
2.3.1	Présentation du modèle	26
2.3.2	Estimation des paramètres pour données censurées	30
2.4	Modèle transformé de la distribution exponentielle linéaire généralisée <i>TGLE</i>	32
2.4.1	Présentation du modèle	33
2.4.2	Estimation des paramètres pour données censurées	35
3	Tests d'ajustement pour données censurées	38
3.1	Introduction	38
3.2	Test de Bagdonavicius et Nikulin	39
3.3	Test d'ajustement pour le modèle <i>GEE</i>	44

3.3.1	Simulation des valeurs du critère de test	46
3.3.2	Exemple de données de survie	48
3.4	Test d'ajustement pour le modèle LE	49
3.4.1	Simulation des valeurs du critère de test	51
3.4.2	Exemple de données de fiabilité	53
3.5	Test d'ajustement pour le modèle $Kw - GEE$	54
3.5.1	Simulation des valeurs du critère de test	56
3.5.2	Application à des données de survie	57
3.6	Test d'ajustement pour le modèle $TGLE$	58
3.6.1	Simulation des valeurs du critère de test	61
3.6.2	Application à des données de survie	64
4	Modèle $AFT - GEE$	66
4.1	Introduction	66
4.2	Le modèle AFT	67
4.3	Construction du modèle $AFT - GEE$	68
4.4	Estimation du maximum de vraisemblance	69
4.5	Test d'ajustement pour le modèle AFT-GEE	70
4.6	Simulation	73
4.6.1	Estimation des paramètres	73
4.6.2	La distribution de la statistique $Y_n^2(\hat{\gamma})$	74
4.6.3	La statistique du test	76
4.7	Exemple de données de fiabilité	76
5	Modèle $AFT - LE$	79
5.1	Construction du modèle $AFT - LE$	79
5.2	Estimation du maximum de vraisemblance	80
5.3	Test d'ajustement pour le modèle AFT-LE	81
5.4	Simulations	84

5.4.1	Estimation des paramètres	84
5.4.2	La distribution de la statistique $Y_n^2(\hat{\gamma})$	85
5.4.3	Statistique du test	87
5.5	Exemple de données de fiabilité	87
6	Modèle $AFT - KwGEE$	90
6.1	Construction du modèle $AFT - KwGEE$	90
6.2	Estimation du maximum de vraisemblance	91
6.3	Test d'ajustement pour le modèle AFT-KwGEE	93
6.4	Simulations	97
6.4.1	Estimation des paramètres	97
6.4.2	La distribution de la statistique $Y_n^2(\hat{\gamma})$	98
6.4.3	La statistique du test	100
6.5	Exemple de données de fiabilité	101
7	Modèle $EKw - GEE$	103
7.1	Introduction	103
7.2	Construction du modèle $EKw - GEE$	104
7.3	Estimation du maximum de vraisemblance pour données censurées	107
7.4	Test d'ajustement pour le modèle $EKw - GEE$	109
7.5	Simulations	114
7.5.1	Estimation des paramètres	114
7.5.2	La distribution de la statistique $Y_n^2(\hat{\gamma})$	115
7.5.3	La statistique du test	117
7.6	Exemple de données de fiabilité	120
8	Conclusion	123

Remerciements

Je rends grâce à Dieu qui m'a donnée la volonté, la patience et le courage pour accomplir ce modeste travail.

Mes remerciements les plus chaleureux à ma directrice de thèse le professeur Seddik Ameur Nacira, pour son implication presque quotidienne et sa grande disponibilité. Elle m'a apportée un soutien moral et scientifique constant. Je lui présente les témoignages de ma sincère reconnaissance.

Mes respectueux remerciements au professeur Remita Med Riad qui m'a fait l'honneur de présider le jury ainsi qu'au professeur Chadli Assia, le professeur Nemouchi Nahima et le professeur Messaci Fatiha pour avoir accepté de faire partie du jury et d'y avoir consacré une partie de leurs temps.

Je souhaiterais exprimer ma gratitude à monsieur Rahmania Nadji pour son accueil chaleureux au Laboratoire de l'université de Lille 1, sa disponibilité à chaque fois que j'ai sollicité son aide, ainsi que pour ses multiples encouragements.

Je remercie également tous les membres du département de mathématiques de l'université Badji Mokhtar Annaba, en particulier ceux du laboratoire de probabilités et statistique LaPS, parce que c'est aussi grâce à eux que ce laboratoire est si agréable et si stimulant.

Mes plus profonds remerciements vont à mes parents. 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.

A titre plus personnel, Je remercie chaleureusement mon mari Hamza, pour la grande patience et ses encouragements. Je tiens à le remercier surtout pour son soutien moral ininterrompu.

A ma sœur Ines , je n'oublie pas mon cœur Mohamed Amir et mon beau-frère Rabie.

A mon frère Ayoub et mon frère Hamdi et ses enfants Oumayma, Iyed et Moataz.

A ma tante Nedjatte et son mari Abdelsadek Abdelhak et ses enfants Sifou, Amine , Abdallah et Mayouma.

A ma tante Aicha, je n'oublie pas ma très cher grand-mère Zakia.

A mes belles-sœurs Leila, Rahima, Habiba et Soraya et bien sûr à mes beaux-frères.

A toute ma famille paternelle et maternelle.

Enfin, je remercie mes beaux-parents pour leur disponibilité toutes les fois que j'ai eu besoin d'eux. Je voudrais également remercier toutes mes plus proches amies.

Les modèles *AFT* ou modèles à temps de vie accéléré sont des modèles où les fonctions d'intérêt sont fonctions de variables dites explicatives représentant les différents facteurs de stress. Ces modèles consistent à réduire ou à augmenter les durées de vie des produits par l'accélération ou la décélération des dégradations provoquant les défaillances dans le but d'obtenir des résultats dans des délais courts. Dans cette thèse, on propose les modèles *AFT* – *GEE*, *AFT* – *LE* et *AFT* – *KwGEE* dont les distributions de base sont respectivement la distribution généralisée exponentielle exponentielle (*GEE*), la distribution exponentielle linéaire (*LE*) et la distribution Kumaraswamy généralisée exponentielle exponentielle *Kw* – *GEE*. Après l'étude des propriétés statistiques, nous construisons des statistiques de tests du type du chi-deux modifié permettant d'ajuster des observations à chacun de ces modèles dans le cas de censure aléatoire droite et de paramètres inconnus. Nous développons aussi les critères de tests d'ajustement pour les distributions de base utilisées dans les modèles *AFT* proposés ainsi que pour la transformée de la distribution linéaire généralisée *TGLE*. Enfin, on introduit une nouvelle distribution, flexible, obtenue par l'exponentiation de la distribution Kumaraswamy exponentielle exponentielle qu'on note *EKw* – *GEE*. En utilisant la même approche que pour les autres modèles, on fournit une statistique de test d'adéquation pour cette distribution.

Mots-clés : Analyse de survie, Données censurées, Modèle généralisé exponentiel exponentié, Modèle exponentiel linéaire, Modèle Kumaraswamy généralisé exponentiel exponentié, Transformée de la distribution linéaire généralisée, Modèle Kumaraswamy exponentiel exponentié, Modèle à temps de vie accéléré, Tests d'ajustement du chi-deux modifié.

Accelerated failure time (*AFT*) models relate survival times to covariates representing different stresses. In order to obtain failures data in a short time, the lifetime of devices are increasing or decreasing. In this thesis, we propose accelerated models *AFT – GEE*, *AFT – LE* and *AFT – KwGEE* where the baseline distributions are the generalized exponential exponentiated (*GEE*), the linear exponential (*LE*), and the Kumaraswamy generalized exponential exponentiated distributions respectively. After studying the statistical properties, we constructed modified chi-squared goodness-of-fit tests for each model when data are right censored and the parameters are unknown. We develop, also, goodness-of-fit test statistics for the baseline distributions of the proposed models and the transmuted generalized exponential linear distribution (*TGLE*). Finally, we introduce a new flexible distribution noted *EKw – GEE* obtained by the exponentiation of the Kumaraswamy generalized exponential exponentiated. Using the same approach as for the other models, we provide a criteria goodness-of-fit test for this distribution.

Keywords : Accelerated failure time model, Censored data, Chi-squared type goodness-of-fit test, generalized exponential exponentiated distribution, linear exponential distribution, Kumaraswamy generalized exponential exponentiated distribution, transmuted generalized exponential linear distribution, Kumaraswamy generalized exponential exponentiated distribution, Survival analysis.

ملخص

نموذج AFT أو نموذج وقت الاخفاق المتسارع هي نماذج تمثل متغيرات تفسيرية لضغوطات مختلفة، هذه النماذج تقوم بتقليل أو بزيادة في مدة حياة المنتج التي يتسارع أو يتباطأ من ضرر سببه الفشل من أجل الحصول على نتائج في وقت قصير.

في هذه الاطروحة نقوم بطرح نماذج AFT-GEE و AFT-LE و AFT-KwGEE ذات توزيعات قاعدية هم على التوالي (GEE) *généralisée exponentielle exponentielle* و *exponentielle* *Kumaraswamy généralisée exponentielle exponentielle* Kw-GEE و *linéaire* (LE).

بعد دراسة الخصائص الاحصائية لهذه النماذج نقوم ببناء اختبار احصائي المعدل كا لضبط الملاحظات على كل هذه النماذج في حالة البيانات المراقبة العشوائية و ببارمترات مجهولة .

كما قمنا أيضا بوضع معايير لائقة بالاختبار من اجل التوزيعات القاعدية المستخدمة في نماذج AFT المقترحة و ايضا من اجل تحويل توزيع الخطي المعمم TGLE و اخيرا نقدم لكم توزيع جديد و مرن الذي تحصلنا عليه عن طريق وضع الاس لتوزيع *Kumaraswamy exponentielle exponentielle* و يرمز له بالرمز EKw-GEE باتباع نفس الطريقة المستخدمة في باقي النماذج و توفير اختبار احصائي ملائم لهذا التوزيع.

كلمات البحث: تحليل البقاء على قيد الحياة، والبيانات للرقابة، نموذج *généralisée exponentielle* *Kumaraswamy généralisée* ، نموذج *exponentielle linéaire* ، نموذج *exponentielle* *Kumaraswamy généralisée* ، تحويل التوزيع الخطي المعمم ، و نموذج *exponentielle exponentielle* *Kumaraswamy* ، نموذج وقت الاخفاق المتسارع، اختبار الضبط المعدل كا.

CHAPITRE 1

Introduction générale

Dans les études de fiabilité, pour optimiser la qualité d'un matériel depuis la phase de conception jusqu'à la phase de production, ensuite proposer les techniques adéquates pour la maintenance du produit, les ingénieurs utilisent les données de retour d'expériences obtenues à partir des tests de vie accélérés (Accelerated Life Testing *ALT*). Dans ces expériences, on met les composants du matériel sous des conditions appelées stress qui influencent sa durée de vie. Les modèles mathématiques utilisés pour décrire et analyser ce type d'observations sont appelés "modèles *AFT*" ou modèles à temps de vie accéléré ou tests de vie accélérés. Ces modèles où les fonctions d'intérêt sont fonctions de variables dites explicatives représentant les différents facteurs de stress, consistent à réduire ou à augmenter les durées de vie des produits par l'accélération ou la décélération des dégradations provoquant les défaillances. Ainsi, on peut estimer la fiabilité du matériel dans des délais courts permettant d'optimiser les performances et la planification de la maintenance de celui-ci.

Les modèles *AFT* sont utilisés dans presque tous les domaines comme par exemple, quand on cherche à évaluer l'impact d'un schéma thérapeutique sur le temps de guérison d'un patient en analyse de survie, ou pour analyser les effets comportementaux de clients dans des études commerciales, ou bien l'âge, le niveau d'éducation, le salaire des individus dans une étude démographique. Dans les modèles *AFT*, les résultats sont facilement interprétables. Par exemple, à partir des résultats d'un essai clinique, on peut estimer l'amélioration de

l'espérance de vie suite à un nouveau traitement par rapport au contrôle. Ces modèles sont proposés comme alternative aux modèles à risques proportionnels de Cox (1972) connus sous le nom de modèles *PH* (proportional hazards) et largement développés en analyse de survie (voir Wei, 1992). Pour les modèles *PH*, l'hypothèse de proportionnalité des risques doit être vérifiée alors qu'il n'y a pas de contraintes pour les modèles *AFT*.

Les modèles *AFT* paramétriques sont construits à partir d'une distribution paramétrique dite distribution de base. Les distributions utilisées dans la littérature statistique sont la distribution classique de Weibull, l'exponentielle, la log-logistique et la log-normale (voir Bagdonavicius et Nikulin, 2002).

Vu les grandes avancées technologiques et la complexité des données de retour d'expérience de systèmes réels des différents secteurs d'une part et les développements des études médicales d'autre part, les modèles *AFT* existants ne sont pas en mesure de décrire toutes les observations. La construction de nouveaux modèles flexibles capables de générer une grande variété de formes de risque fréquemment observées est devenue plus que nécessaire. Ces dernières décennies, les chercheurs ont proposé diverses techniques de généralisation des distributions de probabilité classiques ayant toute forme de risque y compris croissante, décroissante, unimodale et en baignoire. Récemment, Ortega et al. (2009) ont introduit le modèle *AFT* dont la distribution de base est la distribution log-gamma. Goual et Seddik-Ameur (2014) ont étudié le modèle *AFT – WIG* dont la distribution de base est celle de Weibull inverse généralisée (*WIG*). Celle-ci a été introduite par Gusmao et al. (2009) en ajoutant un autre paramètre de forme γ à la distribution Weibull inverse standard .

Comme les résultats de n'importe quelle analyse statistique dépendent du choix de la distribution utilisée, il devient nécessaire de valider le modèle choisi. Jusqu'à la dernière décennie, la technique la plus courante pour ajuster des données à un modèle *AFT* était l'utilisation des tests graphiques des distributions des résidus, tandis que pour choisir entre deux distributions on applique les tests du rapport de vraisemblance, le test de Wald ou les tests de score. Un grand nombre d'articles concernant les modèles *AFT* ont été publiés sans pour autant développer des tests d'ajustement pour ceux-ci (Bagdonavicius and Nikulin

(2002), Kalbeisch and Prentice (2002), Klein and Moeschberger (2003), Antonov and Nikulin (2012)). Ceci est dû essentiellement à la présence de censure ce qui est le cas dans les études de fiabilité et de l'analyse de survie et où les tests d'ajustement classiques ne sont plus adaptés.

Les chercheurs ont commencé à s'intéresser à ce problème en proposant de nouvelles méthodes qui tiennent compte aussi bien de la censure que des paramètres inconnus. Galanova et al. (2012) ont développé des modifications des statistiques classiques de Kolmogorov, Cramer-Von Mises-Smirnov et celle d'Anderson-Darling pour les modèles AFT —exponentiel, AFT —gamma, AFT —Weibull, AFT —Lognormal et AFT —Weibull généralisée. Balakrishnan et al. (2013) ont considéré des statistiques de tests basées sur les résidus. Ils ont étudié les distributions des statistiques de tests classiques pour différents modèles. Bagdonavicius et al. (2013) ont construit un test du type du chi-deux modifié pour les modèles AFT —Weibull, AFT —loglogistique, AFT —lognormal. Goual and Seddik-Ameur (2014) ont fourni un critère de test du type du chi-deux modifié pour valider le modèle AFT —Weibull inverse généralisée. Meideros et al. (2014) ont évalué les performances du test du gradient introduit par Terrell (2002) par rapport au test du rapport de vraisemblance pour les modèles AFT —Weibull, AFT —log-normal et AFT —log-logistique.

Dans ce travail de recherche, nous nous sommes intéressées à la construction de nouveaux modèles AFT basés sur les généralisations de la distribution exponentielle. La distribution exponentielle a des domaines d'application nombreux et diversifiés. En radioactivité, la durée de vie d'un atome radioactif est distribuée selon une loi exponentielle dont le paramètre représente la constante de désintégration. En théorie des files d'attente, elle modélise l'arrivée des clients dans une file. En électronique, elle peut aussi modéliser la durée de vie des composants électroniques qui, en général, ne s'usent pas durant toute la durée de vie du produit dans lequel ils sont installés. Il s'agit par exemple de composants de circuits intégrés de haute qualité comme des transistors, des résistances et des condensateurs. Cette distribution décrit également la longue période où le risque de défaillance est faible, c'est à dire la période qui caractérise la durée de vie utile du produit.

Les généralisations de la distribution exponentielle considérées comme distributions de base pour les modèles AFT , proposés dans cette thèse, sont la distribution généralisée exponentielle exponentielle (GEE), la distribution exponentielle linéaire (LE) et la distribution Kumaraswamy généralisée exponentielle exponentielle $Kw - GEE$. Après l'étude et l'estimation des paramètres inconnus des modèles $AFT - GEE$, $AFT - LE$ et $AFT - KwGEE$, nous construisons des statistiques de tests du type du chi-deux modifié permettant d'ajuster des observations à chacun de ces modèles dans le cas de censure aléatoire droite et de paramètres inconnus. Nous avons aussi développé des tests d'ajustement pour les distributions de base des modèles sus-cités.

D'un autre côté, en combinant deux approches généralisant la distribution de Kumaraswamy (Kumaraswamy, 1980), celle proposée par Lemonte et al. (2013) et celle de Cordeiro et Castro (2011), nous introduisons une nouvelle distribution, flexible, obtenue par l'exponentiation de la distribution $Kw - GEE$ proposée récemment par Mohammed (2014). Cette distribution sera notée par $EKw - GEE$. Pour ces types de généralisations, les paramètres additionnels permettent de modéliser l'aplatissement, l'asymétrie et la forme de la queue des distributions ce qui les rend plus utiles dans les applications. Nous adaptons la même approche que précédemment, pour développer un test d'ajustement du type du chi-deux modifié pour ce modèle ($EKw - GEE$) dans le cas de censure aléatoire droite. Huang et Oluyede (2014) ont proposé la distribution exponentielle de Kumaraswamy-Dagum, Rodriguez et al. (2015) la distribution exponentielle de Kumaraswamy-exponentielle où les distributions de référence sont respectivement la distribution Dagum et l'exponentielle. Les auteurs ont étudié les propriétés statistiques sans toutefois proposer de techniques pour valider ces nouveaux modèles. Ainsi avec la mise en place du test d'ajustement pour la distribution $EKw - GEE$, on peut valider plusieurs sous-modèles tels que la distribution GEE , la distribution $Kw - GEE$, les distributions exponentielles exponentielles (EE) avec différents paramètres, la distribution Kumaraswamy-exponentielle $Kw - Exp$ (Nadarajah et al., 2012), la distribution ($EKw - exp$) proposée par Rodriguez *et al.* (2015) et bien sûr différentes distributions GEE .

Des dizaines de milliers d'échantillons de plusieurs tailles et avec différentes valeurs des paramètres, provenant des distributions étudiées, ont été simulés pour confirmer les résultats théoriques obtenus. Enfin, on a utilisé les statistiques de tests développées dans ce travail pour ajuster aux modèles étudiés plusieurs jeux de données réelles provenant des domaines médical et industriel, montrant ainsi la maniabilité et l'applicabilité de ces tests.

La thèse est structurée de la manière suivante :

Dans le premier chapitre, les différentes généralisations de la distribution exponentielle classique en l'occurrence la distribution généralisée exponentielle exponentielle (GEE), la distribution exponentielle linéaire (LE), la distribution Kumaraswamy généralisée exponentielle exponentielle ($Kw - GEE$) et la transformée de la distribution exponentielle linéaire généralisée ($TGLE$) sont présentées. Les estimateurs du maximum de vraisemblance des paramètres inconnus en présence de censure aléatoire droite sont déterminés par des méthodes itératives numériques.

Etant donné que les tests d'ajustement classiques ne sont pas adaptés pour valider les données censurées, on expose dans le chapitre deux la théorie de tests d'ajustement du type du chi-deux modifié, proposée récemment par Bagdonavicius et Nikulin (2011) pour les modèles paramétriques à données censurées à droite. Ce test est basé sur les estimateurs du maximum de vraisemblance sur les données non regroupées et dont la distribution suit une loi du chi-deux. Comme dans la littérature, les tests utilisés pour valider la distribution GEE , la distribution LE , la distribution $Kw - GEE$ et la distribution $TGLE$ sont des tests graphiques ou ceux du rapport de vraisemblance, ceci nous a motivé à construire des critères de tests pour permettre d'ajuster des observations à chacun de ces modèles. Le calcul des différentes statistiques de test sont développées dans ce chapitre.

Dans le chapitre trois, un modèle à temps de vie accéléré dont la distribution de base est la distribution généralisée exponentielle exponentielle ($AFT - GEE$) est introduit. En utilisant les estimateurs du maximum de vraisemblance des paramètres inconnus de la distribution de base et ceux de la régression, on construit un test d'ajustement du type du chi-deux modifié pour valider un tel modèle.

Quand au quatrième chapitre, il est consacré au modèle $AFT - LE$ dont la distribution de base est la distribution exponentielle linéaire généralisée. Ce modèle utilisé pour les durées de vie à croissance linéaire trouve son application aussi bien en fiabilité qu'en étude médicale. Une analyse statistique complète a été faite. Une statistique de test d'adéquation à ce modèle a été aussi proposée.

Au cinquième chapitre, on propose un modèle $AFT - KwGEE$ basé sur la distribution de Kumaraswamy généralisée ($Kw - GEE$) présentée dans le premier chapitre. La distribution de Kumaraswamy a plusieurs avantages particulièrement en simulations numériques où elle permet de générer des variables aléatoires de plusieurs distributions. La statistique de test du chi-deux modifiée introduite dans le deuxième chapitre a été adaptée au modèle proposé et la forme explicite de tous les éléments du critère de tests a été obtenue.

On termine enfin par l'introduction d'une nouvelle distribution, flexible, obtenue par l'exponentiation de la distribution Kumaraswamy exponentielle exponentiée qu'on note $EKw - GEE$. Après l'étude des caractéristiques statistiques, on détermine les estimateurs du maximum de vraisemblance. En utilisant la même approche que pour les autres modèles, on fournit également une statistique de test d'adéquation pour cette distribution.

D'importantes simulations numériques ont été conduites pour le calcul des estimateurs du maximum de vraisemblance des paramètres et le calcul des critères de tests pour chacun des modèles étudiés. Des applications à des jeux de données réelles provenant de différents domaines ont montré la maniabilité des tests proposés dans ce travail.

CHAPITRE 2

Les généralisations du modèle exponentiel

La modélisation des durées de vie est un problème auquel sont confrontés les analystes des différents domaines d'étude. Par exemple, en fiabilité pour décrire le comportement de la durée de vie du matériel ou de la durée entre deux pannes successives, en médecine pour estimer la durée entre l'apparition d'une maladie et la guérison ou bien la durée de rémission après une intervention chirurgicale, en assurance pour prévoir la durée entre deux sinistres. Les modèles mathématiques utilisés dépendent du type de données et surtout du comportement de la fonction risque ou taux de hasard qui indique la cause des pannes. Si la forme de cette fonction est décroissante, elle décrit les pannes accidentelles ou pannes de jeunesse, si elle croissante, elle représente les pannes de vieillesse ou d'usure, si elle est constante, elle représente la phase de vie utile. La distribution de base utilisée dans la modélisation des durées de vie est la distribution exponentielle. Néanmoins celle-ci dont le taux de hasard est constant, ne peut décrire que des durées de vie d'un système qui ne vieillit pas c'est-à dire pendant la phase de vie utile. Weibull (1951) a proposé la fameuse distribution de Weibull, de taux de hasard monotone, pour calculer la fiabilité d'un matériel réparable, et dont la distribution exponentielle constitue un cas particulier. Les distributions de Weibull sont utilisées aussi en analyse de survie, en démographie et en assurance vie. Seulement ces distributions ne modélisent pas les durées de vie pouvant présenter différentes

causes de panne, ou des durées de rémission de patients ayant plusieurs facteurs vitaux. De nouvelles distributions dont la fonction de hasard est en forme de cloche, ou en forme de baignoire, ont été alors proposées comme la loi Gamma, différentes généralisations de la distribution de Weibull et d'autres familles de lois.

Dans ce chapitre, on s'intéresse à l'étude de quelques généralisations de la distribution exponentielle telles que la distribution généralisée exponentielle exponentielle (*GEE*), la distribution exponentielle linéaire (*LE*), la distribution Kumaraswamy généralisée exponentielle exponentielle (*Kw - GEE*) et la transformée de la distribution exponentielle linéaire généralisée (*TGLE*). Ces distributions à deux et quatre paramètres ont différentes formes de fonction risque et donc pourront modéliser un grand nombre de durées de vie issues de tous les domaines.

2.1 Modèle exponentiel généralisé *GEE*

Les distributions Weibull et gamma sont les plus utilisées pour modéliser les durées de vie et ceci est dû à leurs interprétations physiques et leur flexibilité. Néanmoins, la fonction de fiabilité de la distribution gamma ne peut être exprimée sous une forme explicite si le paramètre de forme n'est pas entier. Tandis que la convergence des estimateurs du maximum de vraisemblance pour la distribution Weibull est faible. Pour contourner ces inconvénients, une nouvelle distribution à deux paramètres ayant les mêmes caractéristiques que celles-ci, a été introduite par Gupta et Kundu (1999). Cette distribution est appelée distribution exponentielle exponentielle généralisée (*GEE*). Ce modèle peut être considéré comme alternative au modèle gamma et au modèle Weibull dans de nombreuses situations. La répartition *GEE* constitue un sous-modèle de la distribution de Weibull exponentielle introduite par Mudholkar et Srivastava (1993), Mudholkar et al. (1995), Mudholkar et Hutson (1996) et Kim et Song (2011).

Les caractéristiques de la distribution et les différentes méthodes d'estimation ont été développées par Gupta et Kundu (2001, 2007), Raqab et Ahsanullah (2001), Raqab (2002).

Pradhan et Kundu (2008) ont étudié le cas de censure progressive alors que Sarhan (2007) a considéré des modèles à risques concurrents basés sur cette distribution. Jaheen (2004), Raqab et Madi (2005), Kundu et Gupta (2008) et Achcar et al.(2015) se sont intéressés à l'analyse bayésienne de cette distribution. Rao (2012) a considéré l'estimation de la fiabilité de la résistance d'un modèle multicomposant lorsque les variations des composants sont décrits par cette distribution avec différents paramètres de forme. Cette distribution est encore au centre d'intérêt de beaucoup de chercheurs (Zheng, 2002, Singh et al., 2008, Kim et Song, 2010, Chen et Lio, 2010 et Yarmohammadi, 2010, Valiollahi et al., 2017, Lee et al., 2017, Abd el Monen et Jaheen, 2017).

Vu l'intérêt de cette distribution et étant donné que dans la majorité des articles, les tests d'ajustement en présence de censure n'ont pas été traités, nous nous proposons de construire un test d'ajustement pour ce modèle. Dans cette section, on commence par présenter les caractéristiques de la distribution exponentielle exponentielle généralisée GEE. Ensuite, on utilise les méthodes numériques pour calculer les estimateurs du maximum de vraisemblance des paramètres. L'étude des tests sera exposée au chapitre deux.

2.1.1 Présentation du modèle

La distribution exponentielle généralisée proposée par Gupta et Kundu (1999) de paramètre de forme α et de paramètre d'échelle λ notée $GEE(\gamma)$, a pour fonction densité

$$f(t; \alpha, \lambda) = \alpha \lambda (1 - e^{-\lambda t})^{\alpha-1} e^{-\lambda t}, \quad t > 0$$

Où $\gamma = (\alpha, \lambda)^T$

La fonction de distribution cumulative est :

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

La fonction de survie et le taux de hasard sont respectivement :

$$S(t; \gamma) = 1 - (1 - e^{-\lambda t})^\alpha$$

et

$$h(t; \gamma) = \frac{\alpha \lambda (1 - e^{-\lambda t})^{\alpha-1} e^{-\lambda t}}{1 - (1 - e^{-\lambda t})^\alpha}$$

Le taux de hasard cumulé est donné par :

$$H(t; \gamma) = -\ln S(t; \alpha, \lambda) = -\ln (1 - (1 - e^{-\lambda t})^\alpha)$$

Pour simuler une variable provenant de la distribution *GEE*, nous utilisons l'équation non linéaire

$$t = -\frac{1}{\lambda} \ln \left(1 - Z^{\frac{1}{\alpha}} \right)$$

Où Z suit la distribution uniforme sur $[0, 1]$.

Quelques exemples des formes de la densité *GEE* sont présentés respectivement dans les figures 1 et 2, et le taux de hasard dans les figures 3 et 4.

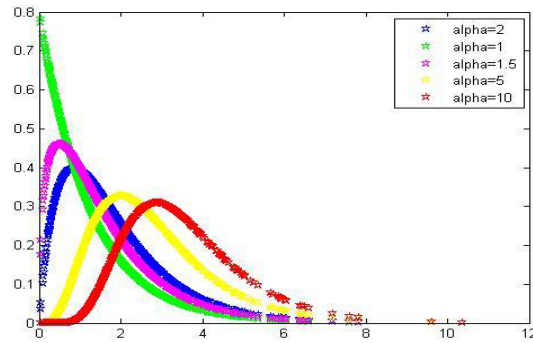


Fig. 1. Densité de la distribution *GEE* pour $\lambda = 0.8$ et $\alpha = (0.8, 1.5, 2, 5, 10)$.

Nous remarquons que :

- Si le paramètre de forme $\alpha = 1$, on retrouve la distribution exponentielle,
- Si $\alpha > 1$, la distribution est unimodale,
- Si $\alpha < 1$ la distribution est décroissante.

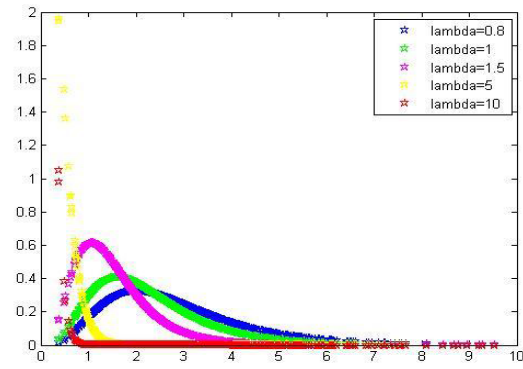


Fig. 2. Densité de la distribution *GEE* pour $\alpha = 3$ et $\lambda = (0.8, 1, 1.5, 5, 10)$

Nous remarquons que :

- Si le paramètre d'échelle $\lambda > 1$, la distribution est décroissante.
- Si $\lambda \leq 1$ la distribution est unimodale.

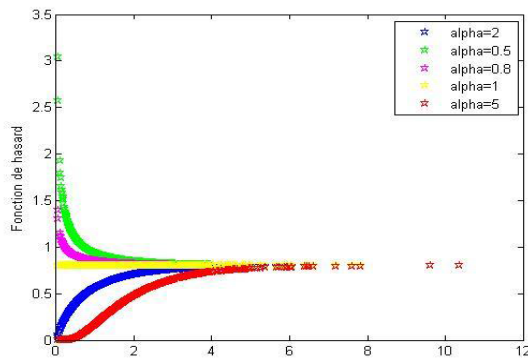


Fig. 3. Taux de hasard de la distribution *GEE* pour $\lambda = 0.8$ et $\alpha = (0.5, 0.8, 1, 2, 5)$

Nous remarquons que :

- lorsque le paramètre de forme $\alpha > 1$, la fonction du taux de panne est croissante,
- décroissante lorsque $\alpha < 1$,
- et bien sûr constante quand $\alpha = 1$ (c'est la loi exponentielle).

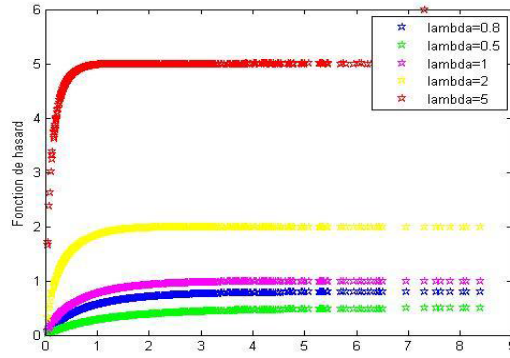


Fig. 4. Taux de hazard de la distribution *GEE* pour $\alpha = 2$ et $\lambda = (0.5, 0.8, 1, 2, 5)$

Nous remarquons que la forme du taux de hazard ne change pas avec la valeur de λ .

2.1.2 Estimation des paramètres dans le cas de censure droite

Nous nous intéressons au calcul des estimateurs du maximum de vraisemblance des paramètres du modèle exponentiel exponentielle généralisé en cas de censure aléatoire droite. Les différentes méthodes d'estimations bayésienne et classiques ont été largement développées par Gupta et Kundu (2001, 2007).

Estimation du maximum de vraisemblance

Soit $T_1, T_2, \dots, T_n$ un échantillon de variables aléatoires censurées indépendantes de loi exponentielle généralisée $GEE(\gamma)$, de vecteur de paramètre $\gamma = (\alpha, \lambda)^T$. Chaque individu i est supposé avoir une durée de vie t_i et un temps de censure C_i , où T_i et C_i sont des variables aléatoires indépendantes. Supposons que les données sont constituées de n observations $t_i = \min(T_i, C_i)$ pour $i = 1, \dots, n$. La distribution de C_i ne dépend pas des paramètres inconnus de T_i .

La fonction de vraisemblance s'écrit :

$$l(t, \gamma) = \prod_{i=1}^n f(t_i)^{\delta_i} S(t_i)^{1-\delta_i}, \quad \delta_i = 1_{T_i < C_i}$$

δ_i est l'indicateur de censure.

La fonction log-vraisemblance du modèle :

$$L(t, \gamma) = \sum_{i=1}^n \delta_i \ln f(t_i) + \sum_{i=1}^n (1 - \delta_i) \ln S(t_i)$$

alors l'équation de la log-vraisemblance devient :

$$\begin{aligned} L(t, \gamma) &= \sum_{i=1}^n \delta_i \ln h(t_i) + \sum_{i=1}^n \ln S(t_i) \\ &= \sum_{i=1}^n \delta_i [\ln f(t_i) - \ln S(t_i)] + \sum_{i=1}^n \ln S(t_i) \end{aligned}$$

Ce que donne :

$$\begin{aligned} L(t, \gamma) &= \sum_{i=1}^n \delta_i [\ln \alpha + \ln \lambda - \lambda t_i + (\alpha - 1) \ln (1 - e^{-\lambda t_i}) - \ln (1 - (1 - e^{-\lambda t_i})^\alpha)] \\ &\quad + \sum_{i=1}^n \ln (1 - (1 - e^{-\lambda t_i})^\alpha) \end{aligned}$$

Les fonctions de score pour les paramètres α et λ sont données par :

$$\frac{\partial L}{\partial \alpha} = \sum_{i=1}^n \delta_i \left[\frac{1}{\alpha} + \frac{\ln (1 - e^{-\lambda t_i})}{1 - (1 - e^{-\lambda t_i})^\alpha} \right] - \sum_{i=1}^n \frac{(1 - e^{-\lambda t_i})^\alpha \ln (1 - e^{-\lambda t_i})}{1 - (1 - e^{-\lambda t_i})^\alpha}$$

$$\frac{\partial L}{\partial \lambda} = \sum_{i=1}^n \delta_i \left[\frac{1}{\lambda} + \frac{t_i (\alpha e^{-\lambda t_i} + (1 - e^{-\lambda t_i})^\alpha - 1)}{(1 - e^{-\lambda t_i}) (1 - (1 - e^{-\lambda t_i})^\alpha)} \right] - \sum_{i=1}^n \frac{\alpha t_i e^{-\lambda t_i} (1 - e^{-\lambda t_i})^{\alpha-1}}{1 - (1 - e^{-\lambda t_i})^\alpha}$$

Pour résoudre ce système de fonctions non linéaires, on fait appel aux méthodes numériques, telles que la méthode de Newton Raphson, l'algorithme Barzilai-Borwein (**BB**) ou d'autres algorithmes.

Simulations

On considère le modèle *GEE*. Les données ont été simulées $N = 10,000$ fois, avec les tailles des échantillons $n_1 = 15$, $n_2 = 25$, $n_3 = 50$, $n_4 = 130$, $n_5 = 350$ et $n_6 = 500$, et les valeurs des paramètres $\alpha = 3$, $\lambda = 2$. En utilisant le logiciel *R* et l'algorithme Barzilai-Borwein (**BB**) (Ravi, 2009) pour les calculs des moyennes des valeurs simulées

des estimateurs du maximum de vraisemblance $\hat{\alpha}, \hat{\lambda}$ des paramètres et leurs erreurs quadratiques moyennes (notées S.M.E), nous obtenons les résultats présentés dans la table 1.

$N = 10.000$	$n_1 = 15$	$n_2 = 25$	$n_3 = 50$	$n_4 = 130$	$n_5 = 350$	$n_6 = 500$
$\hat{\alpha}$	2.8594	2.8897	2.9585	2.9698	2.9898	3.0001
$S.M.E$	0.0057	0.0043	0.0031	0.0022	0.0011	0.0009
$\hat{\lambda}$	1.9541	1.9675	1.9742	1.9869	1.9975	2.0003
$S.M.E$	0.0045	0.0034	0.0022	0.0015	0.0009	0.0006

Tab. 1. Les valeurs moyennes des EMV ($\hat{\alpha}, \hat{\lambda}$) des paramètres et leurs erreurs quadratiques moyennes pour données censurées.

Les résultats des simulations confirment le fait que pour les modèles réguliers, les estimateurs du maximum de vraisemblance sont convergents.

2.2 Modèle exponentiel linéaire généralisé LE

Le modèle exponentiel linéaire (LE) est une généralisation de la distribution exponentielle et de celle Rayleigh. La distribution LE , utilisée pour décrire les phénomènes de fonction risque linéairement croissante, attire l'intérêt de plusieurs auteurs. Broadbent (1958), utilise la distribution LE pour décrire le service des bouteilles de lait qui sont remplies dans une laiterie, distribuées aux clients et renvoyées à la laiterie. Le modèle exponentiel linéaire a également été utilisé par Carbone et al. (1967) pour étudier le schéma de survie des patients atteints de myélome plasmocytaire. Les données censurées de type 2 sont utilisées par Bain (1978) pour discuter des estimations minimales des paramètres α et λ et par Pandey et al. (1993) pour déterminer les estimateurs de Bayes de (α, λ) .

La distribution exponentielle linéaire a de nombreuses applications en analyse de survie et en fiabilité. C'est une distribution qui intéresse beaucoup de chercheurs, Al-Khedhairi (2008) a développée l'estimation bayésienne et celle du maximum de vraisemblance pour les

données groupées, Mahmoud et Alam (2010), Mahmouda et al. (2013), Tian et al. (2014) et d'autres ont proposé des généralisations pour cette distribution telle que la distribution exponentielle linéaire généralisée (GLE), la distribution exponentielle linéaire généralisée exponentielle ($EGLE$), la nouvelle distribution exponentielle linéaire généralisée ($NGLE$). Leurs études concernent les propriétés statistiques et les méthodes d'estimation.

2.2.1 Présentation du modèle

On dit qu'une variable aléatoire T suit la distribution exponentielle linéaire généralisée avec α le paramètre de forme et λ le paramètre d'échelle notée $LE(\gamma)$, si sa fonction densité est

$$f(t; \gamma) = (\alpha + \lambda t) \exp \left(- \left[\alpha t + \frac{\lambda t^2}{2} \right] \right) \quad t \geq 0$$

Où $\gamma = (\alpha, \lambda)^T$.

La fonction de distribution cumulative est :

$$F(t; \gamma) = 1 - \exp \left(- \left[\alpha t + \frac{\lambda t^2}{2} \right] \right)$$

La fonction de survie et le taux de hasard sont respectivement :

$$S(t; \gamma) = \exp \left(- \left[\alpha t + \frac{\lambda t^2}{2} \right] \right)$$

et

$$h(t; \gamma) = \alpha + \lambda t$$

Le taux de hasard cumulé est donné par

$$H(t; \gamma) = \alpha t + \frac{\lambda t^2}{2}$$

Pour simuler les variables aléatoires $t \sim LE(\gamma)$, on utilise l'équation non linéaire :

$$t = -\frac{\alpha}{\lambda} + \frac{1}{\lambda} \sqrt{\alpha^2 - 2\lambda \ln(1 - Z)}$$

Où Z suit la distribution uniforme sur $[0, 1]$.

Des graphes de la densité de la distribution LE sont présentés respectivement dans les figures 5 et 6, et ceux du taux de hasard dans les figures 7 et 8.

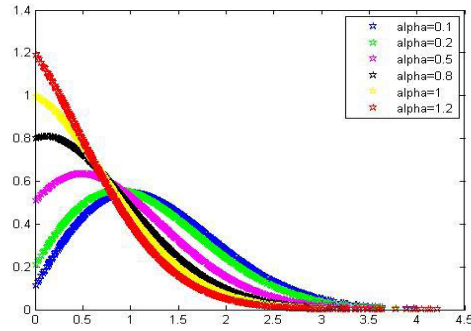


Fig. 5. Densité de la distribution LE pour $\lambda = 0.8$ avec $\alpha = (0.1, 0.2, 0.5, 0.8, 1, 1.2)$.

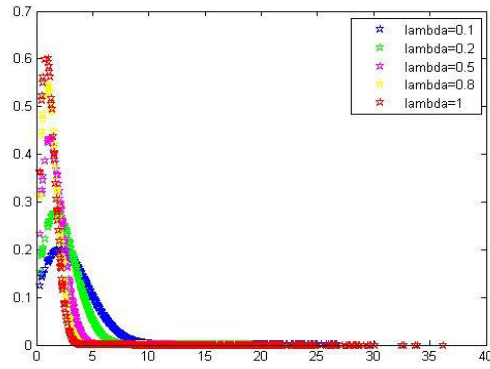


Fig. 6. Densité de la LE pour $\alpha = 3$ avec $\lambda = (0.1, 0.2, 0.5, 0.8, 1)$

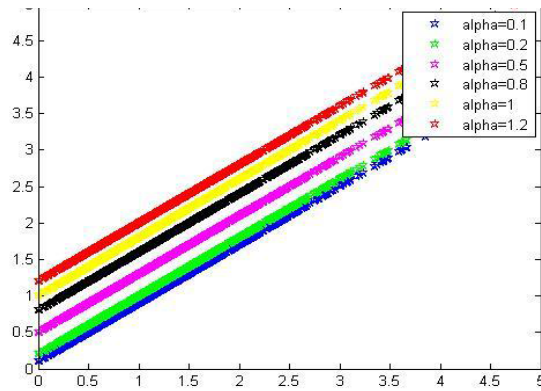


Fig. 7. Taux de hasard de la distribution LE pour $\lambda = 0.8$ avec $\alpha = (0.1, 0.2, 0.5, 0.8, 1, 1.2)$.

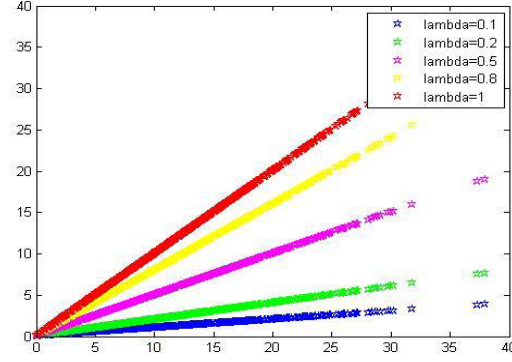


Fig. 8. Taux de hasard de la distribution LE pour $\alpha = 2$ avec $\lambda = (0.1, 0.2, 0.5, 0.8, 1)$.

2.2.2 Estimation des paramètres dans le cas de censure droite

Dans cette section, nous présentons l'estimation des paramètres par la méthode du maximum de vraisemblance en cas de censure aléatoire droite. Les estimateurs trouvés seront utilisés pour tester la validité du modèle.

Estimation du maximum de vraisemblance

On considère un échantillon aléatoire censuré à droite $T_1, T_2, \dots, T_n$, et où T_i est distribué selon une loi exponentielle linéaire généralisée LE avec le vecteur de paramètres $\gamma = (\alpha, \lambda)^T$. Les observations consistent en $t_i = \min(T_i, C_i)$ pour $i = 1, \dots, n$ où T_i sont les taux de pannes et C_i les taux de censures. Ils sont supposés indépendants.

L'équation de la log-vraisemblance pour ce modèle LE est alors donnée par :

$$L(t, \gamma) = \sum_{i=1}^n \delta_i \ln(\alpha + \lambda t_i) - \alpha \sum_{i=1}^n t_i - \frac{\lambda}{2} \sum_{i=1}^n t_i^2$$

δ_i est l'indicateur de censure.

Et les fonctions de score sont obtenues comme suit :

$$\frac{\partial L}{\partial \alpha} = \sum_{i=1}^n \frac{\delta_i}{\alpha + \lambda t_i} - \sum_{i=1}^n t_i$$

$$\frac{\partial L}{\partial \lambda} = \sum_{i=1}^n \frac{\delta_i t_i}{\alpha + \lambda t_i} - \frac{\sum_{i=1}^n t_i^2}{2}$$

Les estimateurs ne peuvent pas être déduit d'une manière explicite, alors on utilise n'importe quelle méthode itérative. Dans notre travail, nous avons appliqué l'algorithme BB solve et le logiciel R.

Simulations cas de données censurées

Pour calculer les estimateurs du maximum de vraisemblance $\hat{\alpha}$ et $\hat{\lambda}$, nous effectuons $N = 10,000$ simulations d'échantillons de données censurées du modèle $LE(\alpha = 3, \lambda = 2)$, de tailles respectives $n_1 = 15, n_2 = 25, n_3 = 50, n_4 = 130, n_5 = 350$ et $n_6 = 500$. Les valeurs moyennes de $\hat{\alpha}$ et $\hat{\lambda}$ et leurs erreurs quadratiques moyennes sont réunies dans la table 2.

$N = 10.000$	$n_1 = 15$	$n_2 = 25$	$n_3 = 50$	$n_4 = 130$	$n_5 = 350$	$n_6 = 500$
$\hat{\alpha}$	2.7895	2.8647	2.9874	2.9954	2.9978	2.9999
$S.M.E$	0.0068	0.0053	0.0048	0.0036	0.0028	0.0009
$\hat{\lambda}$	1.5843	1.6572	1.7489	1.8854	1.9541	1.9954
$S.M.E$	0.0064	0.0055	0.0045	0.0037	0.0021	0.0018

Tab. 2. Estimateurs du maximum de vraisemblance $(\hat{\alpha}, \hat{\lambda})$ des paramètres et leurs erreurs quadratiques moyennes.

Nous remarquons que les erreurs quadratiques sont très faibles et donc les estimateurs du maximum de vraisemblance sont convergents.

2.3 Modèle Kumaraswamy généralisé exponentiel exponentié $Kw - GEE$

En raison de ses multiples avantages notamment en matière de simulation, de nouveaux modèles statistiques reposent sur la distribution de Kumaraswamy (1980) :

$$Kw(t) = 1 - (1 - t^\sigma)^\theta$$

où $0 < t < 1$, et $\theta, \sigma > 0$ sont des paramètres de forme. Cordeiro et Castro (2011) ont proposé une nouvelle famille de distributions, dérivée d'une distribution cumulative continue $G(t)$ par

$$F_1(t) = 1 - \{1 - G^\sigma(t)\}^\theta$$

appelée la distribution de Kumaraswamy généralisée. Avec en plus deux paramètres de forme θ et σ et ceux de la distribution de base $G(t)$, les auteurs ont montré que ces distributions sont très flexibles. Cette nouvelle famille est capable de modéliser plus de données réelles que les modèles classiques. En utilisant la distribution de base une exponentielle exponentiée généralisée GEE , Mohammed (2014) a proposé la distribution Kumaraswamy généralisée exponentielle exponentiée et notée $Kw - GEE$. Grâce aux différentes valeurs prises par ses paramètres de forme, ce modèle peut contrôler l'asymétrie, la dispersion et la queue de la distribution.

2.3.1 Présentation du modèle

Caractérisée par $\lambda > 0$, le paramètre d'échelle et $\alpha, \sigma, \theta > 0$, les paramètres de forme, la fonction de répartition du modèle Kumaraswamy généralisé exponentiel exponentié est donnée par :

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

La fonction de densité de probabilité est

$$f(t) = \alpha\sigma\lambda\theta e^{-\lambda t}(1 - e^{-\lambda t})^{\alpha\sigma-1} \{1 - (1 - e^{-\lambda t})^{\alpha\sigma}\}^{\theta-1}$$

La fonction de survie et le taux de hasard sont respectivement :

$$S(t; \gamma) = \{1 - (1 - e^{-\lambda t})^{\alpha\sigma}\}^{\theta}$$

et

$$h(t; \gamma) = \frac{\alpha\sigma\lambda\theta e^{-\lambda t}(1 - e^{-\lambda t})^{\alpha\sigma-1}}{1 - (1 - e^{-\lambda t})^{\alpha\sigma}}$$

Le taux de hasard cumulé est

$$H(t; \gamma) = -\theta \ln(1 - (1 - e^{-\lambda t})^{\alpha\sigma})$$

Pour générer les variable aléatoires d'une loi de $Kw - GEE$, on utilise l'équation suivante :

$$t = -\frac{1}{\lambda} \ln \left(1 - \left[\left(1 - (1 - Z)^{1/\theta} \right)^{1/\sigma} \right]^{1/\alpha} \right)$$

Où Z suit la distribution uniforme sur $[0, 1]$.

Quelques graphes de la densité de la distribution $Kw - GEE$ sont présentés respectivement dans les figures 9,10, 11 et 12 et ceux du taux de hasard dans les figures 13, 14, 15 et 16.

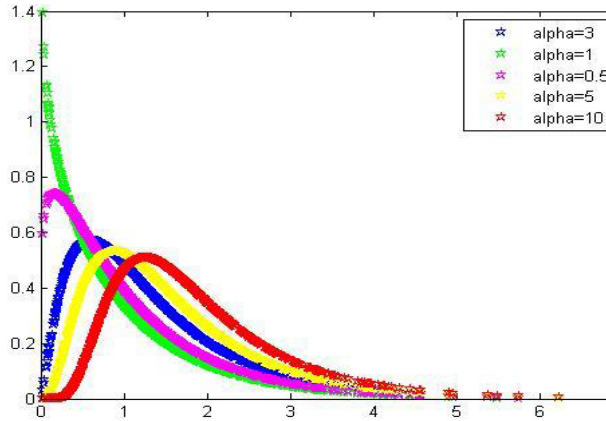


Fig. 9. Densité de la distribution $Kw - GEE$ pour $\lambda = 2$, $\sigma = 0.82$ et $\theta = 0.5$ et $\alpha = (0.5, 1, 3, 5, 10)$.

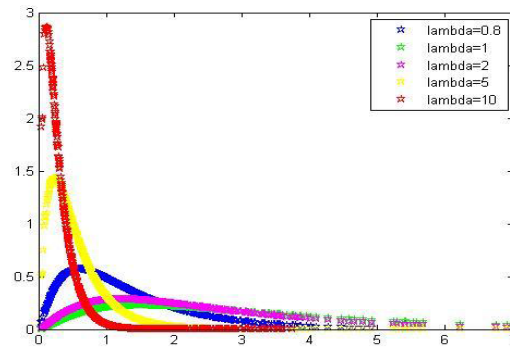


Fig. 10. Densité de la $Kw - GEE$ pour $\alpha = 3, \sigma = 0.82$ et $\theta = 0.5$ avec $\lambda = (0.5, 1, 3, 5, 10)$

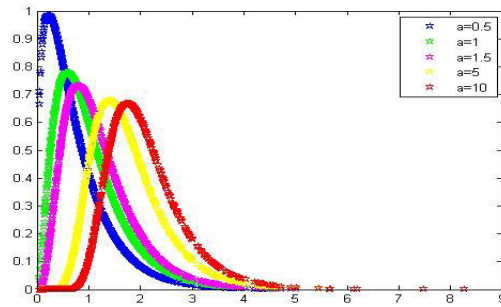


Fig. 11. Densité de la distribution $Kw - GEE$ pour $\alpha = 3, \lambda = 2$ et $\theta = 0.82$ avec $\sigma = (0.5, 1, 1.5, 5, 10)$

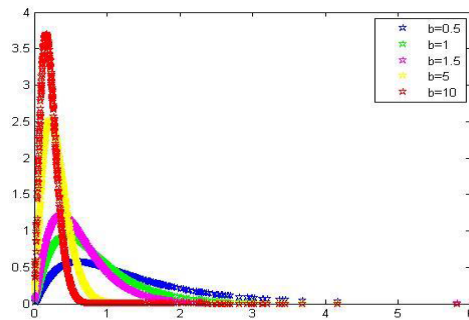


Fig. 12. Densité de la distribution $Kw - GEE$ pour $\alpha = 3, \lambda = 2$, et $\sigma = 0.82$ avec $\theta = (0.5, 1, 1.5, 5, 10)$.

2.3. Modèle Kumaraswamy généralisé exponentiel exponentié $Kw - GEE$

Comme on le voit, la densité de probabilité de cette distribution est décroissante ou bien unimodale plus ou moins aplatie et avec différentes asymétries.

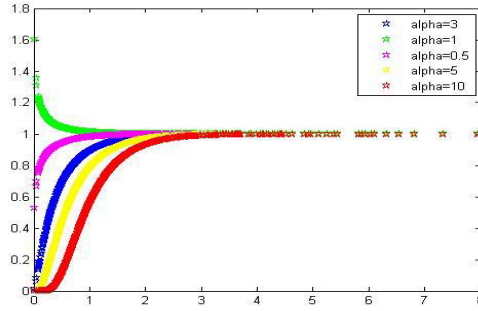


Fig. 13. Taux de hasard de la distribution $Kw - GEE$ pour $\lambda = 2$ et $\alpha = (0.5, 1, 3, 5, 10)$.

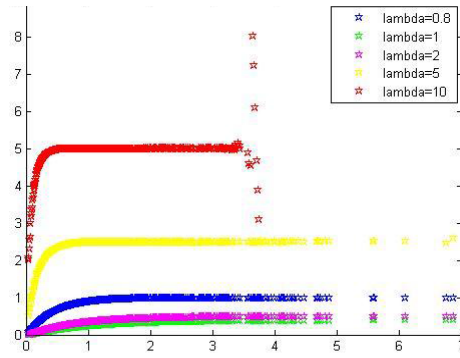


Fig. 14. Taux de hasard de la distribution $Kw - GEE$ pour $\alpha = 3$ et $\lambda = (0.8, 1, 2, 5, 10)$

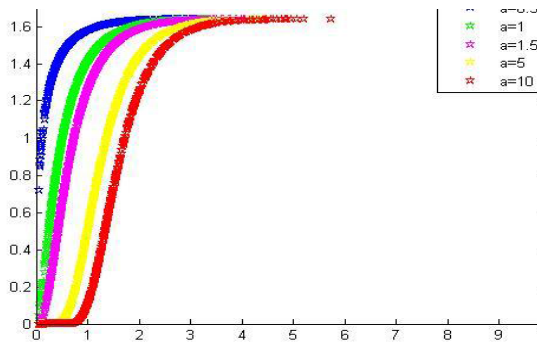


Fig. 15. Taux de hasard de la distribution $Kw - GEE$ pour $\alpha = 3$ et $\sigma = (0.5, 1, 1.5, 5, 10)$.

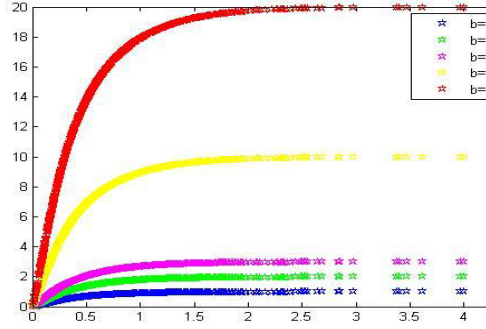


Fig. 16. Taux de hasard de la distribution $Kw - GEE$ pour $\alpha = 3$ et $\theta = (0.5, 1, 1.5, 5, 10)$.

La fonction risque est soit décroissante, soit croissante plus ou moins rapidement.

2.3.2 Estimation des paramètres pour données censurées

De même que pour toutes les distributions exposées dans ce travail, nous avons besoin de déterminer les estimateurs du maximum de vraisemblance des paramètres du modèle $Kw - GEE$ dans le cas de censure aléatoire droite. Mohammed (2014) a étudié les caractéristiques de cette distribution et n'a considéré les estimateurs du maximum de vraisemblance que pour les données complètes.

Estimation du maximum de vraisemblance

Considérons un échantillon de données censurées aléatoirement à droite. L'échantillon est issu de la loi $Kw - GEE(\gamma)$, de vecteur de paramètre $\gamma = (\alpha, \lambda, \sigma, \theta)^T$. Les observations sont représentées par les variables $t_i = \min(T_i, C_i)$ pour $i = 1, \dots, n$. T_i et C_i sont respectivement les taux de panne et les taux de censure. La censure est supposée non informative et indépendante des T_i .

Dans ce cas, la log-vraisemblance est donnée par :

$$L(\gamma) = \sum_{i=1}^n \delta_i \left[\ln \alpha \sigma \theta \lambda - \lambda t_i + (\alpha \sigma - 1) \ln (1 - e^{-\lambda t_i}) - \ln (1 - (1 - e^{-\lambda t_i})^{\alpha \sigma}) \right] + \theta \sum_{i=1}^n \ln (1 - (1 - e^{-\lambda t_i})^{\alpha \sigma})$$

Les fonctions de score sont obtenues :

$$\begin{aligned}\frac{\partial L}{\partial \alpha} &= \sum_{i=1}^n \delta_i \left[\frac{1}{\alpha} + \frac{\sigma \ln(1 - e^{-\lambda t_i})}{1 - (1 - e^{-\lambda t_i})^{\alpha\sigma}} \right] - \sigma \theta \sum_{i=1}^n \frac{(1 - e^{-\lambda t_i})^{\alpha\sigma} \ln(1 - e^{-\lambda t_i})}{1 - (1 - e^{-\lambda t_i})^{\alpha\sigma}} \\ \frac{\partial L}{\partial \lambda} &= \sum_{i=1}^n \delta_i \left[\frac{1}{\lambda} - \frac{t_i (1 - \alpha\sigma e^{-\lambda t_i} - (1 - e^{-\lambda t_i})^{\alpha\sigma})}{(1 - e^{-\lambda t_i}) (1 - (1 - e^{-\lambda t_i})^{\alpha\sigma})} \right] - \alpha\sigma\theta \sum_{i=1}^n \frac{t_i e^{-\lambda t_i} (1 - e^{-\lambda t_i})^{\alpha\sigma-1}}{1 - (1 - e^{-\lambda t_i})^{\alpha\sigma}} \\ \frac{\partial L}{\partial \sigma} &= \sum_{i=1}^n \delta_i \left[\frac{1}{\sigma} + \frac{\alpha \ln(1 - e^{-\lambda t_i})}{1 - (1 - e^{-\lambda t_i})^{\alpha\sigma}} \right] - \alpha\theta \sum_{i=1}^n \frac{(1 - e^{-\lambda t_i})^{\alpha\sigma} \ln(1 - e^{-\lambda t_i})}{1 - (1 - e^{-\lambda t_i})^{\alpha\sigma}} \\ \frac{\partial L}{\partial \theta} &= \sum_{i=1}^n \frac{\delta_i}{\theta} + \sum_{i=1}^n \ln(1 - (1 - e^{-\lambda t_i})^{\alpha\sigma})\end{aligned}$$

Comme les formes analytiques des estimateurs du maximum de vraisemblance des paramètres ne peuvent pas être déduites, alors on applique les méthodes numériques.

Simulations cas de données censurées

Nous proposons une étude numérique pour déterminer les estimateurs du maximum de vraisemblance. Nous avons généré une série de $N = 10,000$ échantillons avec différents taux de censure, et différentes taille $n_1 = 15, n_2 = 25, n_3 = 50, n_4 = 130, n_5 = 350$ et $n_6 = 500$, de variables aléatoires issues d'une loi $Kw - GEE$ de paramètres $\alpha = 2, \lambda = 1.5, \sigma = 0.7, \theta = 0.9$. On calcule les valeurs des estimateurs du maximum de vraisemblance des paramètres pour chacun de ces échantillons en utilisant l'algorithme BB solve et le logiciel R. Les valeurs moyennes simulés $\hat{\alpha}$ et $\hat{\lambda}$, $\hat{\sigma}$, $\hat{\theta}$ et leurs erreurs quadratiques moyennes (notées $S.M.E$) sont réunies dans le tableau suivant :

$N = 10,000$	$n_1 = 15$	$n_2 = 25$	$n_3 = 50$	$n_4 = 130$	$n_5 = 350$	$n_6 = 500$
$\hat{\alpha}$	1.8954	1.9214	1.9543	1.9654	1.9874	1.9897
$S.M.E$	0.0045	0.0041	0.0035	0.0024	0.019	0.0012
$\hat{\lambda}$	1.3547	1.3987	1.4231	1.4513	1.4786	1.4987
$S.M.E$	0.0039	0.0031	0.0022	0.0018	0.0012	0.0008
$\hat{\sigma}$	0.6235	0.6514	0.6599	0.6789	0.6897	0.6964
$S.M.E$	0.0046	0.0037	0.0029	0.0025	0.0015	0.0005
$\hat{\theta}$	0.8698	0.8746	0.8830	0.8948	0.8966	0.8987
$S.M.E$	0.0029	0.0027	0.0020	0.0013	0.0009	0.0003

Tab. 3. Les valeurs moyennes des EMV des paramètres et leurs erreurs quadratiques moyennes de la distribution $Kw - GEE$.

Nous remarquons que les résultats sont très satisfaisants.

2.4 Modèle transformé de la distribution exponentielle linéaire généralisée *TGLE*

La transformée de la distribution exponentielle linéaire généralisée (*TGLE*) caractérisée par quatre paramètres $\gamma = (\alpha, \lambda, \beta, \sigma)$ et proposée récemment par Elbatal et al. (2013) a un grand intérêt pour la modélisation de nombreuses données réelles, que ce soit en fiabilité, en analyse de survie ou en biologie. Elle peut être considérée comme la généralisation d'au moins quatre distributions de probabilité : les transformées de la distribution exponentielle linéaire, exponentielle linéaire généralisée, Weibull et Rayleigh, en plus des transformées des distributions exponentielles linéaires simples.

2.4.1 Présentation du modèle

On dit qu'une variable aléatoire $t \sim TGLE$ de paramètre d'échelle α et de paramètres de forme λ, β si sa fonction de répartition est donnée par

$$F(t; \gamma) = \left(1 - e^{-(\alpha t + \frac{\lambda}{2} t^2)^\beta}\right) \left(1 + \sigma e^{-(\alpha t + \frac{\lambda}{2} t^2)^\beta}\right)$$

où σ est le paramètre de la transformation.

La fonction de densité est

$$f(t; \gamma) = (\alpha + \lambda t) \beta e^{-(\alpha t + \frac{\lambda}{2} t^2)^\beta} \left(\alpha t + \frac{\lambda}{2} t^2\right)^{\beta-1} \left(1 - \sigma + 2\sigma e^{-(\alpha t + \frac{\lambda}{2} t^2)^\beta}\right)$$

La fonction de survie et le taux de hasard sont respectivement :

$$S(t; \gamma) = 1 - \left(1 - e^{-(\alpha t + \frac{\lambda}{2} t^2)^\beta}\right) \left(1 + \sigma e^{-(\alpha t + \frac{\lambda}{2} t^2)^\beta}\right)$$

et

$$h(t; \gamma) = \frac{(\alpha + \lambda t) \beta e^{-(\alpha t + \frac{\lambda}{2} t^2)^\beta} \left(\alpha t + \frac{\lambda}{2} t^2\right)^{\beta-1} \left(1 - \sigma + 2\sigma e^{-(\alpha t + \frac{\lambda}{2} t^2)^\beta}\right)}{1 - \left(1 - e^{-(\alpha t + \frac{\lambda}{2} t^2)^\beta}\right) \left(1 + \sigma e^{-(\alpha t + \frac{\lambda}{2} t^2)^\beta}\right)}$$

Le taux de hasard cumulé est donné par

$$H(t; \gamma) = -\ln \left[1 - \left(1 - e^{-(\alpha t + \frac{\lambda}{2} t^2)^\beta}\right) \left(1 + \sigma e^{-(\alpha t + \frac{\lambda}{2} t^2)^\beta}\right)\right]$$

La transformée de la distribution exponentielle linéaire généralisée est un modèle très flexible qui se rapproche de différentes distributions selon les valeurs des paramètres choisis.

Pour $\sigma = 0$, on obtient la distribution exponentielle linéaire généralisée *GLE* (α, λ, β) , dont le taux de hasard est donné par

$$h(t; \gamma) = \beta (\alpha + \lambda t) \left(\alpha t + \frac{\lambda}{2} t^2\right)^{\beta-1}$$

Pour $\beta = 1$, on obtient la transformée de la distribution exponentielle linéaire *TLE* $(\alpha, \lambda, \sigma)$

$$h(t; \gamma) = \frac{(\alpha + \lambda t) e^{-(\alpha t + \frac{\lambda}{2} t^2)} \left(1 - \sigma + 2\sigma e^{-(\alpha t + \frac{\lambda}{2} t^2)}\right)}{1 - \left(1 - e^{-(\alpha t + \frac{\lambda}{2} t^2)}\right) \left(1 + \sigma e^{-(\alpha t + \frac{\lambda}{2} t^2)}\right)}$$

Pour $\beta = 1, \sigma = 0$, on obtient la distribution linéaire exponentielle *LE* (α, λ)

$$h(t; \gamma) = (\alpha + \lambda t)$$

Pour $\alpha = \frac{1}{\eta}$, $\lambda = 0$, nous obtenons la transformée de la distribution de Weibull $TW(\eta, \beta, \sigma)$

$$h(t; \gamma) = \frac{\frac{\beta}{\eta} \left(\frac{t}{\eta}\right)^{\beta-1} e^{-\left(\frac{t}{\eta}\right)^\beta} \left(1 - \sigma + 2\sigma e^{-\left(\frac{t}{\eta}\right)^\beta}\right)}{1 - \left(1 - e^{-\left(\frac{t}{\eta}\right)^\beta}\right) \left(1 + \sigma e^{-\left(\frac{t}{\eta}\right)^\beta}\right)}$$

Pour $\alpha = \frac{1}{\eta}$, $\sigma = \lambda = 0$, nous obtenons la distribution de Weibull $W(\eta, \beta)$

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

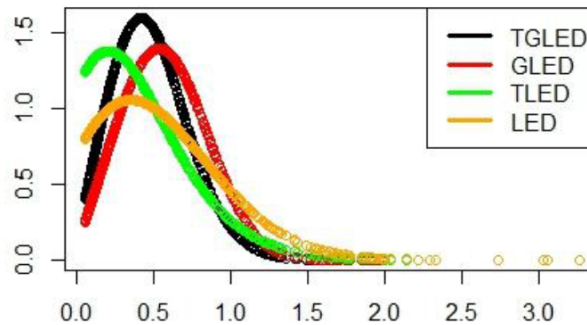
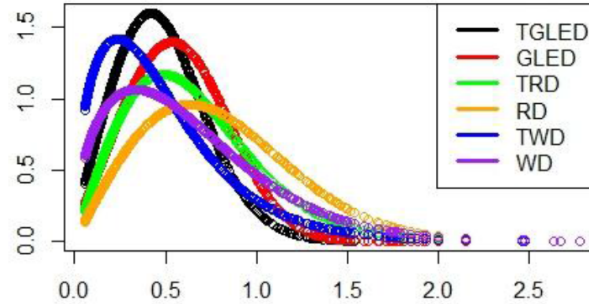
Pour $\beta = 1$, $\alpha = 0$, nous obtenons la transformée de distribution de Rayleigh $TR(\lambda, \sigma)$

$$h(t; \gamma) = \frac{\lambda t e^{-\frac{\lambda}{2}t^2} \left(1 - \sigma + 2\sigma e^{-\frac{\lambda}{2}t^2}\right)}{1 - \left(1 - e^{-\frac{\lambda}{2}t^2}\right) \left(1 + \sigma e^{-\frac{\lambda}{2}t^2}\right)}$$

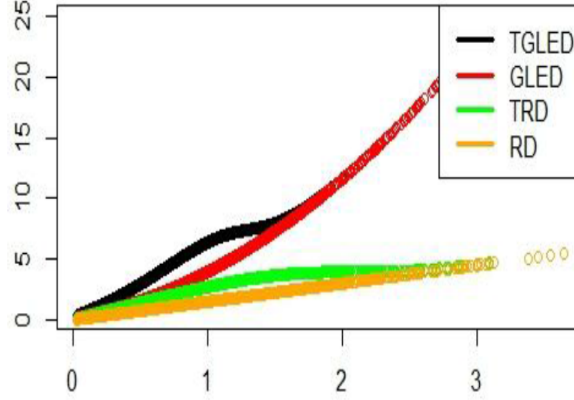
Pour $\beta = 1$, $\alpha = \sigma = 0$, nous obtenons la distribution de Rayleigh $R(\lambda)$

$$h(t; \gamma) = \lambda t$$

Comparaison des formes de la densité de la distribution *TGLE* et les différents sous-modèles :



Comparaison des formes du taux de hasard de la distribution *TGLE* et les différents sous-modèles :



2.4.2 Estimation des paramètres pour données censurées

Les différentes caractéristiques du modèle sont développées dans l'article de Elbatal (2013). Néanmoins, le cas de la censure n'a pas été abordé. Dans cette section, on se propose de calculer les estimateurs du maximum de vraisemblance des paramètres du modèle basés sur un échantillon aléatoire censuré aléatoirement à droite.

Estimation du maximum de vraisemblance

Considérons une suite de variables aléatoires indépendantes et identiquement T_i distribuées selon un modèle transformé de la distribution exponentielle linéaire généralisée (*TGLE*). L'échantillon est composé des variables $t_i = \min(T_i, C_i)$ où C_i représente la censure. La fonction de la log-vraisemblance peut s'écrire dans ce cas sous la forme :

$$\begin{aligned}
 L(\gamma) = & \sum_{i=1}^n \delta_i \left[\ln(\beta) + \ln(\alpha + \lambda t_i) - \left(\alpha t_i + \frac{\lambda}{2} t_i^2 \right)^\beta + (\beta - 1) \ln \left(\alpha t_i + \frac{\lambda}{2} t_i^2 \right) \right. \\
 & \left. + \ln \left(1 - \sigma + 2\sigma e^{-(\alpha t_i + \frac{\lambda}{2} t_i^2)^\beta} \right) - \ln \left[1 - \left(1 - e^{-(\alpha t_i + \frac{\lambda}{2} t_i^2)^\beta} \right) \left(1 + \sigma e^{-(\alpha t_i + \frac{\lambda}{2} t_i^2)^\beta} \right) \right] \right] \\
 & + \sum_{i=1}^n \ln \left[1 - \left(1 - e^{-(\alpha t_i + \frac{\lambda}{2} t_i^2)^\beta} \right) \left(1 + \sigma e^{-(\alpha t_i + \frac{\lambda}{2} t_i^2)^\beta} \right) \right]
 \end{aligned}$$

Après de longs calculs, nous obtenons les fonctions de score pour les paramètres

$$\begin{aligned} \frac{\partial L}{\partial \alpha} = & \sum_{i=1}^n \delta_i \left[\frac{1}{\alpha + \lambda t_i} - \beta t_i u_i^{\beta-1} + (\beta - 1) t_i u_i^{-1} - \frac{2\sigma \beta t_i e^{-u_i^\beta} u_i^{\beta-1}}{1 - \sigma + 2\sigma e^{-u_i^\beta}} \right. \\ & \left. + \beta \frac{t_i e^{-u_i^\beta} u_i^{\beta-1} [1 - \sigma + 2\sigma e^{-u_i^\beta}]}{1 - (1 - e^{-u_i^\beta}) (1 + \sigma e^{-u_i^\beta})} \right] - \beta \sum_{i=1}^n \frac{t_i e^{-u_i^\beta} u_i^{\beta-1} [1 - \sigma + 2\sigma e^{-u_i^\beta}]}{1 - (1 - e^{-u_i^\beta}) (1 + \sigma e^{-u_i^\beta})} \end{aligned}$$

$$\begin{aligned} \frac{\partial L}{\partial \lambda} = & \sum_{i=1}^n \delta_i \left[\frac{t_i}{\alpha + \lambda t_i} - \frac{\beta t_i^2 u_i^{\beta-1}}{2} + \frac{(\beta - 1) t_i^2 u_i^{-1}}{2} - \frac{\sigma \beta t_i^2 e^{-u_i^\beta} u_i^{\beta-1}}{1 - \sigma + 2\sigma e^{-u_i^\beta}} \right. \\ & \left. + \frac{\beta t_i^2 e^{-u_i^\beta} u_i^{\beta-1} [1 - \sigma + 2\sigma e^{-u_i^\beta}]}{2 [1 - (1 - e^{-u_i^\beta}) (1 + \sigma e^{-u_i^\beta})]} \right] - \frac{\beta}{2} \sum_{i=1}^n \frac{t_i^2 e^{-u_i^\beta} u_i^{\beta-1} [1 - \sigma + 2\sigma e^{-u_i^\beta}]}{1 - (1 - e^{-u_i^\beta}) (1 + \sigma e^{-u_i^\beta})} \end{aligned}$$

$$\begin{aligned} \frac{\partial L}{\partial \beta} = & \sum_{i=1}^n \delta_i \left[\frac{1}{\beta} + \ln u_i (1 - u_i^\beta) - \frac{2\sigma u_i^\beta \ln(u_i) e^{-u_i^\beta}}{1 - \sigma + 2\sigma e^{-u_i^\beta}} + \frac{u_i^\beta \ln(u_i) e^{-u_i^\beta} [1 - \sigma + 2\sigma e^{-u_i^\beta}]}{1 - (1 - e^{-u_i^\beta}) (1 + \sigma e^{-u_i^\beta})} \right] \\ & - \sum_{i=1}^n \frac{u_i^\beta \ln(u_i) e^{-u_i^\beta} [1 - \sigma + 2\sigma e^{-u_i^\beta}]}{1 - (1 - e^{-u_i^\beta}) (1 + \sigma e^{-u_i^\beta})} \end{aligned}$$

$$\frac{\partial L}{\partial \sigma} = \sum_{i=1}^n \delta_i \left[\frac{2e^{u_i^{-\beta}} - 1}{1 - \sigma + 2\sigma e^{u_i^\beta}} + \frac{e^{-u_i^\beta} (1 - e^{-u_i^\beta})}{1 - (1 - e^{-u_i^\beta}) (1 + \sigma e^{-u_i^\beta})} \right] - \sum_{i=1}^n \frac{e^{-u_i^\beta} (1 - e^{-u_i^\beta})}{1 - (1 - e^{-u_i^\beta}) (1 + \sigma e^{-u_i^\beta})}$$

En égalisant ces fonctions à zéro, pour calculer les estimateurs du maximum de vraisemblance des paramètres, on obtient des équations non linéaires non résolubles analytiquement. Dans ce cas, on fait appel aux méthodes numériques.

Simulations

La génération d'une variable aléatoire T suivant une loi *TGLE* se ramène à la génération des suites des variables aléatoires indépendantes de loi uniforme sur $[0, 1]$. Si Z représente une variable aléatoire uniforme dans $[0, 1]$, alors

$$T = -\frac{\alpha}{\lambda} + \frac{1}{\lambda} \sqrt{\alpha^2 + 2\lambda \left(-\ln \left\{ \frac{\sigma - 1 + \sqrt{(1 + \sigma)^2 - 4\sigma Z}}{2\sigma} \right\} \right)^{1/\beta}}$$

Pour le calcul des estimateurs du maximum de vraisemblance $\hat{\alpha}, \hat{\gamma}, \hat{\lambda}, \hat{\beta}$ et $\hat{\sigma}$ et leurs erreurs quadratiques moyennes (*S.M.E*) des paramètres du modèle, nous avons simulé les données de $N = 10,000$ échantillons de tailles respectives $n_1 = 15, n_2 = 25, n_3 = 50, n_4 = 130, n_5 = 350$ et $n_6 = 500$, ayant les valeurs suivantes des paramètres : $\alpha = 0.3, \lambda = 0.6, \beta = 0.65, \sigma = 0.7$. Les valeurs obtenues sont résumées dans le tableau suivant :

$N = 10,000$	$n_1 = 15$	$n_2 = 25$	$n_3 = 50$	$n_4 = 130$	$n_5 = 350$	$n_6 = 500$
$\hat{\alpha}$	0.2425	0.2541	0.2625	0.2745	0.2894	0.2978
<i>S.M.E</i>	0.0141	0.0116	0.0107	0.0098	0.0018	0.0010
$\hat{\lambda}$	0.5423	0.5525	0.5714	0.5854	0.5945	0.6012
<i>S.M.E</i>	0.0191	0.0124	0.0121	0.0054	0.0035	0.0022
$\hat{\beta}$	0.6821	0.6754	0.6625	0.6554	0.6562	0.6519
<i>S.M.E</i>	0.0152	0.0148	0.0098	0.0050	0.0035	0.0017
$\hat{\sigma}$	0.7654	0.7425	0.7315	0.7245	0.7194	0.7098
<i>S.M.E</i>	0.0135	0.0130	0.0126	0.0089	0.0052	0.0036

Tab. 4. Les valeurs moyennes des EMV des paramètres et leurs erreurs quadratiques moyennes pour les données censurées

Du tableau 4, nous pouvons remarquer que les erreurs quadratiques moyennes sont faibles confirmant la convergence des estimateurs obtenus.

CHAPITRE 3

Tests d'ajustement pour données censurées

3.1 Introduction

Vérifier si une série d'observations peut être modélisée par une distribution de probabilité est et restera l'un des problèmes les plus importants dans l'analyse statistique. Pour valider le choix d'un modèle, on utilise ce qu'on appelle les tests statistiques d'ajustement. Les chercheurs se sont penchés sur ce problème depuis le début du 20ème siècle où Karl Pearson (1900) a proposé le fameux test du chi-deux. Et depuis, plusieurs techniques et méthodes ont vu le jour comme les statistiques classiques très répandues d'Anderson-Darling, de Kolmogorov-Smirnov, Cramer-VonMises et bien d'autres, sans oublier les tests graphiques. Avec le développement technologique dans tous les domaines que ce soit dans l'industrie par la production de matériels de plus en plus performants, ou dans la recherche médicale avec les nouveaux schémas thérapeutiques, en biologie, en assurance, les échantillons récoltés sont souvent incomplets. Dans ces cas, les méthodes classiques d'ajustement ne sont plus adaptées, alors des modifications de celles-ci ont été proposées ces dernières décennies pour tenir compte de la censure et des paramètres inconnus. Galanova *et al.* (2012) ont donné les modifications des statistiques de Kolmogorov, Cramer-Von Mises-Smirnov, et Anderson-Darling pour quelques modèles où les fonctions de répartitions empiriques sont remplacées par leurs estimateurs de Kaplan-Meier. D'autres tests d'ajustement basés sur la différence entre les

fonctions cumulatives théoriques et estimées par l'estimateur de Kaplan-Meyer et celui du maximum de vraisemblance ont été proposés par Habib et Thomas (1986), Hollander et Pena (1992), alors que Hjort (1990) a basé son test sur les différences entre les nombres de taux de pannes observées et théoriques.

Dans ce chapitre, on expose un test d'ajustement proposé récemment par Bagdonavicius et Nikulin (2011), Bagdonavicius et al. (2013) pour les familles de lois paramétriques continues. Basé sur les différences entre les nombres de taux de pannes observées et théoriques dans les intervalles de groupement des données, ce test utilise toute l'information apportée par les estimateurs du maximum de vraisemblance sur les données initiales non groupées. La statistique de ce type de test est distribuée selon loi du chi-deux. Ensuite nous construisons des tests d'ajustement pour les différentes distributions présentées dans le premier chapitre, en l'occurrence la distribution GEE, la distribution LE, La distribution Kw-GEE et la distribution TGLE.

3.2 Test de Bagdonavicius et Nikulin

La statistique de test pour données censurées, proposée par Bagdonavicius et Nikulin (2011), Bagdonavicius et al. (2013) suit asymptotiquement une distribution du chi-deux et dont le regroupement des données est fonction de variables aléatoires telles que le nombre de pannes attendues soit le même pour toutes les classes choisies.

On considère un échantillon censuré à droite

$$(t_1, \delta_1), (t_2, \delta_2), \dots, (t_n, \delta_n) \quad (1)$$

Supposons que les taux de panne $T_1, T_2, \dots, T_n$ sont n variables aléatoires indépendantes et identiquement distribuées. La fonction de densité de T_i appartient à une famille paramétrique $\{f(t, \gamma), \gamma \in \Theta \subset R^s\}$

$$t_i = T_i \wedge C_i \quad \delta_i = 1_{\{T_i \leq C_i\}}$$

sous une censure non informative C .

$$H(t, \gamma) = -\ln S(t, \gamma) = \int_0^t h(u, \gamma) du$$

est la fonction de hasard cumulative. Et la fonction de la log-vraisemblance est :

$$L(\gamma) = \sum_{i=1}^n \delta_i \ln h(t_i, \gamma) + \sum_{i=1}^n \ln S(t_i, \gamma) \quad (2)$$

alors l'estimateur du maximum de vraisemblance $\hat{\gamma}$ de vérifie :

$$\dot{L}(\hat{\gamma}) = 0$$

où

$$\dot{L}(\gamma) = \frac{\partial}{\partial \gamma} L(\gamma) = \left(\frac{\partial}{\partial \gamma_1} L(\gamma), \dots, \frac{\partial}{\partial \gamma_s} L(\gamma) \right)^T \quad (3)$$

La matrice d'information de Fisher est :

$$I(\gamma) = -E_{\gamma} \ddot{L}(\gamma) \quad (4)$$

où

$$\ddot{L}(\gamma) = \sum_{i=1}^n \delta_i \frac{\partial^2}{\partial \gamma^2} \ln h(t_i, \gamma) + \sum_{i=1}^n \frac{\partial^2}{\partial \gamma^2} H(t_i, \gamma) \quad (5)$$

Si γ_0 est la vraie valeur de γ et sous les conditions de régularité, on a :

$$\hat{\gamma} \xrightarrow{P} \gamma_0; \quad \sqrt{n}(\hat{\gamma} - \gamma_0) \xrightarrow{P} N(0_s, i^{-1}(\gamma_0)), \quad 0_s = (0, 0, \dots, 0)^T;$$

$$\frac{1}{\sqrt{n}} \dot{L}(\gamma_0) \xrightarrow{d} N(0_s, i^{-1}(\gamma_0)); \quad -\frac{1}{n} \ddot{L}(\hat{\gamma}) \xrightarrow{P} i(\gamma_0),$$

L'échantillon (1) peut être représenté en termes de processus de comptage comme suit :

$$(N_1(t), Y_1(t), t \geq 0), \dots, (N_n(t), Y_n(t), t \geq 0) \quad (6)$$

Où

$$N_i(t) = 1_{\{X_i \leq t, \delta_i = 1\}}, \quad Y_i(t) = 1_{\{0 < t \leq X_i\}}$$

En posant

$$N(t) = \sum_{i=1}^n N_i(t) \quad \text{et} \quad Y(t) = \sum_{i=1}^n Y_i(t) \quad (7)$$

Les auteurs ont exprimé les expressions des équations (2), (3), (4) et (5) en termes des processus stochastiques N_i et Y_i . Les trajectoires du processus de comptage $N_i(t)$ ont la forme :

$$N_i(t) = \begin{cases} 0, & 0 \leq t < X_i \\ 1, & t \geq X_i \end{cases}$$

Si $\delta_i = 1$, et $N_i(t) = 0$, si $\delta_i = 0$, les trajectoires du processus de comptage Y_i ont la forme :

$$Y_i(t) = \begin{cases} 0, & 0 \leq t < X_i \\ 1, & t \geq X_i \end{cases}$$

En utilisant ces processus, ils obtiennent les relations suivantes :

$$\int_0^\infty \ln h(u, \gamma) dN_i(u) = \begin{cases} \ln h(t_i, \gamma), & \delta_i = 1, \\ 0, & \delta_i = 0 \end{cases} = \delta_i \ln h(t_i, \gamma) \quad (8)$$

et

$$\int_0^\infty Y_i(u) h(u) du = \int_0^{T_i} h(u) du = -\ln S(t_i, \gamma)$$

Ce qui implique

$$L(\gamma) = \sum_{i=1}^n \int_0^\infty \{\ln h(u, \gamma) dN_i(u) - Y_i(u) h(u, \gamma)\} du$$

La matrice d'information de Fisher est :

$$\begin{aligned} I(\gamma) &= -E_\theta \ddot{L}(\gamma) \\ &= E_\theta \sum_{i=1}^n \int_0^\tau \frac{\partial}{\partial \theta} \ln h(u, \gamma) \left(\frac{\partial}{\partial \theta} \ln h(u, \gamma) \right)^T h(u, \gamma) Y_i(u) du \end{aligned}$$

Sous les conditions suivantes [Hjort (1990)] :

1- Il existe un voisinage Θ_0 de γ_0 la vraie valeur de γ tels que pour tout n et $\gamma_0 \in \Theta_0$, et pour tout $t \in [0, \tau]$, les dérivées partielles d'ordre 1 et 2 de $h(t, \gamma)$ par rapport à γ existent et sont continues et bornées.

On suppose aussi que la fonction de log-vraisemblance est deux fois différentiable par rapport à $\gamma_0 \in \Theta_0$.

2- $h(t, \gamma)$ est bornée dans $[0, \tau] \times \Theta_0$.

3- Il existe une fonction positive $y(t)$ telle que :

$$\sup_{t \in [0, \tau]} |Y(t)/n - y(t)| \longrightarrow 0$$

4- La matrice $i(\gamma_0) = \lim_{n \rightarrow \infty} \frac{I(\gamma_0)}{n}$ (la limite existe sous les conditions 1-3) est définie positive.

On suppose que les processus N_i et Y_i sont observés pendant un temps fini τ . L'intervalle $[0, \tau]$ est partagé en $k > s$ petits intervalles.

$$I_j = (a_{j-1}, a_j], \quad a_0 = 0, \quad a_k = \tau$$

notons par

$$U_j = N(a_j) - N(a_{j-1}) = \sum_{i: t_i \in I_j} \delta_i$$

le nombre de défaillances observées dans le j^{eme} intervalle, $j = 1, 2, \dots, k$.

On prévoit d'observer e_j défaillances

$$e_j = \int_{a_{j-1}}^{a_j} h(u, \hat{\gamma}) Y(u) du = \sum_{i: t_i > a_{j-1}} (H(a_j \wedge t_i, \hat{\gamma}) - H(a_{j-1}, \hat{\gamma}))$$

où $a \wedge b = \min(a, b)$; ici $\hat{\gamma}$ est l'estimateur du maximum de vraisemblance du paramètre γ .

Pour tester l'hypothèse H_0 , Bagdonavicius et Nikulin (2011) ont proposé la statistique :

$$Y_n^2 = Z^T \hat{V}^- Z$$

basée sur le vecteur

$$Z = (Z_1, Z_2, \dots, Z_k)^T, \quad Z_j = \frac{1}{\sqrt{n}} (U_j - e_j), \quad j = 1, \dots, k$$

où

$$\hat{V}^- = \hat{A}^{-1} + \hat{A}^{-1} \hat{C}^T \hat{G}^- \hat{C} \hat{A}^{-1}, \quad \hat{G} = \hat{i} - \hat{C} \hat{A}^{-1} \hat{C}^T$$

est une matrice inverse généralisée de $\hat{V}$

$$\hat{A} = \begin{pmatrix} \hat{A}_1 & 0 & \dots & 0 \\ 0 & \hat{A}_2 & \dots & 0 \\ & & \ddots & \\ 0 & 0 & & \hat{A}_k \end{pmatrix}, \quad \hat{A} = \frac{U_j}{n}$$

$$\hat{C} = \begin{pmatrix} \hat{C}_{11} & \hat{C}_{12} & \dots & \hat{C}_{1k} \\ \hat{C}_{21} & \hat{C}_{22} & \dots & \hat{C}_{2k} \\ . & . & & . \\ . & . & & . \\ \hat{C}_{s1} & \hat{C}_{s2} & \dots & \hat{C}_{sk} \end{pmatrix}, \quad \hat{C} = \frac{1}{n} \sum_{i:t_i \in I_j} \delta_i \frac{\partial}{\partial \theta_l} \ln h(t_i, \hat{\gamma})$$

$$\hat{i} = \begin{pmatrix} \hat{i}_{11} & \hat{i}_{12} & \dots & \hat{i}_{1r} \\ \hat{i}_{21} & \hat{i}_{22} & \dots & \hat{i}_{2r} \\ . & . & & . \\ . & . & & . \\ \hat{i}_{s1} & \hat{i}_{s2} & \dots & \hat{i}_{sr} \end{pmatrix}, \quad \hat{i}_{ll'} = \frac{1}{n} \sum_{i=1}^n \delta_i \frac{\partial \ln h(t_i, \hat{\gamma})}{\partial \gamma_l} \frac{\partial \ln h(t_i, \hat{\gamma})}{\partial \gamma_{l'}}$$

Cette statistique peut s'écrire

$$Y_n^2 = Z^T \hat{A}^{-1} Z + Z^T \hat{A}^{-1} \hat{C}^T \hat{G}^{-1} \hat{C} \hat{A}^{-1} Z = \sum_{j=1}^k \frac{(U_j - e_j)^2}{U_j} + Q$$

où la forme quadratique Q est

$$Q = W^T \hat{G}^{-1} W, \quad \hat{W} = \hat{C} \hat{A}^{-1} Z = (\hat{W}_1, \dots, \hat{W}_s)^T,$$

$$W_l = \sum_{j=1}^k \hat{C}_{lj} \hat{A}_j^{-1} Z_j, \quad \hat{G} = [\hat{g}_{ll'}]_{s \times s}, \quad \hat{g}_{ll'} = \hat{i}_{ll'} - \sum_{j=1}^k \hat{C}_{lj} \hat{C}_{l'j} \hat{A}_j^{-1}$$

$$i = 1, \dots, n, \quad j = 1, \dots, k, \quad l, l' = 1, \dots, s$$

Choix des intervalles du groupement des données

Les auteurs ont proposé de prendre les limites a_j des intervalles I_j comme fonctions de variables aléatoires pour obtenir des effectifs théoriques des taux de pannes tous égaux. L'intervalle $[0, \tau]$ est partagé en k intervalles tels que.

$$\hat{a}_j = H^{-1} \left(\frac{E_j - \sum_{l=1}^{j-1} H(t_l, \gamma)}{n - j + 1}, \hat{\gamma} \right), \quad \hat{a}_k = \max(t_{(n)}, \tau)$$

où H^{-1} est l'inverse de la fonction de hasard cumulative H

Nous avons

$$0 < \hat{a}_1 < \hat{a}_2 < \dots \hat{a}_k = \tau$$

Pour ce choix d'intervalles, on a

$$e_j = \frac{E_k}{k} \text{ pour tout } j,$$

où

$$E_k = \sum_{i=1}^n H(t_i, \hat{\gamma})$$

En remplaçant a_j par $\hat{a}_j$ dans l'expression de la statistique Y_n^2 , la distribution limite de la statistique Y_n^2 est un Chi-deux avec ν degrés de liberté $\nu = tr V^-$, comme dans le cas de a_j fixe.

3.3 Test d'ajustement pour le modèle GEE

Dans cette section nous utilisons l'approche proposée par Bagdonavicius et Nikulin (2011) et exposée ci-dessus, pour construire un test d'ajustement pour la distribution exponentielle exponentiée généralisée *GEE* dans le cas de censure aléatoire droite et de paramètres inconnus. Comme nous l'avons mentionné, toutes les méthodes d'estimation pour cette distribution, ont été développées dans la littérature statistique. Pour les données complètes et paramètres inconnus, Hassan (2005) a fourni des tables de valeurs critiques des statistiques de test classiques de Kolmogorov-smirnov, Anderson-Darling, Watson test et Ln test. Pour les données censurées, ce sont les tests du rapport de vraisemblance qui sont utilisés.

Soit T_i un échantillon censuré à droite. Pour vérifier l'hypothèse nulle H_0 que le taux de défaillance T_i suit une distribution exponentielle exponentiée généralisée *GEE*, on calcule d'abord les estimateurs du maximum de vraisemblance des paramètres α, λ . Pour le choix des limites $\hat{a}_j$ des classes, lorsque la distribution de base est une *GEE*, est obtenu comme suit :

$$\hat{a}_j = -\frac{1}{\hat{\lambda}} \ln \left(1 - \left[1 - \exp \left(\frac{\sum_{l=1}^{i-1} H(t_l, \hat{\gamma}) - E_j}{n - i + 1} \right) \right]^{\frac{1}{\hat{\alpha}}} \right), \quad \text{pour } j = 1, \dots, k-1, \quad \hat{a}_k = t_{(n)},$$

Où

$$E_j = \frac{j}{k-1} \sum_{i=1}^n H(t_i, \hat{\gamma}) = -\frac{j}{k-1} \sum_{i=1}^n \ln \left(1 - \left(1 - \exp \left\{ -\hat{\lambda} t_i \right\} \right)^{\hat{\alpha}} \right), \quad j = 1, \dots, k-1$$

Pour chaque intervalle I_j on a $e_j = \frac{E_k}{k}$

Après calcul, nous trouvons les éléments de la matrice $\hat{C}$ qui sont donnés par :

$$\begin{aligned} \hat{C}_{1j} &= \frac{1}{n} \sum_{i:t_i \in I_j} \delta_i \left[\frac{1}{\hat{\alpha}} + \frac{\ln(1 - e^{-\hat{\lambda} t_i})}{1 - (1 - e^{-\hat{\lambda} t_i})^{\hat{\alpha}}} \right] \\ \hat{C}_{2j} &= \frac{1}{n} \sum_{i:t_i \in I_j} \delta_i \left[\frac{1}{\hat{\lambda}} + \frac{t_i \left(\hat{\alpha} e^{-\hat{\lambda} t_i} + (1 - e^{-\hat{\lambda} t_i})^{\hat{\alpha}} - 1 \right)}{(1 - e^{-\hat{\lambda} t_i}) \left(1 - (1 - e^{-\hat{\lambda} t_i})^{\hat{\alpha}} \right)} \right] \end{aligned}$$

On en déduit la matrice $\hat{W}$ d'éléments :

$$\hat{W}_l = \sum_{j=1}^k \hat{C}_{lj} A_j^{-1} Z_j, \quad l = 1, 2 \quad j = 1, \dots, k$$

Calcul de la matrice de Fisher estimée $\hat{I}$

Dans ce cas, à partir de la fonction de la log-vraisemblance, nous pouvons calculer la matrice d'information de Fisher estimée

$$\begin{aligned} \hat{i}_{11} &= \frac{1}{n} \sum_{i=1}^n \delta_i \left[\frac{1}{\hat{\alpha}} + \frac{\ln(1 - e^{-\hat{\lambda} t_i})}{1 - (1 - e^{-\hat{\lambda} t_i})^{\hat{\alpha}}} \right]^2 \\ \hat{i}_{22} &= \frac{1}{n} \sum_{i=1}^n \delta_i \left[\frac{1}{\hat{\lambda}} + \frac{t_i \left(\hat{\alpha} e^{-\hat{\lambda} t_i} + (1 - e^{-\hat{\lambda} t_i})^{\hat{\alpha}} - 1 \right)}{(1 - e^{-\hat{\lambda} t_i}) \left(1 - (1 - e^{-\hat{\lambda} t_i})^{\hat{\alpha}} \right)} \right]^2 \\ \hat{i}_{21} = \hat{i}_{12} &= \frac{1}{n} \sum_{i=1}^n \delta_i \left[\frac{1}{\hat{\alpha}} + \frac{\ln(1 - e^{-\hat{\lambda} t_i})}{1 - (1 - e^{-\hat{\lambda} t_i})^{\hat{\alpha}}} \right] \left[\frac{1}{\hat{\lambda}} + \frac{t_i \left(\hat{\alpha} e^{-\hat{\lambda} t_i} + (1 - e^{-\hat{\lambda} t_i})^{\hat{\alpha}} - 1 \right)}{(1 - e^{-\hat{\lambda} t_i}) \left(1 - (1 - e^{-\hat{\lambda} t_i})^{\hat{\alpha}} \right)} \right] \end{aligned}$$

Calcul de la matrice $\hat{G}$

Les composants de la matrice $\hat{G} = [\hat{g}_{uv}]_{2 \times 2}$ sont donnés comme suit :

$$\hat{g}_{11} = \frac{1}{n} \sum_{i=1}^n \delta_i \left[\frac{1}{\hat{\alpha}} + \frac{\ln(1 - e^{-\hat{\lambda} t_i})}{1 - (1 - e^{-\hat{\lambda} t_i})^{\hat{\alpha}}} \right]^2 - \frac{A_j^{-1}}{n^2} \sum_{j=1}^k \left[\sum_{i:t_i \in I_j} \delta_i \left(\frac{1}{\hat{\alpha}} + \frac{\ln(1 - e^{-\hat{\lambda} t_i})}{1 - (1 - e^{-\hat{\lambda} t_i})^{\hat{\alpha}}} \right) \right]^2$$

$$\begin{aligned}
 \hat{g}_{12} &= \hat{g}_{21} = \frac{1}{n} \sum_{i=1}^n \delta_i \left(\frac{1}{\hat{\alpha}} + \frac{\ln(1 - e^{-\hat{\lambda}t_i})}{1 - (1 - e^{-\hat{\lambda}t_i})^{\hat{\alpha}}} \right) \left(\frac{1}{\hat{\lambda}} + \frac{t_i \left(\hat{\alpha}e^{-\hat{\lambda}t_i} + (1 - e^{-\hat{\lambda}t_i})^{\hat{\alpha}} - 1 \right)}{(1 - e^{-\hat{\lambda}t_i}) \left(1 - (1 - e^{-\hat{\lambda}t_i})^{\hat{\alpha}} \right)} \right) \\
 &\quad - \frac{A_j^{-1}}{n^2} \sum_{j=1}^k \left[\sum_{i:t_i \in I_j} \delta_i \left(\frac{1}{\hat{\alpha}} + \frac{\ln(1 - e^{-\hat{\lambda}t_i})}{1 - (1 - e^{-\hat{\lambda}t_i})^{\hat{\alpha}}} \right) \right] \left[\sum_{i:t_i \in I_j} \delta_i \left(\frac{1}{\hat{\lambda}} + \frac{t_i \left(\hat{\alpha}e^{-\hat{\lambda}t_i} + (1 - e^{-\hat{\lambda}t_i})^{\hat{\alpha}} - 1 \right)}{(1 - e^{-\hat{\lambda}t_i}) \left(1 - (1 - e^{-\hat{\lambda}t_i})^{\hat{\alpha}} \right)} \right) \right] \\
 \hat{g}_{22} &= \frac{1}{n} \sum_{i=1}^n \delta_i \left[\frac{1}{\hat{\lambda}} + \frac{t_i \left(\hat{\alpha}e^{-\hat{\lambda}t_i} + (1 - e^{-\hat{\lambda}t_i})^{\hat{\alpha}} - 1 \right)}{(1 - e^{-\hat{\lambda}t_i}) \left(1 - (1 - e^{-\hat{\lambda}t_i})^{\hat{\alpha}} \right)} \right]^2 \\
 &\quad - \frac{A_j^{-1}}{n^2} \sum_{j=1}^k \left[\sum_{i:t_i \in I_j} \delta_i \left(\frac{1}{\hat{\lambda}} + \frac{t_i \left(\hat{\alpha}e^{-\hat{\lambda}t_i} + (1 - e^{-\hat{\lambda}t_i})^{\hat{\alpha}} - 1 \right)}{(1 - e^{-\hat{\lambda}t_i}) \left(1 - (1 - e^{-\hat{\lambda}t_i})^{\hat{\alpha}} \right)} \right) \right]^2
 \end{aligned}$$

Enfin, on obtient la forme explicite de la statistique Y_n^2 du test

$$Y_n^2 = \sum_{j=1}^k \frac{(U_j - e_j)^2}{U_j} + Q$$

Cas où $\hat{\gamma} = (\hat{\alpha}, \hat{\lambda})^T$ est inconnu

$$\begin{aligned}
 Y_n^2 &= \sum_{j=1}^k \frac{(U_j - e_j)^2}{U_j} + \frac{W_1^2}{n |G|} \left(\hat{i}_{22} - \sum_{j=1}^k [\hat{C}_{2j}]^2 A_j^{-1} \right) \\
 &\quad + \frac{W_2^2}{n |G|} \left(\hat{i}_{11} - \sum_{j=1}^k [\hat{C}_{1j}]^2 A_j^{-1} \right) - \frac{2W_1W_2}{n |G|} \left(\hat{i}_{12} - \sum_{j=1}^k \hat{C}_{1j}\hat{C}_{2j}A_j^{-1} \right)
 \end{aligned}$$

Où

$|G|$ est le déterminant de la matrice G .

Le cas où α ou bien λ est inconnu

$$Y_n^2(\hat{\gamma}_n) = \begin{cases} \sum_{j=1}^k \frac{(U_j - e_j)^2}{U_j} + \frac{(\sum_{j=1}^k \hat{C}_{1j} A_j^{-1} Z_j)^2}{n(\hat{i}_{22} - \sum_{j=1}^k [\hat{C}_{2j}]^2 A_j^{-1})} & \text{si } \alpha \text{ est connu et } \lambda \text{ est inconnu} \\ \sum_{j=1}^k \frac{(U_j - e_j)^2}{U_j} + \frac{(\sum_{j=1}^k \hat{C}_{2j} A_j^{-1} Z_j)^2}{n(\hat{i}_{11} - \sum_{j=1}^k [\hat{C}_{1j}]^2 A_j^{-1})} & \text{si } \lambda \text{ est connu et } \alpha \text{ est inconnu} \end{cases}$$

3.3.1 Simulation des valeurs du critère de test

Pour tester l'hypothèse nulle H_0 selon laquelle un échantillon de données provient d'un modèle exponentiel exponentié généralisé $GEE(\hat{\gamma})$, nous générons 10,000 échantillons de

3.3. Test d'ajustement pour le modèle GEE

données censurées provenant de cette distribution dont les tailles respectives sont $n_1 = 15$, $n_2 = 25$, $n_3 = 50$, $n_4 = 130$, $n_5 = 350$ et $n_6 = 500$.

Nous calculons le nombre de cas de rejet de l'hypothèse nulle H_0 , c'est à dire lorsque $Y^2 > \chi_\varepsilon^2(k)$ (dans notre cas $\chi_\varepsilon^2(k)$ est le quantile de la distribution du chi-deux à k degrés de liberté), pour les niveaux de signification $\varepsilon = (0.10, 0.05, 0.01)$. Les valeurs des niveaux de signification empiriques sont comparés avec leurs valeurs théoriques correspondantes.

Le cas α et λ sont inconnus

$N = 10,000$	$n_1 = 15$	$n_2 = 25$	$n_3 = 50$	$n_4 = 130$	$n_5 = 350$	$n_6 = 500$
$\varepsilon = 1\%$	0.0081	0.0085	0.0089	0.0095	0.0098	0.0102
$\varepsilon = 5\%$	0.0462	0.0475	0.0479	0.0483	0.0496	0.0501
$\varepsilon = 10\%$	0.0978	0.0980	0.0984	0.0992	0.0995	0.0999

Le cas α est connu

$N = 10,000$	$n_1 = 15$	$n_2 = 25$	$n_3 = 50$	$n_4 = 130$	$n_5 = 350$	$n_6 = 500$
$\varepsilon = 1\%$	0.0080	0.0085	0.0089	0.0092	0.0097	0.0101
$\varepsilon = 5\%$	0.0476	0.0488	0.0492	0.0495	0.0499	0.0502
$\varepsilon = 10\%$	0.0979	0.0982	0.0989	0.0995	0.0998	0.1002

Le cas λ est connu

$N = 10,000$	$n_1 = 15$	$n_2 = 25$	$n_3 = 50$	$n_4 = 130$	$n_5 = 350$	$n_6 = 500$
$\varepsilon = 1\%$	0.0085	0.0089	0.0093	0.0095	0.0098	0.0099
$\varepsilon = 5\%$	0.0489	0.0492	0.0495	0.0498	0.0511	0.0504
$\varepsilon = 10\%$	0.0974	0.0984	0.0989	0.0991	0.0996	0.0999

Tab.5. Comparaison des risques empiriques et risques théoriques correspondants

($\varepsilon = 0.1, 0.05, 0.01$) pour les données censurées

A partir des résultats obtenus, on constate que les niveaux de signification empiriques de la statistique Y_n^2 coïncident avec ceux correspondants aux niveaux théoriques de la distribution du chi-deux à k degrés de liberté. Par conséquent, nous pouvons dire que le test proposé peut ajuster d'une manière satisfaisante des données censurées de la distribution exponentielle généralisée $GEE(\gamma)$.

3.3.2 Exemple de données de survie

Nous considérons les données d'échantillons de 51 patients atteints de leucémie myéloïde aiguë avancée signalée dans le Registre international de transplantation de moelle osseuse. Ces patients avaient reçu une transplantation autologue (auto) de moelle osseuse dans laquelle, après de fortes doses de chimiothérapie, leur propre moelle a été réintroduite pour remplacer leur système immunitaire détruit.

Temps de survie (en mois) pour les transplantations autologues :

0.658, 0.822, 1.414, 2.5, 3.322, 3.816, 4.737, 4.836*, 4.934, 5.033, 5.757, 5.855, 5.987, 6.151, 6.217, 6.447*, 8.651, 8.717, 9.441*, 10.329, 11.48, 12.007, 12.007*, 12.237, 12.401*, 13.059*, 14.474*, 15*, 15.461, 15.757, 16.48, 16.711, 17.204*, 17.237, 17.303*, 17.664*, 18.092, 18.092*, 18.75*, 20.625*, 23.158, 27.73*, 31.184*, 32.434*, 35.921*, 42.237*, 44.638*, 46.48*, 47.467*, 48.322*, 56.086.

Nous appliquons le test proposé ci-dessus, pour tester l'hypothèse nulle H_0 , selon laquelle ces données correspondent à une distribution exponentielle exponentielle généralisée GEE . Nous calculons les estimateurs de maximum de vraisemblance $\hat{\alpha}, \hat{\lambda}$ des paramètres inconnus α, λ :

$$\hat{\theta} = \left(\hat{\alpha}, \hat{\lambda} \right)^T = (0.9190, 0.03)^T$$

Nous choisissons $k = 5$ classes pour grouper ces observations, les calculs intermédiaires du critère Y_n^2 sont donnés dans le tableau 6

$\hat{a}_j$	3.0890	7.1303	12.5666	22.8197	56.0860
U_j	4	10	6	6	2
e_j	5.5863	5.5863	5.5863	5.5863	5.5863
$\hat{C}_{1j}$	-0.1917	-0.2469	-0.0924	-0.0628	-0.0095
$\hat{C}_{2j}$	2.4278	6.1542	3.7326	3.7633	1.2740

Table 6. valeurs des $\hat{a}_j, e_j, U_j, \hat{C}_{1j}, \hat{C}_{2j}$

et la matrice estimée $\hat{G}$

$$\hat{G}_{2 \times 2} = \begin{pmatrix} 0.1198 & -3.7226 \\ -3.7226 & 177.7709 \end{pmatrix}$$

Le vecteur estimé $\hat{W}$ est égal à :

$$\hat{W} = (\hat{W}_1, \hat{W}_2)^T = \begin{pmatrix} -0.1259 & 8.2881 \end{pmatrix}^T$$

Nous obtenons donc la valeur de la statistique de test Y_n^2 :

$$Y_n^2 = X^2 + Q = 9.0652 + 0.4406 = 9.5059$$

Pour le niveau de signification $\varepsilon = 0.05$, la valeur critique de la distribution chi-carré est $\chi_5^2 = 11.07050$. Comme $\chi_5^2 > Y_n^2$, Ainsi, nous pouvons dire que ces données peuvent être modélisées par la distribution GEE .

3.4 Test d'ajustement pour le modèle LE

Pour valider le modèle exponentiel linéaire, les techniques utilisées jusqu'alors sont les tests graphiques ou bien le rapport de vraisemblance. Pour les données censurées, le problème reste ouvert. Nous proposons la construction d'un test d'ajustement pour le modèle LE dans le cas de censure aléatoire droite et de paramètres inconnus, en adaptant la statistique développée précédemment.

Sous l'hypothèse nulle H_0 , qui consiste à supposer qu'une suite t_i de données censurées à droite est distribuée selon une loi exponentielle linéaire LE , on calcule les estimateurs du maximum de vraisemblance des paramètres. Les valeurs estimées $\hat{a}_j$ des bornes des intervalles I_j de groupement des données, sont obtenues comme suit :

$$\hat{a}_j = -\frac{\hat{\alpha}}{\hat{\lambda}} + \frac{1}{\hat{\lambda}} \sqrt{\hat{\alpha}^2 + 2\hat{\lambda} \left(\frac{E_j - \sum_{l=1}^{i-1} \hat{\alpha} t_l + \frac{\hat{\lambda} t_l^2}{2}}{n - i + 1} \right)}, \quad \text{pour } j = 1, \dots, k-1, \quad \hat{a}_k = t_{(n)},$$

où

$$E_j = \frac{j}{k-1} \sum_{i=1}^n \hat{\alpha} t_i + \frac{\hat{\lambda} t_i^2}{2}, \quad j = 1, \dots, k-1$$

Pour chaque intervalle I_j on a

$$e_j = E_k/k.$$

Calcul des matrices $\hat{W}$ et $\hat{C}$

Le calcul de la matrice $\hat{W}$ passe par celui de la matrice $\hat{C}$:

$$\hat{C}_{1j} = \frac{1}{n} \sum_{i:t_i \in I_j}^n \frac{\delta_i}{\hat{\alpha} + \hat{\lambda} t_i}$$

$$\hat{C}_{2j} = \frac{1}{n} \sum_{i:t_i \in I_j}^n \frac{\delta_i t_i}{\hat{\alpha} + \hat{\lambda} t_i}$$

On déduit la matrice $\hat{W}$ d'éléments :

$$\hat{W}_l = \sum_{j=1}^k \hat{C}_{lj} A_j^{-1} Z_j, \quad l = 1, 2 \quad j = 1, \dots, k$$

Calcul de la matrice de Fisher estimée $\hat{I}$

Les éléments de la matrice d'information de Fisher estimée sont donnés comme suit :

$$\hat{i}_{11} = \frac{1}{n} \sum_{i=1}^n \frac{\delta_i}{\left(\hat{\alpha} + \hat{\lambda} t_i \right)^2}$$

$$\hat{i}_{22} = \frac{1}{n} \sum_{i=1}^n \frac{\delta_i t_i^2}{\left(\hat{\alpha} + \hat{\lambda} t_i \right)^2}$$

$$\hat{t}_{12} = \hat{t}_{21} = \frac{1}{n} \sum_{i=1}^n \frac{\delta_i t_i}{(\hat{\alpha} + \hat{\lambda} t_i)^2}$$

Calcul de la matrice $\hat{G}$

les éléments de la matrice $\hat{G} = [\hat{g}_{uv}]_{2 \times 2}$ sont obtenus comme suit :

$$\begin{aligned} \hat{g}_{11} &= \frac{1}{n} \sum_{i=1}^n \frac{\delta_i}{(\hat{\alpha} + \hat{\lambda} t_i)^2} - \frac{A_j^{-1}}{n^2} \sum_{j=1}^k \left(\sum_{i: x_i \in I_j} \frac{\delta_i}{\hat{\alpha} + \hat{\lambda} t_i} \right)^2 \\ \hat{g}_{22} &= \frac{1}{n} \sum_{i=1}^n \frac{\delta_i t_i^2}{(\hat{\alpha} + \hat{\lambda} t_i)^2} - \frac{A_j^{-1}}{n^2} \sum_{j=1}^k \left(\sum_{i: x_i \in I_j} \frac{\delta_i t_i}{\hat{\alpha} + \hat{\lambda} t_i} \right)^2 \\ \hat{g}_{12} &= \frac{1}{n} \sum_{i=1}^n \frac{\delta_i t_i}{(\hat{\alpha} + \hat{\lambda} t_i)^2} - \frac{A_j^{-1}}{n^2} \sum_{j=1}^k \left(\sum_{i: x_i \in I_j} \frac{\delta_i}{\hat{\alpha} + \hat{\lambda} t_i} \right) \left(\sum_{i: x_i \in I_j} \frac{\delta_i t_i}{\hat{\alpha} + \hat{\lambda} t_i} \right) \end{aligned}$$

Enfin, on déduit la statistique Y_n^2 du test proposé :

$$Y_n^2 = \sum_{j=1}^k \frac{(U_j - e_j)^2}{U_j} + Q$$

Cas où $\hat{\gamma} = (\hat{\alpha}, \hat{\lambda})^T$ est inconnu

$$\begin{aligned} Y_n^2 &= \sum_{j=1}^k \frac{(U_j - e_j)^2}{U_j} + \frac{W_1^2}{n |G|} \left(\hat{t}_{22} - \sum_{j=1}^k [\hat{C}_{2j}]^2 A_j^{-1} \right) \\ &\quad + \frac{W_2^2}{n |G|} \left(\hat{t}_{11} - \sum_{j=1}^k [\hat{C}_{1j}]^2 A_j^{-1} \right) - \frac{2W_1 W_2}{n |G|} \left(\hat{t}_{12} - \sum_{j=1}^k \hat{C}_{1j} \hat{C}_{2j} A_j^{-1} \right) \end{aligned}$$

Où $|G|$ est le déterminant de la matrice G .

3.4.1 Simulation des valeurs du critère de test

Afin d'étudier la maniabilité et l'efficacité du test d'ajustement pour le modèle exponentiel linéaire $LE(\hat{\gamma})$ développé dans cette section, on a réalisé une importante étude par simulations numériques.

3.4. Test d'ajustement pour le modèle LE

Nous générons 10,000 échantillons de données censurées dont les tailles respectives des échantillons sont $n_1 = 15, n_2 = 25, n_3 = 50, n_4 = 130, n_5 = 350, n_6 = 500$. On calcule les valeurs du critère de test Y_n^2 pour chaque échantillon, ensuite on compte le pourcentage du nombre de rejet de l'hypothèse nulle H_0 pour différents niveaux de signification que l'on va comparer aux niveaux de signification théoriques correspondants.

Cas où $\hat{\gamma} = (\hat{\alpha}, \hat{\lambda})^T$ est inconnu

$N = 10,000$	$n_1 = 15$	$n_2 = 25$	$n_3 = 50$	$n_4 = 130$	$n_5 = 350$	$n_6 = 500$
$\varepsilon = 1\%$	0.0083	0.0089	0.0092	0.0098	0.0112	0.0104
$\varepsilon = 5\%$	0.0479	0.0488	0.0493	0.0498	0.0513	0.0502
$\varepsilon = 10\%$	0.0972	0.0979	0.0982	0.0988	0.1012	0.1001

Cas où α est connu

$N = 10,000$	$n_1 = 15$	$n_2 = 25$	$n_3 = 50$	$n_4 = 130$	$n_5 = 350$	$n_6 = 500$
$\varepsilon = 1\%$	0.0089	0.0082	0.0094	0.0098	0.0111	0.0103
$\varepsilon = 5\%$	0.0476	0.0488	0.0496	0.0499	0.0507	0.0502
$\varepsilon = 10\%$	0.0981	0.0984	0.0994	0.0999	0.1008	0.1003

Cas où λ est connu

$N = 10,000$	$n_1 = 15$	$n_2 = 25$	$n_3 = 50$	$n_4 = 130$	$n_5 = 350$	$n_6 = 500$
$\varepsilon = 1\%$	0.0085	0.0089	0.0092	0.0095	0.0099	0.0101
$\varepsilon = 5\%$	0.0489	0.0482	0.0495	0.0499	0.0512	0.0504
$\varepsilon = 10\%$	0.0980	0.0984	0.0989	0.0991	0.0999	0.1002

Tab.7. Comparaison des risques empiriques et des risques théoriques correspondants

($\varepsilon = 0, 1, 0, 05, 0, 01$) pour les données censurées.

Les résultats des simulations confirment les résultats théoriques. Nous pouvons conclure que le test proposé est bien adapté pour valider la distribution $LE(\hat{\gamma})$.

3.4.2 Exemple de données de fiabilité

Dans cet exemple, nous appliquons le test d'ajustement à un modèle LE pour un ensemble de données considérées dans l'étude de Whitmore (1983). Il a examiné les taux de pannes de 20 cellules de réduction de l'aluminium. Les nombres de pannes, en unités de 1000 jours sont donnés ci-dessous :

0.468 – 0.725 – 0.838 – 0.853 – 0.965 – 1.139 – 1.142 – 1.304 – 1.317 – 1.427 – 1.554 – 1.658 – 1.764 – 1.776 – 1.990 – 2.010 – 2.224 – 2.279* – 2.244* – 2.286*.

Sous l'hypothèse nulle H_0 selon laquelle ces données suivent une distribution exponentielle linéaire LE , nous trouvons les estimateurs du maximum de vraisemblance des paramètres inconnus :

$$\hat{\alpha} = -0.5350 \quad \text{et} \quad \hat{\lambda} = 1.2950$$

Pour le calcul de la statistique Y^2 , nous choisissons par exemple, $k = 5$ intervalles de regroupement de données, les calculs nécessaires pour le critère de test sont donnés dans la table suivante :

$\hat{a}_j$	0.9896	1.1547	1.7547	2.000	2.286
U_j	5	2	5	3	2
e_j	3.3997	3.3997	3.3997	3.3997	3.3997
$\hat{C}_{1j}$	1.0760	0.1061	0.1889	0.0813	0.0454
$\hat{C}_{2j}$	0.6375	0.1210	0.2711	0.1494	0.0960

Table 8 : Calculs de $\hat{a}_j, e_j, U_j, \hat{C}_{1j}, \hat{C}_{2j}$ pour les données de Whitmore.

La matrice estimée $\hat{W}$:

$$\hat{W} = \left(\hat{W}_1, \hat{W}_2 \right)^T = (1.2875, 0.5314)^T$$

Par conséquent, la matrice $\hat{G}$:

$$\hat{G}_{2 \times 2} = \begin{pmatrix} 5.9967 & 2.4774 \\ 2.4774 & 1.0234 \end{pmatrix}$$

On obtient alors la valeur de la statistique de test Y_n^2 :

$$Y_n^2 = X^2 + Q = 3.0368 + 0.1893 = 3.2261$$

Puisque la valeur critique est $\chi_{0.05}^2(5) = 11.07050 > Y_n^2 = 3.2261$, donc nous pouvons affirmer avec un risque $\varepsilon = 0,05$ que les taux de défaillance des cellules de réduction de l'aluminium peuvent être modélisées par une distribution LE

3.5 Test d'ajustement pour le modèle $Kw - GEE$

La distribution généralisée Kumaraswamy exponentielle exponentiée introduite récemment par Mohammed (2014) est une nouvelle distribution très flexible et très intéressante pour les applications. Pour comparer ce modèle au modèle exponentiel exponentié généralisé GEE , l'auteur a utilisé le test du rapport de vraisemblance et les critères d'information d'AIC, BIC et CAIC.

Dans cette section, nous nous proposons de construire un test d'ajustement du type du chi-deux modifié pour valider un tel modèle en utilisant l'approche de Bagdonavicius et Nikulin (2011). Il faut d'abord estimer les paramètres $\alpha, \lambda, \sigma, \theta$ sur données initiales, puis calculer les matrices $\widehat{W}$, $\widehat{C}$, $\widehat{I} = \left(\widehat{i}_{ll'}\right)_{s \times s}$ et $\widehat{G}$.

Le choix de $\hat{a}_j$ lorsque la distribution est celle d'une $Kw - GEE$, est obtenu comme suit :

$$\hat{a}_j = -\frac{1}{\hat{\lambda}} \ln \left\{ 1 - \left[1 - \exp \left(\frac{\sum_{l=1}^{i-1} H(t_l, \hat{\gamma}) - E_j}{\theta(n-i+1)} \right) \right]^{1/\hat{\alpha}\hat{\sigma}} \right\}, \quad \text{pour } j = 1, \dots, k-1, \quad \hat{a}_k = t_{(n)},$$

où

$$E_j = -\frac{j}{k-1} \sum_{i=1}^n \hat{\theta} \ln \left(1 - (1 - e^{-\hat{\lambda}t_i})^{\hat{\alpha}\hat{\sigma}} \right),$$

Calcul de la matrice d'information de Fisher estimée

Après quelques simplifications, on a obtenu les résultats suivants :

$$\hat{i}_{11} = \frac{1}{n} \sum_{i=1}^n \delta_i \left[\frac{1}{\hat{\alpha}} + \frac{\hat{\sigma} \ln(1 - e^{-\hat{\lambda}t_i})}{1 - (1 - e^{-\hat{\lambda}t_i})^{\hat{\alpha}\hat{\sigma}}} \right]^2$$

$$\hat{i}_{22} = \frac{1}{n} \sum_{i=1}^n \delta_i \left[\frac{1}{\hat{\lambda}} + \frac{t_i \left(1 - \hat{\alpha}\hat{\sigma}e^{-\hat{\lambda}t_i} - (1 - e^{-\hat{\lambda}t_i})^{\hat{\alpha}\hat{\sigma}} \right)}{(1 - e^{-\hat{\lambda}t_i}) \left(1 - (1 - e^{-\hat{\lambda}t_i})^{\hat{\alpha}\hat{\sigma}} \right)} \right]^2$$

$$\hat{i}_{33} = \frac{1}{n} \sum_{i=1}^n \delta_i \left[\frac{1}{\hat{\sigma}} + \frac{\hat{\alpha} \ln(1 - e^{-\hat{\lambda}t_i})}{1 - (1 - e^{-\hat{\lambda}t_i})^{\hat{\alpha}\hat{\sigma}}} \right]^2,$$

$$\hat{i}_{44} = \frac{1}{n} \sum_{i=1}^n \frac{\delta_i}{\hat{\theta}^2}$$

$$\hat{i}_{12} = \hat{i}_{21} = \frac{1}{n} \sum_{i=1}^n \delta_i \left[\frac{1}{\hat{\alpha}} + \frac{\hat{\sigma} \ln(1 - e^{-\hat{\lambda}t_i})}{1 - (1 - e^{-\hat{\lambda}t_i})^{\hat{\alpha}\hat{\sigma}}} \right] \left[\frac{1}{\hat{\lambda}} + \frac{t_i \left(1 - \hat{\alpha}\hat{\sigma}e^{-\hat{\lambda}t_i} - (1 - e^{-\hat{\lambda}t_i})^{\hat{\alpha}\hat{\sigma}} \right)}{(1 - e^{-\hat{\lambda}t_i}) \left(1 - (1 - e^{-\hat{\lambda}t_i})^{\hat{\alpha}\hat{\sigma}} \right)} \right]$$

$$\hat{i}_{13} = \hat{i}_{31} = \frac{1}{n} \sum_{i=1}^n \delta_i \left[\frac{1}{\hat{\alpha}} + \frac{\hat{\sigma} \ln(1 - e^{-\hat{\lambda}t_i})}{1 - (1 - e^{-\hat{\lambda}t_i})^{\hat{\alpha}\hat{\sigma}}} \right] \left[\frac{1}{\hat{\sigma}} + \frac{\hat{\alpha} \ln(1 - e^{-\hat{\lambda}t_i})}{1 - (1 - e^{-\hat{\lambda}t_i})^{\hat{\alpha}\hat{\sigma}}} \right]$$

$$\hat{i}_{14} = \hat{i}_{41} = \frac{1}{n} \sum_{i=1}^n \delta_i \left[\frac{1}{\hat{\alpha}\hat{\theta}} + \frac{\hat{\sigma} \ln(1 - e^{-\hat{\lambda}t_i})}{(1 - (1 - e^{-\hat{\lambda}t_i})^{\hat{\alpha}\hat{\sigma}}) \hat{\theta}} \right]$$

$$\hat{i}_{23} = \hat{i}_{32} = \frac{1}{n} \sum_{i=1}^n \delta_i \left[\frac{1}{\hat{\lambda}} + \frac{t_i \left(1 - \hat{\alpha}\hat{\sigma}e^{-\hat{\lambda}t_i} - (1 - e^{-\hat{\lambda}t_i})^{\hat{\alpha}\hat{\sigma}} \right)}{(1 - e^{-\hat{\lambda}t_i}) \left(1 - (1 - e^{-\hat{\lambda}t_i})^{\hat{\alpha}\hat{\sigma}} \right)} \right] \left[\frac{1}{\hat{\sigma}} + \frac{\hat{\alpha} \ln(1 - e^{-\hat{\lambda}t_i})}{1 - (1 - e^{-\hat{\lambda}t_i})^{\hat{\alpha}\hat{\sigma}}} \right]$$

$$\hat{i}_{24} = \hat{i}_{42} = \frac{1}{n} \sum_{i=1}^n \delta_i \left[\frac{1}{\hat{\lambda}\hat{\theta}} + \frac{t_i \left(1 - \hat{\alpha}\hat{\sigma}e^{-\hat{\lambda}t_i} - (1 - e^{-\hat{\lambda}t_i})^{\hat{\alpha}\hat{\sigma}} \right)}{(1 - e^{-\hat{\lambda}t_i}) \left(1 - (1 - e^{-\hat{\lambda}t_i})^{\hat{\alpha}\hat{\sigma}} \right)} \right]$$

$$\hat{i}_{34} = \hat{i}_{43} = \frac{1}{n} \sum_{i=1}^n \delta_i \left[\frac{1}{\hat{\sigma}\hat{\theta}} + \frac{\hat{\alpha} \ln(1 - e^{-\hat{\lambda}t_i})}{(1 - (1 - e^{-\hat{\lambda}t_i})^{\hat{\alpha}\hat{\sigma}}) \hat{\theta}} \right]$$

Calcul des matrices $\hat{W}$ et $\hat{C}$

De même, nous trouvons les éléments de la matrice $\hat{C}$

$$\begin{aligned}\hat{C}_{1j} &= \frac{1}{n} \sum_{i:t_i \in I_j} \delta_i \left[\frac{1}{\hat{\alpha}} + \frac{\hat{\sigma} \ln(1 - e^{-\hat{\lambda}t_i})}{1 - (1 - e^{-\hat{\lambda}t_i})^{\hat{\alpha}\hat{\sigma}}} \right] \\ \hat{C}_{2j} &= \frac{1}{n} \sum_{i:t_i \in I_j} \delta_i \left[\frac{1}{\hat{\lambda}} - \frac{t_i \left(1 - \hat{\alpha}\hat{\sigma}e^{-\hat{\lambda}t_i} - (1 - e^{-\hat{\lambda}t_i})^{\hat{\alpha}\hat{\sigma}} \right)}{(1 - e^{-\hat{\lambda}t_i}) \left(1 - (1 - e^{-\hat{\lambda}t_i})^{\hat{\alpha}\hat{\sigma}} \right)} \right] \\ \hat{C}_{3j} &= \frac{1}{n} \sum_{i:t_i \in I_j} \delta_i \left[\frac{1}{\hat{\sigma}} + \frac{\hat{\alpha} \ln(1 - e^{-\hat{\lambda}t_i})}{1 - (1 - e^{-\hat{\lambda}t_i})^{\hat{\alpha}\hat{\sigma}}} \right] \\ \hat{C}_{4j} &= \frac{1}{n} \sum_{i:x_i \in I_j} \frac{\delta_i}{\hat{\theta}}\end{aligned}$$

On déduit la matrice $\hat{W}$ d'éléments :

$$\hat{W}_l = \sum_{j=1}^k \hat{C}_{lj} A_j^{-1} Z_j, \quad l = 1, 2, 3, 4 \quad j = 1, \dots, k$$

et donc, la statistique du test :

$$Y_n^2(\hat{\gamma}) = \sum_{j=1}^k \frac{(U_j - e_j)^2}{U_j} + \hat{W}^T \left[\hat{u}_{ll'} - \sum_{j=1}^k \hat{C}_{lj} \hat{C}_{l'j} \hat{A}_j^{-1} \right]^{-1} \hat{W}$$

Sous l'hypothèse nulle H_0 , la statistique Y_n^2 suit asymptotiquement une loi de χ_k^2 à k degrés de liberté.

3.5.1 Simulation des valeurs du critère de test

Pour confirmer les résultats théoriques, on mène une série de simulations numériques. Sous l'hypothèse nulle H_0 selon laquelle la variable suit une distribution $Kw - GEE$, on génère $N = 10.000$ fois, des échantillons de tailles respectives $n_1 = 15$, $n_2 = 25$, $n_3 = 50$, $n_4 = 130$, $n_5 = 350$ et $n_6 = 500$, de données censurées provenant de cette distribution. Nous calculons les différentes valeurs Y^2 du critère de test proposé. Ensuite, les différents niveaux empiriques de rejet de l'hypothèse nulle H_0 , lorsque $Y^2 > \chi_\varepsilon^2$ sont comparés à leurs niveaux

de signification théoriques ε ($\varepsilon = 0, 01, 0, 05, 0, 10$) . Les résultats sont donnés dans la table 9.

$N = 10,000$	$n_1 = 15$	$n_2 = 25$	$n_3 = 50$	$n_4 = 130$	$n_5 = 350$	$n_6 = 500$
$\varepsilon = 1\%$	0.0056	0.0072	0.0078	0.0079	0.0089	0.0098
$\varepsilon = 5\%$	0.0456	0.0479	0.0490	0.0496	0.0501	0.0502
$\varepsilon = 10\%$	0.0984	0.0982	0.0988	0.0992	0.0998	0.1002

Tab. 9. Comparaison des risques théoriques et des risques empiriques correspondants

$$\varepsilon = 0, 01 ; 0, 05 ; 0, 10$$

Nous observons, en tenant compte des erreurs des simulations, que les niveaux simulés pour la statistique Y^2 coïncident avec ceux correspondants des niveaux théoriques de la distribution du chi-deux avec k degrés de liberté. En conséquence, nous pouvons dire que le test proposé dans ce travail, peut ajuster convenablement les données provenant d'un modèle de $Kw - GEE$.

3.5.2 Application à des données de survie

Pike (1966) a donné des données d'une enquête de laboratoire dans laquelle les vagines de rats ont été peintes avec le DMBA cancérigène, et le nombre de jours T jusqu'à apparition d'un carcinome a été enregistré. Les données ci-dessous concernent un groupe de 19 rats (Groupe 1 dans l'article de Pike). Les deux observations avec des astérisques sont des temps de censure.

143, 164, 188, 188, 190, 192, 206, 209, 213, 216, 216*, 220, 227, 230, 234, 244*, 246, 265, 304, .

Sous l'hypothèse nulle H_0 , nous trouvons les estimateurs du maximum de vraisemblance des paramètres inconnus de la distribution $KwGEE$:

$$\hat{\alpha} = 8.7659, \hat{\lambda} = 6.6589, \hat{\sigma} = 2.1589, \hat{\theta} = 4.3521.$$

Pour le calcul de la statistique Y^2 , nous choisissons par exemple, $k = 5$ intervalles de regroupement de données, les résultats intermédiaires sont donnés dans la table10 suivante :

$\hat{a}_j$	192.1	213.85	220.35	262.95	304
U_j	5	4	2	4	2
e_j	5.2658	5.2658	5.2658	5.2658	5.2658
$\hat{C}_{1j}$	2.1458	1.3659	0.2594	1.6258	0.0214
$\hat{C}_{2j}$	-5.2658	0.2154	-4.2365	0.4156	-0.2145
$\hat{C}_{3j}$	2.1478	1.3256	3.0258	1.0245	0.0269
$\hat{C}_{4j}$	-0.2145	-0.2456	0.1235	0.0145	-0.4523

Table 10 : Calculs de $\hat{a}_j, e_j, U_j, \hat{C}_{1j}, \hat{C}_{2j}, \hat{C}_{3j}, \hat{C}_{4j}$ des données de Pike

La matrice estimée $\hat{W}$:

$$\hat{W} = \left(\hat{W}_1, \hat{W}_2, \hat{W}_3, \hat{W}_4 \right)^T = (2.1578, 1.2689, 0.2147, -0.2145)^T$$

La valeur de la statistique de test Y_n^2 est :

$$Y_n^2 = X^2 + Q = 4.215 + 2.1578 = 6.3728$$

Puisque la valeur critique est $\chi_{0.05}^2(5) = 11.07050 > Y_n^2 = 6.3728$, on peut dire que les données de Pike suivent approximativement un modèle *Kw - GEE*.

3.6 Test d'ajustement pour le modèle *TGLE*

De la même manière en appliquant l'approche de Bagdonavicius et Nikulin (2011), on fournit un critère de test d'ajustement capable de vérifier l'adéquation de la transformée de la distribution exponentielle linéaire généralisée *TGLE*. Comme on l'a déjà souligné, cette distribution constitue une généralisation de plusieurs distributions connues. Avec la statistique de test que l'on va construire, on peut déduire des critères de tests d'ajustement pour tous les sous-modèles.

Supposons vraie l'hypothèse nulle H_0 qu'un échantillon censuré provienne d'une distribution *TGLE*. Comme les estimateurs du maximum de vraisemblance des paramètres inconnus α, λ, β et σ , ont été développés dans le premier chapitre, nous calculons les nombres de défaillances théoriques de manière qu'ils soient tous égaux dans les différentes classes de groupement de données I_j .

Pour cela, le choix des limites $\hat{a}_j$ de ces classes est obtenu comme suit :

$$\hat{a}_j = -\frac{\hat{\alpha}}{\hat{\lambda}} + \frac{1}{\hat{\lambda}} \sqrt{\hat{\alpha}^2 + 2\hat{\lambda} \left(-\ln \left\{ \frac{\hat{\sigma} - 1 + \sqrt{(1 - \hat{\sigma})^2 + 4\hat{\sigma} \exp \left(\frac{\sum_{l=1}^{i-1} H(t_l, \hat{\sigma}) - E_j}{(n-i+1)} \right)}{2\hat{\sigma}} \right\} \right)^{1/\hat{\beta}}}$$

pour $j = 1, \dots, k-1$, $\hat{a}_k = \max(t_{(l)}, \tau)$,

Calcul de la matrice $\hat{W}$ et $\hat{C}$

Pour le calcul de la matrice $\hat{W}$ on calcule d'abord les dérivées partielles de la fonction $\ln h(t; \gamma)$ par rapport à α, λ, β et σ :

$$\begin{aligned} \hat{C}_{1j} &= \frac{1}{n} \sum_{i:t_i \in I_j} \delta_i \left[\frac{1}{\hat{\alpha} + \hat{\lambda} t_i} + \frac{(\hat{\beta} - 1) t_i}{u_i} + \hat{\beta} \hat{\sigma} (1 - \hat{\sigma}) t_i u_i^{\hat{\beta}-1} M(t_i, \hat{\gamma}) \right] \\ \hat{C}_{2j} &= \frac{1}{n} \sum_{i:t_i \in I_j} \delta_i \left[\frac{t_i}{\hat{\alpha} + \hat{\lambda} t_i} + \frac{(\hat{\beta} - 1) t_i^2}{2u_i} + \hat{\beta} \hat{\sigma} (1 - \hat{\sigma}) t_i^2 u_i^{\hat{\beta}-1} M(t_i, \hat{\gamma}) \right] \\ \hat{C}_{3j} &= \frac{1}{n} \sum_{i:t_i \in I_j} \delta_i \left[\frac{1}{\hat{\beta}} + \ln(u_i) + \sigma (1 - \sigma) M(t_i, \gamma) u_i^\beta \ln(u_i) \right] \\ \hat{C}_{4j} &= \frac{1}{n} \sum_{i:t_i \in I_j} \delta_i M(t_i, \gamma) \end{aligned}$$

Où

$$M(t_i, \gamma) = e^{-u_i^\beta} \left(1 - \sigma + \sigma e^{-u_i^\beta} \right)^{-1} \left(1 - \sigma + 2\sigma e^{-u_i^\beta} \right)^{-1}$$

On déduit la matrice $\hat{W}$ d'éléments :

$$\hat{W}_l = \sum_{j=1}^r \hat{C}_{lj} A_j^{-1} Z_j, \quad l = 1, 2, 3, 4, \quad j = 1, \dots, k$$

Calcul de la matrice de Fisher estimée $\hat{I}$

Dans ce cas, à partir de la fonction de la log-vraisemblance, nous pouvons estimer la matrice d'information de Fisher :

$$\hat{i}_{11} = \frac{1}{n} \sum_{i=1}^n \delta_i \left(\frac{1}{\hat{\alpha} + \hat{\lambda} t_i} + \frac{(\hat{\beta} - 1) t_i}{u_i} + \hat{\beta} \hat{\sigma} (1 - \hat{\sigma}) t_i u_i^{\hat{\beta}-1} M(t_i, \hat{\gamma}) \right)^2$$

$$\hat{i}_{22} = \frac{1}{n} \sum_{i=1}^n \delta_i \left(\frac{t_i}{\hat{\alpha} + \hat{\lambda} t_i} + \frac{(\hat{\beta} - 1) t_i^2}{2u_i} + \hat{\beta} \hat{\sigma} (1 - \hat{\sigma}) t_i^2 u_i^{\hat{\beta}-1} M(t_i, \hat{\gamma}) \right)^2$$

$$\hat{i}_{33} = \frac{1}{n} \sum_{i=1}^n \delta_i \left(\frac{1}{\hat{\beta}} + \ln(u_i) + \sigma (1 - \sigma) M(t_i, \gamma) u_i^{\beta} \ln(u_i) \right)^2$$

$$\hat{i}_{44} = \frac{1}{n} \sum_{i=1}^n \delta_i M^2(t_i, \gamma)$$

$$\begin{aligned} \hat{i}_{12} = & \frac{1}{n} \sum_{i=1}^n \delta_i \left(\frac{1}{\hat{\alpha} + \hat{\lambda} t_i} + \frac{(\hat{\beta} - 1) t_i}{u_i} + \hat{\beta} \hat{\sigma} (1 - \hat{\sigma}) t_i u_i^{\hat{\beta}-1} M(t_i, \hat{\gamma}) \right) \\ & \times \left(\frac{t_i}{\hat{\alpha} + \hat{\lambda} t_i} + \frac{(\hat{\beta} - 1) t_i^2}{2u_i} + \hat{\beta} \hat{\sigma} (1 - \hat{\sigma}) t_i^2 u_i^{\hat{\beta}-1} M(t_i, \hat{\gamma}) \right) \end{aligned}$$

$$\begin{aligned} \hat{i}_{13} = & \frac{1}{n} \sum_{i=1}^n \delta_i \left(\frac{1}{\hat{\alpha} + \hat{\lambda} t_i} + \frac{(\hat{\beta} - 1) t_i}{u_i} + \hat{\beta} \hat{\sigma} (1 - \hat{\sigma}) t_i u_i^{\hat{\beta}-1} M(t_i, \hat{\gamma}) \right) \\ & \times \left(\frac{1}{\hat{\beta}} + \ln(u_i) + \sigma (1 - \sigma) M(t_i, \gamma) u_i^{\beta} \ln(u_i) \right) \end{aligned}$$

$$\hat{i}_{14} = \frac{1}{n} \sum_{i=1}^n \delta_i \left(\frac{M(t_i, \gamma)}{\hat{\alpha} + \hat{\lambda} t_i} + \frac{(\hat{\beta} - 1) t_i}{u_i} M(t_i, \gamma) + \hat{\beta} \hat{\sigma} (1 - \hat{\sigma}) t_i u_i^{\hat{\beta}-1} M^2(t_i, \hat{\gamma}) \right)$$

$$\hat{i}_{23} = \frac{1}{n} \sum_{i=1}^n \delta_i \left(\frac{t_i}{\hat{\alpha} + \hat{\lambda} t_i} + \frac{(\hat{\beta} - 1) t_i^2}{2u_i} + \hat{\beta} \hat{\sigma} (1 - \hat{\sigma}) t_i^2 u_i^{\hat{\beta}-1} M(t_i, \hat{\gamma}) \right) \\ \times \left(\frac{1}{\hat{\beta}} + \ln(u_i) + \sigma (1 - \sigma) M(t_i, \gamma) u_i^{\beta} \ln(u_i) \right)$$

$$\hat{i}_{24} = \frac{1}{n} \sum_{i=1}^n \delta_i \left(\frac{M(t_i, \gamma) t_i}{\hat{\alpha} + \hat{\lambda} t_i} + \frac{(\hat{\beta} - 1) t_i^2}{2u_i} M(t_i, \gamma) + \hat{\beta} \hat{\sigma} (1 - \hat{\sigma}) t_i^2 u_i^{\hat{\beta}-1} M^2(t_i, \hat{\gamma}) \right)$$

$$\hat{i}_{34} = \frac{1}{n} \sum_{i=1}^n \delta_i \left(\frac{M(t_i, \gamma)}{\hat{\beta}} + M(t_i, \gamma) \ln(u_i) + \sigma (1 - \sigma) M^2(t_i, \gamma) u_i^{\beta} \ln(u_i) \right)$$

On déduit la matrice $\hat{G} = [\hat{g}_{ll'}]_{4 \times 4}$

$$\hat{g}_{ll'} = \hat{i}_{ll'} - \sum_{j=1}^r \hat{C}_{lj} \hat{C}_{l'j} A_j^{-1}$$

Tous les composants de la statistique du test Y_n^2 sont calculés explicitement, on déduit alors :

$$Y_n^2(\hat{\gamma}) = \sum_{j=1}^r \frac{(U_j - e_j)^2}{U_j} + \hat{W}^T \left[\hat{i}_{ll'} - \sum_{j=1}^r \hat{C}_{lj} \hat{C}_{l'j} \hat{A}_j^{-1} \right]^{-1} \hat{W}$$

Notons que selon les différentes paramétrisations de cette distribution, nous pouvons déduire des critères de test d'ajustement pour les différents sous-modèles qui sont les transformées de la distribution exponentielle linéaire *LE*, exponentielle linéaire généralisée *GLE*, Weibull *W* et Rayleigh *Ray*, en plus des transformées des distributions exponentielles linéaires simples *TLE*.

3.6.1 Simulation des valeurs du critère de test

Pour conforter les résultats obtenus, on mène une importante étude de simulations. On génère $N = 10,000$ échantillons de tailles respectives $n_1 = 15$, $n_2 = 25$, $n_3 = 50$, $n_4 = 130$, $n_5 = 350$ et $n_6 = 500$, suivant le modèle *TGLE*. Puis, nous calculons les valeurs de la statistique Y_n^2 pour chaque échantillon. Ensuite, on compte le nombre de cas de rejets de

3.6. Test d'ajustement pour le modèle $TGLE$

l'hypothèse nulle H_0 que l'on reporte dans le tableau suivant en comparaison avec les niveaux théoriques correspondants. Ensuite et selon les différentes paramétrisations, on applique nos résultats aux différents sous-modèles de la distribution $TGLE$ pour calculer les critères de tests correspondants. Les résultats obtenus sont réunis dans les tableaux qui suivent.

Dans le cas $TGLE$

$N = 10,000$	$n_1 = 15$	$n_2 = 25$	$n_3 = 50$	$n_4 = 130$	$n_5 = 350$	$n_6 = 500$
$\varepsilon = 1\%$	0.0082	0.0087	0.0089	0.0094	0.0097	0.0101
$\varepsilon = 5\%$	0.0472	0.0477	0.0479	0.0493	0.0498	0.0500
$\varepsilon = 10\%$	0.0980	0.0985	0.0989	0.0995	0.0996	0.1000

Dans le cas GLE (exponentielle linéaire généralisée)

$N = 10,000$	$n_1 = 15$	$n_2 = 25$	$n_3 = 50$	$n_4 = 130$	$n_5 = 350$	$n_6 = 500$
$\varepsilon = 1\%$	0.0084	0.0089	0.0092	0.0096	0.0099	0.0100
$\varepsilon = 5\%$	0.0486	0.0489	0.0492	0.0497	0.0499	0.0501
$\varepsilon = 10\%$	0.0980	0.0989	0.0992	0.0996	0.0998	0.1001

Dans le cas LE (exponentielle linéaire)

$N = 10,000$	$n_1 = 15$	$n_2 = 25$	$n_3 = 50$	$n_4 = 130$	$n_5 = 350$	$n_6 = 500$
$\varepsilon = 1\%$	0.0083	0.0088	0.0092	0.0094	0.0097	0.0099
$\varepsilon = 5\%$	0.0490	0.0492	0.0494	0.0497	0.0507	0.0502
$\varepsilon = 10\%$	0.0984	0.0989	0.0990	0.0996	0.0998	0.0999

Dans le cas TLE (la transformée de l'exponentielle linéaire)

$N = 10,000$	$n_1 = 15$	$n_2 = 25$	$n_3 = 50$	$n_4 = 130$	$n_5 = 350$	$n_6 = 500$
$\varepsilon = 1\%$	0.0089	0.0045	0.0054	0.0096	0.0101	0.0101
$\varepsilon = 5\%$	0.0486	0.0488	0.0492	0.0495	0.0498	0.0504
$\varepsilon = 10\%$	0.0982	0.0984	0.0991	0.0998	0.01012	0.1003

Dans le cas *TR* (la transformée de Rayleigh)

$N = 10,000$	$n_1 = 15$	$n_2 = 25$	$n_3 = 50$	$n_4 = 130$	$n_5 = 350$	$n_6 = 500$
$\varepsilon = 1\%$	0.0080	0.0084	0.0089	0.0095	0.0098	0.0102
$\varepsilon = 5\%$	0.0475	0.0480	0.0482	0.0495	0.0508	0.0505
$\varepsilon = 10\%$	0.0946	0.0965	0.0979	0.0983	0.0996	0.1003

Dans le cas *Ray* (Rayleigh)

$N = 10,000$	$n_1 = 15$	$n_2 = 25$	$n_3 = 50$	$n_4 = 130$	$n_5 = 350$	$n_6 = 500$
$\varepsilon = 1\%$	0.0083	0.0089	0.0092	0.0098	0.0109	0.0101
$\varepsilon = 5\%$	0.0479	0.0488	0.0493	0.0498	0.0507	0.0502
$\varepsilon = 10\%$	0.0972	0.0979	0.0982	0.0988	0.1009	0.1001

Dans le cas *TW* (transformée de Weibull)

$N = 10,000$	$n_1 = 15$	$n_2 = 25$	$n_3 = 50$	$n_4 = 130$	$n_5 = 350$	$n_6 = 500$
$\varepsilon = 1\%$	0.0085	0.0089	0.0092	0.0095	0.0099	0.0101
$\varepsilon = 5\%$	0.0489	0.0482	0.0495	0.0499	0.0499	0.0504
$\varepsilon = 10\%$	0.0985	0.0989	0.0995	0.0998	0.0999	0.1002

Dans le cas *W* (Weibull)

$N = 10,000$	$n_1 = 15$	$n_2 = 25$	$n_3 = 50$	$n_4 = 130$	$n_5 = 350$	$n_6 = 500$
$\varepsilon = 1\%$	0.0089	0.0082	0.0094	0.0098	0.0101	0.0102
$\varepsilon = 5\%$	0.0476	0.0488	0.0496	0.0499	0.0507	0.0501
$\varepsilon = 10\%$	0.0981	0.0984	0.0994	0.0999	0.1008	0.1001

Tab.11. Comparaison des risques empiriques et des risques théoriques correspondants
($\varepsilon = 0, 1, 0, 05, 0, 01$) pour les données censurées

De ces résultats, et en tenant compte des erreurs de simulation, nous observons que les niveaux de signification simulés de la statistique Y_n^2 du test proposé, coïncident avec ceux correspondant aux niveaux théoriques de la distribution du chi-deux avec k (ici $k = 10$) degrés de liberté et ceci pour la distribution *TGLE* et tous les sous-modèles de celle-ci. Par conséquent, nous pouvons dire que les statistiques de test proposées peuvent ajuster des données des différents modèles de manière très satisfaisante.

3.6.2 Application à des données de survie

Nous considérons des données d'échantillon de 50 patients atteints de leucémie aiguë myéloïde, signalés au registre international des transplantations de moelle osseuse. Ces patients ont eu une transplantation de moelle osseuse allogénique (allo) où la moelle d'un homologue HLA (Histocompatibility Leukocyte Antigen) a été utilisé pour reconstituer leur système immunitaire.

Les données ordonnées de cette étude sont présentées dans la table (12). où (*) représente la censure.

0.030	0.493	0.855	1.184	1.283	1.480	1.776	2.138	2.500
2.763	2.993	3.224	3.421	4.178	4.441*	5.691	5.855*	6.941*
6.941	7.993*	8.882	8.882	9.145*	11.480	11.513	12.105*	12.796
12.993*	13.849*	16.612*	17.138*	20.066	20.329*	22.368*	26.776*	28.717*
28.717*	32.928*	33.783*	34.211*	34.770*	39.539*	41.118*	45.033*	46.053*
46.941*	48.289*	57.401*	58.322*	60.625*				

Tab 12 Taux de survie de la leucémie (en mois) pour les transplantations autologues et allogéniques.

En supposant que ces données sont modélisées par une distribution *TGLE*, on utilise *R* et l'algorithme BB solve pour calculer les estimateurs du vecteur de paramètres $\hat{\gamma}$ de γ maximisant la fonction de log-vraisemblance. On obtient les valeurs suivantes :

$$\hat{\alpha} = 8.7659, \hat{\lambda} = 6.6589, \hat{\beta} = 2.1589, \hat{\sigma} = 4.3521.$$

3.6. Test d'ajustement pour le modèle *TGLE*

Nous choisissons un nombre de classes $k = 5$, les éléments de la statistique Y_n^2 du test sont présentés dans le tableau qui suit :

$\hat{a}_j$	2.143	4.596	11.254	19.258	60.625
U_j	8	6	4	3	1
e_j	7.2489	7.2489	7.2489	7.2489	7.2489
$\hat{C}_{1j}$	0.1478	0.2158	0.1418	0.2356	0.0147
$\hat{C}_{2j}$	-0.1400	0.2415	0.31578	0.0014	0.2145
$\hat{C}_{3j}$	0.0014	-4.2515	0.9562	-2.3654	0.21587
$\hat{C}_{4j}$	0.2241	0.0130	0.1502	0.2430	0.2508

Tab 13 Calculs de $\hat{a}_j$; U_j ; e_j ; $\hat{C}_{1j}$; $\hat{C}_{2j}$; $\hat{C}_{3j}$; $\hat{C}_{4j}$

La matrice estimée $\hat{W}$:

$$\hat{W} = \left(\hat{W}_1, \hat{W}_2, \hat{W}_3, \hat{W}_4 \right)^T = (1.0325, 0.1865, -3.0251, 0.4589)^T$$

Nous déduisons alors la valeur du critère de test :

$$Y_n^2 = X^2 + Q = 6.3728 + 2.6584 = 9.0312$$

Au seuil $\varepsilon = 0.05$, la valeur critique $\chi_{0.05}^2(5) = 11.07050$, On peut dire que ces données peuvent être modélisées par la distribution *TGLE*.

4.1 Introduction

Les modèles appelés "modèles *AFT*" (accelerated life testing) ou modèles à temps de vie accéléré sont des modèles mathématiques utilisés pour décrire le comportement de durées de vie où les fonctions d'intérêt sont fonctions de variables dites explicatives représentant les différents facteurs de stress qui ont un impact direct sur ces durées de vie. L'utilité des modèles *AFT* consiste à réduire ou à augmenter les durées de vie des produits par l'accélération ou la décélération des dégradations provoquant les défaillances en un temps court. Ainsi, on peut estimer la fiabilité du matériel de manière à optimiser ses performances et la planification de sa maintenance. Ces modèles ont des applications dans tous les secteurs : la fiabilité, l'analyse de survie, la biologie, la météorologie. Les modèles *AFT* paramétriques sont construits à partir d'une distribution paramétrique dite distribution de base. Les distributions utilisées dans la littérature statistique sont la distribution classique de Weibull, l'exponentielle, la log-logistique et la log-normale (voir Bagdonavicius et Nikulin, 2002, Nikulin et al., 2007, Bagdonavicius et al., 2013) . La diversification des observations qui ne peuvent être représentées par les distributions classiques, a encouragé les chercheurs à proposer de nouveaux modèles *AFT*. On peut citer par exemple le modèle *AFT* log-gamma introduit

par Ortega et al. (2009), et le modèle *AFT* inverse Weibull généralisée étudié par Goual et Seddik-Ameur (2014).

Dans ce chapitre, on propose un modèle à temps de vie accéléré *AFT – GEE* dont la distribution de base est une exponentielle exponentielle généralisée. Après la présentation du modèle, on utilise la méthode du maximum de vraisemblance pour estimer les paramètres inconnus de la distribution. Nous terminons par la mise en place d'un test d'adéquation pour ce modèle quand les données sont censurées à droite et les paramètres inconnus permettant d'ajuster des observations à ce type de modèle de manière satisfaisante.

4.2 Le modèle AFT

Un modèle paramétrique *AFT* est défini à partir d'une fonction de survie de base notée $S_0(t)$.

Soit E l'ensemble de tous les stress possibles, défini par :

$$E = \{z(.) = (z_1(.), z_2(.), \dots, z_m(.))^T\}$$

Si $z(.)$ est constant, nous écrivons z à la place de $z(.)$, et nous notons par E_1 le sous-ensemble de E de tous les stress constants et $\beta = (\beta_0, \beta_1, \dots, \beta_m)^T$ les paramètres de la régression des covariables représentant les stress éventuels $z = (1, z_1, \dots, z_m)^T$ du modèle *AFT*.

Notons par $T_{z(.)}$ le taux de panne sous $z(.)$ et par $f_{z(.)}(t)$, $S_{z(.)}(t)$ et $F_{z(.)}(t)$ respectivement la fonction de densité, la fonction de survie et la fonction cumulative de la distribution, où :

$$S_{z(.)}(t) = P(T_{z(.)} \geq t) = 1 - F_{z(.)}(t), \quad z(.) \in E$$

La fonction risque de $T_{z(.)}$ sous $z(.)$ est :

$$h_{z(.)}(t) = \frac{f_{z(.)}(t)}{S_{z(.)}(t)} = \frac{S'_{z(.)}(t)}{S_{z(.)}(t)}$$

la fonction cumulative de risque de $T_{z(.)}$ sous $z(.)$ est :

$$H_{z(.)}(t) = -\ln(S_{z(.)}(t))$$

Définition du modèle AFT

Le modèle AFT est défini sur E , si la fonction de survie sous stress $z(.) \in E$, est

$$S_{z(.)}(t) = S_0\left(\int_0^t r(z(u))du\right), \forall z(.) \in E$$

Si $z(.) = z$ est constant, alors le modèle devient :

$$S_z(t) = S_0(r(z)t), \forall z(.) \in E_1$$

Le modèle AFT paramétrique est donné par la formule :

$$S_{z(.)}(t) = S_0\left(\int_0^t \exp\{-\beta^T z(u)\}du\right), \forall z(.) \in E$$

Si les stress sont constants dans le temps, on a :

$$S_{z(.)}(t) = S_0(\exp\{-\beta^T z\}t)$$

Le logarithme du taux de défaillance T_z sous z peut s'écrire :

$$\ln(T_z) = \beta^T z + \epsilon$$

où la fonction de survie de ϵ ne dépend pas de z et est $S(t) = S_0(\ln t)$.

4.3 Construction du modèle AFT – GEE

Pour un modèle à durée de vie accélérée AFT dont la distribution de base est une exponentielle généralisée, noté AFT – GEE, la fonction de survie $S(t)$ est donnée par :

$$\begin{aligned} S(t, \gamma) &= S_0\left(te^{-\beta^T z}\right) \\ &= 1 - \left(1 - \exp\left\{-\lambda te^{-\beta^T z}\right\}\right)^\alpha \end{aligned}$$

où S_0 est la fonction de survie de la distribution GEE.

α est le paramètre de forme et λ le paramètre d'échelle de la distribution GEE, $\beta = (\beta_0, \beta_1, \dots, \beta_m)^T$ sont les paramètres de la régression du modèle AFT et $z = (1, z_1, \dots, z_m)^T$ sont les covariables représentant les stress éventuels.

Ainsi, La fonction densité de probabilité du modèle $AFT - GEE$ est obtenue par la formule suivante :

$$f(t, \gamma) = \alpha \lambda e^{-\beta^T z} \exp \left\{ -\lambda t e^{-\beta^T z} \right\} \left(1 - \exp \left\{ -\lambda t e^{-\beta^T z} \right\} \right)^{\alpha-1}$$

Nous pouvons alors déduire le taux de hasard et le taux de hasard cumulé comme suit :

$$h(t, \gamma) = \frac{\alpha \lambda e^{-\beta^T z} \exp \left\{ -\lambda t e^{-\beta^T z} \right\} \left(1 - \exp \left\{ -\lambda t e^{-\beta^T z} \right\} \right)^{\alpha-1}}{1 - \left(1 - \exp \left\{ -\lambda t e^{-\beta^T z} \right\} \right)^{\alpha}}$$

$$H(t, \gamma) = -\ln \left(1 - \left(1 - \exp \left\{ -\lambda t e^{-\beta^T z} \right\} \right)^{\alpha} \right)$$

4.4 Estimation du maximum de vraisemblance

Soit T_i une variable aléatoire distribuée selon un modèle $AFT - GEE$ avec le vecteur de paramètres $\gamma = (\alpha, \lambda, \beta_0, \beta_1, \dots, \beta_m)^T$.

On considère un échantillon aléatoire censuré à droite composé de n observations indépendantes $t_i = \min(T_i, C_i)$ pour $i = 1, \dots, n$. La distribution de la censure C_i est supposée indépendante de celle de T_i . Dans le cas de censure, la fonction de log-vraisemblance du modèle s'écrit alors :

$$\begin{aligned} L(t, \gamma) &= \sum_{i=1}^n \delta_i \ln h(t_i, \gamma) + \sum_{i=1}^n \ln S(t_i, \gamma) \\ &= \sum_{i=1}^n \delta_i [\ln f(t_i, \gamma) - \ln S(t_i, \gamma)] + \sum_{i=1}^n \ln S(t_i, \gamma) \\ L(t, \gamma) &= \sum_{i=1}^n \delta_i [\ln(\alpha \lambda) - \beta^T z_i - t_i u_i + (\alpha - 1) \ln(1 - e^{-t_i u_i}) \\ &\quad - \ln(1 - (1 - e^{-t_i u_i})^{\alpha})] + \sum_{i=1}^n \ln(1 - (1 - e^{-t_i u_i})^{\alpha}) \end{aligned}$$

Après de longs calculs, nous obtenons les fonctions de score pour les paramètres $\alpha, \lambda, \beta_0, \beta_1, \dots, \beta_m$

$$\frac{\partial L}{\partial \alpha} = \sum_{i=1}^n \delta_i \left[\frac{1}{\alpha} + \frac{\ln(1 - e^{-t_i u_i})}{1 - (1 - e^{-t_i u_i})^{\alpha}} \right] - \sum_{i=1}^n \frac{(1 - e^{-t_i u_i})^{\alpha} \ln(1 - e^{-t_i u_i})}{1 - (1 - e^{-t_i u_i})^{\alpha}}$$

$$\frac{\partial L}{\partial \lambda} = \sum_{i=1}^n \delta_i \left[\frac{1}{\lambda} - \frac{t_i e^{-\beta^T z_i} [1 - \alpha e^{-t_i u_i} - (1 - e^{-t_i u_i})^\alpha]}{(1 - e^{-t_i u_i}) (1 - (1 - e^{-t_i u_i})^\alpha)} \right] - \sum_{i=1}^n \frac{\alpha t_i e^{-\beta^T z_i} \exp \{-t_i u_i\} (1 - e^{-t_i u_i})^{\alpha-1}}{1 - (1 - e^{-t_i u_i})^\alpha}$$

$$\frac{\partial L}{\partial \beta_0} = \sum_{i=1}^n \delta_i \left[-1 + \frac{t_i u_i [1 - \alpha e^{-t_i u_i} - (1 - e^{-t_i u_i})^\alpha]}{(1 - e^{-t_i u_i}) (1 - (1 - e^{-t_i u_i})^\alpha)} \right] + \sum_{i=1}^n \frac{\alpha t_i u_i e^{-t_i u_i} (1 - e^{-t_i u_i})^{\alpha-1}}{1 - (1 - e^{-t_i u_i})^\alpha}$$

$$\frac{\partial L}{\partial \beta_1} = \sum_{i=1}^n \delta_i \left[-z_{i1} + \frac{t_i z_{i1} u_i [1 - \alpha e^{-t_i u_i} - (1 - e^{-t_i u_i})^\alpha]}{(1 - e^{-t_i u_i}) (1 - (1 - e^{-t_i u_i})^\alpha)} \right] + \sum_{i=1}^n \frac{\alpha t_i z_{i1} u_i e^{-t_i u_i} (1 - e^{-t_i u_i})^{\alpha-1}}{1 - (1 - e^{-t_i u_i})^\alpha}$$

.....

$$\frac{\partial L}{\partial \beta_m} = \sum_{i=1}^n \delta_i \left[-z_{im} + \frac{t_i z_{im} u_i [1 - \alpha e^{-t_i u_i} - (1 - e^{-t_i u_i})^\alpha]}{(1 - e^{-t_i u_i}) (1 - (1 - e^{-t_i u_i})^\alpha)} \right] + \sum_{i=1}^n \frac{\alpha t_i z_{im} u_i e^{-t_i u_i} (1 - e^{-t_i u_i})^{\alpha-1}}{1 - (1 - e^{-t_i u_i})^\alpha}$$

Pour trouver les estimateurs du maximum de vraisemblance, on égale ces fonctions au vecteur zéro. Les méthodes utilisées pour les calculs sont numériques.

4.5 Test d'ajustement pour le modèle AFT-GEE

Dans cette section, nous allons développer une statistique de test d'ajustement pour le modèle *AFT - GEE* en utilisant la statistique proposée par Bagdonavicius et Nikulin (2011) décrite dans le chapitre deux de cette thèse, et ceci dans le cas où les paramètres sont inconnus et les données censurées à droite.

Notons que dans le travail de Bagdonavicius et al. (2013), les auteurs ont construit des tests d'ajustement pour les modèles *AFT* exponentiel, Weibull et Log-logistique alors que Goual et seddik-Ameur (2014) l'ont fait pour le modèle AFT- inverse Weibull généralisé.

Pour un temps fixé τ on partage l'intervalle $[0, \tau]$ en k sous-intervalles $I_j = [a_{j-1}, a_j[$ pour grouper les observations. Sous l'hypothèse nulle H_0 , et après avoir calculé les estimateurs du maximum de vraisemblance des paramètres $\alpha, \lambda, \beta_0, \beta_1, \beta_2, \dots, \beta_m$ sur les données initiales, nous calculons les nombres de défaillances théoriques suivants :

$$E_j = -\frac{j}{k-1} \sum_{i=1}^n \ln \left(1 - \left(1 - \exp \left\{ -\hat{\lambda} t_i e^{-\hat{\beta}^T z_i} \right\} \right)^{\hat{\alpha}} \right), \quad j = 1, \dots, k-1$$

Pour pouvoir avoir les mêmes effectifs e_j attendus dans toutes les classes, le choix de $\hat{a}_j$ dans le cas du modèle AFT – GEE , est obtenu comme suit :

$$\hat{a}_j = -\frac{1}{\hat{\lambda}} e^{\hat{\beta}^T z} \ln \left(1 - \left[1 - \exp \left(\frac{\sum_{l=1}^{i-1} H(t_l, \hat{\gamma}) - E_j}{n - i + 1} \right) \right]^{\frac{1}{\hat{\alpha}}} \right)$$

pour $j = 1, \dots, k - 1$, $\hat{a}_k = t_{(n)}$,

Calcul de la matrice $\hat{W}$ et $\hat{C}$

Le calcul de la matrice $\hat{W}$ passe par celui de la matrice $\hat{C}$ obtenue comme suit :

$$\begin{aligned} \hat{C}_{1j} &= \frac{1}{n} \sum_{i:t_i \in I_j} \delta_i \left[\frac{1}{\hat{\alpha}} + \frac{\ln(1 - e^{-t_i u_i})}{1 - (1 - e^{-t_i u_i})^{\hat{\alpha}}} \right] \\ \hat{C}_{2j} &= \frac{1}{n} \sum_{i:t_i \in I_j} \delta_i \left[\frac{1}{\hat{\lambda}} - \frac{t_i e^{-\hat{\beta}^T z_i} [1 - \hat{\alpha} e^{-t_i u_i} - (1 - e^{-t_i u_i})^{\hat{\alpha}}]}{(1 - e^{-t_i u_i}) (1 - (1 - e^{-t_i u_i})^{\hat{\alpha}})} \right] \\ \hat{C}_{3j} &= \frac{1}{n} \sum_{i:t_i \in I_j} \delta_i \left[-1 + \frac{t_i u_i [1 - \hat{\alpha} e^{-t_i u_i} - (1 - e^{-t_i u_i})^{\hat{\alpha}}]}{(1 - e^{-t_i u_i}) (1 - (1 - e^{-t_i u_i})^{\hat{\alpha}})} \right] \\ \hat{C}_{4j} &= \frac{1}{n} \sum_{i:t_i \in I_j} \delta_i \left[-z_{i1} + \frac{t_i z_{i1} u_i [1 - \hat{\alpha} e^{-t_i u_i} - (1 - e^{-t_i u_i})^{\hat{\alpha}}]}{(1 - e^{-t_i u_i}) (1 - (1 - e^{-t_i u_i})^{\hat{\alpha}})} \right] \\ &\dots\dots\dots \\ \hat{C}_{mj} &= \frac{1}{n} \sum_{i:t_i \in I_j} \delta_i \left[-z_{im} + \frac{t_i z_{im} u_i [1 - \hat{\alpha} e^{-t_i u_i} - (1 - e^{-t_i u_i})^{\hat{\alpha}}]}{(1 - e^{-t_i u_i}) (1 - (1 - e^{-t_i u_i})^{\hat{\alpha}})} \right] \end{aligned}$$

On déduit la matrice $\hat{W}$ d'éléments :

$$\hat{W}_l = \sum_{j=1}^k \hat{C}_{lj} A_j^{-1} Z_j, \quad l = 1, \dots, m \quad j = 1, \dots, k$$

Calcul de la matrice de Fisher estimée $\hat{I}$

Sans perdre de généralité, nous n'allons considérer que les coefficients β_0, β_1 de la régression, le reste se fait de manière analogue. Les éléments de la matrice d'information de Fisher estimée sont obtenus comme suit :

$$\hat{i}_{11} = \frac{1}{n} \sum_{i=1}^n \delta_i \left(\frac{1}{\hat{\alpha}} + \frac{\ln(1 - e^{-t_i u_i})}{1 - (1 - e^{-t_i u_i})^{\hat{\alpha}}} \right)^2$$

$$\hat{i}_{22} = \frac{1}{n} \sum_{i=1}^n \delta_i \left(\frac{1}{\hat{\lambda}} - \frac{t_i e^{-\hat{\beta}^T z_i} [1 - \hat{\alpha} e^{-t_i u_i} - (1 - e^{-t_i u_i})^{\hat{\alpha}}]}{(1 - e^{-t_i u_i}) (1 - (1 - e^{-t_i u_i})^{\hat{\alpha}})} \right)^2$$

$$\hat{i}_{33} = \frac{1}{n} \sum_{i=1}^n \delta_i \left(-1 + \frac{t_i u_i [1 - \hat{\alpha} e^{-t_i u_i} - (1 - e^{-t_i u_i})^{\hat{\alpha}}]}{(1 - e^{-t_i u_i}) (1 - (1 - e^{-t_i u_i})^{\hat{\alpha}})} \right)^2$$

$$\hat{i}_{44} = \frac{1}{n} \sum_{i=1}^n \delta_i \left(-z_{i1} + \frac{t_i z_{i1} u_i [1 - \hat{\alpha} e^{-t_i u_i} - (1 - e^{-t_i u_i})^{\hat{\alpha}}]}{(1 - e^{-t_i u_i}) (1 - (1 - e^{-t_i u_i})^{\hat{\alpha}})} \right)^2$$

$$\hat{i}_{12} = \hat{i}_{21} = \frac{1}{n} \sum_{i=1}^n \delta_i \left(\frac{1}{\hat{\alpha}} + \frac{\ln(1 - e^{-t_i u_i})}{1 - (1 - e^{-t_i u_i})^{\hat{\alpha}}} \right) \left(\frac{1}{\hat{\lambda}} - \frac{t_i e^{-\hat{\beta}^T z_i} [1 - \hat{\alpha} e^{-t_i u_i} - (1 - e^{-t_i u_i})^{\hat{\alpha}}]}{(1 - e^{-t_i u_i}) (1 - (1 - e^{-t_i u_i})^{\hat{\alpha}})} \right)$$

$$\hat{i}_{13} = \hat{i}_{31} = \frac{1}{n} \sum_{i=1}^n \delta_i \left(\frac{1}{\hat{\alpha}} + \frac{\ln(1 - e^{-t_i u_i})}{1 - (1 - e^{-t_i u_i})^{\hat{\alpha}}} \right) \left(-1 + \frac{t_i u_i [1 - \hat{\alpha} e^{-t_i u_i} - (1 - e^{-t_i u_i})^{\hat{\alpha}}]}{(1 - e^{-t_i u_i}) (1 - (1 - e^{-t_i u_i})^{\hat{\alpha}})} \right)$$

$$\hat{i}_{14} = \hat{i}_{41} = \frac{1}{n} \sum_{i=1}^n \delta_i \left(\frac{1}{\hat{\alpha}} + \frac{\ln(1 - e^{-t_i u_i})}{1 - (1 - e^{-t_i u_i})^{\hat{\alpha}}} \right) \left(-z_{i1} + \frac{t_i z_{i1} u_i [1 - \hat{\alpha} e^{-t_i u_i} - (1 - e^{-t_i u_i})^{\hat{\alpha}}]}{(1 - e^{-t_i u_i}) (1 - (1 - e^{-t_i u_i})^{\hat{\alpha}})} \right)$$

$$\begin{aligned} \hat{i}_{23} = \hat{i}_{32} &= \frac{1}{n} \sum_{i=1}^n \delta_i \left(\frac{1}{\hat{\lambda}} - \frac{t_i e^{-\hat{\beta}^T z_i} [1 - \hat{\alpha} e^{-t_i u_i} - (1 - e^{-t_i u_i})^{\hat{\alpha}}]}{(1 - e^{-t_i u_i}) (1 - (1 - e^{-t_i u_i})^{\hat{\alpha}})} \right) \\ &\quad \times \left(-1 + \frac{t_i u_i [1 - \hat{\alpha} e^{-t_i u_i} - (1 - e^{-t_i u_i})^{\hat{\alpha}}]}{(1 - e^{-t_i u_i}) (1 - (1 - e^{-t_i u_i})^{\hat{\alpha}})} \right) \end{aligned}$$

$$\begin{aligned}\hat{i}_{24} &= \hat{i}_{42} = \frac{1}{n} \sum_{i=1}^n \delta_i \left(\frac{1}{\hat{\lambda}} - \frac{t_i e^{-\hat{\beta}^T z_i} [1 - \hat{\alpha} e^{-t_i u_i} - (1 - e^{-t_i u_i})^{\hat{\alpha}}]}{(1 - e^{-t_i u_i}) (1 - (1 - e^{-t_i u_i})^{\hat{\alpha}})} \right) \\ &\quad \times \left(-z_{i1} + \frac{t_i z_{i1} u_i [1 - \hat{\alpha} e^{-t_i u_i} - (1 - e^{-t_i u_i})^{\hat{\alpha}}]}{(1 - e^{-t_i u_i}) (1 - (1 - e^{-t_i u_i})^{\hat{\alpha}})} \right) \\ \hat{i}_{34} &= \hat{i}_{43} = \frac{1}{n} \sum_{i=1}^n \delta_i \left(-1 + \frac{t_i u_i [1 - \hat{\alpha} e^{-t_i u_i} - (1 - e^{-t_i u_i})^{\hat{\alpha}}]}{(1 - e^{-t_i u_i}) (1 - (1 - e^{-t_i u_i})^{\hat{\alpha}})} \right) \\ &\quad \times \left(-z_{i1} + \frac{t_i z_{i1} u_i [1 - \hat{\alpha} e^{-t_i u_i} - (1 - e^{-t_i u_i})^{\hat{\alpha}}]}{(1 - e^{-t_i u_i}) (1 - (1 - e^{-t_i u_i})^{\hat{\alpha}})} \right)\end{aligned}$$

Tous les composants de la statistique du test sont déterminés, elle sera alors déduite par la formule quadratique

$$Y_n^2(\hat{\gamma}) = \sum_{j=1}^k \frac{(U_j - e_j)^2}{U_j} + \hat{W}^T \left[\hat{i}_{ll'} - \sum_{j=1}^k \hat{C}_{lj} \hat{C}_{l'j} \hat{A}_j^{-1} \right]^{-1} \hat{W}$$

Sous H_0 , la distribution de la statistique Y_n^2 suit asymptotiquement une loi de chi-deux.

4.6 Simulation

4.6.1 Estimation des paramètres

En utilisant l'algorithme de Newton-Raphson et le logiciel maple, nous allons générer une série de $N = 10.000$ échantillons de taille $n_1 = 15$, $n_2 = 25$, $n_3 = 50$, $n_4 = 130$, $n_5 = 350$, $n_6 = 500$, de variables aléatoires censurées à droite issues d'un modèle $AFT - GEE$ de paramètres $\alpha = 3$, $\lambda = 2$, $\beta_0 = 0.5$, $\beta_1 = 0.8$, et ceci dans le but de montrer la convergence des estimateurs du maximum de vraisemblance. Les valeurs de la moyenne des estimateurs du maximum de vraisemblance $\hat{\alpha}, \hat{\lambda}, \hat{\beta}_0, \hat{\beta}_1$, et leurs erreurs quadratiques moyennes (SME) sont calculées et données dans le tableau suivant :

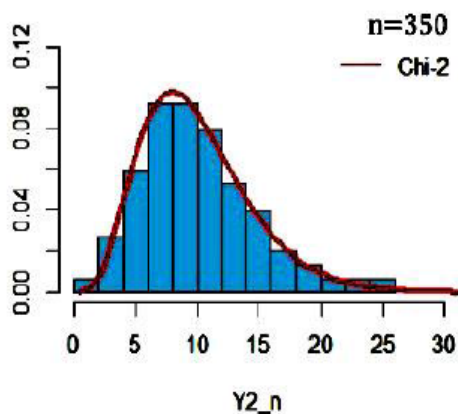
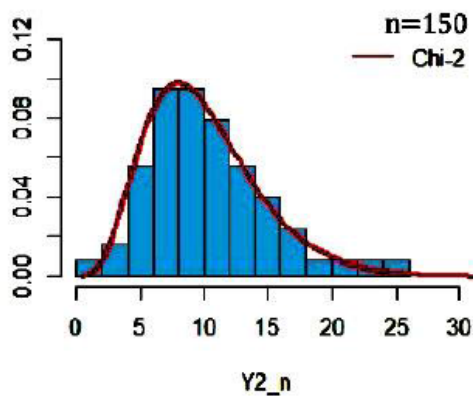
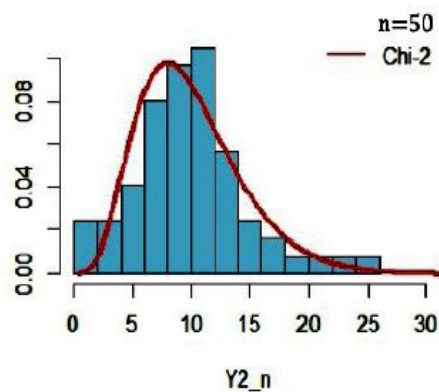
$N = 10,000$	$n_1 = 15$	$n_2 = 25$	$n_3 = 50$	$n_4 = 130$	$n_5 = 350$	$n_6 = 500$
$\hat{\alpha}$	2.8854	2.8985	2.9254	2.9523	2.9785	3.0026
$S.M.E$	0.0040	0.0032	0.0028	0.0021	0.0016	0.0004
$\hat{\lambda}$	1.8654	1.8925	1.9625	1.9745	1.9894	1.9998
$S.M.E$	0.0035	0.0029	0.0025	0.0019	0.0013	0.0008
$\hat{\beta}_0$	0.5721	0.5654	0.5425	0.5254	0.5102	0.5019
$S.M.E$	0.0045	0.0039	0.0026	0.0019	0.0012	0.0006
$\hat{\beta}_1$	0.8423	0.8325	0.8214	0.8154	0.8045	0.8012
$S.M.E$	0.0037	0.0025	0.0017	0.0012	0.0009	0.0005

Tab. 14. Les valeurs moyennes des EMV des paramètres et leurs erreurs quadratiques moyennes pour les données censurées

Les erreurs quadratiques moyennes sont très faibles (< 0.001). Les résultats de la simulation confirment le fait bien connu que pour les modèles réguliers, les estimateurs du maximum de vraisemblance sont convergents.

4.6.2 La distribution de la statistique $Y_n^2(\hat{\gamma})$

Pour montrer que la distribution de la statistique de test est invariante, nous avons effectué des simulations ($N = 10.000$) de la distribution de la statistique $Y^2(\hat{\gamma})$ sous l'hypothèse nulle H_0 , avec différentes valeurs de paramètres et différents nombres d'intervalles k de regroupement de données soumises à une censure droite. Ensuite, nous avons représenté les histogrammes des données des échantillons par rapport aux distributions du chi-deux correspondantes avec k degrés de liberté (voir les exemples dans les figures).



Tenant compte des erreurs des simulations, nous pouvons dire que la distribution limite de la statistique $Y^2(\hat{\gamma})$, est une distribution du chi-deux avec k degrés de liberté.

4.6.3 La statistique du test

Afin d'étudier la maniabilité et l'efficacité du test d'ajustement du type du chi-deux modifié pour le modèle AFT -exponentiel généralisé $AFT - GEE(\hat{\gamma})$ dans le cas où les données sont censurées, proposé dans ce travail, on a réalisé une importante étude par simulations numériques. Pour tester l'hypothèse nulle H_0 , nous générons 10,000 échantillons de données censurées de distribution $AFT - GEE(\hat{\gamma})$ dont les tailles respectives des échantillons sont $n_1 = 15$, $n_2 = 25$, $n_3 = 50$, $n_4 = 130$, $n_5 = 350$ et $n_6 = 500$. Les valeurs du critère de la statistique Y^2 sont calculées comme indiquées précédemment. Ensuite, nous calculons le nombre de cas de rejet de l'hypothèse nulle H_0 , c'est à dire lorsque $Y^2 > \chi_\varepsilon^2(k)$ (dans notre cas $\chi_\varepsilon^2(k)$ est le quantile de la distribution du chi-deux à k degrés de liberté), pour les niveaux de signification $\varepsilon = (0.10, 0.05, 0.01)$. Les valeurs des niveaux de signification empiriques sont comparés avec leurs valeurs théoriques correspondantes.

$N = 10,000$	$n_1 = 15$	$n_2 = 25$	$n_3 = 50$	$n_4 = 130$	$n_5 = 350$	$n_6 = 500$
$\varepsilon = 1\%$	0.0089	0.0045	0.0054	0.0096	0.0101	0.0102
$\varepsilon = 5\%$	0.0486	0.0488	0.0492	0.0495	0.0498	0.0504
$\varepsilon = 10\%$	0.0982	0.0984	0.0991	0.0998	0.01012	0.1004

Tab. 15. Comparaison des risques théoriques et des risques empiriques correspondants

$\varepsilon = 0,01 ; 0,05 ; 0,10$ pour les données censurées.

Au vu des résultats obtenus, on constate que les niveaux de signification empiriques sont très proches des niveaux de signification théoriques. On peut alors déduire que la statistique de test proposée permet de valider le modèle $AFT - GEE$ d'une manière satisfaisante.

4.7 Exemple de données de fiabilité

Nous considérons un ensemble de données de Cox et Oakes (1984), représentant les résultats de tests vie de ressorts, donné par Armstrong. Nous voulons vérifier si ces données peuvent être modélisées par une distribution $AFT - GEE$. Les ressorts sont testés dans des cycles de temps de chargement jusqu'à la rupture. Par commodité, on prend 10 cycles

4.7. Exemple de données de fiabilité

comme l'unité de «temps». Les 60 ressorts ont été attribués, 10 à chacun des six différents niveaux de stress.

<i>Stress</i>										
0.950	0.225	0.171	0.198	0.189	0.189	0.135	0.162	0.135	0.117	0.162
0.900	0.216	0.162	0.153	0.216	0.225	0.216	0.306	0.225	0.243	0.189
0.850	0.324	0.321	0.432	0.252	0.279	0.414	0.396	0.379	0.351	0.333
0.800	0.627	1.051	1.434	2.020	0.525	0.402	0.463	0.431	0.365	0.715
0.750	3.402	9.417	1.802	4.326	11.52*	7.152	2.969	3.012	1.550	11.211
0.700	12.51*	12.505*	3.027	12.505	6.253	8.011	7.795	11.604*	11.604*	12.47*

* Censuré

Table 16. Le temps de rupture des ressorts

En utilisant la méthode de Newton Raphson, nous calculons les estimateurs du maximum de vraisemblance des paramètres :

$$\hat{\theta} = \left(\hat{\alpha}, \hat{\lambda}, \hat{\beta}_0, \hat{\beta}_1 \right)^T = (3.0663, 0.0624, 11.6171, -18.1301)^T.$$

Les résultats nécessaires pour le calcul du critère de test Y_n^2 , avec $k = 6$ intervalles de regroupement de ces données sont donnés dans le tableau 17.

$\hat{a}_j$	0.190	0.296	0.445	1.356	6.124	12.510
U_j	11	10	12	5	9	7
e_j	9.1682	9.1682	9.1682	9.1682	9.1682	9.1682
$\hat{C}_{1j}$	-0.0169	-0.0285	-0.0752	-0.0375	-0.0362	-0.0103
$\hat{C}_{2j}$	0.9726	0.0707	0.9864	0.6638	0.9867	0.4858
$\hat{C}_{3j}$	-0.2478	-0.2540	-0.3735	-0.1662	-0.2487	-0.1551
$\hat{C}_{4j}$	-0.2362	-0.2373	-0.3418	-0.1495	-0.2121	-0.1226

Tableau 17. Valeurs de $\hat{a}_j, e_j, U_j, \hat{C}_{1j}, \hat{C}_{2j}, \hat{C}_{3j}, \hat{C}_{4j}$

Nous obtenons donc la valeur du critère de test Y_n^2

$$Y_n^2 = X^2 + Q = 5.1920 + 7.1284 = 12.3204$$

Pour le niveau de signification $\varepsilon = 0.05$, et comme la valeur de $Y_n^2 = 12.3204 < \chi_6^2 = 12.5915$, nous pouvons conclure que ces données sont distribuées selon un modèle *AFT* – *GEE*.

CHAPITRE 5

Modèle $AFT - LE$

La distribution exponentielle linéaire (LE) est une généralisation de la distribution exponentielle et de la distribution de Rayleigh. L'avantage de cette distribution est qu'elle peut décrire les phénomènes de fonction risque linéairement croissante ce qui n'est pas le cas des autres distributions utilisées dans l'analyse de fiabilité ou de l'analyse de survie. Ceci nous a motivé à introduire un modèle à durée de vie accélérée dont la distribution de base est une exponentielle linéaire $AFT - LE$. Ce modèle pourra être appliqué dans les tests de vie accélérée en intervenant sur les différents facteurs qui peuvent causer les défaillances. Avant de donner les résultats d'une analyse, on doit s'assurer que l'on a choisi le modèle adéquat, c'est pourquoi nous construisons une statistique de test maniable capable d'ajuster à ce modèle des données censurées à droite. Nous proposons une étude par simulations numériques et on termine par l'application de ce modèle à un jeu de données de fiabilité.

5.1 Construction du modèle $AFT - LE$

En suivant la même démarche que dans le chapitre précédent, nous allons introduire un modèle à durée de vie accélérée dont la distribution de base est une distribution exponentielle linéaire qu'on notera $AFT - LE$. Après la présentation du modèle, on estime les paramètres inconnus par la méthode du maximum de vraisemblance. Ensuite en utilisant l'approche

proposée par Bagdonavicius et Nikulin (2011), on fournit une statistique de test d'ajustement du type du chi-deux modifié pour ce modèle en présence de censure aléatoire droite.

Sous le modèle $AFT - LE$, la fonction de survie est donnée par la quantité suivante :

$$S(t; \gamma) = \exp \left(- \left[\alpha t e^{-\beta^T z} + \frac{\lambda}{2} t^2 e^{-2\beta^T z} \right] \right)$$

Où $\gamma = (\alpha, \lambda, \beta_0, \dots, \beta_m)^T$ est le vecteur des paramètres de la distribution exponentielle linéaire et ceux de la régression.

La fonction densité de probabilité est déduite par

$$f(t; \gamma) = (\alpha e^{-\beta^T z} + \lambda t e^{-2\beta^T z}) \exp \left(- \left[\alpha t e^{-\beta^T z} + \frac{\lambda}{2} t^2 e^{-2\beta^T z} \right] \right) \quad t \geq 0$$

Le taux de hasard et le taux de hasard cumulé sont respectivement :

$$h(t; \gamma) = \alpha e^{-\beta^T z} + \lambda t e^{-2\beta^T z}$$

et

$$H(t; \gamma) = \alpha t e^{-\beta^T z} + \frac{\lambda}{2} t^2 e^{-2\beta^T z}$$

5.2 Estimation du maximum de vraisemblance

On considère un échantillon de données censurées provenant d'un modèle $AFT - LE$. On suppose que l'échantillon est constitué de n observations t_i , $i = 1, 2, \dots, n$, tel que $t_i = \min(T_i, C_i)$ où T_i est la durée de vie et C_i est une censure aléatoire droite.

l'équation de la log-vraisemblance peut être écrite sous la formule suivante :

$$L(\gamma) = \sum_{i=1}^n \delta_i \ln \left(\alpha e^{-\beta^T z} + \lambda t_i e^{-2\beta^T z} \right) - \alpha \sum_{i=1}^n t_i e^{-\beta^T z} - \frac{\lambda}{2} \sum_{i=1}^n t_i^2 e^{-2\beta^T z}$$

Les dérivées par rapport à chacun des paramètres $\alpha, \lambda, \beta_0, \beta_1, \dots, \beta_m$, nous donne les fonctions de score

$$\frac{\partial L}{\partial \alpha} = \sum_{i=1}^n \frac{\delta_i e^{-\beta^T z}}{\alpha e^{-\beta^T z} + \lambda t_i e^{-2\beta^T z}} - \sum_{i=1}^n t_i e^{-\beta^T z}$$

$$\frac{\partial L}{\partial \lambda} = \sum_{i=1}^n \frac{\delta_i t_i e^{-2\beta^T z}}{\alpha e^{-\beta^T z} + \lambda t_i e^{-2\beta^T z}} - \frac{1}{2} \sum_{i=1}^n t_i^2 e^{-2\beta^T z}$$

$$\frac{\partial L}{\partial \beta_0} = - \sum_{i=1}^n \frac{\delta_i (\alpha e^{-\beta^T z} + 2\lambda t_i e^{-2\beta^T z})}{\alpha e^{-\beta^T z} + \lambda t_i e^{-2\beta^T z}} + \alpha \sum_{i=1}^n t_i e^{-\beta^T z} + \lambda \sum_{i=1}^n t_i^2 e^{-2\beta^T z}$$

$$\frac{\partial L}{\partial \beta_1} = - \sum_{i=1}^n \frac{\delta_i (\alpha z_{i1} e^{-\beta^T z} + 2z_{i1} \lambda t_i e^{-2\beta^T z})}{\alpha e^{-\beta^T z} + \lambda t_i e^{-2\beta^T z}} + \alpha \sum_{i=1}^n z_{i1} t_i e^{-\beta^T z} + \lambda \sum_{i=1}^n z_{i1} t_i^2 e^{-2\beta^T z}$$

.....

$$\frac{\partial L}{\partial \beta_m} = - \sum_{i=1}^n \frac{\delta_i (\alpha z_{im} e^{-\beta^T z} + 2z_{im} \lambda t_i e^{-2\beta^T z})}{\alpha e^{-\beta^T z} + \lambda t_i e^{-2\beta^T z}} + \alpha \sum_{i=1}^n z_{im} t_i e^{-\beta^T z} + \lambda \sum_{i=1}^n z_{im} t_i^2 e^{-2\beta^T z}$$

Les estimateurs du maximum de vraisemblance sont obtenus par des méthodes itératives comme la méthode de Newton Raphson ou par des logiciels de calculs numériques.

5.3 Test d'ajustement pour le modèle AFT-LE

Nous allons adapter la statistique développée précédemment pour tester l'hypothèse nulle H_0 selon laquelle des données proviennent d'un modèle à durée de vie accélérée dont la distribution de base est une exponentielle linéaire ($AFT - LE$).

Considérons l'hypothèse nulle :

$$H_0 : F(t) \in F_0 = \{F_0(t, \gamma), t \in R^1, \gamma \in \Theta \subset R^s\}$$

Où $\gamma = (\alpha, \gamma, \lambda, \beta_0, \beta_1)^T$. Comme auparavant pour les calculs, on ne considère que deux variables explicatives et donc deux coefficients, sans pour autant perdre de généralités. Prenons un intervalle de temps fini $[0, \tau]$ et divisons-le en $k > s$ intervalles $I_j = (a_{j-1}, a_j]$ où les observations sont groupées,

$$0 = a_0 < a_1 < \dots < a_{k-1} < a_k = +\infty$$

Pour que les fréquences théoriques soit égales dans les classes de groupement des données I_j , la valeur estimée $\hat{a}_j$ des limites des intervalles sera donnée, dans ce cas, par :

$$\hat{a}_j = -\frac{\hat{\alpha}}{\hat{\lambda}}e^{\beta^T z} + \frac{e^{\beta^T z}}{\hat{\lambda}} \sqrt{\alpha^2 + 2\lambda \left(\frac{E_j - \sum_{l=1}^{j-1} H(t_l, \hat{\gamma})}{n - j + 1} \right)}, \quad \hat{a}_k = t_{(n)}, j = 1, \dots, k$$

où $\hat{\gamma}$ est le vecteur des estimateurs du maximum de vraisemblance du vecteur des paramètres γ , et

$$E_j = \frac{j}{k-1} \sum_{i=1}^n \hat{\alpha} t_i e^{-\beta^T z} + \frac{\hat{\lambda}}{2} t_i^2 e^{-2\beta^T z}, \quad j = 1, \dots, k-1.$$

Par un tel choix, on a $e_j = \frac{E_k}{k}$ pour chaque intervalle I_j .

Pour le calcul du critère de test Y^2 permettant de valider un modèle $AFT - LE$, on a besoin de déterminer toutes les composantes correspondant à ce modèle et qui sont les matrices $\widehat{W}$ et $\widehat{I}$. On calcule d'abord les estimateurs du maximum de vraisemblance des paramètres inconnus du modèle sur les données initiales comme indiqué dans la section précédente.

Calcul de la matrice $\widehat{W}$

Les éléments de la matrice estimée $\widehat{W}$ sont obtenus à partir de la matrice $\hat{C}$:

$$\begin{aligned} \hat{C}_{1j} &= \frac{1}{n} \sum_{i:t_i \in I_j} \frac{e^{-\beta^T z}}{\alpha e^{-\beta^T z} + \lambda t_i e^{-2\beta^T z}} \\ \hat{C}_{2j} &= \frac{1}{n} \sum_{i:t_i \in I_j} \frac{t_i e^{-2\beta^T z}}{\alpha e^{-\beta^T z} + \lambda t_i e^{-2\beta^T z}} \\ \hat{C}_{3j} &= \frac{1}{n} \sum_{i:t_i \in I_j} \frac{\alpha e^{-\beta^T z} + 2\lambda t_i e^{-2\beta^T z}}{\alpha e^{-\beta^T z} + \lambda t_i e^{-2\beta^T z}} \\ \hat{C}_{4j} &= \frac{1}{n} \sum_{i:t_i \in I_j} \frac{\alpha z_{i1} e^{-\beta^T z} + 2z_{i1} \lambda t_i e^{-2\beta^T z}}{\alpha e^{-\beta^T z} + \lambda t_i e^{-2\beta^T z}} \end{aligned}$$

Calcul de la matrice $\hat{I}$

Vu la nécessité de la matrice d'information $\hat{I} = (\hat{i}_{ll'})_{4 \times 4}$, nous avons calculé explicitement tous ses éléments :

$$\hat{i}_{11} = \frac{1}{n} \sum_{i=1}^n \delta_i \left(\frac{e^{-\beta^T z}}{\alpha e^{-\beta^T z} + \lambda t_i e^{-2\beta^T z}} \right)^2$$

$$\hat{i}_{22} = \frac{1}{n} \sum_{i=1}^n \delta_i \left(\frac{t_i e^{-2\beta^T z}}{\alpha e^{-\beta^T z} + \lambda t_i e^{-2\beta^T z}} \right)^2$$

$$\hat{i}_{33} = \frac{1}{n} \sum_{i=1}^n \delta_i \left(\frac{\alpha e^{-\beta^T z} + 2\lambda t_i e^{-2\beta^T z}}{\alpha e^{-\beta^T z} + \lambda t_i e^{-2\beta^T z}} \right)^2$$

$$\hat{i}_{44} = \frac{1}{n} \sum_{i=1}^n \delta_i \left(\frac{\alpha z_{i1} e^{-\beta^T z} + 2z_{i1} \lambda t_i e^{-2\beta^T z}}{\alpha e^{-\beta^T z} + \lambda t_i e^{-2\beta^T z}} \right)^2$$

$$\hat{i}_{12} = \hat{i}_{21} = \frac{1}{n} \sum_{i=1}^n \delta_i \left(\frac{e^{-\beta^T z}}{\alpha e^{-\beta^T z} + \lambda t_i e^{-2\beta^T z}} \right) \left(\frac{t_i e^{-2\beta^T z}}{\alpha e^{-\beta^T z} + \lambda t_i e^{-2\beta^T z}} \right)$$

$$\hat{i}_{13} = \hat{i}_{31} = \frac{1}{n} \sum_{i=1}^n \delta_i \left(\frac{e^{-\beta^T z}}{\alpha e^{-\beta^T z} + \lambda t_i e^{-2\beta^T z}} \right) \left(\frac{\alpha e^{-\beta^T z} + 2\lambda t_i e^{-2\beta^T z}}{\alpha e^{-\beta^T z} + \lambda t_i e^{-2\beta^T z}} \right)$$

$$\hat{i}_{14} = \hat{i}_{41} = \frac{1}{n} \sum_{i=1}^n \delta_i \left(\frac{e^{-\beta^T z}}{\alpha e^{-\beta^T z} + \lambda t_i e^{-2\beta^T z}} \right) \left(\frac{\alpha z_{i1} e^{-\beta^T z} + 2z_{i1} \lambda t_i e^{-2\beta^T z}}{\alpha e^{-\beta^T z} + \lambda t_i e^{-2\beta^T z}} \right)$$

$$\hat{i}_{23} = \hat{i}_{32} = \frac{1}{n} \sum_{i=1}^n \delta_i \left(\frac{t_i e^{-2\beta^T z}}{\alpha e^{-\beta^T z} + \lambda t_i e^{-2\beta^T z}} \right) \left(\frac{\alpha e^{-\beta^T z} + 2\lambda t_i e^{-2\beta^T z}}{\alpha e^{-\beta^T z} + \lambda t_i e^{-2\beta^T z}} \right)$$

$$\hat{i}_{24} = \hat{i}_{42} = \frac{1}{n} \sum_{i=1}^n \delta_i \left(\frac{t_i e^{-2\beta^T z}}{\alpha e^{-\beta^T z} + \lambda t_i e^{-2\beta^T z}} \right) \left(\frac{\alpha z_{i1} e^{-\beta^T z} + 2z_{i1} \lambda t_i e^{-2\beta^T z}}{\alpha e^{-\beta^T z} + \lambda t_i e^{-2\beta^T z}} \right)$$

$$\hat{i}_{34} = \hat{i}_{43} = \frac{1}{n} \sum_{i=1}^n \delta_i \left(\frac{\alpha e^{-\beta^T z} + 2\lambda t_i e^{-2\beta^T z}}{\alpha e^{-\beta^T z} + \lambda t_i e^{-2\beta^T z}} \right) \left(\frac{\alpha z_{i1} e^{-\beta^T z} + 2z_{i1} \lambda t_i e^{-2\beta^T z}}{\alpha e^{-\beta^T z} + \lambda t_i e^{-2\beta^T z}} \right)$$

Tous les éléments composant la statistique de test Y_n^2 sont maintenant calculés, on déduit alors sa forme explicite :

$$Y_n^2(\hat{\gamma}) = \sum_{j=1}^k \frac{(U_j - e_j)^2}{U_j} + \hat{W}^T \left[\hat{i}_{ll'} - \sum_{j=1}^k \hat{C}_{lj} \hat{C}_{l'j} \hat{A}_j^{-1} \right]^{-1} \hat{W}$$

Cette statistique est distribuée asymptotiquement selon une loi du chi-deux à k degrés de liberté.

5.4 Simulations

Dans cette section, on propose une étude par simulations numériques pour déterminer les estimateurs du maximum de vraisemblance des paramètres du modèle et des valeurs du critère de test d'adéquation développé pour le modèle $AFT - LE$.

5.4.1 Estimation des paramètres

Etant donné que les formules analytiques des estimateurs des paramètres du modèle $AFT - LE$ ne peuvent pas être déterminés, nous utilisons alors les méthodes itératives pour les calculer. Pour cela, sous l'hypothèse nulle H_0 , on simule $N = 10.000$ fois des échantillons de tailles respectives $n_1 = 15$, $n_2 = 25$, $n_3 = 50$, $n_4 = 130$, $n_5 = 350$, $n_6 = 500$, avec les valeurs suivantes des paramètres $\alpha = 0.7$, $\lambda = 2.5$, $\beta_0 = 0.5$, $\beta_1 = 0.98$. En utilisant l'algorithme BB solve et le logiciel R, on calcule les valeurs moyennes des estimateurs du maximum de vraisemblance $\hat{\alpha}$, $\hat{\lambda}$, $\hat{\beta}_0$, $\hat{\beta}_1$, et leurs erreurs quadratiques moyennes (SME) qu'on regroupe dans le tableau suivant :

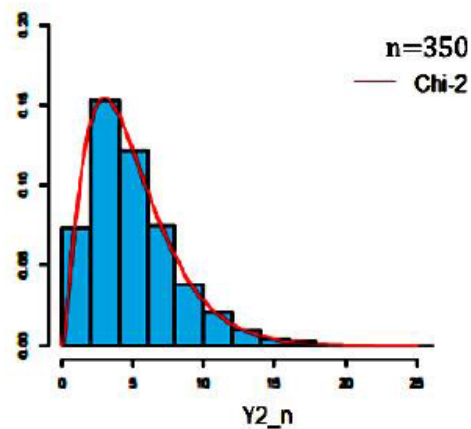
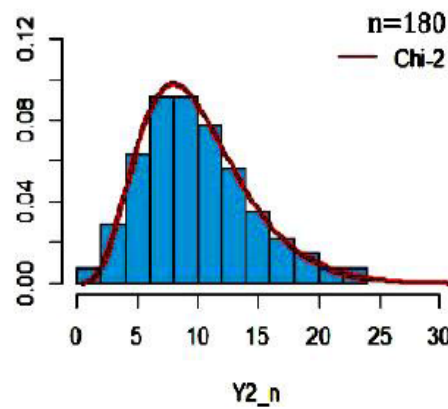
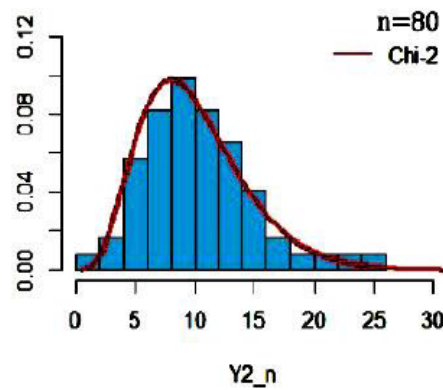
$N = 10,000$	$n_1 = 15$	$n_2 = 25$	$n_3 = 50$	$n_4 = 130$	$n_5 = 350$	$n_6 = 500$
$\hat{\alpha}$	0.5854	0.5985	0.6254	0.6523	0.6785	0.7020
$S.M.E$	0.0040	0.0032	0.0028	0.0021	0.0016	0.0004
$\hat{\lambda}$	2.2654	2.2925	2.3625	2.3745	2.4894	2.4998
$S.M.E$	0.0035	0.0029	0.0025	0.0019	0.0013	0.0008
$\hat{\beta}_0$	0.5721	0.5654	0.5425	0.5254	0.5102	0.5019
$S.M.E$	0.0045	0.0039	0.0026	0.0019	0.0012	0.0006
$\hat{\beta}_1$	0.9423	0.9525	0.9614	0.9654	0.9745	0.9802
$S.M.E$	0.0037	0.0025	0.0017	0.0012	0.0009	0.0005

Tab. 18. Les valeurs moyennes des EMV des paramètres et leurs erreurs quadratiques moyennes du modèle $AFT - LE$.

Nous pouvons remarquer que les erreurs quadratiques moyennes des estimateurs obtenus sont toutes très faibles même pour les petits échantillons. Les résultats sont très satisfaisants.

5.4.2 La distribution de la statistique $Y_n^2(\hat{\gamma})$

Pour étudier la distribution de la statistique $Y_n^2(\hat{\gamma})$ proposée pour tester la validité d'un modèle $AFT - LE$, nous avons représenté les histogrammes des données des échantillons (de différentes tailles) par rapport aux distributions du chi-deux correspondantes avec k degrés de liberté (voir les exemples dans les figures).



Quelque soit la taille de l'échantillon, la statistique Y_n^2 du test de validation du modèle $AFT - LE$ suit à la limite la distribution du chi-deux avec k degrés de liberté.

5.4.3 Statistique du test

Pour vérifier les résultats théoriques obtenus pour tester l'adéquation d'une série d'observations au modèle $AFT - LE$, on compare les risques théoriques ($\varepsilon = 0.01, 0.05, 0.1$) et les risques empiriques des critères de test calculés plus haut, pour 10,000 échantillons simulés de la distribution $AFT - LE(\hat{\gamma})$ dont les tailles respectives sont $n_1 = 15, n_2 = 25, n_3 = 50, n_4 = 130, n_5 = 350$ et $n_6 = 500$.

$N = 10,000$	$n_1 = 15$	$n_2 = 25$	$n_3 = 50$	$n_4 = 130$	$n_5 = 350$	$n_6 = 500$
$\varepsilon = 1\%$	0.0080	0.0084	0.0089	0.0095	0.0098	0.0102
$\varepsilon = 5\%$	0.0475	0.0480	0.0482	0.0495	0.0508	0.0505
$\varepsilon = 10\%$	0.0946	0.0965	0.0979	0.0983	0.0996	0.1003

Tab.19. Comparaison des risques empiriques et des risques théoriques correspondants ($\varepsilon = 0.01, 0.05, 0.1$) pour les données censurées

Nous pouvons dire que la statistique Y_n^2 est bien adaptée pour ajuster des données censurées au modèle $AFT - LE$.

5.5 Exemple de données de fiabilité

On considère les données de Nelson (1990) concernant les temps de défaillance (T_i) de 76 fluides isolants électriques testés à une tension comprise entre 26 et 38 kilovolts (kv). Cette expérience a été exécutée assez longtemps pour observer les taux de défaillance de tous les éléments. Nous supposons l'hypothèse nulle H_0 que la répartition du temps de rupture du fluide est un modèle à temps de vie accéléré dont la distribution de base est une exponentielle linéaire ($AFT - LE$) dépendant de la variable explicative Z_{i1} qui représente les niveaux de tension appliqués v_i .

5.5. Exemple de données de fiabilité

Les données sont reportées dans le tableau suivant :

$v_i(kv)$	Failure times T_i
26	5.79, 1579.52, 2323.7
28	68.85, 426.07, 110.29, 108.29, 1067.6
30	17.05, 22.66, 21.02, 175.88, 139.07, 144.12, 20.46, 43.40, 194.90, 47.30, 7.74
32	0.4, 82.85, 9.88, 89.29, 215.10, 2.75, 0.79, 15.93, 3.91, 0.27, 0.69, 100.58, 27.80, 13.95, 53.2
34	0.96, 4.15, 0.19, 0.78, 8.01, 31.75, 7.35, 6.50, 8.27, 33.91, 32.52, 3.16, 4.85, 2.78, 4.67, 1.31, 12.06, 36.71, 72.89
36	1.97, 0.59, 2.58, 1.69, 2.71, 25.50, 0.35, 0.99, 3.99, 3.67, 2.07, 0.96, 5.35, 2.90, 13.77
38	0.47, 0.73, 1.40, 0.74, 0.39, 1.13, 0.09, 2.38

Données de Nelson

En utilisant le logiciel statistique R, nous trouvons les valeurs des estimateurs MV :

$$\hat{\alpha} = 2.6589, \hat{\lambda} = 4.1856, \hat{\beta}_0 = -12.6788, \hat{\beta}_1 = 10.8954$$

Si nous choisissons $k = 6$ intervalles pour grouper ces données, les résultats des calculs préliminaires obtenus sont exposés dans le tableau suivant :

$\hat{a}_j$	2.4587	9.4521	25.3621	42.1576	127.456	2323.7
U_j	23	19	9	6	10	9
e_j	8.3695	8.3695	8.3695	8.3695	8.3695	8.3695
$\hat{C}_{1j}$	0.8454	0.2154	0.4157	0.6784	0.3178	0.1589
$\hat{C}_{2j}$	-2.6584	-1.3265	-4.1569	-5.1291	-4.3651	-2.3154
$\hat{C}_{3j}$	0.5124	0.6214	0.3124	0.7415	0.2365	0.1245
$\hat{C}_{4j}$	-3.1547	-7.6214	-4.6252	0.1456	0.8952	0.7451

Table 20. valeurs des $\hat{a}_j, e_j, U_j, \hat{C}_{1j}, \hat{C}_{2j}, \hat{C}_{3j}, \hat{C}_{4j}$

Nous obtenons donc la valeur du test Y_n^2

$$Y_n^2 = X^2 + Q = 5.6258 + 4.1592 = 9.785$$

Pour le niveau de signification $\varepsilon = 0.05$, la valeur de $Y_n^2 = 9.785 < \chi_6^2 = 12.3204$, nous pouvons donc dire que le modèle proposé *AFT – LE* décrit convenablement ces données..

CHAPITRE 6

Modèle $AFT - KwGEE$

La distribution Kumaraswamy généralisée exponentielle exponentiée, notée $Kw - GEE$, introduite par Mohammed (2014) et présentée dans le premier chapitre est une distribution très flexible. Ses multiples paramètres de forme font que cette distribution peut avoir différentes formes de queues et des asymétries plus ou moins prononcées. En raison de ses multiples champs d'applications, la distribution $Kw - GEE$ pourrait être une distribution de base pour un modèle à durée de vie accélérée très intéressant. Dans ce chapitre, nous construisons un modèle à temps de vie accéléré à partir de cette distribution qu'on appellera $AFT - KwGEE$. Après l'étude des estimateurs du maximum de vraisemblance, on propose la construction d'un test d'ajustement du type du chi-deux modifié pour ce modèle dans le cas de censure aléatoire droite et de paramètres inconnus et estimés par la méthode du maximum de vraisemblance sur données non regroupées. On confirme nos résultats par des simulations et une application à des données réelles.

6.1 Construction du modèle $AFT - KwGEE$

Pour un modèle AFT dont la distribution de base est la distribution Kumaraswamy généralisée exponentielle exponentiée, la fonction de survie est donnée par :

$$\begin{aligned} S(t, \gamma) &= S_0 \left(te^{-\beta^T z} \right) \\ &= \left\{ 1 - \left(1 - \exp \left\{ -\lambda te^{-\beta^T z} \right\} \right)^{\alpha\sigma} \right\}^\theta \end{aligned}$$

La fonction densité de probabilité correspondante sera donnée alors par la formule suivante :

$$f(t, \gamma) = \alpha\lambda\sigma\theta e^{-\beta^T z} \exp \left\{ -\lambda te^{-\beta^T z} \right\} \left(1 - \exp \left\{ -\lambda te^{-\beta^T z} \right\} \right)^{\alpha\sigma-1} \left\{ 1 - \left(1 - \exp \left\{ -\lambda te^{-\beta^T z} \right\} \right)^{\alpha\sigma} \right\}^{\theta-1}$$

Le taux de hasard et le taux de hasard cumulé sont obtenus par :

$$h(t, \gamma) = \frac{\alpha\lambda\sigma\theta e^{-\beta^T z} \exp \left\{ -\lambda te^{-\beta^T z} \right\} \left(1 - \exp \left\{ -\lambda te^{-\beta^T z} \right\} \right)^{\alpha\sigma-1}}{1 - \left(1 - \exp \left\{ -\lambda te^{-\beta^T z} \right\} \right)^{\alpha\sigma}}$$

et

$$H(t, \gamma) = -\theta \ln \left(1 - \left(1 - \exp \left\{ -\lambda te^{-\beta^T z} \right\} \right)^{\alpha\sigma} \right)$$

6.2 Estimation du maximum de vraisemblance

Supposons que l'on a un n -échantillon $(T_i)_{1,n}$ de variables aléatoires censurées à droite de la distribution $AFT - KwGEE$, les observations consistent en $t_i = \min(T_i, C_i)$ pour $i = 1, \dots, n$ où T_i représentent les taux de pannes et C_i les taux de censure. Les distributions de T_i et de C_i sont supposées indépendantes, la fonction de la log-vraisemblance du modèle s'écrit alors :

$$\begin{aligned} L(\gamma) &= \sum_{i=1}^n \delta_i [\ln \alpha\sigma\theta\lambda - \beta^T z - \lambda e^{-\beta^T z} t_i + (\alpha\sigma - 1) \ln \left(1 - \exp \left\{ -\lambda t_i e^{-\beta^T z} \right\} \right) \\ &\quad - \ln \left(1 - \left(1 - \exp \left\{ -\lambda t_i e^{-\beta^T z} \right\} \right)^{\alpha\sigma} \right)] + \theta \sum_{i=1}^n \ln \left(1 - \left(1 - \exp \left\{ -\lambda t_i e^{-\beta^T z} \right\} \right)^{\alpha\sigma} \right) \end{aligned}$$

Les fonctions de score sont données par :

$$\frac{\partial L}{\partial \alpha} = \sum_{i=1}^n \delta_i \left[\frac{1}{\alpha} + \frac{\sigma \ln(1 - \exp\{-\lambda t_i e^{-\beta^T z}\})}{1 - (1 - \exp\{-\lambda t_i e^{-\beta^T z}\})^{\alpha\sigma}} \right] - \sigma \theta \sum_{i=1}^n \frac{(1 - \exp\{-\lambda t_i e^{-\beta^T z}\})^{\alpha\sigma} \ln(1 - \exp\{-\lambda t_i e^{-\beta^T z}\})}{1 - (1 - \exp\{-\lambda t_i e^{-\beta^T z}\})^{\alpha\sigma}}$$

$$\frac{\partial L}{\partial \lambda} = \sum_{i=1}^n \delta_i \left[\frac{1}{\lambda} - \frac{t_i e^{-\beta^T z} (1 - \alpha\sigma \exp\{-\lambda t_i e^{-\beta^T z}\} - (1 - \exp\{-\lambda t_i e^{-\beta^T z}\})^{\alpha\sigma})}{(1 - \exp\{-\lambda t_i e^{-\beta^T z}\}) (1 - (1 - \exp\{-\lambda t_i e^{-\beta^T z}\})^{\alpha\sigma})} \right] - \alpha\sigma \theta \sum_{i=1}^n \frac{t_i e^{-\beta^T z} \exp\{-\lambda t_i e^{-\beta^T z}\} (1 - \exp\{-\lambda t_i e^{-\beta^T z}\})^{\alpha\sigma-1}}{1 - (1 - \exp\{-\lambda t_i e^{-\beta^T z}\})^{\alpha\sigma}}$$

$$\frac{\partial L}{\partial \sigma} = \sum_{i=1}^n \delta_i \left[\frac{1}{\sigma} + \frac{\alpha \ln(1 - \exp\{-\lambda t_i e^{-\beta^T z}\})}{1 - (1 - \exp\{-\lambda t_i e^{-\beta^T z}\})^{\alpha\sigma}} \right] - \alpha\theta \sum_{i=1}^n \frac{(1 - \exp\{-\lambda t_i e^{-\beta^T z}\})^{\alpha\sigma} \ln(1 - \exp\{-\lambda t_i e^{-\beta^T z}\})}{1 - (1 - \exp\{-\lambda t_i e^{-\beta^T z}\})^{\alpha\sigma}}$$

$$\frac{\partial L}{\partial \theta} = \sum_{i=1}^n \frac{\delta_i}{\theta} + \sum_{i=1}^n \ln(1 - (1 - \exp\{-\lambda t_i e^{-\beta^T z}\})^{\alpha\sigma})$$

$$\frac{\partial L}{\partial \beta_0} = \sum_{i=1}^n \delta_i \left[-1 - \frac{\lambda t_i e^{-\beta^T z} (1 - \alpha\sigma \exp\{-\lambda t_i e^{-\beta^T z}\} - (1 - \exp\{-\lambda t_i e^{-\beta^T z}\})^{\alpha\sigma})}{(1 - \exp\{-\lambda t_i e^{-\beta^T z}\}) (1 - (1 - \exp\{-\lambda t_i e^{-\beta^T z}\})^{\alpha\sigma})} \right] - \alpha\sigma \lambda \theta \sum_{i=1}^n \frac{t_i e^{-\beta^T z} \exp\{-\lambda t_i e^{-\beta^T z}\} (1 - \exp\{-\lambda t_i e^{-\beta^T z}\})^{\alpha\sigma-1}}{1 - (1 - \exp\{-\lambda t_i e^{-\beta^T z}\})^{\alpha\sigma}}$$

$$\frac{\partial L}{\partial \beta_1} = \sum_{i=1}^n \delta_i \left[-z_{i1} - \frac{z_{i1} \lambda t_i e^{-\beta^T z} (1 - \alpha\sigma \exp\{-\lambda t_i e^{-\beta^T z}\} - (1 - \exp\{-\lambda t_i e^{-\beta^T z}\})^{\alpha\sigma})}{(1 - \exp\{-\lambda t_i e^{-\beta^T z}\}) (1 - (1 - \exp\{-\lambda t_i e^{-\beta^T z}\})^{\alpha\sigma})} \right] - \alpha\sigma \lambda \theta \sum_{i=1}^n \frac{z_{i1} t_i e^{-\beta^T z} \exp\{-\lambda t_i e^{-\beta^T z}\} (1 - \exp\{-\lambda t_i e^{-\beta^T z}\})^{\alpha\sigma-1}}{1 - (1 - \exp\{-\lambda t_i e^{-\beta^T z}\})^{\alpha\sigma}}$$

$$\begin{aligned} \frac{\partial L}{\partial \beta_m} = & \sum_{i=1}^n \delta_i \left[-z_{im} - \frac{z_{im} \lambda t_i e^{-\beta^T z} \left(1 - \alpha \sigma \exp \left\{ -\lambda t_i e^{-\beta^T z} \right\} - \left(1 - \exp \left\{ -\lambda t_i e^{-\beta^T z} \right\} \right)^{\alpha \sigma} \right)}{(1 - \exp \left\{ -\lambda t_i e^{-\beta^T z} \right\}) (1 - (1 - \exp \left\{ -\lambda t_i e^{-\beta^T z} \right\})^{\alpha \sigma})} \right] \\ & - \alpha \sigma \lambda \theta \sum_{i=1}^n \frac{z_{im} t_i e^{-\beta^T z} \exp \left\{ -\lambda t_i e^{-\beta^T z} \right\} \left(1 - \exp \left\{ -\lambda t_i e^{-\beta^T z} \right\} \right)^{\alpha \sigma - 1}}{1 - (1 - \exp \left\{ -\lambda t_i e^{-\beta^T z} \right\})^{\alpha \sigma}} \end{aligned}$$

En égalant ces fonctions à zéro, on obtient un système d'équations dont les solutions sont les estimateurs du maximum de vraisemblance des paramètres inconnus. Cependant on n'arrive pas à dégager des formes analytiques de ces estimateurs. On peut utiliser alors n'importe quelle méthode itérative.

6.3 Test d'ajustement pour le modèle AFT-KwGEE

Pour valider le modèle *AFT - KwGEE*, on adapte la statistique de test du type du chi-deux modifié décrit dans le chapitre deux. On procède de la même façon que dans les chapitres précédents pour les calculs.

Supposons l'hypothèse nulle H_0 que le taux de défaillance T_i suit un modèle à durée de vie accélérée, dont la distribution de base est une exponentielle exponentielle généralisée *AFT - KwGEE*.

Après avoir calculé les estimateurs du maximum de vraisemblance des paramètres $\alpha, \lambda, \sigma, \theta, \beta_0, \beta_1, \dots, \beta_m$ sur un échantillon aléatoire censuré à droite, on regroupe les observations dans k classes I_j limitées par a_j . Ces limites sont fonctions de variables aléatoires telles que les classes choisies soient de même fréquence théorique e_j .

La valeur estimée $\hat{a}_j$ pour la distribution *EKw - GEE* est obtenue comme suit :

$$\hat{a}_j = -\frac{e^{\beta^T z}}{\hat{\lambda}} \ln \left\{ 1 - \left[1 - \exp \left(\frac{\sum_{l=1}^{i-1} H(t_l, \hat{\gamma}) - E_j}{\theta (n - i + 1)} \right) \right]^{1/\hat{\alpha}\hat{\sigma}} \right\}$$

pour $j = 1, \dots, k - 1$, $\hat{a}_k = t_{(n)}$, avec

$$E_j = -\frac{j}{k-1} \sum_{i=1}^n \hat{\theta} \ln \left(1 - \left(1 - \exp \left\{ -\hat{\lambda} t_i e^{-\beta^T z} \right\} \right)^{\hat{\alpha}\hat{\sigma}} \right),$$

Avec ce choix, nous calculons les effectifs théoriques e_j qui seront égaux dans toutes les classes I_j de groupement des données.

Calcul de la matrice $\hat{W}$ et $\hat{C}$

pour calculer la forme quadratique Q du test, on doit d'abord déterminer les éléments qui composent les matrices $\hat{W}$ et $\hat{C}$.

Après de longs calculs, on trouve :

$$\hat{C}_{1j} = \frac{1}{n} \sum_{i:t_i \in I_j} \delta_i \left[\frac{1}{\hat{\alpha}} + \frac{\hat{\sigma} \ln(1 - \exp \{-\hat{\lambda} t e^{-\hat{\beta}^T z}\})}{1 - (1 - \exp \{-\hat{\lambda} t e^{-\hat{\beta}^T z}\})^{\hat{\alpha} \hat{\sigma}}} \right]$$

$$\hat{C}_{2j} = \frac{1}{n} \sum_{i:t_i \in I_j} \delta_i \left[\frac{1}{\hat{\lambda}} - \frac{t e^{-\hat{\beta}^T z} \left(1 - \hat{\alpha} \hat{\sigma} \exp \{-\hat{\lambda} t e^{-\hat{\beta}^T z}\} - (1 - \exp \{-\hat{\lambda} t e^{-\hat{\beta}^T z}\})^{\hat{\alpha} \hat{\sigma}} \right)}{(1 - \exp \{-\hat{\lambda} t e^{-\hat{\beta}^T z}\}) \left(1 - (1 - \exp \{-\hat{\lambda} t e^{-\hat{\beta}^T z}\})^{\hat{\alpha} \hat{\sigma}} \right)} \right]$$

$$\hat{C}_{3j} = \frac{1}{n} \sum_{i:t_i \in I_j} \delta_i \left[\frac{1}{\hat{\sigma}} + \frac{\hat{\alpha} \ln(1 - \exp \{-\hat{\lambda} t e^{-\hat{\beta}^T z}\})}{1 - (1 - \exp \{-\hat{\lambda} t e^{-\hat{\beta}^T z}\})^{\hat{\alpha} \hat{\sigma}}} \right]$$

$$\hat{C}_{4j} = \frac{1}{n} \sum_{i:x_i \in I_j} \frac{\delta_i}{\hat{\theta}}$$

$$\hat{C}_{5j} = \frac{1}{n} \sum_{i:t_i \in I_j} \delta_i \left[-1 + \frac{\lambda t e^{-\hat{\beta}^T z} \left(1 - \hat{\alpha} \hat{\sigma} \exp \{-\hat{\lambda} t e^{-\hat{\beta}^T z}\} - (1 - \exp \{-\hat{\lambda} t e^{-\hat{\beta}^T z}\})^{\hat{\alpha} \hat{\sigma}} \right)}{(1 - \exp \{-\hat{\lambda} t e^{-\hat{\beta}^T z}\}) \left(1 - (1 - \exp \{-\hat{\lambda} t e^{-\hat{\beta}^T z}\})^{\hat{\alpha} \hat{\sigma}} \right)} \right]$$

$$\hat{C}_{6j} = \frac{1}{n} \sum_{i:t_i \in I_j} \delta_i \left[-z_{i1} + \frac{z_{i1} \lambda t e^{-\hat{\beta}^T z} \left(1 - \hat{\alpha} \hat{\sigma} \exp \{-\hat{\lambda} t e^{-\hat{\beta}^T z}\} - (1 - \exp \{-\hat{\lambda} t e^{-\hat{\beta}^T z}\})^{\hat{\alpha} \hat{\sigma}} \right)}{(1 - \exp \{-\hat{\lambda} t e^{-\hat{\beta}^T z}\}) \left(1 - (1 - \exp \{-\hat{\lambda} t e^{-\hat{\beta}^T z}\})^{\hat{\alpha} \hat{\sigma}} \right)} \right]$$

On déduit la matrice estimée $\hat{W}$ d'éléments :

$$\hat{W}_l = \sum_{j=1}^k \hat{C}_{lj} A_j^{-1} Z_j, \quad l = 1, 2, 3, 4, 5, 6 \quad j = 1, \dots, k$$

Calcul de la matrice d'information estimée $\hat{I}$:

La matrice symétrique d'information I est nécessaire pour le calcul de la matrice G

$$G = \hat{u}_{ll'} - \sum_{j=1}^r \hat{C}_{lj} \hat{C}_{l'j} \hat{A}_j^{-1}$$

Après un calcul laborieux, on a obtenu tous les éléments de la matrice $\hat{I}$.

$$\hat{i}_{11} = \frac{1}{n} \sum_{i=1}^n \delta_i \left(\frac{1}{\hat{\alpha}} + \hat{\sigma} A(t_i, \gamma) \right)^2$$

$$\hat{i}_{22} = \frac{1}{n} \sum_{i=1}^n \delta_i \left(\frac{1}{\hat{\lambda}} - M(t_i, \gamma) \right)^2$$

$$\hat{i}_{33} = \frac{1}{n} \sum_{i=1}^n \delta_i \left(\frac{1}{\hat{\sigma}} + \hat{\alpha} A(t_i, \gamma) \right)^2$$

$$\hat{i}_{44} = \frac{1}{n} \sum_{i=1}^n \frac{\delta_i}{\hat{\theta}^2}$$

$$\hat{i}_{55} = \frac{1}{n} \sum_{i=1}^n \delta_i (-1 + \lambda M(t_i, \gamma))^2$$

$$\hat{i}_{66} = \frac{1}{n} \sum_{i=1}^n \delta_i (-z_{i1} + z_{i1} \lambda M(t_i, \gamma))^2$$

$$\hat{i}_{12} = \frac{1}{n} \sum_{i=1}^n \delta_i \left(\frac{1}{\hat{\alpha}} + \hat{\sigma} A(t_i, \gamma) \right) \left(\frac{1}{\hat{\lambda}} - M(t_i, \gamma) \right)$$

$$\hat{i}_{13} = \frac{1}{n} \sum_{i=1}^n \delta_i \left(\frac{1}{\hat{\alpha}} + \hat{\sigma} A(t_i, \gamma) \right) \left(\frac{1}{\hat{\sigma}} + \hat{\alpha} A(t_i, \gamma) \right)$$

$$\hat{i}_{14} = \frac{1}{n} \sum_{i=1}^n \frac{\delta_i}{\hat{\theta}} \left(\frac{1}{\hat{\alpha}} + \hat{\sigma} A(t_i, \gamma) \right)$$

$$\hat{i}_{15} = \frac{1}{n} \sum_{i=1}^n \delta_i \left(\frac{1}{\hat{\alpha}} + \hat{\sigma} A(t_i, \gamma) \right) (-1 + \lambda M(t_i, \gamma))$$

$$\hat{i}_{16} = \frac{1}{n} \sum_{i=1}^n \delta_i \left(\frac{1}{\hat{\alpha}} + \hat{\sigma} A(t_i, \gamma) \right) (-z_{i1} + z_{i1} \lambda M(t_i, \gamma))$$

$$\hat{i}_{23} = \frac{1}{n} \sum_{i=1}^n \delta_i \left(\frac{1}{\hat{\lambda}} - M(t_i, \gamma) \right) \left(\frac{1}{\hat{\sigma}} + \hat{\alpha} A(t_i, \gamma) \right)$$

$$\hat{i}_{24} = \frac{1}{n} \sum_{i=1}^n \frac{\delta_i}{\hat{\theta}} \left(\frac{1}{\hat{\lambda}} - M(t_i, \gamma) \right)$$

$$\begin{aligned}
 \hat{i}_{25} &= \frac{1}{n} \sum_{i=1}^n \delta_i \left(\frac{1}{\hat{\lambda}} - M(t_i, \gamma) \right) (-1 + \lambda M(t_i, \gamma)) \\
 \hat{i}_{26} &= \frac{1}{n} \sum_{i=1}^n \delta_i \left(\frac{1}{\hat{\lambda}} - M(t_i, \gamma) \right) (-z_{i1} + z_{i1} \lambda M(t_i, \gamma)) \\
 \hat{i}_{34} &= \frac{1}{n} \sum_{i=1}^n \frac{\delta_i}{\hat{\theta}} \left(\frac{1}{\hat{\sigma}} + \hat{\alpha} A(t_i, \gamma) \right) \\
 \hat{i}_{35} &= \frac{1}{n} \sum_{i=1}^n \delta_i \left(\frac{1}{\hat{\sigma}} + \hat{\alpha} A(t_i, \gamma) \right) (-1 + \lambda M(t_i, \gamma)) \\
 \hat{i}_{36} &= \frac{1}{n} \sum_{i=1}^n \delta_i \left(\frac{1}{\hat{\sigma}} + \hat{\alpha} A(t_i, \gamma) \right) (-z_{i1} + z_{i1} \lambda M(t_i, \gamma)) \\
 \hat{i}_{45} &= \frac{1}{n} \sum_{i=1}^n \frac{\delta_i}{\hat{\theta}} (-1 + \lambda M(t_i, \gamma)) \\
 \hat{i}_{46} &= \frac{1}{n} \sum_{i=1}^n \frac{\delta_i}{\hat{\theta}} (-z_{i1} + z_{i1} \lambda M(t_i, \gamma)) \\
 \hat{i}_{56} &= \frac{1}{n} \sum_{i=1}^n \delta_i (-1 + \lambda M(t_i, \gamma)) (-z_{i1} + z_{i1} \lambda M(t_i, \gamma))
 \end{aligned}$$

où $A(t_i, \gamma)$ et $M(t_i, \gamma)$ représentent les quantités suivantes :

$$\begin{aligned}
 A(t_i, \gamma) &= \frac{\ln(1 - \exp\{-\hat{\lambda} t e^{-\hat{\beta}^T z}\})}{1 - (1 - \exp\{-\hat{\lambda} t e^{-\hat{\beta}^T z}\})^{\hat{\alpha} \hat{\sigma}}} \\
 M(t, \gamma) &= \frac{t e^{-\hat{\beta}^T z} \left(1 - \hat{\alpha} \hat{\sigma} \exp\{-\hat{\lambda} t e^{-\hat{\beta}^T z}\} - (1 - \exp\{-\hat{\lambda} t e^{-\hat{\beta}^T z}\})^{\hat{\alpha} \hat{\sigma}} \right)}{(1 - \exp\{-\hat{\lambda} t e^{-\hat{\beta}^T z}\}) \left(1 - (1 - \exp\{-\hat{\lambda} t e^{-\hat{\beta}^T z}\})^{\hat{\alpha} \hat{\sigma}} \right)}
 \end{aligned}$$

Etant donné que les valeurs des effectifs théoriques e_j sont déjà calculées, les matrices $\hat{W}$, $\hat{C}$, et $\hat{I}$ sont complètement déterminées, alors on déduit la statistique Y_n^2 d'un test d'ajustement pour le modèle AFT - KwGEE :

$$Y_n^2(\hat{\gamma}) = \sum_{j=1}^k \frac{(U_j - e_j)^2}{U_j} + \hat{W}^T \left[\hat{i}_{ll'} - \sum_{j=1}^r \hat{C}_{lj} \hat{C}_{l'j} \hat{A}_j^{-1} \right]^{-1} \hat{W}$$

Sous l'hypothèse nulle H_0 , cette statistique est distribuée selon une loi du chi-deux à k degrés de liberté.

6.4 Simulations

Une étude par simulations numériques s'avère nécessaire pour vérifier les différents résultats obtenus. On commence par calculer les estimateurs du maximum de vraisemblance sur données initiales non groupées pour des dizaines de milliers d'échantillons de différentes tailles, ensuite les valeurs des critères de test correspondants.

6.4.1 Estimation des paramètres

Pour confirmer la propriété de la convergence des estimateurs du maximum de vraisemblance, nous calculons l'erreur quadratique moyenne pour les EMV simulés de tous les paramètres du modèle. Supposons que le modèle $AFT - KwGEE$ est considéré, on simule $N = 10\,000$ fois des échantillons de différentes tailles, avec les valeurs suivantes des paramètres : $(\alpha = 2, \lambda = 1.5, \sigma = 0.7, \theta = 0.9, \beta_0 = 0.6, \beta_1 = 0.8,)$. Nous utilisons le logiciel statistique *R* et l'algorithme Barzilai-Borwein (**BB**) (Ravi, 2009)). Les valeurs de la moyenne des estimateurs du maximum de vraisemblance $\hat{\alpha}, \hat{\gamma}, \hat{\lambda}, \hat{\beta}_1$ et $\hat{\beta}_2$ et leurs erreurs quadratiques moyennes (*S.M.E*) sont calculées et données dans le tableau suivant (tailles des échantillons $n_1 = 15, n_2 = 25, n_3 = 50, n_4 = 130, n_5 = 350$ et $n_6 = 500$) :

$N = 10,000$	$n_1 = 15$	$n_2 = 25$	$n_3 = 50$	$n_4 = 130$	$n_5 = 350$	$n_6 = 500$
$\hat{\alpha}$	1.8854	1.8985	1.9254	1.9523	1.9785	1.9998
$S.M.E$	0.0035	0.0032	0.0029	0.0025	0.0018	0.0009
$\hat{\lambda}$	1.2564	1.2952	1.3256	1.3754	1.4849	1.4998
$S.M.E$	0.0032	0.0027	0.0021	0.0015	0.0010	0.0005
$\hat{\sigma}$	0.6721	0.6754	0.6825	0.6854	0.6982	0.7002
$S.M.E$	0.0036	0.0031	0.0025	0.0018	0.0012	0.0008
$\hat{\theta}$	0.9123	0.9125	0.9114	0.9044	0.9015	0.9002
$S.M.E$	0.0030	0.0024	0.0018	0.0012	0.0009	0.0003
$\hat{\beta}_0$	0.5423	0.5648	0.5742	0.5842	0.5912	0.5998
$S.M.E$	0.0029	0.0021	0.0017	0.0013	0.0008	0.0002
$\hat{\beta}_1$	0.8072	0.8062	0.8045	0.8024	0.8012	0.8002
$S.M.E$	0.0025	0.0018	0.0014	0.0010	0.0005	0.0003

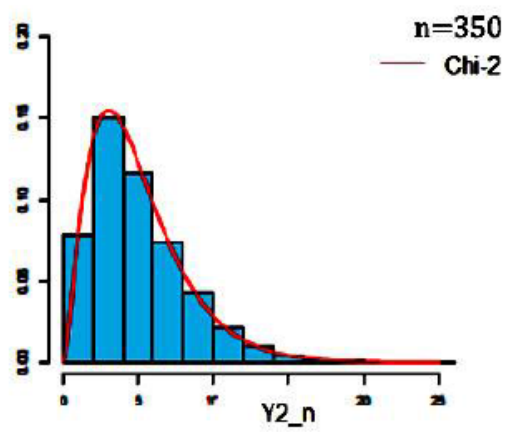
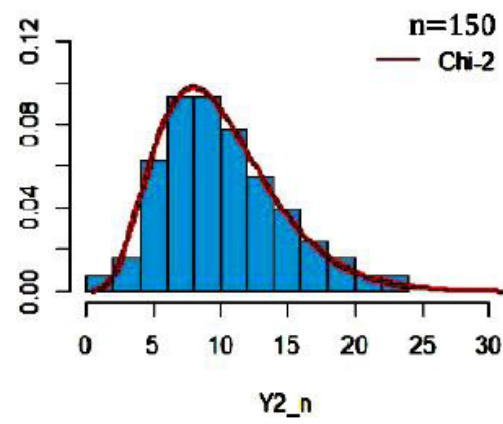
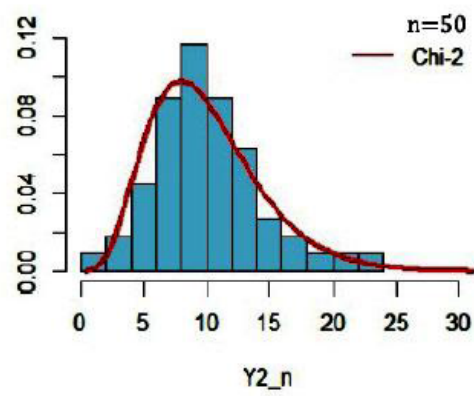
Tab 21 Erreurs quadratiques moyennes des estimateurs du maximum
de vraisemblance $\hat{\alpha}, \hat{\lambda}, \hat{\sigma}, \hat{\theta}, \hat{\beta}_0$ et $\hat{\beta}_1$.

Comme nous le constatons, les erreurs quadratiques moyennes sont très faibles et donc on confirme la convergence des estimateurs du maximum de vraisemblance obtenus.

6.4.2 La distribution de la statistique $Y_n^2(\hat{\gamma})$

Nous simulons $N = 10000$ fois des échantillons du modèle $AFT - KwGEE$, avec différentes valeurs des paramètres et avec des nombres différents de groupement de classes k .

Nous présentons dans les figures les histogrammes des échantillons simulés par rapport à la distribution du Chi-deux avec k ddl.



6.4.3 La statistique du test

Afin d'étudier la maniabilité et l'efficacité du test d'ajustement du type du chi-deux modifié pour le modèle $AFT-KwGEE(\gamma)$ dans le cas où les données sont censurées, proposé dans ce travail, on a réalisé une importante étude par simulations numériques. Ainsi et pour tester l'hypothèse nulle H_0 selon laquelle un échantillon de données provient de ce modèle, nous générons 10,000 échantillons de données censurées de distribution $AFT-KwGEE(\gamma)$ dont les tailles respectives des échantillons sont $n_1 = 15, n_2 = 25, n_3 = 50, n_4 = 130, n_5 = 350$ et $n_6 = 500$. On calcule les valeurs prises par la statistique $Y_n^2(\hat{\gamma})$ pour tous ces échantillons. Sous H_0 , cette statistique est distribuée selon un chi-deux, on compte alors le pourcentage de rejet de l'hypothèse H_0 pour différents seuils de signification que l'on compare avec les seuils théoriques correspondants. Les résultats obtenus sont exposés dans le tableau suivant :

$N = 10,000$	$n_1 = 15$	$n_2 = 25$	$n_3 = 50$	$n_4 = 130$	$n_5 = 350$	$n_6 = 500$
$\varepsilon = 1\%$	0.0089	0.0092	0.0093	0.0095	0.0098	0.0101
$\varepsilon = 5\%$	0.0467	0.0478	0.0489	0.0492	0.0495	0.0499
$\varepsilon = 10\%$	0.0983	0.0986	0.0992	0.0996	0.0998	0.1001

Tab.22. Comparaison des risques empirique et du risques théoriques correspondant
($\varepsilon = 0, 1, 0, 05, 0, 01$) pour les données censurées

A partir de ces résultats, on constate que les niveaux de signification empiriques de la statistique Y_n^2 coïncident avec ceux correspondants aux niveaux théoriques de la distributions du chi-deux à k degrés de liberté. Par conséquent, nous pouvons dire que le test proposé peut ajuster d'une manière efficace des données censurées de la distribution $AFT-KwGEE(\gamma)$.

6.5 Exemple de données de fiabilité

Pour déterminer la relation entre les taux de pannes de motos et la température de fonctionnement, Nelson et Hahn (1972) ont étudié les durées de vie (en heures) jusqu'aux pannes d'un ensemble de motos fonctionnant sous différentes températures. Ils utilisent les tests de vie accélérée où la température est un facteur de contrainte (stress) pour accélérer les taux de pannes par rapport à ce qui serait observé dans des conditions de fonctionnement normales, la température normale de fonctionnement étant de 130° . Le facteur de stress (température) est constant dans le temps pour n'importe quelle motorisation particulière. Les données sont présentées dans le tableau suivant :

Stress	
150°	Les 10 motos sans censurées 8064 heures
170°	1764, 2772, 3444, 3542, 3780, 4860, 5196
	les 3 motos sans censurées 5448 heures
190°	408, 408, 1344, 1344, 1440
	les 5 motos sans censurées 1680 heures
220°	408, 408, 504, 504, 504
	Les 5 motos sans censurées 528 heures

Table . durées de vie des motos (Nelson, 1972)

Supposons l'hypothèse H_0 que ces données suivent un modèle à durée de vie accéléré $AFT - KwGEE$, l'estimateur du vecteur de paramètres $\hat{\gamma}$ de γ obtenu est :

$$\hat{\gamma} = (\hat{\alpha}, \hat{\lambda}, \hat{\sigma}, \hat{\theta}, \hat{\beta}_0, \hat{\beta}_1)^T = (9.5684, 7.2548, -4.2135, 2.259, 1.2356, 0.0456)^T$$

6.5. Exemple de données de fiabilité

Nous allons regrouper ces observations dans $k = 4$ intervalles I_j , tous les calculs intermédiaires sont présentés dans le tableau suivant :

$\hat{a}_j$	8356	2185	6687	8064
U_j	5	5	3	10
$\hat{C}_{1j}$	0.0124	-2.1544	0.7842	-3.546
$\hat{C}_{2j}$	2.1434	1.2398	2.4578	1.6520
$\hat{C}_{3j}$	-9.6544	-7.1584	-4.2589	-6.3214
$\hat{C}_{4j}$	0.1478	0.0012	0.0074	0.0436
$\hat{C}_{5j}$	-7.1548	-5.2146	-6.3254	-8.1245
$\hat{C}_{6j}$	0.9654	0.8456	0.7458	0.0142

Tab 23 Calculs des $\hat{a}_j$; U_j ; $\hat{C}_{1j}$; $\hat{C}_{2j}$; $\hat{C}_{3j}$; $\hat{C}_{4j}$; $\hat{C}_{5j}$; $\hat{C}_{6j}$

Les effectifs théoriques obtenus sont $e_j = 7.2569$ pour toutes les classes.

Ainsi, on obtient la valeur du critère de test Y_n^2

$$Y_n^2 = X^2 + Q = 2.3569 + 4.2578 = 6.6147$$

Au seuil $\varepsilon = 0.05$, la valeur critique $\chi_{0.05}^2(4) = 9.488$ largement supérieure à la valeur de $Y_n^2 = 6.6147$, on conclut que les données de Nelson sont distribuées selon le modèle proposé *AFT - KwGEE*. Le modèle étant validé, on peut estimer la fiabilité ou bien toutes autres caractéristiques relatives au fonctionnement de ces types de motos.

7.1 Introduction

Comme les distributions existantes ne sont plus en mesure de modéliser toutes les données réelles obtenues dans les différents domaines, de nouvelles généralisations sont proposées dans la littérature récente. Grâce aux multiples avantages de la distribution de Kumaraswamy (1980), particulièrement dans les simulations, beaucoup de nouveaux modèles sont basés sur cette distribution. Etant donné que la variable aléatoire t de la distribution de Kumaraswamy est comprise entre 0 et 1, Cordeiro et Castro (2011) proposent de la remplacer par une fonction de répartition G , ce qui permet d'avoir différents modèles selon la distribution de base choisie. Ils obtiennent alors les distributions Kumaraswamy généralisées ($Kw - G$) en particulier celles dont les distributions de base G sont les distributions de Weibull, gamma, Gumbel, Kumaraswamy et normale. Mohammed (2014) a considéré la distribution Kumaraswamy-exponentielle exponentiée généralisée $Kw - GEE$. D'un autre côté, Lemonte *et al.* (2013) ont suggéré l'exponentiation de la distribution Kumaraswamy (EKw). En combinant les deux approches, Huang et Oluyede (2014) ont proposé la distribution de Kumaraswamy-Dagum exponentiée ($EKw - D$), tandis que Rodriguez et al. (2015), ils ont considéré l'exponentiation de la distribution Kumaraswamy-exponentielle ($EKw - \exp$). Les auteurs ont étudié les propriétés statistiques des ces distributions et l'estimation du maximum de vraisemblance

dans le cas de données complètes sans toutefois proposer de techniques pour valider ces nouveaux modèles.

En suivant la même approche, nous introduisons une nouvelle distribution, flexible, obtenue par l'exponentiation de la distribution $Kw - GEE$ proposée par Mohammed (2014). Cette distribution sera notée par $EKw - GEE$. Pour ces types de généralisations, les paramètres additionnels permettent de modéliser l'aplatissement, l'asymétrie et la forme de la queue des distributions ce qui les rend plus utiles dans les applications. Nous utilisons la statistique introduite par Bagdonavicius et Nikulin (2011) pour développer un test d'ajustement du type du chi-deux modifié pour ce modèle, dans le cas de censure aléatoire droite. Ainsi avec la mise en place de ce test, on peut vérifier l'adéquation de plusieurs sous-modèles tels que la distribution $Kw - GEE$, les distributions exponentielles exponentiées (EE) avec différents paramètres, la distribution Kumaraswamy-exponentielle $Kw - Exp$ (Nadarajah et al., 2012), et la distribution $(EKw - exp)$ proposée par Rodriguez *et al.* (2015).

7.2 Construction du modèle $EKw - GEE$

La distribution Kumaraswamy exponentiée exponentielle généralisée $EKw - GEE$ est définie par une puissance positive ($\theta > 0$) d'une distribution généralisée de Kumaraswamy avec la distribution exponentielle exponentiée généralisée GEE comme distribution de base. La fonction de densité cumulée correspondante (cdf) est définie par

$$F(t, \gamma) = [1 - \{1 - (1 - e^{-\lambda t})^{\alpha\beta}\}^{\sigma}]^{\theta}$$

avec le vecteur des paramètres $\gamma = (\lambda, \alpha, \beta, \sigma, \theta)^T$ où $\lambda > 0$ est le paramètre d'échelle, et $\alpha, \beta, \sigma, \theta > 0$ sont des paramètres de forme. La fonction de densité de probabilité (pdf) est définie comme suit

$$f(t, \gamma) = \alpha\beta\lambda\theta\sigma e^{-\lambda t}(1 - e^{-\lambda t})^{\alpha\beta-1} \{1 - (1 - e^{-\lambda t})^{\alpha\beta}\}^{\sigma-1} [1 - \{1 - (1 - e^{-\lambda t})^{\alpha\beta}\}^{\sigma}]^{\theta-1}.$$

La fonction de survie et le taux de hasard sont respectivement :

$$S(t; \gamma) = 1 - [1 - \{1 - (1 - e^{-\lambda t})^{\alpha\beta}\}^{\sigma}]^{\theta}$$

et

$$h(t; \gamma) = \frac{\alpha\beta\lambda\theta\sigma e^{-\lambda t}(1 - e^{-\lambda t})^{\alpha\beta-1} \{1 - (1 - e^{-\lambda t})^{\alpha\beta}\}^{\sigma-1} [1 - \{1 - (1 - e^{-\lambda t})^{\alpha\beta}\}^{\sigma}]^{\theta-1}}{1 - [1 - \{1 - (1 - e^{-\lambda t})^{\alpha\beta}\}^{\sigma}]^{\theta}}$$

$$H(t; \gamma) = -\ln \left(1 - [1 - \{1 - (1 - e^{-\lambda t})^{\alpha\beta}\}^{\sigma}]^{\theta} \right)$$

Pour simuler une variable de loi de $EKw - GEE$, on utilise l'équation non linéaire

$$T = -\frac{1}{\lambda} \ln \left(1 - \left\{ 1 - (1 - Z^{1/\theta})^{1/\sigma} \right\}^{1/\alpha\beta} \right)$$

Où Z suit la distribution uniforme sur $[0, 1]$.

Selon les valeurs prises par les paramètres de la distribution, la fonction du taux de hasard et la fonction de densité de probabilité ont différentes formes, comme illustré dans les figures 17, 18 et 19. La fonction de densité de probabilité est décroissante ou en $\cap$ avec une queue plus ou moins aplatie et un taux de hasard de croissance plus ou moins rapide.

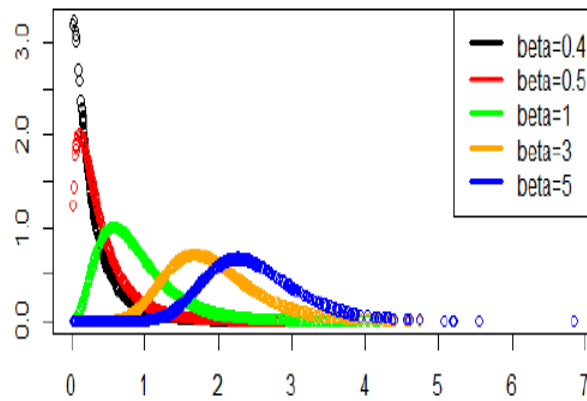


Fig. 17. Densité de la $EKw - GEE$ pour $\beta = (0.4, 0.5, 1, 3, 5)$

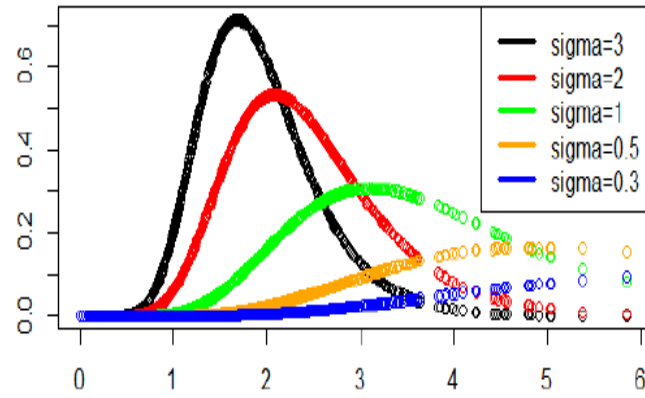


Fig. 18. Densité de la $EKw - GEE$ pour $\sigma = (3, 2, 1, 0.5, 0.3)$

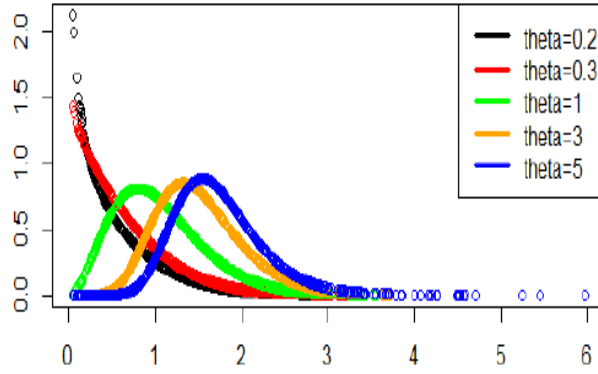


Fig. 19. Densité de la $EKw - GEE$ pour $\theta = (0.2, 0.3, 1, 3, 5)$

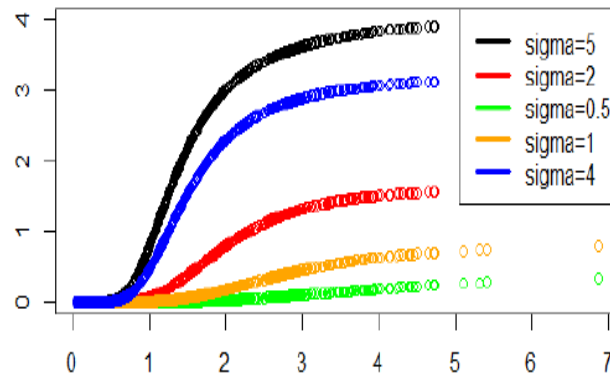


Fig. 20. Taux de hasard de la $EKw - GEE$ pour $\sigma = (5, 2, 0.5, 1, 4)$

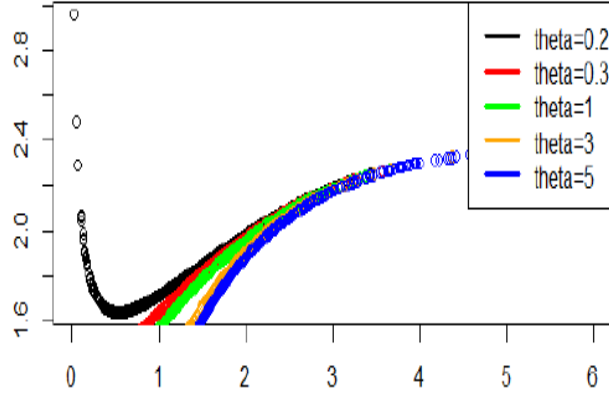


Fig. 21. Taux de hasard de la $EKw - GEE$ pour $\theta = (0.2, 0.3, 1, 3, 5)$

Notons que si $\theta = 1$, la distribution $EKw - GEE$ est réduite à la distribution de Kumaraswamy exponentielle exponentiée généralisée $Kw - GEE$ proposée récemment par Mohammed (2014).

7.3 Estimation du maximum de vraisemblance pour données censurées

Pour un modèle $EKw - GEE$, l'équation de la log-vraisemblance sur un échantillon aléatoire avec censure aléatoire droite non informative, est donnée par la formule suivante :

$$\begin{aligned}
 L(t; \gamma) = & \sum_{i=1}^n \delta_i [\ln \alpha \beta \lambda \theta \sigma - \lambda t_i + (\alpha \beta - 1) \ln (1 - e^{-\lambda t_i}) + (\sigma - 1) \ln (1 - (1 - e^{-\lambda t_i})^{\alpha \beta}) \\
 & + (\theta - 1) \ln (1 - \{1 - (1 - e^{-\lambda t_i})^{\alpha \beta}\}^{\sigma}) - \ln (1 - [1 - \{1 - (1 - e^{-\lambda t_i})^{\alpha \beta}\}^{\sigma}]^{\theta})] \\
 & + \sum_{i=1}^n \ln (1 - [1 - \{1 - (1 - e^{-\lambda t_i})^{\alpha \beta}\}^{\sigma}]^{\theta})
 \end{aligned}$$

on pose

$$u = 1 - e^{-\lambda t}$$

Les fonctions de score pour les paramètres $\alpha, \lambda, \beta, \sigma$ et θ sont obtenues par :

$$\begin{aligned} \frac{\partial L}{\partial \alpha} = & \sum_{i=1}^n \delta_i \left[\frac{1}{\alpha} + \beta \ln u_i - \beta (\sigma - 1) \frac{u_i^{\alpha\beta} \ln u_i}{1 - u_i^{\alpha\beta}} + \sigma \beta (\theta - 1) \frac{u_i^{\alpha\beta} \ln u_i (1 - u_i^{\alpha\beta})^{\sigma-1}}{1 - (1 - u_i^{\alpha\beta})^\sigma} \right. \\ & + \theta \beta \sigma \frac{u_i^{\alpha\beta} \ln u_i (1 - u_i^{\alpha\beta})^{\sigma-1} [1 - (1 - u_i^{\alpha\beta})^\sigma]^{\theta-1}}{1 - [1 - (1 - u_i^{\alpha\beta})^\sigma]^{\theta-1}} \left. \right] \\ & - \theta \beta \sigma \sum_{i=1}^n \frac{u_i^{\alpha\beta} \ln u_i (1 - u_i^{\alpha\beta})^{\sigma-1} [1 - (1 - u_i^{\alpha\beta})^\sigma]^{\theta-1}}{1 - [1 - (1 - u_i^{\alpha\beta})^\sigma]^{\theta-1}} \end{aligned}$$

$$\begin{aligned} \frac{\partial L}{\partial \lambda} = & \sum_{i=1}^n \delta_i \left[\frac{1}{\lambda} - t_i + (\alpha\beta - 1) \frac{t_i e^{-\lambda t_i}}{1 - e^{-\lambda t_i}} - \alpha\beta (\sigma - 1) \frac{t_i e^{-\lambda t_i} u_i^{\alpha\beta}}{1 - u_i^{\alpha\beta}} \right. \\ & + \alpha\sigma(\theta - 1) \frac{t_i e^{-\lambda t_i} u_i^{\alpha\beta-1} (1 - u_i^{\alpha\beta})^{\sigma-1}}{1 - (1 - u_i^{\alpha\beta})^\sigma} \\ & + \alpha\theta\beta\sigma \frac{t_i e^{-\lambda t_i} u_i^{\alpha\beta-1} (1 - u_i^{\alpha\beta})^{\sigma-1} [1 - (1 - u_i^{\alpha\beta})^\sigma]^{\theta-1}}{1 - [1 - (1 - u_i^{\alpha\beta})^\sigma]^{\theta-1}} \left. \right] \\ & - \alpha\theta\beta\sigma \sum_{i=1}^n \frac{t_i e^{-\lambda t_i} u_i^{\alpha\beta-1} (1 - u_i^{\alpha\beta})^{\sigma-1} (1 - (1 - u_i^{\alpha\beta})^\sigma)^{\theta-1}}{1 - [1 - (1 - u_i^{\alpha\beta})^\sigma]^{\theta-1}} \end{aligned}$$

$$\begin{aligned} \frac{\partial L}{\partial \beta} = & \sum_{i=1}^n \delta_i \left[\frac{1}{\beta} + \alpha \ln u_i - \alpha (\sigma - 1) \frac{u_i^{\alpha\beta} \ln u_i}{1 - u_i^{\alpha\beta}} + \alpha\sigma(\theta - 1) \frac{u_i^{\alpha\beta} \ln u_i (1 - u_i^{\alpha\beta})^{\sigma-1}}{1 - (1 - u_i^{\alpha\beta})^\sigma} \right. \\ & + \alpha\sigma\theta \frac{u_i^{\alpha\beta} \ln u_i (1 - u_i^{\alpha\beta})^{\sigma-1} [1 - (1 - u_i^{\alpha\beta})^\sigma]^{\theta-1}}{1 - [1 - (1 - u_i^{\alpha\beta})^\sigma]^{\theta-1}} \left. \right] \\ & - \alpha\sigma\theta \sum_{i=1}^n \frac{u_i^{\alpha\beta} \ln u_i (1 - u_i^{\alpha\beta})^{\sigma-1} [1 - (1 - u_i^{\alpha\beta})^\sigma]^{\theta-1}}{1 - [1 - (1 - u_i^{\alpha\beta})^\sigma]^{\theta-1}} \end{aligned}$$

$$\begin{aligned}
 \frac{\partial L}{\partial \sigma} &= \sum_{i=1}^n \delta_i \left[\frac{1}{\sigma} + \ln(1 - u_i^{\alpha\beta}) - (\theta - 1) \frac{(1 - u_i^{\alpha\beta})^\sigma \ln(1 - u_i^{\alpha\beta})}{1 - (1 - u_i^{\alpha\beta})^\sigma} \right. \\
 &\quad \left. - \theta \frac{(1 - u_i^{\alpha\beta})^\sigma \ln(1 - u_i^{\alpha\beta}) [1 - (1 - u_i^{\alpha\beta})^\sigma]^{\theta-1}}{1 - [1 - (1 - u_i^{\alpha\beta})^\sigma]^{\theta-1}} \right] \\
 &\quad + \theta \sum_{i=1}^n \frac{(1 - u_i^{\alpha\beta})^\sigma \ln(1 - u_i^{\alpha\beta}) [1 - (1 - u_i^{\alpha\beta})^\sigma]^{\theta-1}}{1 - [1 - (1 - u_i^{\alpha\beta})^\sigma]^{\theta-1}} \\
 \\
 \frac{\partial L}{\partial \theta} &= \sum_{i=1}^n \delta_i \left(\frac{1}{\theta} + \frac{\ln(1 - (1 - u_i^{\alpha\beta})^\sigma)}{1 - [1 - (1 - u_i^{\alpha\beta})^\sigma]^\theta} \right) \\
 &\quad + \sum_{i=1}^n \frac{[1 - (1 - u_i^{\alpha\beta})^\sigma]^\theta \ln(1 - (1 - u_i^{\alpha\beta})^\sigma)}{1 - [1 - (1 - u_i^{\alpha\beta})^\sigma]^\theta}
 \end{aligned}$$

Nous obtenons un système d'équations non linéaires en égalant ces fonctions à zéro. Nous utilisons alors des méthodes numériques pour déterminer les estimateurs du maximum de vraisemblance des paramètres solutions de ces équations.

7.4 Test d'ajustement pour le modèle $EKw - GEE$

Vu l'importance de la distribution $EKw - GEE$ du fait qu'elle constitue des généralisations de plusieurs distributions, nous proposons un test d'ajustement basé sur l'approche de Bagdonavicius et Nikulin (2011). Le critère de test déterminé sera utilisé aussi pour vérifier l'adéquation des différents sous-modèles.

Supposons H_0 que le taux de défaillance T_i suit une distribution $EKw - GEE$. La fonction de survie est :

$$S(t; \hat{\gamma}) = 1 - \left[1 - \left\{ 1 - (1 - e^{-\hat{\lambda}t})^{\hat{\alpha}\hat{\beta}} \right\}^{\hat{\sigma}} \right]^{\hat{\theta}}$$

Après avoir calculé les estimateurs du maximum de vraisemblance des paramètres $\alpha, \lambda, \beta, \theta, \sigma$, nous calculons les nombres de défaillances théoriques :

$$E_j = -\frac{j}{k-1} \sum_{i=1}^n \ln \left(1 - \left[1 - \left\{ 1 - (1 - e^{-\hat{\lambda} t_i})^{\hat{\alpha} \hat{\beta}} \right\}^{\hat{\sigma}} \right]^{\hat{\theta}} \right),$$

Le choix de $\hat{a}_j$ lorsque la distribution de base est une $EKw - GEE$, est obtenu comme suit :

$$\hat{a}_j = -\frac{1}{\hat{\lambda}} \ln \left\{ 1 - \left[1 - \left(1 - \left\{ 1 - \exp \left(\frac{\sum_{l=1}^{i-1} H(t_l, \hat{\gamma}) - E_j}{n - i + 1} \right) \right\}^{1/\hat{\theta}} \right)^{1/\hat{\sigma}} \right]^{1/\hat{\alpha} \hat{\beta}} \right\}$$

pour $j = 1, \dots, k-1$, $\hat{a}_k = t_{(n)}$. Les intervalles $I_j = (a_{j-1}, a_j]$ de groupement des données obtenus auront les mêmes effectifs e_j .

Calcul de la matrice $\hat{W}$ et $\hat{C}$

Pour le calcul de la matrice $\hat{W}$, on a besoin des dérivées partielles de la fonction $\ln h(t; \hat{\gamma})$ par rapport à $\alpha, \lambda, \beta, \sigma$ et θ .

Après quelques simplifications, nous trouvons les éléments de la matrice $\hat{C}$:

$$\begin{aligned} \hat{C}_{1j} &= \frac{1}{n} \sum_{i:t_i \in I_j} \delta_i \left[\frac{1}{\alpha} + \beta \ln u_i - \beta(\sigma - 1) \frac{u_i^{\alpha\beta} \ln u_i}{1 - u_i^{\alpha\beta}} + \sigma\beta(\theta - 1) \frac{u_i^{\alpha\beta} \ln u_i (1 - u_i^{\alpha\beta})^{\sigma-1}}{1 - (1 - u_i^{\alpha\beta})^\sigma} \right. \\ &\quad \left. + \theta\beta\sigma \frac{u_i^{\alpha\beta} \ln u_i (1 - u_i^{\alpha\beta})^{\sigma-1} [1 - (1 - u_i^{\alpha\beta})^\sigma]^{\theta-1}}{1 - [1 - (1 - u_i^{\alpha\beta})^\sigma]^{\theta-1}} \right] \\ \hat{C}_{2j} &= \frac{1}{n} \sum_{i:t_i \in I_j} \delta_i \left[\frac{1}{\lambda} - t_i + (\alpha\beta - 1) \frac{t_i e^{-\lambda t_i}}{1 - e^{-\lambda t_i}} - \alpha\beta(\sigma - 1) \frac{t_i e^{-\lambda t_i} u_i^{\alpha\beta}}{1 - u_i^{\alpha\beta}} \right. \\ &\quad \left. + \alpha\sigma(\theta - 1) \frac{t_i e^{-\lambda t_i} u_i^{\alpha\beta-1} (1 - u_i^{\alpha\beta})^{\sigma-1}}{1 - (1 - u_i^{\alpha\beta})^\sigma} \right. \\ &\quad \left. + \alpha\theta\beta\sigma \frac{t_i e^{-\lambda t_i} u_i^{\alpha\beta-1} (1 - u_i^{\alpha\beta})^{\sigma-1} [1 - (1 - u_i^{\alpha\beta})^\sigma]^{\theta-1}}{1 - [1 - (1 - u_i^{\alpha\beta})^\sigma]^{\theta-1}} \right] \end{aligned}$$

$$\begin{aligned}\hat{C}_{3j} = & \frac{1}{n} \sum_{i:t_i \in I_j} \delta_i \left[\frac{1}{\beta} + \alpha \ln u_i - \alpha(\sigma - 1) \frac{u_i^{\alpha\beta} \ln u_i}{1 - u_i^{\alpha\beta}} + \alpha\sigma(\theta - 1) \frac{u_i^{\alpha\beta} \ln u_i (1 - u_i^{\alpha\beta})^{\sigma-1}}{1 - (1 - u_i^{\alpha\beta})^\sigma} \right. \\ & \left. + \alpha\sigma\theta \frac{u_i^{\alpha\beta} \ln u_i (1 - u_i^{\alpha\beta})^{\sigma-1} [1 - (1 - u_i^{\alpha\beta})^\sigma]^{\theta-1}}{1 - [1 - (1 - u_i^{\alpha\beta})^\sigma]^{\theta-1}} \right]\end{aligned}$$

$$\begin{aligned}\hat{C}_{4j} = & \frac{1}{n} \sum_{i:x_i \in I_j} \delta_i \left[\frac{1}{\sigma} + \ln(1 - u_i^{\alpha\beta}) - (\theta - 1) \frac{(1 - u_i^{\alpha\beta})^\sigma \ln(1 - u_i^{\alpha\beta})}{1 - (1 - u_i^{\alpha\beta})^\sigma} \right. \\ & \left. - \theta \frac{(1 - u_i^{\alpha\beta})^\sigma \ln(1 - u_i^{\alpha\beta}) [1 - (1 - u_i^{\alpha\beta})^\sigma]^{\theta-1}}{1 - [1 - (1 - u_i^{\alpha\beta})^\sigma]^{\theta-1}} \right]\end{aligned}$$

$$\hat{C}_{5j} = \frac{1}{n} \sum_{i:x_i \in I_j} \delta_i \left(\frac{1}{\theta} + \frac{\ln(1 - (1 - u_i^{\alpha\beta})^\sigma)}{1 - [1 - (1 - u_i^{\alpha\beta})^\sigma]^\theta} \right)$$

On déduit la matrice $\hat{W}$ d'éléments :

$$\hat{W}_l = \sum_{j=1}^r \hat{C}_{lj} A_j^{-1} Z_j, \quad l = 1, 2, 3, 4, 5 \quad j = 1, \dots, k$$

Calcul de la matrice d'information estimée $\hat{I}$

Si on pose

$$\begin{aligned}A(t_i, \gamma) &= \sigma(\theta - 1) \frac{u_i^{\alpha\beta} (1 - u_i^{\alpha\beta})^{\sigma-1}}{1 - (1 - u_i^{\alpha\beta})^\sigma} \\ M(t_i, \gamma) &= \sigma\theta \frac{u_i^{\alpha\beta} (1 - u_i^{\alpha\beta})^{\sigma-1} [1 - (1 - u_i^{\alpha\beta})^\sigma]^{\theta-1}}{1 - [1 - (1 - u_i^{\alpha\beta})^\sigma]^{\theta-1}} \\ R(t_i, \gamma) &= \frac{(1 - u_i^{\alpha\beta})^\sigma \ln(1 - u_i^{\alpha\beta}) [1 - (1 - u_i^{\alpha\beta})^\sigma]^{\theta-1}}{1 - [1 - (1 - u_i^{\alpha\beta})^\sigma]^{\theta-1}}\end{aligned}$$

après de longs calculs, les éléments de la matrice d'information de Fisher estimée $\hat{I}$ sont obtenues comme suit :

$$\begin{aligned}\hat{i}_{11} &= \frac{1}{n} \sum_{i=1}^n \delta_i \left(\frac{1}{\alpha} + \frac{\beta \ln u_i (1 - \sigma u_i^{\alpha\beta})}{1 - u_i^{\alpha\beta}} + \beta \ln u_i [A(t_i, \gamma) + M(t_i, \gamma)] \right)^2 \\ \hat{i}_{22} &= \frac{1}{n} \sum_{i=1}^n \delta_i \left(\frac{1}{\lambda} - \frac{t_i (1 - \alpha\beta e^{-\lambda t_i})}{1 - e^{-\lambda t_i}} - \alpha\beta (\sigma - 1) \frac{t_i e^{-\lambda t_i} u_i^{\alpha\beta}}{1 - u_i^{\alpha\beta}} + \alpha\beta \frac{t_i e^{-\lambda t_i}}{u_i} [A(t_i, \gamma) + M(t, \gamma)] \right)^2 \\ \hat{i}_{33} &= \frac{1}{n} \sum_{i=1}^n \delta_i \left(\frac{1}{\beta} + \frac{\alpha \ln u_i (1 - \sigma u_i^{\alpha\beta})}{1 - u_i^{\alpha\beta}} + \alpha \ln u_i [A(t_i, \gamma) + M(t_i, \gamma)] \right)^2 \\ \hat{i}_{44} &= \frac{1}{n} \sum_{i=1}^n \delta_i \left(\frac{1}{\sigma} + \frac{\ln (1 - u_i^{\alpha\beta}) (1 - \theta (1 - u_i^{\alpha\beta})^\sigma)}{1 - (1 - u_i^{\alpha\beta})^\sigma} - \theta R(t_i, \gamma) \right)^2 \\ \hat{i}_{55} &= \frac{1}{n} \sum_{i=1}^n \delta_i \left(\frac{1}{\theta} + \frac{\ln (1 - (1 - u_i^{\alpha\beta})^\sigma)}{1 - [1 - (1 - u_i^{\alpha\beta})^\sigma]^\theta} \right)^2 \\ \hat{i}_{12} &= \frac{1}{n} \sum_{i=1}^n \delta_i \left(\frac{1}{\alpha} + \frac{\beta \ln u_i (1 - \sigma u_i^{\alpha\beta})}{1 - u_i^{\alpha\beta}} + \beta \ln u_i [A(t_i, \gamma) + M(t_i, \gamma)] \right) \\ &\quad \times \left(\frac{1}{\lambda} - \frac{t_i (1 - \alpha\beta e^{-\lambda t_i})}{1 - e^{-\lambda t_i}} - \alpha\beta (\sigma - 1) \frac{t_i e^{-\lambda t_i} u_i^{\alpha\beta}}{1 - u_i^{\alpha\beta}} + \alpha\beta \frac{t_i e^{-\lambda t_i}}{u_i} [A(t_i, \gamma) + M(t, \gamma)] \right) \\ \hat{i}_{13} &= \frac{1}{n} \sum_{i=1}^n \delta_i \left(\frac{1}{\alpha} + \frac{\beta \ln u_i (1 - \sigma u_i^{\alpha\beta})}{1 - u_i^{\alpha\beta}} + \beta \ln u_i [A(t_i, \gamma) + M(t_i, \gamma)] \right) \\ &\quad \times \left(\frac{1}{\beta} + \frac{\alpha \ln u_i (1 - \sigma u_i^{\alpha\beta})}{1 - u_i^{\alpha\beta}} + \alpha \ln u_i [A(t_i, \gamma) + M(t_i, \gamma)] \right) \\ \hat{i}_{14} &= \frac{1}{n} \sum_{i=1}^n \delta_i \left(\frac{1}{\alpha} + \frac{\beta \ln u_i (1 - \sigma u_i^{\alpha\beta})}{1 - u_i^{\alpha\beta}} + \beta \ln u_i [A(t_i, \gamma) + M(t_i, \gamma)] \right) \\ &\quad \times \left(\frac{1}{\sigma} + \frac{\ln (1 - u_i^{\alpha\beta}) (1 - \theta (1 - u_i^{\alpha\beta})^\sigma)}{1 - (1 - u_i^{\alpha\beta})^\sigma} - \theta R(t_i, \gamma) \right)\end{aligned}$$

$$\begin{aligned}
 \hat{i}_{15} &= \frac{1}{n} \sum_{i=1}^n \delta_i \left(\frac{1}{\alpha} + \frac{\beta \ln u_i (1 - \sigma u_i^{\alpha\beta})}{1 - u_i^{\alpha\beta}} + \beta \ln u_i [A(t_i, \gamma) + M(t_i, \gamma)] \right) \\
 &\quad \times \left(\frac{1}{\theta} + \frac{\ln (1 - (1 - u_i^{\alpha\beta})^\sigma)}{1 - [1 - (1 - u_i^{\alpha\beta})^\sigma]^\theta} \right) \\
 \hat{i}_{23} &= \frac{1}{n} \sum_{i=1}^n \delta_i \left(\frac{1}{\lambda} - \frac{t_i (1 - \alpha\beta e^{-\lambda t_i})}{1 - e^{-\lambda t_i}} - \alpha\beta (\sigma - 1) \frac{t_i e^{-\lambda t_i} u_i^{\alpha\beta}}{1 - u_i^{\alpha\beta}} + \alpha\beta \frac{t_i e^{-\lambda t_i}}{u_i} [A(t_i, \gamma) + M(t, \gamma)] \right) \\
 &\quad \times \left(\frac{1}{\beta} + \frac{\alpha \ln u_i (1 - \sigma u_i^{\alpha\beta})}{1 - u_i^{\alpha\beta}} + \alpha \ln u_i [A(t_i, \gamma) + M(t_i, \gamma)] \right) \\
 \hat{i}_{24} &= \frac{1}{n} \sum_{i=1}^n \delta_i \left(\frac{1}{\lambda} - \frac{t_i (1 - \alpha\beta e^{-\lambda t_i})}{1 - e^{-\lambda t_i}} - \alpha\beta (\sigma - 1) \frac{t_i e^{-\lambda t_i} u_i^{\alpha\beta}}{1 - u_i^{\alpha\beta}} + \alpha\beta \frac{t_i e^{-\lambda t_i}}{u_i} [A(t_i, \gamma) + M(t, \gamma)] \right) \\
 &\quad \times \left(\frac{1}{\sigma} + \frac{\ln (1 - u_i^{\alpha\beta}) (1 - \theta (1 - u_i^{\alpha\beta})^\sigma)}{1 - (1 - u_i^{\alpha\beta})^\sigma} - \theta R(t_i, \gamma) \right) \\
 \hat{i}_{25} &= \frac{1}{n} \sum_{i=1}^n \delta_i \left(\frac{1}{\lambda} - \frac{t_i (1 - \alpha\beta e^{-\lambda t_i})}{1 - e^{-\lambda t_i}} - \alpha\beta (\sigma - 1) \frac{t_i e^{-\lambda t_i} u_i^{\alpha\beta}}{1 - u_i^{\alpha\beta}} + \alpha\beta \frac{t_i e^{-\lambda t_i}}{u_i} [A(t_i, \gamma) + M(t, \gamma)] \right) \\
 &\quad \times \left(\frac{1}{\theta} + \frac{\ln (1 - (1 - u_i^{\alpha\beta})^\sigma)}{1 - [1 - (1 - u_i^{\alpha\beta})^\sigma]^\theta} \right) \\
 \hat{i}_{34} &= \frac{1}{n} \sum_{i=1}^n \delta_i \left(\frac{1}{\beta} + \frac{\alpha \ln u_i (1 - \sigma u_i^{\alpha\beta})}{1 - u_i^{\alpha\beta}} + \alpha \ln u_i [A(t_i, \gamma) + M(t_i, \gamma)] \right) \\
 &\quad \times \left(\frac{1}{\sigma} + \frac{\ln (1 - u_i^{\alpha\beta}) (1 - \theta (1 - u_i^{\alpha\beta})^\sigma)}{1 - (1 - u_i^{\alpha\beta})^\sigma} - \theta R(t_i, \gamma) \right) \\
 \hat{i}_{35} &= \frac{1}{n} \sum_{i=1}^n \delta_i \left(\frac{1}{\beta} + \frac{\alpha \ln u_i (1 - \sigma u_i^{\alpha\beta})}{1 - u_i^{\alpha\beta}} + \alpha \ln u_i [A(t_i, \gamma) + M(t_i, \gamma)] \right) \\
 &\quad \times \left(\frac{1}{\theta} + \frac{\ln (1 - (1 - u_i^{\alpha\beta})^\sigma)}{1 - [1 - (1 - u_i^{\alpha\beta})^\sigma]^\theta} \right)
 \end{aligned}$$

$$\begin{aligned} \hat{i}_{45} = & \frac{1}{n} \sum_{i=1}^n \delta_i \left(\frac{1}{\sigma} + \frac{\ln \left(1 - u_i^{\alpha\beta} \right) \left(1 - \theta \left(1 - u_i^{\alpha\beta} \right)^\sigma \right)}{1 - \left(1 - u_i^{\alpha\beta} \right)^\sigma} - \theta R(t_i, \gamma) \right) \\ & \times \left(\frac{1}{\theta} + \frac{\ln \left(1 - \left(1 - u_i^{\alpha\beta} \right)^\sigma \right)}{1 - \left[1 - \left(1 - u_i^{\alpha\beta} \right)^\sigma \right]^\theta} \right) \end{aligned}$$

Les fréquences théoriques e_j et empiriques U_j des taux de pannes dans les classes I_j de groupement des données, sont calculées, les estimateurs du maximum de vraisemblance, la matrice d'information et les autres composants sont tous déterminés, on obtient alors la statistique du test d'ajustement Y_n^2 pour le modèle $EKw - GEE$:

$$Y_n^2(\hat{\gamma}) = \sum_{j=1}^r \frac{(U_j - e_j)^2}{U_j} + \hat{W}^T \left[\hat{i}_{ll'} - \sum_{j=1}^r \hat{C}_{lj} \hat{C}_{l'j} \hat{A}_j^{-1} \right]^{-1} \hat{W}$$

dont la distribution est un chi-deux à k degrés de libertés.

7.5 Simulations

Pour confirmer les résultats obtenus plus haut, nous conduisons une importante étude par simulations numériques. On calcule en premier lieu, les estimateurs du maximum de vraisemblance des paramètres des différents échantillons, ensuite les critères de tests respectifs.

7.5.1 Estimation des paramètres

Nous effectuons $N = 10,000$ simulations d'échantillons de données censurées du modèle $EKw - GEE$ ($\alpha = 2, \lambda = 1.5, \beta = 3, \sigma = 4, \theta = 0.5$), de tailles respectives $n_1 = 15, n_2 = 25, n_3 = 50, n_4 = 130, n_5 = 350$ et $n_6 = 500$. Les valeurs moyennes des estimateurs du maximum de vraisemblance $\hat{\alpha}, \hat{\lambda}, \hat{\beta}, \hat{\sigma}, \hat{\theta}$ et leurs erreurs quadratiques moyennes sont calculées pour tous les échantillons et sont réunies dans la table 24.

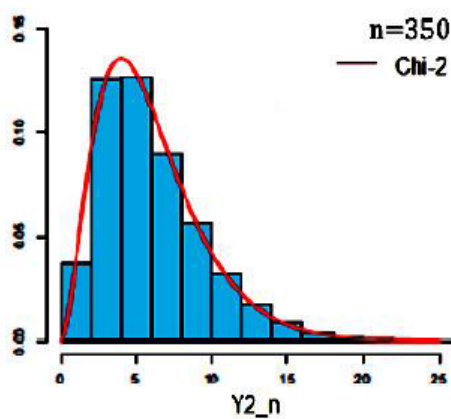
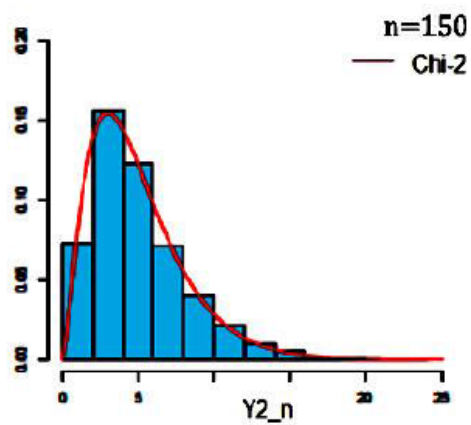
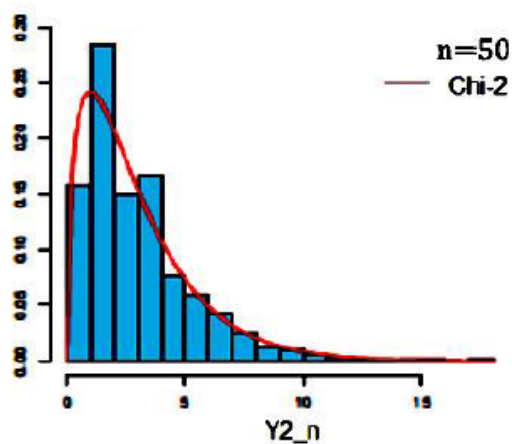
$N = 10,000$	$n_1 = 15$	$n_2 = 25$	$n_3 = 50$	$n_4 = 130$	$n_5 = 350$	$n_6 = 500$
$\hat{\alpha}$	1.9124	1.9389	1.9459	1.9683	1.9889	1.9999
$S.M.E$	0.0035	0.0028	0.0022	0.0019	0.0015	0.0008
$\hat{\lambda}$	1.5254	1.9547	1.9789	1.9874	1.9921	1.9935
$S.M.E$	0.0042	0.0035	0.0029	0.0020	0.0015	0.0011
$\hat{\beta}$	0.5523	0.5415	0.5369	0.5221	0.5147	0.5012
$S.M.E$	0.0040	0.0032	0.0027	0.0019	0.0010	0.0005
$\hat{\sigma}$	0.8547	0.8496	0.8320	0.8298	0.8196	0.8007
$S.M.E$	0.0031	0.0029	0.0025	0.0019	0.0015	0.0004
$\hat{\theta}$	0.5430	0.5369	0.5322	0.5296	0.5201	0.5102
$S.M.E$	0.0021	0.0019	0.0015	0.0012	0.0008	0.0007

Tab. 24. Les valeurs moyennes des EMV des paramètres et leurs erreurs
quadratiques moyennes pour les données censurées

Les résultats des simulations confirment le fait que les estimateurs du maximum de vraisemblance obtenus sont convergents.

7.5.2 La distribution de la statistique $Y_n^2(\hat{\gamma})$

Pour montrer que la distribution de la statistique de test est invariante, nous avons effectué des simulations ($N = 10.000$) de la distribution de la statistique $Y^2(\hat{\gamma})$ sous l'hypothèse nulle H_0 , avec différentes valeurs de paramètres et différents nombres d'intervalles k de regroupement de données soumises à une censure droite. Ensuite, nous avons représenté les histogrammes des données des échantillons par rapport aux distributions du chi-deux correspondantes avec k degrés de liberté (voir les exemples dans les figures).



Tenant compte des erreurs des simulations, nous pouvons dire que la distribution limite de la statistique $Y^2(\hat{\gamma})$, est une distribution du chi-deux avec k degrés de liberté.

7.5.3 La statistique du test

Pour tester l'hypothèse nulle H_0 selon laquelle un échantillon de données censurées à droite, provient d'un modèle Kumaraswamy exponentielle exponentielle généralisée $EKw - GEE(\hat{\gamma})$, nous utilisons les échantillons générés plus haut dont les tailles respectives sont $n_1 = 15$, $n_2 = 25$, $n_3 = 50$, $n_4 = 130$, $n_5 = 350$ et $n_6 = 500$. On calcule les valeurs prises par la statistique de test $Y^2(\hat{\gamma})$ pour chacun de tous les échantillons. Comme cette statistique suit une distribution du chi-deux à k degrés de libertés sous l'hypothèse nulle H_0 , on compte le nombre de cas où $Y^2 \succ \chi_k^2$ pour différents seuils de signification que l'on compare aux seuils de signification théoriques correspondants. Les résultats sont exposés dans le tableau suivant :

$N = 10,000$	$n_1 = 15$	$n_2 = 25$	$n_3 = 50$	$n_4 = 130$	$n_5 = 350$	$n_6 = 500$
$\varepsilon = 1\%$	0.0086	0.0092	0.0095	0.0099	0.0122	0.0102
$\varepsilon = 5\%$	0.0456	0.0472	0.0486	0.0492	0.0495	0.0502
$\varepsilon = 10\%$	0.0952	0.0963	0.0976	0.0984	0.0997	0.1002

Table 25. Comparaison des risques empiriques et des risques théoriques correspondants

($\varepsilon = 0,1, 0,05, 0,01$) pour les données censurées

Nous remarquons que les résultats sont très satisfaisants et donc que cette statistique, distribuée selon un chi-deux à k degrés de libertés, nous permet d'ajuster une série de données au modèle $EKw - GEE$.

Comme on l'a précisé plus haut, cette distribution généralise plusieurs autres. On va utiliser ce critère de test pour déduire les différentes statistiques qui permettent de vérifier l'ajustement des distributions en question. En suivant le même procédé, on calcule les valeurs des statistiques de test selon les distributions à valider.

$N = 10,000$	$n_1 = 15$	$n_2 = 25$	$n_3 = 50$	$n_4 = 130$	$n_5 = 350$	$n_6 = 500$
$\varepsilon = 1\%$	0.0075	0.0082	0.0089	0.0092	0.0097	0.0099
$\varepsilon = 5\%$	0.0463	0.0475	0.0485	0.0488	0.0495	0.0498
$\varepsilon = 10\%$	0.0948	0.0953	0.0979	0.0986	0.0992	0.0995

Table 26. Simulated levels of significance for $Y_n^2(\hat{\gamma})$ test for $Kw - GEE(\lambda, \alpha, \beta, \sigma)$

$N = 10,000$	$n_1 = 15$	$n_2 = 25$	$n_3 = 50$	$n_4 = 130$	$n_5 = 350$	$n_6 = 500$
$\varepsilon = 1\%$	0.0079	0.0084	0.0088	0.0094	0.0098	0.0101
$\varepsilon = 5\%$	0.0478	0.0489	0.0492	0.0505	0.0508	0.0502
$\varepsilon = 10\%$	0.0898	0.0942	0.0975	0.0995	0.103	0.1003

Table 27. Simulated levels of significance for $Y_n^2(\hat{\gamma})$ test for $EKw - E(\lambda, \beta, \sigma, \theta)$.

$N = 10,000$	$n_1 = 15$	$n_2 = 25$	$n_3 = 50$	$n_4 = 130$	$n_5 = 350$	$n_6 = 500$
$\varepsilon = 1\%$	0.0081	0.0089	0.0092	0.0097	0.0099	0.0103
$\varepsilon = 5\%$	0.0489	0.0492	0.0498	0.0512	0.0505	0.0502
$\varepsilon = 10\%$	0.0921	0.0955	0.0965	0.0978	0.0989	0.0995

Table 28. Simulated levels of significance for $Y_n^2(\hat{\gamma})$ test for $Kw - E(\lambda, \beta, \sigma)$.

$N = 10,000$	$n_1 = 15$	$n_2 = 25$	$n_3 = 50$	$n_4 = 130$	$n_5 = 350$	$n_6 = 500$
$\varepsilon = 1\%$	0.0089	0.0092	0.0095	0.0092	0.0099	0.0103
$\varepsilon = 5\%$	0.0456	0.0478	0.0489	0.0492	0.0498	0.0504
$\varepsilon = 10\%$	0.0932	0.0956	0.0978	0.0989	0.0992	0.0999

Table 29. Simulated levels of significance for $Y_n^2(\hat{\gamma})$ test for $Exp(\lambda)$.

$N = 10,000$	$n_1 = 15$	$n_2 = 25$	$n_3 = 50$	$n_4 = 130$	$n_5 = 350$	$n_6 = 500$
$\varepsilon = 1\%$	0.0082	0.0089	0.0091	0.0094	0.0098	0.0099
$\varepsilon = 5\%$	0.0432	0.0466	0.0477	0.0488	0.0492	0.0499
$\varepsilon = 10\%$	0.0943	0.0975	0.0986	0.0992	0.0098	0.1003

Table 30. Simulated levels of significance for $Y_n^2(\hat{\gamma})$ test for $Exp(\lambda\theta)$

$N = 10,000$	$n_1 = 15$	$n_2 = 25$	$n_3 = 50$	$n_4 = 130$	$n_5 = 350$	$n_6 = 500$
$\varepsilon = 1\%$	0.0083	0.0085	0.0089	0.0093	0.0098	0.0105
$\varepsilon = 5\%$	0.0432	0.0449	0.0462	0.0478	0.0489	0.0502
$\varepsilon = 10\%$	0.0945	0.0978	0.0989	0.0995	0.1012	0.1004

Table 31. Simulated levels of significance for $Y_n^2(\hat{\gamma})$ test for $GEE(\lambda, \alpha)$.

$N = 10,000$	$n_1 = 15$	$n_2 = 25$	$n_3 = 50$	$n_4 = 130$	$n_5 = 350$	$n_6 = 500$
$\varepsilon = 1\%$	0.0081	0.0084	0.0087	0.0094	0.0099	0.0102
$\varepsilon = 5\%$	0.0412	0.0434	0.0456	0.0471	0.0482	0.0498
$\varepsilon = 10\%$	0.0932	0.0945	0.0975	0.0986	0.0988	0.0999

Table 32. Simulated levels of significance for $Y_n^2(\hat{\gamma})$ test for $GEE(\lambda, \theta)$.

$N = 10,000$	$n_1 = 15$	$n_2 = 25$	$n_3 = 50$	$n_4 = 130$	$n_5 = 350$	$n_6 = 500$
$\varepsilon = 1\%$	0.0078	0.0083	0.0086	0.0092	0.0097	0.0099
$\varepsilon = 5\%$	0.0456	0.0478	0.0482	0.0489	0.0495	0.0499
$\varepsilon = 10\%$	0.0912	0.0932	0.0965	0.0979	0.989	0.999

Table 33. Simulated levels of significance for $Y_n^2(\hat{\gamma})$ test for $GEE(\lambda, \alpha\beta)$.

$N = 10,000$	$n_1 = 15$	$n_2 = 25$	$n_3 = 50$	$n_4 = 130$	$n_5 = 350$	$n_6 = 500$
$\varepsilon = 1\%$	0.0084	0.0087	0.0091	0.0096	0.0099	0.0103
$\varepsilon = 5\%$	0.0465	0.0472	0.0483	0.0488	0.0495	0.0502
$\varepsilon = 10\%$	0.0935	0.0955	0.0977	0.0982	0.0994	0.1001

Table 34. Simulated levels of significance for $Y_n^2(\hat{\gamma})$ test for $GEE(\lambda\sigma, \theta)$.

$N = 10,000$	$n_1 = 15$	$n_2 = 25$	$n_3 = 50$	$n_4 = 130$	$n_5 = 350$	$n_6 = 500$
$\varepsilon = 1\%$	0.0091	0.0094	0.0098	0.0123	0.0112	0.0102
$\varepsilon = 5\%$	0.0484	0.0491	0.0499	0.0532	0.0511	0.0504
$\varepsilon = 10\%$	0.0945	0.0975	0.0989	0.0997	0.1013	0.1003

Table 35. Simulated levels of significance for $Y_n^2(\hat{\gamma})$ test for $GEE(\lambda, \alpha\theta\beta)$.

Comme on peut l'observer, la statistique de test proposée ajuste de manière très satisfaisante les différentes distributions. On peut conclure que le test construit peut être utilisé pour valider les distributions de Kumaraswamy, les distributions exponentielles exponentiées (EE) avec différents paramètres, les distributions Kumaraswamy-exponentielles $Kw - Exp$ (Nadarajah et al., 2012) et la distribution ($EKw - exp$) proposée par Rodriguez et al. (2015).

7.6 Exemple de données de fiabilité

Dans cette section, nous appliquons les résultats obtenus dans cette étude, à des données réelles de la fiabilité (Crowder et al., 1991). Dans une expérience, pour obtenir des informations sur la force d'un certain type de cordon tressé après les intempéries, les forces de 48 morceaux de cordon ayant résisté pendant un temps déterminé, ont été étudiées. Les valeurs de force censurées à droite observées sont données ci-dessous :

26.8*, 29.6*, 33.4*, 35*, 36.3, 40*, 41.7, 41.9*, 42.5*, 43.9, 49.9, 50.1, 50.8, 51.9, 52.1, 52.3, 52.3, 52.4, 52.6, 52.7, 53.1, 53.6, 53.6, 53.9, 53.9, 54.1, 54.6, 54.8, 54.8, 55.1, 55.4, 55.9, 56, 56.1, 56.5, 56.9, 57.1, 57.1, 57.3, 57.7, 57.8, 58.1, 58.9, 59, 59.1, 59.6, 60.4, 60.7.

Nous voulons vérifier si ces données peuvent être modélisées par la distribution $EKw - GEE$. Sous l'hypothèse nulle H_0 , en utilisant le logiciel R et la méthode de Newton Raphson, on calcule les estimateurs du maximum de vraisemblance des paramètres inconnus :

$$\hat{\theta} = \left(\hat{\alpha}, \hat{\lambda}, \hat{\beta}, \hat{\sigma}, \hat{\theta} \right)^T = (5.5252, 0.5181, 0.2579, 0.0853, 5.3104)^T.$$

On regroupe ces observations dans $k = 5$ Intervalles I_j et on calcule tous les éléments nécessaires pour déterminer la statistique Y_n^2 :

$\hat{a}_j$	49.987	53.243	54.987	57.125	60.700
U_j	4	10	8	9	10
$\hat{C}_{1j}$	-0.0007	-0.0013	-0.0009	-0.0002	-0.0011
$\hat{C}_{2j}$	1.0324	0.2083	1.6127	1.7439	1.8439
$\hat{C}_{3j}$	-0.0160	-0.0283	-0.0214	-0.0229	-0.0241
$\hat{C}_{4j}$	2.0169	4.7217	3.7487	4.1944	4.6390
$\hat{C}_{5j}$	-0.0083	-0.0124	-0.0089	-0.0090	-0.0088
e_j	7.8575	7.8575	7.8575	7.8575	7.8575

Table 36. valeurs des $\hat{a}_j, e_j, U_j, \hat{C}_{1j}, \hat{C}_{2j}, \hat{C}_{3j}, \hat{C}_{4j}, \hat{C}_{5j}$

On obtient alors la valeur de la statistique Y_n^2 comme suit :

$$Y_n^2 = X^2 + Q = 4.7857 + 2.9779 = 7.7636$$

Pour le niveau de signification $\varepsilon = 0.05$, la valeur de $Y_n^2 = 7.7636 \prec \chi_5^2 = 11.0705$, nous pouvons affirmer que ces données suivent une distribution $EKw - GEE$.

Les statistiques de test Y_n^2 relatives aux sous-modèles de la distribution $EKw - GEE$ sont aussi calculées et présentées dans la table suivante :

<i>Modeling distribution</i>	Y_n^2
$EKw - GEE(\lambda, \alpha, \beta, \sigma, \theta)$	7.7636
$Kw - GEE(\lambda, \beta, \sigma, \theta)$	8.8456
$EKw - E(\lambda, \beta, \sigma, \theta)$	6.9856
$Kw - E(\beta, \sigma, \theta)$	6.5231
$Exp(\lambda)$	7.2365
$Exp(\lambda\theta)$	9.2587
$GEE(\lambda, \alpha)$	7.2365
$GEE(\lambda, \theta)$	7.0256
$GEE(\lambda, \alpha\beta)$	9.7895
$GEE(\lambda\sigma, \theta)$	11.0258
$GEE(\lambda, \alpha\theta\beta)$	12.1256

Table 37. Valeurs des Y_n^2 pour les sous-modèles de la distribution

$$EKw - GEE(\lambda, \alpha, \beta, \sigma, \theta).$$

CHAPITRE 8

Conclusion

La modélisation mathématique est et restera le moyen essentiel d'analyse de tous les phénomènes naturels, industriel, médical, socio-démographique, etc. Le choix de la distribution qui peut décrire convenablement les observations est crucial. Avec les grandes avancées technologiques, les distributions existantes ne sont plus en mesure de modéliser la quantité volumineuse et la diversité des données. De nouveaux modèles sont alors proposés dans la littérature statistique. Basés sur des généralisation de la distribution exponentielle, on a introduit dans ce travail des modèles à durée de vie accélérée (*AFT*) permettant d'obtenir en un temps court, les taux de panne de matériels de plus en plus performants ou les durées de vie en analyse médicale. Après l'étude des caractéristiques statistiques des modèles proposés *AFT-GEE*, *AFT-LE* et *AFT-KwGEE*, et l'estimation des paramètres inconnus, nous construisons des statistiques de tests du type du chi-deux modifié permettant d'ajuster des observations à chacun de ces modèles dans le cas de censure aléatoire droite et de paramètres inconnus. De plus, on a proposé une nouvelle distribution *EKw-GEE* très flexible, basée sur une généralisation de la distribution de Kumaraswamy. Grâce aux paramètres additifs, on peut contrôler l'asymétrie, l'aplatissement et la queue de la distribution. Après l'étude statistique, un critère de test d'adéquation a été aussi mis en place pour ce modèle. Sans oublier pour autant qu'on a fourni des statistiques de tests pour valider le modèle exponentiel exponentié généralisé (*GEE*), le modèle exponentiel linéaire généralisé (*GLE*), la distribu-

tion Kumaraswamy généralisée exponentielle exponentiée ($Kw - GEE$) et la transformée de la distribution exponentielle linéaire généralisée ($TGLE$) et ceci dans le cas de censure aléatoire droite et de paramètres inconnus.

En perspectives, on s'intéresse à de nouvelles distributions obtenues par les nouvelles méthodes de généralisations comme les G-distributions, l'exponentiation des fonctions de répartitions. Aussi, on se propose l'étude des distributions des statistiques de tests classiques, Anderson-Darling, Kolmogorov-Smirnov et Von-Mises pour les modèles étudiés.

- [1] Abd el Monemg. H. et Jaheen Z. F. (2017). Bayesian estimation for a multistep-stress accelerated generalized exponential model with type-ii censored data. J. Math. Comput. Sci. 7, No. 2, 292-306 ISSN : 1927-5307.
- [2] Achcar, J.A., and Moala, F. A., and Boleata, J., (2015). Generalized exponential distribution : A Bayesian approach using MCMC methods, International Journal of Industrial Engineering Computations, Vol. 6, pp.1-14.
- [3] Antonov , A. V..and NikulinM. C.(2012). Statistical Models in Reliability Theory : Textbook (Abris, Moscow).
- [4] Bagdonavičius, V., Nikulin, M. (2002). Accelereted Life Models : Modeling and Statistical Analysis, Chapman and Hall.
- [5] Bagdonavičius, V., Nikulin, M. (2011). Chi-squared tests for general composite hypotheses from censored samples Comptes Rendus de l'académie des Sciences de Paris, Mathématiques, V.349, N.3 – 4, 219 – 223.
- [6] Bagdonavičius, V., Nikulin, M. (2011a). Chi-squared Goodness-of-fit Test for Right Censored Data. International Journal of Applied Mathematics and Statistics, vol. 24, p30 – 50.
- [7] Bagdonavičius, V., Levulienė, R., J., and Nikulin, M., (2013). Chi-squared goodness-of-fit tests for parametric accelerated failure time models, Communications in Statistics-Theory and Methods, Volume : 42 Issue : 15, Pages : 2768 – 2785.

- [8] Bain, L.J. (1978). Statistical Analysis of Reliability and Life Testing Models, Macel Dekker.
- [9] Balakrishnan, N., Chimitova, E., Galanova N. and Vedernikova, M., (2013), Testing Goodness of Fit of Parametric AFT and PH Models with Residuals, communications in statistics, simulation and computation, volume 42, issue 6.
- [10] Broadbent, S., (1958). Simple mortality rates, Journal of Applied Statistics, 7, 86.
- [11] Carbone, P., Kellerthouse, L. and Gehan, E. (1967). Plasmacytic Myeloma : A study of the relationship of survival to various clinical manifestations and anomalous protein type in 112 patients, American Journal of Medicine, 42, 937-948.
- [12] Coolen FPA, Yan KJ. (2004) Non parametric predictive inference with right-censored data J Stat Plan Infer. 126 : 25 – 54.
- [13] Cordeiro G.M., Castro M., (2011) A new family of generalized distributions, Journal of Statistical Computation and Simulation, vol. 81, n 7;pp. 883-898.
- [14] Chouia, S. and Seddik-Ameur, N. (2017), A modified chi-square test for Bertholon model with censored data, communications in statistics-simulation and computation, Volume 46, Issue 1, 593 – 602.
- [15] Cox, D.R., and Oakes, D., (1984), Analysis of Survival Data, Monographs on Statistics and Applied Probabilities, Ed. Chapman & Hall/CRC.
- [16] Crowder M.J., Kimber A.C., Smith R.L& T.J Sweeting, (1991) Statistical analysis of reliability data, CHAPMAN & HALL/CRC.
- [17] Elbatal I., Diab L.S., Abdul Alim N.A., (2013), Transmuted Generalized Linear Exponential Distribution, International Journal of Computer Applications (0975 8887), Volume 83 - No. 17.
- [18] Francisco M. C. Medeiros¹, Antonio H. M. da Silva-J´unior¹, Dione M. Valenc,a¹ & Silvia L. P. Ferrari²,(2014) Testing Inference in Accelerated Failure Time Models; International Journal of Statistics and Probability; Vol. 3, No. 2.

- [19] Freireich, E. J. Gehan, E. Frei, E. Schroeder, L. R. Wolman, I. J. Anbari, R. Burgert, E. O. Mills, S. D. Pinkel, D. Selawry, O. S. et al. (1963) The effect of 6-mercaptopurine on the duration of steroid-induced remissions in acute leukemia : A model for evaluation of other potentially useful therapy. *Blood*, 21(6) : 699 – 716.
- [20] Galanova N. Lemesko S., B. Yu., and Chimitova E. V., (2012), Using Nonparametric Goodness-of-Fit Tests to Validate Accelerated Failure Time Models, *Optoelectronics, Instrumentation and Data Processing*, Vol. 48, No. 6, pp. 580–592.
- [21] Goual, H. and Seddik Ameer, N. (2014) Chi-squared type test for the AFT-generalized inverse Weibull distribution. *Communication in Statistics-Theory and Method*, 43, 13, 2605 – 2617.
- [22] Gupta, A., Mukherjee, B., Upadhyay, S.K. (2008) Weibull extension model : A Bayes study using Markov chain Monte Carlo simulation . *Reliability Engineering & System Safety* ,93, 1434 – 1443.
- [23] Gupta, R.D., and Kundu, D. (1999), Generalized exponential distributions. *Austr.New Zealand J. Statist.* 41, 173{188}.
- [24] Gupta, R.D., and Kundu D. (2007a), Generalized exponential distribution : existing results and some recent developments, *J. Statist. Plann. Inference*, 137(2007), 3537-3547.
- [25] Gupta, R.D., and Kundu D. (2007b), A convenient way of generating gamma random variables using generalized exponential distribution. *Comput. Statist. Data Anal.* 51, 2796{2802.
- [26] Habib, M.G., Thomas, D.R. (1986) Chi-squared Goodness-of-Fit Tests For Randomly Censored Data, *Annals of Statistics*, V.14, N 2, p 759 – 765.
- [27] Hassan, A. S., (2005), Goodness-of- t tests for the generalized exponential distribution ; *Inter. Stat. Electronic Journal*, pp. 1-15.
- [28] Hjort, N.L. (1990). Goodness of t tests in models for life history data based on cumulative hazard rates. *The Annals of Statistics*, 18, 1221-1258.

- [29] Hollander, M., Pena, E. (1992), Chi-square goodness-of-fit test for randomly censored data. *JASA* 87(417) : 458–463.
- [30] Hougaard, P. (1986) Survival models for heterogeneous populations derived from stable distributions, *Biometrika* 73, 387 – 396.
- [31] Huang S., and Oluyede BO., (2014). Exponentiated Kumaraswamy-Dagum distribution with applications to income and lifetime data, *Journal of statistical distributions and Applications*, 2014,1,1-18.
- [32] Kalbeisch J. Dand R. L. Prentice,(2002). *The Statistical Analysis of Failure Time Data* (John Wiley and Sons, New York).
- [33] Klein J. P and M. L. Moeschberger, (2003). *Survival Analysis : Techniques for Censored and Truncated Data* (Springer,New York).
- [34] Kumaraswamy P., (1980). A generalized probability density for double-bounded random processes, *Journal of Hydrology*, 46 : 79-88.
- [35] Lemonte A.J., Souza W.B., Cordeiro G.M,(2013) The exponentiated Kumaraswamy distribution and its log-transform, *Brazilian Journal of Probability and Statistics*, 27 : 31-53.
- [36] Mahmouda M.A.W., Farouq Mohammad A. Alamb, (2013). Exponentiated generalized linear exponential distribution, *Applied Mathematical Modelling* Volume 37, Issue 5, 1 March 2013, Pages 2838–2849.
- [37] Mederios, F.M.C, da Silva-Junior, A.H.M., Valen ca, D.M., and Ferrari, S.L.P.(2014), Testing inference in Accelerated Failure Time Models, *International Journal of Statistics and Probability*, Vol. 3, No 2.
- [38] Mohammed, B.E., (2014), Statistical properties of Kumaraswamy-Generalized Exponentiated exponential distribution, *International Journal of Computer Applications*, Volume 94, N 4.

- [39] Muhammad Yameen Danish and Muhammad Aslam (2013) Bayesian Analysis of Randomly Censored Generalized Exponential Distribution. Austrian journal of statistics, Volume 42 , Number 1, 47–62
- [40] Nelson, W., & Hahn, G. J. (1972). Linear Estimation of a Regression Relationship from Censored Data Part I—Simple Methods and Their Application. *Technometrics*, 14(2), 247-269.
- [41] Nelson, W.(1990) Accelerated Testing : Statistical Models, Test Plans, and Data Analysis. Wiley, New York
- [42] Nikulin, M.S. and Tran ; X. Q., (2013), Chi-squared Goodness-of-ft test for Generalized Birnbaum Saunders Models for right censored data and its reliability Applications, RT&A # 02 (29), Vol.8.
- [43] Orbe, J., Ferreira, E., & Nunez-Anton, V. (2002). Comparing proportional hazards and accelerated failure time models for survival analysis. *Statistics in Medicine*, 21, 3493-3510. [http ://dx.doi.org/10.1002/sim.1251](http://dx.doi.org/10.1002/sim.1251)
- [44] Ortega, E. M. M., Cancho, V. G. and Paula G. A. (2009). Generalized log-gamma regression models with cure fraction. *Lifetime Data Analysis*. 15, 79 – 106.
- [45] Pandey, A., Singh, A. and Zimmer, W.J.(1993). Bayes estimation of the linear hazard-rate model, *IEEE Transactions on Reliability*, 42, 636-640.
- [46] Pike M. C. , (1966). A method of analysis of a certain class of experiments in carcinogenesis, *Biometrics* 22, 142-161.
- [47] Rao G. S. (2012), Estimation of reliability in multicomponent stress-strength based on generalized exponential distribution, *Revista Colombiana de Estadística*, vol. 35, pp 67-76.
- [48] Rao, K. C., Robson, D. S., (1974) A Chi-square statistic for goodness-of-fit tests within the exponential family. *Communication in Statistics*, 3, 1139 – 1153.

- [49] Ravi, V., Gilbert, P. D. (2009) BB : An R package for solving a large system of nonlinear equations and for optimizing a high-dimensional nonlinear objective function. *J. Statist. Software*, 32(4).
- [50] Rinne, H. (2008) *The Weibull Distribution : A Handbook*. Taylor & Francis.
- [51] Rodriguez J.A., and Silva, A.P.C.M, The exponentiated Kumaraswamy- exponential distribution, *British Journal of Applied Science & Technology*, 10(5) :1-12, (2015).
- [52] Santos, P. B., & Valenc,a, D. M.(2012). Accelerated failure time models with random effects to estimate operation time of oil wells. *Advances and Applications in Statistics*, 28, 23-44.
- [53] Terrell, G. R. (2002). The gradient statistic. *Computing Science and Statistics*, 34, 206-215.
- [54] Voinov, V., Nikulin, M., Balakrishnan, N. (2013) *Chi-Squared Goodness of Fit Tests with Applications*, Academic Press, Elsevier.
- [55] Valiollahi, R., Asgharzadeh, A., & Kundu, D. (2017). Pr vision des d faillances futures pour la distribution exponentielle g n ralis e sous-type I ou de type II censures hybride. *Journal br silien de Probabilit s et statistiques* , 31 (1), 41-61.
- [56] Wei, L. J.(1992). The accelerated failure time model : a useful alternative to the Cox regression model in survival analysis. *Statistics in Medicine*, 11, 1871-1879.
- [57] Woo Dong Lee, Dal Ho Kim et Sang Gil Kang (2017) Objective Bayesian multiple hypothesis testing for the shape parameter of generalized exponential distribution. *Journal of the Korean Data & Information Science Society* 2017 ;28 :217-25