

وزارة التعليم العالي والبحث العلمي

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

BADJI MOKHTAR-ANNABA
UNIVERSITY
UNIVERSITE BADJI MOKHTAR
ANNABA



جامعة باجي مختار
- عنابة -

Faculté des Sciences

Département de Mathématiques

Année : 2024/2025



THÈSE

Présentée en vue de l'obtention du diplôme de Doctorat

Le Titre

Modélisation de données de Fiabilité et Applications

Filière

Mathématiques Appliquées

Spécialité

Probabilités et statistique

Par

Nouadri Rochdi

DIRECTEUR DE THÈSE : SEDDIK AMEUR Nacira

Prof. U.B.M. Annaba

Devant le jury

PRESIDENT: Chadli Assia

Prof. U.B.M - Annaba

EXAMINATEUR : Aidi Khaoula

MCA. U. B.M -Annaba

EXAMINATEUR : Arrar Nawel-Khadidja

MCA. E.N.S.I.A - Alger

EXAMINATEUR: Merzougui Mouna

Prof. U.B.M - Annaba

Table des matières

Remerciements	i
Résumé	ii
Résumé en arabe	iii
Abstract	iv
Introduction	v
1 Estimations et Tests	1
1.1 Méthodes d'estimation	1
1.1.1 Maximum de vraisemblance	1
1.1.2 Information de Fisher	2
1.1.3 Le produit maximum des espacements (MPS)	2
1.1.4 Méthodes du minimum	3
Méthode de Cramér–von Mises	3
Méthode d'Anderson–Darling	4
Méthodes d'Anderson–Darling à droite et à gauche	4
1.1.5 Méthode de Kolmogorov–Smirnov	5
1.1.6 Méthode du percentile	5
1.2 Statistiques de test d'ajustement	6
1.2.1 Statistique de Dzhaparidze-Nikulin	6
1.2.2 Statistique Nikulin-Rao-Robson	7
1.2.3 Statistique de McCulloch	9
1.2.4 Statistique de Singh	9
2 La distribution exponentielle généralisée unitaire	11
2.1 Distribution Exponentielle Généralisée unitaire (UGE)	13

2.2	Estimation	16
2.2.1	Moments	16
2.2.2	Méthode du produit maximum des espacements (MPS)	17
2.2.3	Méthodes d'estimation du type minimum	17
2.3	Validation du modèle <i>UGE</i>	19
2.4	Simulations et Applications	23
2.4.1	Simulations	23
2.4.2	Applications	30
3	La nouvelle distribution à queue lourde	42
3.1	Introduction	42
3.2	Distribution Weighted-ND	44
3.3	Moments et méthodes d'estimation	47
3.4	Validation du modèle	51
3.5	Simulations	53
3.6	Applications	55
3.7	Exemple.1	57
3.8	Exemple.2	59
4	La distribution exponentielle généralisée	62
4.1	La distribution extension exponentielle	63
4.2	Estimation des paramètres	65
4.2.1	Methode de maximum de vraisemblance et Information de Fisher	65
4.2.2	Methode du produit maximum des espacements (MPS)	66
4.2.3	Méthode d'estimation du minimum	67
	Estimation de Cramer-Von Mises	67
	Estimation d'Anderson-Darling	67
	Méthode d'Anderson Darling à droite et à gauche	68
4.2.4	Méthode de Kolmogorov-Smirnov	69
4.3	Validation du modèle	69
4.4	Simulations numériques	70
4.4.1	Estimation :	70
4.4.2	Critères de test	72
4.5	Applications sur des ensembles de données réelles	73
	Conclusion	79

Table des figures

1	Représentations graphiques de la pdf et du taux de hasard de la distribution UGE.	15
2	pdf, CDF, PP-plot et QQ-plot pour les données de la force exercée sur le verre des fenêtres de l'avion.	31
3	Ajustement des distributions avec l'estimation par la méthode 'MLE' pour les valeurs de la force exercée sur le verre de la fenêtre de l'avion.	33
4	Ajustement des distributions avec l'estimation par la méthode 'AD' pour les valeurs de la force exercée sur le verre de la fenêtre de l'avion.	33
5	pdf, cdf, qq-plot et pp-plot du temps de défaillance du turbocompresseur.	38
6	Ajustement des distributions avec l'estimation par la méthode 'MLE' pour le temps de défaillance (103h) du turbocompresseur d'un type de moteur.	38
7	Ajustement des distributions avec l'estimation par la méthode d'Anderson Darling pour le temps de défaillance (103h) du turbocompresseur d'un type de moteur.	39
8	Weighted-ND Density distribution	46
9	<i>weighted</i> – ND Cumulative density distribution	47
10	<i>weighted</i> – ND hazard function	47
11	Ajustement des distributions aux temps de survie de 121 patients atteints de cancer du sein	57
12	Ajustement des distributions aux valeurs des mesures de résistance prises sur 69 fibres de carbone simples	59
13	Fonction de densité de probabilité de la distribution extension exponentielle	64
14	Fonction de risque de la distribution extension exponentielle	64
15	Fonction de densité et fonction de risque de la distribution extension exponentielle avec $\alpha > 1$	65
16	pdf empirique et cdf EE versus distributions alternatives théoriques	74
17	pdf et cdf empiriques versus distributions alternatives	76

Liste des tableaux

1	Valeurs des estimateurs, leur biais et MSE.	24
2	Valeurs des estimateurs, leur biais et MSE.. . . .	25
3	Valeurs des estimateurs, leur biais et MSE.. . . .	26
4	Valeurs de Y^2 et S^2 pour les classes é quiprobables.	28
5	Valeurs de Q^2 et U^2 pour les estimateurs MLE , AD , MPS et CM pour classes de données é quiprobables.	28
6	Valeurs de Q^2 et U^2 pour les estimateurs MLE , AD , MPS et CM dans le cas de classes de données non-è quiprobables.	29
7	Critères de sélection de modèle pour les données de Fuller (1994) pour le modèle UGE par rapport aux alternatives.	32
8	Valeurs des statistiques pour les estimateurs MLE	35
9	Sélection classique de modèles de l'UGE par rapport aux alternatives.	37
10	Valeurs des critères de test pour les estimateurs MLE et AD	40
11	Estimation des paramètres par la méthode MLE	53
12	Estimation des paramètres par la méthode MPS	54
13	Tests statistiques critiques comparant différentes méthodes d'estimation.	58

Remerciements

Cette thèse est l'aboutissement de plusieurs années de travail, de questionnements, de remises en question, mais aussi de découvertes, de rencontres précieuses et de soutiens constants, que je souhaite ici remercier avec sincérité.

Je tiens tout d'abord à exprimer ma profonde gratitude à mon directeur de thèse pour son accompagnement rigoureux et bienveillant tout au long de ce parcours. Sa disponibilité, ses conseils éclairés et sa confiance ont été essentiels à la progression de ce travail, tant sur le plan scientifique que personnel.

Je remercie également les membres du jury pour avoir accepté d'évaluer cette thèse. Leurs observations, remarques et suggestions ont été particulièrement enrichissantes et constructives.

Mes pensées reconnaissantes vont aussi à l'ensemble des enseignants, chercheurs et doctorants pour leur soutien et les échanges stimulants tout au long de ces années.

Un grand merci à mes collègues et compagnons de thèse, pour leur présence, leur humour et pour les moments de complicité qui ont rendu ce parcours plus léger, même dans les périodes les plus intenses.

Je n'oublie pas ma famille, dont le soutien indéfectible a été pour moi une force constante. Merci à mes parents, pour leur patience, leur écoute et leurs encouragements sans faille. Merci également à mes proches, pour leur présence, leur compréhension et leur confiance toujours renouvelée.

Enfin, je tiens à remercier toutes les personnes, proches ou moins proches, qui ont contribué, de près ou de loin, à la réalisation de cette thèse.

Résumé

Ce travail est consacré à l'introduction et l'étude de nouvelles distributions de probabilité capables de mieux décrire différents types de durée de vie de fiabilité, d'analyse de survie et des phénomènes extrêmes, que les distributions classiques existantes. Outre l'utilisation des différentes méthodes pour estimer les paramètres inconnus, on a développé de nouvelles statistiques efficaces et puissantes pour valider les modèles proposés. Les résultats théoriques obtenus sont corroborés par des milliers d'échantillons de données simulées et des applications aux données réelles provenant de plusieurs domaines d'application.

Résumé en arabe

يركز هذا العمل على تقديم ودراسة توزيعات احتمالية جديدة يمكنها نمذجة بيانات الاعتمادية والبقاء بشكل أفضل من التوزيعات الكلاسيكية. بالإضافة إلى استخدام طرق مختلفة لتقدير العلامات المجهولة، قمنا ببناء إحصاءات جديدة فعالة وقوية للتحقق من صحة هذه النماذج الاحتمالية. وقد تم تأكيد النتائج المتحصل عليها من خلال بيانات محاكاة وعينات حقيقية.

Abstract

This work is devoted to the introduction and study of new probability distributions capable of better describing different types of reliability lifetime, survival analysis and extreme phenomena than existing classical distributions. Besides the use of different methods to estimate unknown parameters, new efficient and powerful statistics have been developed to validate the proposed models. The theoretical results obtained are corroborated by thousands of simulated data samples and applications to real data from several application areas.

Introduction

La modélisation des données de fiabilité représente un défi central dans des domaines où la précision statistique se heurte à la complexité croissante des systèmes modernes. Si les distributions classiques ont longtemps servi de fondement à l'analyse des durées de vie, des taux de défaillance ou des événements extrêmes, leurs limites deviennent criantes face à des données asymétriques, à support borné, ou marquées par des queues lourdes. Ces lacunes, couplées aux faiblesses des tests d'ajustement traditionnels, souvent peu puissants ou inadaptés aux paramètres estimés appellent à une refonte méthodologique. Cette thèse répond à ce besoin en proposant une approche innovante, structurée autour de nouvelles distributions, chacune conçue pour capturer des dynamiques spécifiques, et accompagnée de procédures rigoureuses d'estimation et de validation.

L'impulsion initiale de ce travail réside dans la nécessité de modèles flexibles capables de s'adapter à des supports contraints, comme l'intervalle $[0, 1]$, où les lois classiques échouent à décrire des comportements unimodaux ou asymétriques. Une première contribution introduit ainsi une distribution exponentielle généralisée unitaire, combinant la robustesse des modèles existants avec une adaptation aux données normalisées. Pour en valider l'adéquation, des tests d'ajustement modifiés exploitant des estimateurs \sqrt{n} -convergeants et des méthodes d'estimation diversifiées (maximum de vraisemblance, produit des espacements, Cramér–von Mises) sont développés et testés sur des milliers de simulations, confirmant leur supériorité en puissance et en stabilité. Des applications à des données industrielles et environnementales illustrent leur utilité pratique.

Parallèlement, la modélisation des événements extrêmes — des crashes financiers aux pannes critiques — exige des distributions à queues lourdes capables de quantifier des risques rares mais dévastateurs. Une deuxième contribution étend une distribution récente via une pondération inspirée de la méthode d'Azzalini,

offrant une flexibilité inédite pour les données hétérogènes. Comparée aux modèles établis (Weibull, Fréchet, GEV), cette loi démontre sa pertinence à travers des critères d'information (AIC, BIC) et des tests d'ajustement, tout en intégrant des méthodes d'estimation robustes (maximum de vraisemblance, produit maximum des espacements). Ses performances, validées sur des données synthétiques et réelles, en font un outil privilégié pour les secteurs financiers ou les systèmes à haute criticité.

Enfin, un troisième volet se concentre sur la validation rigoureuse d'un modèle capable de décrire des taux de défaillance croissants ou constants, malgré une densité monotone, une lacune majeure des distributions traditionnelles. En proposant des estimateurs adaptés aux petits échantillons (méthode du produit des espacements, Anderson–Darling) et un test d'ajustement basé sur la statistique de Nikulin–Rao–Robson, cette contribution évite les écueils des approches comparatives. Des études de simulation exhaustives, couplées à des applications à des données de précipitations journalières, révèlent non seulement la robustesse du modèle, mais aussi sa capacité à concurrencer les lois établies dans des contextes exigeants.

Au-delà de ces avancées théoriques, cette thèse s'ancre dans une démarche empirique systématique. Chaque modèle est soumis à une batterie de tests, simulations Monte Carlo, comparaisons de biais et d'erreurs quadratiques moyennes, analyses sur jeux de données réelles garantissant sa fiabilité opérationnelle. L'accent est mis sur la reproductibilité des méthodes et leur applicabilité transversale, des systèmes industriels aux études environnementales. En unifiant innovation distributionnelle, rigueur statistique et pragmatisme appliqué, ce travail ouvre des perspectives pour des domaines émergents tels que l'intelligence artificielle explicable ou l'ingénierie de la résilience, où la modélisation fine des incertitudes devient un impératif.

Cette thèse est structurée de la manière suivante :

Après avoir présenté toutes les méthodes classiques d'estimation, on expose la théorie des modifications des tests d'ajustement du type khi-deux dans un premier chapitre.

Le deuxième est dédié à l'introduction d'un nouveau modèle de probabilité, nommé

la distribution exponentielle généralisée unitaire (UGE), à support borné, obtenu à partir de la distribution exponentielle. Après le calcul des différents estimateurs des paramètres inconnus, on a construit plusieurs critères de test puissants et efficaces qui permettent de valider le modèle proposé.

Le chapitre trois concerne l'introduction d'une nouvelle distribution à queue lourde appelée *weighted new-distribution* (WND). Celle-ci décrit des données aussi bien symétriques qu'asymétriques, ce qui lui permet d'élargir ses champs d'application, comme pour les études actuarielles ou d'assurance.

Le quatrième chapitre est consacré à une distribution très intéressante introduite par Nadarajah et Haghghi, dont le mode est égal à zéro alors que le taux de hasard peut être croissant ou décroissant, ce qui n'existe pas dans les distributions classiques. Tout d'abord, on a calculé les estimateurs des paramètres par cinq méthodes différentes ; ensuite, on a développé un test efficace pour ajuster ce modèle.

Enfin, une importante étude par simulations numériques a été menée pour confirmer les résultats obtenus. De plus, des applications sur des données réelles provenant de plusieurs domaines ont été utilisées pour illustrer la flexibilité des modèles proposés et des statistiques construites pour valider ceux-ci.

Estimations et Tests

Dans ce chapitre, on présente différentes méthodes d'estimation telles que le maximum de vraisemblance (MLE), le maximum des espacements (MPS), les méthodes du minimum, Kolmogorov–Smirnov, et la méthode du percentile. Ensuite, on expose la théorie de certains critères de tests d'ajustement modifiant le célèbre test du khi-deux.

1.1 Méthodes d'estimation

1.1.1 Maximum de vraisemblance

Nous considérons un échantillon de n variables indépendantes et identiquement distribuées t_1, t_2, \dots, t_n , issues d'une loi de densité $f(t_i; \theta)$, où θ est le vecteur des paramètres. La fonction de vraisemblance est donnée par :

$$L(t_1, t_2, \dots, t_n; \theta) = \prod_{i=1}^n f(t_i; \theta) \quad (1.1)$$

et la log-vraisemblance est :

$$\ell(\theta) = \ln L(t_1, t_2, \dots, t_n; \theta) = \sum_{i=1}^n \ln f(t_i; \theta)$$

Les estimateurs du maximum de vraisemblance (MLE) sont obtenus en annulant les dérivées premières de la fonction de log-vraisemblance par rapport aux paramètres. Ces estimateurs sont asymptotiquement convergents et suivent une loi normale.

1.1.2 Information de Fisher

L'information de Fisher mesure la quantité d'information qu'une variable aléatoire observée fournit sur un paramètre inconnu du modèle. Elle est définie à partir de la fonction de vraisemblance et joue un rôle fondamental dans la théorie de l'estimation, notamment dans le cadre de l'estimation par maximum de vraisemblance.

Soit T une variable aléatoire suivant une loi de densité $f(t; \theta)$ dépendant d'un paramètre $\theta \in \Theta$. L'information de Fisher, notée $\mathcal{I}(\theta)$, est définie par :

$$\mathcal{I}(\theta) = \mathbb{E} \left[\left(\frac{\partial}{\partial \theta} \ln f(T; \theta) \right)^2 \right]$$

Autrement dit, il s'agit de l'espérance du carré de la dérivée du logarithme de la fonction de vraisemblance par rapport au paramètre θ . Sous certaines conditions de régularité, cette définition est équivalente à :

$$\mathcal{I}(\theta) = -\mathbb{E} \left[\frac{\partial^2}{\partial \theta^2} \ln f(T; \theta) \right]$$

Dans le cas multivarié, lorsque le modèle dépend de plusieurs paramètres $\theta = (\theta_1, \theta_2, \dots, \theta_k)$, l'information de Fisher s'écrit sous forme matricielle :

$$\mathcal{I}(\theta) = [\mathcal{I}_{ij}(\theta)] = -\mathbb{E} \left[\frac{\partial^2}{\partial \theta_i \partial \theta_j} \ln f(T; \theta) \right]$$

Chaque élément $\mathcal{I}_{ij}(\theta)$ correspond à une dérivée seconde croisée de la log-vraisemblance. Cette matrice est symétrique et semi-définie positive. Son inverse, lorsqu'il existe, fournit une approximation asymptotique de la matrice de variance-covariance des estimateurs du maximum de vraisemblance.

1.1.3 Le produit maximum des espacements (MPS)

Introduite par Cheng et Amin (1979, 1983), la méthode d'estimation par le produit des espacements (MPS) est une alternative à la méthode d'estimation par maximum de vraisemblance. Cette méthode a également été proposée par Ramnaby (1984) comme une approximation de la mesure d'information de Kullback-Leibler.

La méthode MPS repose sur les différences entre les valeurs successives croissantes de la fonction de répartition théorique $F(t_i; \theta)$, avec $F(t_0) = 0$ et $F(t_{n+1}) = 1$:

$$P_i(\theta) = F(t_i; \theta) - F(t_{i-1}; \theta) \quad (1.2)$$

Les estimateurs MPS sont obtenus en maximisant la quantité $M(\theta)$:

$$M(\theta) = \frac{1}{n+1} \sum_{i=1}^{n+1} \ln(P_i(\theta)) \quad (1.3)$$

c'est-à-dire les solutions de l'équation :

$$\frac{\partial M(\theta)}{\partial \theta} = \frac{1}{n+1} \sum_{i=1}^{n+1} \frac{D(t_i; \theta) - D(t_{i-1}; \theta)}{P_i} = 0 \quad (1.4)$$

où D est le vecteur des dérivées premières de la fonction de répartition par rapport au vecteur des paramètres inconnus :

$$\frac{\partial D(t_i; \theta)}{\partial \theta} \quad (1.5)$$

Par ailleurs, Cheng et Amin (1983) ont démontré que les estimateurs MPS sont aussi efficaces et convergents que les estimateurs MLE, sous des conditions encore plus générales.

1.1.4 Méthodes du minimum

Méthode de Cramér-von Mises

Parmi les méthodes d'estimation de type minimum proposées dans la littérature statistique, on trouve la méthode de Cramér-von Mises (CM), qui consiste à minimiser la distance entre les fonctions de répartition théorique et empirique du modèle étudié. Cette méthode est reconnue pour produire des estimateurs présentant un biais plus faible que les autres estimateurs de ce type (voir Mac-Donald, 1971). La statistique correspondante est définie par :

$$C(\theta) = \frac{1}{12n} + \sum_{i=1}^n \left(F(t_i; \theta) - \frac{2i-1}{2n} \right)^2 \quad (1.6)$$

Pour calculer les estimateurs $\widehat{\theta}_{CV}$, on annule les dérivées partielles de la statistique $C(\theta)$ par rapport aux paramètres inconnus, ce qui revient à résoudre l'équation suivante :

$$\sum_{i=1}^n \left(F(t_i; \theta) - \frac{2i-1}{2n} \right) D = 0 \quad (1.7)$$

où D est défini comme dans l'équation 1.5.

Méthode d'Anderson–Darling

La méthode d'Anderson–Darling est également une méthode d'estimation de type minimum. Les estimateurs $(\widehat{\alpha}_{AD}, \widehat{\lambda}_{AD})$ sont obtenus en minimisant la célèbre statistique A^2 d'Anderson et Darling (1952). Boos (1981) en a présenté la forme calculatoire suivante :

$$A^2(\alpha, \lambda) = -n - \frac{1}{n} \sum_{i=1}^n (2i-1) [\ln(F(t_i)) + \ln(1 - F(t_{i^*}))] \quad (1.8)$$

avec $i^* = n - i + 1$.

Les estimateurs $(\widehat{\alpha}_{AD}, \widehat{\lambda}_{AD})$ sont alors obtenus en résolvant l'équation suivante :

$$\sum_{i=1}^n (2i-1) \left[\frac{D(t_i)}{F(t_i)} - \frac{D(t_{i^*})}{1 - F(t_{i^*})} \right] = 0 \quad (1.9)$$

où D est défini dans l'équation 1.5.

Méthodes d'Anderson–Darling à droite et à gauche

Luceno (2006) a proposé des modifications de la statistique AD . La plus utilisée est la statistique Anderson–Darling à droite (RAD) pour les distributions avec une queue à droite (voir Dey et al., 2014 ; Louzada et al., 2016 ; Ramos et Louzada, 2016). Les estimateurs $(\widehat{\theta}_{RAD})$ sont obtenus en minimisant :

$$RAD = \frac{n}{2} - 2 \sum_{i=1}^n \ln(F(t_i)) - \frac{1}{n} \sum_{i=1}^n (2i-1) (1 - F(t_{i^*})) \quad (1.10)$$

Ce qui mène à la résolution du système suivant :

$$-2 \sum_{i=1}^n \frac{D(t_i)}{F(t_i)} + \frac{1}{n} \sum_{i=1}^n (2i-1) \frac{D(t_{i^*})}{1 - F(t_{i^*})} = 0 \quad (1.11)$$

Les estimateurs Anderson–Darling à gauche ($\widehat{\theta}_{LAD}$), ou *left-tail Anderson–Darling*, sont obtenus en minimisant :

$$LAD = -\frac{3n}{2} + 2 \sum_{i=1}^n \ln(F(t_i)) - \frac{1}{n} \sum_{i=1}^n (2i-1)(1-F(t_{i^*})) \quad (1.12)$$

1.1.5 Méthode de Kolmogorov–Smirnov

Weber et al. (2006) ont proposé une méthode d'estimation appelée estimation Kolmogorov–Smirnov minimale (KS), dont le but est de minimiser la célèbre statistique de Kolmogorov–Smirnov :

$$KS = d(F_n, F) = \max_{t \in \mathbb{R}} |F_n(t) - F(t)| \quad (1.13)$$

où n est la taille de l'échantillon, $F(t)$ est la fonction de répartition théorique (CDF) de la distribution étudiée, et $F_n(t)$ est la fonction de répartition empirique, une fonction en escalier qui augmente de $1/n$ à chaque donnée observée.

La statistique KS est largement utilisée pour ajuster des données à des distributions continues. Son intérêt réside dans sa simplicité de calcul et sa propriété de distribution libre.

Les composantes D^+ et D^- sont données par :

$$D^+ = \max_{i=1, \dots, n} \left(\frac{i}{n} - F(t_i) \right) \quad (1.14)$$

$$D^- = \max_{i=1, \dots, n} \left(F(t_i) - \frac{i-1}{n} \right) \quad (1.15)$$

$$KS = \max \{ D^+, D^- \} \quad (1.16)$$

Les $t_{(1)}, t_{(2)}, \dots, t_{(n)}$ représentent les statistiques d'ordre. Les paramètres qui maximisent la statistique KS sont notés ($\widehat{\theta}_{KS}$).

1.1.6 Méthode du percentile

Une autre méthode d'estimation consiste à minimiser les différences entre les observations x_i et les quantiles théoriques $Q(u_i)$, par rapport aux paramètres inconnus :

$$U(\theta) = \sum_{i=1}^n \{x_i - Q(u_i)\}^2$$

1.2 Statistiques de test d'ajustement

Pour tester la validité d'un modèle hypothétique, les méthodes couramment utilisées reposent sur les fonctions de distribution cumulative empirique (EDF), les critères classiques de sélection de modèles, ainsi que la statistique de Pearson X^2 .

Comme le test du khi-carré ne peut pas être appliqué lorsque les paramètres sont inconnus, différentes adaptations ont été proposées dans la littérature statistique. Ces tests d'ajustement sont très intéressants, mais ils restent encore insuffisamment développés.

Dans ce travail, nous nous intéressons aux statistiques proposées par Dzhaparidze et Nikulin (1974), Nikulin (1973), Rao et Robson (1974), McCulloch (1985) et Singh (1987).

1.2.1 Statistique de Dzhaparidze-Nikulin

Supposons qu'un échantillon (t_1, t_2, \dots, t_n) , suit une distribution de probabilité $F(t, \theta)$:

$$H_0 : t \sim F(x, \theta) / P(X_i < t) = F(t, \theta)$$

$\theta = (\theta_1, \dots, \theta_s)^T$ est le vecteur de paramètres inconnus.

Si nous regroupons les observations en r classes disjointes $\Delta_j = [a_{j-1}, a_j[$, nous notons :

$$p_j(\theta) = \int_{\Delta_j} dF(x; \theta), \quad j = 1, \dots, r.$$

et $v = (v_1, v_2, \dots, v_r)^T$ le vecteur des fréquences obtenues. Pour tester l'hypothèse nulle H_0 , on utilise généralement la statistique du khi-carré de Pearson Z^2 qui est basée sur les différences entre les fréquences observées v_j et les fréquences attendues np_j des observations dans les classes Δ_j :

$$Z^2 = \sum_{j=1}^r \frac{(v_j - np_j(\theta))^2}{np_j(\theta)}$$

Lorsque la distribution hypothétique est spécifiée, Z^2 suit une distribution du khi-carré avec $(r - 1)$ degrés de liberté. Mais lorsque les paramètres sont inconnus et remplacés par leurs estimations, ce test devient inapplicable. Alors, Dzharidze et Nikulin (1974) ont proposé une nouvelle forme de $Z^2(\theta)$, appelée la statistique DN et notée U^2 . Cette statistique est définie par :

$$U^2(\theta) = Z_n^T(\theta) \left[I - B(\theta) (B^T(\theta)B(\theta))^{-1} B^T(\theta) \right] Z_n(\theta)$$

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

$B(\theta)$ est la matrice $(r \times s)$ avec les composantes b_{jk} :

$$b_{jk} = \frac{1}{\sqrt{p_j(\theta)}} \int_{\Delta_j} \frac{\partial f(x, \theta)}{\partial \theta_k}, \quad j = 1, \dots, r; \quad k = 1, \dots, s$$

s représente le nombre de paramètres estimés et I est la matrice identité. La distribution limite de cette statistique suit une loi du khi-carré avec $r - s - 1$ degrés de liberté. L'avantage de ce test d'ajustement U^2 est que les paramètres inconnus sont remplacés par n'importe quels estimateurs \sqrt{n} -convergeants $\tilde{\theta}$, en plus, il n'y a aucune condition sur le regroupement des données.

1.2.2 Statistique Nikulin-Rao-Robson

D'autre part, Nikulin (1973) et Rao et Robson (1974) ont proposé une modification très intéressante de la statistique de Pearson $Z^2(\theta)$, appelée la statistique NRR (Nikulin-Rao-Robson) et notée Y^2 . Basé sur les estimateurs du maximum de vraisemblance sur des données non groupées, ce test d'ajustement récupère l'information perdue lors du regroupement des données.

Cette statistique notée Y^2 est définie comme suit (voir Greenwood et Nikulin, 1996) :

$$Y^2(\hat{\theta}) = Z_n^T(\hat{\theta}) (I - BI_n^{-1}(\hat{\theta})B^T)^{-1} Z_n(\hat{\theta})$$

$\hat{\theta}$ est l'estimateur du maximum de vraisemblance de θ , I est la matrice identité, $I_n(\hat{\theta})$ est la matrice d'information de Fisher sur les données non groupées, la matrice B est définie ci-dessus, et les observations (T_1, T_2, \dots, T_n) sont regroupées en r classes disjointes et équiprobables Δ'_j , c'est-à-dire :

$$p_j(\theta) = \int_{\Delta'_j} dF(x; \theta) = \frac{1}{r}, \quad \text{pour tout } j = 1, \dots, r.$$

Pour le calcul, Moore et Spruill (1975) réécrivent cette statistique comme la somme de la statistique de Pearson Z^2 et d'une quantité R^2 qui est considérée comme un terme de correction :

$$Y^2(\hat{\theta}) = Z^2(\hat{\theta}) + R^2(\hat{\theta})$$

avec

$$\begin{aligned} Z^2(\hat{\theta}) &= Z_n^T(\hat{\theta}) Z_n(\hat{\theta}) \\ R^2(\hat{\theta}) &= Z_n^T(\hat{\theta}) B(I_n - J_n)^{-1} B^T Z_n(\hat{\theta}) \end{aligned}$$

$$J_n(\theta) = B^T(\theta)B(\theta)$$

J_n représente la matrice d'information de Fisher sur les données groupées.

La statistique $Y^2(\hat{\theta})$ suit une loi du khi-deux avec $(r - 1)$ degrés de liberté, r est le nombre de classes choisies Δ'_j .

Plus tard, Dzhaparidze et Nikulin (1992) ont proposé d'améliorer tout estimateur \sqrt{n} -convergent pour qu'il soit aussi efficace que l'estimateur du maximum de vraisemblance afin qu'il puisse être utilisé pour la statistique NRR .

1.2.3 Statistique de McCulloch

Des années plus tard, McCulloch (1985) a présenté une autre décomposition très intéressante de $Y^2(\theta)$. Il a montré que celle-ci peut s'écrire comme la somme de deux statistiques indépendantes, U^2 , la statistique DN , et S^2 , la statistique de McCulloch :

$$Y^2(\theta) = U^2(\theta) + S^2(\theta)$$

et ceci en utilisant la transformation suivante :

$$[I - BI_n^{-1}B^T]^{-1} = [I - B(B^TB)^{-1}B^T] + B[(I_n - B^TB)^{-1} + (B^TB)^{-1}]B^T$$

S^2 est obtenue comme suit :

$$S_n^2 = Z_n^T(\hat{\theta}) B \left[(I_n(\theta) - J_n(\theta))^{-1} + (B^TB)^{-1} \right] B^T Z_n(\hat{\theta})$$

I_n et J_n sont respectivement les matrices d'information de Fisher pour les données initiales et groupées, et la matrice B est définie comme précédemment.

La statistique $S^2(\theta)$, appelée test de McCulloch (noté *McCu test*), suit une loi du khi-deux avec s degrés de liberté.

Il a été démontré (McCulloch, 1985) que ce test possède la plus faible variance et le taux de convergence le plus élevé, ce qui le rend donc une bonne alternative aux autres tests d'ajustement. En plus, cette statistique permet de récupérer l'information de Fisher perdue lors du regroupement des données.

1.2.4 Statistique de Singh

Comme les estimateurs du maximum de vraisemblance (MLE) ne peuvent pas être obtenus dans la plupart des cas, les paramètres inconnus peuvent être estimés par d'autres méthodes. En se basant sur l'approche de Dzhaparidze et Nikulin (1974), Singh (1987) a proposé de modifier la statistique NRR de manière à ce que tout estimateur \sqrt{n} -convergent $\tilde{\theta}$ puisse être utilisé. Cette statistique de test, notée Q^2 , est définie par :

$$Q_n^2 = Z_*^T(\tilde{\theta}) (I - BI_n^{-1}B^T)^{-1} Z_*(\tilde{\theta})$$

avec

$$Z_*(\tilde{\theta}) = Z(\tilde{\theta}) - BI_n^{-1}W(\tilde{\theta})$$

et

$$W(\tilde{\theta}) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\partial \ln f(x, \theta)}{\partial \theta_k} \Big|_{\theta=\tilde{\theta}}; \quad k = 1, \dots, s$$

La loi limite de la statistique Q^2 suit également un khi-deux avec $r - 1$ degrés de liberté, r est le nombre de classes choisies lors du regroupement des données.

Notez qu'en plus d'utiliser tout estimateur \sqrt{n} -convergent $\tilde{\theta}$, il n'y a pas de restriction sur les classes de regroupement des données pour le calcul de Q^2 . Cette statistique coïncide avec la statistique NRR dans le cas des estimateurs du maximum de vraisemblance.

La forme calculatoire de cette statistique est donnée par :

$$Q_n^2 = Z_*^T(\tilde{\theta}) Z_*(\tilde{\theta}) + (B^T Z_*(\tilde{\theta}))^T (I_n - J_n)^{-1} (B^T Z_*(\tilde{\theta}))$$

La distribution exponentielle généralisée unitaire

Introduction

Depuis le début de ce siècle, plusieurs nouvelles familles de distributions ont été introduites dans la littérature statistique pour décrire les données collectées à partir des études expérimentales. La plupart des nouveaux modèles sont définis sur \mathbb{R}^+ , mais les distributions sur un support fini ne sont pas bien étudiées.

Les distributions de probabilité à support $[0, 1]$ jouent un rôle fondamental dans divers domaines de la statistique, notamment en inférence bayésienne, en théorie des probabilités et dans de nombreuses applications pratiques où les valeurs d'intérêt sont naturellement contraintes entre ces bornes. Ces distributions permettent de modéliser des phénomènes variés, allant des probabilités de succès dans des expériences aléatoires à des modélisations plus complexes dans les domaines de l'apprentissage automatique, de la simulation et de la finance.

Parmi les premières et les plus connues de ces distributions, on trouve la distribution Beta, qui a été introduite au début du XXe siècle par des statisticiens comme Fisher et Edgeworth. Cette loi a été largement adoptée pour modéliser des probabilités dans des contextes où les résultats sont restreints à un intervalle entre 0 et 1, notamment dans le cadre des *distributions a priori* en inférence bayésienne. La fonction densité de probabilité de la distribution Beta, qui dépend de deux paramètres de forme, α et β , est formulée comme suit :

$$f(x; \alpha, \beta) = \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha, \beta)} \quad \text{pour } x \in [0, 1].$$

$B(\alpha, \beta)$ est la fonction Beta de normalisation. Cette distribution est remarqua-

blement flexible, permettant de moduler la forme de la distribution en fonction des valeurs des paramètres.

Une autre loi de probabilité notable, la distribution Kumaraswamy, a été introduite par le statisticien indien Kumaraswamy en 1980. Cette distribution offre une alternative à la loi Beta avec une forme de densité plus simple à manipuler, tout en conservant une flexibilité similaire. Sa densité est exprimée par :

$$f(x; a, b) = abx^{a-1}(1 - x^a)^{b-1} \quad \text{pour } x \in [0, 1]$$

Les paramètres a et b contrôlent la forme de la distribution. La distribution Kumaraswamy est souvent utilisée dans des situations où la simplicité numérique est un avantage, en particulier dans les simulations.

D'autres distributions ont également été développées pour mieux répondre à des besoins spécifiques. La distribution triangulaire, dont la densité est souvent utilisée dans des contextes de modélisation avec une probabilité maximale au centre de l'intervalle $[0, 1]$, est définie par une fonction en forme de triangle, et est couramment utilisée dans la simulation de systèmes incertains.

Une autre famille importante de distributions unitaires est la loi logistique unitaire, qui est souvent employée dans des modèles *sigmoïdes*, comme ceux utilisés en régression logistique et en apprentissage automatique. Bien que la loi logistique soit traditionnellement définie sur \mathbb{R} , sa version unitaire restreint les valeurs à l'intervalle $[0, 1]$, ce qui la rend très utile pour les modèles dont les résultats doivent être contraints entre ces deux bornes.

La loi de Pareto unitaire et la distribution de Rayleigh unitaire sont également des variantes intéressantes des distributions classiques, qui ont été adaptées pour avoir un support sur $[0, 1]$. Ces lois sont souvent utilisées dans les domaines de la finance et des télécommunications pour modéliser des phénomènes où les valeurs extrêmes sont rares mais significatives.

De plus, des distributions plus récentes ont vu le jour pour répondre à des besoins spécifiques de modélisation. Parmi celles-ci, on peut citer la loi uniforme généralisée, qui permet une plus grande flexibilité que la distribution uniforme classique, tout en conservant un support sur $[0, 1]$. Une autre distribution généralisée est la loi logistique unitaire. Enfin, la loi Beta modifiée permet d'adapter

la forme de la distribution Beta tout en conservant un support entre 0 et 1.

Et plus récemment, des lois spécifiques à support $[0, 1]$ ont été introduites pour mieux capturer certaines dynamiques dans les données. Ces distributions incluent la distribution de Weibull unitaire (Mazucheli et al., 2018a), la distribution de Birnbaum-Saunders unitaire (Mazucheli et al., 2018c), la distribution de Gompertz unitaire (Mazucheli et al., 2019), et la distribution de Weibull inverse généralisée unitaire (Aidi et al., 2021). Ces modèles ont été développés pour mieux représenter des processus de défaillance ou d'événements extrêmes avec des comportements plus complexes que ceux capturés par les distributions classiques.

Ainsi, les distributions à support sur $[0, 1]$ permettent d'adapter les modèles aux caractéristiques spécifiques des données, que ce soit dans un contexte bayésien, une simulation, ou un modèle d'apprentissage automatique. Elles enrichissent les outils disponibles pour les chercheurs et les utilisateurs dans de multiples disciplines.

En se basant sur la distribution exponentielle généralisée (GE), nous introduisons un nouveau modèle à deux paramètres sur $[0, 1]$ appelé la distribution exponentielle généralisée unitaire (UGE). Avec une fonction de densité de probabilité unimodale, décroissante et asymétrique à gauche, cette distribution peut être considérée comme une bonne alternative aux modèles cités ci-dessus.

2.1 Distribution Exponentielle Généralisée unitaire (UGE)

La distribution exponentielle généralisée unitaire (UGE) est une extension de la distribution exponentielle généralisée (GE) permettant de modéliser des variables aléatoires bornées dans l'intervalle $(0, 1)$. Grâce à sa flexibilité, elle s'adapte à une large variété de formes de densité, allant d'une décroissance stricte à une distribution unimodale ou présentant une queue lourde à gauche. Cette diversité lui confère un grand potentiel d'application dans divers domaines scientifiques et industriels.

En fiabilité et en analyse des durées de vie, elle permet d'expliquer des phénomènes où le risque évolue avec le temps, notamment les systèmes soumis à une

usure progressive ou les processus de défaillance avec une phase de mortalité infantile suivie d'une stabilisation. En sciences économiques et financières, elle est utile pour modéliser des rendements, des pertes ou des indices normalisés dont la distribution présente des asymétries marquées. Dans le domaine médical et biologique, elle peut être appliquée à l'étude des concentrations de certaines substances ou des probabilités de survie. Elle trouve également un intérêt en hydrologie et en sciences de l'environnement, notamment pour l'analyse des proportions de précipitations ou des indices de sécheresse normalisés.

Un de ses principaux atouts réside dans la diversité des formes que peut prendre son taux de hasard. Selon les paramètres choisis, celui-ci peut être croissant, traduisant un risque qui augmente avec le temps, ou en forme de baignoire, caractérisant des phénomènes où le risque est élevé au départ, diminue, puis réaugmente. Cette capacité d'adaptation fait de la distribution UGE un outil puissant pour l'analyse statistique de données bornées, offrant une alternative aux modèles classiques souvent trop rigides pour certaines applications complexes.

La distribution Exponentielle Généralisée (GE)

La distribution exponentielle généralisée introduite par Gupta et Kundu (1999) est définie par :

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

Les paramètres α et λ représentent les paramètres de forme et d'échelle.

En utilisant la transformation :

$$t = \frac{x}{1-x}, \quad \text{avec } x \in (0, 1),$$

on obtient une densité en termes d'une variable bornée sur $(0, 1)$:

$$f(x) = \alpha\lambda(1-x)^{-2} \exp\left(-\lambda\frac{x}{1-x}\right) \left[1 - \exp\left(-\lambda\frac{x}{1-x}\right)\right]^{\alpha-1}.$$

Cette distribution est appelée distribution exponentielle généralisée unitaire et notée *UGE*. Sa fonction de répartition (CDF) est alors obtenue comme suit :

$$F(x) = \left[1 - \exp \left(-\lambda \frac{x}{1-x} \right) \right]^\alpha, \quad \alpha > 0, \lambda > 0, x \in (0, 1).$$

Cette fonction de répartition décrit la probabilité accumulée jusqu'à une valeur donnée x dans le support $(0, 1)$.

Cette transformation permet d'adapter la distribution exponentielle généralisée à un support borné sur $(0, 1)$, ce qui est utile pour modéliser des phénomènes avec des variables naturelles comprises dans cet intervalle. Alors le taux de hasard de cette distribution est défini comme suit :

$$h(x) = \frac{\alpha \lambda (1-x)^{-2} \exp \left(-\lambda \frac{x}{1-x} \right) \left[1 - \exp \left(-\lambda \frac{x}{1-x} \right) \right]^{\alpha-1}}{1 - \left[1 - \exp \left(-\lambda \frac{x}{1-x} \right) \right]^\alpha}.$$

Pour générer des données à partir de cette distribution, nous utilisons la fonction quantile suivante :

$$Q(u) = \frac{-\log(1 - u^{1/\alpha})}{\lambda - \log(1 - u^{1/\alpha})},$$

où u est uniformément distribuée sur l'intervalle $[0, 1]$.

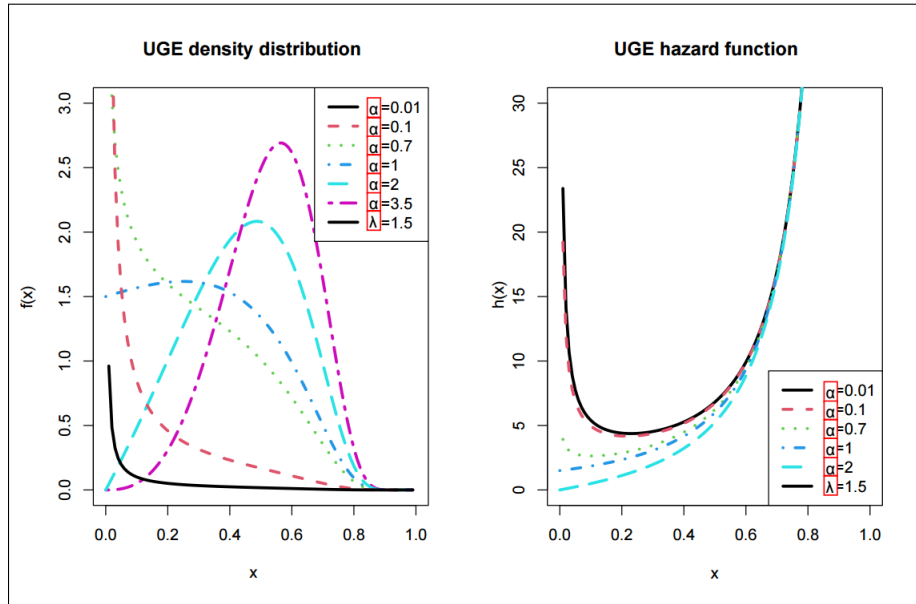


FIGURE 1 – Représentations graphiques de la pdf et du taux de hasard de la distribution UGE.

Selon les valeurs des paramètres, la fonction de densité de probabilité (pdf) du modèle *UGE* et la fonction de taux de défaillance affichées dans la Figure 1

montrent la flexibilité de ce nouveau modèle.

Pour de petites valeurs de α , la densité et la forme de la fonction de taux de défaillance sont décroissantes, comme dans le cas de la loi exponentielle, ce qui décrit les défaillances précoces. Il est à noter que pour des valeurs plus élevées de α , la densité devient progressivement unimodale avec une fonction de risque en forme de baignoire ou croissante, ce qui caractérise de nombreux systèmes réels : les défaillances se produisent fréquemment en début de vie (en raison de défauts), se stabilisent ensuite, puis augmentent de nouveau à cause du vieillissement ou de l'usure.

2.2 Estimation

En plus de certaines propriétés statistiques, nous proposons différentes méthodes d'estimation pour calculer les estimateurs des paramètres inconnus.

2.2.1 Moments

Les coefficients d'asymétrie et d'aplatissement de toute distribution de probabilité donnent une idée sur le comportement du phénomène, par exemple sur la présence d'outliers ou sur l'asymétrie de la distribution. Nous fournissons dans cette section les moments de la distribution *UGE* nécessaires pour le calcul de ces coefficients.

Le moment général d'ordre s est donné par :

$$E[X^s] = \int_0^1 x^s \alpha \lambda (1-x)^{-2} \exp\left(-\lambda \frac{x}{1-x}\right) \left[1 - \exp\left(-\lambda \frac{x}{1-x}\right)\right]^{\alpha-1} dx.$$

Après quelques simplifications, nous obtenons :

$$E[X^s] = 1 - \frac{s(-1)^\alpha}{\alpha^{\alpha+1} \lambda k^\alpha} [(s+1)\Gamma(\alpha+1)\beta(2, s) - (s-1)\Gamma(\alpha+1)\beta(2, s-1)].$$

Ensuite, nous déduisons les coefficients d'asymétrie sk et d'aplatissement kur :

$$sk = \frac{E[X^3]}{Var[X]^{3/2}} = \frac{1 - \frac{3(-1)^\alpha}{\alpha^{\alpha+1}\lambda k^\alpha} [4\Gamma(\alpha + 1)\beta(2, 3) - 2\Gamma(\alpha + 1)\beta(2, 2)]}{Var[X]^{3/2}},$$

$$kur = \frac{E[X^4]}{(E[X^2] - E[X]^2)^2} = \frac{1 - \frac{4(-1)^\alpha}{\alpha^{\alpha+1}\lambda k^\alpha} [5\Gamma(\alpha + 1)\beta(2, 4) - 3\Gamma(\alpha + 1)\beta(2, 3)]}{(E[X^2] - E[X]^2)^2}.$$

où Γ représente la fonction Gamma.

2.2.2 Méthode du produit maximum des espacements (MPS)

Comme indiqué dans le premier chapitre, les estimateurs par cette méthode, notés $\hat{\alpha}_{MPS}$ et $\hat{\lambda}_{MPS}$, sont les solutions des équations suivantes :

$$\frac{\partial M(\theta)}{\partial \alpha} = \frac{1}{n+1} \sum_{i=1}^{n+1} \frac{f_\alpha'(x_i, \theta) - f_\alpha'(x_{i-1}, \theta)}{P_i} = 0,$$

$$\frac{\partial M(\theta)}{\partial \lambda} = \frac{1}{n+1} \sum_{i=1}^{n+1} \frac{f_\lambda'(x_i, \theta) - f_\lambda'(x_{i-1}, \theta)}{P_i} = 0,$$

où f_α' et f_λ' représentent respectivement :

$$f_\alpha' = \frac{\partial F}{\partial \alpha} = \left[1 - \exp\left(-\lambda \frac{x}{1-x}\right) \right]^\alpha \cdot \log\left(1 - \exp\left(-\lambda \frac{x}{1-x}\right) \right), \quad (2.1)$$

$$f_\lambda' = \frac{\partial F}{\partial \lambda} = \frac{\alpha x}{1-x} \exp\left(-\lambda \frac{x}{1-x}\right) \left[1 - \exp\left(-\lambda \frac{x}{1-x}\right) \right]^{\alpha-1}. \quad (2.2)$$

2.2.3 Méthodes d'estimation du type minimum

Pour le modèle proposé, les estimateurs $(\hat{\alpha}_{CVM}, \hat{\lambda}_{CVM})$ et $(\hat{\alpha}_{AD}, \hat{\lambda}_{AD})$ sont obtenus en résolvant les systèmes d'équations de Cramér-von Mises (CVM) et Anderson-Darling (AD) suivants :

Méthode de Cramér-von Mises

Pour les estimateurs $(\widehat{\alpha}_{CVM}, \widehat{\lambda}_{CVM})$, on résout le système d'équations suivant :

$$\sum_{i=1}^n \left(\left[1 - \exp \left(-\lambda \frac{x_i}{1-x_i} \right) \right]^\alpha - \frac{2i-1}{2n} \right) f_\alpha' = 0$$

$$\sum_{i=1}^n \left(\left[1 - \exp \left(-\lambda \frac{x_i}{1-x_i} \right) \right]^\alpha - \frac{2i-1}{2n} \right) f_\lambda' = 0.$$

Méthode d'Anderson-Darling

Et pour $(\widehat{\alpha}_{AD}, \widehat{\lambda}_{AD})$, on va résoudre le système suivant :

$$\sum_{i=1}^n (2i-1) \left[\frac{f_\alpha'(x_i)}{\left[1 - \exp \left(-\lambda \frac{x_i}{1-x_i} \right) \right]^\alpha} - \frac{f_\alpha'(x_{i^*})}{1 - \left[1 - \exp \left(-\lambda \frac{x_{i^*}}{1-x_{i^*}} \right) \right]^\alpha} \right] = 0$$

$$\sum_{i=1}^n (2i-1) \left[\frac{f_\lambda'(x_i)}{\left[1 - \exp \left(-\lambda \frac{x_i}{1-x_i} \right) \right]^\alpha} - \frac{f_\lambda'(x_{i^*})}{1 - \left[1 - \exp \left(-\lambda \frac{x_{i^*}}{1-x_{i^*}} \right) \right]^\alpha} \right] = 0$$

Méthode du percentile

Pour ce modèle, nous avons :

$$U(\theta) = \sum_{i=1}^n \left\{ x_i - \frac{-\log(1-u^{1/\alpha})}{\lambda - \log(1-u^{1/\alpha})} \right\}^2$$

Les estimateurs des percentiles $(\widehat{\alpha}_{PE}, \widehat{\lambda}_{PE})$ sont alors obtenus en résolvant les équations suivantes :

$$2 \sum_{i=1}^n \left\{ x_i - \frac{\log(1-u_i^{1/\alpha})}{\lambda - \log(1-u_i^{1/\alpha})} \right\} \left[\frac{-\lambda u_i^{1/\alpha} \log(u_i)}{(1-u_i^{1/\alpha})(\lambda - \log(1-u_i^{1/\alpha}))^2} \right] = 0$$

$$2 \sum_{i=1}^n \left\{ x_i - \frac{\log(1-u_i^{1/\alpha})}{\lambda - \log(1-u_i^{1/\alpha})} \right\} \left[\frac{-\log(1-u_i)}{(\lambda - \log(1-u_i^{1/\alpha}))^2} \right] = 0$$

2.3 Validation du modèle UGE

En se basant sur les différentes approches décrites dans le premier chapitre, on va construire des statistiques de test du type du khi-deux modifié pour tester l'hypothèse nulle H_0 qu'un échantillon (X_1, X_2, \dots, X_n) provienne de la distribution UGE avec le vecteur de paramètres inconnus $\theta = (\alpha, \lambda)^T$.

Critères Dzhaparidze-Nikulin

Pour construire la statistique Dzhaparidze-Nikulin (DN), nous procédons comme suit. Tout d'abord, nous calculons les estimateurs \sqrt{n} -convergeants $\tilde{\theta}$ des paramètres inconnus. Ensuite, nous regroupons les observations (X_1, X_2, \dots, X_n) en r classes disjointes $\Delta_j = [a_{j-1}; a_j[$, où

$$p_j(\theta) = \int_{\Delta_j} dF(x; \theta), \quad j = 1, \dots, r.$$

Si $v = (v_1, v_2, \dots, v_r)^T$ représente le vecteur des fréquences observées dans ces classes Δ_j , alors

$$\begin{aligned} p_j(\theta) &= F(a_j) - F(a_{j-1}) \\ &= \left[1 - \exp\left(-\lambda \frac{a_j}{1 - a_j}\right) \right]^\alpha - \left[1 - \exp\left(-\lambda \frac{a_{j-1}}{1 - a_{j-1}}\right) \right]^\alpha \end{aligned}$$

et

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

La matrice $B(\theta) = (b_{jk})_{r \times 2}$ est obtenue comme suit :

$$b_{j1} = \frac{1}{\sqrt{p_j(\theta)}} \int_{\Delta_j} \frac{\partial f(x, \theta)}{\partial \alpha} dx, \quad j = 1, \dots, r,$$

$$b_{j1} = \frac{1}{\sqrt{p_j(\theta)}} \left[\left(1 - \exp \left(-\lambda \frac{a_j}{1 - a_j} \right) \right)^\alpha \log \left(1 - \exp \left(-\lambda \frac{a_j}{1 - a_j} \right) \right) \right. \\ \left. - \left(1 - \exp \left(-\lambda \frac{a_{j-1}}{1 - a_{j-1}} \right) \right)^\alpha \log \left(1 - \exp \left(-\lambda \frac{a_{j-1}}{1 - a_{j-1}} \right) \right) \right]$$

$$b_{j2} = \frac{1}{\sqrt{p_j(\theta)}} \int_{\Delta_j} \frac{\partial f(x, \theta)}{\partial \lambda} dx, \quad j = 1, \dots, r,$$

$$b_{j2} = \frac{1}{\sqrt{p_j(\theta)}} \left[\frac{\alpha a_j}{1 - a_j} \exp \left(-\lambda \frac{a_j}{1 - a_j} \right) \left(1 - \exp \left(-\lambda \frac{a_j}{1 - a_j} \right) \right)^\alpha \right. \\ \left. - \frac{\alpha a_{j-1}}{1 - a_{j-1}} \exp \left(-\lambda \frac{a_{j-1}}{1 - a_{j-1}} \right) \left(1 - \exp \left(-\lambda \frac{a_{j-1}}{1 - a_{j-1}} \right) \right)^\alpha \right]$$

La matrice d'information de Fisher I_n a été calculée plus haut. Ainsi, tous les composants de $U^2(\tilde{\theta})$ sont obtenus, on en déduit alors la statistique :

$$U^2(\theta) = Z_n^T(\theta) \left[I - B(\theta) (B^T(\theta)B(\theta))^{-1} B^T(\theta) \right] Z_n(\theta).$$

L'hypothèse nulle H_0 est rejetée pour

$$U^2 > \chi_{r-3}^2$$

avec $(r - 2 - 1) = (r - 3)$ degrés de liberté.

Critères Nikulin-Rao-Robson

Pour valider l'hypothèse nulle H_0 selon laquelle un échantillon appartient à la distribution UGE , on adapte la statistique Y^2 à notre modèle. On utilise la forme de calcul suivante :

$$Y^2(\hat{\theta}) = X^2(\hat{\theta}) + \frac{1}{n} L^T(\hat{\theta}_n) \left(I_n(\hat{\theta}_n) - J(\hat{\theta}_n) \right)^{-1} L(\hat{\theta}_n)$$

On calcule d'abord l'estimateur du maximum de vraisemblance (EMV) du paramètre θ . Ensuite, les données doivent être regroupées en r classes équiprobables $\Delta'_j = [a'_{j-1}, a'_j]$, ce qui équivaut à :

$$p_j(\theta) = \int_{\Delta'_j} dF(x; \theta) = \frac{1}{r}, \quad \text{pour tout } j = 1, \dots, r.$$

Les bornes des intervalles de ces classes peuvent être obtenues par :

$$a'_j = F^{-1}(j/r)$$

Ce qui donne :

$$a'_j = \frac{-\log(1 - u^{1/\alpha})}{\lambda - \log(1 - u^{1/\alpha})}$$

où u est une variable uniformément distribuée sur $(0, 1)$. On a alors :

$$p_j(\theta) = F(a'_j) - F(a'_{j-1})$$

L'information de Fisher sur les données groupées est donnée par $J(\hat{\theta}) = B^T(\hat{\theta})B(\hat{\theta})$, calculée sur les classes équiprobables Δ'_j :

$$b_{j1}(\theta) = \frac{1}{\sqrt{p_j}} \cdot \frac{\partial p_j(\theta)}{\partial \alpha}, \quad b_{j2}(\theta) = \frac{1}{\sqrt{p_j}} \cdot \frac{\partial p_j(\theta)}{\partial \lambda}, \quad j = 1, \dots, r$$

$$J_{11} = \sum_j \frac{1}{p_j} \left(\frac{\partial p_j(\theta)}{\partial \alpha} \right)^2, \quad J_{22} = \sum_j \frac{1}{p_j} \left(\frac{\partial p_j(\theta)}{\partial \lambda} \right)^2,$$

$$J_{12} = J_{21} = \sum_j \frac{1}{p_j} \left(\frac{\partial p_j(\theta)}{\partial \alpha} \cdot \frac{\partial p_j(\theta)}{\partial \lambda} \right)$$

$$L = (L_1, L_2)^T, \quad L_1 = \sum_{j=1}^r \frac{v_j}{p_j(\theta)} \cdot \frac{dp_j(\theta)}{d\alpha}, \quad L_2 = \sum_{j=1}^r \frac{v_j}{p_j(\theta)} \cdot \frac{dp_j(\theta)}{d\lambda}$$

Les dérivées partielles de $p_j(\theta)$ sont :

$$f_{j1} = \frac{\partial p_j}{\partial \alpha} = \left[1 - \exp \left(-\lambda \cdot \frac{a_j}{1 - a_j} \right) \right]^\alpha \log \left[1 - \exp \left(-\lambda \cdot \frac{a_j}{1 - a_j} \right) \right] \\ - \left[1 - \exp \left(-\lambda \cdot \frac{a_{j-1}}{1 - a_{j-1}} \right) \right]^\alpha \log \left[1 - \exp \left(-\lambda \cdot \frac{a_{j-1}}{1 - a_{j-1}} \right) \right]$$

$$f_{j2} = \frac{\partial p_j}{\partial \lambda} = \alpha \cdot \frac{a_j}{1 - a_j} \cdot \exp \left(-\lambda \cdot \frac{a_j}{1 - a_j} \right) \cdot \left[1 - \exp \left(-\lambda \cdot \frac{a_j}{1 - a_j} \right) \right]^{\alpha-1} \\ - \alpha \cdot \frac{a_{j-1}}{1 - a_{j-1}} \cdot \exp \left(-\lambda \cdot \frac{a_{j-1}}{1 - a_{j-1}} \right) \cdot \left[1 - \exp \left(-\lambda \cdot \frac{a_{j-1}}{1 - a_{j-1}} \right) \right]^{\alpha-1}$$

Les formules générales de $\frac{\partial p_j(\theta)}{d\alpha} = f_{j1}$ et $\frac{\partial p_j(\theta)}{d\lambda} = f_{j2}$ sont adaptées aux classes Δ'_j . Tous les composants de la statistique sont donc calculés, par conséquent, Y^2 peut être déduite.

Critère de McCulloch

Étant donné que la statistique de McCulloch S^2 est basée sur les estimateurs du maximum de vraisemblance $\hat{\theta}$, la matrice d'information de Fisher sur les données initiales $I_n(\hat{\theta})$ est obtenue de la même manière que pour la statistique Y^2 . Ensuite, on calcule $J_n(\hat{\theta})$ et la matrice correspondante B pour les données regroupées dans les classes choisies Δ_j . Par conséquent, on déduit le critère *McCu*-test :

$$S_n^2 = Z_n^T(\hat{\theta}) \cdot B \left[(I_n(\hat{\theta}) - J_n(\hat{\theta}))^{-1} + (B^T B)^{-1} \right] B^T Z_n(\hat{\theta})$$

qui suit une loi du khi-deux avec 2 degrés de liberté.

Statistique de Singh

La statistique de Singh Q^2 peut être adaptée pour valider la distribution *UGE* dans des conditions plus générales que la statistique Y^2 . Dans ce cas, on peut choisir la méthode d'estimation pour calculer les paramètres inconnus et les classes de regroupement. Si $\tilde{\theta}$ est un estimateur \sqrt{n} -convergent, on détermine le vecteur $Z_*(\tilde{\theta})$:

$$Z_*(\tilde{\theta}) = Z(\tilde{\theta}) - BI_n^{-1}W(\tilde{\theta})$$

où B est défini comme précédemment. Les éléments de $W(\tilde{\theta})$ sont :

$$W_1(\tilde{\theta}) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\partial \ln f(x_i, \theta)}{\partial \alpha} \Big|_{\alpha=\tilde{\alpha}} = \frac{1}{\sqrt{n}} \sum_{i=1}^n \left[\frac{1}{\tilde{\alpha}} + \ln \left(1 - \exp \left(-\tilde{\lambda} \cdot \frac{x_i}{1-x_i} \right) \right) \right]$$

$$W_2(\tilde{\theta}) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{\partial \ln f(x_i, \theta)}{\partial \lambda} \Big|_{\lambda=\tilde{\lambda}} = \frac{1}{\sqrt{n}} \sum_{i=1}^n \left[\frac{1}{\tilde{\lambda}} + \frac{x_i}{1-x_i} \left(\frac{(\tilde{\alpha}-1) \exp \left(-\tilde{\lambda} \cdot \frac{x_i}{1-x_i} \right)}{1 - \exp \left(-\tilde{\lambda} \cdot \frac{x_i}{1-x_i} \right)} - 1 \right) \right]$$

Ainsi, on obtient la valeur de $Q_n^2(\tilde{\theta})$ qui suit une loi du khi-deux avec $(r-1)$ degrés de liberté. L'hypothèse H_0 est rejetée si :

$$Q_n^2 > \chi_{(r-1, \alpha)}^2$$

2.4 Simulations et Applications

Dans ce paragraphe, on présente une étude par simulation numérique pour confirmer les résultats théoriques et montrer l'utilité du nouveau modèle.

2.4.1 Simulations

Estimation

Une importante étude par simulations est réalisée pour choisir des estimateurs \sqrt{n} -convergeants afin de calculer les statistiques de test proposées dans ce travail pour valider le modèle *UGE*.

À cet effet, nous avons généré $N = 10\,000$ échantillons de tailles différentes ($n = 20, 30, 50, 100$ et $n = 150$), avec des valeurs de paramètres différentes :

$$(\alpha = 0,3; \lambda = 0,7), \quad (\alpha = 0,5; \lambda = 0,3), \quad \text{et} \quad (\alpha = 1,2; \lambda = 0,7)$$

de la distribution *UGE*.

Nous avons calculé les estimateurs *MLE*, *MPS*, *CM*, *AD*, *RAD*, *LAD*, *KS* et les estimateurs de percentile des paramètres, ainsi que leurs biais et leurs erreurs quadratiques moyennes (voir les Tables 1, 2 et 3).

n	(0.5, 0.3)	MLE	MPS	CM	AD	RAD	LAD	KS	PE
20	$\hat{\alpha}$	0.5449	0.5564	0.5656	0.5336	0.5528	0.5540	0.5733	0.5961
	<i>MSE</i>	0.0166	0.0351	0.0445	0.0300	0.0444	0.0357	0.0487	0.0464
	<i>bias</i>	0.0449	0.0564	0.0656	0.0336	0.0528	0.0540	0.0733	0.0961
	$\hat{\lambda}$	0.3462	0.3449	0.3389	0.3151	0.3179	0.3558	0.3479	0.3999
	<i>MSE</i>	0.0142	0.0181	0.0235	0.0110	0.0182	0.0486	0.0246	0.0425
	<i>bias</i>	0.0462	0.0449	0.0389	0.0151	0.0179	0.0558	0.0479	0.0999
30	$\hat{\alpha}$	0.5192	0.5250	0.5299	0.5116	0.5242	0.5136	0.5326	0.5860
	<i>MSE</i>	0.0077	0.0203	0.0236	0.0151	0.0187	0.0187	0.0245	0.0358
	<i>bias</i>	0.0192	0.0250	0.0299	0.0116	0.0112	0.0136	0.0326	0.0860
	$\hat{\lambda}$	0.3230	0.3257	0.3460	0.3114	0.3178	0.3397	0.3444	0.3757
	<i>MSE</i>	0.0059	0.0084	0.0214	0.0160	0.0103	0.0103	0.0220	0.0332
	<i>bias</i>	0.0230	0.0257	0.0460	0.0114	0.0178	0.0397	0.0444	0.0757
50	$\hat{\alpha}$	0.5085	0.5236	0.5278	0.5081	0.5241	0.5115	0.5268	0.5542
	<i>MSE</i>	0.0033	0.0057	0.0150	0.0094	0.0138	0.0108	0.0141	0.0229
	<i>bias</i>	0.0085	0.0236	0.0278	0.0081	0.0241	0.0115	0.0268	0.0542
	$\hat{\lambda}$	0.3197	0.3240	0.3243	0.3097	0.3167	0.3138	0.3210	0.3353
	<i>MSE</i>	0.0027	0.0062	0.0092	0.0067	0.0076	0.0100	0.0083	0.0120
	<i>bias</i>	0.0197	0.0240	0.0243	0.0097	0.0167	0.0138	0.0210	0.0353
100	$\hat{\alpha}$	0.5019	0.5138	0.5113	0.5044	0.5103	0.5074	0.5076	0.5203
	<i>MSE</i>	0.0023	0.0029	0.0048	0.0048	0.0060	0.0042	0.0048	0.0088
	<i>bias</i>	0.0019	0.0138	0.0113	0.0044	0.0103	0.0074	0.0076	0.0203
	$\hat{\lambda}$	0.3142	0.3101	0.3125	0.3051	0.3069	0.3123	0.3126	0.3160
	<i>MSE</i>	0.0012	0.0028	0.0039	0.0039	0.0025	0.0051	0.0040	0.0074
	<i>bias</i>	0.0142	0.0101	0.0125	0.0051	0.0069	0.0123	0.0126	0.0160
150	$\hat{\alpha}$	0.5009	0.5074	0.5083	0.5002	0.5044	0.5044	0.5038	0.5079
	<i>MSE</i>	0.0016	0.0025	0.0029	0.0023	0.0029	0.0029	0.0029	0.0061
	<i>bias</i>	0.0009	0.0074	0.0083	0.0002	0.0044	0.0044	0.0038	0.0079
	$\hat{\lambda}$	0.2984	0.3047	0.3091	0.3043	0.3040	0.3094	0.3090	0.3044
	<i>MSE</i>	0.0010	0.0014	0.0021	0.0016	0.0016	0.0029	0.0029	0.0040
	<i>bias</i>	0.0016	0.0047	0.0091	0.0043	0.0040	0.0094	0.0090	0.0044

TABLE 1 – Valeurs des estimateurs, leur biais et MSE.

n	(0.3, 0.7)	MLE	MPS	CM	AD	RAD	LAD	KS	PE
20	$\hat{\alpha}$	0.3282	0.2759	0.3438	0.3134	0.3345	0.3193	0.3520	0.3022
	<i>MSE</i>	0.0070	0.0048	0.0103	0.0066	0.0129	0.0065	0.0117	0.00119
	<i>biais</i>	0.0282	0.0241	0.0438	0.0134	0.0345	0.0193	0.0520	0.0022
	$\hat{\lambda}$	0.9673	0.6044	0.9942	0.7812	0.8133	0.9802	0.9943	0.6977
	<i>MSE</i>	0.3128	0.0819	0.6686	0.1760	0.1639	0.9537	0.4459	0.00024
	<i>biais</i>	0.2673	0.0956	0.2942	0.0812	0.1133	0.2802	0.2943	0.0023
30	$\hat{\alpha}$	0.3148	0.2849	0.3185	0.3076	0.3193	0.3145	0.3207	0.3020
	<i>MSE</i>	0.0044	0.0041	0.0096	0.0059	0.0096	0.0085	0.0092	0.00118
	<i>biais</i>	0.0148	0.0151	0.0185	0.0076	0.0193	0.0145	0.0207	0.0020
	$\hat{\lambda}$	0.8285	0.6232	0.7935	0.7435	0.7581	0.8564	0.8144	0.6979
	<i>MSE</i>	0.1136	0.0590	0.1748	0.1082	0.1097	0.5268	0.1968	0.00024
	<i>biais</i>	0.1285	0.0768	0.0935	0.0435	0.0581	0.1564	0.1144	0.0021
50	$\hat{\alpha}$	0.3136	0.2850	0.3139	0.3053	0.3110	0.3071	0.3167	0.3019
	<i>MSE</i>	0.0032	0.0018	0.0037	0.0026	0.0038	0.0027	0.0039	0.00111
	<i>biais</i>	0.0136	0.0150	0.0139	0.0053	0.0110	0.0071	0.0167	0.0019
	$\hat{\lambda}$	0.7621	0.6497	0.7915	0.7348	0.7553	0.7813	0.8065	0.6980
	<i>MSE</i>	0.0501	0.0343	0.0934	0.0576	0.0594	0.1132	0.0931	0.00021
	<i>biais</i>	0.0621	0.0541	0.0915	0.0348	0.0553	0.0813	0.1065	0.0020
100	$\hat{\alpha}$	0.3046	0.2886	0.3094	0.3036	0.3066	0.3052	0.3107	0.3018
	<i>MSE</i>	0.0012	0.0014	0.0027	0.0019	0.0026	0.0022	0.0026	0.00108
	<i>biais</i>	0.0046	0.0114	0.0094	0.0036	0.0066	0.0052	0.0107	0.0018
	$\hat{\lambda}$	0.7343	0.6545	0.7573	0.7269	0.7331	0.7505	0.7652	0.6981
	<i>MSE</i>	0.0215	0.0263	0.0595	0.0371	0.0405	0.0667	0.0565	0.00021
	<i>biais</i>	0.0343	0.0455	0.0573	0.0269	0.0331	0.0505	0.0652	0.0019
150	$\hat{\alpha}$	0.3026	0.2898	0.3088	0.3032	0.3063	0.3048	0.3103	0.3016
	<i>MSE</i>	0.0007	0.0013	0.0026	0.0019	0.0025	0.0021	0.0026	0.00107
	<i>biais</i>	0.0026	0.0102	0.0088	0.0032	0.0063	0.0048	0.0103	0.0016
	$\hat{\lambda}$	0.7214	0.6568	0.7518	0.7234	0.7295	0.7466	0.7625	0.6982
	<i>MSE</i>	0.0146	0.0233	0.0534	0.0353	0.0370	0.0579	0.0515	0.00021
	<i>biais</i>	0.0214	0.0432	0.0518	0.0234	0.0295	0.0466	0.0625	0.0018

TABLE 2 – Valeurs des estimateurs, leur biais et MSE..

n	(1.2, 0.7)	MLE	MPS	CM	AD	RAD	LAD	KS	PE
20	$\hat{\alpha}$	1.2944	1.1328	1.2980	1.2439	1.2817	1.2600	1.3040	1.2320
	<i>MSE</i>	0.0824	0.0530	0.1252	0.0769	0.1219	0.0895	0.1153	0.0285
	<i>biais</i>	0.0944	0.0672	0.0980	0.0439	0.0817	0.0600	0.1040	0.0320
	$\hat{\lambda}$	0.7364	0.7413	0.7435	0.7199	0.7283	0.7326	0.7473	0.6836
	<i>MSE</i>	0.0272	0.0149	0.0280	0.0191	0.0219	0.0281	0.0262	0.00197
	<i>biais</i>	0.0364	0.0413	0.0435	0.0199	0.0283	0.0326	0.0473	0.0164
30	$\hat{\alpha}$	1.2831	1.1352	1.2941	1.2404	1.2800	1.2589	1.3025	1.2260
	<i>MSE</i>	0.0736	0.0508	0.1248	0.0743	0.1214	0.0888	0.1083	0.0142
	<i>biais</i>	0.0831	0.0648	0.0941	0.0404	0.0800	0.0589	0.1025	0.0260
	$\hat{\lambda}$	0.7373	0.6610	0.7394	0.7162	0.7250	0.7297	0.7447	0.6870
	<i>MSE</i>	0.0190	0.0145	0.0275	0.0185	0.0216	0.0274	0.0257	0.00152
	<i>biais</i>	0.0373	0.0390	0.0394	0.0162	0.0250	0.0297	0.0447	0.0130
50	$\hat{\alpha}$	1.2763	1.1375	1.2926	1.2383	1.2776	1.2559	1.2999	1.2236
	<i>MSE</i>	0.0685	0.0499	0.1208	0.0743	0.1209	0.0724	0.1049	0.0099
	<i>biais</i>	0.0736	0.0625	0.0926	0.0383	0.0776	0.0559	0.0999	0.0236
	$\hat{\lambda}$	0.7354	0.6630	0.7313	0.7159	0.7265	0.7279	0.7429	0.6959
	<i>MSE</i>	0.0180	0.0141	0.0248	0.0194	0.0211	0.0265	0.0247	0.00081
	<i>biais</i>	0.0354	0.0370	0.0313	0.0159	0.0265	0.0279	0.0429	0.0041
100	$\hat{\alpha}$	1.2410	1.1531	1.2479	1.2241	1.2356	1.2351	1.2577	1.2223
	<i>MSE</i>	0.0307	0.0257	0.0485	0.0349	0.0487	0.0400	0.0499	0.0060
	<i>biais</i>	0.0410	0.0469	0.0479	0.0241	0.0356	0.0351	0.0577	0.0223
	$\hat{\lambda}$	0.7140	0.6703	0.7162	0.7051	0.7074	0.7131	0.7207	0.6967
	<i>MSE</i>	0.0077	0.0069	0.0120	0.0088	0.0099	0.0123	0.0122	0.00033
	<i>biais</i>	0.0140	0.0297	0.0162	0.0051	0.0074	0.0131	0.0207	0.0033
150	$\hat{\alpha}$	1.2118	1.1575	1.2041	1.1963	1.2147	1.1976	1.2100	1.2128
	<i>MSE</i>	0.0231	0.0060	0.0324	0.0263	0.0377	0.0262	0.0344	0.0037
	<i>biais</i>	0.0118	0.0425	0.0041	0.0037	0.0147	0.0024	0.0100	0.0128
	$\hat{\lambda}$	0.7055	0.6743	0.6979	0.6954	0.7006	0.6982	0.6997	0.6969
	<i>MSE</i>	0.0067	0.0062	0.0084	0.0073	0.0082	0.0095	0.0091	0.00015
	<i>biais</i>	0.0055	0.0257	0.0021	0.0046	0.0006	0.0018	0.0003	0.0031

TABLE 3 – Valeurs des estimateurs, leur biais et MSE..

À partir de ces tableaux, nous voyons que les estimateurs obtenus par les méthodes du maximum de vraisemblance, d'Anderson-Darling, *MPS* et de Cramér-von Mises sont les plus efficaces. Nous allons donc les utiliser pour calculer les statistiques U^2 , Y^2 , S^2 et Q^2 .

Critères de test

Pour tous les échantillons simulés et pour différents niveaux de signification ($\epsilon = 0,025, 0,05$ et $0,1$), nous calculons le nombre de rejets de l'hypothèse nulle H_0 et les comparons avec leurs valeurs théoriques dans les classes de données équiprobables ou non équiprobables, selon la statistique choisie.

Les résultats sont résumés dans le Tableau 2.4 pour les statistiques Y^2 et S^2 , et dans les Tableaux 5 et 6 pour les statistiques U^2 et Q^2 , en utilisant différentes méthodes d'estimation et des classes de données équiprobables et non équiprobables.

Classes de données équiprobables			
n	$\epsilon = 0.025$	$\epsilon = 0.05$	$\epsilon = 0.1$
Y^2			
20	(0.0238)	(0.0439)	(0.1131)
30	(0.0243)	(0.0441)	(0.1092)
50	(0.0247)	(0.0447)	(0.1081)
100	(0.0251)	(0.0516)	(0.1080)
150	(0.0249)	(0.0506)	(0.1056)
S^2			
20	(0.0259)	(0.0410)	(0.0857)
30	(0.0244)	(0.0416)	(0.0904)
50	(0.0246)	(0.0470)	(0.0923)
100	(0.0254)	(0.0476)	(0.1069)
150	(0.0251)	(0.0511)	(0.1000)

TABLE 4 – Valeurs de Y^2 et S^2 pour les classes é quiprobables.

n	$\epsilon = 0.025$	$\epsilon = 0.05$	$\epsilon = 0.1$	$\epsilon = 0.025$	$\epsilon = 0.05$	$\epsilon = 0.1$
$Q^2(\theta_{MLE})$			$U^2(\theta_{MLE})$			
20	(0.0257)	(0.0550)	(0.0924)	(0.0340)	(0.0864)	(0.1530)
30	(0.0254)	(0.0521)	(0.0933)	(0.0313)	(0.0554)	(0.1164)
50	(0.0253)	(0.0520)	(0.0943)	(0.0309)	(0.0485)	(0.1137)
100	(0.0248)	(0.0511)	(0.0988)	(0.0290)	(0.0486)	(0.1090)
150	(0.0252)	(0.0494)	(0.1009)	(0.0246)	(0.0533)	(0.0991)
$Q^2(\theta_{AD})$			$U^2(\theta_{AD})$			
20	(0.0262)	(0.0520)	(0.0932)	(0.0256)	(0.0598)	(0.1063)
30	(0.0255)	(0.0516)	(0.0956)	(0.0243)	(0.0527)	(0.1057)
50	(0.0245)	(0.0509)	(0.0967)	(0.0254)	(0.0511)	(0.1030)
100	(0.0253)	(0.0494)	(0.0982)	(0.0252)	(0.0505)	(0.1016)
150	(0.0249)	(0.0498)	(0.0984)	(0.0250)	(0.0503)	(0.1003)
$Q^2(\theta_{MPS})$			$U^2(\theta_{MPS})$			
20	(0.0229)	(0.0573)	(0.0914)	(0.0262)	(0.0544)	(0.1042)
30	(0.0239)	(0.0564)	(0.0960)	(0.0259)	(0.0534)	(0.1040)
50	(0.0256)	(0.0559)	(0.1022)	(0.0244)	(0.0514)	(0.1035)
100	(0.0255)	(0.0496)	(0.0981)	(0.0247)	(0.0510)	(0.1019)
150	(0.0248)	(0.0499)	(0.1010)	(0.0252)	(0.0501)	(0.1012)
$Q^2(\theta_{CM})$			$U^2(\theta_{CM})$			
20	(0.0231)	(0.0588)	(0.0954)	(0.0259)	(0.0552)	(0.1070)
30	(0.0265)	(0.0546)	(0.0955)	(0.0258)	(0.0550)	(0.1060)
50	(0.0261)	(0.0523)	(0.0958)	(0.0243)	(0.0544)	(0.0952)
100	(0.0243)	(0.0492)	(0.1018)	(0.0255)	(0.0484)	(0.1048)
150	(0.0252)	(0.0474)	(0.0984)	(0.0246)	(0.0494)	(0.1004)

TABLE 5 – Valeurs de Q^2 et U^2 pour les estimateurs MLE , AD , MPS et CM pour classes de données é quiprobables.

n	$\epsilon = 0.025$	$\epsilon = 0.05$	$\epsilon = 0.1$	$\epsilon = 0.025$	$\epsilon = 0.05$	$\epsilon = 0.1$
$Q^2(\theta_{MLE})$			$U^2(\theta_{MLE})$			
20	(0.0269)	(0.0459)	(0.0908)	(0.0237)	(0.0443)	(0.1084)
30	(0.0241)	(0.0464)	(0.0931)	(0.0246)	(0.0567)	(0.1064)
50	(0.0257)	(0.0469)	(0.0945)	(0.0247)	(0.0535)	(0.1034)
100	(0.0252)	(0.0476)	(0.0952)	(0.0252)	(0.0533)	(0.1027)
150	(0.0249)	(0.0491)	(0.0965)	(0.0249)	(0.0522)	(0.0991)
$Q^2(\theta_{AD})$			$U^2(\theta_{AD})$			
20	(0.0263)	(0.0446)	(0.0904)	(0.0263)	(0.0566)	(0.0948)
30	(0.0241)	(0.0454)	(0.0907)	(0.0262)	(0.0534)	(0.0966)
50	(0.0258)	(0.0460)	(0.0911)	(0.0247)	(0.0531)	(0.1054)
100	(0.0246)	(0.0474)	(0.0939)	(0.0257)	(0.0515)	(0.1018)
150	(0.0250)	(0.0477)	(0.0952)	(0.0253)	(0.0510)	(0.1011)
$Q^2(\theta_{MPS})$			$U^2(\theta_{MPS})$			
20	(0.0263)	(0.0532)	(0.0919)	(0.0225)	(0.0554)	(0.1075)
30	(0.0258)	(0.0516)	(0.0938)	(0.0263)	(0.0542)	(0.1067)
50	(0.0257)	(0.0493)	(0.1054)	(0.0240)	(0.0461)	(0.1040)
100	(0.0248)	(0.0494)	(0.0957)	(0.0246)	(0.0507)	(0.0961)
150	(0.0249)	(0.0497)	(0.0966)	(0.0249)	(0.0503)	(0.1033)
$Q^2(\theta_{CM})$			$U^2(\theta_{CM})$			
20	(0.0264)	(0.0451)	(0.0918)	(0.0237)	(0.0530)	(0.1075)
30	(0.0253)	(0.0471)	(0.0958)	(0.0262)	(0.0512)	(0.1067)
50	(0.0248)	(0.0482)	(0.0964)	(0.0242)	(0.0491)	(0.1040)
100	(0.0249)	(0.0510)	(0.0975)	(0.0249)	(0.0494)	(0.0961)
150	(0.0251)	(0.0504)	(0.0983)	(0.0250)	(0.0499)	(0.1033)

TABLE 6 – Valeurs de Q^2 et U^2 pour les estimateurs MLE , AD , MPS et CM dans le cas de classes de données non-è quiprobables.

Nous observons que les niveaux de signification simulés pour les statistiques Y^2 , S^2 , U^2 et Q^2 sont très proches de leurs niveaux théoriques correspondants sous la loi du Khi-deux. Ainsi, on peut affirmer que toutes ces statistiques peuvent être utilisées pour valider la distribution UGE . Nous notons que la méthode d'estimation d'Anderson-Darling a donné les meilleurs résultats pour la statistique Q^2 , tandis que les estimateurs MPS et de Cramér-von Mises ont fourni de meilleurs résultats que le MLE pour la statistique U^2 .

À partir de cette étude, nous pouvons conclure que les utilisateurs ont le choix entre les méthodes d'estimation AD , CV , MPS ou MLE , ainsi que parmi l'ensemble des statistiques proposées, pour ajuster ce modèle.

2.4.2 Applications

Dans cette section, nous démontrons la puissance et la flexibilité des statistiques de test construites ainsi que la flexibilité de la distribution proposée, à travers deux ensembles de données réelles. Nous mettons en évidence l'utilité de ce modèle comme alternative aux distributions bornées concurrentes, telles que :

- **Distribution de Weibull inverse généralisée unitaire** : $F_{IGUW}(x) = e^{-A\left(\frac{C(1-x)}{x}\right)^B}$;
- **Distribution de Gompertz** : $F_{UG}(x) = e^{-C(x^{-D}-1)}$;
- **Distribution Burr-III unitaire** : $F_{UBurr-III}(x) = \left(1 + \left(\frac{1}{x} - 1\right)^d\right)^{-k}$;
- **Distribution de Kumaraswamy** : $F_{KUM}(x) = 1 - (1 - x^n)^s$.

Nous commencerons par estimer les paramètres des différentes distributions afin d'adapter correctement chaque modèle aux données réelles. Une fois cette étape réalisée, nous effectuerons des représentations graphiques pour chaque exemple, en affichant l'histogramme des données et en superposant les distributions ajustées. Cette approche permettra d'analyser visuellement le comportement des différentes distributions.

Ensuite, nous comparerons les distributions proposées en utilisant des critères de test classiques (présentés dans les tableaux 2.7 et 2.9). Enfin, nous appliquerons les tests de Khi-deux modifiés, à savoir les statistiques U^2 , Y^2 , S^2 et Q^2 , pour évaluer la qualité de l'ajustement de la distribution proposée aux données. Les

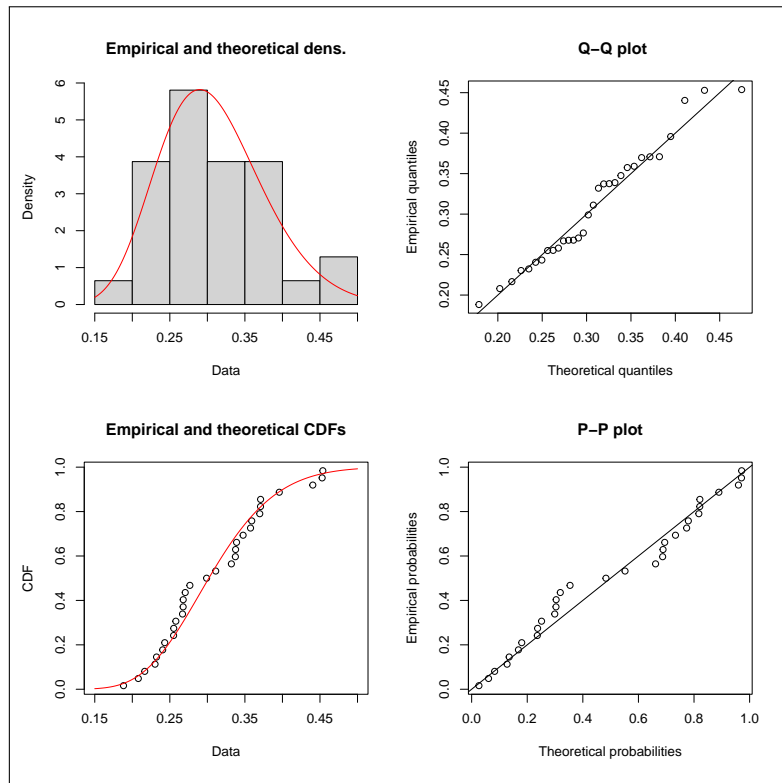


FIGURE 2 – pdf, CDF, PP-plot et QQ-plot pour les données de la force exercée sur le verre des fenêtres de l’avion.

calculs détaillés pour ces tests seront fournis en fin de chaque exemple.

Exemple 1 :

Les observations correspondent aux valeurs de la force exercée sur le verre des fenêtres de l’avion rapportées par Fuller et al. (1994), divisées par 100 afin d’obtenir des données comprises entre $[0, 1]$:

- 18.83, 20.80, 21.657, 23.03, 23.23, 24.05, 24.321, 25.50, 25.52, 25.80,
- 26.69, 26.77, 26.78, 27.05, 27.67, 29.90, 31.11, 33.20, 33.73, 33.76, 33.89,
- 34.76, 35.75, 35.91, 36.98, 37.08, 37.09, 39.58, 44.045, 45.29, 45.381.

Les pdf et cdf empiriques et théoriques de ces données sont représentées dans la Figure 2.2 ainsi que les PP-plots et QQ-plots.

Distributions	Estimateurs	LL	AIC	BIC
UGE(α, λ)	$\alpha_{MLE} = 21.0305$ $\lambda_{MLE} = 7.9277$ $\alpha_{AD} = 15.2618$ $\lambda_{AD} = 7.2007$	38.7757	-73.5515	-70.6835
		38.5395	-73.0791	-70.2111
UGIW(B, A, C)	$B_{MLE} = 3.3441$ $A_{MLE} = 0.3805$ $C_{MLE} = 0.4938$ $B_{AD} = 6.07 \times 10^{-10}$ $A_{AD} = 0.6930$ $C_{AD} = 0.8426$	38.2989	-70.5978	-66.2958
		38.2989	-70.5978	-66.2958
Kum(s, n)	$s_{MLE} = 152.05$ $n_{MLE} = 4.6213$ $s_{AD} = 152.05$ $n_{AD} = 4.6213$	37.2578	-70.5157	-67.6477
		37.2578	-70.5157	-67.6477
UBurr – III(k, d)	$k_{MLE} = 0.0170$ $d_{MLE} = 70.8399$ $k_{AD} = 0.0173$ $d_{AD} = 54.0221$	23.6859	-43.3718	-40.5038
		22.8039	-41.6078	-38.7398
UW(b, a)	$b_{MLE} = 5.8993$ $a_{MLE} = 0.2133$ $a_{AD} = 5.5019$ $b_{AD} = 0.2376$	38.7082	-73.4164	-70.5485
		38.5797	-73.1594	-70.2914
UG(D, C)	$D_{MLE} = 4.6205$ $C_{MLE} = 0.0022$ $D_{AD} = 4.4654$ $C_{AD} = 0.0028$	37.5053	-71.0106	-68.1426
		37.4722	-70.9444	-68.0764

TABLE 7 – Critères de sélection de modèle pour les données de Fuller (1994) pour le modèle UGE par rapport aux alternatives.

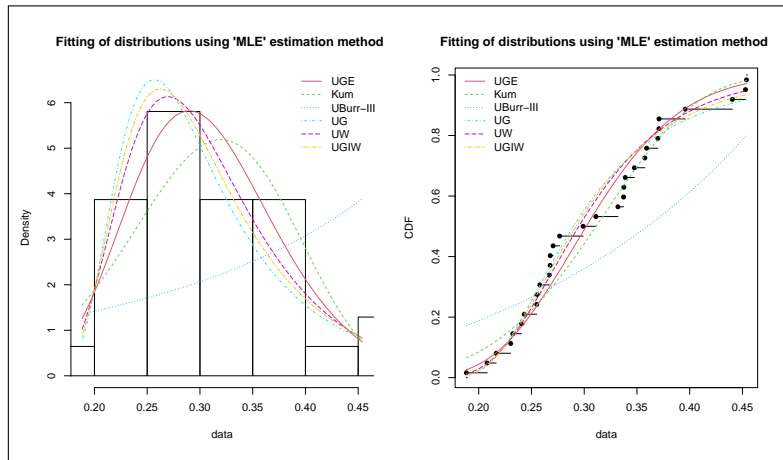


FIGURE 3 – Ajustement des distributions avec l’estimation par la méthode ‘MLE’ pour les valeurs de la force exercée sur le verre de la fenêtre de l’avion.

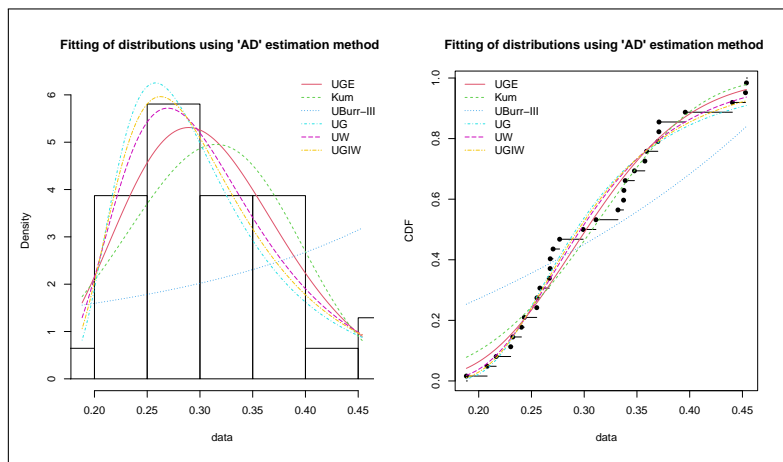


FIGURE 4 – Ajustement des distributions avec l’estimation par la méthode ‘AD’ pour les valeurs de la force exercée sur le verre de la fenêtre de l’avion.

Basé sur les estimations par maximum de vraisemblance (MLE) et Anderson-Darling (AD), les distributions ajustées sont comparées par les critères classiques de sélection de modèles. Les résultats obtenus (Tableau 2.7), ainsi que les graphiques des pdf et des cdf pour les différents modèles (Figures 2.3 et 2.4), prouvent que ce nouveau modèle peut être utilisé à la place de ses alternatives pour décrire les valeurs de la force exercée sur le verre des fenêtres de l'avion.

Pour les nouvelles statistiques de test du khi-deux modifié Y^2 , U^2 , S^2 et Q^2 , nous choisissons, par exemple, $r = 7$ classes de regroupement des données et les estimateurs MLE. Les calculs intermédiaires sont donnés ci-dessous :

$$B(\hat{\theta}_{MLE}) = \begin{bmatrix} -0.02295 & -0.01005 \\ -0.00065 & 0.00545 \\ 0.00999 & 0.01361 \\ 0.01307 & 0.17182 \\ 0.11499 & 0.04882 \\ -0.00599 & -0.05866 \\ -0.11726 & -0.15573 \end{bmatrix}$$

$$L = (L_1 = 0.01006, \quad L_2 = 0.37676)$$

$$J(\hat{\theta}_{MLE}) = B^T(\hat{\theta}_{MLE})B(\hat{\theta}_{MLE}) = \begin{bmatrix} 0.02781 & 0.02684 \\ 0.02684 & 0.05992 \end{bmatrix}$$

$$I_n(\hat{\theta}_{MLE}) = \begin{bmatrix} 2.026 \times 10^{-10} & -3.660 \times 10^{-5} \\ -3.660 \times 10^{-5} & 6.610947 \end{bmatrix}$$

$$Z = (0.72820, 0.27153, 0.27153, -1.15403, -0.20365, 0.74672, 0.24321)$$

$$W = \frac{1}{\sqrt{31}}(-2.131 \times 10^{-5}, 14.31)$$

Les résultats obtenus sont donnés dans le Tableau 2.8.

	Y^2	S^2	U^2	Q^2
Valeurs des statistiques	2.9251	1.2952	1.1152	2.5793
Valeurs critiques	$X_{r-1}^2 = 12.59$	$X_s^2 = 5.99$	$X_{r-s-1}^2 = 9.49$	$X_{r-1}^2 = 12.59$

TABLE 8 – Valeurs des statistiques pour les estimateurs MLE

Pour un niveau de signification $\varepsilon = 0.05$, les valeurs des statistiques sont inférieures à celles des valeurs critiques correspondantes (Tableau 8). Ainsi, nous concluons que tous les tests ont confirmé l'hypothèse nulle H_0 selon laquelle la distribution UGE décrit convenablement ces observations.

Dans la deuxième partie, en se basant sur les estimateurs d'Anderson-Darling, nous calculons U^2 et Q^2 . Les composants des matrices correspondantes sont :

$$B(\hat{\theta}_{AD}) = \begin{bmatrix} -0.02528 & -0.01385 \\ -0.00090 & 0.00751 \\ 0.01377 & 0.01875 \\ 0.01647 & 0.14108 \\ 0.11700 & 0.05218 \\ -0.00261 & -0.05624 \\ -0.11711 & -0.14285 \end{bmatrix}$$

$$L = \left(L_1 = 0.16886, \quad L_2 = -0.08795 \right)$$

$$J(\hat{\theta}_{AD}) = B^T(\hat{\theta}_{AD})B(\hat{\theta}_{AD}) = \begin{bmatrix} 0.02851 & 0.02591 \\ 0.02591 & 0.04679 \end{bmatrix}$$

$$I_n(\hat{\theta}_{AD}) = \begin{bmatrix} 0.00046 & 0.05505 \\ 0.05505 & 6.46184 \end{bmatrix}$$

$$Z = (-0.07968, 1.22191, 0.27153, -1.15403, -0.20365, 0.74672, 0.41184)$$

$$W = \frac{1}{\sqrt{31}}(0.1206, 14.15)$$

Nous obtenons alors :

$$Q^2(\tilde{\theta}_{AD}) = 3.5605 \quad \text{et} \quad U^2(\tilde{\theta}_{AD}) = 2.0282.$$

Pour un niveau de signification $\varepsilon = 0.05$, les valeurs critiques correspondantes du khi-deux sont $X_{r-1}^2 = 12.59$ et $X_{r-s-1}^2 = 9.49$. Ainsi, l'hypothèse nulle H_0 ne peut pas être rejetée, ce qui confirme les résultats obtenus par les autres méthodes.

Exemple 2

Cet exemple concerne la flexibilité de ce nouveau modèle par rapport à des données asymétriques à gauche. Nous supposons que les observations du temps jusqu'à la défaillance d'un type de turbocompresseur étudié par Xu et al. (2003) suivent la distribution UGE . Les temps de défaillance observés sont donnés ci-dessous et divisés par 10, en utilisant la même méthode appliquée aux premières données pour obtenir des valeurs dans $[0, 1]$:

4.8, 6.5, 7.0, 7.3, 7.7, 8.0, 8.4, 1.6, 3.5, 3.0, 4.6, 2.0, 3.9, 5.0, 5.6, 6.1, 6.5, 7.1, 7.3, 7.8,
8.1, 8.4, 2.6, 4.5, 5.1, 5.8, 6.3, 6.7, 7.3, 7.7, 7.9, 8.3, 8.5, 5.3, 6.0, 8.7, 8.8, 5.4, 6.0, 9.0.

Pour tester cette hypothèse, et sans perte de généralité, nous appliquons nos résultats aux cas des estimateurs de maximum de vraisemblance et d'Anderson-Darling. Mais d'abord, nous représentons les pp-plots et les qq-plots de ces observations (Figure 2.5). Les méthodes classiques de sélection de modèles ont montré que notre modèle convient mieux à ces observations de temps de défaillance que les alternatives mentionnées plus haut (Tableau 2.9).

Distributions	Estimateurs	LL	AIC	BIC
$UGE(\alpha, \lambda)$	$\alpha_{MLE} = 1.5853$ $\lambda_{MLE} = 0.5122$	13.8294	-23.6588	-20.2811
	$\alpha_{AD} = 1.4531$ $\lambda_{AD} = 0.4825$	13.7543	-23.5087	-20.1309
$UGIW(B, A, C)$	$B_{MLE} = 1.0286$ $A_{MLE} = 0.7545$ $C_{MLE} = 1.4651$	7.7719	-9.5439	-4.4773
	$B_{AD} = 2.9 \times 10^{-09}$ $A_{AD} = 0.6931$ $C_{AD} = 0.8347$	-761.72	1529.44	1534.51
$Kum(s, n)$	$s_{MLE} = 2.4924$ $n_{MLE} = 3.1564$	12.3428	-20.6857	-17.3079
	$s_{AD} = 2.2269$ $n_{AD} = 3.0377$	12.2268	-20.4536	-17.0758
$UBurr - III(k, d)$	$k_{MLE} = 1.8753$ $d_{MLE} = 1.5006$	11.1585	-18.3171	-14.9394
	$k_{AD} = 1.9567$ $d_{AD} = 1.4651$	11.1054	-18.2109	-14.8331
$UW(b, a)$	$b_{MLE} = 1.4611$ $a_{MLE} = 2.1273$	10.6883	-17.3767	-13.9989
	$b_{AD} = 1.4614$ $a_{AD} = 2.2538$	10.6205	-17.2410	-13.8632
$UG(D, C)$	$D_{MLE} = 0.6951$ $C_{MLE} = 1.9321$	8.1274	-12.2549	-8.8771
	$D_{AD} = 0.9971$ $C_{AD} = 1.2052$	7.6749	-11.3498	-7.9721

TABLE 9 – Sélection classique de modèles de l'UGE par rapport aux alternatives.

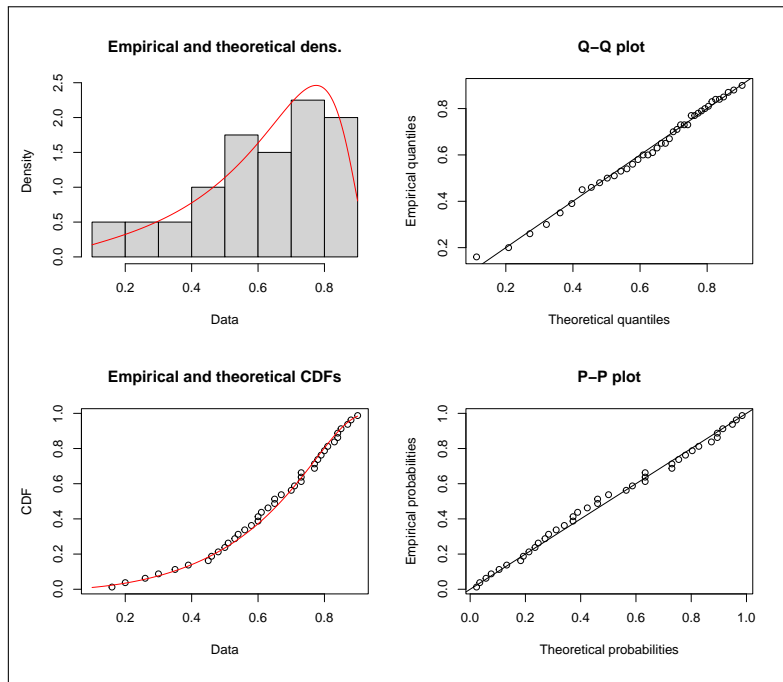


FIGURE 5 – pdf, cdf, qq-plot et pp-plot du temps de défaillance du turbocompresseur.

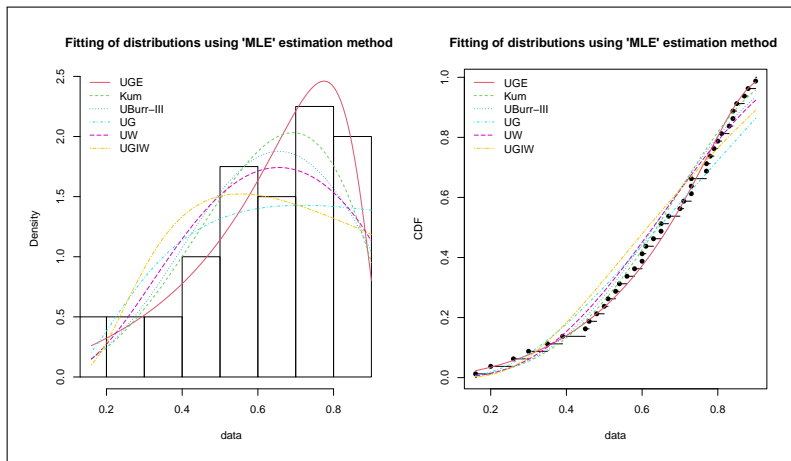


FIGURE 6 – Ajustement des distributions avec l'estimation par la méthode 'MLE' pour le temps de défaillance (103h) du turbocompresseur d'un type de moteur.

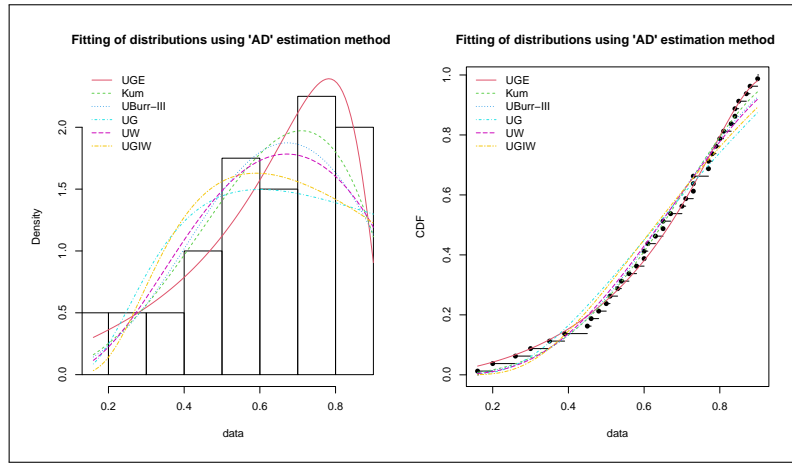


FIGURE 7 – Ajustement des distributions avec l'estimation par la méthode d'Anderson Darling pour le temps de défaillance (103h) du turbocompresseur d'un type de moteur.

Comme montré dans la Figure 2.6 et la Figure 2.7, la distribution UGE s'adapte mieux à ces données que ses concurrentes. Pour confirmer ce fait, nous calculons les statistiques de test correspondantes Y^2 , U^2 , S^2 et Q^2 . Si nous choisissons par exemple $r = 6$ classes de regroupement, les calculs nécessaires sont donnés pour les méthodes d'estimation MLE et AD respectivement :

$$B(\hat{\theta}) = \begin{bmatrix} -0.32672 & -0.10441 \\ 0.03033 & 0.11783 \\ 0.18290 & 0.21065 \\ 0.86594 & 0.71828 \\ 0.40619 & 0.03509 \\ -0.48356 & -1.36337 \end{bmatrix}$$

$$L = \left(L_1 = 1.305785, \quad L_2 = 3.416131 \right)$$

$$J(\hat{\theta}) = B^T(\hat{\theta})B(\hat{\theta}) = \begin{bmatrix} 1.2898 & 1.3717 \\ 1.3717 & 2.4451 \end{bmatrix}$$

$$I_n(\hat{\theta}) = \begin{bmatrix} 8.701 \times 10^{-09} & -0.00152 \\ -0.00152 & 266.422 \end{bmatrix}$$

$$Z = (0.1081749, 0.5163978, 0.1290994, -0.2581989, -0.2581989, 0.3916068)$$

$$W = \frac{1}{\sqrt{40}}(-5.90 \times 10^{-4}, 103.23)$$

$$B(\hat{\theta}_{AD}) = \begin{bmatrix} -0.32961 & -0.11391 \\ 0.03309 & 0.12855 \\ 0.19954 & 0.22465 \\ 0.82445 & 0.73146 \\ 0.42394 & 0.05082 \\ -0.47927 & -1.34270 \end{bmatrix}$$

$$L = \left(L_1 = 1.6284, \quad L_2 = 4.0360 \right)$$

$$J(\hat{\theta}_{AD}) = B^T(\hat{\theta}_{AD})B(\hat{\theta}_{AD}) = \begin{bmatrix} 1.2387 & 1.3547 \\ 0.5093 & 0.5733 \end{bmatrix}$$

$$I_n(\hat{\theta}_{AD}) = \begin{bmatrix} 0.0228 & 2.4923 \\ 2.4923 & 271.470 \end{bmatrix}$$

$$Z = (0.2131557, 0.5163978, 0.1290994, -0.2581989, -0.2581989, 0.4469937)$$

$$W = \frac{1}{\sqrt{40}}(-2.2024, 53.5212)$$

Nous obtenons, pour un niveau de signification $\epsilon = 0.05$:

	Y^2	S^2	U^2	Q^2
valeurs des statistiques (MLE)	0.7864	0.0112	0.3657	0.3936
valeurs des statistiques (AD)			0.3912	0.4242
valeurs critiques	$X_{r-1}^2 = 11.07$	$X_s^2 = 5.99$	$X_{r-s-1}^2 = 7.81$	$X_{r-1}^2 = 11.07$

TABLE 10 – Valeurs des critères de test pour les estimateurs MLE et AD

Comme prévu, l'hypothèse nulle H_0 n'est pas rejetée. Nous pouvons en déduire que cette nouvelle distribution peut également modéliser des données asymétriques.

La nouvelle distribution à queue lourde

3.1 Introduction

La modélisation statistique des phénomènes extrêmes constitue un enjeu majeur dans de nombreux domaines d'application, tels que la finance, l'ingénierie, les télécommunications et les sciences de l'environnement. Dans ces contextes, les distributions classiques, souvent « à queue légère », s'avèrent insuffisantes pour représenter adéquatement la probabilité d'événements rares mais critiques. C'est dans ce cadre que les *distributions à queue lourde* trouvent toute leur pertinence, en permettant de mieux saisir la fréquence et l'intensité de tels événements.

Parmi les familles de lois les plus utilisées, on distingue deux grandes catégories : d'une part, les distributions issues de la théorie des valeurs extrêmes, telles que la distribution généralisée des valeurs extrêmes (GEV) (Gumbel, 1958), les lois de Fréchet (1927) ou de Lévy (1937), et d'autre part, des lois de durée de vie flexibles comme la Weibull (1939), la log-normale, la log-logistique ou l'« Inverse Weibull » (Dey, 2014). Chacune de ces distributions présente des propriétés spécifiques en termes de forme de queue, de comportement asymptotique, ou encore de structure de la fonction de risque.

Malgré cette diversité, la littérature souligne certaines limites : complexité des fonctions de vraisemblance, difficulté d'interprétation paramétrique, ou encore rigidité dans l'ajustement de données asymétriques ou multimodales. Pour répondre à ces défis, plusieurs extensions ont été proposées, notamment via des mécanismes de pondération ou de biais. Parmi elles, l'approche d'Azzalini (1985), fondée sur une généralisation de la symétrie, a permis de construire des familles de lois plus

flexibles tout en conservant des propriétés analytiques maîtrisables.

Dans ce contexte, notre travail s'inscrit dans la continuité des recherches récentes autour de la *New Distribution (ND)*, introduite par Doostmoradi (2018) et approfondie par Aidi et Seddik-Ameur (2021), qui ont démontré sa capacité à modéliser efficacement des données de durée de vie à l'aide de seulement deux paramètres. Toutefois, cette distribution reste encore limitée face à certaines formes extrêmes de queue ou d'asymétrie.

Nous proposons ici une extension originale de la ND, fondée sur une pondération paramétrée inspirée de la méthode d'Azzalini (1985). Initialement appliquée à la distribution normale pour introduire une asymétrie contrôlée. Cette approche s'est révélée particulièrement efficace pour développer des distributions asymétriques, aussi bien à gauche qu'à droite. Depuis, de nombreux travaux ont étendu cette méthode à d'autres lois de probabilité, soulignant sa flexibilité et sa pertinence dans la modélisation de données asymétriques.

Dans notre étude, cette technique de pondération est utilisée pour construire une nouvelle loi que nous appelons *Weighted New Distribution (WND)*. Cette distribution à deux paramètres offre une plus grande souplesse dans la modélisation des queues lourdes et des asymétries, tout en conservant une structure paramétrique simple. Elle est particulièrement bien adaptée aux situations pratiques où les données présentent à la fois des durées extrêmes, de la censure, et des formes de risque atypiques.

L'analyse théorique de la WND est accompagnée d'une étude empirique approfondie, dans laquelle nous comparons ses performances à celles de modèles de référence, tels que les distributions GEV, log-normale, Inverse Weibull, Weibull ou log-logistique. L'estimation des paramètres est assurée par des méthodes robustes, notamment le maximum de vraisemblance et le *Maximum Product of Spacings*, et l'évaluation des ajustements s'appuie sur des critères standards (AIC, BIC, log-vraisemblance). Afin de compléter cette validation, deux nouveaux tests statistiques basés sur U^2 et Q^2 , sont développés pour juger de l'adéquation du modèle aux données observées.

Ce chapitre apporte ainsi une double contribution. Sur le plan *théorique*, il enrichit le corpus des lois à queue lourde par l'introduction d'une distribution

flexible et rigoureuse. Sur le plan *pratique*, il fournit aux chercheurs et aux utilisateurs un outil performant pour analyser des données extrêmes complexes, en particulier dans des contextes de censure ou de comportement non standard de la fonction de risque.

3.2 Distribution Weighted-ND

La distribution ND est un modèle probabiliste flexible introduit par Doostmoadi (2018) pour décrire un large éventail de phénomènes dans diverses disciplines. Sa fonction de densité de probabilité (pdf) est définie comme suit :

$$g(t) = \alpha^2 \beta t^{2\beta-1} e^{-\alpha t^\beta},$$

avec $t > 0$, $\alpha > 0$ et $\beta > 0$ des paramètres qui contrôlent la forme et l'échelle de la distribution. Le paramètre α régit principalement l'échelle de la distribution, tandis que β influence le comportement de la queue et sa flexibilité globale. La fonction de répartition (CDF) correspondante s'exprime ainsi :

$$G(t) = 1 - (1 + \alpha t^\beta) e^{-\alpha t^\beta}.$$

Dans cette section, nous présentons une méthode puissante pour améliorer la flexibilité des distributions de probabilité standard afin de mieux modéliser des données présentant des caractéristiques de longues queues. Cette méthode, proposée à l'origine par Azzalini (1985), consiste à modifier la distribution en y incorporant un mécanisme de pondération ainsi que des paramètres supplémentaires, permettant ainsi une meilleure adaptation à diverses structures de données.

L'idée principale est de partir d'une distribution de base, notée $g(t)$, et de construire une nouvelle distribution à queue lourde en introduisant un paramètre d'échelle $\lambda > 0$ et un coefficient de normalisation $K > 0$. La fonction de répartition (CDF) est définie par :

$$F_W(t) = \int_0^t K \cdot g(x) S(\lambda x) dx$$

où $S(\lambda x)$ est la version mise à l'échelle de la fonction de survie de la distribution de base $S(t)$, λ agit comme un paramètre supplémentaire qui ajuste le

comportement de la distribution, et K assure la normalisation correcte de la nouvelle distribution pondérée. Le coefficient K est déterminé de la manière suivante pour satisfaire la condition de probabilité totale :

$$K = \frac{1}{\int_0^{\infty} g(x)S(\lambda x) dx}.$$

Cette modification ne préserve pas seulement les propriétés structurelles de la distribution originale $g(t)$, mais elle introduit également une flexibilité supplémentaire grâce au paramètre λ , permettant de capturer une gamme plus étendue de motifs dans les données, y compris des comportements à longues queues ou asymétriques. La formulation d'Azzalini offre un cadre systématique pour développer des distributions pondérées qui conservent leur validité probabiliste tout en s'adaptant aux complexités des données réelles. En appliquant cette approche à une distribution classique $g(t)$, une nouvelle famille de distributions émerge, pouvant être ajustée pour mieux correspondre à des ensembles de données caractérisés par des valeurs extrêmes, un manque de symétrie ou d'autres irrégularités.

• **Définir le coefficient de normalisation de la distribution WND:**

$$f(t) = k\alpha^2\beta t^{2\beta-1} (1 + \alpha(\lambda t)^\beta) e^{-(\alpha+\alpha\lambda^\beta)t^\beta} \quad (3.1)$$

$$F(t) = \int_0^t f(x)dx = \int_0^t k\alpha^2\beta x^{2\beta-1} (1 + \alpha(\lambda x)^\beta) e^{-(\alpha+\alpha\lambda^\beta)x^\beta} dx \quad (3.2)$$

Pour trouver le coefficient de normalisation k , nous devons calculer l'équation suivante :

$$\begin{aligned} k &= \frac{1}{\int_0^{\infty} \alpha^2\beta x^{2\beta-1} (1 + \alpha(\lambda x)^\beta) e^{-(\alpha+\alpha\lambda^\beta)x^\beta} dx} \\ u &= (\lambda t)^\beta, \quad x = \frac{u^{\frac{1}{\beta}}}{\lambda}, \quad dx = \frac{u^{\frac{1}{\beta}-1}}{\beta\lambda} \\ &= \frac{\alpha^2}{\lambda^{2\beta}} \int_0^t \frac{\alpha^2}{\lambda^{2\beta}} u (1 + \alpha u) e^{-(\frac{\alpha}{\lambda^\beta} + \alpha)u} du \\ &= \frac{\alpha^2}{\lambda^{2\beta}} \int_0^t \left[u e^{-(\frac{\alpha}{\lambda^\beta} + \alpha)u} + \alpha u^2 e^{-(\frac{\alpha}{\lambda^\beta} + \alpha)u} \right] du \\ &= \frac{\alpha^2}{\lambda^{2\beta}} \left(\frac{\alpha}{\lambda^\beta} + \alpha \right)^{-1} \int_0^t \left[\left(\frac{\alpha}{\lambda^\beta} + \alpha \right) u e^{-(\frac{\alpha}{\lambda^\beta} + \alpha)u} \right] du + \\ &= \frac{\alpha^3}{\lambda^{2\beta}} \left(\frac{\alpha}{\lambda^\beta} + \alpha \right)^{-2} \int_0^t \left[\left(\frac{\alpha}{\lambda^\beta} + \alpha \right)^2 u^2 e^{-(\frac{\alpha}{\lambda^\beta} + \alpha)u} \right] du \\ &= \frac{\alpha^2}{\lambda^{2\beta}} \left(\frac{\alpha}{\lambda^\beta} + \alpha \right)^{-1} \int_0^t \left[\left(\frac{\alpha}{\lambda^\beta} + \alpha \right) u e^{-(\frac{\alpha}{\lambda^\beta} + \alpha)u} \right] du + \\ &= \frac{\alpha^3}{\lambda^{2\beta}} \left(\frac{\alpha}{\lambda^\beta} + \alpha \right)^{-2} \int_0^t \left[\left(\left(\frac{\alpha}{\lambda^\beta} + \alpha \right) u \right)^2 e^{-(\frac{\alpha}{\lambda^\beta} + \alpha)u} \right] du \\ z &= \left(\frac{\alpha}{\lambda^\beta} + \alpha \right) u, \quad u = \frac{z}{\left(\frac{\alpha}{\lambda^\beta} + \alpha \right)}, \quad du = \frac{dz}{\left(\frac{\alpha}{\lambda^\beta} + \alpha \right)} \end{aligned}$$

$$\begin{aligned}
 &= \frac{\alpha^2}{\lambda^{2\beta}} \left(\frac{\alpha}{\lambda^\beta} + \alpha \right)^{-2} \int_0^t [ze^{-z}] dz + \frac{\alpha^3}{\lambda^{2\beta}} \left(\frac{\alpha}{\lambda^\beta} + \alpha \right)^{-3} \int_0^t [z^2 e^{-z}] dz \\
 &\quad \int_0^\infty [z^{n-1} e^{-z}] dz = \Gamma(n)
 \end{aligned}$$

$$k = \left[\frac{\alpha^2}{(\alpha + \alpha\lambda^\beta)^2} + \frac{2\alpha^3\lambda^\beta}{(\alpha + \alpha\lambda^\beta)^3} \right]^{-1} = \frac{(1 + \lambda^\beta)^3}{1 + 3\lambda^\beta}$$

Alors la fonction de densité de la distribution WND est dè fini comme suit:

$$f(t) = \frac{(1 + \lambda^\beta)^3}{1 + 3\lambda^\beta} \alpha^2 \beta t^{2\beta-1} (1 + \alpha(\lambda t)^\beta) e^{-(\alpha + \alpha\lambda^\beta)t^\beta} \quad (3.3)$$

En intégrant cette fonction, on obtient la fonction de répartition suivante :

$$F(t) = 1 - \left[\frac{\lambda^\beta(\alpha + \alpha\lambda^\beta)^2}{1 + 3\lambda^\beta} t^{2\beta} + (\alpha + \alpha\lambda^\beta)t^\beta + 1 \right] e^{-(\alpha + \alpha\lambda^\beta)t^\beta} \quad (3.4)$$

Utilisant la fonction de densité et de ré partition on trouve les fonctions de fiabilité $S(t)$ et de hasard $h(t)$ suivantes:

$$S(t) = \left[\frac{\lambda^\beta(\alpha + \alpha\lambda^\beta)^2}{1 + 3\lambda^\beta} t^{2\beta} + (\alpha + \alpha\lambda^\beta)t^\beta + 1 \right] e^{-(\alpha + \alpha\lambda^\beta)t^\beta} \quad (3.5)$$

$$h(t) = \frac{g(t)}{S(t)} = \frac{k\alpha^2\beta t^{2\beta-1} (1 + \alpha(\lambda t)^\beta)}{\frac{\lambda^\beta(\alpha + \alpha\lambda^\beta)^2}{1 + 3\lambda^\beta} t^{2\beta} + (\alpha + \alpha\lambda^\beta)t^\beta + 1} \quad (3.6)$$

On présente quelques formes de la densité probabilitèet du taux de hasard de la distribution *weighted - ND* :

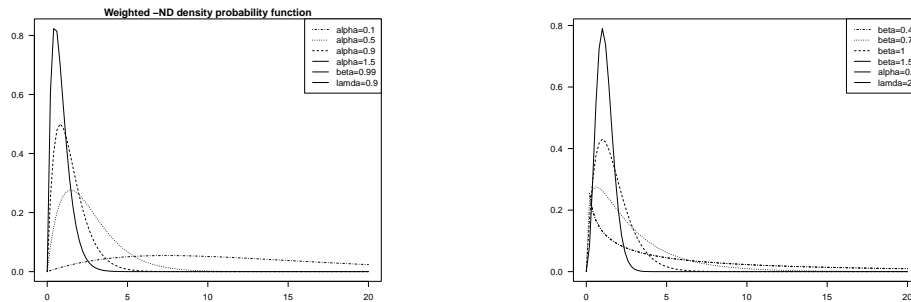
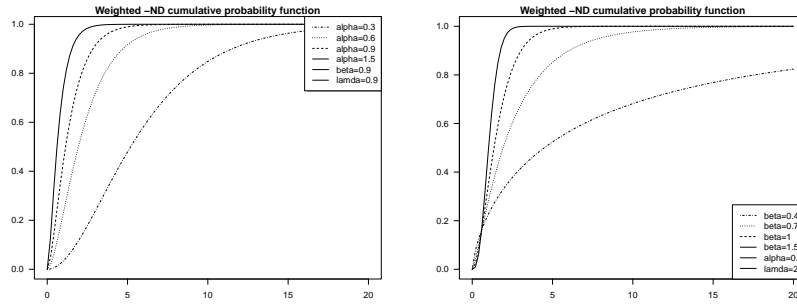
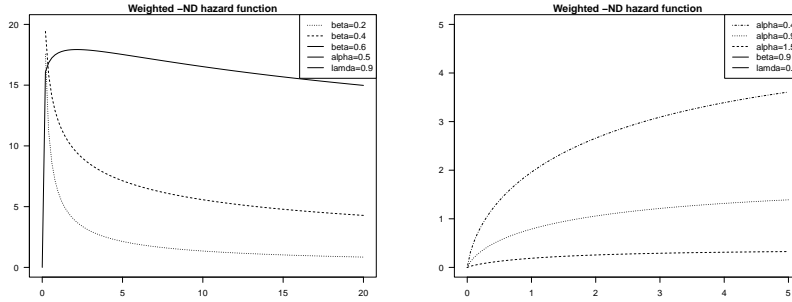


Figure 8 : Weighted-ND Density distribution


 Figure 9 : *weighted* – *ND* Cumulative density distribution

 Figure 10 : *weighted* – *ND* hazard function

3.3 Moments et méthodes d'estimation

$$E[t^s] = \int_0^{\infty} t^s g(t) dt \quad (3.7)$$

En appliquant cette définition à notre distribution, nous trouvons :

$$E[t^s] = \int_0^{\infty} t^s k \alpha^2 \beta t^{2\beta-1} \exp^{-(\alpha+\alpha\lambda^\beta)t^\beta} dt \quad (3.8)$$

Pour trouver cette intégrale, nous devons effectuer un changement de variable, puis essayer d'obtenir une forme de la distribution gamma. En appliquant ces deux astuces, nous obtenons le moment d'ordre s comme suit :

$$E[t^s] = \frac{k\alpha^2}{(\alpha + \alpha\lambda^\beta)^{1+\frac{s}{\beta}}} \Gamma\left(2 + \frac{s}{\beta}\right) + \frac{k\alpha^3\lambda^\beta}{(\alpha + \alpha\lambda^\beta)^{2+\frac{s}{\beta}}} \Gamma\left(3 + \frac{s}{\beta}\right). \quad (3.9)$$

Nous allons ensuite fournir les premier, deuxième, troisième et quatrième moments en fonction du moment d'ordre s , afin de déterminer clairement les coefficients d'échelle et d'aplatissement.

$$m_1 = E[t] = \frac{k\alpha^2}{(\alpha + \alpha\lambda^\beta)^{1+\frac{1}{\beta}}} \Gamma\left(2 + \frac{1}{\beta}\right) + \frac{k\alpha^3\lambda^\beta}{(\alpha + \alpha\lambda^\beta)^{2+\frac{1}{\beta}}} \Gamma\left(3 + \frac{1}{\beta}\right) \quad (3.10)$$

$$m_2 = E[t^2] = \frac{k\alpha^2}{(\alpha + \alpha\lambda^\beta)^{1+\frac{2}{\beta}}} \Gamma\left(2 + \frac{2}{\beta}\right) + \frac{k\alpha^3\lambda^\beta}{(\alpha + \alpha\lambda^\beta)^{2+\frac{2}{\beta}}} \Gamma\left(3 + \frac{2}{\beta}\right) \quad (3.11)$$

$$m_3 = E[t^3] = \frac{k\alpha^2}{(\alpha + \alpha\lambda^\beta)^{1+\frac{3}{\beta}}} \Gamma\left(2 + \frac{3}{\beta}\right) + \frac{k\alpha^3\lambda^\beta}{(\alpha + \alpha\lambda^\beta)^{2+\frac{3}{\beta}}} \Gamma\left(3 + \frac{3}{\beta}\right) \quad (3.12)$$

$$m_4 = E[t^4] = \frac{k\alpha^2}{(\alpha + \alpha\lambda^\beta)^{1+\frac{4}{\beta}}} \Gamma\left(2 + \frac{4}{\beta}\right) + \frac{k\alpha^3\lambda^\beta}{(\alpha + \alpha\lambda^\beta)^{2+\frac{4}{\beta}}} \Gamma\left(3 + \frac{4}{\beta}\right) \quad (3.13)$$

Ensuite, nous pouvons calculer les coefficients d'aplatissement et d'échelle en remplaçant les moments précédents dans les formules suivantes :

Coefficient d'asymétrie :

$$S = \frac{E[t^3]}{Var[t]^{\frac{3}{2}}} = \frac{\frac{k\alpha^2}{(\alpha+\alpha\lambda^\beta)^{1+\frac{3}{\beta}}} \Gamma\left(2 + \frac{3}{\beta}\right) + \frac{k\alpha^3\lambda^\beta}{(\alpha+\alpha\lambda^\beta)^{2+\frac{3}{\beta}}} \Gamma\left(3 + \frac{3}{\beta}\right)}{(E[t^2] - E[t]^2)^{\frac{3}{2}}} \quad (3.14)$$

Puisque la variance est toujours positive, le signe du coefficient d'aplatissement dépend du troisième moment, qui est également positif dans notre cas.

Coefficient d'aplatissement :

$$Kr = \frac{E[t^4]}{(E[t^2] - E[t]^2)^2} = \frac{\frac{k\alpha^2}{(\alpha+\alpha\lambda^\beta)^{1+\frac{4}{\beta}}} \Gamma\left(2 + \frac{4}{\beta}\right) + \frac{k\alpha^3\lambda^\beta}{(\alpha+\alpha\lambda^\beta)^{2+\frac{4}{\beta}}} \Gamma\left(3 + \frac{4}{\beta}\right)}{Var[t]^2} \quad (3.15)$$

Méthode de maximum de vraisemblance

Pour t_1, t_2, \dots, t_n , une réalisation d'un échantillon aléatoire T_1, T_2, \dots, T_n , on cherche à estimer $\theta = (\alpha, \beta, \lambda)^T$, le vecteur des paramètres de la distribution *weighted-ND*, en utilisant la méthode du maximum de vraisemblance. Ainsi, la log-vraisemblance de notre distribution est donnée comme suit :

$$l = \ln(L(t_1, t_2, \dots, t_n; \theta)) = \sum_{i=1}^n \left[\ln(k) + 2 \ln(\alpha) + \ln(\beta) + (2\beta - 1) \ln(t_i) - (\alpha + \alpha\lambda^\beta) t_i^\beta \right] \quad (3.16)$$

Ensuite, on va annuler les fonctions de score suivantes pour obtenir l'estimateur $\hat{\theta} = (\hat{\alpha}, \hat{\beta}, \hat{\lambda})^T$:

$$\begin{cases} \sum_{i=1}^n \left[\frac{2}{\alpha} + \frac{\lambda^\beta t_i^\beta}{(1 + \alpha\lambda^\beta t_i^\beta)} - (1 + \lambda^\beta) t_i^\beta \right] \\ \sum_{i=1}^n \left[\frac{3\lambda^\beta \ln(\lambda)}{(1 + \lambda^\beta)} - \frac{3\lambda^\beta \ln(\lambda)}{(1 + 3\lambda^\beta)} + \frac{1}{\beta} + 2 \ln(t_i) + \frac{\alpha(\lambda t_i)^\beta \ln(\lambda t_i)}{(1 + \alpha(\lambda t_i)^\beta)} - \alpha t_i^\beta \ln(t_i) + \alpha(\lambda t_i)^\beta \ln(\lambda t_i) \right] \\ \sum_{i=1}^n \left[\frac{3\beta\lambda^{\beta-1}}{1 + \lambda^\beta} - \frac{3\beta\lambda^{\beta-1}}{1 + 3\lambda^\beta} + \frac{\alpha\beta t_i^\beta \lambda^{\beta-1}}{(1 + \alpha(\lambda t_i)^\beta)} - \alpha\beta t_i^\beta \lambda^{\beta-1} \right] \end{cases} \quad (3.17)$$

Information de Fisher ” I_n ”

Après de longs calculs, les éléments de la matrice d'information de Fisher sont obtenus:

$$I_{11} = \frac{-((\lambda t)^\beta (\lambda t)^\beta)}{(1 + \alpha(\lambda t)^\beta)^2} + \frac{2}{\alpha^2}$$

$$\begin{aligned} I_{22} = & \frac{3(\lambda^\beta \log(\lambda)) \log(\lambda)}{(1 + \lambda^\beta)} - \frac{3\lambda^\beta \log(\lambda)(\lambda^\beta \log(\lambda))}{(1 + \lambda^\beta)^2} \\ & - \frac{3(\lambda^\beta \log(\lambda)) \log(\lambda)}{(1 + 3\lambda^\beta)} - \frac{3\lambda^\beta \log(\lambda)(3(\lambda^\beta \log(\lambda)))}{(1 + 3\lambda^\beta)^2} \\ & - \frac{1}{\beta^2} + \frac{\alpha((\lambda t)^\beta \log((\lambda t))) \log(\lambda t)}{(1 + \alpha(\lambda t)^\beta)} \\ & - \frac{\alpha(\lambda t)^\beta \log(\lambda t)(\alpha((\lambda t)^\beta \log((\lambda t))))}{(1 + \alpha(\lambda t)^\beta)^2} \\ & - \alpha(t^\beta \log(t)) \log(t) + \alpha((\lambda t)^\beta \log((\lambda t))) \log(\lambda t) \end{aligned}$$

$$\begin{aligned} I_{33} = & \frac{3\beta(\lambda^{(\beta-1)-1}(\beta-1))}{1 + \lambda^\beta} - \frac{3\beta\lambda^{\beta-1}(\lambda^{\beta-1}\beta)}{(1 + \lambda^\beta)^2} \\ & - \frac{3\beta(\lambda^{(\beta-1)-1}(\beta-1))}{1 + 3\lambda^\beta} - \frac{3\beta\lambda^{\beta-1}(3(\lambda^{\beta-1}\beta))}{(1 + 3\lambda^\beta)^2} \\ & + \frac{\alpha\beta(\lambda^{(\beta-1)-1}(\beta-1))t^\beta}{1 + \alpha(\lambda t)^\beta} \\ & - \frac{(\alpha\beta\lambda^{\beta-1}t^\beta)(\alpha((\lambda t)^{\beta-1}(\beta t)))}{(1 + \alpha(\lambda t)^\beta)^2} \\ & - \alpha\beta(\lambda^{(\beta-1)-1}(\beta-1))t^\beta \end{aligned}$$

$$\begin{aligned} I_{12} = & \frac{(\lambda t)^\beta \log((\lambda t))}{1 + \alpha(\lambda t)^\beta} - \frac{(\lambda t)^\beta (\alpha((\lambda t)^\beta \log((\lambda t))))}{(1 + \alpha(\lambda t)^\beta)^2} \\ & - ((\lambda)^\beta \log((\lambda))t^\beta + (1 + (\lambda)^\beta)(t^\beta \log(t))) \end{aligned}$$

$$I_{13} = \frac{(\lambda t)^{(\beta)-1}((\beta)t)}{(1 + \alpha(\lambda t)^\beta)} - \frac{(\lambda t)^\beta (\alpha((\lambda t)^{(\beta)-1}((\beta)t)))}{(1 + \alpha(\lambda t)^\beta)^2} - (\lambda)^{\beta-1}(\beta)t^\beta$$

$$\begin{aligned}
 I_{23} = & \frac{3(\lambda^{\beta-1}\beta) \log(\lambda) + 3\lambda^\beta \cdot \frac{1}{\lambda}}{1 + \lambda^\beta} - \frac{3\lambda^\beta \log(\lambda)(\lambda^{\beta-1}\beta)}{(1 + \lambda^\beta)^2} \\
 & - \frac{3(\lambda^{\beta-1}\beta) \log(\lambda) + 3\lambda^{\beta-1}}{1 + 3\lambda^\beta} - \frac{3\lambda^\beta \log(\lambda)(3(\lambda^{\beta-1}\beta))}{(1 + 3\lambda^\beta)^2} \\
 & + \alpha((\lambda t)^{\beta-1}(\beta t)) \log(\lambda t) + \frac{\alpha(\lambda t)^\beta \cdot \frac{t}{\lambda t}}{1 + \alpha(\lambda t)^\beta} \\
 & - \frac{\alpha(\lambda t)^\beta \log(\lambda t)(\alpha((\lambda t)^{\beta-1}(\beta t)))}{(1 + \alpha(\lambda t)^\beta)^2} \\
 & + \alpha((\lambda t)^{\beta-1}(\beta t)) \log(\lambda t) + \alpha(\lambda t)^\beta \cdot \frac{t}{\lambda t}
 \end{aligned}$$

Comme les expressions explicites des MLE $\hat{\alpha}_{\text{MLE}}$ et $\hat{\lambda}_{\text{MLE}}$ ne peuvent généralement pas être obtenues en raison de la complexité des équations résultantes, il est nécessaire de recourir à des méthodes numériques itératives. À cet effet, plusieurs algorithmes et packages implémentés dans le langage **R**, tels que **bbsolve**, permettent de résoudre efficacement ces systèmes non linéaires.

En pratique, l'obtention des estimateurs du maximum de vraisemblance peut se heurter à des difficultés computationnelles ou à des problèmes de convergence. Pour pallier ces limites, de nombreuses méthodes alternatives ont été proposées dans la littérature. Dans notre cas, on cite la méthode du maximum du produit des espacements (MPS).

Méthode de produit maximum des espacements (MPS)

Vu les propriétés de cette méthode (voir le Chapitre 1), on calcule les estimateurs correspondants pour notre modèle. Les estimateurs MPS sont obtenus en maximisant la fonction suivante $M(\theta)$:

$$M(\theta) = \frac{1}{n+1} \sum_{i=1}^{n+1} \ln(P_i(\theta)) \quad (3.18)$$

Dans le cas de la distribution *Weighted-ND*, cette fonction prend la forme spécifique suivante :

$$P_i(\theta) = \left[1 - \left(\frac{\lambda^\beta(\alpha + \alpha\lambda^\beta)^2}{1 + 3\lambda^\beta} t_i^{2\beta} + (\alpha + \alpha\lambda^\beta)t_i^\beta + 1 \right) e^{-(\alpha + \alpha\lambda^\beta)t_i^\beta} \right] - \left[1 - \left(\frac{\lambda^\beta(\alpha + \alpha\lambda^\beta)^2}{1 + 3\lambda^\beta} t_{i-1}^{2\beta} + (\alpha + \alpha\lambda^\beta)t_{i-1}^\beta + 1 \right) e^{-(\alpha + \alpha\lambda^\beta)t_{i-1}^\beta} \right] \quad (3.19)$$

Les valeurs $\hat{\alpha}_{\text{MPS}}$, $\hat{\beta}_{\text{MPS}}$ et $\hat{\lambda}_{\text{MPS}}$ sont les solutions du système d'équations obtenues en annulant les dérivées de la statistique $M(\theta)$ par rapport aux paramètres α , β et λ .

3.4 Validation du modèle

Dans cette partie, nous construisons deux nouvelles statistiques de test, notées U^2 et Q^2 , afin de valider le modèle étudié.

$$U^2(\theta) = Z_n^T(\theta) \left[I - B(\theta) (B^T(\theta)B(\theta))^{-1} B^T(\theta) \right] Z_n(\theta)$$

$$Q_n^2 = Z_*^T(\tilde{\theta}) \left(I - B \left(I_n^{-1}(\hat{\theta}_n) \right) B^T \right) Z_*(\tilde{\theta})$$

Un avantage clé de ce test d'adéquation U^2 est qu'il permet de substituer les paramètres inconnus par n'importe quel estimateur \sqrt{n} -convergent $\tilde{\theta}$, sans exiger de conditions spécifiques sur le regroupement des données en intervalles. Cette propriété est également valable pour Q_n^2 . Les matrices $I_n(\hat{\theta}_n)$ et $J(\hat{\theta}_n)$ sont les matrices d'information estimées respectivement pour les données initiales et les données regroupées en r classes I_j choisies.

$$I_j = [a_j, a_{j+1}[; \quad \text{pour } j = 1, \dots, r-1$$

Étant donné que les classes regroupées ne sont pas nécessairement équiprobables, les limites a_j sont déterminées à l'aide des méthodes classiques de la statistique descriptive. Dans notre cas, la fonction quantile (ou l'inverse de la fonction de répartition) n'étant pas disponible ou difficile à inverser analytiquement, nous avons recours à ces méthodes classiques pour calculer les bornes des classes, telles que l'utilisation des percentiles empiriques ou des intervalles fondés sur la distribution observée des données.

Si v_1, v_2, \dots, v_r sont les fréquences des données regroupées dans ces classes, on a :

$$v_j = \text{card} \{i \mid T_i \in I_j, i = 1, \dots, n\}$$

avec

$$Z_n(\hat{\theta}) = \left(\frac{v_1 - np_1(\theta)}{\sqrt{np_1(\theta)}}, \dots, \frac{v_r - np_r(\theta)}{\sqrt{np_r(\theta)}} \right)$$

et

$$Z_* (\tilde{\theta}) = Z (\tilde{\theta}) - BI_n^{-1}W(\tilde{\theta})$$

$$\begin{aligned} p_j &= F(a_j) - F(a_{j-1}) \\ &= \left(1 - \left[\frac{\lambda^\beta(\alpha + \alpha\lambda^\beta)^2}{1 + 3\lambda^\beta} a_j^{2\beta} + (\alpha + \alpha\lambda^\beta)a_j^\beta + 1 \right] e^{-(\alpha + \alpha\lambda^\beta)a_j^\beta} \right) \\ &\quad - \left(1 - \left[\frac{\lambda^\beta(\alpha + \alpha\lambda^\beta)^2}{1 + 3\lambda^\beta} a_{j-1}^{2\beta} + (\alpha + \alpha\lambda^\beta)a_{j-1}^\beta + 1 \right] e^{-(\alpha + \alpha\lambda^\beta)a_{j-1}^\beta} \right) \end{aligned}$$

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

Étant donné que la matrice I_n est déjà fournie, on calcule la matrice $J(\theta)$ de la manière suivante :

$$J(\theta) = B^T(\theta)B(\theta) \quad \text{avec} \quad B(\theta) = \left(\frac{1}{\sqrt{p_j(\theta)}} \frac{dp_j(\theta)}{d\alpha}, \frac{1}{\sqrt{p_j(\theta)}} \frac{dp_j(\theta)}{d\beta}, \frac{1}{\sqrt{p_j(\theta)}} \frac{dp_j(\theta)}{d\lambda} \right)$$

$$\frac{dp_j(\theta)}{d\alpha} = \frac{e^{-\alpha(1+\lambda^\beta)a_j^\beta} \left((3\lambda^{2\beta} + \lambda^{3\beta} + 3\lambda^\beta + 1) a_j^{2\beta} + \alpha a_j^{3\beta} (\lambda^\beta + 3\lambda^{2\beta} + 3\lambda^{3\beta} + \lambda^{4\beta}) \right) \alpha}{1 + 3\lambda^\beta}$$

$$\frac{dp_j(\theta)}{d\beta} = \frac{1}{(1 + 3\lambda^\beta)^2} \left(e^{-\alpha(1+\lambda^\beta)a_j^\beta} \left(\left(\left(\ln(\lambda) + 6 \ln(a_j) \right) \lambda^{2\beta} + \left(5 \ln(\lambda) + 12 \ln(a_j) \right) \lambda^{3\beta} \right. \right. \right.$$

$$\left. \left. + \left(7 \ln(\lambda) + 10 \ln(a_j) \right) \lambda^{4\beta} + \left(3 \ln(\lambda) + 3 \ln(a_j) \right) \lambda^{5\beta} + \lambda^\beta \ln(a_j) \right) \alpha a_j^{3\beta} \right.$$

$$\left. + \left(\left(3 \ln(\lambda) + 12 \ln(a_j) \right) \lambda^{2\beta} + \left(6 \ln(\lambda) + 10 \ln(a_j) \right) \lambda^{3\beta} \right. \right.$$

$$\left. + \left(3 \ln(\lambda) + 3 \ln(a_j) \right) \lambda^{4\beta} + \ln(a_j) \left(6\lambda^\beta + 1 \right) \right) a_j^{2\beta} \alpha^2 \right)$$

$$\begin{aligned} \frac{dp_j(\theta)}{d\lambda} &= \frac{\beta \alpha^2 e^{-\alpha(1+\lambda^\beta)a_j^\beta}}{(1 + 3\lambda^\beta)^2} \left(6a_j^{2\beta} \lambda^{3\beta-1} + 3a_j^{2\beta} \lambda^{4\beta-1} + 5\lambda^{3\beta-1} a_j^{3\beta} \alpha \right. \\ &\quad \left. + 7\lambda^{4\beta-1} a_j^{3\beta} \alpha + 3\lambda^{5\beta-1} a_j^{3\beta} \alpha + 3\lambda^{2\beta-1} a_j^{2\beta} + \lambda^{2\beta-1} a_j^{3\beta} \alpha \right) \end{aligned}$$

Ainsi, tous les éléments des statistiques U^2 et Q^2 sont fournis, on en déduit leurs valeurs respectives. La praticabilité et la faisabilité de ces tests sont démontrées par son application sur différents ensembles de données.

3.5 Simulations

Pour estimer les paramètres de la nouvelle distribution, nous avons utilisé les méthodes du maximum de vraisemblance et du produit maximum des espacements. Nous avons généré $N = 10.000$ échantillons avec différentes tailles ($n = 20, 30, 50, 100$ et $n = 150$) à partir de la distribution weighted-ND.

n	$\hat{\theta}_{MLE}$	$MSE(\hat{\theta}_{MLE})$	$Bias(\hat{\theta}_{MLE})$
$\theta_{MLE} = (0.1, 1.5, 0.5)$			
20	$\hat{\alpha} = (0.0873)$	(0.0032)	(-0.0126)
	$\hat{\beta} = (1.6346)$	(0.0761)	(0.1346)
	$\hat{\lambda} = (0.5735)$	(0.2957)	(0.0735)
30	$\hat{\alpha} = (0.0903)$	(0.0020)	(-0.0096)
	$\hat{\beta} = (1.5809)$	(0.0437)	(0.0809)
	$\hat{\lambda} = (0.6060)$	(0.3181)	(0.1060)
50	$\hat{\alpha} = (0.0933)$	(0.0011)	(-0.0066)
	$\hat{\beta} = (1.5407)$	(0.0293)	(0.0407)
	$\hat{\lambda} = (0.5828)$	(0.2545)	(0.0828)
100	$\hat{\alpha} = (0.0842)$	(0.0005)	(-0.0157)
	$\hat{\beta} = (1.5608)$	(0.0126)	(0.0608)
	$\hat{\lambda} = (0.5949)$	(0.2525)	(0.0949)
150	$\hat{\alpha} = (0.0845)$	(0.0004)	(-0.0154)
	$\hat{\beta} = (1.5490)$	(0.0088)	(0.0490)
	$\hat{\lambda} = (0.6415)$	(0.2416)	(0.1415)
$\theta_{MLE} = (1, 0.61, 0.5)$			
5	$\hat{\alpha} = (1.1460)$	(0.3181)	(0.1460)
	$\hat{\beta} = (0.9283)$	(0.2919)	(0.3183)
	$\hat{\lambda} = (0.2839)$	(0.1202)	(-0.2160)
15	$\hat{\alpha} = (1.1866)$	(0.1242)	(0.1866)
	$\hat{\beta} = (0.6887)$	(0.0268)	(0.0787)
	$\hat{\lambda} = (0.3358)$	(0.2564)	(-0.1641)
30	$\hat{\alpha} = (1.1496)$	(0.0859)	(0.1496)
	$\hat{\beta} = (0.6575)$	(0.0076)	(0.0475)
	$\hat{\lambda} = (0.3824)$	(0.2720)	(-0.1175)
80	$\hat{\alpha} = (1.0700)$	(0.0594)	(0.0700)
	$\hat{\beta} = (0.6396)$	(0.0033)	(0.0296)
	$\hat{\lambda} = (0.5867)$	(0.3467)	(0.0867)
150	$\hat{\alpha} = (1.0507)$	(0.0546)	(0.0507)
	$\hat{\beta} = (0.6371)$	(0.0018)	(0.0271)
	$\hat{\lambda} = (0.5434)$	(0.3519)	(0.0434)

Table 11 : Estimation des paramètres par la méthode MLE

n	$\hat{\theta}_{MPS}$	$MSE(\hat{\theta}_{MPS})$	$Bias(\hat{\theta}_{MPS})$
$\theta_{MPS} = (0.1, 1.5, 0.09)$			
20	$\hat{\alpha} = (0.1809)$	(0.00685)	(0.0809)
	$\hat{\beta} = (1.3694)$	(0.07308)	(-0.1305)
	$\hat{\lambda} = (0.2228)$	(0.15189)	(0.1328)
30	$\hat{\alpha} = (0.1636)$	(0.00471)	(0.0636)
	$\hat{\beta} = (1.4316)$	(0.06634)	(-0.0684)
	$\hat{\lambda} = (0.1322)$	(0.05347)	(0.0422)
50	$\hat{\alpha} = (0.1575)$	(0.00303)	(0.0575)
	$\hat{\beta} = (1.4418)$	(0.02765)	(-0.0581)
	$\hat{\lambda} = (0.1125)$	(0.00452)	(0.0225)
100	$\hat{\alpha} = (0.1571)$	(0.00214)	(0.0571)
	$\hat{\beta} = (1.4564)$	(0.01985)	(-0.0435)
	$\hat{\lambda} = (0.0949)$	(0.00376)	(0.0049)
150	$\hat{\alpha} = (0.1449)$	(0.00135)	(0.0449)
	$\hat{\beta} = (1.4791)$	(0.01249)	(-0.0208)
	$\hat{\lambda} = (0.0944)$	(0.00299)	(0.0044)
$\theta_{MPS} = (0.3, 2.5, 0.2)$			
20	$\hat{\alpha} = (0.3786)$	(0.00693)	(0.0786)
	$\hat{\beta} = (2.3182)$	(0.12715)	(-0.2817)
	$\hat{\lambda} = (0.2961)$	(0.08423)	(0.1328)
30	$\hat{\alpha} = (0.3611)$	(0.00575)	(0.0611)
	$\hat{\beta} = (2.3890)$	(0.07022)	(-0.2109)
	$\hat{\lambda} = (0.2577)$	(0.07776)	(0.0577)
50	$\hat{\alpha} = (0.3284)$	(0.00314)	(0.0284)
	$\hat{\beta} = (2.4940)$	(0.05583)	(-0.1059)
	$\hat{\lambda} = (0.2231)$	(0.06310)	(0.0231)
100	$\hat{\alpha} = (0.3366)$	(0.00312)	(0.0366)
	$\hat{\beta} = (2.4530)$	(0.05116)	(-0.1469)
	$\hat{\lambda} = (0.2443)$	(0.08833)	(0.0443)
150	$\hat{\alpha} = (0.3332)$	(0.00226)	(0.0332)
	$\hat{\beta} = (2.4824)$	(0.03522)	(-0.1175)
	$\hat{\lambda} = (0.1726)$	(0.06014)	(-0.0273)

Table 12 : Estimation des paramètres par la méthode MPS

Description des tableaux :

Comme vous le savez, lorsqu'on dérive une forme implicite des équations dans les méthodes d'estimation, des techniques numériques sont nécessaires pour résoudre ces équations. Obtenir la convergence vers les valeurs initiales pour toutes les méthodes peut s'avérer difficile. Pour cette raison, les tableaux de simulation fournis incluent différentes valeurs de départ pour chaque méthode d'estimation. Cette approche garantit la robustesse du processus d'estimation en permettant à plusieurs méthodes de converger vers des valeurs optimales des paramètres.

Au cours du processus de simulation, nous avons observé que chaque méthode d'estimation rencontrait des difficultés avec certaines valeurs initiales des paramètres. Pour remédier à ce problème, nous proposons une approche alternative qui améliore la convergence des paramètres. Les tableaux ci-dessous démontrent l'efficacité de cette approche, montrant que tous les paramètres initiaux convergent avec succès avec des valeurs d'Erreur Quadratique Moyenne (MSE) relativement faibles, garantissant ainsi la fiabilité des estimations.

Dans l'ensemble, les tableaux offrent une comparaison exhaustive des différentes méthodes d'estimation, incluant leurs comportements de convergence et leurs performances sur des ensembles de données complètes. Cette comparaison souligne la robustesse et l'utilité pratique des méthodes proposées pour l'estimation des paramètres.

3.6 Applications

Dans cette section, nous avons appliqué notre nouvelle distribution à plusieurs ensembles de données, en comparant ses performances à une gamme de distributions de probabilité pondérées classiques, y compris les distributions de valeurs extrêmes généralisées (GEV), Gumbel, Fréchet, Weibull, Gamma, Lognormale et Cauchy. De plus, nous avons inclus une comparaison avec la distribution normale pour évaluer l'adéquation de notre Weighted ND comme alternative pour les ensembles de données symétriques.

Pour l'estimation des paramètres, nous avons utilisé la méthode du maximum de vraisemblance (MLE), car elle a démontré une performance supérieure par rapport à la méthode du maximum du produit des espacements (MPS) dans notre contexte. Pour évaluer et comparer la qualité de l'ajustement des différentes distributions de probabilité, nous avons calculé des mesures statistiques standard telles que le critère d'information d'Akaike (AIC), le critère d'information bayésien (BIC) et la log-vraisemblance (LL).

Pour renforcer l'évaluation de la robustesse de notre nouvelle distribution, nous menons une analyse complémentaire, nous avons mis en œuvre deux tests statistiques du chi-carré modifié. Ces tests ont fourni une confirmation de l'adéquation du modèle proposé pour les ensembles de données analysés.

Vous trouverez ci-dessous une présentation détaillée des résultats, y com-

pris des tableaux et des analyses graphiques pour chaque ensemble de données. Ces résultats mettent en évidence la flexibilité et la précision de notre distribution, en faisant un concurrent solide face aux distributions classiques existantes dans des contextes symétriques et non symétriques.

Les distributions retenues dans cette étude, en raison de leur pertinence théorique et de leur large utilisation en modélisation statistique, sont définies comme suit :

- **Distribution des valeurs extrêmes généralisées (GEV)** : pour $\xi \neq 0$

$$f(x; \mu, \sigma, \xi) = \frac{1}{\sigma} \left[1 + \xi \left(\frac{x - \mu}{\sigma} \right) \right]^{-1 - \frac{1}{\xi}} \exp \left(- \left[1 + \xi \left(\frac{x - \mu}{\sigma} \right) \right]^{-\frac{1}{\xi}} \right),$$

$$\text{pour } 1 + \xi \left(\frac{x - \mu}{\sigma} \right) > 0$$

- **Distribution de Gumbel** (cas particulier de GEV avec $\xi = 0$) :

$$f(x; \mu, \sigma) = \frac{1}{\sigma} \exp \left(- \frac{x - \mu}{\sigma} - \exp \left(- \frac{x - \mu}{\sigma} \right) \right)$$

- **Distribution de Fréchet** :

$$f(x; \alpha, s) = \frac{\alpha}{s} \left(\frac{x}{s} \right)^{-\alpha-1} \exp \left(- \left(\frac{x}{s} \right)^{-\alpha} \right), \quad x > 0$$

- **Distribution de Weibull** :

$$f(x; \lambda, k) = \frac{k}{\lambda} \left(\frac{x}{\lambda} \right)^{k-1} \exp \left(- \left(\frac{x}{\lambda} \right)^k \right), \quad x \geq 0$$

- **Distribution Gamma** :

$$f(x; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, \quad x > 0$$

- **Distribution Lognormale** :

$$f(x; \mu, \sigma) = \frac{1}{x\sigma\sqrt{2\pi}} \exp \left(- \frac{(\ln x - \mu)^2}{2\sigma^2} \right), \quad x > 0$$

- **Distribution de Cauchy** :

$$f(x; x_0, \gamma) = \frac{1}{\pi\gamma \left[1 + \left(\frac{x - x_0}{\gamma} \right)^2 \right]}, \quad x \in \mathbb{R}$$

3.7 Exemple.1

Le premier ensemble de données correspond aux temps de survie de 121 patientes atteintes d'un cancer du sein, recueillis dans un grand hôpital entre 1929 et 1938 (Lee et Wang, 2003). Afin de faciliter leur ajustement à la distribution étudiée, ces données ont été préalablement divisées par 10 et sont présentées comme suit :

15.4	13.9	12.9	12.9	12.7	12.6	12.5	11.7	11.5	11.1	10.9	10.9	10.5
10.3	9.6	9.3	9.0	8.9	8.8	8.3	8.0	7.8	6.9	6.8	6.7	6.7
6.5	6.5	6.2	6.1	6.0	6.0	6.0	5.9	5.8	5.7	5.6	5.5	5.4
5.2	5.1	5.1	5.1	4.9	4.8	4.7	4.6	4.6	4.5	4.5	4.4	4.3
4.3	4.3	4.2	4.1	4.1	4.1	4.0	4.0	4.0	3.9	3.9	3.8	3.8
3.8	3.7	3.7	3.7	3.5	3.5	3.2	3.1	3.1	3.0	2.91	2.82	2.79
2.4	2.4	2.36	2.34	2.3	2.11	2.1	2.1	2.09	2.04	1.98	1.79	1.75
1.73	1.72	1.68	1.65	1.63	1.62	1.57	1.55	1.48	1.44	1.44	1.35	1.23
1.22	1.18	1.10	1.03	0.84	0.84	0.75	0.74	0.68	0.66	0.63	0.62	0.56
0.50	0.40	0.03	0.03									

On voudrait vérifier si les données peuvent être décrits par la distribution UGE.

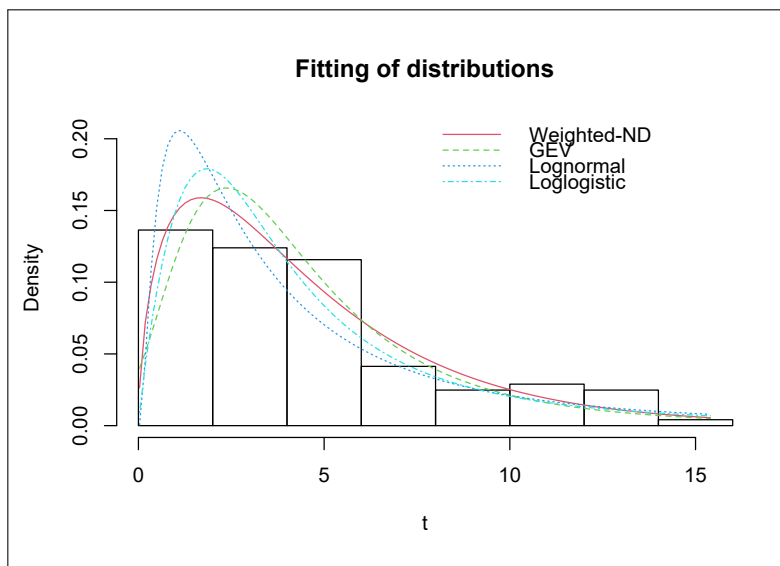


Figure 11 : Ajustement des distributions aux temps de survie de 121 patients atteints de cancer du sein

Distributions	paramètres	MLE	LL	AIC	BIC
WND	α	0.1958			
	β	0.7922	-301.5837	609.1673	617.5547
	λ	4.2285			
GEV	μ	2.7998			
	σ	2.2674			
	ξ	0.2133	-304.90	615.81	624.20
lognormal	μ	1.1632			
	σ	1.0334	-316.4206	636.8412	642.4328
loglogistique	β	1.8559			
	λ	3.5172	-308.9868	621.9736	627.5652

Table 13 : Tests statistiques critiques comparant différentes méthodes d'estimation.

Description du tableau : Le tableau ci-dessus compare les performances de plusieurs distributions de probabilité, notamment les distributions WND, GEV (Valeurs Extrêmes Généralisées), lognormale et loglogistique, en se basant sur leurs estimations par Maximum de Vraisemblance (MLE), les valeurs de log-vraisemblance (LL), le Critère d'Information d'Akaike (AIC) et le Critère d'Information Bayésien (BIC). Cette comparaison met en évidence la flexibilité et la précision supérieures de la distribution WND par rapport aux modèles alternatifs.

Validation du modèle D'autre part, nous supposons l'hypothèse nulle H_0 selon laquelle ces données suivent une distribution WND . Si nous choisissons $r = 5$ classes de regroupement, les calculs intermédiaires des critères de tests Q^2 et U^2 sont les suivants :

$$B(\hat{\theta}_{MLE}) = \begin{bmatrix} 8.4317367 & -0.15300067 & -2.8791879 \\ -2.3740340 & -1.45341523 & 3.1094918 \\ 4.1109634 & 3.12778653 & 1.9403626 \\ 1.0892323 & 0.22685486 & 0.2249913 \\ 0.1476651 & 0.08321904 & 0.0435715 \end{bmatrix}$$

$$L = \left(\sum_{j=1}^r \frac{v_j}{p_j(\theta)} \frac{dp_j(\theta)}{d\alpha} = 388.91, \sum_{j=1}^r \frac{v_j}{p_j(\theta)} \frac{dp_j(\theta)}{d\beta} = 454.89, \sum_{j=1}^r \frac{v_j}{p_j(\theta)} \frac{dp_j(\theta)}{d\lambda} = 27.375 \right)$$

$$J(\hat{\theta}_{MLE}) = B^T(\hat{\theta}_{MLE})B(\hat{\theta}_{MLE}) = \begin{bmatrix} 94.83847 & 15.277998 & -23.430331 \\ 15.27800 & 11.977262 & 2.044841 \\ -23.43033 & 2.044841 & 21.776189 \end{bmatrix}$$

$$I_n(\hat{\theta}_{MLE}) = \begin{bmatrix} 2.03310^{-2} & 19.52 & 5.73610^{-4} \\ 19.52 & 1.87410^{+4} & 0.5506 \\ 5.73610^{-4} & 0.5506 & 1.61710^{-5} \end{bmatrix}$$

$$V = (0.2956125, 0.5798558, -1.4664870, 0.1136953, 1.6960007)$$

$$W = \frac{1}{\text{sqrt}(121)} \left(\frac{\partial l}{\partial \alpha} = 1.5687, \frac{\partial l}{\partial \beta} = 1.505810^3, \frac{\partial l}{\partial \lambda} = 4.424310^{-2} \right)$$

Nous obtenons $Q^2(\theta_{MLE}) = 4.767 < X_{r-1}^2 = 9.487$ et $U^2(\theta_{MLE}) = 3.044 < X_{r-p-1}^2 = 3.841$, ce qui confirme l'hypothèse H_0 .

3.8 Exemple.2

L'application suivante concerne des données de fiabilité, pour tester l'hypothèse selon laquelle ce jeu de données est distribué selon la distribution WND , on procède comme suit. Les valeurs des mesures de résistance prises sur 69 fibres de carbone simples en (GPa) sont les suivantes :

0.312	0.314	0.479	0.552	0.700	0.803	0.861	0.865	0.944	0.958	0.966	0.977
1.006	1.021	1.027	1.055	1.063	1.098	1.140	1.179	1.224	1.240	1.253	1.270
1.272	1.274	1.301	1.301	1.359	1.382	1.382	1.426	1.434	1.435	1.478	1.490
1.511	1.514	1.535	1.554	1.566	1.570	1.586	1.629	1.633	1.642	1.648	1.684
1.697	1.726	1.770	1.773	1.800	1.809	1.818	1.821	1.848	1.880	1.954	2.012
2.067	2.084	2.090	2.096	2.128	2.233	2.433	2.585	2.585			

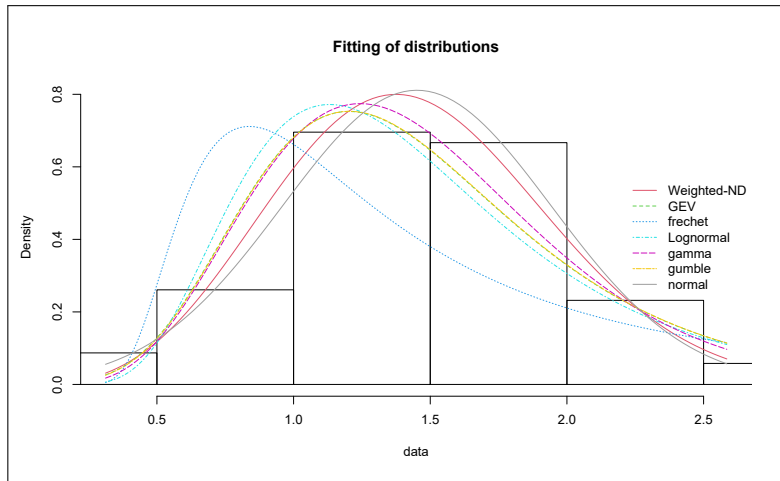


Figure 12 : Ajustement des distributions aux valeurs des mesures de résistance prises sur 69 fibres de carbone simples

Distributions	paramètres	MLE	LL	AIC	BIC
WND	α	0.2593			
	β	1.9710	-49.84	105.69	112.39
	λ	1.8454			
$Gumble$	α	1.2040			
	λ	0.4883	-54.45	112.91	117.38
GEV	μ	1.2039			
	σ	0.4884			
	ξ	1.009×10^{-6}	-54.4563	114.9128	121.6151
$Normal$	μ	1.4510			
	σ	0.4918	-48.94	101.88	106.35
$Frechet$	α	1.0671			
	β	1.8122	-77.8759	159.75	164.22
$Lognormal$	$\mu \log$	0.2990			
	$\sigma \log$	0.4182	-58.39	120.79	125.26
$Gamma$	β	6.9878			
	λ	4.8158	-53.11	110.22	114.68

Description du tableau : Ce tableau présente une analyse comparative de plusieurs distributions citées plus haut. En résumé, la distribution WND offre le meilleur ajustement aux données selon les critères LL , AIC et BIC , ce qui en fait un choix préféré par rapport aux modèles concurrents tels que Gumbel, GEV et autres.

Validation du modèle D'autre part, nous supposons l'hypothèse nulle H_0 selon laquelle ces observations sont décrites par la distribution WND . Si nous choisissons $r = 6$ classes, les calculs intermédiaires des tests de critères Q^2 et U^2 sont les suivants :

$$B(\hat{\theta}_{MLE}) = \begin{bmatrix} 1.33903969 & 3.2309978 & 2.1256717 \\ -1.25417292 & -2.7518162 & -1.8701721 \\ 0.07308304 & 0.6133125 & 1.3322269 \\ 1.39220435 & 0.8325496 & 0.3091155 \\ 0.29218970 & 0.9628729 & 1.4116193 \\ 1.15623962 & 0.5825048 & 0.1901120 \end{bmatrix}$$

$$L = \left(\sum_{j=1}^r \frac{v_j}{p_j(\theta)} \frac{dp_j(\theta)}{d\alpha} = 176.69, \sum_{j=1}^r \frac{v_j}{p_j(\theta)} \frac{dp_j(\theta)}{d\beta} = 524.86, \sum_{j=1}^r \frac{v_j}{p_j(\theta)} \frac{dp_j(\theta)}{d\lambda} = 516.29 \right)$$

$$J(\hat{\theta}_{MLE}) = B^T(\hat{\theta}_{MLE})B(\hat{\theta}_{MLE}) = \begin{bmatrix} 6.7318 & 9.9364 & 6.3518 \\ 9.93646 & 20.347 & 14.558 \\ 6.3518 & 14.558 & 11.915 \end{bmatrix}$$

$$I_n(\hat{\theta}_{MLE}) = \begin{bmatrix} 0.00081 & 1.079278 & 1.699 \times 10^{-4} \\ 1.07927 & 1427.739 & 0.2248329 \\ 0.00016 & 0.224832 & 3.540 \times 10^{-5} \end{bmatrix}$$

$$V = (0.117723, 0.009154, -0.682021, 1.026445, -0.241709, 0.247015)$$

$$W = \frac{1}{\text{sqr}(69)} \left(\frac{\partial l}{\partial \alpha} = 0.23726, \frac{\partial l}{\partial \beta} = 313.86, \frac{\partial l}{\partial \lambda} = 0.04942 \right)$$

Nous obtenons $Q^2(\theta_{MLE}) = 0.3065 < X_{r-1}^2 = 11.0705$ et

$U^2(\theta_{MLE}) = 0.2955 < X_{r-p-1}^2 = 5.9914$, ceci nous permet d'affirmer que ces données de fiabilités peuvent être modélisées par la distribution WND.

De plus, cette distribution se distingue particulièrement par sa performance supérieure par rapport aux distributions à queue lourde (telles que Fréchet, Lognormale ou GEV) lorsque les données sont presque symétriques et très proches d'une distribution normale. Cette capacité d'offrir un excellent compromis entre flexibilité et précision dans ce type de configuration confère un avantage considérable à ce nouveau modèle.

La distribution exponentielle généralisée

Introduction

Ce chapitre concerne l'étude d'un modèle de probabilité très intéressant, c'est un modèle introduit par Nadarajah et Haghghi (2011) et nommé EE. Avec deux paramètres, ce modèle offre une bonne alternative aux distributions gamma, Weibull et exponentielle exponentielle pour lesquelles les fonctions de risque sont uniquement constantes ou décroissantes lorsque la fonction de densité de probabilité (pdf) est monotone décroissante, alors que le taux de défaillance de la nouvelle distribution peut être croissant. De plus, cette distribution est capable de décrire des données avec un mode fixé à zéro lorsque ses taux de risque peuvent être croissants, décroissants et constants, ce qui n'est pas le cas pour les autres modèles. En effectuant un simple changement de variable, le nouveau modèle se réduit à une distribution de Weibull tronquée. La distribution proposée a plus d'applications que les modèles classiques, comme l'ajustement des données de précipitations journalières (par exemple), qui ont une fonction de densité de probabilité (pdf) monotone décroissante lorsque le taux de défaillance empirique (hrf) est croissant. Les auteurs ont étudié des propriétés mathématiques telles que les moments, les courbes de Bonferroni et Lorenz, l'entropie et les statistiques d'ordre. Pour estimer les paramètres inconnus, ils ont proposé les méthodes d'estimation par maximum de vraisemblance (*MLE*) et par moments (*MME*).

Ces méthodes ne peuvent pas donner des formes explicites pour les estimateurs, elles nécessitent donc des méthodes itératives, et comme il est bien connu, les résultats sont performants pour les grands échantillons. Nous proposons alors différentes méthodes classiques pour déterminer ceux-ci. Il convient également de noter que la validation de ce modèle n'a pas encore été étudiée jusqu'à présent, ce qui nous motive à proposer une nouvelle technique pour tester l'ajustement de cette distribution.

Dans ce travail, nous proposons, premièrement, différentes méthodes d'estimation utilisables pour les petits et grands échantillons telles que le maximum du produit des espacements (*MPS*), la méthode de

Cramer-von-Mises (*CM*), la méthode d'Anderson-Darling (*AD*), les méthodes d'Anderson-Darling pour les queues droite et gauche (*RAD*, *LAD*) et la méthode de Kolmogorov-Smirnov (*KS*) pour estimer les paramètres inconnus comme alternatives à la méthode d'estimation par maximum de vraisemblance. D'autre part, pour obtenir des résultats fiables dans toute analyse statistique, il est nécessaire de vérifier si les données observées sont réellement décrites par le modèle choisi. Pour cela, nous adaptons la fameuse statistique Y^2 au modèle en question. Vue son efficacité, cette approche a été récemment développée pour certaines nouvelles distributions. Récemment, cette méthode a été appliquée à certaines nouvelles distributions, voir Seddik-Ameur et Chouia (2014), Treidi et Seddik-Ameur (2016), Tilbi et Seddik-Ameur (2017). Une étude de simulation importante est réalisée pour calculer les différents estimateurs, leurs biais, leurs erreurs quadratiques moyennes et leurs largeurs moyennes estimées. Une étude de comparaison entre les performances des méthodes est également fournie. Outre les critères classiques de sélection de modèles tels que AIC, BIC, LL et la statistique de Kolmogorov-Smirnov, nous avons fourni tous les éléments du nouveau test statistique utilisés pour distinguer entre cette distribution et ses concurrents. Des applications sur des données réelles sont utilisées à titre d'illustration.

4.1 La distribution extension exponentielle

En fonction de deux paramètres α et λ (forme et échelle), l'extension de la distribution exponentielle introduite par Nadarajah et Haghghi (2011) notée *EE* (Extension Exponentielle) est caractérisée par la fonction de répartition cumulative :

$$F(t) = 1 - \exp(1 - (1 + \lambda t)^\alpha) \quad (4.1)$$

La fonction de densité de probabilité (pdf) est donnée par :

$$f(t) = \alpha\lambda(1 + \lambda t)^{\alpha-1} \exp(1 - (1 + \lambda t)^\alpha) \quad \alpha > 0 \text{ et } \lambda > 0. \quad (4.2)$$

Les fonctions de risque instantané et de risque cumulé sont :

$$h(t) = \alpha\lambda(1 + \lambda t)^{\alpha-1} \quad (4.3)$$

$$H(t) = -\log(S(t)) = (1 + \lambda t)^\alpha - 1 \quad (4.4)$$

Les différentes formes de la pdf et des fonctions de taux de risque sont représentées dans les Figures 1 et 2.

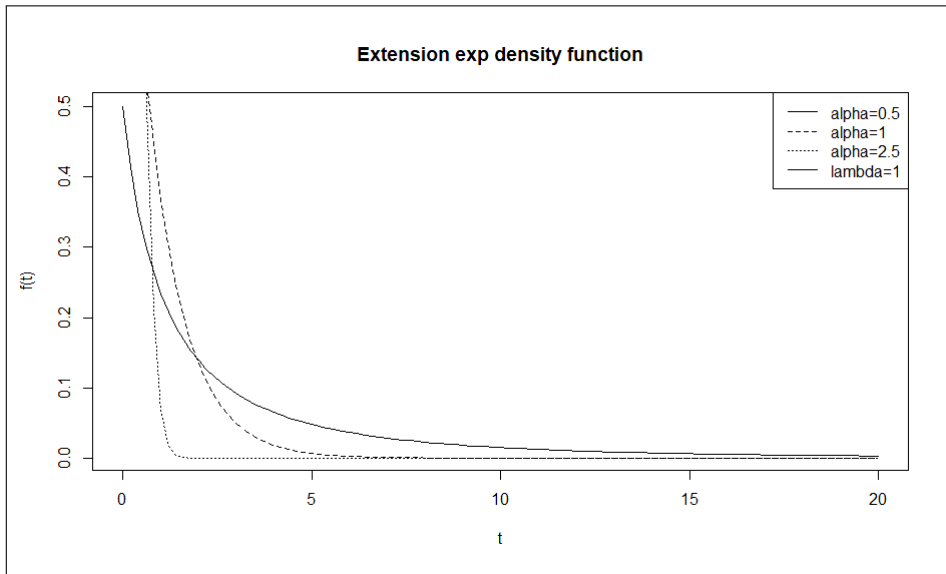


Figure 13 : Fonction de densité de probabilité de la distribution extension exponentielle

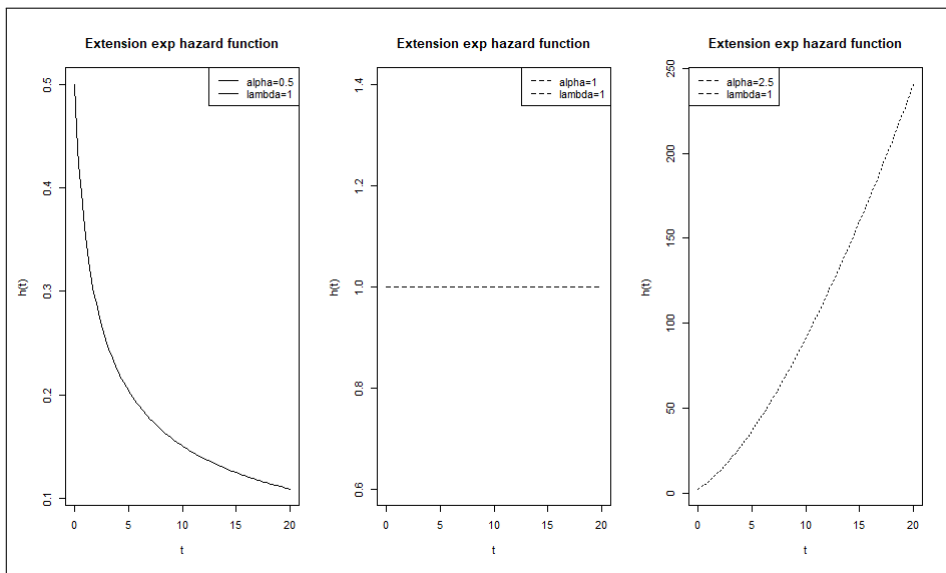


Figure 14 : Fonction de risque de la distribution extension exponentielle

Cette distribution prend uniquement des formes décroissantes alors que la fonction de risque peut être croissante (voir Figure 3).

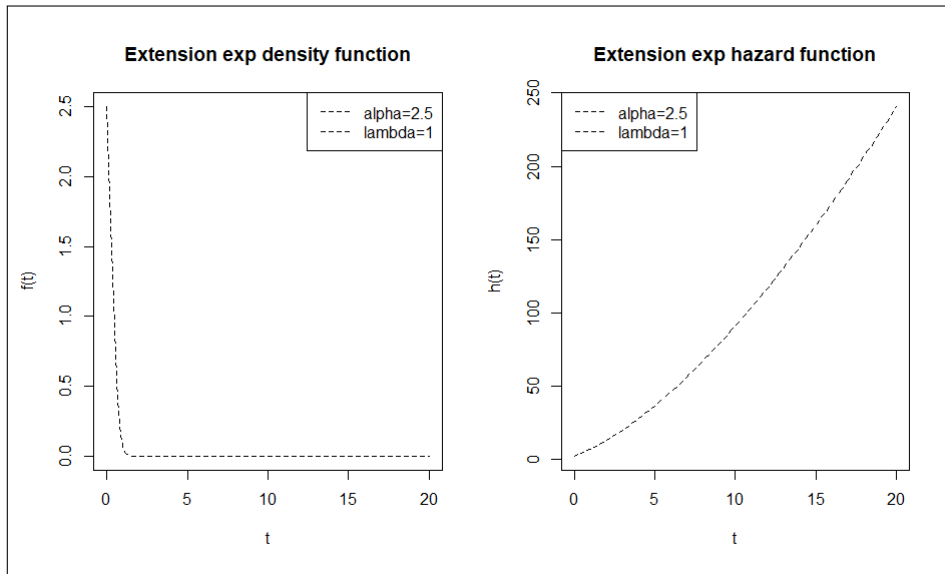


Figure 15 : Fonction de densité et fonction de risque de la distribution extension exponentielle avec $\alpha > 1$

Les auteurs ont étudié des propriétés mathématiques telles que les moments, les courbes de Bonferroni et de Lorenz, qui ont des applications dans plusieurs domaines, l'entropie et les statistiques d'ordre. Pour estimer les paramètres inconnus, ils ont proposé les méthodes d'estimation par maximum de vraisemblance et par la méthode des moments.

4.2 Estimation des paramètres

Considérons un échantillon de n variables indépendantes identiquement t_1, t_2, \dots, t_n provenant de la distribution extension exponentielle avec le vecteur de paramètres $\theta = (\alpha, \lambda)$.

4.2.1 Méthode de maximum de vraisemblance et Information de Fisher

La fonction de vraisemblance est donnée par :

$$L(t_1, t_2, \dots, t_n; \theta) = \prod_{i=1}^n f(t_i; \theta) = \prod_{i=1}^n \alpha \lambda (1 + \lambda t_i)^{\alpha-1} \exp(1 - (1 + \lambda t_i)^\alpha) \quad (4.5)$$

La fonction de log-vraisemblance est :

$$l = n + n \ln(\alpha) + n \ln(\lambda) + (\alpha - 1) \sum_{i=1}^n \ln(1 + \lambda t_i) - \sum_{i=1}^n (1 + \lambda t_i)^\alpha]$$

Les estimateurs du maximum de vraisemblance (*MLE*) sont obtenus en annulant les dérivées premières suivantes:

$$\frac{\partial l}{\partial \alpha} = \frac{n}{\alpha} + \sum_{i=1}^n \ln(1 + \lambda t_i) - \sum_{i=1}^n (1 + \lambda t_i)^\alpha \ln(1 + \lambda t_i) = 0 \quad (4.6)$$

$$\frac{\partial l}{\partial \lambda} = \frac{n}{\lambda} + (\alpha - 1) \sum_{i=1}^n t_i (1 + \lambda t_i)^{-1} - \alpha \sum_{i=1}^n t_i (1 + \lambda t_i)^{\alpha-1} = 0 \quad (4.7)$$

Comme les formes explicites des *MLE* ($\widehat{\alpha}_{MLE}, \widehat{\lambda}_{MLE}$) ne peuvent pas être obtenues, on utilise alors des méthodes numériques.

Les composantes de la matrice d'information de Fisher sont obtenus comme suit:

$$I_{11} = \left(\frac{-1}{n} \right) \sum_{i=1}^n - \left(\frac{1}{\alpha^2} + (1 + \lambda t_i)^\alpha (\ln(1 + \lambda t_i))^2 \right) \quad (4.8)$$

$$I_{22} = \left(\frac{-1}{n} \right) \sum_{i=1}^n - \left(\frac{1}{\lambda^2} + \frac{(\alpha - 1)t_i^2}{(1 + \lambda t_i)^2} + (1 + \lambda t_i)^{\alpha-2} (\alpha - 1) \alpha t_i^2 \right) \quad (4.9)$$

$$I_{12} = \left(\frac{-1}{n} \right) \sum_{i=1}^n \frac{t_i}{1 + \lambda t_i} + \frac{t(1 + \lambda t_i)^\alpha}{(1 + \lambda t_i)} - \alpha t_i \ln(1 + \lambda t_i) (1 + \lambda t_i)^{\alpha-1} \quad (4.10)$$

$$I_{21} = I_{12} = \left(\frac{-1}{n} \right) \sum_{i=1}^n \frac{t_i}{1 + \lambda t_i} + \frac{t(1 + \lambda t_i)^\alpha}{(1 + \lambda t_i)} - \alpha t_i \ln(1 + \lambda t_i) (1 + \lambda t_i)^{\alpha-1} \quad (4.11)$$

4.2.2 Methode du produit maximum des espacements (MPS)

Les estimateurs *MPS* sont obtenus en maximisant la quantité $M(\theta)$:

$$M(\theta) = \frac{1}{n+1} \sum_{i=1}^{n+1} \ln(P_i(\theta)) \quad (4.12)$$

Pour la distribution extension exponentielle, cette fonction s'écrit :

$$M(\theta) = \frac{1}{n+1} \sum_{i=1}^{n+1} \ln((1 - \exp(1 - (1 + \lambda t_i)^\alpha)) - (1 - \exp(1 - (1 + \lambda t_{i-1})^\alpha))) \quad (4.13)$$

$$M(\theta) = \frac{1}{n+1} \sum_{i=1}^{n+1} \ln(\exp(1 - (1 + \lambda t_{i-1})^\alpha) - \exp(1 - (1 + \lambda t_i)^\alpha)) \quad (4.14)$$

Les estimateurs *MPS* sont donc les solutions des équations suivantes :

$$\frac{\partial M(\theta)}{\partial \alpha} = \frac{1}{n+1} \sum_{i=1}^{n+1} \frac{D1(t_i, \theta) - D1(t_{i-1}, \theta)}{P_i} = 0 \quad (4.15)$$

$$\frac{\partial M(\theta)}{\partial \lambda} = \frac{1}{n+1} \sum_{i=1}^{n+1} \frac{D2(t_i, \theta) - D2(t_{i-1}, \theta)}{P_i} = 0 \quad (4.16)$$

où *D1* et *D2* sont les dérivées premières de la fonction de répartition par rapport aux paramètres inconnus α et λ :

$$D1 = \frac{dF(t)}{d\alpha} = (1 + \lambda t)^\alpha \ln(1 + \lambda t) \exp(1 - (1 + \lambda t)^\alpha) \quad (4.17)$$

$$D2 = \frac{dF(t)}{d\lambda} = \alpha t (1 + \lambda t)^{\alpha-1} \exp(1 - (1 + \lambda t)^\alpha) \quad (4.18)$$

4.2.3 Méthode d'estimation du minimum

Estimation de Cramer-Von Mises

Pour ce modèle, cette statistique est définie par :

$$C(\alpha, \lambda) = \frac{1}{12} + \sum_{i=1}^n \left((1 - \exp(1 - (1 + \lambda t_i)^\alpha)) - \frac{2i-1}{2n} \right)^2 \quad (4.19)$$

les estimateurs *CM* ($\hat{\alpha}_{CV}, \hat{\lambda}_{CV}$) sont alors solutions des équations suivantes:

$$\sum_{i=1}^n \left((1 - \exp(1 - (1 + \lambda t_i)^\alpha)) - \frac{2i-1}{2n} \right) D1 = 0 \quad (4.20)$$

$$\sum_{i=1}^n \left((1 - \exp(1 - (1 + \lambda t_i)^\alpha)) - \frac{2i-1}{2n} \right) D2 = 0 \quad (4.21)$$

où *D1* et *D2* sont définies comme précédemment.

Estimation d'Anderson-Darling

les estimateurs ($\hat{\alpha}_{AD}, \hat{\lambda}_{AD}$) sont obtenus en minimisant la statistique A^2 :

$$A^2(\alpha, \lambda) = -n - \frac{1}{n} \sum_{i=1}^n (2i-1) [\ln(1 - \exp(1 - (1 + \lambda t_i)^\alpha)) - \ln(1 - (1 - \exp(1 - (1 + \lambda t_{i*})^\alpha)))] \quad (4.22)$$

où $i^* = n - i + 1$.

Pour notre modèle, les estimateurs $(\widehat{\alpha}_{AD}, \widehat{\lambda}_{AD})$ sont obtenus par la résolution des équations suivantes :

$$\sum_{i=1}^n (2i-1) \left[\frac{D1(t_i)}{(1 - \exp(1 - (1 + \lambda t_i)^\alpha))} - \frac{D1(t_{i^*})}{(1 - (1 - \exp(1 - (1 + \lambda t_{i^*})^\alpha)))} \right] = 0 \quad (4.23)$$

$$\sum_{i=1}^n (2i-1) \left[\frac{D2(t_i)}{(1 - \exp(1 - (1 + \lambda t_i)^\alpha))} - \frac{D2(t_{i^*})}{(1 - (1 - \exp(1 - (1 + \lambda t_{i^*})^\alpha)))} \right] = 0 \quad (4.24)$$

Méthode d'Anderson Darling à droite et à gauche

Les estimateurs Anderson-Darling à droite $(\widehat{\alpha}_{RAD}, \widehat{\lambda}_{RAD})$ sont obtenus en minimisant l'expression suivante :

$$RAD = \frac{n}{2} - 2 \sum_{i=1}^n \ln(F(t_i)) - \frac{1}{n} \sum_{i=1}^n (2i-1)(1 - (F(t_{i^*}))) \quad (4.25)$$

$$RAD = \frac{n}{2} - 2 \sum_{i=1}^n \ln(1 - \exp(1 - (1 + \lambda t_i)^\alpha)) - \frac{1}{n} \sum_{i=1}^n (2i-1)(1 - (1 - \exp(1 - (1 + \lambda t_{i^*})^\alpha))) \quad (4.26)$$

ce qui implique la résolution du système d'équations suivant :

$$-2 \sum_{i=1}^n \frac{D1(t_i)}{(1 - \exp(1 - (1 + \lambda t_i)^\alpha))} + \frac{1}{n} \sum_{i=1}^n (2i-1) \frac{D1(t_{i^*})}{(1 - (1 - \exp(1 - (1 + \lambda t_{i^*})^\alpha)))} = 0 \quad (4.27)$$

$$-2 \sum_{i=1}^n \frac{D2(t_i)}{(1 - \exp(1 - (1 + \lambda t_i)^\alpha))} + \frac{1}{n} \sum_{i=1}^n (2i-1) \frac{D2(t_{i^*})}{(1 - (1 - \exp(1 - (1 + \lambda t_{i^*})^\alpha)))} = 0 \quad (4.28)$$

Les estimateurs Anderson-Darling à gauche $(\widehat{\alpha}_{LAD}, \widehat{\lambda}_{LAD})$ sont obtenus en minimisant l'expression suivante :

$$LAD = \frac{-3n}{2} + 2 \sum_{i=1}^n \ln(F(t_i)) - \frac{1}{n} \sum_{i=1}^n (2i-1)(1 - (F(t_{i^*}))) \quad (4.29)$$

$$LAD = \frac{-3n}{2} + 2 \sum_{i=1}^n \ln(1 - \exp(1 - (1 + \lambda t_i)^\alpha)) - \frac{1}{n} \sum_{i=1}^n (2i-1)(1 - (1 - \exp(1 - (1 + \lambda t_{i^*})^\alpha))) \quad (4.30)$$

4.2.4 Méthode de Kolmogorov-Smirnov

cette méthode repose sur la statistique :

$$KS = d(F_n, F) = \max_{x \in \mathfrak{R}} |F_n(t) - F(t)| \quad (4.31)$$

Si $F(t)$ est la fonction de répartition de la distribution extension exponentielle, nous avons

$$KS = d(F_n, F) = \max_{t \in \mathfrak{R}} |F_n(t) - (1 - \exp(1 - (1 + \lambda t)^a))| \quad (4.32)$$

où n est la taille de l'échantillon, $F(t)$ est la fonction de répartition de la distribution proposé (cdf), et $F_n(t)$ est la fonction de répartition empirique, une fonction en escalier qui augmente de $1/n$ à chaque valeur de donnée. La statistique KS a été largement utilisée pour ajuster des données à des populations continues. L'intérêt de cette statistique inclut le calcul simple de la statistique de test et sa caractéristique de distribution libre.

Les formules de D^+ et D^- pour déterminer le KS sont :

$$D^+ = \max_{i=1,2,\dots,n} \left(\frac{i}{n} - (1 - \exp(1 - (1 + \lambda t_{(i)}^a)) \right) \quad (4.33)$$

$$D^- = \max_{i=1,2,\dots,n} \left((1 - \exp(1 - (1 + \lambda t_{(i)}^a)) - \frac{i-1}{n} \right) \quad (4.34)$$

et

$$KS = \max \{D^+, D^-\} \quad (4.35)$$

où $t_{(1)}, t_{(2)}, \dots, t_{(n)}$ sont les statistiques d'ordre. Les valeurs des paramètres qui maximisent la statistique KS sont notées $(\hat{\alpha}_{KS}, \hat{\lambda}_{KS})$.

4.3 Validation du modèle

Considérons l'hypothèse nulle selon laquelle un échantillon T_1, T_2, \dots, T_n suit la distribution extension exponentielle $P(T_i < t) = F(t, \theta)$ avec le vecteur de paramètres inconnus $\theta = (\alpha, \lambda)^T$. On construit le critère Y^2 pour valider le modèle EE.

$$Y^2(\hat{\theta}) = Z_n^2(\hat{\theta}) + Q_n(\hat{\theta}) \quad (4.36)$$

La formule de calcul de $Q(\hat{\theta})$ est donnée comme suit (voir Van der Vaart, 1998) :

$$Q_n(\hat{\theta}) = \frac{1}{n} L^T(\hat{\theta}_n) \left(I(\hat{\theta}_n) - J(\hat{\theta}_n) \right)^{-1} L(\hat{\theta}_n) \quad (4.37)$$

où $I(\hat{\theta}_n)$ est la matrice d'information estimée de Fisher.

Les bornes des intervalles du regroupement des observations s'obtiennent comme suit:

$$a_j = F^{-1}(j/r) = \frac{(1 - (\log(1 - \frac{j}{r}))^{1/\alpha} - 1)}{\lambda} \quad (4.38)$$

Si v_1, v_2, \dots, v_r sont les fréquences théoriques des données groupées dans ces intervalles

$$Z_n(\hat{\theta}) = \left(\frac{v_1 - np_1(\theta)}{\sqrt{np_1(\theta)}}, \dots, \frac{v_r - np_r(\theta)}{\sqrt{np_r(\theta)}} \right) \quad (4.39)$$

et

$$p_j = (1 - \exp(1 - (1 + \lambda a_j)^\alpha)) - (1 - \exp(1 - (1 + \lambda a_{j-1})^\alpha)) \text{ avec } j = 1, \dots, r \quad (4.40)$$

La matrice d'information de Fisher étant calculée, les éléments de la forme quadratique pour la notre distribution sont obtenus comme suit :

$$L = \left(L_1 = \sum_{j=1}^r \frac{v_j}{p_j(\theta)} \frac{dp_j(\theta)}{d\alpha}, \quad L_2 = \sum_{j=1}^r \frac{v_j}{p_j(\theta)} \frac{dp_j(\theta)}{d\lambda} \right) \quad (4.41)$$

$$J(\theta) = B^T(\theta)B(\theta) \text{ avec } B(\theta) = \left(\frac{1}{\sqrt{p_j(\theta)}} \frac{dp_j(\theta)}{d\alpha}, \frac{1}{\sqrt{p_j(\theta)}} \frac{dp_j(\theta)}{d\lambda} \right) \quad (4.42)$$

Ainsi, on obtient tous les éléments de la statistique de test Y^2 pour la distribution EE sont fournis.

4.4 Simulations numériques

4.4.1 Estimation :

Pour évaluer les performances des différentes méthodes proposées dans ce travail, nous avons généré $N = 10\,000$ échantillons de tailles variées ($n = 15, 25, 40, 80$ et $n = 150$) à partir de la distribution EE . Nous avons calculé les estimateurs des paramètres inconnus, leurs biais, leurs erreurs quadratiques moyennes (MSE) et leurs amplitudes moyennes estimées (AW) par différentes méthodes pour les valeurs des paramètres $\alpha = 0.2$ et $\lambda = 0.1$. Les résultats sont résumés dans les tableaux suivants (Tableaux 1, 2, 3, 4, 5, 6 et 7).

n	$(\hat{\alpha}, \hat{\lambda})$	$MSE(\hat{\theta})$	$Bias(\hat{\theta})$	AW
15	(0.2252, 0.1162)	(0.00368, 0.01256)	(0.02527, 0.01620)	(0.1188, 0.2196)
25	(0.2195, 0.1148)	(0.00265, 0.01129)	(0.01954, 0.01482)	(0.1008, 0.2082)
40	(0.2105, 0.1105)	(0.00115, 0.00540)	(0.01054, 0.01055)	(0.0664, 0.1440)
80	(0.2054, 0.1040)	(0.00049, 0.00212)	(0.00542, 0.00407)	(0.0433, 0.0902)
150	(0.2023, 0.1028)	(0.00023, 0.00101)	(0.00230, 0.00283)	(0.0297, 0.0622)

Tableau 1 : Estimateurs MLE pour les valeurs des paramètres $\alpha = 0.2$,
 $\lambda = 0.1$

n	$(\hat{\alpha}, \hat{\lambda})$	$MSE(\hat{\theta})$	$Bias(\hat{\theta})$	AW
15	(0.1890, 0.2097)	(0.00205, 0.05728)	(0.01093, 0.10977)	(0.0887, 0.4687)
25	(0.1902, 0.1785)	(0.00164, 0.03929)	(0.00976, 0.07854)	(0.0793, 0.3885)
40	(0.1918, 0.1441)	(0.00089, 0.01032)	(0.00811, 0.04416)	(0.0584, 0.1991)
80	(0.1944, 0.1217)	(0.00044, 0.00301)	(0.00555, 0.02177)	(0.0411, 0.1075)
150	(0.1963, 0.1116)	(0.00022, 0.00121)	(0.00360, 0.01165)	(0.0290, 0.0681)

Tableau 2 : Estimateurs MPS pour les valeurs des paramètres $\alpha = 0.2$,
 $\lambda = 0.1$

n	$(\hat{\alpha}, \hat{\lambda})$	$MSE(\hat{\theta})$	$Bias(\hat{\theta})$	AW
15	(0.2425, 0.1425)	(0.02460, 0.05036)	(0.04252, 0.04250)	(0.3074, 0.4398)
25	(0.2193, 0.1270)	(0.00393, 0.01534)	(0.01931, 0.02707)	(0.1228, 0.2427)
40	(0.2128, 0.1142)	(0.00225, 0.00663)	(0.01280, 0.01424)	(0.0929, 0.1595)
80	(0.2054, 0.1074)	(0.00077, 0.00270)	(0.00549, 0.00747)	(0.0543, 0.1018)
150	(0.2032, 0.1033)	(0.00038, 0.00124)	(0.00322, 0.00332)	(0.0382, 0.0690)

Tableau 3 : Estimateurs CvM pour les valeurs des paramètres $\alpha = 0.2$,
 $\lambda = 0.1$

n	$(\hat{\alpha}, \hat{\lambda})$	$MSE(\hat{\theta})$	$Bias(\hat{\theta})$	AW
15	(0.2179, 0.1631)	(0.00545, 0.03055)	(0.01796, 0.06314)	(0.1446, 0.3425)
25	(0.2090, 0.1329)	(0.00228, 0.01512)	(0.00908, 0.03295)	(0.0935, 0.2410)
40	(0.2055, 0.1184)	(0.00122, 0.00643)	(0.00557, 0.01844)	(0.0684, 0.1571)
80	(0.2021, 0.1093)	(0.00054, 0.00235)	(0.00215, 0.00930)	(0.0455, 0.0950)
150	(0.2012, 0.1049)	(0.00028, 0.00117)	(0.00127, 0.00499)	(0.0327, 0.0670)

Tableau 4 : Estimateurs AD pour les valeurs des paramètres $\alpha = 0.2$,
 $\lambda = 0.1$

n	$(\hat{\alpha}, \hat{\lambda})$	$MSE(\hat{\theta})$	$Bias(\hat{\theta})$	AW
15	(0.2237, 0.1801)	(0.00921, 0.07814)	(0.02378, 0.08012)	(0.1880, 0.5478)
25	(0.2115, 0.1270)	(0.00232, 0.01405)	(0.01152, 0.02700)	(0.0944, 0.2323)
40	(0.2070, 0.1197)	(0.00133, 0.00739)	(0.00708, 0.01972)	(0.0714, 0.1684)
80	(0.2032, 0.1092)	(0.00055, 0.00280)	(0.00325, 0.00922)	(0.0459, 0.1037)
150	(0.2018, 0.1040)	(0.00028, 0.00123)	(0.00186, 0.00406)	(0.0327, 0.0687)

Tableau 5 : Estimateurs RAD pour les valeurs des paramètres $\alpha = 0.2$,
 $\lambda = 0.1$

n	$(\hat{\alpha}, \hat{\lambda})$	$MSE(\hat{\theta})$	$Bias(\hat{\theta})$	AW
15	(0.2227, 0.1652)	(0.00519, 0.06445)	(0.02277, 0.06527)	(0.1412, 0.4975)
25	(0.2243, 0.1373)	(0.00610, 0.02161)	(0.02433, 0.03737)	(0.1530, 0.2881)
40	(0.2120, 0.1213)	(0.00271, 0.00849)	(0.01201, 0.02133)	(0.1020, 0.1805)
80	(0.2057, 0.1095)	(0.00097, 0.00301)	(0.00576, 0.00951)	(0.0610, 0.1075)
150	(0.2030, 0.1046)	(0.00047, 0.00135)	(0.00303, 0.00465)	(0.0424, 0.0720)

Tableau 6 : Estimateurs LAD pour les valeurs des paramètres $\alpha = 0.2$, $\lambda = 0.1$

n	$(\hat{\alpha}, \hat{\lambda})$	$MSE(\hat{\theta})$	$Bias(\hat{\theta})$	AW
15	(0.2417, 0.1501)	(0.01127, 0.06134)	(0.04173, 0.05015)	(0.2080, 0.4854)
25	(0.2229, 0.1222)	(0.00441, 0.01555)	(0.02292, 0.02227)	(0.0411, 0.2444)
40	(0.2136, 0.1127)	(0.00204, 0.00677)	(0.01368, 0.01274)	(0.0885, 0.1612)
80	(0.2065, 0.1066)	(0.00084, 0.00282)	(0.00652, 0.00665)	(0.0568, 0.1040)
150	(0.2036, 0.1033)	(0.00040, 0.00130)	(0.00366, 0.00336)	(0.0392, 0.0706)

Tableau 7 : Estimateurs KS pour les valeurs des paramètres $\alpha = 0.2$, $\lambda = 0.1$

Comme prévu, on observe que les biais et les erreurs quadratiques moyennes (MSE) diminuent à mesure que la taille de l'échantillon augmente, et ce, pour toutes les méthodes. On peut voir que pour les petites tailles d'échantillon, les méthodes MPS, AD et RAD ont fourni des résultats très satisfaisants, mais les estimateurs MPS sont les meilleurs par rapport à toutes les autres méthodes. Pour les grandes tailles, toutes les méthodes ont donné des résultats approximativement similaires ; cependant, les estimateurs AD et RAD sont meilleurs que tous les autres, y compris ceux de la méthode du maximum de vraisemblance (MLE).

4.4.2 Critères de test

Pour démontrer la faisabilité du test statistique Y^2 pour la distribution EE. On calcule le nombre de rejets de l'hypothèse nulle H_0 que l'on compare aux valeurs théoriques correspondantes ($\epsilon = 0, 01; 0, 05$ et $0, 1$). Les résultats sont présentés dans le Tableau 8.

n	$\epsilon = 0.05$	$\epsilon = 0.01$	$\epsilon = 0.1$
15	(0.03879)	(0.01697)	(0.1230)
25	(0.04918)	(0.01513)	(0.1080)
40	(0.05638)	(0.01503)	(0.1068)
80	(0.05204)	(0.01222)	(0.1022)
150	(0.05073)	(0.01124)	(0.1004)

Tableau 8. Valeurs critiques théoriques et leurs valeurs empiriques correspondantes pour Y^2

Comme prévu, les valeurs obtenues (Tableau 8) sont très proches de leurs valeurs critiques, ce qui montre que ce critère peut ajuster ce modèle de manière efficace.

4.5 Applications sur des ensembles de données réelles

Dans cette section, nous allons appliquer les résultats obtenus dans ce travail à deux séries de mesures. Dans une première étape, nous calculons les estimateurs obtenus par les différentes méthodes. Ensuite, on trace les graphes des densités et des fonctions de répartitions relatives à la distribution EE et les autres candidates pour tester si les observations sont régies par ce modèle, on utilise aussi les critères classiques et la statistique proposée dans ce travail.

Exemple 1: Le premier concerne les temps de défaillance du système de climatisation d'un avion, étudiés pour la première fois par Linhart et Zuccini (1986), et Gupta et Kundu (2001) qui ont ajusté ces données respectivement aux distributions Weibull et exponentielle exponentielle. Récemment, Dinesh et al. (2017) ont montré que la distribution Weibull modifiée est plus appropriée. Les valeurs observées sont les suivantes : 23, 261, 87, 7, 120, 14, 62, 47, 225, 71, 246, 21, 42, 20, 5, 12, 120, 11, 3, 14, 71, 11, 14, 11, 16, 90, 1, 16, 52, 95 .

La fonction de répartition de la distribution Weibull modifiée considérée ici est :

$$F_{MW} = \frac{2F_W}{1 + F_W} = \frac{2(1 - e^{-(\frac{t}{\lambda})^\alpha})}{2 - e^{-(\frac{t}{\lambda})^\alpha}}$$

On calcule tout d'abord les estimateurs des paramètres des distributions concurrentes citées dans le tableau.

Distributions	paramètres	MLE	CVM	KS	AD	LAD	RAD	MPS
Extension exp	α	0.5985	0.4866	0.6947	0.5233	0.5531	0.5081	0.5057
	λ	4.3385	6.3226	3.8403	5.4853	5.0388	5.8283	5.7787
Weibull	α	0.8535	0.7854	0.9000	0.7994	0.8775	0.7560	0.7805
	λ	54.612	0.5183	0.4496	0.5324	0.4906	0.5265	0.5577
M.Weibull	α	0.9869	0.8989	1.0260	0.9215	0.9909	0.8829	0.9023
	λ	84.097	0.8633	0.7095	0.8664	0.7808	0.8807	0.9050
Exponentiated exp	α	0.8090	0.6979	0.8451	0.7082	0.8528	0.6243	0.6990
	λ	0.0145	1.3880	1.9274	1.3489	1.8123	1.1922	1.2200

Tableau 9. Valeurs des estimateurs des paramètres inconnus des distributions concurrentes pour les données de défaillance du système de climatisation des avions

Comme on peut le voir (tableau 10), pour la méthode d'estimation du maximum de vraisemblance, seules les valeurs de AIC, BIC et LL du modèle

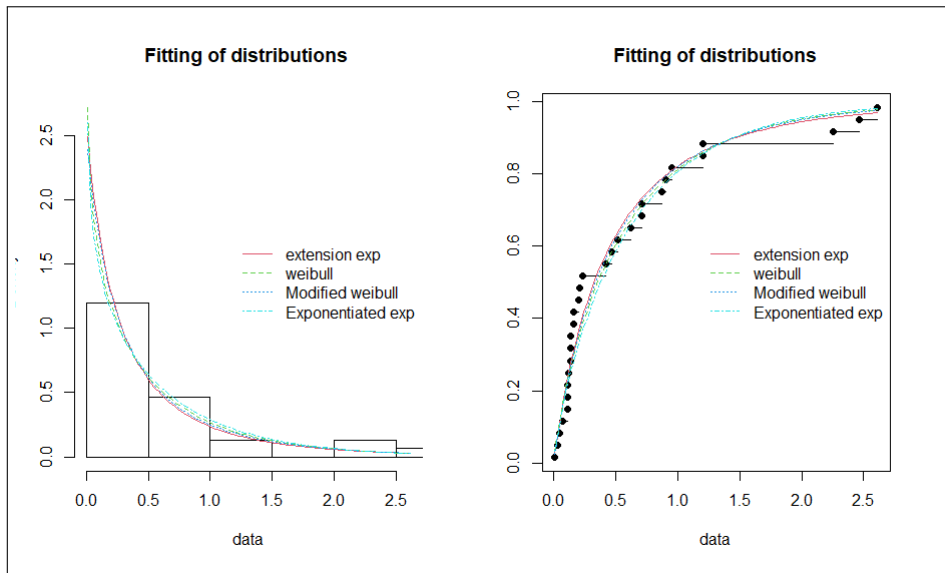


Figure 16 : pdf empirique et cdf EE versus distributions alternatives théoriques

Weibull modifié sont les plus petites. Mais pour les autres méthodes, tous les critères de sélection de modèles AIC, BIC, LL, KS, CvM et AD de la distribution EE donnent les valeurs les plus faibles. Ainsi, on peut dire que la distribution extension exponentielle pourrait mieux modéliser ces données que les autres distributions. Ce fait est aussi confirmé par les graphiques des différentes fonctions de densité de probabilité (voir Figure 4).

	méthodes	exp. étendue	Weibull	Exp. exponentiée	M. Weibull
AIC	MLE	(30.8527)	(31.5635)	(32.0910)	(30.6859)
	CvM	(31.3652)	(31.9047)	(32.5510)	(31.1571)
	KS	(31.6769)	(33.0019)	(33.9210)	(34.2721)
	LAD	(30.9216)	(31.9584)	(33.0875)	(30.8079)
	RAD	(31.1693)	(32.2839)	(33.2811)	(31.3592)
	AD	(31.0697)	(31.7777)	(32.4168)	(30.9512)
	MPS	(31.2078)	(32.0362)	(32.595)	(31.186)
BIC	MLE	(33.6551)	(34.3659)	(34.8934)	(33.4883)
	CvM	(34.1676)	(34.7071)	(35.3534)	(33.9595)
	KS	(34.4793)	(35.8043)	(36.7234)	(34.2721)
	LAD	(33.7240)	(34.7608)	(35.8899)	(33.6103)
	RAD	(33.9717)	(35.0863)	(36.0835)	(34.1616)
	AD	(33.8721)	(34.5801)	(35.2191)	(33.7536)
	MPS	(34.0101)	(34.8385)	(35.3973)	(33.9883)
LL	MLE	(-13.4263)	(-13.7817)	(-14.0455)	(-13.3429)
	CvM	(-13.6826)	(-13.9523)	(-14.2755)	(-13.5785)
	KS	(-13.8384)	(-14.5009)	(-14.9605)	(-13.7348)
	LAD	(-13.4608)	(-13.9792)	(-14.5437)	(-13.4039)
	RAD	(-13.5846)	(-14.1419)	(-14.6405)	(-13.6796)
	AD	(-13.5348)	(-13.8888)	(-14.2084)	(-13.4756)
	MPS	(-13.6039)	(-14.0181)	(-14.2975)	(-13.5930)
KS	MLE	(0.1318)	(0.1533)	(0.1719)	(0.1427)
	CvM	(0.1206)	(0.1229)	(0.1289)	(0.1202)
	KS	(0.1089)	(0.1120)	(0.1135)	(0.1080)
	LAD	(0.1215)	(0.1312)	(0.1339)	(0.1237)
	RAD	(0.1157)	(0.1303)	(0.1367)	(0.1234)
	AD	(0.1203)	(0.1331)	(0.1411)	(0.1267)
	MPS	(0.1198)	(0.1393)	(0.1595)	(0.5943)

Tableau 10 : Valeurs des statistiques AIC, BIC, LL et KS

D'un autre coté, On utilise la statistique Y^2 construite dans ce travail, nous avons regroupé les données en $r = 8$ classes. Après avoir calculé les limites de ces intervalles, nous obtenons :

$$L = \left(L_1 = -2.5073, L_2 = -0.6273 \right)$$

$$J(\hat{\theta}) = \begin{bmatrix} 2.1928 & 0.1684 \\ 0.1684 & 0.0145 \end{bmatrix}$$

$$I_n(\hat{\theta}) = \begin{bmatrix} 7.5394 & 0.4651 \\ 0.4651 & 0.0330 \end{bmatrix}$$

On déduit alors la valeur de $Y^2 = 9.0753$. Pour un niveau de signification $\epsilon = 0.05$, la valeur critique du chi-carré est $X_{r-1}^2 = 14.07$. Comme $Y^2 = 9.0753 < X_{r-1}^2 = 14.07$, l'hypothèse nulle H_0 ne peut pas être rejetée . Donc, les temps de défaillance du système de climatisation d'un avion sont effectivement modélisés par la distribution EE . D'après ces résultats, nous

pouvons dire que la distribution extension exponentielle s'adapte mieux à ces données que toutes ses alternatives.

Exemple 2: Ici, nous considérons un ensemble de données classiques sur les crues de la Rivière Floyd, présenté dans Mudholkar et Huston (1996). Dans l'article d'Akisente et al. (2008), les auteurs ont calculé le critère du rapport de vraisemblance log-vraisemblance de ces données pour différentes distributions, et les valeurs obtenues sont respectivement -376,35 pour la distribution de Weibull exponentiée, -382,13 pour la distribution bêta-Weibull et -365,45 pour le modèle bêta-Pareto. Dans ce deuxième exemple, nous proposons la distribution extension exponentielle. Nous calculons les différents estimateurs des paramètres (Tableau 11) ainsi que les critères de sélection des modèles AIC, BIC, LL et KS (Tableau 12).

Distributions	Paramètres	MLE	CVM	KS	AD	LAD	RAD	MPS
Extension exp	α	0.6093	1.2844	0.8757	0.8301	4.0602	0.6793	0.5295
	λ	0.3358	1.1373	0.2053	0.2355	0.0364	0.3358	0.4625
Weibull	α	0.8714	1.1196	1.0977	1.0011	1.2663	0.8886	0.8088
	λ	6.1888	5.0977	5.6478	5.5515	4.8755	5.4194	6.3671
Gamma	α	0.9195	1.2261	1.1642	1.0716	1.4655	0.8615	0.8074
	n	0.1358	0.2498	0.2131	0.1970	0.3194	0.1523	0.1109
Exponentiated exp	α	0.9688	1.2485	1.1758	1.0996	1.5061	0.8690	0.8357
	λ	0.1445	0.2337	0.2026	0.1963	0.2783	0.1617	0.1213

Tableau 11. Estimations des paramètres inconnus pour les distributions concurrentes pour les données du Floyd River

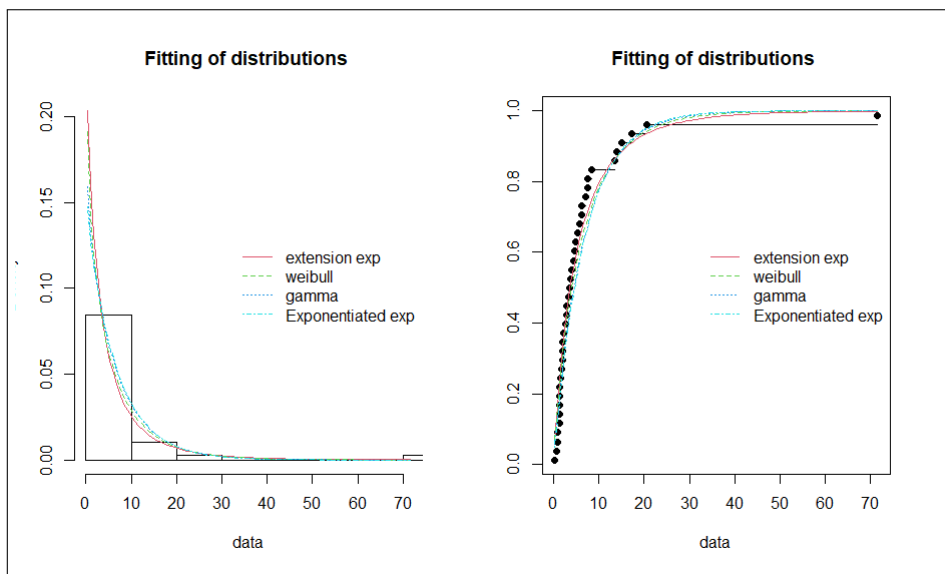


Figure 17 : pdf et cdf empiriques versus distributions alternatives

	methods	Extension exp	Weibull	Gamma	Exponentiated exp
AIC	MLE	(225.9119)	(229.4550)	(231.0047)	(231.167)
	CvM	(246.0783)	(243.3696)	(238.8346)	(237.5375)
	KS	(229.0334)	(237.9247)	(234.7716)	(234.005)
	AD	(228.4938)	(232.8804)	(233.7901)	(233.7134)
	LAD	(571.5404)	(265.8108)	(247.2532)	(243.5087)
	RAD	(226.4296)	(230.1609)	(232.2683)	(232.7168)
	MPS	(226.4262)	(230.0223)	(231.5818)	(231.7741)
BIC	MLE	(229.2391)	(232.7821)	(234.3318)	(234.4941)
	CvM	(249.4054)	(246.6968)	(242.1618)	(240.8647)
	KS	(232.3605)	(241.2518)	(238.0988)	(237.3321)
	AD	(231.8209)	(236.2075)	(237.1172)	(237.0406)
	LAD	(574.8675)	(269.1379)	(250.5803)	(246.8356)
	RAD	(229.7567)	(233.4880)	(235.5954)	(236.0439)
	MPS	(229.7533)	(233.3495)	(234.9089)	(235.1012)
LL	MLE	(-110.956)	(-112.7275)	(-113.5023)	(-113.5835)
	CvM	(-121.0392)	(-119.6848)	(-117.4173)	(-116.7688)
	KS	(-112.5167)	(-116.9624)	(-115.3858)	(-115.0025)
	AD	(-112.2469)	(-114.4402)	(-114.895)	(-114.8567)
	LAD	(-283.7702)	(-130.9054)	(-121.6266)	(-119.7543)
	RAD	(-111.2148)	(-113.0804)	(-114.1341)	(-114.3584)
	MPS	(-111.2131)	(-113.0112)	(-113.7909)	(-113.887)
KS	MLE	(0.1366)	(0.1270)	(0.1471)	(0.1502)
	CvM	(0.1067)	(0.1015)	(0.1003)	(0.0996)
	KS	(0.1034)	(0.0782)	(0.0768)	(0.0764)
	AD	(0.1172)	(0.1059)	(0.0950)	(0.0916)
	LAD	(0.1360)	(0.1264)	(0.1200)	(0.1179)
	RAD	(0.1409)	(0.1425)	(0.1360)	(0.1332)
	MPS	(0.6684)	(0.6715)	(0.7197)	(0.7314)

Tableau 12: Valeurs des critères AIC, BIC, LL et KS pour les données du Floyd River

Comparés aux valeurs LL obtenues pour les différents modèles mentionnés ci-dessus, nous trouvons la valeur LL=(-110.956) pour EE, ce qui prouve que ce modèle est meilleur que toutes les distributions alternatives.

D'autre part, pour utiliser la statistique Y^2 , On choisit $r = 8$ classes pour regrouper ces données, les calculs intermédiaires du critères de test Y^2 sont:

$$L = \left(L_1 = 10.2593, L_2 = 5.3251 \right)$$

$$J(\theta)(\hat{\theta}) = \begin{bmatrix} 4.4775 & 4.0355 \\ 4.0355 & 3.9302 \end{bmatrix}$$

$$I_n(\hat{\theta}) = \begin{bmatrix} 6.9144 & 5.1848 \\ 5.1848 & 5.9629 \end{bmatrix}$$

la valeur obtenue est égale $Y^2 = 6,6724$, donc pour le niveau de signification $\epsilon = 0.05$, la valeur critique $X_{r-1}^2 = 14.07$, nous ne pouvons pas

rejeter H_0 . De plus, les courbes des fonctions de répartition empiriques sont plus proches de la distribution EE que les autres modèles. À partir de cette étude, nous concluons que les utilisateurs peuvent se baser sur les estimateurs MPS et AD respectivement pour des échantillons modérés et grands dans l'analyse. On peut aussi appliquer la statistique Y^2 pour valider le choix de ce modèle.

Conclusion

La modélisation des durées de vie est cruciale pour anticiper les défaillances, optimiser les systèmes complexes et prévenir les risques extrêmes ou bien dans le suivi des patients et les essais thérapeutiques, ou bien pour prévenir la survenue d'un sinistre dans les études actuarielles. Elle exige des outils statistiques à la fois robustes et adaptatifs. Dans cette thèse, on a essayé d'apporter notre contribution en proposant de nouvelles distributions flexibles capables de modéliser des données aussi bien symétriques qu'asymétriques. Nous avons utilisé différentes méthodes pour estimer les paramètres inconnus. De plus, on a construit de nouveaux critères statistiques permettant de valider les modèles proposés sans être contraint de faire appel à de possibles modèles alternatifs comme ce qui se fait dans les méthodes de sélection de modèles classiques.

Perspectives

Comme dans les analyses de fiabilité et de survie, les données sont souvent censurées, nous proposons dans un travail à venir de construire des tests modifiés pour les échantillons censurés à droite pour les modèles proposés dans cette thèse, d'une part et de proposer de nouvelles distributions en utilisant des générateurs de lois.

Bibliographie

- [1] Aidi Khaoula, Abd El-Baset Ahmed, Seddik-Ameur Nacira, A New Two-parameter Distribution-G Family: Properties and Applications, *Electronic Journal of Applied Statistical Analysis*, EJASA (2023), Vol. 16, Issue 3, 2023, 614-638.
- [2] Aidi Khaoula, seddik-Ameur N., Goodness-of-fit test and power comparison for the new two-parameter distribution with unknown parameters and censorship, *Journal of econometrics and statistics JES* (2021), volume 1, issue 1, pp.61-73.
- [3] Aidi, K., Seddik-Ameur, N., Ahmad, A.E.B.A., Khaleel, M. (2022). The Topp-Leone Extended Exponential Distribution: Estimation Methods and Applications. *Pakistan Journal of Statistics Operation Research*, 18(4).
- [4] Alzaatreh, A., Lee, C., Famoye, F. (2013). A new method for generating families of continuous distributions. *Metron*, 71, 63–79.
- [5] Alzaatreh, A., Famoye, F., Lee, C. (2014). The Kumaraswamy-G distribution. *Journal of Statistical Distributions and Applications*, 1, 1–22.
- [6] Anderson, T.W., Darling, D.A. (1952). Asymptotic theory of goodness-of-fit criteria based on stochastic processes. *Annals of Mathematical Statistics*, 23, 193–212.
- [7] Bagdonavičius, V., Clerjoud, L., Nikulin, M. (2006). Power generalized Weibull in accelerated life testing. Preprint 0602, Université Victor Segalen Bordeaux 2.
- [8] Balakrishnan, N., Kundu, D. (2019). New weighted distributions and their applications. *Communications in Statistics – Theory and Methods*.
- [9] Balakrishnan, N., Mitra, D. (2012). Progressive censoring: Theory, methods and applications. *Statistics in Biosciences*.
- [10] Balakrishnan, N., Voinov, V., and Nikulin, M. S. (2013). *Chi-squared goodness of fit tests with applications*. Academic Press.
- [11] Barlow, R., Proschan, F. (1975). *Statistical Theory of Reliability and Life Testing*. Holt, Rinehart and Winston.

- [12] Boos, D.D. (1981). Minimum distance estimators for location and goodness of fit. *Journal of the American Statistical Association*, 76, 663–670.
- [13] Casella, G., Berger, R. (2002). *Statistical Inference*. Duxbury Press.
- [14] Cheng, R., Amin, N. (1983). Estimating parameters in continuous univariate distributions with a shifted origin. *J. Royal Statistical Society B*, 45, 394–403.
- [15] Chouia, S., Seddik-Ameur, N. (2014). A goodness-of-fit test for a competing risk model. *International Journal of Applied Mathematics and Statistics*, 52(2), 34–41.
- [16] Cooray, K., Ananda, M.M.A. (2008). A generalized Weibull distribution. *Communications in Statistics – Theory and Methods*, 37, 1323–1337.
- [17] Cordeiro, G.M., de Castro, M. (2010). A new family of generalized distributions. *Journal of Statistical Computation and Simulation*, 81, 883–898.
- [18] Cordeiro, G.M., Lemonte, A.J. (2011). The Beta Weibull Distribution. *Journal of Statistical Theory and Applications*, 10, 1–23.
- [19] Crow, E.L., Shimizu, K. (1988). *Lognormal Distributions*. Marcel Dekker.
- [20] D’Agostino, R., Stephens, M.A. (1986). *Goodness-of-Fit Techniques*. Marcel Dekker.
- [21] Dzhaparidze, K.O., Nikulin, M.S. (1974). On a modification of the Pearson statistic. *Theory of Probability and its Applications*, 19, 851–853.
- [22] Elbatal, I., Aryal, G.R. (2013). Generalized modified Weibull distribution. *Journal of Probability and Statistics*.
- [23] Eugene, N., Lee, C., Famoye, F. (2002). The Beta-normal distribution. *Statistics Probability Letters*, 54, 177–183.
- [24] Famoye, F., Lee, C., Olumolade, O. (2005). The beta-Weibull distribution. *Journal of Statistical Theory and Practice*, 2, 121–136.
- [25] Fisher, R.A. (1922). On the mathematical foundations of statistics. *Philosophical Transactions of the Royal Society A*, 222, 309–368.
- [26] Fuller, E.J., Frieman, S., Quinn, J. (1994). Fracture mechanics approach to the design of glass aircraft windows. *SPIE Proc.*, 2286, 419–430.
- [27] Ghitany, M.E. (2005). Exponentiated modified Weibull distribution. *Statistical Papers*, 46, 511–523.
- [28] Gómez-Déniz, E., Calderín-Ojeda, E. (2011). The gamma-Lindley distribution. *Journal of Statistical Computation and Simulation*.
- [29] Gupta, R.D., Kundu, D. (1999). Generalized exponential distribution. *Australian New Zealand Journal of Statistics*, 41, 173–188.

- [30] Johnson, N.L., Kotz, S., Balakrishnan, N. (1994). Continuous Univariate Distributions, Vol. 1. Wiley.
- [31] Johnson, N.L., Kotz, S., Balakrishnan, N. (1995). Continuous Univariate Distributions, Vol. 2. Wiley.
- [32] Kumaraswamy, P. (1980). Generalized probability density function for double-bounded variables. *Journal of Hydrology*, 462, 79–88.
- [33] Kundu, D., Raqab, M.Z. (2005). Generalized Rayleigh distribution and its applications. *Communications in Statistics – Theory and Methods*, 34, 1201–1215.
- [34] Leemis, L.M. (2008). *Reliability: Probabilistic Models and Statistical Methods*. Prentice Hall.
- [35] Lemonte, A.J., Cordeiro, G.M. (2013). The exponentiated Kumaraswamy distribution. *Journal of the Korean Statistical Society*, 42, 181–193.
- [36] Marshall, A.W., Olkin, I., Arnold, B.C. (2011). *Inequalities: Theory of Majorization and its Applications*. Springer.
- [37] Marshall, A.W., Olkin, I. (1997). A new method for adding parameters to a family of distributions. *Biometrika*, 84, 641–652.
- [38] Mazucheli, J. (2018). The unit Weibull distribution. *Journal of Applied Probability and Statistics*, 13(2), 1–22.
- [39] Mazucheli, J., Menezes, A.F., Dey, S. (2019). The Unit-Gompertz distribution. *Statistica*, 79(1), 25–43.
- [40] McCulloch, C.E. (1985). Relationships among chi-squared goodness-of-fit statistics. *Communications in Statistics – Theory and Methods*, 14, 593–603.
- [41] Meribout, K.K., Seddik-Ameur, N., Goual, H. (2023). Goodness-of-fit tests for the new Weibull-G family. *Advances in Mathematics: Scientific Journal*.
- [42] Mudholkar, G.S., Srivastava, D.K. (1995). Exponentiated Weibull family. *Technometrics*, 37, 436–445.
- [43] Murthy, D.N.P., Xie, M., Jiang, R. (2004). *Weibull Models*. Wiley.
- [44] Nadarajah, S., Haghghi, F. (2011). An extension of the exponential distribution. *Journal of Theoretical and Applied Statistics*, 45(6).
- [45] Nadarajah, S., Kotz, S. (2011). Exponential Weibull distribution. *Annals of Applied Statistics*, 5, 120–145.
- [46] Nelson, W. (1982). *Applied Life Data Analysis*. Wiley.
- [47] Nikulin, M.S. (1973). Chi-square test for continuous distributions. *Theory of Probability and its Applications*, 18, 638–639.

- [48] Nikulin, M.S. (1973). Chi-square test for normality. International Vilnius Conference on Probability and Statistics, 119–122.
- [49] O'Connor, P., Kleyner, A. (2012). Practical Reliability Engineering. Wiley.
- [50] Rao, K.C., Robson, D.S. (1974). A chi-square statistic for exponential family tests. *Communications in Statistics*, 3, 1139–1153.
- [51] Ramnaby, B. (1984). The Maximum Spacing Method. *Scandinavian Journal of Statistics*, 11, 93–112.
- [52] Ristić, M.M., Balakrishnan, N. (2015). Gamma-exponentiated exponential distribution. *Journal of Statistical Computation and Simulation*, 85, 2334–2349.
- [53] Sarhan, A.M., Kundu, D. (2009). Generalized linear failure rate distribution. *Communications in Statistics – Theory and Methods*, 38, 233–248.
- [54] Sarhan, A.M., Balakrishnan, N. (2020). *Advanced Methods in Reliability and Lifetime Data Analysis*. Springer.
- [55] Seddik-Ameur N., Treidi W., On testing the fit of accelerated failure time and proportional hazard Weibull extension models, *Journal of Statistical Theory and Practice*, Volume 12, Issue 2, pp. 397- 411, (2018).
- [56] Shaw, W.T., Buckley, I.R. (2007). Generalized lambda distributions. *Statistical Science*, 22, 274–304.
- [57] Silva, R.B., Cordeiro, G.M. (2013). The Burr XII power series distributions. *Computational Statistics Data Analysis*, 65, 41–52.
- [58] Singh, A.C. (1987). Generalization of Rao–Robson statistic. *Communications in Statistics – Theory and Methods*, 16, 3255–3273.
- [59] Singh, K.P., Maddala, G.S. (1977). A function for size distribution of incomes. *Econometrica*, 45, 963–970.
- [60] Tahir, M.H., Cordeiro, G.M. (2015). The odd Lindley-G family. *Journal of Applied Statistics*, 42, 1691–1718.
- [61] Tilbi, D., Seddik-Ameur, N. (2017). Chi-square GOF tests for generalized Rayleigh distribution. *Journal of Statistical Theory and Practice*, 11(4), 594–603.
- [62] Weber, M.D., Leemis, L.M., Kincaid, R.K. (2006). Minimum Kolmogorov–Smirnov estimators. *Journal of Statistical Computation and Simulation*, 76, 195–206.
- [63] Xu, Y., et al. (2003). Application of neural networks in forecasting reliability. *Applied Soft Computing*, 2, 255–268.
- [64] Yousof, H.M., Afify, A.Z., Cordeiro, G.M. (2016). The Weibull-G family of distributions. *Journal of Data Science*, 14, 509–536.