

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 –

Faculté des Sciences
Département de Mathématiques
Laboratoire LaPS



THÈSE

Présentée en vue de l'obtention du diplôme de
DOCTORAT EN MATHÉMATIQUES
Option : Modélisation Mathématique

*Sur l'étude des modèles paramétriques PH et
les modèles AFT correspondants*

Présentée par
Mlle TREIDI WAFA

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

Devant le jury

PRÉSIDENT : CHADLI Assia PROF. U.B.M. ANNABA

EXAMINATEUR : DJELLAB Natalia PROF. U.B.M. ANNABA

EXAMINATEUR : SADKI Ourida PROF. U.S.T.H.B. ALGER

Remerciements

Je tiens avant tout à remercier Allah pour la force et la volonté qu'il m'a données pour pouvoir achever ce travail.

J'exprime toute ma reconnaissance à ma directrice de thèse le professeur Seddik ameur Nacira, pour son aide, son soutien, ses conseils ainsi que la confiance qu'elle m'a faite en acceptant de m'encadrer.

Je tiens également à remercier professeur.Chadli Assia qui m'a fait l'honneur de présider le jury ainsi que Mme.Djellab Natalia, Mme.Sadki Ourida pour avoir accepté de faire partie du jury et d'y avoir consacré une partie de leurs temps.

Mes sincères remerciements vont à mes très chères amies Fairouz, Soumaya qui m'ont accompagnée et soutenue moralement tout au long de mon parcours.

Je remercie également mes frères, mes soeurs pour leur soutien et contribution à la réalisation de ce travail. Je remercie aussi d'autres personnes qui m'ont encouragée à finir ce travail par des gestes d'amitié dont je suis reconnaissante.

Enfin, les mots les plus simples étant les plus forts, j'adresse toute mon affection à ma famille, et je ne saurais oublier l'apport de ma mère pour l'accomplissement de ce travail, je tiens à leur rendre hommage à travers cette thèse. Merci pour avoir fait de moi ce que je suis aujourd'hui.

Une pensée pour terminer ces remerciements pour mon père qui n'a pas vu mon travail mais je sais que tu en aurais été très fier de ta fille!!!

TABLE DES MATIÈRES

1	Présentation du modèle Extension de Weibull (EW)	10
1.1	Fonction de répartition et densité	11
1.2	La fonction de survie et le taux de hasard	12
1.3	Fonction de hasard cumulé	14
1.4	L'espérance et la variance	14
1.5	Estimation du maximum de vraisemblance en cas de données complètes . . .	15
1.5.1	Estimation du maximum de vraisemblance	15
1.5.2	Simulations cas de données complètes	16
1.6	Estimation du maximum de vraisemblance en cas de données censurées . . .	16
1.6.1	Estimation du maximum de vraisemblance	16
1.6.2	Simulations cas de données censurées	18
2	Tests d'ajustement	19
2.1	Introduction	21
2.2	Nikulin-Rao-Robson test pour données complètes	22
2.3	Nikulin-Rao-Robson statistique pour le modèle <i>Extension de Weibull</i>	23
2.3.1	Calcul de la matrice d'information des données groupées $J(\theta)$	24
2.3.2	Calcul de la matrice d'information de Fisher $I(\theta)$	25
2.3.3	Simulation (NRR statistique Y_n^2)	25

TABLE DES MATIÈRES

2.3.4	Application à des données réelles	26
2.4	Tests d'ajustement classiques	27
2.4.1	Test de Kolmogorov-Smirnov	27
2.4.2	Test d'Anderson-Darling	28
2.4.3	Test de Cramer-Von Mises	28
2.5	Tests d'ajustement pour données censurées	28
2.6	Test d'ajustement en présence de censure pour la distribution EW	33
2.6.1	Calcul de la matrice \hat{W}	34
2.6.2	Calcul de la matrice \hat{I}	34
2.6.3	Simulation de la distribution de la statistique Y_n^2	35
2.6.4	Simulation des valeurs du critère de test	36
2.6.5	Analyse de données de survie	37
3	Construction du modèle EW-AFT	40
3.1	Introduction	41
3.2	Le modèle AFT	41
3.3	Construction du modèle EW-AFT	43
3.4	Estimation du maximum de vraisemblance en cas de données censurées . . .	45
3.4.1	Calcul des estimateurs du maximum de vraisemblance	45
3.4.2	Simulations	46
3.5	Test d'ajustement pour le modèle EW-AFT en cas de données censurées . .	47
3.5.1	Choix des intervalles	48
3.5.2	Calcul de la matrice \hat{W}	48
3.5.3	Calcul de la matrice \hat{I}	49
3.5.4	Simulation (NRR statistique Y_n^2)	51
3.5.5	Analyse de données de fiabilité	52
4	Modèle à hasard proportionnel (PH)	55
4.1	Introduction	56

TABLE DES MATIÈRES

4.2	Principe du modèle	56
4.3	Construction du modèle EW-PH	57
4.4	Estimation du maximum de vraisemblance en cas de données censurées . . .	59
4.4.1	estimateurs du maximum de vraisemblance	59
4.4.2	Simulations	61
4.5	Test d'ajustement pour le modèle EW-PH en cas de données censurées . . .	62
4.5.1	Choix des intervalles	62
4.5.2	Calcul de la matrice \widehat{W}	63
4.5.3	Calcul de la matrice \widehat{I}	63
4.5.4	Simulation de la distribution de Y_n^2	65
4.5.5	Simulation de la statistique Y_n^2	66
4.5.6	Application à des données de survie	66

Cette thèse traite le problème de la modélisation des données de durées de vie en analyse statistique. Elle est consacrée particulièrement à une généralisation de la distribution classique de Weibull introduite par Xie et al. (2002) appelée "*modèle Extension de Weibull (EW)*". Nous proposons, dans un premier temps, la construction de tests d'ajustement du type du chi-deux modifié basés sur la statistique NRR pour ce modèle dans le cas de données complètes et de données censurées à droite. Dans un deuxième temps, on construit un modèle à temps de vie accéléré dont la distribution de base est une Extension de Weibull (EW-AFT). Ensuite on introduit un modèle à taux de hasards proportionnels (EW-PH) dont la fonction du taux de hasard est celle d'une distribution Extension de Weibull. Après l'étude des caractéristiques statistiques de ces distributions, on construit des tests d'ajustement du type du chi-deux modifié pour ces deux modèles dans le cas où les paramètres sont inconnus et les données censurées à droite. Une importante étude par simulations numériques et des applications réelles ont été proposées pour confirmer les résultats obtenus dans ce travail.

Mots-clés : Analyse de survie, Données censurées, Modèle extension de Weibull, Modèle de temps de défaillance accéléré, Modèle de Cox, Statistique NRR, Tests d'ajustement du type du Chi-deux..

This thesis deals with the problem of the modeling of lifetimes data in statistical analysis. It is devoted particularly to a generalization of the classical Weibull distribution, introduced by Xie et al. (2002), and so-called the Weibull Extension Model (EW). First, we propose the construction of modified chi-squared type goodness-of-fit tests based on the NRR statistic for this model in both of complete and right censored data. Secondly, we construct an accelerated lifetime model with this distribution as baseline (EW-AFT). Then, we introduce a proportional hazards model (EW-PH) with a Weibull Extension rate baseline function. After studying the statistical characteristics of these distributions, we construct modified chi-squared goodness-of-fit tests for these models when the parameters are unknown and data are right censored. An important study by numerical simulations and applications to real data have been proposed to confirm the results obtained in this work.

Keywords : Accelerated failure time model, Censored data, Chi-squared type goodness-of-fit test, Extension Weibull distribution, Hazard ratio (**HR**), NRR statistic, Survival analysis.

تتناول هذه الأطروحة مشكلة نمذجة عمر من البيانات في التحليل الإحصائي . و هو مكرس بشكل خاص لتعميم توزيع واييل الكلاسيكي الذي وضعه (Xie et al. (2002 , بعنوان " modèle Extension de Weibull (EW) " . نقترح في البداية ، اختبار الضبط المعدل كما يستند على إحصائية NRR لهذا النموذج في حالة البيانات المراقبة و البيانات الكاملة. ثانياً, نقوم ببناء نموذج وقت الإخفاق المتسارع على (modèle Extension de Weibull (EW-AFT) .

ثم قدم نمودجا لتقييم مخاطر النسبية (EW-PH) ، وظيفة معدل الخطورة التي لتوزيع تمديد ييل. بعد دراسة الخصائص الإحصائية لهذه التوزيعات , نبني اختبار الضبط المعدل كما لهذين النموذجين في حال المعلمات غير معروفة والبيانات المراقبة على اليمين. وقد اقترحت دراسة رئيسية عن طريق المحاكاة العددية وتطبيقات حقيقية لتأكيد النتائج التي تم الحصول عليها في هذا العمل.

كلمات الرئيسية : إحصائية NRR , اختبار الضبط المعدل , البيانات المراقبة , تحليل البقاء على قيد الحياة , extension de Weibull , نموذج كوكس , نموذج وقت الإخفاق المتسارع .

Résoudre les problèmes auxquels on est confronté, est une activité qui a une histoire aussi ancienne que la race humaine. De nombreux types de modèles ont été utilisés pour y faire face. Il s'agit notamment des modèles physiques, des modèles descriptifs et des modèles mathématiques. L'utilisation de modèles mathématiques est relativement récente. Initialement, les modèles mathématiques ont été utilisés pour résoudre des problèmes des sciences physiques (par exemple, prédire le mouvement des planètes, le calendrier des marées hautes et basses), mais ces deux derniers siècles, les modèles mathématiques ont été largement utilisés pour résoudre les problèmes biologiques (modéliser les informations disponibles sur des protéines, sur des gènes, l'étude de l'ADN,...), en sciences sociales (les enquêtes d'emploi, la dynamique des populations,...), en ingénierie pour l'étude de la fiabilité des composants des systèmes, en analyse de survie pour l'étude des effets des schémas thérapeutiques sur les patients ou n'importe quel événement ayant un effet sur la santé des patients. Les modèles mathématiques interviennent dans toutes les disciplines et tous les secteurs.

Un modèle paramétrique approprié est souvent intéressant dans l'analyse des données de survie, car il fournit un aperçu des caractéristiques des temps de défaillance et des fonctions de risque.

La distribution de Weibull (1951) est l'une des familles les plus utilisées pour la modélisation de telles données. Cependant ce modèle à deux paramètres est inadéquat lorsque la forme du risque est unimodale ou en baignoire.

Diverses techniques paramétriques ont été développées pour intégrer une grande variété de modèles dans l'analyse de survie. Certains des modèles paramétriques proposés ont introduit un paramètre de forme dans la distribution classique de Weibull pour tenir compte des différentes formes de la fonction du taux de hasard. L'un des premiers modèles de ce type a été proposé par Kalbfleisch (Kalbfleisch et Prentice 1980). Mudholkar et Srivastava (1993) ont introduit une version exponentielle du modèle de Weibull qui comprend un paramètre de forme supplémentaire. Ces modèles flexibles sont capables de générer une grande variété de formes de risque fréquemment observées, y compris unimodale et en baignoire. Mudholkar, Srivastava et Kollia (1996) ont proposé une autre généralisation du modèle de Weibull, qui est capable de générer des types similaires de formes de risque comme le modèle exponentié, mais le support de la distribution devient dépendant de l'espace des paramètres. Une généralisation à quatre paramètres de la distribution de Weibull a été introduite par Jeong (2006) à partir des distributions stables proposées par Hougaard (1986), induites par le modèle de fragilité semi-paramétrique. Ce modèle peut aussi être considéré comme une généralisation du modèle de Mudholkar et al. (1996).

Une monographie de Murthy et al. (2004) et le manuel de Rinne (2008) contiennent presque toutes les généralisations concernant la distribution weibull. Lai et al. (2011) fournissent également une vue d'ensemble sur le sujet. Gusmão et al. (2009) ont proposé une distribution Weibull inverse généralisée en ajoutant un autre paramètre γ à la distribution Weibull inverse standard. Bebbington et al. (2007a) ont présenté une distribution Weibull-Flexible simple et très utilisée dans l'étude de fiabilité des pompes de réacteurs secondaires (Bebbington et al. 2007a), et dans l'étude de la mortalité humaine (Bebbington et al. 2007b). La littérature statistique montre que les généralisations de la distribution de Weibull sont appliquées aux données de cancer de la tête et du cou, le cancer du sein, les taux de défaillance de composants mécaniques, pannes accidentelles des systèmes parmi tant d'autres données de survie.

Parmi ces nouvelles distributions, nous avons considéré une généralisation proposée par Xie et al.(2002) appelé "modèle Extension de Weibull *EW*". Cette distribution peut être

considérée comme une généralisation du modèle proposé par Chen (2000) ainsi que celle de la distribution puissance exponentielle étudiée par Smith et Bain (1975). Dans leurs travaux, Tang et al. (2004) ont étudié les propriétés statistiques de la distribution EW, tandis que Gupta et al. (2008) ont utilisé les méthodes de Monte Carlo pour une analyse bayésienne et des tests graphiques pour montrer l'adéquation de ce modèle à des données réelles. Ils ont montré que cette distribution à trois paramètres dont deux sont des paramètres de forme, permet de caractériser beaucoup plus de durées de vie que la distribution classique de Weibull et ses domaines d'application sont très nombreux.

En fiabilité, dans les tests de vie accélérés (*ALT*), pour mesurer les taux de défaillance en un temps court, on met le matériel sous des conditions appelées stress pour accélérer le processus provoquant les défaillances, les modèles utilisés sont dits modèles à temps de vie accéléré *AFT* (accelerated failure time). Ces stress sont représentés par des covariables qui interviennent sur les fonctions d'intérêt du modèle. Les modèles *AFT* sont proposés en alternative aux modèles à risques proportionnels *PH* (proportional hazards) de Cox (1972), largement répandus en analyse de survie, particulièrement en épidémiologie et aussi en industrie pour les tests de vie accélérés (*ALT*). Dans ces modèles, les fonctions d'intérêt dépendent aussi de covariables qui peuvent influencer la survie. Les modèles *PH* exigent la proportionnalité des taux de hasard des individus, ce qui n'est pas le cas dans les modèles *AFT*.

Dans ce travail, nous introduisons dans un premier temps un modèle à temps de vie accéléré (*EW – AFT*) dont la distribution de base est une Extension de Weibull. Ensuite on construit un modèle à taux proportionnels (*EW – PH*) dont la fonction du taux de hasard est celle d'une distribution Extension de Weibull. Après l'étude des caractéristiques statistiques de ces distributions, nous construisons des tests d'ajustement du type du chi-deux modifié pour ces modèles dans le cas où les paramètres sont inconnus et les données censurées à droite.

Tout d'abord, on se propose de construire un test d'ajustement du type du chi-deux modifié pour le modèle "extension de Weibull" dans le cas où les données sont complètes et

les paramètres inconnus. Ce test est basé sur la statistique de Nikulin-Rao-Robson (*NRR*) proposée séparément par Nikulin(1973) et Rao et Robson (1974). Basée sur l'estimateur du maximum de vraisemblance sur les données initiales, cette statistique Y^2 est une modification naturelle du test classique X^2 du chi-deux.

Ensuite, on développe un autre test d'ajustement pour ce modèle dans le cas où les paramètres sont inconnus et les données censurées à droite. Nous utilisons l'approche proposée par Bagdonavicius et Nikulin (2011), basée sur les estimateurs du maximum de vraisemblance sur les données non groupées. Ce test du type du chi-deux modifié Y_2^2 est une modification de la statistique *NRR* qui tient compte aussi bien des paramètres inconnus que de la censure. Aussi, nous construisons des tests d'ajustement du type du chi-deux modifié basés sur cette statistique pour les modèles proposés (*EW – AFT*) et (*EW – PH*). Nous calculons les estimateurs du maximum de vraisemblance en cas de censure droite et tous les éléments constituant les critères de test. Nous menons une étude par simulation numérique très importante. On termine par des applications sur des données réelles.

Le manuscrit est composé de quatre chapitres.

On débute par une étude détaillée du modèle extension de Weibull (**EW**) proposé par Xie et al. (2002). Les propriétés mathématiques, les caractéristiques du modèle et l'estimation des paramètres inconnus dans le cas de données complètes et le cas de données censurées sont présentées.

Dans le deuxième chapitre, après un historique sur les tests du chi-deux, on développe la statistique Y^2 de Nikulin-Rao-Robson (*NRR*) et sa modification Y_2^2 proposée Bagdonavicius et Nikulin (2011) pour données censurées à droite. Ensuite, on construit des tests d'ajustement du type du chi-deux modifié, basé sur la statistique *NRR* et sur sa modification pour le modèle extension de Weibull dans le cas de données complètes et de données censurées.

Dans le troisième chapitre, nous proposons la construction d'un modèle de durée de vie à temps de vie accéléré dont la fonction de base est une extension de Weibull qu'on appelle (*EW – AFT*). Nous présentons une étude du modèle, fonction de fiabilité, fonction risque selon ses différentes formes. Nous utilisons la méthode du maximum de vraisemblance sur les

données initiales pour estimer les paramètres inconnus du modèle, en tenant compte de la censure. Ensuite on construit pour celui-ci, un test d'ajustement du type du chi-deux modifié dans le cas de censure droite. Une étude par simulations numériques a été proposée pour montrer la maniabilité du test proposé.

Quand au dernier chapitre, il est consacré à la construction d'un modèle PH à risques proportionnels, dont la distribution de base est une extension de Weibull, noté ($EW - PH$). Comme pour les autres modèles, on commence par la présentation des propriétés statistiques et l'estimation des paramètres inconnus par la méthode du maximum de vraisemblance sur les données initiales en présence de censure droite, ensuite on construit un test d'adéquation pour ajuster une série d'observations à ce type de modèle. On termine par des simulations numériques et des applications à des données réelles qui confirment les résultats obtenus.

CHAPITRE 1

Présentation du modèle Extension de Weibull (EW)

Sommaire

1.1	Fonction de répartition et densité	11
1.2	La fonction de survie et le taux de hasard	12
1.3	Fonction de hasard cumulé	14
1.4	L'espérance et la variance	14
1.5	Estimation du maximum de vraisemblance en cas de données complètes	15
1.5.1	Estimation du maximum de vraisemblance	15
1.5.2	Simulations cas de données complètes	16
1.6	Estimation du maximum de vraisemblance en cas de données censurées	16
1.6.1	Estimation du maximum de vraisemblance	16
1.6.2	Simulations cas de données censurées	18

Parmi les généralisations de la distribution de Weibull, on s'intéresse à celle proposée par Xie et al.(2002) appelée "modèle Extension de Weibull *EW*". Cette distribution peut être considérée comme une généralisation du modèle proposé par Chen (2000) ainsi que celle de la distribution puissance exponentielle étudiée par Smith et Bain (1975). Le modèle *EW* à trois paramètres dont deux sont des paramètres de forme, permet de caractériser beaucoup plus de durées de vie que la distribution classique de Weibull et ses domaines d'application sont nombreux et divers. La forme de la fonction du taux de hasard peut être croissante ou en baignoire. Dans leurs travaux, Tang et al. (2004) ont étudié les propriétés statistiques de cette distribution, tandis que Gupta et al. (2008) ont utilisé les méthodes de Monte Carlo pour une analyse bayésienne et des tests graphiques pour montrer l'adéquation de ce modèle à des données réelles. Cependant les tests d'ajustement n'ont pas été suffisamment développés dans la littérature statistique. Dans ce chapitre, nous commençons par présenter la distribution extension de Weibull et ses différentes formes de fonction de risque, ensuite on calcule les estimateurs du maximum de vraisemblance des paramètres inconnus à partir des données initiales dans le cas complet et en présence de censure aléatoire droite.

1.1 Fonction de répartition et densité

La fonction de répartition du modèle extension de Weibull notée *EW* , caractérisée par $\alpha > 0$ le paramètre d'échelle et $\gamma, \lambda > 0$ les paramètres de forme, est donnée par :

$$F(t) = 1 - \exp \left\{ -\lambda \alpha \left(e^{\left(\frac{t}{\alpha}\right)^\gamma} - 1 \right) \right\} \quad (1.1)$$

La fonction de densité de probabilité est

$$f(t) = \lambda \gamma \left(\frac{t}{\alpha}\right)^{\gamma-1} \exp \left\{ \left(\frac{t}{\alpha}\right)^\gamma + \lambda \alpha \left(1 - e^{\left(\frac{t}{\alpha}\right)^\gamma}\right) \right\} \quad (1.2)$$

Pour simuler une variable provenant de la distribution *EW* , nous utilisons l'équation :

$$t = \left[\alpha^\gamma \ln \left(1 - \frac{\ln(1-u)}{\lambda \alpha} \right) \right]^{\frac{1}{\gamma}},$$

où u suit la distribution uniforme $U(0, 1)$.

Quelques exemples de la densité EW sont présentés respectivement dans les figures (1.1) et (1.2).

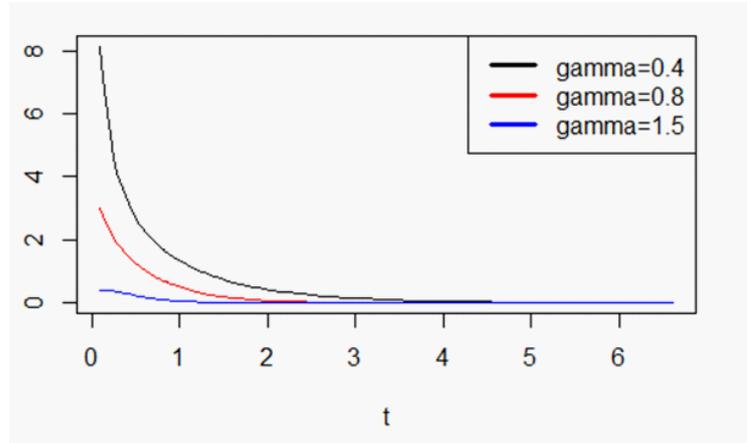


Fig. 1.1. Densité de la distribution EW pour le paramètre de forme $\gamma(0.4; 0.8; 1.5)$ avec $\alpha = 3.5$ et $\lambda = 2.$

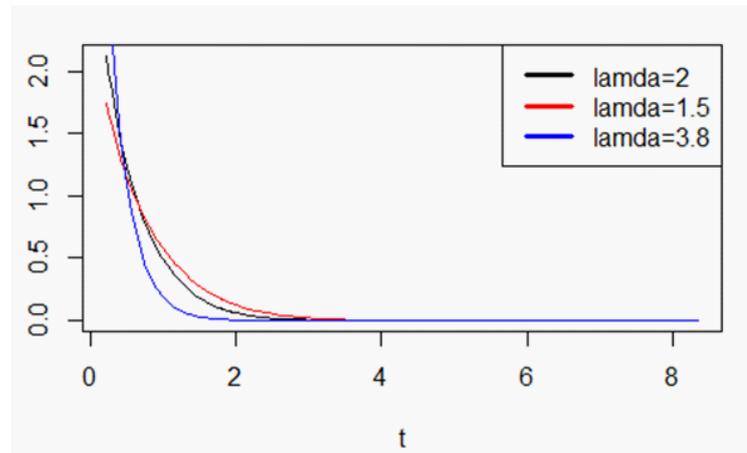


Fig. 1.2. Densité de la distribution EW pour le paramètre de forme $\gamma = 0.8$ avec $\alpha = 3.5$ et différentes valeurs de λ .

1.2 La fonction de survie et le taux de hasard

La fonction de survie du modèle l'extension de Weibull est donnée par

$$S(t) = \exp \left\{ -\lambda \alpha \left(e^{\left(\frac{t}{\alpha}\right)^\gamma} - 1 \right) \right\} \quad (1.3)$$

Sa fonction de hazard est donnée par :

$$h(t) = \lambda \gamma \left(\frac{t}{\alpha}\right)^{\gamma-1} \exp \left\{ -\left(\frac{t}{\alpha}\right)^\gamma \right\} \quad (1.4)$$

Des graphes du taux de hasard de cette distribution EW sont présentés respectivement dans les figures (1.3) et (1.4).

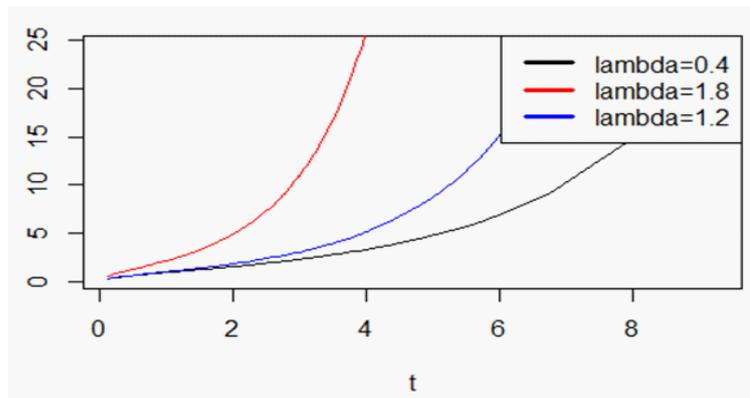


Fig. 1.3. Taux de hasard pour $\gamma = 0.8$ avec différentes valeurs de α et λ .

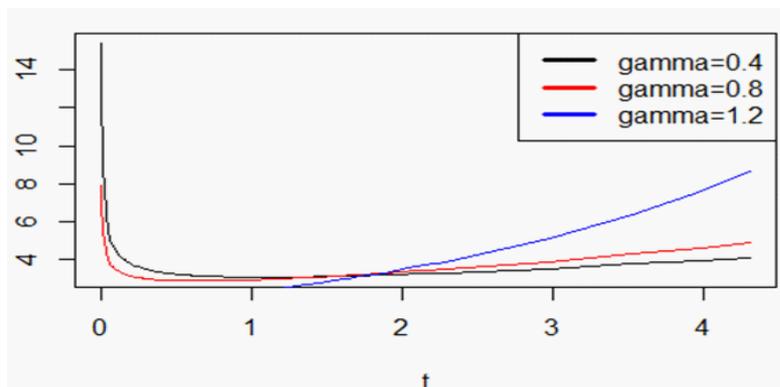


Fig. 1.4. Taux de hasard pour les valeurs de $\gamma(0.4; 0.8; 1.2)$

avec $\alpha = 3.5$ et $\lambda = 2$.

Nous remarquons que :

- Si $\gamma \geq 1$, le taux de panne de la distribution EW est croissant.
- Si $0 < \gamma < 1$, le taux de panne de la distribution EW est décroissant.

Ces différentes formes lui permettent d'être utilisée dans des applications de fiabilité et de l'analyse de survie.

1.3 Fonction de hasard cumulé

La fonction de hasard cumulé du modèle extension de weibull EW se calcule facilement en utilisant (1.3),

$$\begin{aligned} H(t) &= -\ln\{S(t)\} \\ H(t) &= -\ln\{\exp\{-\lambda\alpha(e^{\frac{t}{\alpha}})^\gamma - 1\}\} \end{aligned}$$

Donc

$$H(t) = \lambda\alpha(e^{\frac{t}{\alpha}})^\gamma - 1 \quad (1.5)$$

1.4 L'espérance et la variance

L'espérance est définie comme

$$E(T) = \int_0^{+\infty} t dF(t) = \int_0^{+\infty} S(t) dt = \int_0^{+\infty} \exp\{-\lambda\alpha(e^{\frac{t}{\alpha}})^\gamma - 1\} dt$$

Comme pour la distribution de Weibull, le calcul ci-dessus nécessite l'intégration numérique.

La variance peut être obtenue par les formules suivantes :

$$\begin{aligned} V(T) &= E(T^2) - E(T)^2 \\ V(T) &= \int_0^{+\infty} t^2 dF(t) - E(T)^2 = 2 \int_0^{+\infty} tS(t) dt - E(T)^2 \\ V(T) &= 2 \int_0^{+\infty} t \exp\{-\lambda\alpha(e^{\frac{t}{\alpha}})^\gamma - 1\} dt - E(T)^2 \end{aligned}$$

1.5 Estimation du maximum de vraisemblance en cas de données complètes

1.5.1 Estimation du maximum de vraisemblance

Soit une variable aléatoire $t \sim EW(\theta)$ de vecteur de paramètres $\theta = (\alpha, \gamma, \lambda)$. Afin de calculer les estimateurs du maximum de vraisemblance des paramètres inconnus de cette distribution, on a besoin de la fonction de vraisemblance donnée par l'équation suivante :

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

$$L(t, \theta) = \prod_{i=1}^n \left[\lambda \gamma \left(\frac{t}{\alpha} \right)^{\gamma-1} \exp \left\{ \left(\frac{t}{\alpha} \right)^{\gamma} + \lambda \alpha (1 - e^{(\frac{t}{\alpha})^{\gamma}}) \right\} \right]$$

Et l'expression de la fonction log-vraisemblance est :

$$l(t, \theta) = \sum_{i=1}^n \ln f(t_i, \theta)$$

$$l(t, \theta) = \sum_{i=1}^n \ln \left[\lambda \gamma \left(\frac{t}{\alpha} \right)^{\gamma-1} \exp \left\{ \left(\frac{t}{\alpha} \right)^{\gamma} + \lambda \alpha (1 - e^{(\frac{t}{\alpha})^{\gamma}}) \right\} \right]$$

Pour notre distribution EW, les fonctions de score sont obtenues comme suit :

$$\frac{\partial l(\theta)}{\partial \alpha} = \frac{-n\gamma}{\alpha} + \frac{n}{\alpha} + n\lambda + \sum_{i=1}^n \left[\lambda \gamma \left(\frac{t_i}{\alpha} \right)^{\gamma} e^{(\frac{t_i}{\alpha})^{\gamma}} - \lambda e^{(\frac{t_i}{\alpha})^{\gamma}} - \frac{\gamma (\frac{t_i}{\alpha})^{\gamma}}{\alpha} \right] = 0$$

$$\frac{\partial l(\theta)}{\partial \gamma} = \frac{n}{\gamma} + \sum_{i=1}^n \left[-\lambda \alpha \left(\frac{t_i}{\alpha} \right)^{\gamma} \ln \left(\frac{t_i}{\alpha} \right) e^{(\frac{t_i}{\alpha})^{\gamma}} + \left(\frac{t_i}{\alpha} \right)^{\gamma} \ln \left(\frac{t_i}{\alpha} \right) + \ln \left(\frac{t_i}{\alpha} \right) \right] = 0$$

$$\frac{\partial l(\theta)}{\partial \lambda} = \frac{n}{\lambda} + n\alpha + \sum_{i=1}^n \left[-\alpha e^{(\frac{t_i}{\alpha})^{\gamma}} \right] = 0$$

Puisque, nous ne pouvons pas trouver les formules explicites des estimateurs du maximum de vraisemblance des paramètres, nous utilisons alors des méthodes numériques telles que la méthode de *Newton Raphson*, la méthode de *Monte Carlo*, l'*algorithme BB* ou autres.

1.5.2 Simulations cas de données complètes

On considère le modèle *EW*. Les données ont été simulées $N = 10,000$ fois (avec les tailles des échantillons $n = 30; n = 50; n = 100; n = 250; n = 500$) et les valeurs des paramètres $\alpha = 0.8, \gamma = 2.5, \lambda = 3.5$. 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 $\hat{\alpha}, \hat{\gamma}, \hat{\lambda}$ des paramètres et leurs erreurs quadratiques moyennes (notées *MSE*), nous obtenons les résultats présentés dans la table 1.

$N = 10,000$	$n = 30$	$n = 50$	$n = 100$	$n = 250$	$n = 500$
$\hat{\alpha}$	0.8087	0.8210	0.7916	0.7939	0.7922
<i>MSE</i>	$7.59.e^{-5}$	0.0004	$6.99.e^{-5}$	$3.60.e^{-05}$	$6.03.e^{-05}$
$\hat{\gamma}$	2.4964	2.5341	2.4730	2.4770	2.5164
<i>MSE</i>	$1.25.e^{-5}$	0.0011	0.00072	0.0005	0.0002
$\hat{\lambda}$	3.7150	3.5069	3.5906	3.5165	3.5018
<i>MSE</i>	0.0462	$4.87.e^{-5}$	0.0082	0.0002	$3.25.e^{-06}$

Tab .1. Estimateurs du maximum de vraisemblance ($\hat{\alpha}, \hat{\gamma}, \hat{\lambda}$) des paramètres et leurs erreurs quadratiques moyennes.

Du tableau 1, nous pouvons remarquer que les estimateurs du maximum de vraisemblance sont convergents.

1.6 Estimation du maximum de vraisemblance en cas de données censurées

1.6.1 Estimation du maximum de vraisemblance

Considérons cette fois-ci le cas où les données peuvent être censurées à droite. Soit une variable aléatoire T distribuée selon une loi *EW* de vecteur de paramètres $\theta = (\alpha, \gamma, \lambda)^T$. 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 indépendantes $t_i = \min(T_i, C_i)$ pour $i = 1, \dots, n$. La censure est supposée non informative, alors la distribution de C_i ne dépend pas des paramètres inconnus de T_i .

La fonction de vraisemblance dans le cas de données censurées peut être donnée par :

$$L(t, \theta) = \prod_{i=1}^n f^{\delta_i}(t_i, \theta) S^{1-\delta_i}(t_i, \theta), \delta_i = 1_{\{T_i \leq C_i\}}$$

La fonction log-vraisemblance est :

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

Comme

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

Alors l'équation de la log-vraisemblance devient :

$$l(t, \theta) = \sum_{i=1}^n \delta_i \ln h(t_i, \theta) + \sum_{i=1}^n \ln S(t_i, \theta)$$

Dans notre cas, l'équation de la log-vraisemblance est :

$$l(t, \theta) = \sum_{i=1}^n \delta_i \ln \left[\lambda \gamma \left(\frac{t}{\alpha} \right)^{\gamma-1} \exp \left\{ \left(\frac{t}{\alpha} \right)^\gamma \right\} \right] + \sum_{i=1}^n \ln \left[\exp \left\{ -\lambda \alpha \left(e^{\left(\frac{t}{\alpha} \right)^\gamma} - 1 \right) \right\} \right]$$

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

$$\frac{\partial l(\theta)}{\partial \alpha} = \sum_{i=1}^n \delta_i \left[\frac{1-\gamma}{\alpha} - \frac{\gamma \left(\frac{t_i}{\alpha} \right)^\gamma}{\alpha} \right] + \sum_{i=1}^n \left[-\lambda \left(e^{\left(\frac{t_i}{\alpha} \right)^\gamma} - 1 \right) + \lambda \gamma \left(\frac{t_i}{\alpha} \right)^\gamma e^{\left(\frac{t_i}{\alpha} \right)^\gamma} \right] = 0$$

$$\frac{\partial l(\theta)}{\partial \gamma} = \sum_{i=1}^n \delta_i \left[\frac{1}{\gamma} + \ln \left(\frac{t_i}{\alpha} \right) + \left(\frac{t_i}{\alpha} \right)^\gamma \ln \left(\frac{t_i}{\alpha} \right) \right] + \sum_{i=1}^n \left[-\lambda \alpha \left(\frac{t_i}{\alpha} \right)^\gamma \ln \left(\frac{t_i}{\alpha} \right) e^{\left(\frac{t_i}{\alpha} \right)^\gamma} \right] = 0$$

$$\frac{\partial l(\theta)}{\partial \lambda} = \sum_{i=1}^n \delta_i \left[\frac{1}{\lambda} \right] + \sum_{i=1}^n \left[-\alpha \left(e^{\left(\frac{t_i}{\alpha} \right)^\gamma} - 1 \right) \right] = 0$$

Pour résoudre le système de fonctions de score, assez compliquées, on fait appel aux méthodes numériques, telles que la méthode de *Monte Carlo*, *l'algorithme BB* ou autres.

1.6.2 Simulations cas de données censurées

On considère le modèle EW. Les données ont été simulées $N = 10,000$ fois (avec les tailles des échantillons $n = 30; n = 50; n = 100; n = 250; n = 500$) et les valeurs des paramètres $\alpha = 0.87, \gamma = 3.5, \lambda = 1.5$. En utilisant le logiciel *R* et l'algorithme Barzilai-Borwein (BB) (Ravi. V 2009) pour les calculs, les moyennes des valeurs simulées des estimateurs du maximum de vraisemblance $\hat{\alpha}, \hat{\gamma}, \hat{\lambda}$ des paramètres et leurs erreurs quadratiques moyennes (notées *MSE*) sont calculées et présentées dans la table 2.

$N = 10,000$	$n = 30$	$n = 50$	$n = 100$	$n = 250$	$n = 500$
$\hat{\alpha}$	0.8821	0.9078	0.9021	0.8810	0.8743
<i>MSE</i>	$9.85.e^{-06}$	0.0008	0.0005	$4.25.e^{-06}$	$2.14.e^{-05}$
$\hat{\gamma}$	3.6429	3.5879	3.5427	3.5078	3.4953
<i>MSE</i>	0.0204	0.0077	0.00205	$5.27.e^{-06}$	$2.14.e^{-06}$
$\hat{\lambda}$	1.4824	1.5698	1.5318	1.4739	1.4430
<i>MSE</i>	0.0003	0.0048	0.0010	0.0006	0.0003

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

Du tableau 2, nous pouvons remarquer que les erreurs quadratiques moyennes sont très faibles ce qui confirme la convergence des estimateurs du maximum de vraisemblance.

CHAPITRE 2

Tests d'ajustement

Sommaire

2.1	Introduction	21
2.2	Nikulin-Rao-Robson test pour données complètes	22
2.3	Nikulin-Rao-Robson statistique pour le modèle <i>Extension de Weibull</i>	23
2.3.1	Calcul de la matrice d'information des données groupées $J(\theta)$	24
2.3.2	Calcul de la matrice d'information de Fisher $I(\theta)$	25
2.3.3	Simulation (NRR statistique Y_n^2)	25
2.3.4	Application à des données réelles	26
2.4	Tests d'ajustement classiques	27
2.4.1	Test de Kolmogorov-Smirnov	27
2.4.2	Test d'Anderson-Darling	28
2.4.3	Test de Cramer-Von Mises	28
2.5	Tests d'ajustement pour données censurées	28
2.6	Test d'ajustement en présence de censure pour la distribution <i>EW</i>	33
2.6.1	Calcul de la matrice \hat{W}	34
2.6.2	Calcul de la matrice \hat{I}	34

2.6.3	Simulation de la distribution de la statistique Y_n^2	35
2.6.4	Simulation des valeurs du critère de test	36
2.6.5	Analyse de données de survie	37

2.1 Introduction

Dans le cas de données complètes, diverses techniques sont utilisées pour vérifier l'adéquation de modèles mathématiques à des données issues de l'observation. Les tests les plus répandus sont ceux basés sur la statistique du chi-deux de Pearson. Néanmoins, ceux-ci ne peuvent être appliqués dans toutes les situations, particulièrement quand les paramètres du modèle sont inconnus ou bien quand les données sont censurées. Depuis la moitié du siècle dernier, les chercheurs ont commencé à proposer des modifications des statistiques existantes pour tenir compte des paramètres inconnus d'une part et de la censure d'autre part. Pour les données complètes, Nikulin (1973) et Rao et Robson (1974) ont proposé séparément une statistique connue aujourd'hui sous le nom de la statistique NRR (Nikulin-Rao-Robson). Ce test statistique qui suit une distribution du chi-deux, est une modification naturelle de la statistique de Pearson.

Si en plus des paramètres inconnus, les données sont censurées, les tests classiques sont inadéquats pour vérifier une hypothèse H_0 selon laquelle une suite d'observations provient d'une famille paramétrique $F(t)$. Habib et Thomas (1986) ont considéré les modifications naturelles de la statistique NRR. Ces tests sont basés sur les différences entre deux estimateurs des probabilités, l'un est basé sur l'estimateur de Kaplan-Meier, l'autre basé sur les estimateurs du maximum de vraisemblance des paramètres inconnus de la fonction de distribution cumulative du modèle testé. Quand à Bagdonavičius et Nikulin (2011), Bagdonavičius et al. (2013), ils ont proposé une modification de la statistique NRR qui tient compte de la censure aléatoire droite. Cette statistique, basée sur les estimateurs du maximum de vraisemblance sur les données initiales, suit également à la limite une distribution du chi-deux. Pour plus de détails sur la construction de ces statistiques, on peut voir Voinov et al. (2013). Ces techniques ont été utilisées pour ajuster des observations aux modèles AFT de Weibull, log-normal et loglogistique (Bagdonavičius et al. 2013), au modèle inverse Weibull généralisé (Goual et Seddik-Ameur 2014), à la distribution de Birbaurn Saunders (Nikulin et Tran 2013), au modèle à risques concurrents de Bertholon (Chouia et Seddik-Ameur 2017).

Dans ce travail nous construisons des tests d'ajustement du type du chi-deux modifié pour le modèle extension de Weibull (EW) dans le cas de données complètes et dans le cas de la censure aléatoire droite et les paramètres inconnus. On se base sur la statistique NRR pour le cas de données complètes. En présence de censure, on se base sur la modification de la statistique NRR proposée par Bagdonavičius et Nikulin (2011).

2.2 Nikulin-Rao-Robson test pour données complètes

Pour tester l'hypothèse H_0 selon laquelle T_1, T_2, \dots, T_n , un n -échantillon est issu d'une famille paramétrique $F(t, \theta)$

$$H_0 : P \{T_i \leq t \mid H_0\} = F(t, \theta), \quad t \geq 0,$$

où $\theta = (\theta_1, \theta_2, \dots, \theta_s)^T \in \Theta \subset \mathbb{R}^s$ représente le vecteur de paramètres inconnus, Nikulin (1973) et Rao et Robson (1974) ont proposé Y^2 la statistique NRR définie comme suit :

Les observations T_1, T_2, \dots, T_n sont groupées en r sous-intervalles I_1, I_2, \dots, I_r mutuellement disjoints $I_j =]a_{j-1}, a_j]$, où $j = \overline{1, r}$.

Les bornes a_j des intervalles I_j sont obtenues telles que

$$p_j(\theta) = \int_{a_{j-1}}^{a_j} f(t, \theta) dt = \frac{1}{r}, \quad j = 1, 2, \dots, r$$

Alors

$$a_j = F^{-1} \left(\frac{j}{r} \right), \quad j = \overline{1, r-1}.$$

Si $\nu = (\nu_1, \nu_2, \dots, \nu_r)^T$ est le vecteur des fréquences obtenu par le regroupement de données dans ces intervalles I_j

$$\nu_j = \text{card} \{i : T_i \in I_j, i = 1, 2, 3, \dots, n\}$$

La statistique NRR est donnée par

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

où

$$X_n^2(\theta) = X_n^T(\theta) X_n(\theta) \text{ avec } X_n(\theta) = \left(\frac{\nu_1 - np_1(\theta)}{\sqrt{np_1(\theta)}}, \dots, \frac{\nu_r - np_r(\theta)}{\sqrt{np_r(\theta)}} \right)^T$$

$J(\theta)$ est la matrice d'information pour les données groupées définie par

$$J = J(\theta) = B(\theta)^T B(\theta) \text{ avec } B(\theta) = \left[\frac{1}{\sqrt{p_j(\theta)}} \frac{\partial p_j(\theta)}{\partial \theta_j} \right]_{r \times s}$$

$$L(\hat{\theta}_n) = \left(L_1(\hat{\theta}_n), L_2(\hat{\theta}_n), \dots, L_s(\hat{\theta}_n) \right)^T \text{ avec } L_k(\hat{\theta}_n) = \sum_{j=1}^r \frac{\nu_j}{p_j} \frac{\partial p_j(\hat{\theta}_n)}{\partial \theta_k}, \quad k = 1, \dots, s$$

$I_n(\hat{\theta}_n)$ représente la matrice d'information de Fisher estimée et $\hat{\theta}_n$ est l'estimateur du maximum de vraisemblance du vecteur θ des paramètres. La statistique Y_n^2 suit à la limite une distribution du chi-carré χ_{r-1}^2 avec $(r-1)$ degrés de liberté.

2.3 Nikulin-Rao-Robson statistique pour le modèle *Extension de Weibull*

Considérons un échantillon $T = (T_1, T_2, \dots, T_n)^T$. Pour vérifier si ces données sont distribuées selon le modèle extension de Weibull, $P(T_i \leq t) = F_{EW}(t, \theta)$, avec des paramètres inconnus $\theta = (\alpha, \gamma, \lambda)^T$, on construit un test d'ajustement du chi-deux en adaptant la statistique NRR développée dans la section précédente.

Les données sont regroupées dans des intervalles I_j . Sous l'hypothèse nulle H_0 , les limites a_j de ces intervalles sont obtenues comme suit :

$$\hat{a}_j = F_{EW}^{-1}\left(\frac{j}{r}\right) \quad \text{alors,} \quad \hat{a}_j = \left(\alpha^\gamma \ln\left(1 - \frac{\ln(1 - \frac{j}{r})}{\lambda \alpha}\right)\right)^{\frac{1}{\gamma}}$$

La forme de la statistique $X_n^2(\theta_n)$ pour l'extension Weibull peut être déduite de sa fonction de fiabilité $S(t)$, car

$$p_j = F_{EW}(a_j) - F_{EW}(a_{j-1}) = S(a_{j-1}) - S(a_j)$$

d'où

$$X_n^2(\theta_n) = \frac{1}{n} \sum_{j=1}^r \frac{(\nu_j - np_j)^2}{p_j} = \frac{1}{n} \sum_{j=1}^r \frac{(\nu_j - n[S(a_{j-1}) - S(a_j)])^2}{S(a_{j-1}) - S(a_j)}$$

Le vecteur $(\nu_j)_r$ représente les fréquences observées obtenues dans ce groupement.

2.3.1 Calcul de la matrice d'information des données groupées

$$J(\theta)$$

Les composants de la matrice symétrique estimée $J(\hat{\theta}) = B(\hat{\theta})^T B(\hat{\theta})$ sont obtenus de la manière suivante :

$$\begin{aligned} J_{11} &= \sum_j \frac{1}{p_j} \left(\frac{\partial p_j}{\partial \alpha} \right)^2, & J_{22} &= \sum_j \frac{1}{p_j} \left(\frac{\partial p_j}{\partial \gamma} \right)^2, & J_{33} &= \sum_j \frac{1}{p_j} \left(\frac{\partial p_j}{\partial \lambda} \right)^2, \\ J_{12} = J_{21} &= \sum_j \frac{1}{p_j} \left(\frac{\partial p_j}{\partial \alpha} \right) \left(\frac{\partial p_j}{\partial \gamma} \right), & J_{13} = J_{31} &= \sum_j \frac{1}{p_j} \left(\frac{\partial p_j}{\partial \alpha} \right) \left(\frac{\partial p_j}{\partial \lambda} \right), \\ J_{23} = J_{32} &= \sum_j \frac{1}{p_j} \left(\frac{\partial p_j}{\partial \gamma} \right) \left(\frac{\partial p_j}{\partial \lambda} \right), \end{aligned}$$

et le vecteur

$$L(\hat{\theta}_n) = \left(L_1(\hat{\theta}_n) = \sum_{j=1}^r \frac{\nu_j}{p_j} u_{j1}, L_2(\hat{\theta}_n) = \sum_{j=1}^r \frac{\nu_j}{p_j} u_{j2}, L_3(\hat{\theta}_n) = \sum_{j=1}^r \frac{\nu_j}{p_j} u_{j3} \right)^T$$

$$J(\theta) = B(\theta)^T B(\theta), \quad B(\theta) = (b_{jk}(\theta))_{r \times s}$$

$$b_{jk} = \frac{1}{\sqrt{p_j}} \frac{\partial p_j(\theta)}{\partial \theta_i}, \quad j = 1, \dots, r, \quad k = 1, \dots, s$$

On a obtenu les quantités $(u_{jk})_{r \times s}$ pour la distribution extension de Weibull sous une forme simple :

$$\begin{aligned} u_{j1} &= \frac{\partial p_j(\theta)}{\partial \alpha} = S(a_{j-1})[-\lambda(\exp\{(\frac{a_{j-1}}{\alpha})^\gamma\} - 1) + (\frac{a_{j-1}}{\alpha})h(a_{j-1})] \\ &\quad - S(a_j)[- \lambda(\exp\{(\frac{a_j}{\alpha})^\gamma\} - 1) + (\frac{a_j}{\alpha})h(a_j)] \end{aligned}$$

$$u_{j2} = \frac{\partial p_j(\theta)}{\partial \gamma} = -(\frac{a_{j-1}}{\alpha}) \frac{\alpha}{\gamma} h(a_{j-1}) S(a_{j-1}) \ln(\frac{a_{j-1}}{\alpha}) + (\frac{a_j}{\alpha}) \frac{\alpha}{\gamma} h(a_j) S(a_j) \ln(\frac{a_j}{\alpha})$$

$$u_{j3} = \frac{\partial p_j(\theta)}{\partial \lambda} = -\alpha(\exp\{(\frac{a_{j-1}}{\alpha})^\gamma\} - 1) S(a_{j-1}) + \alpha(\exp\{(\frac{a_j}{\alpha})^\gamma\} - 1) S(a_j)$$

où

$$S(a_j) = \exp \left\{ -\lambda \alpha \left(e^{(\frac{a_j}{\alpha})^\gamma} - 1 \right) \right\}$$

$S(a_j)$ est la fonction de la fiabilité et $h(a_j)$ la fonction du taux de risque de la distribution *EW*.

2.3.2 Calcul de la matrice d'information de Fisher $I(\theta)$

Les éléments de la matrice d'information de Fisher définie par :

$$I_n(\theta)_{k,l} = -E\left[\frac{\partial^2 \ln f(t, \theta)}{\partial \theta_k \partial \theta_l}\right]$$

sont nécessaires pour la construction de la statistique Y^2 . Après plusieurs simplifications, on obtient tous les éléments :

$$\begin{aligned} \frac{\partial^2 l(\theta)}{\partial \alpha^2} &= \frac{n(\gamma - 1)}{\alpha^2} + \sum_{i=1}^n \left[\frac{\left(\frac{t_i}{\alpha}\right) h(t_i) (-\gamma - \gamma \left(\frac{t_i}{\alpha}\right)^\gamma + 1)}{\alpha} + \frac{\left(\frac{t_i}{\alpha}\right)^\gamma \gamma (\gamma + 1)}{\alpha^2} \right] \\ \frac{\partial^2 l(\theta)}{\partial \gamma^2} &= \left(-\frac{n}{\gamma^2}\right) + \sum_{i=1}^n \left[\left(1 + \left(\frac{t_i}{\alpha}\right)^\gamma\right) (-\lambda \alpha \left(\frac{t_i}{\alpha}\right)^\gamma \ln \left(\frac{t_i}{\alpha}\right)^2 \exp \left(\frac{t_i}{\alpha}\right)^\gamma) + \left(\frac{t_i}{\alpha}\right)^\gamma \ln \left(\frac{t_i}{\alpha}\right)^2 \right] \\ \frac{\partial^2 l(\theta)}{\partial \lambda^2} &= -\frac{n}{\lambda^2} \\ \frac{\partial^2 l(\theta)}{\partial \gamma \partial \lambda} &= \frac{\partial^2 l(\theta)}{\partial \lambda \partial \gamma} = \sum_{i=1}^n \left[-\alpha \left(\frac{t_i}{\alpha}\right)^\gamma \ln \left(\frac{t_i}{\alpha}\right) \exp \left(\frac{t_i}{\alpha}\right)^\gamma \right] \\ \frac{\partial^2 l(\theta)}{\partial \alpha \partial \gamma} &= \frac{\partial^2 l(\theta)}{\partial \gamma \partial \alpha} = -\frac{n}{\alpha} + \sum_{i=1}^n \left[\left(1 + \left(\frac{t_i}{\alpha}\right)^\gamma - \frac{1}{\gamma}\right) \lambda \gamma \left(\frac{t_i}{\alpha}\right)^\gamma \ln \left(\frac{t_i}{\alpha}\right) \exp \left(\frac{t_i}{\alpha}\right)^\gamma \right. \\ &\quad \left. + \lambda \left(\frac{t_i}{\alpha}\right)^\gamma \exp \left(\frac{t_i}{\alpha}\right)^\gamma - \frac{\left(\frac{t_i}{\alpha}\right)^\gamma (1 + \ln \left(\frac{t_i}{\alpha}\right) \gamma)}{\alpha} \right] \\ \frac{\partial^2 l(\theta)}{\partial \alpha \partial \lambda} &= \frac{\partial^2 l(\theta)}{\partial \lambda \partial \alpha} = n + \sum_{i=1}^n \left[\frac{\left(\frac{t_i}{\alpha}\right)}{\lambda} h(t_i) - \exp \left(\frac{t_i}{\alpha}\right)^\gamma \right] \end{aligned}$$

Les estimateurs du maximum de vraisemblance $\hat{\theta}_n$ des paramètres inconnus de la distribution EW sont calculés sur les données initiales. La statistique Y^2 ne dépend pas des paramètres, on peut donc utiliser la matrice d'information de Fisher estimée $I_n(\hat{\theta}_n)$. Toutes les composantes de la statistique Y^2 , pour la distribution extension de Weibull sont fournies, par conséquent Y^2 peut être déduite facilement.

2.3.3 Simulation (NRR statistique Y_n^2)

Pour appuyer les résultats obtenus dans ce travail, on mène une étude intensive par simulations numériques. Ainsi, pour tester l'hypothèse nulle H_0 qu'un échantillon appartienne au modèle extension de Weibull, nous calculons Y^2 la statistique NRR de 10,000 échantillons

simulés avec des tailles $n = 30; n = 50; n = 100; n = 250; n = 500$, respectivement. Pour différents niveaux théoriques ($\epsilon = 0.02, 0.05, 0.01, 0.1$), on calcule la moyenne des nombres de non-rejet de l'hypothèse nulle, lorsque $Y_n^2 \leq \chi_\epsilon^2(r-1)$ puis, nous présentons les résultats des niveaux empiriques et théoriques correspondants dans la Table 3.

$N = 10,000$	$\epsilon = 0.02$	$\epsilon = 0.05$	$\epsilon = 0.01$	$\epsilon = 0.1$
$n = 30$	0.9828	0.9517	0.9953	0.9061
$n = 50$	0.9824	0.9576	0.9932	0.9039
$n = 100$	0.9818	0.9522	0.9912	0.9019
$n = 250$	0.9809	0.9501	0.9921	0.9009
$n = 500$	0.9801	0.9499	0.9909	0.8999

Tab.3. niveaux empiriques et niveaux théoriques correspondants ($\epsilon = 0.02, 0.05, 0.01, 0.1$)

Comme il peut être vu, les valeurs des niveaux empiriques calculés sont très proches de ceux de leurs niveaux théoriques correspondants. Donc, nous concluons que le test proposé est bien adapté à la distribution extension de Weibull.

2.3.4 Application à des données réelles

Nous considérons les durées de vie de 50 appareils, fournies par Aarset (1987) et étudiées par d'autres auteurs. Pour tester l'hypothèse nulle H_0 que ces données (tableau 4)

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

Tab .4. Durée de vie de 50 appareils d'Aarset (1987).

sont ajustées par une distribution extension de weibull , nous utilisons la statistique NRR obtenue dans ce travail. En utilisant le logiciel *R* et l'algorithme Barzilai-Borwein (BB) (Ravi, 2009), nous calculons d'abord les estimateurs du maximum de vraisemblance $\hat{\alpha} = 101.0917$, $\hat{\gamma} = 0.8402$, $\hat{\lambda} = 0.01388$ des paramètres inconnus α, γ, λ . La matrice d'information Fisher estimée est alors

$$I(\hat{\theta}) = \begin{pmatrix} 0.0002546459 & -0.0255712941 & 6.2368342619 \\ -0.0255712941 & -94.14084510 & 2220.92706251 \\ 6.2368342619 & 2220.92706251 & -2.592642e + 05 \end{pmatrix}$$

Les quantités W et $X_n^2(\hat{\theta})$ sont obtenues

$$W = \frac{1}{n} L^T(\theta_n) (I_n(\hat{\theta}_n) - J(\hat{\theta}_n))^{-1} L(\hat{\theta}_n) = -0.001128842 \quad ; \quad X_n^2(\hat{\theta}) = 8.8$$

On déduit alors la valeur de $Y_n^2 = 8.798871$. Pour niveau de signification $\epsilon = 0.02$, la valeur critique est $\chi_{0.02}^2(5 - 1) = 11.6678$, alors, la statistique NRR Y_n^2 est inférieure à la valeur critique, ceci nous permet de dire que ces données correspondent convenablement au modèle extension de weibull.

2.4 Tests d'ajustement classiques

Quand les paramètres sont connus, il existe des tests d'ajustement basés sur les différences entre les fonctions de répartition théoriques et les fonctions de répartition empiriques. Nous en citons les plus couramment utilisés.

2.4.1 Test de Kolmogorov-Smirnov

Ce test est basé sur la plus grande distance verticale entre la fonction de répartition empirique $F_n(t)$ et la fonction de distribution théorique $F(t)$.

La statistique est

$$\begin{aligned} KS &= \sup_{t \in \mathbb{R}} |F(t, \theta) - F_n(t)| \\ &= \max_{i=1}^n (F(t_i, \theta) - (i-1)/n) \vee (i/n - F(t_i, \theta)) \end{aligned}$$

avec $(t_i)_{1,\dots,n}$ un n-échantillon que l'on veut ajuster à la distribution $F(t)$. Pour un seuil ϵ donné, si la valeur de la statistique KS est supérieure au point critique, nous rejetons l'hypothèse nulle [voir Bolshev et Smirnov (1983)].

2.4.2 Test d'Anderson–Darling

La statistique d'Anderson - Darling (AD) est définie par la quantité :

$$AD = -n - 2 \sum_{i=1}^n \left\{ \frac{2i-1}{2n} \ln F(t_i, \theta) + \left(1 - \frac{2i-1}{2n}\right) \ln(1 - F(t_i, \theta)) \right\}$$

Si la valeur du test statistique AD est supérieure au point critique, nous rejetons l'hypothèse nulle.

2.4.3 Test de Cramer-Von Mises

La statistique w^2 de Cramér-Von Mises peut s'écrire :

$$S_w = \frac{1}{12n} + \sum_{i=1}^n \left\{ F(t_i, \theta) - \frac{2i-1}{2n} \right\}^2$$

De même, on rejette l'hypothèse nulle H_0 , si la valeur calculée est supérieure à la valeur critique.

2.5 Tests d'ajustement pour données censurées

Pour vérifier l'adéquation du modèle extension de Weibull quand les paramètres sont inconnus et les données censurées, on utilise l'approche proposée par Bagdonavičius et Nikulin (2011), Bagdonavičius et al. (2013) que nous développons dans ce paragraphe. C'est un test du type du chi-deux modifié basé sur une modification de la statistique NRR dont la distribution est un chi-deux.

Soit l'échantillon

$$(t_1, \delta_1), (t_2, \delta_2), \dots, (t_n, \delta_n) \tag{2.1}$$

où

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

2.5. Tests d'ajustement pour données censurées

Supposons que les durées de vie 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 $\{f(t, \theta), \theta \in \Theta \subset R^m\}$. et le taux de hasard cumulé est noté par :

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

La fonction de la log-vraisemblance est :

$$l(\theta) = \sum_{i=1}^n \delta_i \ln h(t_i, \theta) + \sum_{i=1}^n \ln S(t_i, \theta) \quad (2.2)$$

l'estimateur du maximum de vraisemblance $\hat{\theta}$ de θ vérifie :

$$\dot{l}(\hat{\theta}) = 0$$

où

$$\dot{l}(\theta) = \frac{\partial}{\partial \theta} l(\theta) = \left(\frac{\partial}{\partial \theta_1} l(\theta), \dots, \frac{\partial}{\partial \theta_m} l(\theta) \right)^T \quad (2.3)$$

La matrice d'information de Fisher est :

$$I(\theta) = -E_{\theta} \ddot{l}(\theta) \quad (2.4)$$

où

$$\ddot{l}(\theta) = \sum_{i=1}^n \delta_i \frac{\partial^2}{\partial \theta^2} \ln h(t_i, \theta) - \sum_{i=1}^n \frac{\partial^2}{\partial \theta^2} H(t_i, \theta) \quad (2.5)$$

Si θ_0 est la vraie valeur de θ et sous les conditions de régularité, on a :

$$\hat{\theta} \xrightarrow{P} \theta_0; \quad \sqrt{n} (\hat{\theta} - \theta_0) \xrightarrow{P} N(0_m, i^{-1}(\theta_0)), \quad 0_m = (0, 0, \dots, 0)^T;$$

$$\frac{1}{\sqrt{n}} \dot{l}(\theta_0) \xrightarrow{d} N(0_m, i(\theta_0)); \quad -\frac{1}{n} \ddot{l}(\hat{\theta}) \xrightarrow{P} i(\theta_0),$$

L'échantillon (2.1) peut être représenté en termes de processus de comptage comme suit :

$$(N_1(t), Y_1(t), t \geq 0), \dots, (N_n(t), Y_n(t), t \geq 0) \quad (2.6)$$

Où

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

Si

$$N(t) = \sum_{i=1}^n N_i(t) \quad \text{et} \quad Y(t) = \sum_{i=1}^n Y_i(t) \quad (2.7)$$

les expressions des équations (2.2), (2.3), (2.4) et (2.5) peuvent être écrites 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 :

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

En utilisant ces processus, les auteurs ont obtenu les relations suivantes :

$$\int_0^\infty \ln h(u, \theta) dN_i(u) = \begin{cases} \ln h(t_i, \theta), & \delta_i = 1, \\ 0, & \delta_i = 0 \end{cases} = \delta_i \ln h(t_i, \theta) \quad (2.8)$$

et

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

Ce qui implique

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

La matrice d'information de Fisher est :

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

Sous les conditions suivantes [Hjort (1990)] :

1- Il existe un voisinage Θ_0 de θ_0 la vraie valeur de θ tel que pour tout n et $\theta \in \Theta_0$, et pour tout $t \in [0, \tau]$, les dérivées partielles d'ordre 1 et 2 de $h(t, \theta)$ par rapport à θ existent et sont continues et bornées. On suppose aussi que la fonction de log-vraisemblance est trois fois différentiables par rapport à $\theta \in \theta_0$.

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

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

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

4- La matrice $i(\theta_0) = \lim_{n \rightarrow \infty} \frac{I(\theta_0)}{n}$ (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 $r > s$ petits intervalles.

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

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

U_j représente le nombre des défaillances observées dans le j^{eme} intervalle, $j = 1, 2, \dots, r$.

On prévoit d'observer E_j défaillances

$$E_j = \int_{a_{j-1}}^{a_j} h(u, \hat{\theta}) Y(u) du = \sum_{i: X_i > a_{j-1}} \left(H(a_j \wedge t_i, \hat{\theta}) - H(a_{j-1}, \hat{\theta}) \right)$$

ici $\hat{\theta}$ est l'estimateur du maximum de vraisemblance du paramètre θ .

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

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

basée sur le vecteur

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

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 un inverse généralisé de la matrice \hat{V}

$$\hat{A} = \begin{pmatrix} \hat{A}_1 & 0 & \cdots & 0 \\ 0 & \hat{A}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \hat{A}_r \end{pmatrix}, \quad \hat{A}_j = \frac{U_j}{n}$$

$$\hat{C} = \begin{pmatrix} \hat{C}_{11} & \hat{C}_{12} & \cdots & \hat{C}_{1r} \\ \hat{C}_{21} & \hat{C}_{22} & \cdots & \hat{C}_{2r} \\ \vdots & \vdots & \cdots & \vdots \\ \hat{C}_{m1} & \hat{C}_{m2} & \cdots & \hat{C}_{mr} \end{pmatrix}, \quad \hat{C}_{lj} = \frac{1}{n} \sum_{i:t_i \in I_j} \delta_i \frac{\partial}{\partial \theta_l} \ln h(t_i, \hat{\theta})$$

$$\hat{i} = \begin{pmatrix} \hat{i}_{11} & \hat{i}_{12} & \cdots & \hat{i}_{1m} \\ \hat{i}_{21} & \hat{i}_{22} & \cdots & \hat{i}_{2m} \\ \vdots & \vdots & \cdots & \vdots \\ \hat{i}_{m1} & \hat{i}_{m2} & \cdots & \hat{i}_{mm} \end{pmatrix}, \quad \hat{i}_{ll'} = \frac{1}{n} \sum_{i=1}^n \delta_i \frac{\partial \ln h(t_i, \hat{\theta})}{\partial \theta_l} \frac{\partial \ln h(t_i, \hat{\theta})}{\partial \theta_{l'}}$$

cette statistique peut s'écrire

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

où la forme quadratique Q est

$$Q = W^T \hat{G}^{-1} W, \quad W = \hat{C} \hat{A}^{-1} Z$$

Sous l'hypothèse H_0 et pour n suffisamment grand, la statistique Y_n^2 suit une loi de χ_r^2 à r degrés de liberté.

Choix de \hat{a}_j :

Soit

$$b_i = (n - i) H(t_{(i)}, \hat{\theta}) + \sum_{l=1}^i H(t_{(l)}, \hat{\theta})$$

où $t_{(i)}$ est le i^{ieme} élément de la statistique d'ordre $(t_{(1)}, \dots, t_{(n)})$.

Si i est le plus petit nombre naturel vérifiant $E_j \in [b_{i-1}, b_i]$, $j = 1, \dots, r - 1$ alors :

$$(n - i + 1) H(a, \hat{\theta}) + \sum_{l=1}^{i-1} H(t_{(l)}, \hat{\theta}) = E_j$$

d'où

$$\hat{a}_j = H^{-1} \left(\frac{E_j - \sum_{l=1}^{i-1} H(t_{(l)}, \hat{\theta})}{n - i + 1}, \hat{\theta} \right), \quad \hat{a}_r = \max(t_{(n)}, \tau)$$

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

2.6 Test d'ajustement en présence de censure pour la distribution EW

Dans cette section, nous étudions la validité du modèle *extension de Weibull EW*, par un test d'ajustement basé sur Y_n^2 la statistique NRR modifiée présentée dans la section précédente.

Supposons H_0 vérifiée, c'est-à-dire que le taux de défaillance T_i suit une distribution EW.

La fonction de survie est :

$$S(t, \theta) = 1 - F(t, \theta) = \exp \left\{ -\lambda \alpha \left(e^{\left(\frac{t}{\alpha}\right)^\gamma} - 1 \right) \right\}$$

Après avoir calculé les estimateurs du maximum de vraisemblance des paramètres α, γ, λ de cette distribution, le choix de \hat{a}_j est obtenu comme suit :

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

où

$$E_j = \frac{j}{r-1} \sum_{i=1}^n H(t_{(i)}, \hat{\theta})$$

Pour chaque intervalle I_j on a $e_j = \frac{E_r}{r}$.

2.6.1 Calcul de la matrice \hat{W}

Le calcul de la matrice \hat{W} passe par celui de la matrice \hat{C} . Calculons les dérivées partielles de la fonction $\ln h(t, \theta)$ par rapport à α, γ et λ :

$$\begin{aligned}\frac{\partial \ln h(t, \theta)}{\partial \alpha} &= \frac{1 - \hat{\gamma}(1 + (\frac{t}{\hat{\alpha}})^{\hat{\gamma}})}{\hat{\alpha}} \\ \frac{\partial \ln h(t, \theta)}{\partial \gamma} &= \frac{1}{\hat{\gamma}} + [(1 + (\frac{t}{\hat{\alpha}})^{\hat{\gamma}}) \ln(\frac{t}{\hat{\alpha}})] \\ \frac{\partial \ln h(t, \theta)}{\partial \lambda} &= \frac{1}{\hat{\lambda}}\end{aligned}$$

Après calcul, nous trouvons les éléments de la matrice \hat{C} :

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

et la matrice \hat{W} peut être déduite :

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

2.6.2 Calcul de la matrice \hat{I}

Les éléments de la matrice d'information de Fisher estimée définie par :

$$\hat{I}_{ll'} = \frac{1}{n} \sum_{i=1}^n \delta_i \frac{\partial \ln h(t_i, \hat{\theta})}{\partial \theta_l} \frac{\partial \ln h(t_i, \hat{\theta})}{\partial \theta_{l'}}, \quad l, l' = 1, 2, 3$$

sont donnés comme suit :

$$\hat{I} = \begin{pmatrix} \hat{i}_{11} & \hat{i}_{12} & \hat{i}_{13} \\ \hat{i}_{21} & \hat{i}_{22} & \hat{i}_{23} \\ \hat{i}_{31} & \hat{i}_{32} & \hat{i}_{33} \end{pmatrix}$$

où

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

$$\widehat{I}_{22} = \frac{1}{n} \sum_{i=1}^n \delta_i \left[\frac{1}{\widehat{\gamma}} + \left[\left(1 + \left(\frac{t}{\widehat{\alpha}} \right)^{\widehat{\gamma}} \right) \ln \left(\frac{t}{\widehat{\alpha}} \right) \right]^2 \right]$$

$$\widehat{I}_{33} = \frac{1}{n} \sum_{i=1}^n \delta_i \left[\frac{1}{\widehat{\lambda}} \right]^2$$

$$\widehat{I}_{12} = \widehat{I}_{21} = \frac{1}{n} \sum_{i=1}^n \delta_i \left[\frac{1 - \widehat{\gamma} \left(1 + \left(\frac{t_i}{\widehat{\alpha}} \right)^{\widehat{\gamma}} \right)}{\widehat{\alpha}} \right] \left[\frac{1}{\widehat{\gamma}} + \left[\left(1 + \left(\frac{t}{\widehat{\alpha}} \right)^{\widehat{\gamma}} \right) \ln \left(\frac{t}{\widehat{\alpha}} \right) \right] \right]$$

$$\widehat{I}_{13} = \widehat{I}_{31} = \frac{1}{n} \sum_{i=1}^n \delta_i \left[\frac{1 - \widehat{\gamma} \left(1 + \left(\frac{t_i}{\widehat{\alpha}} \right)^{\widehat{\gamma}} \right)}{\widehat{\alpha}} \right] \left[\frac{1}{\widehat{\lambda}} \right]$$

$$\widehat{I}_{23} = \widehat{I}_{32} = \frac{1}{n} \sum_{i=1}^n \delta_i \left[\frac{1}{\widehat{\gamma}} + \left[\left(1 + \left(\frac{t}{\widehat{\alpha}} \right)^{\widehat{\gamma}} \right) \ln \left(\frac{t}{\widehat{\alpha}} \right) \right] \right] \left[\frac{1}{\widehat{\lambda}} \right]$$

Enfin, on obtient la forme explicite de la statistique Y_n^2 qui nous permet de tester l'ajustement d'une série d'observations à une distribution extension de Weibull :

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

2.6.3 Simulation de la distribution de la statistique Y_n^2

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_n^2 sous l'hypothèse nulle H_0 , avec différentes valeurs des paramètres et différents nombres d'intervalles r de regroupement de données soumises à une censure à droite. Ensuite, nous avons représenté les histogrammes des données des échantillons par rapport aux distributions du chi-deux correspondantes avec r degrés de liberté (voir les exemples dans les figures 2.1, 2.2 et 2.3).

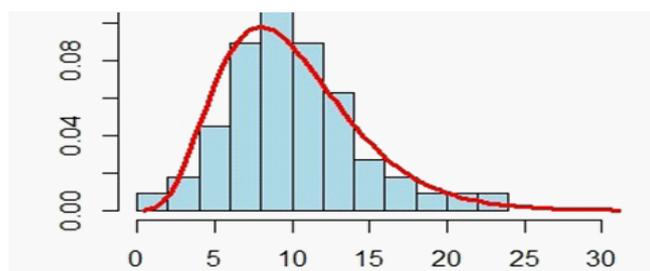


Fig. 2.1. Histogramme de Y^2 pour $n = 30$

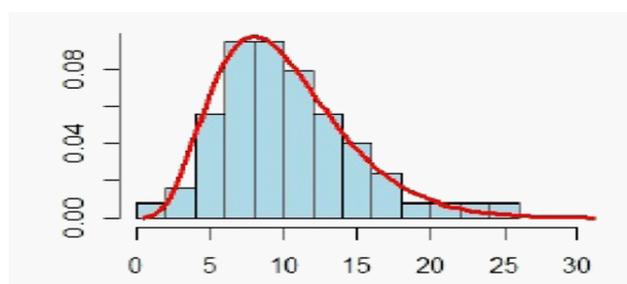


Fig. 2.2. Histogramme de Y^2 pour $n = 100$

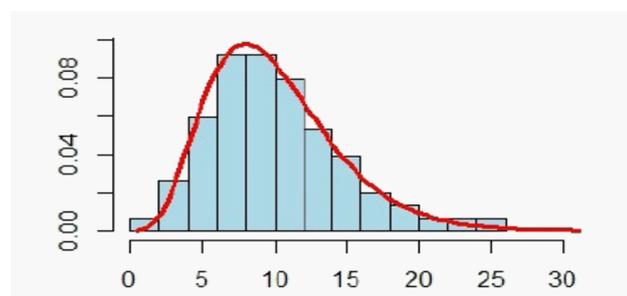


Fig. 2.3. Histogramme de Y^2 pour $n = 400$

Tenant compte des erreurs de simulations, nous pouvons dire que la distribution limite de la statistique Y_n^2 , est une distribution du chi-deux avec r degrés de liberté.

2.6.4 Simulation des valeurs du critère de test

Pour tester l'hypothèse nulle H_0 qu'un échantillon provient d'un modèle extension de Weibull, nous calculons Y^2 la statistique NRR de 10,000 échantillons simulés avec des tailles $n = 30$; $n = 100$; $n = 250$; $n = 500$, respectivement. Pour différents niveaux de signification

2.6. Test d'ajustement en présence de censure pour la distribution EW

($\epsilon = 0.02, 0.05, 0.01, 0.1$), on calcule la moyenne du nombre de non-rejet de l'hypothèse nulle, lorsque $Y_n^2 \leq \chi_\epsilon^2(r)$ puis, nous présentons les résultats des valeurs empiriques et des valeurs théoriques correspondantes dans la Table 5.

$N = 10,000$	$\epsilon = 0.02$	$\epsilon = 0.05$	$\epsilon = 0.01$	$\epsilon = 0.1$
$n = 30$	0.9823	0.9518	0.9943	0.9032
$n = 100$	0.9811	0.9505	0.9937	0.9022
$n = 250$	0.9801	0.9501	0.9915	0.9025
$n = 500$	0.9799	0.9497	0.9903	0.9003

Tab.5. niveaux empiriques et niveaux théoriques correspondants ($\epsilon = 0.02, 0.05, 0.01, 0.1$)

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 distributions de chi-deux à r degrés de liberté. Par conséquent, nous pouvons dire que le test proposé peut ajuster convenablement des données censurées de la distribution EW.

2.6.5 Analyse de données de survie

Nous considérons des données d'échantillon de 50 patients atteints de leucémie aiguë myéloïde aiguë 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 utilisée pour reconstituer leur système immunitaire.

Les données ordonnées de cette étude sont présentées dans la table (6), 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 .6. Taux de survie de la leucémie (en mois) pour les transplantations autologues et allogéniques.

Si on suppose que ces données sont distribuées selon la distribution (EW), l'estimateur du maximum de vraisemblance $\hat{\theta}$ du vecteur de paramètres θ est :

$$\hat{\theta} = (\hat{\alpha}, \hat{\gamma}, \hat{\lambda})^T = (0.0002, 4.3878, 1.3378)^T$$

Nous choisissons un nombre de classes $r = 4$. Les éléments de la statistique Y_n^2 du test sont présentés comme suit :

\hat{a}_j	2.309	6.453	11.423	20.311
U_j	8	7	3	4
e_j	6.2714	6.2714	6.2714	6.2714
\hat{C}_{1j}	-0.1400	0.2430	0.2508	0.0130
\hat{C}_{2j}	0.1501	0.2241	0.1813	0.2513
\hat{C}_{3j}	0.0045	0.0045	0.0045	0.0045

Tab .7. Calculs de \hat{a}_j ; U_j ; e_j ; \hat{C}_{1j} ; \hat{C}_{2j} ; \hat{C}_{3j} .

La matrice d'information de Fisher estimée est :

$$\hat{I} = \begin{pmatrix} -0.6431123 & -1.6349201 & 0.3987176 \\ -1.6349201 & 6.318782 & -0.8563852 \\ 0.3987176 & -0.8563852 & 1.2479566 \end{pmatrix}$$

Après calcul, nous trouvons $Y_n^2 = 9.1577$.

Au seuil $\epsilon = 0.05$, la valeur critique $\chi_{0.05}^2(4) = 9.4877 \succ Y_n^2 = 9.1577$, on peut dire que ces données peuvent être modélisées par la distribution extension de Weibull.

CHAPITRE 3

Construction du modèle EW-AFT

Sommaire

3.1	Introduction	41
3.2	Le modèle AFT	41
3.3	Construction du modèle EW-AFT	43
3.4	Estimation du maximum de vraisemblance en cas de données censurées	45
3.4.1	Calcul des estimateurs du maximum de vraisemblance	45
3.4.2	Simulations	46
3.5	Test d'ajustement pour le modèle EW-AFT en cas de données censurées	47
3.5.1	Choix des intervalles	48
3.5.2	Calcul de la matrice \widehat{W}	48
3.5.3	Calcul de la matrice \widehat{T}	49
3.5.4	Simulation (NRR statistique Y_n^2)	51
3.5.5	Analyse de données de fiabilité	52

3.1 Introduction

Les modèles **AFT** ou modèles à temps de vie accéléré ou tests de vie accélérés (accelerated life testing) [Bagdonavicius et Nikulin (2002), Meeker (1998), Nikulin et al. (2007)] sont des modèles où la fonction de survie dépend de variables dites explicatives représentant les différents facteurs appelés souvent stress pouvant influencer sur la durée de vie du matériel. Ils consistent à réduire les durées de vie des produits par l'accélération des dégradations provoquant les défaillances. Ainsi, on peut estimer la fiabilité du matériel dans des délais courts permettant d'optimiser les performances de celui-ci.

Vu le développement technologique et pour produire de nouveaux matériels de plus en plus performants, les ingénieurs utilisent les données obtenues à partir des expériences de tests de vie accélérés, ce qui a encouragé les chercheurs à proposer de nouveaux modèles *AFT* pour décrire les différentes durées de vie observées. Parmi les modèles étudiés, nous trouvons les modèles AFT de Weibull, exponentiel, log-logistique (Bagdonavicius et al. 2011), le modèle log-gamma (Ortega et al. 2009), le modèle *AFT* inverse Weibull généralisée (Goual et Seddik-Ameur 2014).

Dans ce chapitre, on propose un modèle à temps de vie accéléré *EW – AFT* dont la distribution de base est une extension de Weibull. Ce modèle peut mieux modéliser différentes durées de vie en fiabilité ou en analyse de survie que la distribution *Weibull – AFT*. Après la présentation du modèle, on utilise la méthode d'estimation 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 et les paramètres inconnus permettant d'ajuster des observations à ce type de modèle de manière satisfaisante.

3.2 Le modèle AFT

Un modèle AFT dont la fonction de survie est $S_0(t)$ se définit comme suit :

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

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

3.2. Le modèle AFT

Si $z(\cdot)$ est constant, nous écrivons z à la place de $z(\cdot)$, et nous notons par E_1 le sous-ensemble de E de tous les stress constants.

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

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

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

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

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

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

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

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

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

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

Le modèle AFT paramétrique est donné par la formule :

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

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

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

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 $S(t) = S_0(\ln t)$.

3.3 Construction du modèle EW-AFT

Dans cette section, nous proposons la construction d'un modèle à temps de vie accéléré dont la distribution de base est une distribution extension de Weibull (*EW – AFT*). La fonction de survie de base $S_0(t)$ est donnée par :

$$S_0(t) = \exp \left\{ -\lambda \alpha \left(e^{\left(\frac{t}{\alpha}\right)^\gamma} - 1 \right) \right\}, t > 0, \alpha > 0, \gamma > 0, \lambda > 0$$

La fonction de distribution cumulative du modèle *EW – AFT* est déduite comme suit :

$$\begin{aligned} F(t) &= 1 - S_0(\exp \{ -\beta^T z \} t) \\ F(t) &= 1 - \exp \left[-\lambda \alpha \left(\exp \left\{ \left(\frac{\exp \{ -\beta^T z \} t}{\alpha} \right)^\gamma \right\} - 1 \right) \right] \end{aligned}$$

où α, γ et λ sont les paramètres de la distribution EW, $\beta = (\beta_0, \dots, \beta_m)^T$ sont les paramètres de la régression du modèle AFT, et $z = (1, z_1, \dots, z_m)^T$ sont les covariables représentant les stress eventuels.

Ainsi, la fonction de survie pour le modèle *EW – AFT* est obtenue comme suit :

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

Les fonctions risque et risque cumulé sont respectivement :

$$\begin{aligned} h(t) &= -\frac{S'(t)}{S(t)} \\ h(t) &= \lambda \gamma \alpha^{1-\gamma} u^\gamma t^{\gamma-1} \phi(t) \exp \{ -\lambda \alpha (\phi(t) - 1) \} \end{aligned}$$

où $u = \exp \{ -\beta^T z \}$, et $\phi(t) = \exp \left\{ \left(\frac{tu}{\alpha} \right)^\gamma \right\}$.

Et

$$\begin{aligned} H(t) &= -\ln S(t) \\ H(t) &= -\ln \left(\exp \left[-\lambda \alpha (\phi(t) - 1) \right] \right) \end{aligned}$$

3.3. Construction du modèle EW-AFT

Nous présentons dans la **Fig. 3.1**, les différentes formes du taux de hasard de la distribution EW-AFT en variant la valeur du paramètre de forme γ .

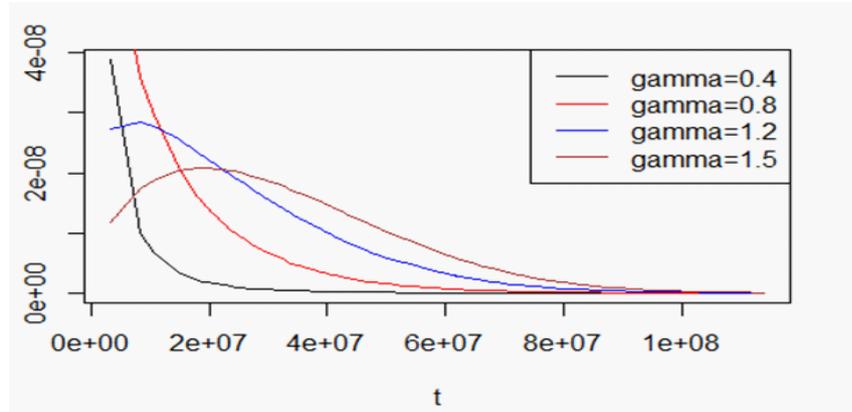


Fig. 3.1. Taux de hasard de la distribution EW-AFT pour $\alpha = 3.5$ et $\lambda = 2, \beta_0 = 6, \beta_1 = 0.82$ avec $\gamma(0.4; 0.8; 1.2; 1.5)$.

Nous présentons dans la **Fig. 3.2**, les différentes formes du taux de hasard du modèle EW – AFT en variant la valeur du paramètre de forme λ .

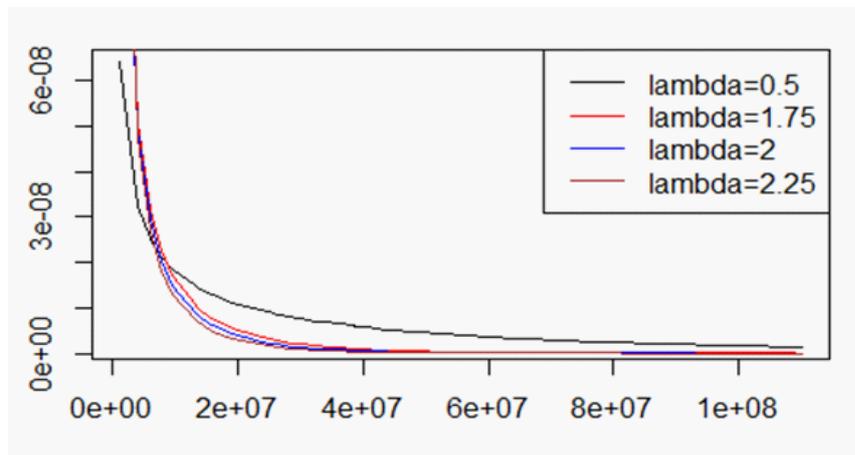


Fig. 3.2. Taux de hasard de la distribution EW-AFT pour $\alpha = 3.5$ et $\gamma = 0.8, \beta_0 = 6, \beta_1 = 0.82$ avec $\lambda(0.5; 1.75; 2; 2.25)$.

3.4 Estimation du maximum de vraisemblance en cas de données censurées

3.4.1 Calcul des estimateurs du maximum de vraisemblance

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

Supposons que les données se composent de n observations indépendantes $t_i = \min(T_i, C_i)$ pour $i = 1, \dots, n$. La censure est supposée non informative, donc la distribution de C_i ne dépend pas des paramètres inconnus de T_i .

La fonction de vraisemblance

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

est équivalente

$$L(t, \theta) = \prod_{i=1}^n h^{\delta_i}(t_i, \theta) S(t_i, \theta), \quad \delta_i = 1_{\{T_i \leq C_i\}}$$

Dans le cas de la distribution EW-AFT, nous avons :

$$L(t, \theta) = \prod_{i=1}^n [\lambda \gamma \alpha^{1-\gamma} u_i^\gamma t_i^{\gamma-1} \phi(t_i) \exp\{-\lambda \alpha (\phi(t_i) - 1)\}]^{\delta_i} [\exp\{-\lambda \alpha (\phi(t_i) - 1)\}]$$

et la fonction de la log-vraisemblance est :

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

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

$$\begin{aligned} \frac{\partial l(\theta)}{\partial \alpha} &= \sum_{i=1}^n \delta_i \left[\frac{1 - \gamma \left[1 + \left(\frac{e^{-\beta^T z_i t_i}}{\alpha}\right)^\gamma (1 - \lambda \alpha e^{\left(\frac{e^{-\beta^T z_i t_i}}{\alpha}\right)^\gamma})\right]}{\alpha} - \lambda \left(e^{\left(\frac{e^{-\beta^T z_i t_i}}{\alpha}\right)^\gamma} - 1\right) \right] \\ &\quad - \sum_{i=1}^n [\lambda \left(e^{\left(\frac{e^{-\beta^T z_i t_i}}{\alpha}\right)^\gamma} - 1\right) - \lambda \gamma \left(\frac{e^{-\beta^T z_i t_i}}{\alpha}\right)^\gamma e^{\left(\frac{e^{-\beta^T z_i t_i}}{\alpha}\right)^\gamma}] \end{aligned}$$

$$\begin{aligned} \frac{\partial l(\theta)}{\partial \gamma} &= \sum_{i=1}^n \delta_i \left[\frac{1}{\gamma} - \ln(\alpha) - \beta^T z_i + \ln(t_i) + \left(\frac{e^{-\beta^T z_i t_i}}{\alpha} \right)^\gamma \ln \left(\frac{e^{-\beta^T z_i t_i}}{\alpha} \right) [1 - \lambda \alpha e^{(\frac{e^{-\beta^T z_i t_i}}{\alpha})^\gamma}] \right. \\ &\quad \left. - \sum_{i=1}^n [\lambda \alpha \left(\frac{e^{-\beta^T z_i t_i}}{\alpha} \right)^\gamma e^{(\frac{e^{-\beta^T z_i t_i}}{\alpha})^\gamma} \ln \left(\frac{e^{-\beta^T z_i t_i}}{\alpha} \right)] \right] \end{aligned}$$

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

$$\frac{\partial l(\theta)}{\partial \beta_0} = \sum_{i=1}^n \delta_i \left[-\gamma + \gamma \left(\frac{e^{-\beta^T z_i t_i}}{\alpha} \right)^\gamma (1 - \lambda \alpha e^{(\frac{e^{-\beta^T z_i t_i}}{\alpha})^\gamma}) \right] - \sum_{i=1}^n [\alpha \lambda \gamma \left(\frac{e^{-\beta^T z_i t_i}}{\alpha} \right)^\gamma e^{(\frac{e^{-\beta^T z_i t_i}}{\alpha})^\gamma}]$$

$$\begin{aligned} \frac{\partial l(\theta)}{\partial \beta_1} &= \sum_{i=1}^n \delta_i \left[-\gamma z_{i1} + \gamma z_{i1} \left(\frac{e^{-\beta^T z_i t_i}}{\alpha} \right)^\gamma (1 - \lambda \alpha e^{(\frac{e^{-\beta^T z_i t_i}}{\alpha})^\gamma}) \right] \\ &\quad - \sum_{i=1}^n [\alpha \lambda \gamma z_{i1} \left(\frac{e^{-\beta^T z_i t_i}}{\alpha} \right)^\gamma e^{(\frac{e^{-\beta^T z_i t_i}}{\alpha})^\gamma}] \end{aligned}$$

⋮

$$\begin{aligned} \frac{\partial l(\theta)}{\partial \beta_m} &= \sum_{i=1}^n \delta_i \left[-\gamma z_{im} + \gamma z_{im} \left(\frac{e^{-\beta^T z_i t_i}}{\alpha} \right)^\gamma (1 - \lambda \alpha e^{(\frac{e^{-\beta^T z_i t_i}}{\alpha})^\gamma}) \right] \\ &\quad - \sum_{i=1}^n [\alpha \lambda \gamma z_{im} \left(\frac{e^{-\beta^T z_i t_i}}{\alpha} \right)^\gamma e^{(\frac{e^{-\beta^T z_i t_i}}{\alpha})^\gamma}] \end{aligned}$$

3.4.2 Simulations

Pour confirmer la propriété de la convergence des estimateurs du maximum de vraisemblance, nous calculons l'erreur quadratique moyenne pour les EMV simulés des paramètres du modèle **EW-AFT**.

Supposons que le modèle **EW-AFT** est considéré. Les données ont été simulées $N = 10,000$ fois, avec les valeurs suivantes des paramètres : ($\alpha = 3$; $\gamma = 2$; $\lambda = 2$; $\beta_0 = 0.5$; $\beta_1 = 0.8$). 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 $\hat{\alpha}$, $\hat{\gamma}$, $\hat{\lambda}$, $\hat{\beta}_0$ et $\hat{\beta}_1$ et leurs erreurs quadratiques moyennes (**SME**) sont calculées et données

dans le tableau 8 (les tailles des échantillons considérés sont : $n = 30$; $n = 50$; $n = 100$; $n = 250$; $n = 500$).

$N = 10,000$	$n = 30$	$n = 50$	$n = 100$	$n = 250$	$n = 500$
$\hat{\alpha}$	2.8985	2.9254	2.9523	2.9785	3.0026
MSE	0.0032	0.0028	0.0021	0.0016	0.0004
$\hat{\gamma}$	1.8654	1.9917	2.0693	2.0887	2.0015
MSE	0.0035	0.0013	0.0045	0.0026	0.0004
$\hat{\lambda}$	1.8924	1.9629	1.9854	1.9896	1.9999
MSE	0.0029	0.0025	0.0019	0.0013	0.0008
$\hat{\beta}_0$	0.5472	0.5319	0.5254	0.5102	0.5019
MSE	0.0039	0.0026	0.0019	0.0012	0.0006
$\hat{\beta}_1$	0.8265	0.8141	0.8154	0.8045	0.8012
MSE	0.0025	0.0017	0.0012	0.0009	0.0005

Tab .8. Erreurs quadratiques moyennes des estimateurs du maximum de vraisemblance $\hat{\alpha}$, $\hat{\gamma}$, $\hat{\lambda}$, $\hat{\beta}_0$ et $\hat{\beta}_1$.

Les résultats de la simulation confirment le fait bien connu que pour les modèles réguliers, les estimateurs du maximum de vraisemblance sont convergents.

3.5 Test d'ajustement pour le modèle EW-AFT en cas de données censurées

Dans cette section, nous allons développer une statistique de test d'ajustement pour le modèle EW-AFT 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.

Considérons l'hypothèse nulle :

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

Où

$$F_0(t, \theta) = 1 - \exp[-\lambda\alpha(\exp\{\frac{\exp\{-\beta^T z\}}{\alpha}t\}^\gamma - 1)]$$

est la fonction de répartition du modèle **EW-AFT**, et $\theta = (\alpha, \gamma, \lambda, \beta_0, \beta_1)^T$. Prenons un intervalle de temps fini $[0, \tau]$ et divisons-le en $r > s$ intervalles $I_j = (a_{j-1}, a_j]$

où

$$0 = a_0 < a_1 < \dots < a_{r-1} < a_r = \tau$$

3.5.1 Choix des intervalles

Les valeurs estimées \hat{a}_j des bornes des intervalles choisis pour le groupement des données pour le modèle EW-AFT sont données par la formule suivante :

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

pour $j = 1, \dots, r-1$; $\hat{a}_r = t_{(n)}$,

où

$$E_j = \frac{j}{r-1} \sum_{i=1}^n [-\ln(\exp[-\hat{\lambda}\hat{\alpha}((\frac{\exp\{-\beta^T z\}}{\hat{\alpha}}t_i)^\gamma - 1)))]$$

et

$$E_r = \sum_{i=1}^n [-\ln(\exp[-\hat{\lambda}\hat{\alpha}((\frac{\exp\{-\beta^T z\}}{\hat{\alpha}}t_i)^\gamma - 1)))]$$

Pour un tel choix d'intervalles, nous avons $e_j = \frac{E_r}{r}$ pour tout $j = 1, \dots, r$.

3.5.2 Calcul de la matrice \widehat{W}

Les éléments de la matrice estimée \widehat{W} définie par :

$$\widehat{W}_l = \sum_{j=1}^r \widehat{C}_{lj} \widehat{A}_j^{-1} Z_j, \quad l = 1, 2, \dots, 5. \quad j = 1, \dots, r$$

sont obtenus comme suit :

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

$$\begin{aligned}\widehat{C}_{2j} &= \frac{1}{n} \sum_{i:t_i \in I_j} \delta_i \left[\frac{1}{\widehat{\gamma}} - \ln(\widehat{\alpha}) - \beta^T z_i + \ln(t_i) + \left(\frac{e^{-\beta^T z_i t_i}}{\widehat{\alpha}} \right)^{\widehat{\gamma}} \ln \left(\frac{e^{-\beta^T z_i t_i}}{\widehat{\alpha}} \right) [1 - \widehat{\lambda} \widehat{\alpha} e^{(\frac{e^{-\beta^T z_i t_i}}{\widehat{\alpha}})^{\widehat{\gamma}}}] \right] \\ \widehat{C}_{3j} &= \frac{1}{n} \sum_{i:t_i \in I_j} \delta_i \left[\frac{1}{\widehat{\lambda}} - \widehat{\alpha} (e^{(\frac{e^{-\beta^T z_i t_i}}{\widehat{\alpha}})^{\widehat{\gamma}}} - 1) \right] \\ \widehat{C}_{4j} &= \frac{1}{n} \sum_{i:t_i \in I_j} \delta_i \left[-\widehat{\gamma} + \widehat{\gamma} \left(\frac{e^{-\beta^T z_i t_i}}{\widehat{\alpha}} \right)^{\widehat{\gamma}} (1 - \widehat{\lambda} \widehat{\alpha} e^{(\frac{e^{-\beta^T z_i t_i}}{\widehat{\alpha}})^{\widehat{\gamma}}}) \right] \\ \widehat{C}_{5j} &= \frac{1}{n} \sum_{i:t_i \in I_j} \delta_i \left[-\widehat{\gamma} z_i + \widehat{\gamma} z_i \left(\frac{e^{-\beta^T z_i t_i}}{\widehat{\alpha}} \right)^{\widehat{\gamma}} (1 - \widehat{\lambda} \widehat{\alpha} e^{(\frac{e^{-\beta^T z_i t_i}}{\widehat{\alpha}})^{\widehat{\gamma}}}) \right]\end{aligned}$$

3.5.3 Calcul de la matrice \widehat{I}

Les composants de la matrice d'information $\widehat{I} = (\widehat{I}_{uv})_{5 \times 5}$ sont nécessaires pour le calcul de la matrice G .

Après simplifications, nous obtenons les éléments de la matrice de Fisher estimée :

$$\begin{aligned}\widehat{I}_{11} &= \frac{1}{n} \sum_{i=1}^n \delta_i \left[\frac{1 - \widehat{\gamma} \left[1 + \left(\frac{e^{-\beta^T z_i t_i}}{\widehat{\alpha}} \right)^{\widehat{\gamma}} (1 - \widehat{\lambda} \widehat{\alpha} e^{(\frac{e^{-\beta^T z_i t_i}}{\widehat{\alpha}})^{\widehat{\gamma}}}) \right]}{\widehat{\alpha}} - \widehat{\lambda} (e^{(\frac{e^{-\beta^T z_i t_i}}{\widehat{\alpha}})^{\widehat{\gamma}}} - 1) \right]^2 \\ \widehat{I}_{22} &= \frac{1}{n} \sum_{i=1}^n \delta_i \left[\frac{1}{\widehat{\gamma}} - \ln(\widehat{\alpha}) - \beta^T z_i + \ln(t_i) + \left(\frac{e^{-\beta^T z_i t_i}}{\widehat{\alpha}} \right)^{\widehat{\gamma}} \ln \left(\frac{e^{-\beta^T z_i t_i}}{\widehat{\alpha}} \right) [1 - \widehat{\lambda} \widehat{\alpha} e^{(\frac{e^{-\beta^T z_i t_i}}{\widehat{\alpha}})^{\widehat{\gamma}}}] \right]^2 \\ \widehat{I}_{33} &= \frac{1}{n} \sum_{i=1}^n \delta_i \left[\frac{1}{\widehat{\lambda}} - \widehat{\alpha} (e^{(\frac{e^{-\beta^T z_i t_i}}{\widehat{\alpha}})^{\widehat{\gamma}}} - 1) \right]^2 \\ \widehat{I}_{44} &= \frac{1}{n} \sum_{i=1}^n \delta_i \left[-\widehat{\gamma} + \widehat{\gamma} \left(\frac{e^{-\beta^T z_i t_i}}{\widehat{\alpha}} \right)^{\widehat{\gamma}} (1 - \widehat{\lambda} \widehat{\alpha} e^{(\frac{e^{-\beta^T z_i t_i}}{\widehat{\alpha}})^{\widehat{\gamma}}}) \right]^2 \\ \widehat{I}_{55} &= \frac{1}{n} \sum_{i=1}^n \delta_i \left[-\widehat{\gamma} z_i + \widehat{\gamma} z_i \left(\frac{e^{-\beta^T z_i t_i}}{\widehat{\alpha}} \right)^{\widehat{\gamma}} (1 - \widehat{\lambda} \widehat{\alpha} e^{(\frac{e^{-\beta^T z_i t_i}}{\widehat{\alpha}})^{\widehat{\gamma}}}) \right]^2 \\ \widehat{I}_{12} &= \frac{1}{n} \sum_{i=1}^n \delta_i \left[\widehat{\alpha}^{-1} \left(-1 - \left(\frac{e^{-\beta^T z_i t_i}}{\widehat{\alpha}} \right)^{\widehat{\gamma}} [\widehat{\gamma} \ln \left(\frac{e^{-\beta^T z_i t_i}}{\widehat{\alpha}} \right) - 1] \right) \right. \\ &\quad \left. + w(t_i, \widehat{\theta}) \left[-1 + 2\widehat{\gamma} \left(\frac{e^{-\beta^T z_i t_i}}{\widehat{\alpha}} \right)^{\widehat{\gamma}} + \widehat{\gamma} + \frac{1}{\ln \left(\frac{e^{-\beta^T z_i t_i}}{\widehat{\alpha}} \right)} \right] \right] \\ \widehat{I}_{13} &= \frac{1}{n} \sum_{i=1}^n \delta_i \left[1 + \left(\widehat{\gamma} \left(\frac{e^{-\beta^T z_i t_i}}{\widehat{\alpha}} \right)^{\widehat{\gamma}} - 1 \right) e^{(\frac{e^{-\beta^T z_i t_i}}{\widehat{\alpha}})^{\widehat{\gamma}}} \right]\end{aligned}$$

$$\begin{aligned}\widehat{I}_{14} &= \frac{1}{n} \sum_{i=1}^n \delta_i [-\hat{\alpha}^{-1} \hat{\gamma}^2 \left(\frac{e^{-\beta^T z_i t_i}}{\hat{\alpha}} \right)^{\hat{\gamma}} + m(t_i, \hat{\theta}) [-1 + 2\hat{\gamma} \left(\frac{e^{-\beta^T z_i t_i}}{\hat{\alpha}} \right)^{\hat{\gamma}} + \hat{\gamma}]] \\ \widehat{I}_{15} &= \frac{1}{n} \sum_{i=1}^n \delta_i [-\hat{\alpha}^{-1} \hat{\gamma}^2 z_i \left(\frac{e^{-\beta^T z_i t_i}}{\hat{\alpha}} \right)^{\hat{\gamma}} + m(t_i, \hat{\theta}) [-z_i + 2\hat{\gamma} z_i \left(\frac{e^{-\beta^T z_i t_i}}{\hat{\alpha}} \right)^{\hat{\gamma}} + \hat{\gamma} z_i]] \\ \widehat{I}_{23} &= \frac{1}{n} \sum_{i=1}^n \delta_i [-\hat{\alpha} \left(\frac{e^{-\beta^T z_i t_i}}{\hat{\alpha}} \right)^{\hat{\gamma}} e^{(\frac{e^{-\beta^T z_i t_i}}{\hat{\alpha}})^{\hat{\gamma}}} \ln \left(\frac{e^{-\beta^T z_i t_i}}{\hat{\alpha}} \right)] \\ \widehat{I}_{24} &= \frac{1}{n} \sum_{i=1}^n \delta_i [-1 + \left(\frac{e^{-\beta^T z_i t_i}}{\hat{\alpha}} \right)^{\hat{\gamma}} [\hat{\gamma} \ln \left(\frac{e^{-\beta^T z_i t_i}}{\hat{\alpha}} \right) + 1]] \\ &\quad - \hat{\alpha} m(t_i, \hat{\theta}) [2 \left(\frac{e^{-\beta^T z_i t_i}}{\hat{\alpha}} \right)^{\hat{\gamma}} \ln \left(\frac{e^{-\beta^T z_i t_i}}{\hat{\alpha}} \right) + \ln \left(\frac{e^{-\beta^T z_i t_i}}{\hat{\alpha}} \right) + \hat{\gamma}^{-1}] \\ \widehat{I}_{25} &= \frac{1}{n} \sum_{i=1}^n \delta_i [-z_i + \left(\frac{e^{-\beta^T z_i t_i}}{\hat{\alpha}} \right)^{\hat{\gamma}} [\hat{\gamma} z_i \ln \left(\frac{e^{-\beta^T z_i t_i}}{\hat{\alpha}} \right) + z_i]] \\ &\quad - \hat{\alpha} z_i m(t_i, \hat{\theta}) [2 \left(\frac{e^{-\beta^T z_i t_i}}{\hat{\alpha}} \right)^{\hat{\gamma}} \ln \left(\frac{e^{-\beta^T z_i t_i}}{\hat{\alpha}} \right) + \ln \left(\frac{e^{-\beta^T z_i t_i}}{\hat{\alpha}} \right) + \hat{\gamma}^{-1}] \\ \widehat{I}_{34} &= \frac{1}{n} \sum_{i=1}^n \delta_i [-\hat{\alpha} \hat{\gamma} \left(\frac{e^{-\beta^T z_i t_i}}{\hat{\alpha}} \right)^{\hat{\gamma}} e^{(\frac{e^{-\beta^T z_i t_i}}{\hat{\alpha}})^{\hat{\gamma}}}] \\ \widehat{I}_{35} &= \frac{1}{n} \sum_{i=1}^n \delta_i [-\hat{\alpha} \hat{\gamma} z_i \left(\frac{e^{-\beta^T z_i t_i}}{\hat{\alpha}} \right)^{\hat{\gamma}} e^{(\frac{e^{-\beta^T z_i t_i}}{\hat{\alpha}})^{\hat{\gamma}}}] \\ \widehat{I}_{45} &= \frac{1}{n} \sum_{i=1}^n \delta_i [\hat{\gamma}^2 z_i \left(\frac{e^{-\beta^T z_i t_i}}{\hat{\alpha}} \right)^{\hat{\gamma}} - \hat{\alpha} \hat{\gamma} z_i m(t_i, \hat{\theta}) [2 \left(\frac{e^{-\beta^T z_i t_i}}{\hat{\alpha}} \right)^{\hat{\gamma}} + 1]]\end{aligned}$$

Où

$$w(t_i, \hat{\theta}) = \hat{\lambda} \left(\frac{e^{-\beta^T z_i t_i}}{\hat{\alpha}} \right)^{\hat{\gamma}} e^{(\frac{e^{-\beta^T z_i t_i}}{\hat{\alpha}})^{\hat{\gamma}}} \ln \left(\frac{e^{-\beta^T z_i t_i}}{\hat{\alpha}} \right)$$

Et

$$m(t_i, \hat{\theta}) = \hat{\lambda} \hat{\gamma} \left(\frac{e^{-\beta^T z_i t_i}}{\hat{\alpha}} \right)^{\hat{\gamma}} e^{(\frac{e^{-\beta^T z_i t_i}}{\hat{\alpha}})^{\hat{\gamma}}}$$

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

$$Y_n^2 = \sum_{j=1}^r \frac{(U_j - e_j)^2}{U_j} + W^T \times [\hat{w} - \sum_{j=1}^r \hat{C}_{lj} \hat{C}_{lj} \hat{A}_j^{-1}]^{-1} \times W$$

3.5.4 Simulation (NRR statistique Y_n^2)

Nous avons simulé $N = 10,000$ fois des échantillons du modèle **EW-AFT**, avec différentes valeurs des paramètres et avec des nombres différents de groupement de classes r .

Pour tester l'hypothèse nulle H_0 qu'un échantillon est distribué selon un modèle **EW-AFT**, nous calculons la statistique Y_n^2 des 10,000 échantillons simulés avec des tailles $n = 30$; $n = 100$; $n = 250$; $n = 500$, respectivement.

Ensuite, nous avons représenté les histogrammes des Y_n^2 des échantillons par rapport aux distributions du chi-deux correspondantes avec r degrés de liberté (voir les exemples dans les figures 3.3, 3.4 et 3.5)

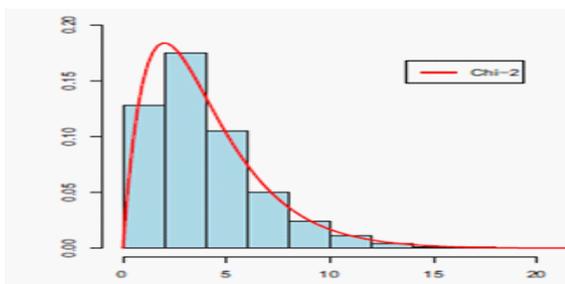


Fig. 3.3. Histogramme de Y^2 pour $n = 30$

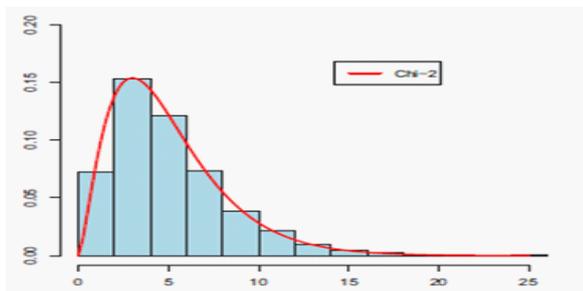


Fig. 3.4. Histogramme de Y^2 pour $n = 150$

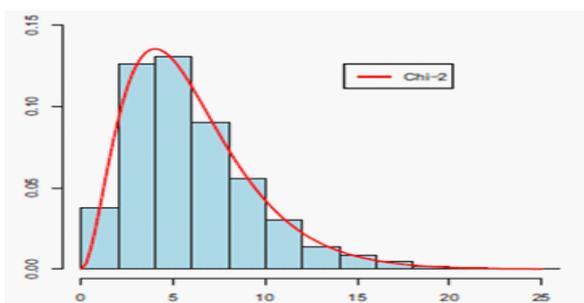


Fig. 3.5. Histogramme de Y^2 pour $n = 400$

3.5. Test d'ajustement pour le modèle EW-AFT en cas de données censurées

Le tableau 9, représente pour chaque risque d'erreur ($\epsilon = 0.02$, 0.05 , 0.01 , 0.1), les résultats des niveaux de signification de la statistique Y_n^2 du test NRR, et leurs valeurs théoriques correspondantes.

$N = 10,000$	$\epsilon = 0.02$	$\epsilon = 0.05$	$\epsilon = 0.01$	$\epsilon = 0.1$
$n = 30$	0.0168	0.0490	0.0068	0.0977
$n = 100$	0.0185	0.0495	0.0076	0.0980
$n = 250$	0.0192	0.0499	0.0090	0.0978
$n = 500$	0.0202	0.0501	0.0097	0.0999

Tab .9. Niveaux de significations simulés de $Y_n^2(\hat{\theta})$ et leurs valeurs théoriques correspondantes.

3.5.5 Analyse de données de fiabilité

Pour appliquer le test construit dans ce chapitre, nous utilisons les données de Nelson (1990) qui représentent 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 extension de Weibull (EW-AFT) dépendant de la variable explicative Z_1 qui représente les niveaux de tension appliqués v_i . Les données sont reportées dans le tableau suivant :

3.5. Test d'ajustement pour le modèle EW-AFT en cas de données censurées

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	0.40, 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.

Tab .10. Données du fluide isolant électrique (Nelson,1990).

Sous l'hypothèse nulle H_0 , et en utilisant le logiciel R, on obtient l'estimateur du maximum de vraisemblance $\hat{\theta}$ du vecteur de paramètres θ :

$$\hat{\theta} = (\hat{\alpha}, \hat{\gamma}, \hat{\lambda}, \hat{\beta}_0, \hat{\beta}_1)^T = (4.7341, 6.2207, 2.0368, 33.2531, -17.5938)^T$$

Si on choisit un nombre de classes $r = 7$, on calcule les éléments de la statistique Y_n^2 qui sont réunis dans le tableau suivant :

\hat{a}_j	1.297	3.87	10.066	21.010	40.913	125.566	2323.700
U_j	17	13	13	6	8	11	8
e_j	9.63	9.63	9.63	9.63	9.63	9.63	9.63
Z_j	0.845	0.386	0.386	-0.416	-0.186	0.157	-0.186
\hat{C}_{1j}	-0.0185	-0.0412	-0.0445	-0.0371	-0.0230	-0.0412	-0.0245
\hat{C}_{2j}	0.5185	0.8663	0.8846	0.8867	0.3158	0.7651	0.4987
\hat{C}_{3j}	0.1371	0.4340	0.4076	0.3402	0.1457	0.4239	0.1907
\hat{C}_{4j}	-0.3298	-0.3949	-0.2545	-0.2787	-0.2112	-0.3666	-0.3087
\hat{C}_{5j}	-0.3193	-0.3567	-0.2876	-0.2329	-0.2398	-0.2442	-0.1768

Tab .11. Calculs de \hat{a}_j ; U_j ; e_j ; Z_j ; \hat{C}_{1j} ; \hat{C}_{2j} ; \hat{C}_{3j} ; \hat{C}_{4j} ; \hat{C}_{5j} .

Le vecteur \hat{W} est alors :

$$\hat{W} = (0.6170 ; -0.0933 ; -1.3528 ; 1.1656 ; 0.0736)^T$$

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

$$Y_n^2 = X^2 + Q = 4.5288 + 6.7649 = 11.2937$$

Au seuil $\epsilon = 0.05$, la valeur critique $\chi_{0.05}^2(7) = 14.0671 \succ Y_n^2 = 11.2937$, on peut donc conclure que ces données de fiabilité sont distribuées selon un modèle *EW - AFT*.

CHAPITRE 4

Modèle à hasard proportionnel (PH)

Sommaire

4.1	Introduction	56
4.2	Principe du modèle	56
4.3	Construction du modèle EW-PH	57
4.4	Estimation du maximum de vraisemblance en cas de données censurées	59
4.4.1	estimateurs du maximum de vraisemblance	59
4.4.2	Simulations	61
4.5	Test d'ajustement pour le modèle EW-PH en cas de données censurées	62
4.5.1	Choix des intervalles	62
4.5.2	Calcul de la matrice \widehat{W}	63
4.5.3	Calcul de la matrice \widehat{I}	63
4.5.4	Simulation de la distribution de Y_n^2	65
4.5.5	Simulation de la statistique Y_n^2	66
4.5.6	Application à des données de survie	66

4.1 Introduction

Le modèle à risques proportionnels (PH) de Cox (1972) largement répandu en analyse de survie, particulièrement en épidémiologie et aussi en industrie pour les tests de vie accélérés (*ALT*) est un modèle de régression. Dans ces modèles, l'objectif est de modéliser le logarithme du risque instantané ou taux de hasard en fonction d'un ensemble de variables explicatives pour évaluer l'effet de ces covariables sur les fonctions d'intérêt. Ils 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.

La principale propriété de ce modèle, à l'origine du nom de "proportional hazards" est que le ratio des taux de panne de deux individus est constant dans le temps.

Dans la littérature statistique, pour tester si un facteur est réellement explicatif, on utilise la statistique de Wald, ou bien le rapport de vraisemblance partielle LR (Likelihood Ratio).

Dans ce chapitre, on introduit un modèle PH dont la fonction de risque est celle d'une distribution extension de Weibull. On y expose d'abord les caractéristiques du modèle, ensuite on calcule les estimateurs du maximum de vraisemblance des paramètres inconnus dans le cas de censure droite. Finalement, pour vérifier si un jeu de données provient de ce modèle, on construit un test d'ajustement du type du chi-deux modifié dans le cas de paramètres inconnus et de censure droite, en utilisant l'approche proposée par Bagdonavicius et Nikulin (2011), et décrite dans le chapitre deux de cette thèse.

4.2 Principe du modèle

Soit un échantillon constitué de n individus indexés par i ($i = 1, \dots, n$). Nous posons $z = (z_{i,1}, \dots, z_{i,m})'$ le vecteur des m variables explicatives pour l'individu i et $\beta = (\beta_1, \dots, \beta_m)'$ le vecteur des coefficients de régression associés à ces variables. Le modèle paramétrique à

risques proportionnels s'écrit :

$$h(t/z_i) = h_0(t) \exp \{ \beta^T z_i \},$$

où $h_0(t)$ est la fonction de risque instantané de base ou taux de risque, où toutes les variables explicatives étant nulles (population de référence). La fonction $h_0(t)$ peut être prédéfinie, dans ce cas on parle de modèle paramétrique de Cox, sinon on a un modèle semi-paramétrique de Cox.

Une hypothèse importante à vérifier dans le modèle de Cox est l'hypothèse de proportionnalité des risques qui sous-entend que le rapport des risques instantanés (ou « hazard ratio (**HR**) » en anglais) est constant au cours du temps.

Soient deux individus i et j dont les variables explicatives sont z_i et z_j respectivement, le **HR** associé à la variable explicative z pour ces individus s'écrit de la façon suivante :

$$HR = \frac{h_0(t) \exp \{ \beta^T z_i \}}{h_0(t) \exp \{ \beta^T z_j \}} = \frac{\exp \{ \beta^T z_i \}}{\exp \{ \beta^T z_j \}} = \exp \{ \beta^T (z_i - z_j) \}$$

4.3 Construction du modèle EW-PH

Sous le modèle paramétrique extension de Weibull à hasard proportionnel (*EW - PH*), la fonction de risque d'un patient particulier avec covariables $(z_{i,1}, \dots, z_{i,m})$ est donnée par :

$$\begin{aligned} h(t/z) &= \lambda \gamma \left(\frac{t}{\alpha}\right)^{\gamma-1} \exp\left\{\left(\frac{t}{\alpha}\right)^\gamma\right\} \exp \{ \beta^T z_i \} \\ h(t/z) &= \lambda \exp \{ \beta^T z_i \} \gamma \left(\frac{t}{\alpha}\right)^{\gamma-1} \exp\left\{\left(\frac{t}{\alpha}\right)^\gamma\right\} \end{aligned}$$

Nous pouvons considérer que le temps de survie de ce patient suit la distribution extension de Weibull avec le paramètre d'échelle $\lambda \exp \{ \beta^T z_i \}$, et γ et α sont les paramètres de forme.

La fonction de hasard cumulé du modèle *EW - PH* est :

$$\begin{aligned} H(t/z) &= H_0(t) \exp \{ \beta^T z_i \} \\ H(t/z) &= \lambda \exp \{ \beta^T z_i \} \alpha \left(e^{\left(\frac{t}{\alpha}\right)^\gamma} - 1 \right) \end{aligned}$$

La fonction de survie correspondante est donnée par :

$$\begin{aligned} S(t/z) &= \exp(-H(t/z)) \\ S(t/z) &= \exp\left(-\lambda \exp \{ \beta^T z_i \} \alpha \left(e^{\left(\frac{t}{\alpha}\right)^\gamma} - 1 \right)\right) \end{aligned}$$

La densité de probabilité est définie par :

$$f(t/z) = h(t/z)S(t/z)$$

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

Des exemples des différentes formes de la densité de la distribution *EW - PH* sont présentés respectivement dans les figures (4.1) et (4.2).

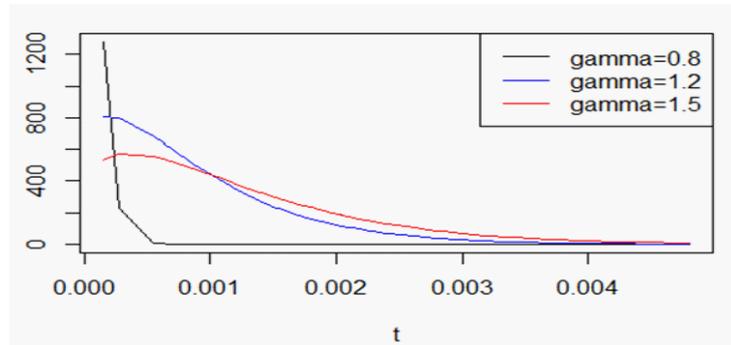


Fig. 4.1. Densité de la distribution EW-PH pour $\gamma(0.8; 1.2; 1.5)$

avec $\alpha = 3.5$, $\lambda = 2$, $\beta_1 = 6$ et $\beta_2 = 0.82$.

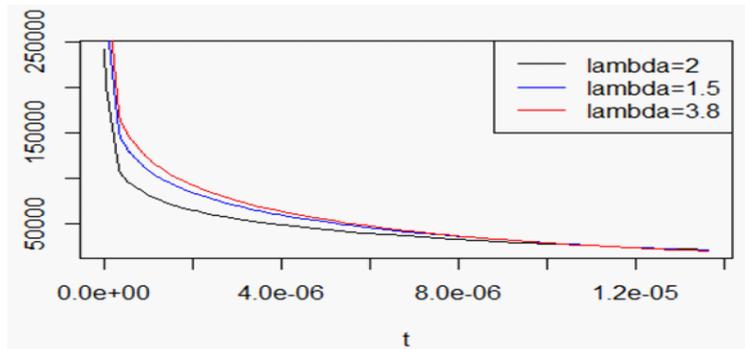


Fig. 4.2. Densité de la distribution EW-PH pour $\gamma = 0.8$, $\beta_1 = 6$ et $\beta_2 = 0.82$

avec $\alpha = 3.5$ et différentes valeurs de λ .

Des graphes du taux de hasard de EW-PH sont présentés respectivement dans les figures (4.3) et (4.4).

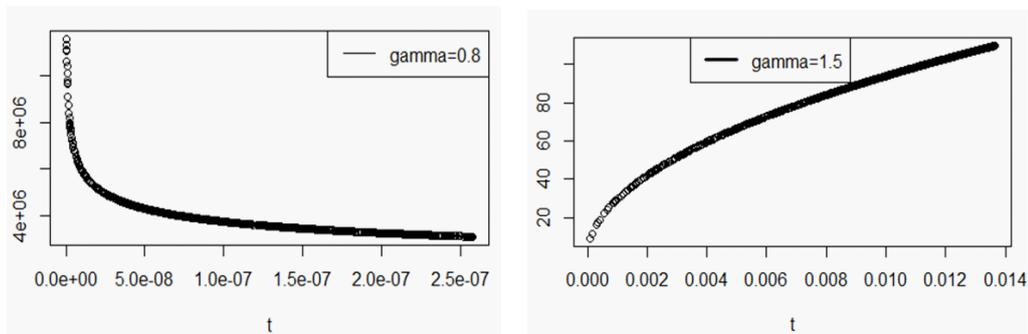


Fig. 4.3. Taux de hasard de la distribution EW-PH pour $\gamma(0.8; 1.5)$

avec $\alpha = 3.5$, $\lambda = 2$, $\beta_1 = 6$ et $\beta_2 = 0.82$.

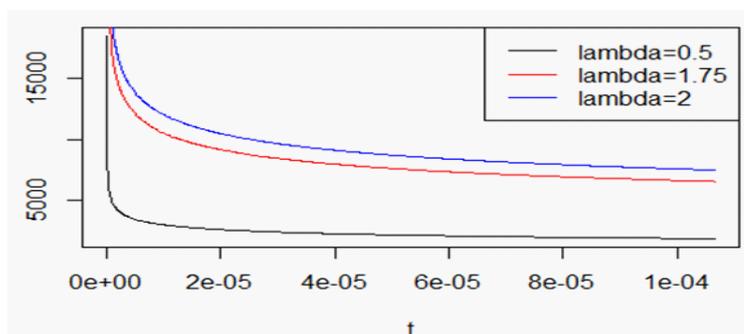


Fig. 4.4. Taux de hasard de la distribution EW-PH pour $\gamma = 0.8$, $\beta_1 = 6$ et $\beta_2 = 0.82$

avec $\alpha = 3.5$ et les différentes valeurs de λ .

4.4 Estimation du maximum de vraisemblance en cas de données censurées

4.4.1 estimateurs du maximum de vraisemblance

Soit T_i une variable aléatoire distribuée selon une loi $EW - PH$ de vecteur de paramètres $\theta = (\alpha, \gamma, \lambda, \beta_1, \dots, \beta_m)^T$.

Supposons que les données se composent de n observations indépendantes $t_i = \min(T_i, C_i)$ pour $i = 1, \dots, n$. La censure C_i étant non informative, la distribution de C_i ne dépend pas des paramètres inconnus de T_i .

La fonction de vraisemblance

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

est équivalente

$$L(t, \theta) = \prod_{i=1}^n h^{\delta_i}(t_i, \theta) S(t_i, \theta), \quad \delta_i = 1_{\{T_i \leq C_i\}}$$

$$L(t, \theta) = \prod_{i=1}^n [\lambda \exp \{\beta^T z_i\} \gamma \left(\frac{t_i}{\alpha}\right)^{\gamma-1} \exp\left\{\left(\frac{t_i}{\alpha}\right)^\gamma\right\}^{\delta_i} [\exp(-\lambda \exp \{\beta^T z_i\} \alpha (e^{(\frac{t_i}{\alpha})^\gamma} - 1))]]$$

La fonction de la log-vraisemblance dans ce cas est :

$$l(t, \theta) = \sum_{i=1}^n \delta_i [\ln(\lambda) + \beta^T z_i + \ln(\gamma) + (\gamma - 1) \ln\left(\frac{t_i}{\alpha}\right) + \left(\frac{t_i}{\alpha}\right)^\gamma] - \sum_{i=1}^n [\lambda \alpha \exp \{\beta^T z_i\} (e^{(\frac{t_i}{\alpha})^\gamma} - 1)]$$

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

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

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

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

$$\frac{\partial l(\theta)}{\partial \beta_1} = \sum_{i=1}^n \delta_i [z_{i1}] - \sum_{i=1}^n [\lambda \alpha z_{i1} \exp \{\beta^T z_i\} (e^{(\frac{t_i}{\alpha})^\gamma} - 1)]$$

⋮

$$\frac{\partial l(\theta)}{\partial \beta_m} = \sum_{i=1}^n \delta_i [z_{im}] - \sum_{i=1}^n [\lambda \alpha z_{im} \exp \{\beta^T z_i\} (e^{(\frac{t_i}{\alpha})^\gamma} - 1)]$$

Pour déterminer les estimateurs du maximum de vraisemblance, on résout le système en égalant les fonctions de score obtenues à zéro. Les formes explicites ne peuvent être obtenues, on utilise alors les méthodes numériques.

4.4.2 Simulations

Pour confirmer la propriété de la convergence des estimateurs du maximum de vraisemblance, nous calculons l'erreur quadratique moyenne pour les EMV simulés des paramètres du modèle $EW - PH$. Pour cela, on considère le modèle $EW - PH$. Les données ont été simulées $N = 10,000$ fois, avec les valeurs suivantes des paramètres : ($\alpha = 2$; $\gamma = 1.5$; $\lambda = 3.5$; $\beta_1 = 6$; $\beta_2 = 0.86$). Nous utilisons le logiciel statistique R et l'algorithme Barzilai-Borwein (**BB**) (Ravi, 2009)). Les valeurs moyennes des estimateurs du maximum de vraisemblance $\hat{\alpha}$, $\hat{\gamma}$, $\hat{\lambda}$, $\hat{\beta}_1$ et $\hat{\beta}_2$ et leurs erreurs quadratiques moyennes (**SME**) sont calculées et données dans le tableau 11 pour les tailles d'échantillon $n = 30$; $n = 50$; $n = 100$; $n = 250$; $n = 500$.

$N = 10,000$	$n = 30$	$n = 50$	$n = 100$	$n = 250$	$n = 500$
$\hat{\alpha}$	1.8932	1.8983	1.8896	1.9934	2.0007
MSE	0.0512	0.0222	0.0693	0.0019	0.0008
$\hat{\gamma}$	1.5433	1.5295	1.4785	1.4921	1.5001
MSE	0.0062	0.0008	0.0010	0.0002	$3.12.e^{-06}$
$\hat{\lambda}$	3.6429	3.5879	3.5427	3.5078	3.4953
MSE	0.0204	0.0077	0.00205	$5.27.e^{-06}$	$2.14.e^{-06}$
$\hat{\beta}_1$	6.2287	6.1425	6.0855	6.0356	5.9973
MSE	0.0367	0.0096	0.0034	0.0004	$5.65.e^{-06}$
$\hat{\beta}_2$	0.8867	0.8814	0.8488	0.8715	0.8622
MSE	0.0145	0.0017	0.0038	0.0002	$2.76.e^{-06}$

Tab .11. Erreurs quadratiques moyennes des estimateurs du maximum de vraisemblance $\hat{\alpha}$, $\hat{\gamma}$, $\hat{\lambda}$, $\hat{\beta}_1$ et $\hat{\beta}_2$.

4.5 Test d'ajustement pour le modèle EW-PH en cas de données censurées

Considérons l'hypothèse nulle :

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

Où

$$F_0(t, \theta) = 1 - \exp(-\lambda \exp\{\beta^T z_i\} \alpha(e^{(\frac{t}{\alpha})^\gamma} - 1))$$

est la fonction de répartition du modèle *EW-PH* de vecteur des paramètres $\theta = (\alpha, \gamma, \lambda, \beta_1, \beta_2)^T$.

Sans perdre de généralité, on considère le vecteur de régression $\beta^T z_i = \beta_1 z_{i1} + \beta_2 z_{i2}$.

Pour valider cette hypothèse H_0 , on utilise la statistique Y_n^2 développée plus haut. Prenons un intervalle de temps fini $[0, \tau]$ et divisons-le en $r > s$ intervalles $I_j = (a_{j-1}, a_j]$ dans lesquels on regroupe les données observées $t_i (i = 1, \dots, n)$

où

$$0 = a_0 < a_1 < \dots < a_{r-1} < a_r = t_{(n)}$$

4.5.1 Choix des intervalles

Le principe du test est que les nombres de défaillances théoriques doivent être tous égaux. Pour vérifier cette condition, le choix des estimateurs des bornes des intervalles \hat{a}_j pour le modèle *EW-PH* est obtenu comme suit :

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

pour $j = 1, \dots, r-1$; $\hat{a}_r = t_{(n)}$,

où

$$E_j = \frac{j}{r-1} \sum_{i=1}^n [\hat{\lambda} \exp\{\beta^T z_i\} \hat{\alpha} (e^{(\frac{t_i}{\hat{\alpha}})^\gamma} - 1)]$$

et

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

Pour un tel choix d'intervalles, nous avons $e_j = \frac{E_r}{r}$ pour tout $j = 1, \dots, r$.

4.5.2 Calcul de la matrice \widehat{W}

Les éléments de la matrice estimée \widehat{W} définie par :

$$\widehat{W}_l = \sum_{j=1}^r \widehat{C}_{lj} \widehat{A}_j^{-1} Z_j, \quad l = 1, 2, \dots, 5. \quad j = 1, \dots, r$$

sont obtenus de la manière suivante :

$$\begin{aligned} \widehat{C}_{1j} &= \frac{1}{n} \sum_{i:t_i \in I_j} \delta_i \left[\frac{-\hat{\gamma} \left(1 + \left(\frac{t_i}{\hat{\alpha}}\right)^{\hat{\gamma}}\right) + 1}{\hat{\alpha}} \right] \\ \widehat{C}_{2j} &= \frac{1}{n} \sum_{i:t_i \in I_j} \delta_i \left[\frac{1}{\hat{\gamma}} + \left(1 + \left(\frac{t_i}{\hat{\alpha}}\right)^{\hat{\gamma}}\right) \ln\left(\frac{t_i}{\hat{\alpha}}\right) \right] \\ \widehat{C}_{3j} &= \frac{1}{n} \sum_{i:t_i \in I_j} \delta_i \left[\frac{1}{\hat{\lambda}} \right] \\ \widehat{C}_{4j} &= \frac{1}{n} \sum_{i:t_i \in I_j} \delta_i z_{i1} \\ \widehat{C}_{5j} &= \frac{1}{n} \sum_{i:t_i \in I_j} \delta_i z_{i2} \end{aligned}$$

4.5.3 Calcul de la matrice \widehat{I}

Les composants de la matrice d'information $\widehat{I} = (\hat{i}_{ll'})_{5 \times 5}$ estimée sont obtenus par

$$\begin{aligned} \widehat{I}_{11} &= \frac{1}{n} \sum_{i=1}^n \delta_i \left[\frac{-\hat{\gamma} \left(1 + \left(\frac{t_i}{\hat{\alpha}}\right)^{\hat{\gamma}}\right) + 1}{\hat{\alpha}} \right]^2 \\ \widehat{I}_{22} &= \frac{1}{n} \sum_{i=1}^n \delta_i \left[\frac{1}{\hat{\gamma}} + \left(1 + \left(\frac{t_i}{\hat{\alpha}}\right)^{\hat{\gamma}}\right) \ln\left(\frac{t_i}{\hat{\alpha}}\right) \right]^2 \\ \widehat{I}_{33} &= \frac{1}{n} \sum_{i=1}^n \delta_i \left[\frac{1}{\hat{\lambda}} \right]^2 \\ \widehat{I}_{44} &= \frac{1}{n} \sum_{i=1}^n \delta_i [z_{i1}]^2 \\ \widehat{I}_{55} &= \frac{1}{n} \sum_{i=1}^n \delta_i [z_{i2}]^2 \end{aligned}$$

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

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

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

$$\widehat{I}_{15} = \widehat{I}_{51} = \frac{1}{n} \sum_{i=1}^n \delta_i \left[\frac{-\hat{\gamma}(1 + (\frac{t_i}{\hat{\alpha}})^{\hat{\gamma}}) + 1}{\hat{\alpha}} \cdot z_{i2} \right]$$

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

$$\widehat{I}_{24} = \widehat{I}_{42} = \frac{1}{n} \sum_{i=1}^n \delta_i \left[\frac{z_{i1}}{\hat{\gamma}} + z_{i1} (1 + (\frac{t_i}{\hat{\alpha}})^{\hat{\gamma}}) \ln(\frac{t_i}{\hat{\alpha}}) \right]$$

$$\widehat{I}_{25} = \widehat{I}_{52} = \frac{1}{n} \sum_{i=1}^n \delta_i \left[\frac{z_{i2}}{\hat{\gamma}} + z_{i2} (1 + (\frac{t_i}{\hat{\alpha}})^{\hat{\gamma}}) \ln(\frac{t_i}{\hat{\alpha}}) \right]$$

$$\widehat{I}_{34} = \widehat{I}_{43} = \frac{1}{n} \sum_{i=1}^n \delta_i \left[\frac{z_{i1}}{\hat{\lambda}} \right]$$

$$\widehat{I}_{35} = \widehat{I}_{53} = \frac{1}{n} \sum_{i=1}^n \delta_i \left[\frac{z_{i2}}{\hat{\lambda}} \right]$$

$$\widehat{I}_{45} = \widehat{I}_{54} = \frac{1}{n} \sum_{i=1}^n \delta_i [z_{i1} \cdot z_{i2}]$$

Comme tous les composants sont calculés, on déduit alors la forme explicite de la statistique Y_n^2 du test

$$Y_n^2 = \sum_{j=1}^r \frac{(U_j - e_j)^2}{U_j} + W^T \times [\hat{i}_W - \sum_{j=1}^r \hat{C}_{lj} \hat{C}_{Uj} \hat{A}_j^{-1}]^{-1} \times W$$

4.5.4 Simulation de la distribution de Y_n^2

Dans cette section, nous montrons par simulations numériques que la distribution de la statistique du test pour le modèle $EW - PH$ est un chi-deux à r degrés de liberté. Nous simulons alors $N = 10,000$ fois des échantillons du modèle $EW - PH$, avec différentes valeurs des paramètres et avec des nombres différents de groupement de classes r .

Pour tester l'hypothèse nulle H_0 qu'un échantillon appartient à $EW - PH$, nous calculons la statistique Y_n^2 des 10,000 échantillons simulés avec respectivement des tailles $n = 30$; $n = 100$; $n = 250$; $n = 500$.

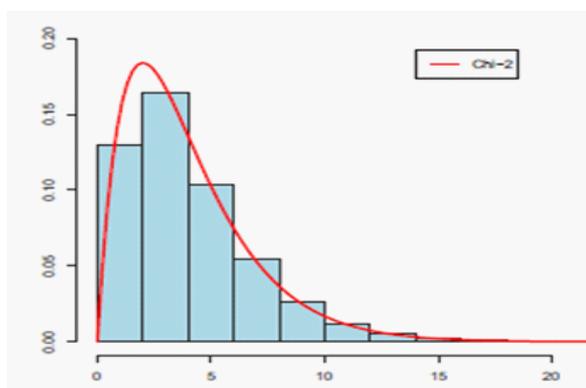


Fig. 4.5. Histogramme de Y^2 pour $n = 30$

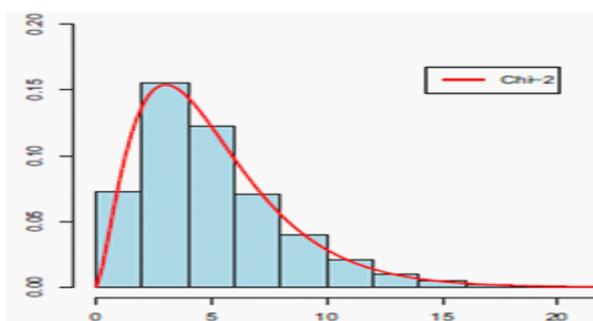


Fig. 4.6. Histogramme de Y^2 pour $n = 150$

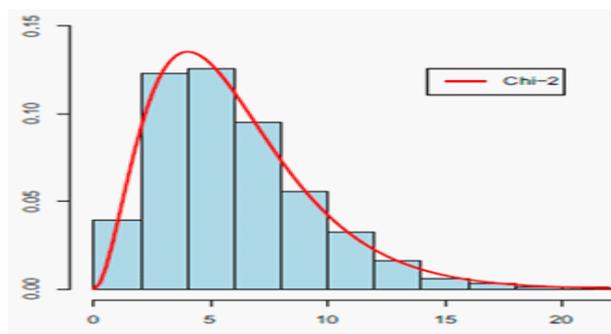


Fig. 4.7. Histogramme de Y^2 pour $n = 400$

4.5.5 Simulation de la statistique Y_n^2

Les valeurs des statistiques Y_n^2 des différents échantillons simulés plus haut sont calculées. Le tableau 12, représente pour chaque risque d'erreur ($\epsilon = 0.02$, 0.05 , 0.01 , 0.1), les résultats des niveaux de signification de la statistique Y_n^2 obtenus et leurs valeurs théoriques correspondantes.

$N = 10,000$	$\epsilon = 0.02$	$\epsilon = 0.05$	$\epsilon = 0.01$	$\epsilon = 0.1$
$n = 30$	0.0151	0.0469	0.0073	0.0968
$n = 100$	0.0172	0.0481	0.0089	0.0983
$n = 250$	0.0190	0.0493	0.0092	0.0993
$n = 500$	0.0198	0.0499	0.0096	0.0998

Tab.12. Niveaux de significations simulés de $Y_n^2(\hat{\theta})$ et les valeurs théoriques correspondantes.

4.5.6 Application à des données de survie

Nous allons appliquer le test proposé dans ce chapitre pour vérifier si des observations provenant d'une analyse de survie sont modélisées par la distribution extension de Weibull à risque proportionnel ($EW - PH$). On considère les données de Freireich et al. (1963), représentant les résultats d'un essai thérapeutique comparant les durées de rémission, en semaines, de patients atteints de leucémie selon qu'ils aient reçu ou non un médicament

4.5. Test d'ajustement pour le modèle EW-PH en cas de données censurées

appelé 6-MP. Les données de durées de rémission dépendent de deux variables explicatives : la covariable binaire Z_1 ($Z_1 = 1$ si le patient a reçu un placebo et $Z_1 = 0$ si le patient a reçu le médicament 6-MP) et la covariable Z_2 qui représente le logarithme du nombre de globules blancs $Z_2 = \log(WBC)$. Les résultats sont reportés dans le tableau suivant :

No	Drug	Length of Remission	log(WBC)	No	Drug	Length of Remission	log(WBC)
1	Placebo	1	2.80	22	6-MP	10	2.96
2	Placebo	5	3.49	23	6-MP	20*	2.01
3	Placebo	22	2.73	24	6-MP	7	4.43
4	Placebo	4	4.36	25	6-MP	19*	2.05
5	Placebo	3	4.01	26	6-MP	32*	2.20
6	Placebo	15	2.30	27	6-MP	6	4.06
7	Placebo	12	1.50	28	6-MP	23	2.57
8	Placebo	8	2.32	29	6-MP	17*	2.16
9	Placebo	8	3.52	30	6-MP	22	2.32
10	Placebo	23	1.97	31	6-MP	35*	1.45
11	Placebo	17	2.95	32	6-MP	6	2.31
12	Placebo	5	3.97	33	6-MP	6	3.28
13	Placebo	2	4.91	34	6-MP	16	3.60
14	Placebo	11	2.12	35	6-MP	13	2.88
15	Placebo	11	3.49	36	6-MP	34*	1.47
16	Placebo	4	2.42	37	6-MP	9*	2.80
17	Placebo	8	3.05	38	6-MP	32*	2.53
18	Placebo	1	5.00	39	6-MP	6*	3.20
19	Placebo	12	3.06	40	6-MP	25*	1.78
20	Placebo	8	3.26	41	6-MP	10*	2.70
21	Placebo	2	4.48	42	6-MP	11*	2.60

Tab .14. Durée de la rémission de 6 – MP versus placebo chez les patients atteints de leucémie (Freireich, 1963).

4.5. Test d'ajustement pour le modèle EW-PH en cas de données censurées

Les nombres suivis du signe * correspondent aux données censurées.

Nous voulons tester l'hypothèse nulle H_0 selon laquelle ces données sont modélisées par une distribution extension de Weibull à hasard proportionnel (EW – PH). En utilisant le logiciel R, on obtient l'estimateur du maximum de vraisemblance $\hat{\theta}$ du vecteur de paramètres θ :

$$\hat{\theta} = (\hat{\alpha}, \hat{\gamma}, \hat{\lambda}, \hat{\beta}_1, \hat{\beta}_2)^T = (2.435, 6.112, 1.788, 1.489, 1.803)^T.$$

Les calculs intermédiaires pour obtenir la statistique du test Y_n^2 , pour un choix du nombre de classes $r = 7$, sont réunis dans le tableau suivant :

\hat{a}_j	3.12	4.34	7.05	10.76	13.09	18.23	23
U_j	5	2	6	5	5	3	4
e_j	4.234	4.234	4.234	4.234	4.234	4.234	4.234
Z_j	0.139	-0.407	0.322	0.139	0.139	-0.225	-0.042
\hat{C}_{1j}	-0.0286	-0.0210	-0.0499	-0.0235	-0.0192	-0.0276	-0.0302
\hat{C}_{2j}	0.4541	0.6115	0.6088	0.7935	0.5107	0.5754	0.6034
\hat{C}_{3j}	0.0333	0.0275	0.0143	0.0365	0.0598	0.0262	0.0313
\hat{C}_{4j}	0.0109	0.0577	0.0342	0.0279	0.0152	0.0225	0.0423
\hat{C}_{5j}	0.0298	0.0731	0.0289	0.0233	0.0154	0.0134	0.0798

Tab .15. Calculs de \hat{a}_j ; U_j ; e_j ; Z_j ; \hat{C}_{1j} ; \hat{C}_{2j} ; \hat{C}_{3j} ; \hat{C}_{4j} ; \hat{C}_{5j} .

Le vecteur \hat{W} est alors :

$$\hat{W} = (0.0672 ; -2.9145 ; 1.0987 ; -0.4556 ; 0.08733)^T$$

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

$$Y_n^2 = X^2 + Q = 7.537 + 1.958 = 9.495$$

Au seuil $\epsilon = 0.05$, la valeur critique du chi-deux $\chi_{0.05}^2(7) = 14.0671$ est supérieure à la valeur du critère de test $Y_n^2 = 9.495$, ainsi l'hypothèse nulle H_0 selon laquelle ces données suivent un modèle *EW - PH* ne peut pas être rejetée.

En analyse de survie ce type de modèle nous permet de calculer le ratio *HR* pour deux individus j_1 et j_2 qui ne diffèrent que par une seule covariable. Par exemple, si on considère que ces deux individus ont le même nombre de globules blancs et que l'un a reçu le traitement 6-MP et l'autre a reçu un placebo, c'est-à-dire qu'ils ne diffèrent par la covariable Z_1 qui vaut 0 chez j_1 et 1 chez j_2 . Alors pour tout t , le rapport des taux de hasard (*HR*) de ces deux individus est égal à $\exp(\hat{\beta}_1)$:

$$HR = \exp(\hat{\beta}_1) = 4.432.$$

Ceci veut dire que les patients qui n'ont pas pris le traitement ont un risque instantané de mourir 4.432 fois plus que ceux qui ont pris le 6-MP.

- les groupes de patients $z_1^1 = 1$ si le patient a reçu un placebo et $z_1^2 = 0$ si le patient a reçu 6-MP
- $\log(WBC)$ est maintenu constant ($z_2^1 = z_2^2 = \log(WBC)$)

$$\begin{aligned} HR &= \frac{h_0(t) \exp \left\{ \hat{\beta}_1 z_1^1 + \hat{\beta}_2 z_2^1 \right\}}{h_0(t) \exp \left\{ \hat{\beta}_1 z_1^2 + \hat{\beta}_2 z_2^2 \right\}} = \frac{\exp \left\{ \hat{\beta}_1 z_1^1 + \hat{\beta}_2 z_2^1 \right\}}{\exp \left\{ \hat{\beta}_1 z_1^2 + \hat{\beta}_2 z_2^2 \right\}} \\ &= \exp \left\{ \hat{\beta}_1 (z_1^1 - z_1^2) + \hat{\beta}_2 (z_2^1 - z_2^2) \right\} \\ &= \exp \left\{ \hat{\beta}_1 (1 - 0) + \hat{\beta}_2 (\log(WBC) - \log(WBC)) \right\} \\ &= \exp(\hat{\beta}_1 + \hat{\beta}_2(0)) \\ &= \exp(\hat{\beta}_1) \end{aligned}$$

Pour mieux comprendre le ratio, on peut aussi considérer par exemple les courbes de survie des deux groupes de patients pour voir l'effet du traitement 6-MP sur la durée de rémission.

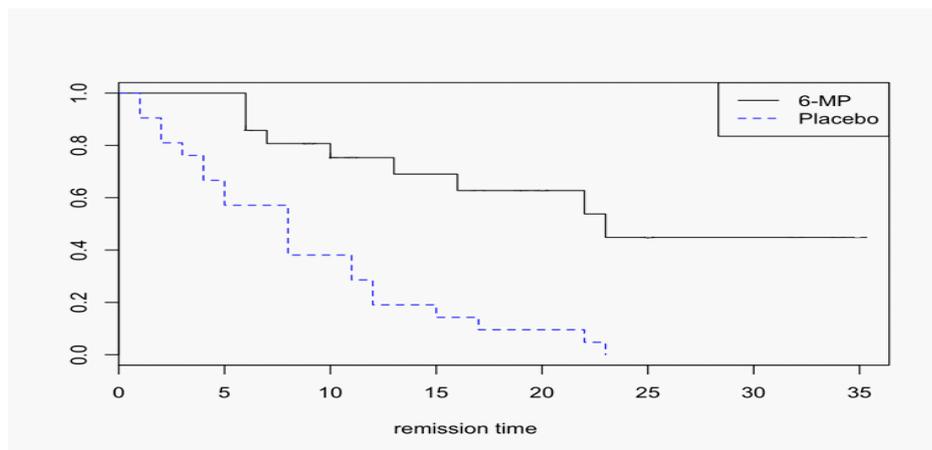


Fig. 4.8. Les courbes de survie en fonction de $\log(WBC)$

Pour ces données, selon la covariable qui représente le nombre de globules blancs $\log(WBC)$, les courbes de survie montrent que le groupe qui a pris le traitement 6-MP a des probabilités de survie plus élevées que le groupe placebo.

Dans ce travail, nous nous sommes intéressés au modèle "*extension de Weibull*" qui est une généralisation de la distribution classique de Weibull et dont la forme mathématique de sa densité de probabilité est maniable permettant ainsi de calculer ses différentes caractéristiques. Ce modèle flexible peut décrire différentes durées de vie provenant de la fiabilité, de l'analyse de survie et d'autres domaines. Basés sur cette distribution, nous avons construit des modèles qui font intervenir des variables de régression qui peuvent avoir des effets directs sur la survie. Le premier modèle est un modèle à temps de vie accéléré **EW-AFT** et le deuxième est un modèle paramétrique à risque proportionnel **EW-PH**.

Comme les résultats de n'importe quelle analyse statistique dépendent du modèle choisi, alors, nous avons construit des tests d'ajustement du type du chi-deux modifié pour permettre aux utilisateurs de vérifier l'adéquation de leurs observations à ces types de distributions. Les tests utilisés tiennent compte des paramètres inconnus des modèles, de la censure droite présente généralement dans les études de fiabilité et d'analyse de survie, et utilisent toute l'information apportée par l'échantillon.

En perspectives, nous proposons des généralisations de ce modèle en utilisant les méthodes d'exponentiation ou en se basant sur d'autres fonctions génératrices de densité et en déduire leurs modèles correspondants à temps de vie accéléré **AFT** et à risque proportionnel **PH**.

- [1] Akritas, G. Michael (1988), Pearson-Type Goodness-of-Fit Tests : The Univariate Case, *Journal of the American Statistical Association*, 83 (401) : 222 – 230.
- [2] Aarset, MV. (1987), How to identify bathtub hazard rate, *IEEE Trans Reliab*, 36 : 106 – 108.
- [3] Bagdonavičius, V., Nikulin, M. (2002), *Accelerated Life Models : Modeling and Statistical Analysis*, Chapman and Hall.
- [4] 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.
- [5] 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.
- [6] 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.
- [7] Bebbington, M., Lai, C. D., & Zitikis, R. (2007a), A flexible Weibull extension, *Reliability Engineering and System Safety*, 92, 719 – 726.

- [8] Bebbington, M., Lai, C. D., & Zitikis, R. (2007*b*), Modeling human mortality using mixtures of bathtub shaped failure rate distributions, *Journal of Theoretical Biology*, 245, 528 – 538.
- [9] Bolshev, L.N., Smirnov, N.V. (1983), *Tables of Mathematical Statistics*. Science, Moscow, [in Russian].
- [10] Chen, Z. (2000), A new two-parameter lifetime distribution with bathtub shape or increasing failure rate function, *Statistics and Probability Letters* , 49, 155 – 161.
- [11] 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.
- [12] Coolen FPA, Yan KJ. (2004), Non parametric predictive inference with right-censored data, *J.Stat.Plan Infer.* 126 : 25 – 54.
- [13] Cox, DR. (1972), Regression models and life table (with discussion), *Journal of the Royal Statistical Society, Series B* 34, p 187 – 202.
- [14] 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. (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.
- [15] Goual, H. and Seddik Ameur, N. (2014), Chi-squared type test for the AFT-generalized inverse Weibull distribution, *Communication in Statistics-Theory and Method*, 43, 13, 2605 – 2617.
- [16] 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.
- [17] Gusmão, F.R.S., Ortega, E.M.M., Cordeiro, G.M. (2009), The generalized inverse Weibull distribution, *Statistical Papers* doi :10.1007/s00362 – 009 – 0271 – 3.

- [18] 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.
- [19] Hjort, N.L. (1990), Goodness of Fit Test in Models for Life History Data Based on Cumulative Hazard Rates, *The annals of statistics*, V.18, N 3, p 1221 – 1258.
- [20] Hougaard, P. (1986), Survival models for heterogeneous populations derived from stable distributions, *Biometrika* 73, 387 – 396.
- [21] Jeong, J. H. (2006), A new parametric family for modelling cumulative incidence function : application to breast cancer data, *Journal of the Royal Statistical Society : Series A*, 169(2), 289 – 303.
- [22] Kalbfleisch, JD, Prentice, R. (1980), *The statistical analysis of failure time data*. Wiley, New York.
- [23] Lai, C. D., Murthy, D. N. P., & Xie, M. (2011), Weibull distributions, *Wiley Interdisciplinary Reviews*, 3(3), 282 – 287.
- [24] Meeker, W.Q., Escobar, L.A. (1998), *Statistical Methods for reliability Data*, John Wiley and Sons INC.
- [25] Mudholkar, G.S., Srivastava, D.K. (1993), Exponentiated Weibull family for analyzing bathtub failure-rate data, *IEEE Trans. Reliab.*, 42, *pp.* 299 – 302.
- [26] Mudholkar, G. S. and Hutson, A. D. (1996), The exponentiated Weibull family : Some properties and a flood data application, *Communications in Statistics-Theory and Methods*, 25(12), 3059 – 3083.
- [27] Mudholkar, G. S., Srivastava, D.K, Kollia, G.D (1996), A generalization of the Weibull distribution with application to the analysis of survival data, *J. Am Stat Assoc*, 91, p. 1575 – 1583.
- [28] Murthy, D. N. P., Xie, M., and Jiang, R. (2004), *Weibull models*, Hoboken : John Wiley & Sons, Inc.
- [29] Nelson, W.(1990), *Accelerated Testing : Statistical Models, Test Plans, and Data Analysis*, Wiley, New York.

- [30] Nikulin, M.S., (1973*a*), Chi-square Test For Continuous Distributions with Shift and Scale Parameters, *teor. Veroyatn. Primen.*, 18, No. 3, p 559 – 568.
- [31] Nikulin, M.S., (1973*b*), On a Chi-square test for continuous distributions, *Theory of Probability and its Applications*, 18, p 638 – 639.
- [32] Nikulin, M.S., (1973*c*), On a Chi-squared test for continuous distributions, *Theory of Probability and its Applications*. vol.18, 3, p 638 – 639.
- [33] Nikulin, M.S., Gerville-Réache, L. and Tran, X. Q. (2013), *Statistical Models and Methods for Reliability and Survival Analysis*, chapter On Chi-Squared Goodness-of-Fit Test for Normality, pages 213 – 227, Wiley Online Library, New York.
- [34] Nikulin, M.S., Gerville-Réache, L., Couallier, V. (2007), *Statistiques des Essais Accélérés*, Hermes : London.
- [35] 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.
- [36] 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.
- [37] 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).
- [38] Rinne, H. (2008), *The Weibull Distribution, A Handbook*, Taylor & Francis.
- [39] Smith, R.M., Bain, L.J. (1975), An exponential power life-testing distribution, *Communication in Statistics*, 4 : 469 – 481.
- [40] Tang, Y., Xie, M., & Goh, T. N. (2003), Statistical analysis of a Weibull extension model. *Communications in Statistics-Theory and Methods*, 32(5), 913 – 928.
- [41] Tang, Y., Xie, M., Goh, T.N. (2004), On changing points of mean residual life and failure rate function for some generalized Weibull distributions, *Reliability Engineering and System Safety* 84, 293 – 299.

BIBLIOGRAPHIE

- [42] Voinov, V., Nikulin, M., Balakrishnan, N. (2013), *Chi-Squared Goodness of Fit Tests with Applications*, Academic Press, Elsevier.
- [43] Xie, M., Goh, T.N., Tang, Y. (2002), A modified Weibull extension with bathtub-shaped failure rate function, *Reliability Engineering and System Safety*, 76, 279 – 285.

