



THÈSE

Pour l'obtention du grade de

DOCTEUR DE L'UNIVERSITE DE BORDEAUX1

ÉCOLE DOCTORALE DE MATHÉMATIQUES ET INFORMATIQUE

DOMAINE DE RECHERCHE: STATISTIQUES

Présentée par

Noureddine SAAIDIA

**Sur les Familles des Lois de Fonction de Hasard Unimodale.
Applications en Fiabilité et Analyse de Survie**

Soutenue le 24 Juin 2013

Devant le jury Composé de:

M. Mounir MESBAH	Professeur, Univ. de Paris VI	Rapporteur
M. Mikhaïl NIKULIN	Professeur, Univ. de Bordeaux	Directeur
M. Alaa CHATEAUNEUF	Professeur, Univ. Blaise Pascal, Clermont Ferrant	Rapporteur
Mme. Catherine HUBER-CAROL	Professeur, Univ. René Descartes-Paris 5	Examinatrice
M. Fabrice GUERIN	Professeur, Univ. d'Angers	Examineur
M. Vincent COUALLIER	Maître de conférences, Univ. de Bordeaux	Examineur
M. Léo GERVILLE-REACHE	Maître de conférences, HDR, Univ. de Bordeaux	Examineur

Dedicace

A mes chers parents.

A ma femme et mes enfants.

A mes frères et sœurs.

Remerciement

Je commence tout d'abord par remercier mon directeur de thèse professeur M.S. Nikulin pour sa confiance, son humanité, sa patience sa générosité, son amitié, ses conseils constructifs, le temps et la disponibilité dont il a fait preuve pendant ces années, aussi ses suggestions importantes et les corrections nécessaires. J'apprécie la pleine confiance et la liberté scientifique qu'il m'a accordé pour la réalisation de cette thèse, ainsi que d'avoir cru en mes capacités et en me donnant une véritable liberté d'action et ce dans d'excellentes conditions.

Il l'a dirigée avec une grande efficacité. Qu'il trouve ici mon admiration et mes profonds respects. Je pense avoir reçu son message sur une vision de la recherche et j'espère pouvoir l'honorer dans le futur.

Je remercie chaleureusement et particulièrement M. Mounir MESBAH, Professeur à l'Université Pierre et Marie Curie, Paris VI de m'avoir fait le grand honneur en acceptant de rapporter ma thèse. Qu'il trouve ici toute de ma profonde gratitude

Je remercie infiniment M. Alaa CHATEAUNEUF, Professeur à l'Université Blaise Pascal, Clermont Ferrant d'avoir accepté de rapporter ma thèse. Je lui adresse tous mes profonds respects.

Mes remerciements vont également à M. Fabrice GUERIN, Professeur à l'Université d'Angers pour avoir accepté d'examiner cette thèse et présider le jury.

Je tiens à exprimer ma profonde reconnaissance et toute ma gratitude à Madame Catherine HUBER-CAROL, Professeur à l'université René Descartes, Paris V. Je suis très sensible à l'honneur qu'elle m'a fait en acceptant d'examiner et juger cette thèse. Qu'elle trouve ici l'expression de ma profonde gratitude.

J'adresse mes chaleureux remerciements et ma gratitude à M. Léo Gerville-Réache, Maître de conférences et HDR à l'Université de Bordeaux d'avoir accepté d'examiner ce travail. Qu'il trouve ici, l'expression de ma très haute considération.

Je tiens à remercier chaleureusement et fortement M. Vincent Couallier, Maître de conférences à l'Université de Bordeaux pour l'intérêt qu'il a accordé à ce travail depuis le début, pour son aide et ses conseils, et pour la spontanéité avec laquelle il a accepté d'examiner et juger ce travail.

Mes vifs remerciements vont aussi à Mme. N. Seddik-Ameur, Professeur au département de Mathématiques à l'université Badji Mokhtar-Annaba, Algérie pour ses conseils constructifs et ses suggestions importantes.

Je remercie mes parents, mes frères et sœurs pour leur encouragements.

J'adresse mes sincères remerciements et mes profonds respects à ma femme l'étoile brillante de ma vie pour son encouragement et sa patience, ainsi à mes enfants Oualla Maissoune, Med Zakaria et le petit Raïd.

Mes vifs remerciements vont plus particulièrement à M. Daniel Commenges, équipe de biostatistique, ISPED, Université Victor Segalen, Bordeaux 2.

Je remercie chaleureusement et particulièrement, Pr. Badr- Eddine Ainseba, M. Pierrick Legrand, M. Ramzan Tahir, M. Quang Tran, M. Alejandro Lopez Rincon, Melle Ingrid Rochelle, et toute l'équipe de "Fiabilité et Statistique Mathématique", Bordeaux 2, pour leur aide et leur soutien.

Mes vifs remerciements vont aussi à Yacine Ait Ammar, Redha Souakri, Nabil Bouadjila, B. Trouzine, H. Trouzine, Foued Mansouri et à tous mes amis que je ne peux pas compter...

Table des matières

Table des figures	ix
Liste des tableaux	xi
Introduction générale	1
1 Les familles des lois de fonction de hasard unimodale	5
1 Introduction	5
2 La distribution log-normale	5
3 La distribution log-logistique	7
4 La distribution de Weibull exponentielle	8
5 La distribution de Weibull généralisée	9
6 La distribution de Birnbaum-Saunders	12
7 La distribution Gaussienne inverse	13
7.1 Densité de probabilité	13
7.2 Fonction caractéristique et moments	15
7.3 La fonction de répartition	17
7.4 La fonction de hasard	18
7.5 Analogies avec la loi Normale	19
7.6 Caractéristiques	20
7.7 Propriétés statistiques	20
7.8 Quelques applications	21
7.8.1 Science actuarielle	21
7.8.2 Démographie	22
7.8.3 Hydrologie	22

2	Test d'ajustement du Chi-deux pour les familles de fonction de hasard unimodale	27
1	Introduction	27
2	La théorie du test du Chi-deux.	27
2.1	Test du Chi-deux de Pearson.	27
2.2	Les méthodes d'estimation	29
2.2.1	La méthode du maximum de vraisemblance (MV)	29
2.2.2	La méthode du minimum du Chi-deux	31
2.2.3	La méthode des moments	32
2.3	Tests d'ajustement du Chi-deux modifiés	34
3	Tests d'ajustement pour la loi log-normale	37
4	Tests d'ajustement pour la loi log-logistique	38
5	Tests d'ajustement pour la loi de Weibull exponentielle	40
6	Tests d'ajustement pour la loi de Weibull généralisée	43
7	Tests d'ajustement pour la loi de Birnbaum-Saunders	46
8	Tests d'ajustement pour la loi Gaussienne inverse	49
8.1	Simulation	56
8.2	Autres tests d'ajustement pour la loi Gaussienne inverse	61
9	Test de Chi-deux de la loi Gaussienne inverse contre la loi log-normale	62
10	Test du Chi-deux de la loi Gaussienne inverse contre la loi de Weibull généralisée	65
11	Etude de la puissance des tests : Simulation	66
12	Test d'ajustement du Chi-deux pour les données censurées	69
12.1	Introduction	69
12.2	Test d'ajustement du Chi-deux pour la famille des lois Gaussiennes inverses	70
12.3	Test de Habib et Thomas	72
12.4	Test de Bagdonavičius et Nikulin	74
12.4.1	Application à la loi Gaussienne inverse	79
12.4.2	Réanalyse des données de "Arm A" du cancer de la tête et du cou [Efron (1988)].	82
3	Le modèle de Régression AFT.	85
1	Introduction	85
2	Le modèle AFT	86
2.1	Introduction	86
2.2	Le principe du modèle de Sedyakin	87
2.3	Définition du modèle AFT	88
2.4	Plan d'expériences	89
2.5	Estimation dans les modèles AFT-LN et AFT-WG	91
2.6	Estimation dans le modèle AFT-IG	91

2.7	Test d'ajustement pour le modèle AFT paramétrique	98
2.7.1	Construction du test	98
2.7.2	La distribution asymptotique de Z	99
2.7.3	Le test Statistique	102
2.7.4	Choix de $\hat{\alpha}_j$:	102
2.7.5	Application à la loi Gaussienne inverse	103
4	Analyse des systèmes redondants	107
1	Introduction	107
2	Système redondant $S(1, 1)$	107
3	Système redondant $S(1, m - 1)$	109
4	L'estimation de $K_j(t)$	110
5	L'intervalle de confiance asymptotique de $K_j(t)$	111
6	La distribution Gaussienne inverse	112
6.1	Simulation	119
7	La distribution log-normale	119
7.1	Simulation	123
8	La distribution log-logistique	123
8.1	Simulation	125
9	La distribution de Weibull généralisée	126
9.1	Simulation	131
	Conclusion générale et perspectives	133
	Bibliographie	135
	Liste de Publications et de Communications	145

Table des figures

1.1	Fonctions de densité de la distribution LN	6
1.2	Fonctions de survie de la distribution LN	6
1.3	Fonctions de hasard de la distribution LN	7
1.4	Fonctions de densité de la distribution LL	8
1.5	Fonctions de survie de la distribution LL	8
1.6	Fonctions de hasard de la distribution LL	9
1.7	Fonctions de densité de la distribution EW	10
1.8	Fonctions de survie de la distribution EW	10
1.9	Fonctions de hasard de la distribution EW	10
1.10	Fonctions de densité de la distribution GW	11
1.11	Fonctions de survie de la distribution GW	11
1.12	Fonctions de hasard de la distribution GW	11
1.13	Fonctions de densité de la distribution BS	12
1.14	Fonctions de survie de la distribution BS	13
1.15	Fonctions de hasard de la distribution BS	13
1.16	Fonctions de densité de la distribution IG	14
1.17	Fonctions de survie de la distribution IG	17
1.18	Fonctions de hasard de la distribution IG	19
2.1	Comparaison des fonctions de densité pour IG	55
2.2	Comparaison des fonctions de survie pour IG	55
2.3	Histogrammes de Y_2^n pour $(\mu, \lambda) = (1, 1)$, et $n = 20, 30, 50, 80, 100, 150$	59
2.4	Histogrammes de Y_2^n pour $(\mu, \lambda) = (1, 1)$, et $n = 200, 300, 400, 500$	60
2.5	Comparaison des densités de IG et de LN pour les données d'ERCA.	64
2.6	Comparaison des densités de IG et de GW pour les données d'ERCA.	66

2.7	Puissances estimées des 3 tests $Y_n^2(\hat{\theta}_n)$, $U_n^2(\hat{\theta}_n)$ et $M_n^2(\hat{\theta}_n)$ en fonction de r contre . .	67
2.8	Puissances estimées des 3 tests $Y_n^2(\hat{\theta}_n)$, $U_n^2(\hat{\theta}_n)$ et $M_n^2(\hat{\theta}_n)$ en fonction de r contre . .	68
2.9	Puissances estimées des 3 tests $Y_n^2(\hat{\theta}_n)$, $U_n^2(\hat{\theta}_n)$ et $M_n^2(\hat{\theta}_n)$ en fonction de r contre . .	68
2.10	Puissances estimées des 3 tests $Y_n^2(\hat{\theta}_n)$, $U_n^2(\hat{\theta}_n)$ et $M_n^2(\hat{\theta}_n)$ en fonction de r contre . .	69
2.11	Puissances estimées des 3 tests $Y_n^2(\hat{\theta}_n)$, $U_n^2(\hat{\theta}_n)$ et $M_n^2(\hat{\theta}_n)$ en fonction de r contre . .	69
4.1	Graphes des estimateurs des fonctions de répartition \hat{F}_1 , \hat{K}_2 , \hat{K}_3 et \hat{K}_4 ,	116
4.2	Graphes des estimateurs des fonctions de répartition \hat{F}_1 , \hat{K}_2 , \hat{K}_3 et \hat{K}_4 ,	121
4.3	Graphes des estimateurs de fonctions de répartition \hat{F}_1 , \hat{K}_2 , \hat{K}_3 et \hat{K}_4 ,	130

Liste des tableaux

1.1	Les quatre premiers moments de la distribution IG	15
1.2	Les estimateurs de $MVUE$ de quelques fonctions de paramètres pour la distribution IG	26
2.1	Données d'entretien d'ERCA.	53
2.2	Valeurs des 3 tests pour les Données d'entretien d'ERCA.	56
2.3	Niveau empirique pour $\mu = 0.25$ et $\lambda = (0.5, 1, 5, 10)$	57
2.4	Niveau empirique pour $\mu = 1$ et $\lambda = (0.5, 1, 5, 10)$	57
2.5	Niveau empirique pour $\mu = 4$ et $\lambda = (0.5, 1, 5, 10)$	58
2.6	Niveau empirique pour $\mu = 10$ et $\lambda = (0.5, 1, 5, 10)$	58
2.7	Quelques tests d'ajustement pour les données censurées aléatoires à droite.	71
2.8	Données d'Efron.	82
4.1	Niveau de confiance pour $K_2(t)$,	119
4.2	Niveau de confiance pour $K_2(t)$,	123
4.3	Niveau de confiance pour $K_2(t)$,	125
4.4	Niveau de confiance pour $K_2(t)$,	131

Introduction générale

La présente thèse est consacrée aux études des modèles statistiques appliqués en fiabilité et en analyse de survie, en particulier aux études des familles paramétriques de fonction de hasard unimodale (croissante puis décroissante) dont nous pouvons citer : les familles des distributions : log-normales, log-logistiques, de Birnbaum-Saunders [Birnbaum et Saunders (1969a,1969b)], de Weibull exponentielle [Mudholkar et al. (1995, 1996)], de Weibull généralisée [Bagdonavičius et Nikulin (2002)] et Gaussiennes inverses. Il faut noter que la monographie de Bagdonavičius et Nikulin (2002) a mis l'accent sur les études des modèles semi-paramétriques. En même temps, la pratique montre que souvent, grâce aux études du comportement des familles semi-paramétriques, il est raisonnable de les approximer par les modèles paramétriques. De ce point de vue, il faut dire que la famille des lois Gaussiennes inverses est très importante par la richesse de ses propriétés ; et c'est pour cela que souvent elle apparaît comme un modèle compétitif vis-à-vis des distributions de fonction de hasard unimodale. Estimant que cette famille aux applications très diversifiées mérite d'être davantage approfondie, nous nous proposons de construire le test d'ajustement du Chi-deux pour cette famille, ainsi la construction des modèles avec covariables basés sur cette famille.

La distribution Gaussienne inverse tient son nom du fait qu'elle établit une relation directe avec la loi normale : la fonction génératrice des cumulants de la loi Gaussienne inverse est l'inverse de celle de la loi normale. Pendant des décennies, cette distribution a intéressé des chercheurs de différents domaines, notamment en fiabilité, actuariat, histomorphométrie, réseaux électriques, hydrologie, management, démographie, tests de vie accélérés, météorologie, santé mentale, physiologie, économie, cardiologie, etc.

L'origine de cette distribution remonte au célèbre botaniste Robert Brown (1773-1858). Il s'est intéressé à l'étude du mouvement des particules (qui est maintenant bien connu sous le nom du

mouvement brownien). En 1905, Albert Einstein a utilisé la distribution normale comme modèle pour le mouvement brownien. En 1915, Schrödinger a montré que la distribution du temps du premier passage suit une loi Gaussienne inverse. L'utilisation de la loi Gaussienne inverse comme un modèle de durée de vie est particulièrement intéressante [Chhikara et Folks (1989)]. La fonction de hasard de la loi Gaussienne inverse est unimodale, de même que les distributions log-normale, log-logistique, de Birnbaum-Saunders, de Weibull exponentielle et de Weibull généralisée. Sachant que les distributions présentant une fonction de hasard unimodale sont très peu nombreuses, il s'avère que la distribution Gaussienne inverse est très concurrente à ces distributions et offre certains avantages par rapport à celles-ci, parce que pour les distributions log-normale, log-logistique, de Weibull exponentielle et de Weibull généralisée, la fonction de hasard croît de zéro à son maximum puis diminue jusqu'à zéro. Par contre, pour les distributions Gaussienne inverse et de Birnbaum-Saunders, la fonction de hasard croît de zéro à son maximum puis tend asymptotiquement vers une constante non nulle. Le comportement asymptotique de la fonction de hasard de la distribution Gaussienne inverse implique que l'occurrence de la panne devient purement aléatoire et indépendante du passé (sans mémoire). Par contre les autres distributions escamotent la panne, ce qui n'est pas possible pour la plupart des systèmes.

Le test d'ajustement du Chi-deux a été introduit par Karl Pearson en 1900. Ce test est bien adapté à une loi discrète, mais si la loi est continue, il faut regrouper les données (ou l'échantillon) en classes dont le nombre et l'amplitude posent problème. Pearson a montré que pour une hypothèse simple, la statistique de Pearson suit asymptotiquement une loi du Chi-deux à $r - 1$ degrés de liberté, où r est le nombre de classes choisi. En général les paramètres de la loi sont inconnus et doivent être estimés à partir de l'échantillon ; dans ce cas la statistique de Pearson ne suit plus asymptotiquement une loi du Chi-deux et dépend à la fois des paramètres en question et de la méthode d'estimation.

Fisher (1928) a montré que si on estime les paramètres par la méthode du minimum du Chi-deux pour les données groupées, la statistique de Pearson suit asymptotiquement une loi du Chi-deux à $(r - s - 1)$ degrés de liberté, où s est le nombre de paramètres estimés. Chernoff et Lehmann (1954) ont montré que si on estime les paramètres de la loi par la méthode du maximum de vraisemblance, la limite de la statistique de Pearson change dramatiquement. Dans les années 70, ce résultat a conduit plusieurs auteurs Nikulin (1973a,1973b), Dzhaparidze et Nikulin (1974), Moore et Spruill (1975), Rao et Robson (1974) à s'intéresser à des modifications de la statistique de Pearson qui suivent asymptotiquement une loi du Chi-deux après estimation des paramètres à partir des données non regroupées. Plus de détails sur la construction des tests de type du Chi-deux pour les données complètes, voir la monographie de Greenwood et Nikulin (1996).

Le test d'ajustement du type du Chi-deux a été étendu même aux données censurées aléatoires. Plusieurs travaux ont été réalisés dans ce cadre, on peut citer Habib et Thomas (1986), Akritas (1988),

Hjort (1990), Kim (1993), Bagdonavičius et Nikulin (2012). Plus de détails sur la constructions des tests du type du Chi-deux pour les données censurées, voir la monographie récente de Bagdonavičius, Kruopis et Nikulin (2010a).

Dans ce travail, nous nous sommes intéressés essentiellement à la construction et la mise en œuvre de test d'ajustement du type du Chi-deux modifié pour les familles des lois de fonction de hasard unimodale et plus particulièrement pour la famille des lois Gaussiennes inverses. Ensuite, nous avons élaboré un logiciel de calcul qui nous a permis de faire des simulations aussi bien pour les données complètes que pour les données censurées. D'autre part, nous avons fait une étude comparative [Lemeshko et al. (2010)] entre le test de Nikulin-Rao-Robson et les tests non-paramétriques de Kolmogorov-Smirnov, Cramér von-Mises et d'Anderson-Darling.

Et pour montrer l'importance de la distribution Gaussienne inverse par rapport aux autres distributions, nous avons construit le test de la loi Gaussienne inverse contre la loi log-normale et le test de la loi Gaussienne inverse contre la loi de Weibull généralisée.

D'autres travaux concernant le test d'ajustement pour la distribution Gaussienne inverse ont été réalisés. On peut citer : Gunes et al. (1997), Mudholkar et al.(2001), O'Reilly et al. (1992), Henze et al. (2002), Edgeman et al. (1988).

Le manuscrit s'articule autour de quatre chapitres.

Dans le premier chapitre, nous donnons une description générale des familles de distributions de fonction de hasard unimodale et la description de la distribution Gaussienne inverse et ses quelques propriétés et caractéristiques, puis nous donnons le tableau des estimateurs non biaisés de variance minimale pour quelques fonctions de paramètres en termes de la statistique exhaustive minimale, en utilisant toute la puissance de la théorie d'estimation paramétrique, voir par exemple, Voinov et Nikulin (1993).

Dans le deuxième chapitre, nous exposons la théorie du test du Chi-deux en donnant les méthodes d'estimation, puis nous construisons les tests d'ajustement modifiés : Nikulin-Rao-Robson, Dzhabaridze et Nikulin et McCulloch (1985) pour les familles citées ci-dessous. Ensuite, nous construisons le test de type du Chi-deux de la loi Gaussienne inverse contre la loi log-normale et le test de type du Chi-deux de la loi Gaussienne inverse contre la loi de Weibull généralisée. Nous avons calculé la puissance de ces tests.

Dans le troisième chapitre, nous construisons des modèles avec covariables basés sur la distribution Gaussienne inverse, en particulier le modèle AFT et le modèle de Sed'yakin pour étudier la fiabilité

des systèmes redondants.

Dans le dernier chapitre, nous exposons les systèmes redondants basés sur les familles paramétriques de fonction de hasard unimodale avec simulation.

Chapitre 1

Les familles des lois de fonction de hasard unimodale

1 Introduction

Dans ce chapitre, nous illustrons brièvement les distributions qui ont une fonction de hasard unimodale et nous donnons quelques caractéristiques et propriétés statistiques de la distribution Gaussienne inverse. Pour plus de détails concernant cette dernière, on peut voir Chhikara et Folks (1989), Seshadri (1993) et Seshadri (1999).

2 La distribution log-normale

La distribution log-normale (LN) est aussi appelée loi de Galton. On dit d'une variable aléatoire X suit une loi log-normale de paramètres ν et σ^2 si son logarithme suit une loi normale.

La fonction de densité, la fonction survie et la fonction de hasard de la loi log-normale sont respectivement données par :

$$f_{LN}(x, \theta_{LN}) = \frac{1}{\sigma\sqrt{2\pi x}} \exp\left\{-\frac{(\ln x - m)^2}{2\sigma^2}\right\}, \quad x > 0, \quad \theta_{LN} = (m, \sigma)^T, \quad \sigma > 0, \quad m \in \mathbf{R}, \quad (1.1)$$

$$S_{LN}(x, \theta_{LN}) = 1 - F_{LN}(\theta_{LN}) = 1 - \Phi\left(\frac{\ln x - m}{\sigma}\right), \quad (1.2)$$

$$h_{LN}(x, \theta_{LN}) = \frac{\frac{1}{\sigma\sqrt{2\pi x}} \exp\left\{-\frac{(\ln x - m)^2}{2\sigma^2}\right\}}{1 - \Phi\left(\frac{\ln x - m}{\sigma}\right)}. \quad (1.3)$$

Où Φ est la fonction de répartition de la loi normale standard $N(0, 1)$. Rappelons qu'une fonc-

tion de hasard est unimodale (ou \cap -shape) si elle augmente de zéro à un maximum puis diminue à zéro (ou à une constante non nulle). Pour la distribution LN , la fonction de hasard (1.3) est unimodale.

Dans les figures 1.1, 1.2 et 1.3, nous représentons les fonctions de densité, de survie et de hasard en fixant un paramètre et en variant l'autre.

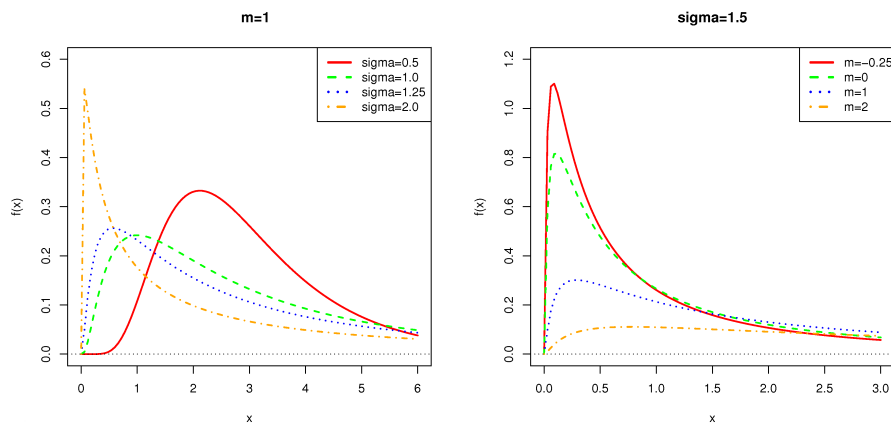


Figure 1.1 – Fonctions de densité de la distribution LN .

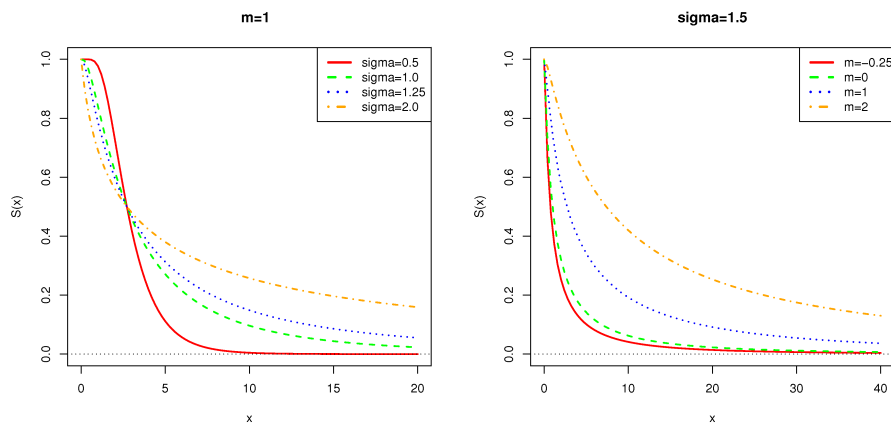


Figure 1.2 – Fonctions de survie de la distribution LN .

Cette loi trouve ses applications notamment en fiabilité, en économie, en biologie pour modéliser le poids des organismes vivants et en hydrologie pour modéliser les débits mensuels de petits bassins versants à régimes pluviaux.

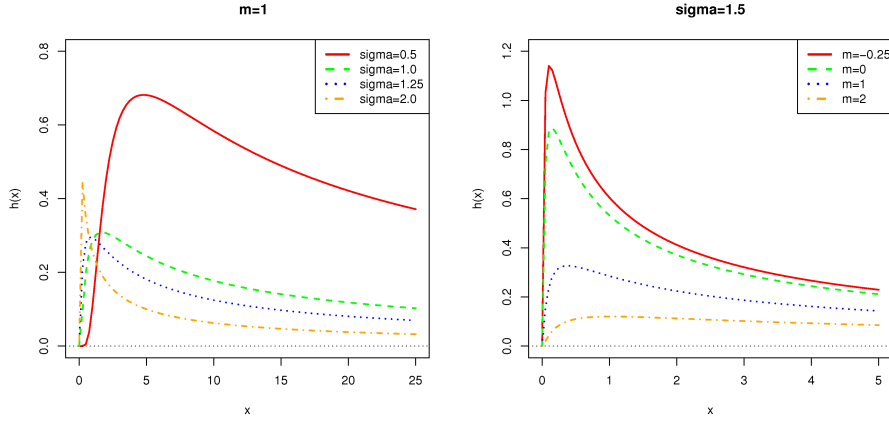


Figure 1.3 – Fonctions de hasard de la distribution LN .

3 La distribution log-logistique

La distribution log-logistique (LL) est largement utilisée dans l'analyse de survie lorsque la fonction de hasard présente une forme unimodale. Elle ressemble beaucoup la distribution LN , mais elle a l'avantage d'avoir des expressions simples pour les fonctions de survie et de hasard. Nous pouvons utiliser la distribution LL pour modéliser le taux de mortalité dû au cancer après diagnostic ou traitement, et aussi en hydrologie pour modéliser le débit d'un cours d'eau ou le niveau des précipitations et en économie pour modéliser l'inégalité des revenus.

Les fonctions de survie et de densité de la distribution LL sont :

$$S_{LL}(x, \theta_{LL}) = 1 - F_{LL}(\theta_{LL}) = \frac{1}{1 + \left(\frac{x}{\sigma}\right)^\nu}, \quad x \geq 0, \quad \theta_{LL} = (\sigma, \nu)^T, \quad \sigma > 0, \quad \nu > 0, \quad (1.4)$$

$$f_{LL}(x, \theta_{LL}) = \frac{\nu}{\sigma^\nu} x^{\nu-1} \left(1 + \left(\frac{x}{\sigma}\right)^\nu\right)^{-2}, \quad (1.5)$$

et la fonction de hasard est

$$h_{LL}(x, \theta_{LL}) = \frac{\nu}{\sigma^\nu} x^{\nu-1} \frac{1}{1 + \left(\frac{x}{\sigma}\right)^\nu}. \quad (1.6)$$

Dans les figures 1.4, 1.5 et 1.6, nous représentons les fonctions de densité, de survie et de hasard pour la distribution LL en fixant un paramètre et en variant l'autre.

Notons que si $\nu < 1$, la fonction de hasard h_{LL} (1.6) est monotone décroissante de l'infini à 0. Si $\nu > 1$, h_{LL} est unimodale. Notons aussi que Santana et al. (2012) ont construit une distribution appelée "Kumaraswamy-log-logistic distribution" basée sur les distributions log-logistique et Kumaraswamy,

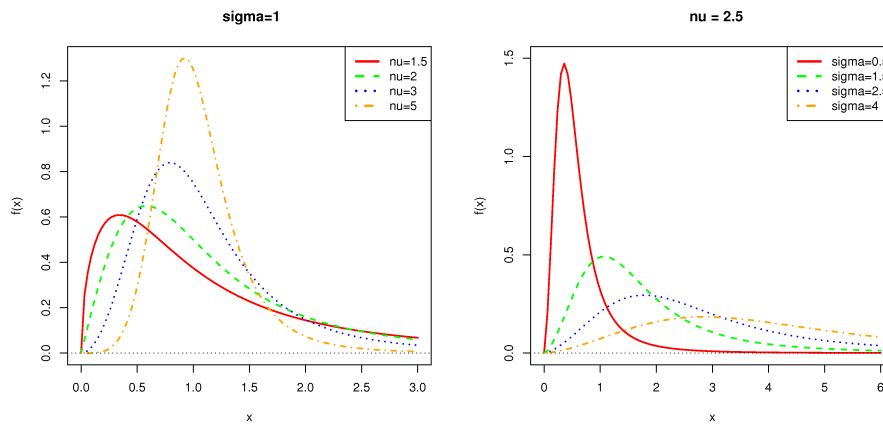


Figure 1.4 – Fonctions de densité de la distribution LL .

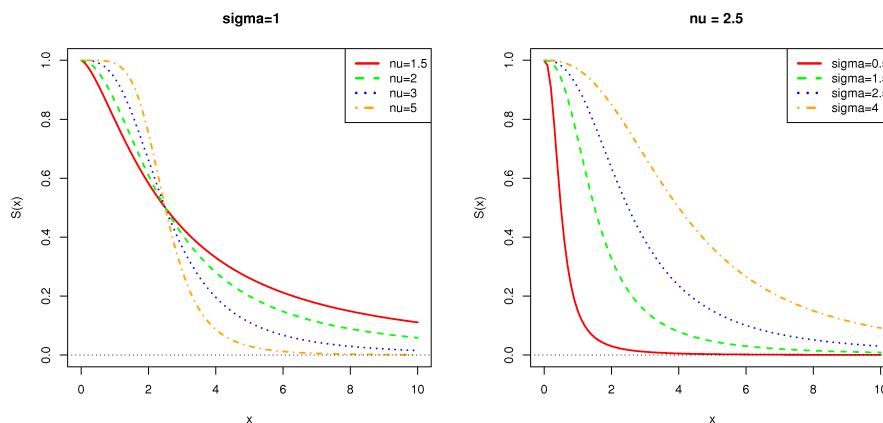
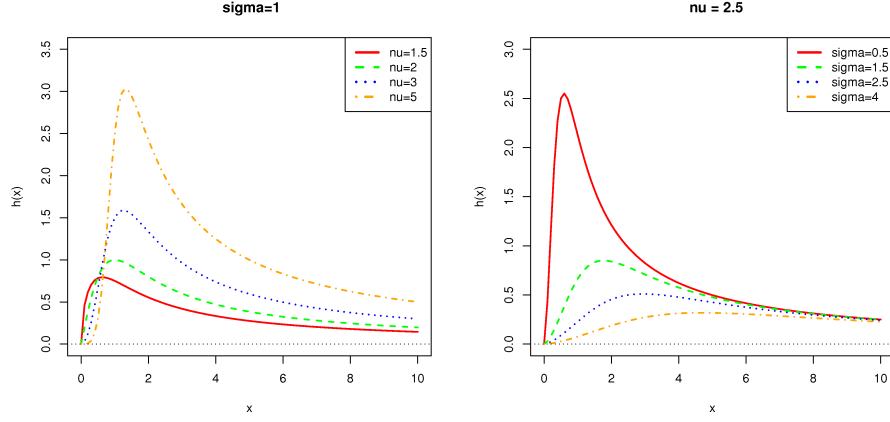


Figure 1.5 – Fonctions de survie de la distribution LL .

sa fonction de hasard pour cette nouvelle distribution peut être monotone (croissante ou décroissante), unimodale ou en forme de baignoire (i.e. \cup -shape).

4 La distribution de Weibull exponentielle

La distribution de Weibull exponentielle (Exponentiated Weibull) (EW) a été introduite par Efron(1988). Ses propriétés ont été étudiées par Mudholkar et al. (1995, 1996). Cette distribution possède de bonnes caractéristiques comme la distribution de Weibull généralisée, sa fonction de hasard peut être constante, monotone (croissante ou décroissante), unimodale ou en forme de baignoire. Si $\gamma < \nu < 1$, la fonction de hasard h_{EW} (1.9) est unimodale


 Figure 1.6 – Fonctions de hasard de la distribution LL .

La fonction de densité de la distribution EW est :

$$f_{EW}(x, \theta_{EW}) = \frac{\nu}{\gamma \sigma^\nu} x^{\nu-1} e^{-\left(\frac{x}{\sigma}\right)^\nu} \left\{ 1 - e^{-\left(\frac{x}{\sigma}\right)^\nu} \right\}^{\frac{1}{\gamma}-1}, \quad x \geq 0, \quad \theta_{EW} = (\sigma, \nu, \gamma)^T, \quad \sigma, \nu, \gamma > 0, \quad (1.7)$$

et les fonctions de survie et de hasard sont :

$$S_{EW}(x, \theta_{EW}) = 1 - F_{EW}(x, \theta_{EW}) = 1 - \left\{ 1 - e^{-\left(\frac{x}{\sigma}\right)^\nu} \right\}^{\frac{1}{\gamma}}, \quad (1.8)$$

et

$$h_{EW}(x, \theta_{EW}) = \frac{\nu \left(\frac{x}{\sigma}\right)^{\nu-1} e^{-\left(\frac{x}{\sigma}\right)^\nu} \left\{ 1 - e^{-\left(\frac{x}{\sigma}\right)^\nu} \right\}^{\frac{1}{\gamma}-1}}{\sigma \gamma \left\{ 1 - \left(1 - e^{-\left(\frac{x}{\sigma}\right)^\nu} \right)^{\frac{1}{\gamma}} \right\}}. \quad (1.9)$$

Dans les figures 1.7, 1.8 et 1.9, nous représentons respectivement les fonction de densité, de survie et de hasard pour la distribution EW en fixant un paramètre et en variant les deux autres.

5 La distribution de Weibull généralisée

La distribution de Weibull généralisée (GW) est proposée par Bagdonavičius et Nikulin (2002). A cause de sa flexibilité, la distribution de Weibull généralisée possède de bonnes propriétés : la fonction de hasard peut être constante, monotone (croissante ou décroissante), unimodale ou en forme de baignoire, ceci explique pourquoi cette loi est la plus utilisée dans les modèles de vie accélérés. Dans le cas où la fonction de hasard est unimodale, nous aurons nécessairement $\gamma > \nu > 1$.

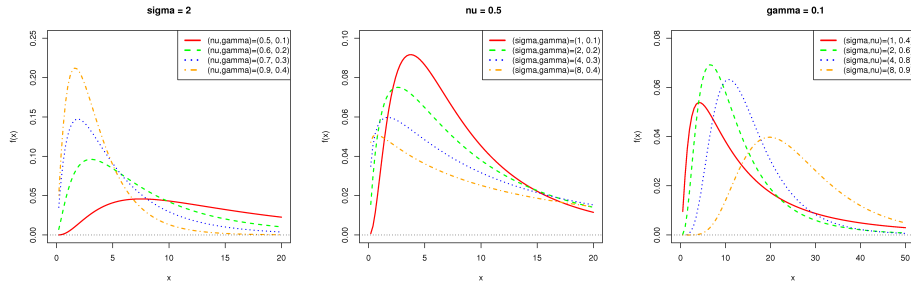


Figure 1.7 – Fonctions de densité de la distribution EW .

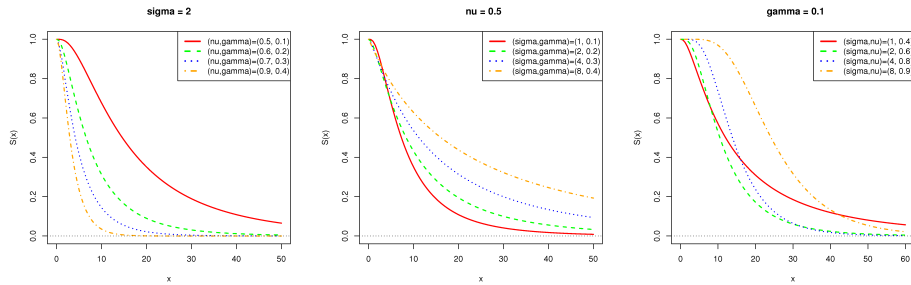


Figure 1.8 – Fonctions de survie de la distribution EW .

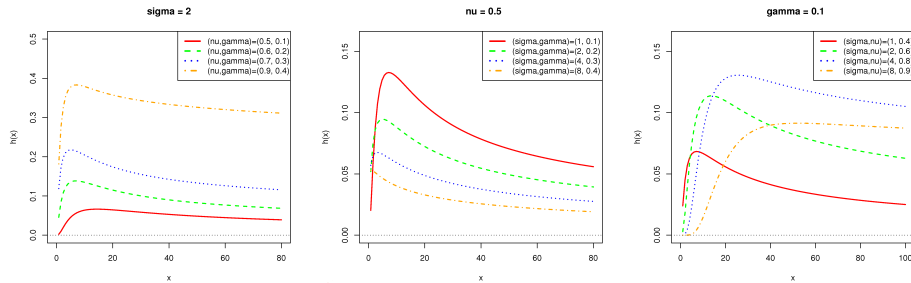


Figure 1.9 – Fonctions de hasard de la distribution EW .

La fonction de survie de la distribution de GW est donnée par :

$$S_{GW}(x, \theta_{GW}) = 1 - F_{GW}(\theta_{GW}) = \exp \left\{ 1 - \left(1 + \left(\frac{x}{\sigma} \right)^\nu \right)^{\frac{1}{\gamma}} \right\}, \quad x \geq 0, \quad \theta_{GW} = (\sigma, \nu, \gamma)^T, \quad \sigma, \nu, \gamma > 0, \quad (1.10)$$

et la fonction de densité est :

$$f_{GW}(x, \theta_{GW}) = \frac{\nu}{\gamma \sigma^\nu} x^{\nu-1} \left(1 + \left(\frac{x}{\sigma} \right)^\nu \right)^{\frac{1}{\gamma}-1} \exp \left\{ 1 - \left(1 + \left(\frac{x}{\sigma} \right)^\nu \right)^{\frac{1}{\gamma}} \right\}. \quad (1.11)$$

1.5 La distribution de Weibull généralisée

La fonction de hasard est :

$$h_{GW}(x, \theta_{GW}) = \frac{\nu}{\gamma \sigma^\nu} x^{\nu-1} \left(1 + \left(\frac{x}{\sigma} \right)^\nu \right)^{\frac{1}{\gamma}-1}. \quad (1.12)$$

Dans les figures 1.10, 1.11 et 1.12 nous représentons respectivement les fonction de densité, de survie et de hasard pour la distribution GW en fixant un paramètre et en variant les deux autres.

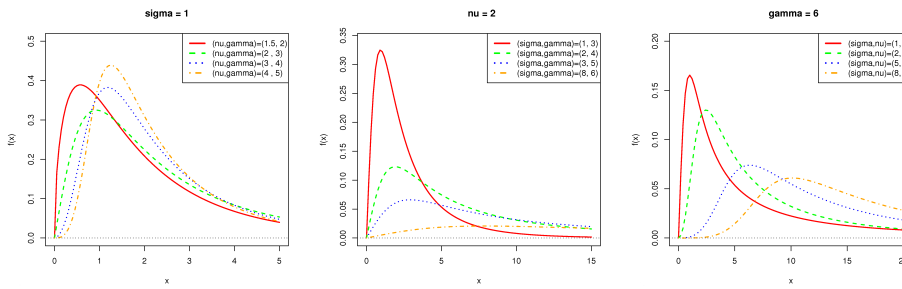


Figure 1.10 – Fonctions de densité de la distribution GW .

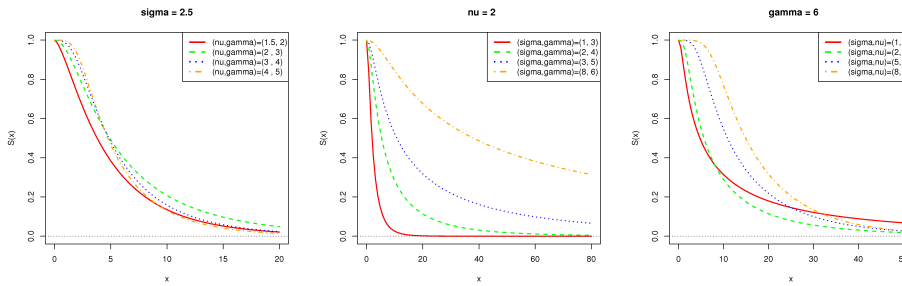


Figure 1.11 – Fonctions de survie de la distribution GW .

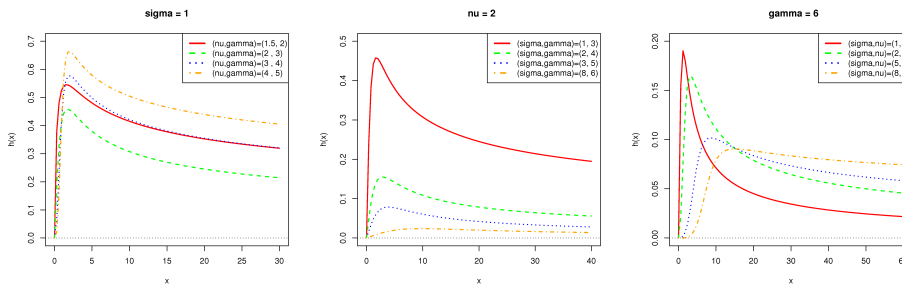


Figure 1.12 – Fonctions de hasard de la distribution GW .

6 La distribution de Birnbaum-Saunders

La distribution de Birnbaum-Saunders (*BS*) [Birnbaum et Saunders (1969a,1969b)] est récemment devenu un modèle très populaire dans le domaine des distributions de fiabilité et de durée de vie. En particulier, cette distribution est utilisée pour modéliser la durée de vie des matériaux soumis à un stress cyclique où la défaillance est due à la croissance d'une fissure dominante.

$$S_{BS}(x, \theta_{BS}) = 1 - F_{BS}(x, \theta_{BS}) = \Phi \left[\frac{1}{\alpha} \left\{ \left(\frac{\beta}{x} \right)^{\frac{1}{2}} - \left(\frac{x}{\beta} \right)^{\frac{1}{2}} \right\} \right], \quad x > 0, \quad \theta_{BS} = (\alpha, \beta)^T, \quad \alpha, \beta > 0, \quad (1.13)$$

et la fonction de densité est

$$f_{BS}(x, \theta_{BS}) = \frac{1}{2\sqrt{2\pi\alpha\beta}} \left\{ \left(\frac{\beta}{x} \right)^{\frac{1}{2}} + \left(\frac{x}{\beta} \right)^{\frac{3}{2}} \right\} \exp \left[-\frac{1}{2\alpha^2} \left(\frac{x}{\beta} + \frac{\beta}{x} - 2 \right) \right]. \quad (1.14)$$

La fonction de hasard est

$$h_{BS}(x, \theta_{BS}) = \frac{\frac{1}{2\sqrt{2\pi\alpha\beta}} \left\{ \left(\frac{\beta}{x} \right)^{\frac{1}{2}} + \left(\frac{x}{\beta} \right)^{\frac{3}{2}} \right\} \exp \left[-\frac{1}{2\alpha^2} \left(\frac{x}{\beta} + \frac{\beta}{x} - 2 \right) \right]}{\Phi \left[\frac{1}{\alpha} \left\{ \left(\frac{\beta}{x} \right)^{\frac{1}{2}} - \left(\frac{x}{\beta} \right)^{\frac{1}{2}} \right\} \right]}. \quad (1.15)$$

Nous pouvons montrer que :

$$\lim_{x \rightarrow \infty} h_{BS}(x, \theta_{BS}) = \frac{1}{2\beta^2}. \quad (1.16)$$

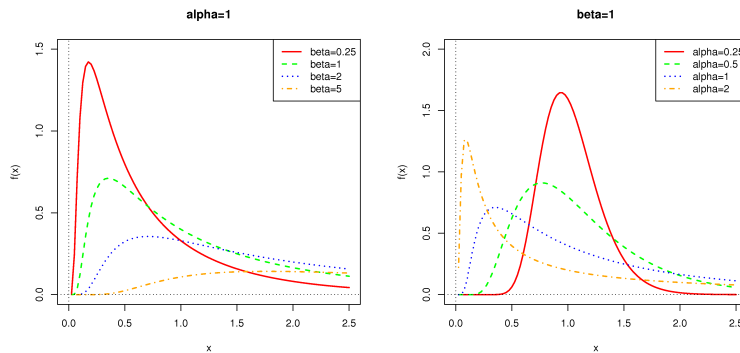


Figure 1.13 – Fonctions de densité de la distribution *BS*.

Dans les figures 1.13, 1.14 et 1.15, nous représentons respectivement les fonctions de densité, de survie et de hasard pour la distribution *BS* en fixant un paramètre et en variant l'autre.

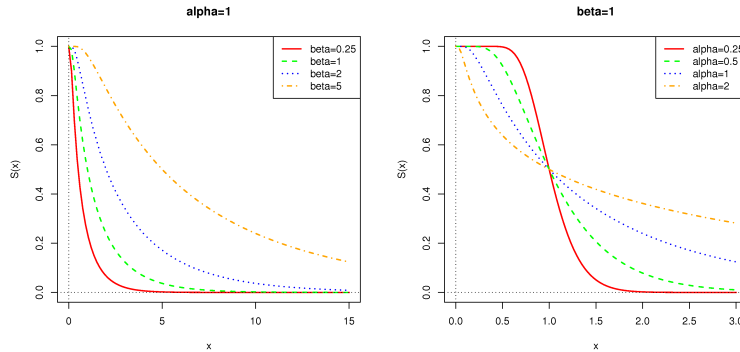


Figure 1.14 – Fonctions de survie de la distribution *BS*.

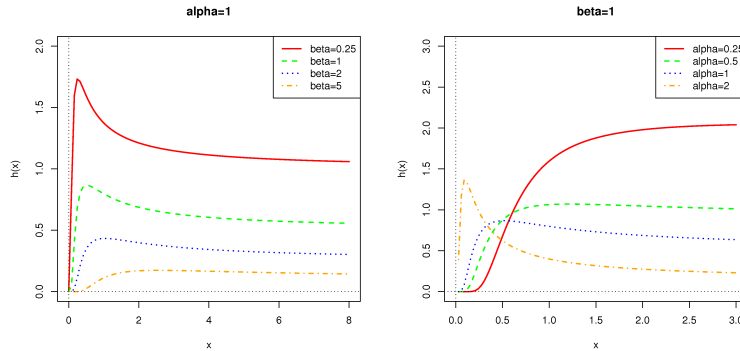


Figure 1.15 – Fonctions de hasard de la distribution *BS*.

7 La distribution Gaussienne inverse

7.1 Densité de probabilité

On dit qu'une variable aléatoire X suit une loi *Gaussienne inverse* (*IG*), et on note $X \sim IG(\mu, \lambda)$ si sa fonction de densité s'écrit :

$$f_{IG}(x, \theta_{IG}) = \left(\frac{\lambda}{2\pi x^3} \right)^{\frac{1}{2}} \exp \left\{ -\frac{\lambda(x - \mu)^2}{2\mu^2 x} \right\}, \quad x \geq 0, \quad \theta_{IG} = (\mu, \lambda)^T, \quad \mu, \lambda > 0. \quad (1.17)$$

Où μ et λ sont deux paramètres strictement positifs, et μ représente la moyenne de X . La fonction de densité (1.17) peut être paramétrisée de plusieurs façons :

$$f_1(x, \alpha, \lambda) = \left(\frac{\lambda}{2\pi x^3} \right)^{\frac{1}{2}} \exp \left\{ -\alpha \lambda x + \lambda \sqrt{2\alpha} - \frac{\lambda}{2x} \right\},$$

$$f_2(x, \mu, \phi) = \left(\frac{\mu\phi}{2\pi x^3} \right)^{\frac{1}{2}} \exp\left\{-\frac{\phi x}{2\mu} + \phi - \frac{\mu\phi}{2x}\right\},$$

$$f_3(x, \phi, \lambda) = \left(\frac{\lambda}{2\pi x^3} \right)^{\frac{1}{2}} \exp\left\{-\frac{\phi^2 x}{2\lambda} + \phi - \frac{\lambda}{2x}\right\},$$

où (μ, λ) est remplacé par (α, λ) , (μ, ϕ) ou (ϕ, λ) , avec : $\phi = \frac{\lambda}{\mu}$, $\phi > 0$, $\mu = (2\alpha)^{-1/2}$, $\alpha > 0$. Chacune de ces formes est adaptée à l'analyse des données dans différentes disciplines. On peut facilement vérifier les relations suivantes :

$$f(x, \theta_{IG}) = \mu^{-1} f_2\left(\frac{x}{\mu}, 1, \phi\right) = \lambda^{-1} f_3\left(\frac{x}{\lambda}, \phi, 1\right).$$

Le paramètre $\phi = \frac{\lambda}{\mu}$ est souvent appelé le paramètre de forme (shape parameter). Plus ϕ prend des valeurs modérées plus la densité (1.17) est fortement dissymétrique, et plus ϕ est grand, plus la densité (1.17) est proche d'une densité normale. Dans la figure (1.16), nous représentons la densité de $IG(\mu, \lambda)$ en fixant un paramètre et en variant l'autre.

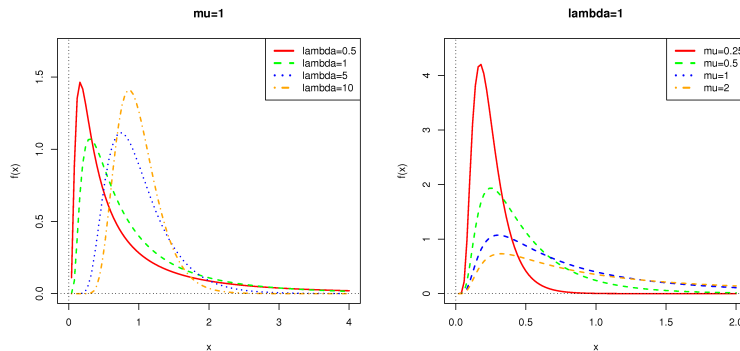


Figure 1.16 – Fonctions de densité de la distribution IG .

La fonction (1.17) est unimodale avec un mode égal à :

$$M_o = \mu \left\{ \left(1 + \frac{9}{4\phi^2} \right)^{\frac{1}{2}} - \frac{3}{2\phi} \right\}.$$

Les coefficients d'asymétrie (skewness coefficient) et d'aplatissement (kurtosis coefficient) de $IG(\mu, \lambda)$ sont respectivement :

$$\gamma_1 = \frac{\mu_3}{\sigma^3} = \frac{3}{\sqrt{\phi}}, \quad \gamma_2 = \frac{\mu_4}{\sigma^4} = 15/\phi + 3.$$

D'après γ_2 , on voit clairement que si le paramètre ϕ devient très grand, la densité de la distribution Gaussienne inverse tend vers celle de la loi normale, et que cette distribution possède plusieurs pro-

Ordre	Moment simple	Moment centré	Cumulant
1	μ	-	μ
2	$\mu^2 + \mu^3/\lambda$	$\mu^3/\lambda = Var(X)$	μ^3/λ
3	$\mu^3 + 3\mu^4/\lambda + 3\mu^5/\lambda^2$	$3\mu^5/\lambda^2$	$3\mu^5/\lambda^2$
4	$\mu^4 + 6\mu^5/\lambda + 15\mu^6/\lambda^2 + 15\mu^7/\lambda^3$	$3\mu^6/\lambda^2 + 15\mu^7/\lambda^3$	$15\mu^7/\lambda^3$

Tableau 1.1 – Les quatre premiers moments de la distribution IG .

propriétés analogues à la distribution normale.

7.2 Fonction caractéristique et moments

Soit $X \sim IG(\mu, \lambda)$. La fonction *génératrice des moments* $M_X(t)$ de X est donnée par :

$$M_X(t) = E(e^{tX}) = \int_0^{+\infty} e^{tx} f(x, \mu, \lambda) dx = \exp \left\{ \frac{\lambda}{\mu} \left(1 - \left(1 - \frac{2\mu^2 t}{\lambda} \right)^{\frac{1}{2}} \right) \right\}, \quad (1.18)$$

et la fonction *caractéristique* $C_X(t)$ de X est donnée par :

$$C_X(t) = E(e^{itX}) = \exp \left\{ \frac{\lambda}{\mu} \left(1 - \left(1 - \frac{2i\mu^2 t}{\lambda} \right)^{\frac{1}{2}} \right) \right\}.$$

Tous les moments positifs et les moments négatifs existent. Les moments positifs (moments négatifs) peuvent être calculés en prenant les dérivées (en intégrant) de l'équation (1.18) au point $t = 0$.

Proposition 1.1. [Seshadri (1993)] Pour tout k un entier strictement non négatif, on a :

$$E(X^k) = \mu^k \sum_{i=0}^{k-1} \frac{(k-1+i)!}{i!(k-1-i)!} (2\phi)^{-i}, \quad k \geq 1,$$

et

$$E(X^{-k}) = \frac{E(X^{k+1})}{\mu^{2k+1}}.$$

Si $k = 1$, on trouve : $E(X) = \mu$, $E\left(\frac{1}{X}\right) = \frac{E(X^2)}{\mu^3} = \frac{1}{\mu} + \frac{1}{\lambda}$.

Définition 1.1. Etant donné une variable aléatoire X de moyenne μ et de variance σ^2 . La fonction *génératrice des cumulants* $g(t)$ est le logarithme de la fonction génératrice des moments, i.e.

$$g(t) = \ln E(e^{tX}) = \sum_{r=1}^{\infty} \kappa_r \frac{t^r}{r!} = \mu t + \sigma^2 \frac{t^2}{2} + \dots, \quad (1.19)$$

cette fonction est intimement liée à la fonction génératrice des moments et à la fonction caractéristique de la variable X .

Les cumulants κ_r s'obtiennent en prenant les dérivées de la fonction (1.19) au point $t = 0$, i.e. $\kappa_r = g^{(r)}(0)$.

En utilisant l'équation (1.18), la fonction génératrice des cumulants de $IG(\mu, \lambda)$ est donnée par :

$$g_{IG}(t) = \frac{\lambda}{\mu} \left(1 - \left(1 - \frac{2\mu^2 t}{\lambda} \right)^{\frac{1}{2}} \right),$$

la dérivée d'ordre r de la fonction génératrice des cumulants g_{IG} est :

$$g_{IG}^{(r)}(t) = \frac{\mu^{2r-1}}{\lambda^{r-1}} 1.3.5.7.9 \dots (2r-3) \left(1 - \frac{2\mu^2 t}{\lambda} \right)^{-r+\frac{1}{2}},$$

et les cumulants κ_r sont donnés par la relation suivante :

$$\text{Pour } r \geq 2, \quad \kappa_r = 1.3.5 \dots (2r-3) \frac{\mu^{2r-1}}{\lambda^{r-1}}.$$

Dans le tableau (1.1), nous donnons les quatre premiers moments simples, moments centrés ainsi que les quatre premiers cumulants.

Remarque 1.1. On dit que les deux variables aléatoires X et Y sont *inverses* si la fonction génératrice des cumulants de X est l'inverse de celle de Y . La fonction génératrice des cumulants de la loi normale $N(m, \sigma^2)$ est donnée par

$$g_N(t) = -mt + \frac{1}{2}\sigma^2 t^2.$$

Nous pouvons vérifier facilement que la fonction g_N est l'inverse de g_{IG} , d'où l'origine du nom "inverse Gaussian". Nous pouvons aussi vérifier que la loi binomiale est l'inverse de la loi binomiale négative, et que la loi de Poisson est l'inverse de la loi Gamma.

7.3 La fonction de répartition

Si $X \sim IG(\mu, \lambda)$, la fonction de répartition de X peut s'écrire :

$$\begin{aligned}
 F_{IG}(x, \theta_{IG}) &= \int_0^x \left(\frac{\lambda}{2\pi t^3} \right)^{\frac{1}{2}} \exp \left\{ -\frac{\lambda(t - \mu)^2}{2\mu^2 t} \right\} dt \\
 &= \Phi \left(\sqrt{\frac{\lambda}{x}} \left(\frac{x}{\mu} - 1 \right) \right) + \exp \left(\frac{2\lambda}{\mu} \right) \Phi \left(-\sqrt{\frac{\lambda}{x}} \left(\frac{x}{\mu} + 1 \right) \right),
 \end{aligned} \tag{1.20}$$

La fonction de survie est :

$$S_{IG}(x, \theta_{IG}) = 1 - F_{IG}(x, \theta_{IG}) = \Phi \left(-\sqrt{\frac{\lambda}{x}} \left(\frac{x}{\mu} - 1 \right) \right) - \exp \left(\frac{2\lambda}{\mu} \right) \Phi \left(-\sqrt{\frac{\lambda}{x}} \left(\frac{x}{\mu} + 1 \right) \right). \tag{1.21}$$

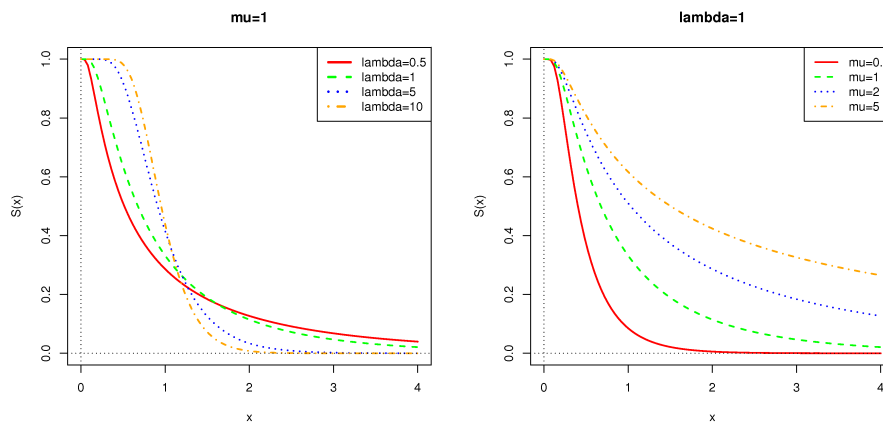


Figure 1.17 – Fonctions de survie de la distribution IG .

Dans la figure 1.17 nous représentons la fonction de survie de la distribution IG pour diverses combinaisons de μ et λ .

Proposition 1.2 Si le rapport $\frac{\lambda}{\mu} = \phi$ est très grand alors la fonction de répartition de X peut être approximée par :

$$F_{IG}(x, \mu, \lambda) \approx \Phi \left(\sqrt{\frac{\lambda}{x}} \left(\frac{x}{\mu} - 1 \right) \right).$$

Preuve :

Il suffit de montrer que :

$$\lim_{\frac{\lambda}{\mu} \rightarrow \infty} \exp\left(\frac{2\lambda}{\mu}\right) \Phi\left(-\sqrt{\frac{\lambda}{x}}\left(\frac{x}{\mu} + 1\right)\right) = 0, \quad (1.22)$$

soit $y = \sqrt{\frac{\lambda}{x}}\left(\frac{x}{\mu} - 1\right)$ alors (1.20) devient :

$$F_{IG}(x, \mu, \lambda) = \Phi(y) + \exp\left(\frac{2\lambda}{\mu}\right) \Phi\left(-\sqrt{y^2 + \frac{4\lambda}{\mu}}\right), \quad (1.23)$$

d'où suivant (1.23) :

$$\begin{aligned} \lim_{\frac{\lambda}{\mu} \rightarrow \infty} \exp\left(\frac{2\lambda}{\mu}\right) \Phi\left(-\sqrt{\frac{\lambda}{x}}\left(\frac{x}{\mu} + 1\right)\right) &= \lim_{\phi \rightarrow \infty} e^{2\phi} \Phi\left(-\sqrt{y^2 + 4\phi}\right) = \\ \lim_{\phi \rightarrow \infty} \frac{\Phi\left(-\sqrt{y^2 + 4\phi}\right)}{e^{-2\phi}} &= \lim_{\phi \rightarrow \infty} \frac{\frac{-2}{\sqrt{y^2 + 4\phi}} \frac{1}{\sqrt{2\pi}} e^{-\left(\frac{1}{2}y^2 + 4\phi\right)}}{-2e^{-2\phi}} = 0. \end{aligned}$$

Ce dernier résultat est obtenu en appliquant la règle de L'Hospital.

7.4 La fonction de hasard

La fonction de *hasard* de la loi *IG* est donnée par :

$$h_{IG}(x, \theta_{IG}) = \frac{f(x, \mu, \lambda)}{S(x, \mu, \lambda)} = \frac{\left(\frac{\lambda}{2\pi x^3}\right)^{\frac{1}{2}} \exp\left\{-\frac{\lambda(x-\mu)^2}{2\mu^2 x}\right\}}{\Phi\left(-\sqrt{\frac{\lambda}{x}}\left(\frac{x}{\mu} - 1\right)\right) - \exp\left(\frac{2\lambda}{\mu}\right) \Phi\left(-\sqrt{\frac{\lambda}{x}}\left(\frac{x}{\mu} + 1\right)\right)}, \quad x \geq 0. \quad (1.24)$$

On peut facilement vérifier que cette fonction est unimodale, de plus nous avons :

$$\lim_{x \rightarrow \infty} h_{IG}(x, \theta_{IG}) = \frac{\lambda}{2\mu^2}. \quad (1.25)$$

Notons que les fonctions de hasard (1.15) et (1.24) des distributions *BS* et *IG* possèdent à l'infinie les limites (1.16) et (1.25) respectivement, cela signifie que la fonction de hasard de chaque une de ces distributions croît de 0 à son maximum puis tend asymptotiquement vers une constante non nulle. En revanche, pour les distributions *LN*, *LL*, *EW* et *PGW* la fonction de hasard croit de 0 à son maximum puis diminue jusqu'à 0. Le comportement asymptotique des fonctions de hasard (1.15) et (1.24) implique que l'occurrence de la panne devient purement aléatoire et indépendante du passé (sans mémoire)[Gunes et al. (1997)], par contre les autres distributions escamotent la panne ce qui n'est pas possible pour la plupart des systèmes.

Dans la figure 1.18 nous représentons la fonction de hasard de la distribution *IG* pour diverses combinaisons des paramètres μ et λ .

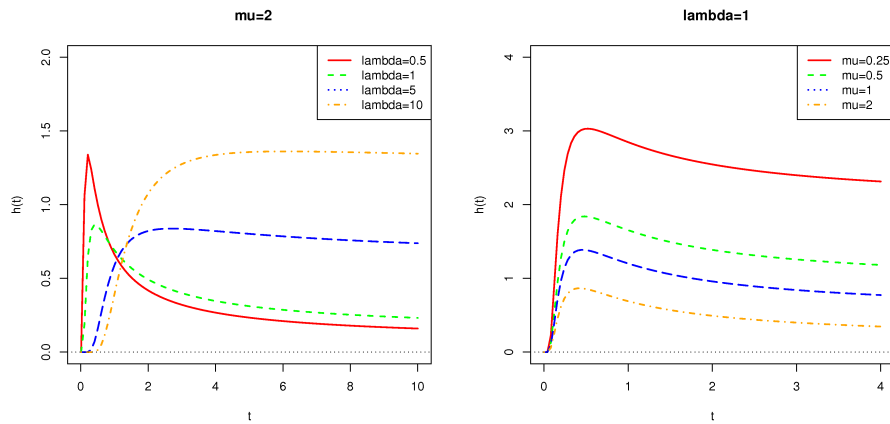


Figure 1.18 – Fonctions de hasard de la distribution *IG*.

7.5 Analogies avec la loi Normale

Proposition 1.3. [Seshadri (1993)]. Soit $\mathbb{X} = (X_1, X_2, \dots, X_n)^T$ un n -échantillon (i.e. n variables aléatoires indépendantes et identiquement distribuées) tel que $X_i \sim IG(\mu, \lambda)$. On considère la variable aléatoire :

$$Y = \frac{\lambda}{\mu^2} \frac{(X_i - \mu)^2}{X_i},$$

alors :

$$Y \sim \Gamma\left(\frac{1}{2}, \frac{1}{2}\right) = \chi_1^2.$$

Par généralisation, on obtient :

$$\sum_{i=1}^n \frac{\lambda(X_i - \mu)^2}{\mu^2 X_i} \sim \chi_n^2.$$

Proposition 1.4. [Seshadri (1993)]. Si $X_i \sim IG(\mu, \lambda)$, alors $X_i^{-1} \sim IG(1/\mu, \lambda/\mu^2) * \Gamma(1/2, \lambda/2)$.

7.6 Caractéristiques

Théorème 1.1. [Chhikara et Folks (1989)]. Soit $\mathbb{X} = (X_1, X_2, \dots, X_n)^T$ un n -échantillon. On suppose que :

$$\mathbf{E}(X_i), \mathbf{E}(X_i^2), \mathbf{E}\left(\frac{1}{X_i}\right), \mathbf{E}\left(\sum_{i=1}^n X_i\right)^{-1},$$

existent et sont non nulles, alors :

$$\forall i = 1, \dots, n, X_i \sim IG(\mu, \lambda) \Leftrightarrow Y = \sum_{i=1}^n X_i \text{ et } Z = \sum_{i=1}^n X_i^{-1} - n^2 Y^{-1} \text{ sont indépendantes.}$$

A la base de ce théorème, Mudholkar et al. (2001) ont construit un test d'ajustement pour la distribution IG .

7.7 Propriétés statistiques

Calculons la fonction de vraisemblance de l'échantillon X_1, X_2, \dots, X_n :

$$L(\theta_{IG}) = \prod_{i=1}^n f_{IG}(X_i, \theta_{IG}) = \lambda^{\frac{n}{2}} (2\pi)^{\frac{n}{2}} \prod_{i=1}^n X_i^{-\frac{3}{2}} \exp\left(\sum_{i=1}^n \frac{-\lambda(X_i - \mu)^2}{2\mu^2 X_i}\right), \quad \theta_{IG} = (\mu, \lambda)^T,$$

la fonction de log-vraisemblance est :

$$\ell(\theta_{IG}) = \frac{n}{2} \ln \lambda - \frac{n}{2} \ln(2\pi) - \frac{3}{2} \sum_{i=1}^n \ln X_i - \sum_{i=1}^n \frac{\lambda(X_i - \mu)^2}{2\mu^2 X_i}, \quad (1.26)$$

les fonctions de score sont :

$$\frac{\partial \ell}{\partial \mu} = \frac{\lambda}{\mu^3} \sum_{i=1}^n (X_i - \mu) = n \frac{\lambda}{\mu^3} (\bar{X} - \mu),$$

$$\frac{\partial \ell}{\partial \lambda} = \frac{n}{2\lambda} - \left(\sum_{i=1}^{i=n} \frac{(X_i - \mu)^2}{2\mu^2 X_i} \right) = \frac{n}{2\lambda} - \frac{1}{2\mu^2} (n\bar{X}) + \frac{2n}{\mu} - \frac{1}{2} \sum_{i=1}^n X_i^{-1}.$$

Les estimateurs du maximum de vraisemblance (MV) de μ et λ sont successivement

$$\begin{cases} \hat{\mu} = \bar{X} \\ \hat{\lambda} = \frac{n}{\sum_{i=1}^n X_i^{-1} - n(\bar{X})^{-1}} = \frac{n}{V} \end{cases},$$

où

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i, \quad V = \sum_{i=1}^n X_i^{-1} - n(\bar{X})^{-1}.$$

Proposition 1.5. [Seshadri (1993)]. Pour la famille $IG(\mu, \lambda)$, la statistique $T = (\bar{X}, V)^T$ est exhaustive minimale et complète pour θ_{IG} .

Remarque 1.2. Les estimateurs non biaisés à variance minimale (*MVUE*) de μ et λ sont respectivement :

$$\hat{\mu} = \bar{X} \quad \text{et} \quad \hat{\lambda} = \frac{n-3}{V}, \quad n > 3.$$

Dans le tableau (1.2), on donne les estimateurs non biaisés à variance minimale de quelques fonctions de paramètres avec $F(\cdot, \cdot, \cdot, \cdot)$ représente la fonction hypergéométrique de Gauss (voir Voinov et Nikulin (1993)).

7.8 Quelques applications

Dans ce paragraphe nous présentons quelques applications majeures de la loi Gaussienne inverse. Comme le mouvement Brownien est la base de cette loi, il est naturel qu'on pense d'abord à des applications dans les sciences naturelles et physiques.

7.8.1 Science actuarielle

Dans le domaine de mathématiques des assurances, il est important de spécifier la distribution de probabilité d'un seul coût d'indemnisation. Pour cette spécification, on calcule la distribution de probabilité du nombre total d'indemnités. Seal (1969) a proposé comme modèle, la distribution log-normale, par contre Seal (1978) et Berg (1994) ont préconisé l'utilisation de la distribution Gaussienne inverse pour la modélisation des coûts d'indemnisation.

7.8.2 Démographie

Une variété de modèles mathématiques ont été proposés dans la littérature pour la modélisation de courbes de fertilité dont le plus fréquemment utilisé est le modèle de Hadwiger (1940). La fonction de Hadwiger est donnée par :

$$f(x) = \frac{ab}{c} \left(\frac{c}{x}\right)^{\frac{3}{2}} \exp\left\{-b^2\left(\frac{c}{x} + \frac{x}{c} - 2\right)\right\}$$

où x est l'âge de la mère à la naissance et a , b , c sont les trois paramètres à estimer. Ces trois paramètres peuvent avoir une interprétation démographique comme suit : le paramètre a est associé à la fertilité totale, le paramètre c est lié à l'âge moyen de la maternité et b est le paramètre déterminant la hauteur de la courbe.

On voit clairement que la fonction de Hadwiger n'est que la densité de la distribution Gaussienne inverse.

7.8.3 Hydrologie

L'analyse des crues par les méthodes statistiques a commencé dès 1913. Bien que Holton (1913) a utilisé la loi normale pour décrire les inondations annuelles. Comme ces données sont asymétriques, alors d'autres modèles sont nécessaires. Hazen (1914) a utilisé la loi log-normale avec un certain succès, et plus tard, la théorie de Gumbel des valeurs extrêmes est devenue "le véhicule" de l'analyse de données des crues. Dans les premières études, il s'avère que la distribution des inondations doit être dotée d'un minimum de trois paramètres ce qui a motivé Halphen à développer la distribution Gaussienne inverse généralisée en 1941.

On rappelle que la distribution Gaussienne inverse généralisée à trois paramètres possède la forme :

$$f(x, \gamma, \mu, \lambda) = \left[2\mu^\gamma K_\gamma\left(\frac{\lambda}{\mu}\right)\right]^{-1} x^{\gamma-1} \exp\left\{-\frac{\lambda x^{-1} + \frac{\lambda}{\mu^2}x}{2}\right\}, \quad x \geq 0, \quad -\infty < \gamma < \infty, \quad \mu, \lambda > 0,$$

où $K_\gamma(\cdot)$ est la fonction de Bessel modifiée de 3^{eme} classe d'ordre γ .

Si $\gamma = -1/2$, on obtient la densité de la distribution Gaussienne inverse $IG(\mu, \lambda)$.

La distribution Gaussienne inverse peut éventuellement être appliquée en management, météorologie, santé mentale, physiologie, pharmacocinétique, photogrammétrie, l'intensité du bruit du trafic, écologie, entomologie, histomorphométrie, réseaux électriques et en cardiologie.

1.7 La distribution Gaussienne inverse

Pour plus de détails, on peut voir Takagi et al. (1997), Onara et al. (2000), Nadarajah (2009), Silva (2006), Shunsuke Sato et al. (2007), Chhikara et Folks (1989), Seshadri (1993), Seshadri (1999), Saunders (2007), Voinov et Nikulin (1989), Voinov et Nikulin (1993), Xiao et al.(2010), Takagi et al. (1997), Shunsuke et al.(2007), Henze et al. (2002), Onara et Padgett (2000), O'Reilly et al. (1992).

le paramètre	MVUE
$\hat{F}(x) = \mathbf{P}(X \leq x)$	<p>a)</p> $\begin{cases} 0, & x < x_1 \\ 1 - F_{n-2}(-\omega_1) + \frac{n-2}{n} \left\{ 1 + \frac{4(n-1)}{nV\bar{X}} \right\}^{\frac{n-3}{2}} F_{n-2}(-\omega_2), & x_1 \leq x \leq x_2 \\ 1, & x > x_2 \end{cases}, n > 2$ $\omega_1 = \frac{\sqrt{n(n-2)}(x - \bar{X})}{\sqrt{V\bar{X}(n\bar{X} - x)x - n(x - \bar{X})^2}},$ $\omega_2 = \frac{\sqrt{n(n-2)} \left[\bar{X} + \frac{n-2}{n}x \right]}{\sqrt{V\bar{X}(n\bar{X} - x)x - n(x - \bar{X})^2}},$ $F_{n-2}(x) = \mathbf{P}(\chi_{n-2}^2 \leq x) = \frac{\Gamma(\frac{n-1}{2})}{\sqrt{\pi}(n-2)\Gamma(\frac{n-2}{2})} \int_{-\infty}^x \left(1 + \frac{u^2}{n-2} \right)^{-\frac{n-1}{2}} du,$ $x_{1,2} = \frac{\bar{X}}{2(n + V\bar{X})} \left\{ n(2 + V\bar{X}) \pm \sqrt{4n(n-1)V\bar{X} + n^2V^2\bar{X}^2} \right\}.$ <p>b)</p> $\begin{cases} 0, & x < x_1 \\ \frac{\sqrt{n}(n-1)(n+V\bar{X})^{\frac{n-4}{2}} \Gamma(\frac{n-1}{2})}{\sqrt{\pi}V^{\frac{n-3}{2}}(\bar{X})^{\frac{n-7}{2}} \Gamma(\frac{n-2}{2})} \int_{x_1}^x (t - x_1)^{\frac{n-4}{2}} (x_2 - t)^{\frac{n-4}{2}} (t)^{-\frac{n-1}{2}} (n\bar{X} - t)^{-\frac{n-1}{2}} dt, & x_1 \leq x \leq x_2 \\ 1, & x > x_2 \end{cases}$ <p>$n > 2$</p>
$\hat{f}(x, \mu, \lambda)$	$\hat{f}(x, \mu, \lambda) = \begin{cases} 0, & 0 < V < C \\ \frac{n(n-1)\sqrt{\bar{X}^3} \Gamma(\frac{n-1}{2})(V-C)^{\frac{n-4}{2}}}{\sqrt{\pi} V^{\frac{n-3}{2}} \Gamma(\frac{n-2}{2}) \sqrt{nx^3(n\bar{X} - x)^3}}, & V > C \end{cases},$ $C = \frac{n(x - \bar{X})^2}{\bar{X}x(n\bar{X} - x)}, \quad n > 2.$

1.7 La distribution Gaussienne inverse

$\hat{f}(x, \mu, \lambda)$ μ connu	$\hat{f}(x, \mu, \lambda) = \begin{cases} 0, & 0 < V < \frac{(x-\mu)^2}{\mu^2 x} \\ \frac{\Gamma(\frac{n-1}{2}) \left[V - \frac{(x-\mu)^2}{\mu^2 x} \right]^{\frac{n-4}{2}}}{\sqrt{nx^3} V^{\frac{n-3}{2}} \Gamma(\frac{n-2}{2})}, & V > \frac{(x-\mu)^2}{\mu^2 x} \end{cases},$ <p style="margin-left: 20px;">$n > 2$</p>
$\hat{f}(x, \mu, \lambda)$ λ connu	$\hat{f}(x, \mu, \lambda) = \begin{cases} 0, & \bar{X}n < x \\ \frac{n(n-1)\sqrt{\lambda\bar{X}^3}}{\sqrt{2\pi n x^3 (n\bar{X} - x)^3}} \exp\left\{-\frac{n\lambda(\bar{X}-x)^2}{2\bar{X}x(n\bar{X}-x)}\right\}, & \bar{X}n \geq x \end{cases},$
μ	\bar{X}
λ	$\frac{n-3}{V}$
λ^α	$\frac{2^\alpha \Gamma(\frac{n-1}{2})}{V^\alpha \Gamma(\frac{n-1}{2} - \alpha)}$
λ^{-1}	$\frac{V}{n-1}$
$\frac{1}{\mu}$	
λ connu	$\frac{1}{\bar{X}} - \frac{1}{n\lambda}$
$\frac{1}{\mu}$	
λ inconnu	$\frac{1}{\bar{X}} - \frac{V}{n(n-1)}$
$\phi = \frac{\lambda}{\mu}$	$\frac{V}{n-1}$
$\frac{1}{\phi} = \frac{\mu}{\lambda}$	$\frac{\bar{X}V}{n-1}, n > 3$
$\frac{3}{\sqrt{\phi}}$	$\frac{3\Gamma(\frac{n-1}{2})\sqrt{\bar{X}V}}{\sqrt{2}\Gamma(\frac{n}{2})} F\left(-\frac{1}{4}, \frac{1}{4}; \frac{n}{2}, -\frac{\bar{X}V}{n}\right)$
$\frac{15}{\phi}$	$\frac{15\bar{X}V}{n-1}, n > 3$

μ^3/λ	<p>a)</p> $n\bar{X}^2 \left[1 - F\left(1, \frac{1}{2}, \frac{n-1}{2}, -\frac{\bar{X}V}{n}\right) \right], \quad n > 1$ <p>b)</p> $\frac{\bar{X}^3 V}{n-1} F\left(1, \frac{3}{2}, \frac{n+1}{2}, -\frac{\bar{X}V}{n}\right), \quad n > 1$
$\sqrt{\mu^3/\lambda}$	$\begin{cases} \frac{\sqrt{\bar{X}^3 V} \Gamma(\frac{n-1}{2})}{\sqrt{2} \Gamma(\frac{n}{2})} F\left(\frac{3}{4}, \frac{1}{4}, \frac{n}{2}, -\frac{\bar{X}V}{n}\right), & \bar{X}V < n \\ \frac{\sqrt{\bar{X}^3 V} n^{\frac{3}{4}} \Gamma(\frac{n-1}{2})}{\sqrt{2}(n + \bar{X}V)^{\frac{3}{4}} \Gamma(\frac{n}{2})} F\left(\frac{3}{4}, \frac{2n-1}{4}, \frac{n}{2}, -\frac{\bar{X}V}{n+\bar{X}V}\right), & \bar{X}V \geq n \end{cases},$ <p>$n \geq 2$</p>
le paramètre	MVUE
$E(X^k)$	$\Gamma\left(\frac{n-1}{2}\right) \sum_{m=0}^{k-1} \frac{2^{-2m} \Gamma(k+m) \bar{X}^{k+m} V^m}{m! \Gamma(k-m) \Gamma(\frac{n+2m-1}{2})} F\left(\frac{m+k-1}{2}, \frac{m+k}{2}, \frac{n+2m-1}{2}, -\frac{\bar{X}V}{n}\right)$
$C_X(t) = E(e^{itX})$	$\frac{(n-1) \sqrt{\bar{X}^3} \Gamma(\frac{n-1}{2})}{\sqrt{\pi n^3 V} \Gamma(\frac{n-2}{2})} \int_0^{+\infty} \tau^{-3/2} (\bar{X} - \tau)^{-3/2} e^{int\tau} \left[\left(1 - \frac{(n\tau - \bar{X}^2)^2}{n\tau \bar{X}V (\bar{X} - \tau)} \right)_+ \right] d\tau,$ <p>où $a_+ = \max(0, a)$</p>

Tableau 1.2 – Les estimateurs de *MVUE* de quelques fonctions de paramètres pour la distribution *IG*.

Chapitre 2

Test d'ajustement du Chi-deux pour les familles de fonction de hasard unimodale

1 Introduction

Dans ce chapitre, nous donnons un aperçu du test du Chi-deux de Pearson ainsi que les méthodes d'estimation et les propriétés des estimateurs. Ensuite nous construisons les tests du Chi-deux modifiés pour les distributions vues précédemment ainsi que la puissance de ces tests en mettant l'accent sur la distribution *IG*. Comme la distribution *IG* est très concurrente aux autres distributions, nous construisons alors le test de la loi *IG* contre la loi *LN* et le test de la loi *IG* contre la loi de *PGW* et ceci en utilisant la statistique de Nikulin-Rao-Robson (NRR) avec les classes de Neyman-Pearson.

2 La théorie du test du Chi-deux.

2.1 Test du Chi-deux de Pearson.

Soient (X_1, X_2, \dots, X_n) un n -échantillon (i.e. n variables aléatoires indépendantes et identiquement distribuées). On considère le problème de tester l'hypothèse H_0 suivant :

$$H_0 : \mathbf{P}(X_i \leq x) = F(x, \theta) \text{ , } \theta = (\theta_1, \theta_2, \dots, \theta_s) \in \Theta \subset \mathbf{R}^s, \text{ } x \in \mathbf{R}^1, \quad (2.1)$$

où $F(x, \theta)$ est la fonction de distribution de X_i . Soit $f(x, \theta)$ la fonction de densité correspondante à $F(x, \theta)$.

On partage la droite réelle en r sous-intervalles : I_1, I_2, \dots, I_r mutuellement disjoints, par les points :

$$-\infty = a_0 < a_1 < \dots < a_{r-1} < a_r = +\infty, \text{ } I_j =]a_{j-1}, a_j].$$

Chapitre 2. Test d'ajustement du Chi-deux pour les familles de fonction de hasard unimodale

Soit $p(\theta) = (p_1(\theta), p_2(\theta), \dots, p_r(\theta))^T$, avec

$$p_i(\theta) = \int_{a_{i-1}}^{a_i} f(x, \theta) dx = F(a_i, \theta) - F(a_{i-1}, \theta), \quad i = 1, 2, \dots, r.$$

Il résulte le vecteur de fréquences $\nu = (\nu_1, \nu_2, \dots, \nu_r)^T$ obtenu en groupant X_1, X_2, \dots, X_n dans les sous-intervalles I_j , avec

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

La matrice d'information de Fisher pour le vercteur ν est $nJ = nB^T B$, où

$$J = J(\theta) = \left[\sum_{l=1}^r \frac{1}{\sqrt{p_l(\theta)}} \frac{\partial p_l(\theta)}{\partial \theta_i} \frac{\partial p_l(\theta)}{\partial \theta_j} \right]_{s \times s} = B^T(\theta) B(\theta), \quad \text{rang}(J) = s,$$

et

$$B(\theta) = \left[\frac{1}{\sqrt{p_l(\theta)}} \frac{\partial p_l(\theta)}{\partial \theta_j} \right]_{r \times s}.$$

La statistique du Chi-deux de Pearson standard est :

$$X_n^2(\theta) = X_n^T(\theta) X_n(\theta) = \sum_{i=1}^r \frac{(\nu_i - np_i(\theta))^2}{np_i(\theta)}, \quad (2.2)$$

où

$$X_n(\theta) = \left(\frac{\nu_1 - np_1(\theta)}{\sqrt{np_1(\theta)}}, \frac{\nu_2 - np_2(\theta)}{\sqrt{np_2(\theta)}}, \dots, \frac{\nu_r - np_r(\theta)}{\sqrt{np_r(\theta)}} \right)^T. \quad (2.3)$$

sous H_0 , si θ est connu, K. Pearson a montré que pour n assez grand, on a :

Théorème 2.1. (K. Pearson)

$$\lim_{n \rightarrow \infty} \mathbf{P} \left(X_n^2(\theta) > x | H_0 \right) = \mathbf{P} \left(\chi_{r-1}^2 > x \right),$$

où χ_{r-1}^2 est la distribution de Chi-deux à $r - 1$ degrés de liberté.

Généralement θ est inconnu. Si on remplace θ dans l'équation (2.2) par un estimateur θ_n^* , la limite de distribution de la statistique de Pearson (2.2) change gramatiquement et elle n'est plus χ_{r-1}^2 et dépend de la méthode d'estimation de θ .

On peut écrire l'équation (2.3) comme suit :

$$X_n(\theta) = \frac{1}{\sqrt{n}} \mathbf{P}^{-1/2}(\nu - np),$$

où

$$\mathbf{P} = \mathbf{P}(\theta) = \begin{pmatrix} p_1(\theta) & 0 & \cdots & 0 \\ 0 & p_2(\theta) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & p_r(\theta) \end{pmatrix},$$

et l'équation (2.2) peut s'écrire :

$$X_n^2(\theta) = \frac{1}{n} (\nu - np)^T \mathbf{P}^{-1} (\nu - np).$$

2.2 Les méthodes d'estimation

Définition 2.1. On dit que l'estimateur θ_n^* de θ est consistant si :

$$\forall \epsilon, \mathbf{P}(|\theta_n^* - \theta| > \epsilon) \rightarrow 0 \text{ quand } n \rightarrow +\infty.$$

Définition 2.2. On dit que l'estimateur θ_n^* de θ est \sqrt{n} -consistant si :

$$\forall \epsilon, \exists C_\epsilon, n_\epsilon \text{ telque } \mathbf{P}(\sqrt{n}|\theta_n^* - \theta| > C_\epsilon) < \epsilon \text{ pour tout } n > n_\epsilon.$$

2.2.1 La méthode du maximum de vraisemblance (MV)

Soit $\mathbb{X} = (X_1, X_2, \dots, X_n)^T$ un n -échantillon, on appelle vraisemblance de \mathbb{X} la fonction L_n définie par :

$$L_n(\theta) = \prod_{i=1}^n f(X_i, \theta), \quad \theta \in \Theta. \quad (2.4)$$

Pour la famille $\{f(x, \theta), \theta \in \Theta\}$ on suppose que pour n très grand les conditions de LeCam (1956) de la normalité asymptotique locale (local asymptotic normality) (LAN) et différentiabilité asymptotique de L_n au point θ sont satisfaites :

$$1) L_n(\theta + \frac{1}{\sqrt{n}}h) - L_n(\theta) = \frac{1}{\sqrt{n}}h^T \Lambda_n(\theta) - \frac{1}{2}h^T I^{-1}(\theta)h + o_p(1), \quad h \in \mathbf{R}^s,$$

$$2) \mathcal{L}(\sqrt{n}(\hat{\theta}_n - \theta)) \rightarrow N(0_s, I^{-1}(\theta)),$$

3) Pour tout \sqrt{n} - consistant estimateur θ_n^* de θ :

$$\frac{1}{\sqrt{n}} (\Lambda_n(\theta_n^*) - \Lambda_n(\theta)) = \sqrt{n}I(\theta)(\theta_n^* - \theta) + o(1_s),$$

où $\Lambda_n(\theta) = \text{grad} \ln L_n(\theta)$ est le vecteur informant basé sur l'échantillon \mathbb{X} , $1_s = (1, \dots, 1)^T$ et $0_s = (0, \dots, 0)^T$,

et

$$I(\theta) = \frac{1}{n} E_{\theta}(\Lambda_n(\theta)\Lambda_n^T(\theta)),$$

est la matrice d'information de Fisher de X_i .

Soit C^* la classe de tous les estimateurs θ_n^* \sqrt{n} - consistants de θ , i.e. si $\theta_n^* \in C^*$, alors :

$$\mathcal{L}(\sqrt{n}(\theta_n^* - \theta)) \rightarrow N(0_s, U),$$

où $0_s = (0, \dots, 0)^T \subset \mathbf{R}^s$, $U = U(\theta)$ est une matrice de covariances non dégénérée. On remarque que les conditions de Lecam sont satisfaites. On considère la suite des estimateurs $\{\hat{\theta}_n\}$ \sqrt{n} - consistants du maximum de vraisemblance

$$L_n(\hat{\theta}_n) = \arg \max_{\theta \in \Theta} L_n(\theta),$$

qui satisfait :

$$\Lambda_n(\hat{\theta}_n) = \text{grad} \ln L_n(\hat{\theta}_n) = 0_s,$$

et on a :

$$\mathcal{L}(\sqrt{n}(\hat{\theta}_n - \theta)) \rightarrow N(0_s, I^{-1}(\theta)). \quad (2.5)$$

On note par $C_{I^{-1}}$ le sous-ensemble de C^* la classe de tous les estimateurs $\{\hat{\theta}_n\}$ \sqrt{n} - consistants vérifiant l'équation (2.5). Si $\theta_n^* \in C^*$ alors d'après les conditions de régularité LAN et d'après LeCam et al. (1983) la suite des estimateurs :

$$\left\{ \theta_n^* + \frac{1}{\sqrt{n}} I^{-1}(\theta_n^*) \Lambda_n(\theta_n^*) \right\}$$

appartient à la classe $C_{I^{-1}}$, et donc tout élément de la classe C^* sera noté par $\{\hat{\theta}_n\}$.

Théorème 2.2. [Chernoff et Lehmann (1954)]. Sous les conditions de régularité de LeCam :

$$\lim_{n \rightarrow \infty} \mathbf{P} \left(X_n^2(\hat{\theta}_n) > x | H_0 \right) = \mathbf{P} \left(\chi_{r-s-1}^2 + \sum_{i=1}^s \lambda_i(\theta) \xi_i^2 > x \right),$$

où χ_{r-s-1}^2 , $\xi_1, \xi_1, \dots, \xi_s$ sont indépendantes, $\xi_i \sim N(0, 1)$, et $0 < \lambda_i(\theta) < 1$, $i = 1, 2, \dots, s$ sont les racines de l'équation :

$$|(1 - \lambda)I(\theta) - J(\theta)| = 0.$$

2.2.2 La méthode du minimum du Chi-deux

On suppose que $\tilde{\theta}_n$ est un estimateur de θ qui minimise l'équation (2.2), i.e.

$$X_n^2(\tilde{\theta}_n) = \arg \min_{\theta \in \Theta} X_n^2(\theta).$$

D'après les conditions de Cramer (1946) on suppose que :

- 1) $p_i(\theta) > c > 0$, $i = 1, 2, \dots, r$, ($r \geq s + 1$)
- 2) $\frac{\partial p_i^2(\theta)}{\partial \theta_i \partial \theta_j}$ sont des fonction continues,
- 3) rang $(J) = s$.

Selon le théorème de Fisher [Cramer (1946)] on obtient :

Théorème 2.3. Sous les conditions de régularité de Cramer :

$$\lim_{n \rightarrow \infty} \mathbf{P} \left(X_n^2(\tilde{\theta}_n) > x | H_0 \right) = \mathbf{P} \left(\chi_{r-s-1}^2 > x \right), \quad r \geq s + 2.$$

L'estimateur $\tilde{\theta}_n$ de θ peut être déterminé par la résolution du système d'équations :

$$\sum_{i=1}^r \frac{\nu_i}{np_i(\theta)} \frac{\partial p_i(\theta)}{\partial \theta_j} = 0, \quad j = 1, 2, \dots, s, \tag{2.6}$$

et sous H_0 , on a :

si $n \rightarrow \infty$, le vecteur $\sqrt{n}(\tilde{\theta}_n - \theta)$ satisfait la relation asymptotique

$$\sqrt{n}(\tilde{\theta}_n - \theta) = J^{-1}(\theta)B(\theta)^T X_n(\theta) + o(1_s),$$

où $o(1_s)$ est un vecteur aléatoire convergent vers 0_s , par conséquent

$$\mathcal{L}(\sqrt{n}(\tilde{\theta}_n - \theta)) \rightarrow N(0_s, J^{-1}(\theta)), \tag{2.7}$$

i.e. $\tilde{\theta}_n$ est un estimateur \sqrt{n} -consistant de θ .

On note par C_{J-1} le sous-ensemble de C^* la classe de tous les estimateurs $\{\tilde{\theta}_n\}$ vérifiant (2.7). Trouver l'estimateur de Fisher $\tilde{\theta}_n$ de l'équation (2.7) était très difficile, alors Cramer (1946) a proposé d'utiliser un estimateur asymptotiquement équivalent $\tilde{\tilde{\theta}}_n$ qui est l'estimateur du maximum de vraisemblance multinomial qui maximise la fonction de vraisemblance :

$$L_n(\theta) = \frac{n!}{\prod_{i=1}^r \nu_i} \prod_{i=1}^r p_i^{\nu_i}(\theta),$$

qui satisfait l'équation :

$$\text{grad} \ln L_n(\tilde{\tilde{\theta}}_n) = 0, \quad (2.8)$$

et on a $\tilde{\tilde{\theta}}_n \in C_{J-1}$, i.e.

$$\mathcal{L}(\sqrt{n}(\tilde{\tilde{\theta}}_n - \theta)) \rightarrow N(0_s, J^{-1}(\theta)).$$

On peut facilement montrer que la solution de (2.6) est équivalente à celle de (2.8) et donc les estimateurs $\tilde{\theta}_n$ et $\tilde{\tilde{\theta}}_n$ sont équivalents.

Remarque 2.1. Notons que pour construire l'estimateur $\tilde{\theta}_n$, le vecteur de fréquences ν est utilisé, et ainsi les données groupées sont utilisées. En général, le vecteur ν n'est pas une statistique exhaustive pour θ , ce qui entraîne que $\tilde{\theta}_n$ n'est pas nécessairement le meilleur estimateur pour θ .

2.2.3 La méthode des moments

La méthode des moments (MM) est plus particulièrement adaptée pour la famille exponentielle et elle peut être utilisée à la place de la méthode du maximum de vraisemblance lorsque la fonction de vraisemblance est infinie, ou quand le maximum local n'existe pas (voir Voinov et al. (2008)). Cette méthode peut être résumée comme suit :

Soit $g(x) = (g_1(x), g_2(x), \dots, g_s(x))^T$ telle que :

$$m(\theta) = \bar{g},$$

où

$$\bar{g} = (\bar{g}_1, \bar{g}_2, \dots, \bar{g}_s)^T, \quad m(\theta) = (m_1(\theta), m_2(\theta), \dots, m_s(\theta))^T,$$

$$\bar{g}_i = \frac{1}{n} \sum_{i=1}^n g_j(X_i), \quad m_i(\theta) = \int g_i(x) f(x, \theta) dx,$$

et vérifiant $\bar{\theta}_n = m^{-1}(\bar{g})$. En général on prend $g_i(x) = x^i, i = 1, 2, \dots, s$.

Selon les conditions de Hsuan et Robson (1976) on suppose que :

- 1) l'estimateur $\bar{\theta}_n$ de θ est \sqrt{n} -consistant,
- 2) la matrice K (définie par la formule (2.9)) est non singulière,
- 3) $\int g_i(x)f(x, \theta)dx$, $\int g_j(x)\frac{\partial f(x, \theta)}{\partial \theta_j}dx$, $\int g_j(x)\frac{\partial^2 f(x, \theta)}{\partial \theta_j \partial \theta_k}dx$,

existent, finies et continues par rapport à θ pour tout $i, j, k = 1, \dots, s$ au voisinage de la valeur réelle du paramètre θ .

Soient

$$K = K(\theta) = \left[\int g_i(x) \frac{\partial f(x, \theta)}{\partial \theta_j} dx \right]_{s \times s}, \quad (2.9)$$

$$C = C(\theta) = \left[\frac{1}{\sqrt{p_i(\theta)}} \left\{ \int_{I_i} g_j(x) f(x, \theta) dx - p_i(\theta) m_j(\theta) \right\} \right]_{r \times s},$$

$$V = V(\theta) = [m_{ij} - m_i m_j]_{s \times s},$$

où

$$m_{ij} = \mathbf{E}(g_i(X)g_j(X)), \quad i, j = 1, 2, \dots, s,$$

$$A = A(\theta) = I_r - qq^T + C(V - C^T C)^{-1} C^T,$$

et

$$L = L(\theta) = V + (C - BK^{-1}V)^T A (C - BK^{-1}V).$$

Hsuan et Robson (1976) ont démontré que si on remplace θ dans l'équation (2.2) par un estimateur \sqrt{n} -consistant $\bar{\theta}_n$ de θ alors on obtient :

Théorème 2.4. (Hsuan et Robson (1976)) Sous les conditions de régularité :

$$\lim_{n \rightarrow \infty} \mathbf{P} \left(X_n^2(\bar{\theta}_n) > x | H_0 \right) = \mathbf{P} \left(\sum_{i=1}^{r-1} \lambda_i(\theta) \chi_i^2 > x \right),$$

où χ_i^2 sont des variables aléatoires indépendantes obéissant à une loi de Chi-deux à un degré de liberté, et $\lambda_i(\theta)$ sont les vecteurs propres non nuls de la matrice :

$$\Sigma(\theta) = I_r - qq^T + BK^{-1}V(K^{-1})^T B^T - C(K^{-1})^T B^T - BK^{-1}C^T, \quad \text{rang}\Sigma(\theta) = r - 1.$$

2.3 Tests d'ajustement du Chi-deux modifiés

1) En 1973, Nikulin (1973a, 1973b) a proposé une modification du test de Pearson du Chi-deux pour la famille de distributions continues avec des paramètres de décalage et d'échelle "Shift and Scale parameters". De leur côté, Rao et Robson (1974) ont obtenu le même résultat pour les familles exponentielles, et depuis 1998, le test est bien connu sous le nom : test de Nikulin-Rao-Robson (noté NRR)(voir par exemple : van der vaart (1998), Drost (1988)). La statistique de ce test s'écrit :

$$Y_n^2(\hat{\theta}_n) = X_n^T(\hat{\theta}_n)W^-(\hat{\theta}_n)X_n(\hat{\theta}_n), \quad (2.10)$$

où $W^-(\theta)$ est la matrice inverse généralisée de $W(\theta)$, et

$$W(\theta) = I_r - q(\theta)q^T(\theta) - B(\theta)I^{-1}(\theta)B^T(\theta), \quad \text{rang}(W) = r - 1,$$

$$\text{où } q(\theta) = \left(\sqrt{p_1(\theta)}, \sqrt{p_2(\theta)}, \dots, \sqrt{p_r(\theta)} \right)^T.$$

et I_r est la matrice unité d'ordre r .

En utilisant l'équation (2.10), Nikulin (1973) a montré que la matrice inverse généralisée W^- de W peut s'écrire :

$$W^-(\theta) = I_r + B(\theta)(I(\theta) - J(\theta))^{-1}B^T(\theta),$$

et donc l'équation (2.10) devient :

$$Y_n^2(\hat{\theta}_n) = X_n^2(\hat{\theta}_n) + X_n^T(\hat{\theta}_n)B(\hat{\theta}_n)(I(\hat{\theta}_n) - J(\hat{\theta}_n))^{-1}B^T(\hat{\theta}_n)X_n(\hat{\theta}_n).$$

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

$$Y_n^2(\hat{\theta}_n) = X_n^2(\hat{\theta}_n) + Q_n(\hat{\theta}_n) = X_n^2(\hat{\theta}_n) + \frac{1}{n}L^T(\hat{\theta}_n)(I(\hat{\theta}_n) - J(\hat{\theta}_n))^{-1}L(\hat{\theta}_n), \quad (2.11)$$

où

$$L(\theta) = (L_1(\theta), L_2(\theta), \dots, L_s(\theta))^T, \quad L_j(\theta) = \sum_{i=1}^r \frac{\nu_i}{p_i} \frac{\partial p_i(\theta)}{\partial \theta_j}.$$

Le comportement asymptotique de la statistique $Y_n^2(\hat{\theta}_n)$ est donné par le théorème suivant.

Théorème 2.5. [Nikulin (1973a)] pour n suffisamment grand on a :

$$\lim_{n \rightarrow \infty} \mathbf{P} \left(Y_n^2(\hat{\theta}_n) > x | H_0 \right) = \mathbf{P} \left(\chi_{r-1}^2 > x \right).$$

2) En 1974, Dzhaparidze et Nikulin (1974) ont proposé une autre modification du test du Chi-deux de Pearson. Cette Statistique est bien connue sous le nom de "statistique de Dzhaparidze-Nikulin (DN)" :

$$U_n^2(\hat{\theta}_n) = X_n^2(\hat{\theta}_n) - X_n^T(\hat{\theta}_n)B(\hat{\theta}_n)J^{-1}(\hat{\theta}_n)B^T(\hat{\theta}_n)X_n(\hat{\theta}_n). \quad (2.11)$$

Le comportement asymptotique de la statistique $U_n^2(\hat{\theta}_n)$ est donné par ce théorème.

Théorème 2.6. [Dzhaparidze et Nikulin (1974)] sous les conditions de régularité :

$$\lim_{n \rightarrow \infty} \mathbf{P} \left(U_n^2(\hat{\theta}_n) > x | H_0 \right) = \mathbf{P} \left(\chi_{r-s-1}^2 > x \right), \quad r > s + 2.$$

3) Une décomposition très importante de (2.10) a été établie par McCulloch (1985) en utilisant la formule

$$(I_r - BI^{-1}B^T)^{-1} = I_r - B(B^TB)^{-1}B^T + B((I - B^TB)^{-1} + (B^TB)^{-1})B^T.$$

McCulloch a présenté (2.10) comme suit :

$$Y_n^2(\hat{\theta}_n) = U_n^2(\hat{\theta}_n) + M_n^2(\hat{\theta}_n),$$

d'où

$$M_n^2(\hat{\theta}_n) = Y_n^2(\hat{\theta}_n) - U_n^2(\hat{\theta}_n), \quad (2.12)$$

avec $U_n^2(\hat{\theta}_n)$ est la statistique DN (2.11).

L'équation (2.11) peut s'écrire :

$$M_n^2(\hat{\theta}_n) = X_n^T(\hat{\theta}_n)B(\hat{\theta}_n)(I(\hat{\theta}_n) - J(\hat{\theta}_n))^{-1}B^T(\hat{\theta}_n)X_n(\hat{\theta}_n) + X_n^T(\hat{\theta}_n)B(\hat{\theta}_n)J^{-1}(\hat{\theta}_n)B^T(\hat{\theta}_n)X_n(\hat{\theta}_n).$$

La statistique $M_n^2(\hat{\theta}_n)$ est appelée la statistique de McCulloch. McCulloch (1985) montre que les statistiques $U_n^2(\hat{\theta}_n)$ et $M_n^2(\hat{\theta}_n)$ sont asymptotiquement indépendantes et que la statistique $M_n^2(\hat{\theta}_n)$ peut être utilisée comme un test d'ajustement indépendamment et on a :

Théorème 2.7. Sous les conditions de régularité de Moore et Spruill (1975) :

$$\lim_{n \rightarrow \infty} \mathbf{P} \left(M_n^2(\hat{\theta}_n) > x | H_0 \right) = \mathbf{P}(\chi_s^2 > x).$$

4) En 1976, Hsuan et Robson (1976) ont proposé une modification Y_n^2 du test standard du Chi-deux de Pearson (2.2) en remplaçant θ par $\bar{\theta}_n$ l'estimateur obtenu par la méthode des moments, mais ils n'ont pas donné la forme explicite du test modifié. D'autre part, Mirvaliev (2001) a obtenu la forme explicite du test modifié. Le test est bien connu sous le nom Hsuan-Robson-Mirvaliev (HRM), et la statistique Y_n^2 s'écrit :

$$Y_n^2(\bar{\theta}_n) = X_n^2(\bar{\theta}_n) + R_n^2(\bar{\theta}_n) - Q_n^2(\bar{\theta}_n),$$

où

$$R_n^2(\bar{\theta}_n) = X_n^T(\bar{\theta}_n) C_n(\bar{\theta}_n) (V_n(\bar{\theta}_n) - C_n^T(\bar{\theta}_n) C_n(\bar{\theta}_n))^{-1} C_n^T(\bar{\theta}_n) X_n(\bar{\theta}_n),$$

et

$$Q_n^2(\bar{\theta}_n) = X_n^T(\bar{\theta}_n) A_n(\bar{\theta}_n) \left[C_n(\bar{\theta}_n) - B_n(\bar{\theta}_n) K_n^{-1}(\bar{\theta}_n) X_n(\bar{\theta}_n) \right] \times \\ L_n^{-1}(\bar{\theta}_n) \left[C_n(\bar{\theta}_n) - B_n(\bar{\theta}_n) K_n^{-1}(\bar{\theta}_n) X_n(\bar{\theta}_n) \right]^T A_n(\bar{\theta}_n) X_n(\bar{\theta}_n).$$

Le comportement asymptotique de la statistique $Y_n^2(\bar{\theta}_n)$ est donné par le théorème suivant.

Théorème 2.8. [Hsuan et Robson (1976)] Sous les conditions de régularité :

$$\lim_{n \rightarrow \infty} \mathbf{P} \left(Y_n^2(\bar{\theta}_n) > x | H_0 \right) = \mathbf{P} \left(\chi_{r-1}^2 > x \right).$$

Remarque 2.2. On peut montrer que pour la famille exponentielle, les deux estimateurs $\hat{\theta}_n$ et $\bar{\theta}_n$ sont identiques.

Nous allons maintenant construire les tests $Y_n^2(\hat{\theta}_n)$, $U_n^2(\hat{\theta}_n)$ et $M_n^2(\hat{\theta}_n)$ pour les distributions LN , LL , EW , PGW , BS et IG

3 Tests d'ajustement pour la loi log-normale

Nous voulons tester l'hypothèse composite H_0 selon laquelle la distribution de l'échantillon $\mathbb{X} = (X_1, X_2, \dots, X_n)^T$ vérifie :

$$H_0 : P(X_i \leq x) = F_{LN}(x, \theta_{LN}), \quad x \geq 0, \quad \theta_{LN} = (m, \sigma)^T,$$

où $F_{LN}(x, \theta_{LN})$ est la fonction de répartition de la loi LN définie par l'équation (1.2).

La fonction de log-vraisemblance est :

$$\ell(\theta_{LN}) = -n \ln \sigma - \sum_{i=1}^n \ln X_i - \frac{1}{2\sigma^2} \sum_{i=1}^n (\ln X_i - m)^2,$$

et l'estimateur du MV $\hat{\theta}_{LN}$ dans ce cas est donné par

$$\hat{m} = \frac{1}{n} \sum_{i=1}^n \ln X_i \quad \text{et} \quad \hat{\sigma} = \frac{(\ln X_i - \hat{m})^2}{n}$$

Pour tester l'hypothèse H_0 , nous pouvons utiliser les 3 tests statistiques $Y_n^2(\hat{\theta}_{LN})$, $U_n^2(\hat{\theta}_{LN})$ et $M_n^2(\hat{\theta}_{LN})$,

avec

$$J(\theta_{LN}) = B^T(\theta_{LN})B(\theta_{LN}), \quad B(\theta_{LN}) = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ \vdots & \vdots \\ b_{r1} & b_{r2} \end{pmatrix},$$

$$b_{i1}(\theta_{LN}) = \frac{1}{\sqrt{p_i(\theta_{LN})}} \frac{\partial p_i(\theta_{LN})}{\partial m}, \quad b_{i2}(\theta_{LN}) = \frac{1}{\sqrt{p_i(\theta_{LN})}} \frac{\partial p_i(\theta_{LN})}{\partial \sigma}, \quad i = 1, 2, \dots, r,$$

avec

$$p_i(\theta_{LN}) = \Phi\left(\frac{\ln a_i - m}{\sigma}\right) - \Phi\left(\frac{\ln a_{i-1} - m}{\sigma}\right),$$

$$\frac{\partial p_i(\theta_{LN})}{\partial m} = -\frac{1}{\sigma} \varphi\left(\frac{\ln a_i - m}{\sigma}\right) + \frac{1}{\sigma} \varphi\left(\frac{\ln a_{i-1} - m}{\sigma}\right),$$

$$\frac{\partial p_i(\theta_{LN})}{\partial \sigma} = -\frac{\ln a_i - m}{\sigma^2} \varphi\left(\frac{\ln a_i - m}{\sigma}\right) + \frac{\ln a_{i-1} - m}{\sigma^2} \varphi\left(\frac{\ln a_{i-1} - m}{\sigma}\right),$$

où φ est la densité de la loi normale standard $N(0, 1)$.

La matrice d'information de Fisher $I(\theta_{LN})$ de X_i est donnée par :

$$I(\theta_{LN}) = \begin{pmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{2}{\sigma^2} \end{pmatrix}.$$

Sous H_0 , et pour n suffisamment grand :

- • La statistique $Y_n^2(\hat{\theta}_{LN})$ suit la distribution χ_{r-1}^2 à $r - 1$ degrés de liberté.
- • La statistique $U_n^2(\hat{\theta}_{LN})$ suit la distribution χ_{r-3}^2 à $r - 3$ degrés de liberté.
- • La statistique $M_n^2(\hat{\theta}_{LN})$ suit la distribution χ_2^2 à 2 degrés de liberté.

Pour un seuil α , H_0 est acceptée si :

- • $Y_n^2(\hat{\theta}_{LN}) \leq C_\alpha$, $C_\alpha = \chi_{r-1, 1-\alpha}^2$, sinon H_0 est refusée.
- • $U_n^2(\hat{\theta}_{LN}) \leq C_\alpha$, $C_\alpha = \chi_{r-3, 1-\alpha}^2$, sinon H_0 est refusée.
- • $M_n^2(\hat{\theta}_{LN}) \leq C_\alpha$, $C_\alpha = \chi_{2, 1-\alpha}^2$, sinon H_0 est refusée.

4 Tests d'ajustement pour la loi log-logistique

Nous voulons tester l'hypothèse composite H_0 selon laquelle la distribution de l'échantillon $\mathbb{X} = (X_1, X_2, \dots, X_n)^T$ vérifie :

$$H_0 : P(X_i \leq x) = F_{LL}(x, \theta_{LL}), \quad x \geq 0, \quad \theta_{LL} = (\sigma, \nu)^T,$$

où $F_{LL}(x, \theta_{LL})$ est la fonction de répartition de la loi LL donnée par l'équation (1.4).

La fonction de log-vraisemblance est :

$$\ell(\theta_{LL}) = (\ln \nu - \nu \ln \sigma) n + (\nu - 1) \sum_{i=1}^n \ln X_i - 2 \sum_{i=1}^n \ln \left(1 + \left(\frac{X_i}{\sigma} \right)^\nu \right),$$

ainsi les fonctions de scores sont

$$\frac{\partial \ell(\theta_{LL})}{\partial \sigma} = \frac{-n\nu}{\sigma} + \frac{2\nu}{\sigma} \sum_{i=1}^n \frac{K_i}{1 + K_i},$$

$$\frac{\partial \ell(\theta_{LL})}{\partial \nu} = \left(\frac{1}{\nu} - \ln \sigma \right) n + \sum_{i=1}^n \ln X_i - 2 \sum_{i=1}^n \frac{\ln \frac{X_i}{\sigma} K_i}{1 + K_i}.$$

Les dérivées secondes de $\ell(\theta_{LL})$ sont données par

$$\frac{\partial^2 \ell(\theta_{LL})}{\partial \sigma^2} = -\frac{n\nu}{\sigma^2} - \frac{2\nu}{\sigma^2} \sum_{i=1}^n \frac{K_i}{1 + K_i} - \frac{2\nu^2}{\sigma^2} \sum_{i=1}^n \frac{K_i}{(1 + K_i)^2},$$

$$\frac{\partial^2 \ell(\theta_{LL})}{\partial \sigma \partial \nu} = -\frac{n}{\sigma} + \frac{2}{\sigma} \sum_{i=1}^n \frac{K_i}{1 + K_i} + \frac{2\nu}{\sigma} \sum_{i=1}^n \frac{K_i \ln \frac{X_i}{\sigma}}{(1 + K_i)^2},$$

$$\frac{\partial^2 \ell(\theta_{LL})}{\partial \nu^2} = -\frac{n}{\nu^2} - 2 \sum_{i=1}^n \left(\ln \frac{X_i}{\sigma} \right)^2 \frac{K_i}{(1 + K_i)^2},$$

avec $K_i = \left(\frac{X_i}{\sigma} \right)^\nu$.

Pour tester l'hypothèse H_0 , nous pouvons utiliser les 3 tests statistiques $Y_n^2(\hat{\theta}_{LL})$, $U_n^2(\hat{\theta}_{LL})$ et $M_n^2(\hat{\theta}_{LL})$,

avec

$$J(\theta_{LL}) = B^T(\theta_{LL})B(\theta_{LL}), \quad B(\theta_{LL}) = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ \vdots & \vdots \\ b_{r1} & b_{r2} \end{pmatrix},$$

$$b_{i1}(\theta_{LL}) = \frac{1}{\sqrt{p_i(\theta_{LL})}} \frac{\partial p_i(\theta_{LL})}{\partial \sigma}, \quad b_{i2}(\theta_{LL}) = \frac{1}{\sqrt{p_i(\theta_{LL})}} \frac{\partial p_i(\theta_{LL})}{\partial \nu}, \quad i = 1, 2, \dots, r,$$

avec

$$p_i(\theta_{LL}) = \frac{1}{1 + \left(\frac{a_{i-1}}{\sigma} \right)^\nu} - \frac{1}{1 + \left(\frac{a_i}{\sigma} \right)^\nu},$$

$$\frac{\partial p_i(\theta_{LL})}{\partial \sigma} = \frac{\nu}{\sigma} \frac{\left(\frac{a_{i-1}}{\sigma}\right)^\nu}{\left(1 + \left(\frac{a_{i-1}}{\sigma}\right)^\nu\right)^2} - \frac{\nu}{\sigma} \frac{\left(\frac{a_i}{\sigma}\right)^\nu}{\left(1 + \left(\frac{a_i}{\sigma}\right)^\nu\right)^2},$$

$$\frac{\partial p_i(\theta_{LL})}{\partial \nu} = -\frac{\left(\frac{a_{i-1}}{\sigma}\right)^\nu \ln\left(\frac{a_{i-1}}{\sigma}\right)}{\left(1 + \left(\frac{a_{i-1}}{\sigma}\right)^\nu\right)^2} + \frac{\left(\frac{a_i}{\sigma}\right)^\nu \ln\left(\frac{a_i}{\sigma}\right)}{\left(1 + \left(\frac{a_i}{\sigma}\right)^\nu\right)^2},$$

et la matrice d'information de Fisher $I(\theta_{LL})$ peut être remplacée par

$$\hat{I}(\hat{\theta}_{LL}) = -\frac{1}{n} \begin{pmatrix} \frac{\partial^2 \ell(\hat{\theta}_{LL})}{\partial \sigma^2} & \frac{\partial^2 \ell(\hat{\theta}_{LL})}{\partial \sigma \partial \nu} \\ \frac{\partial^2 \ell(\hat{\theta}_{LL})}{\partial \sigma \partial \nu} & \frac{\partial^2 \ell(\hat{\theta}_{LL})}{\partial \nu^2} \end{pmatrix}.$$

Sous H_0 , et pour n suffisamment grand :

- • La statistique $Y_n^2(\hat{\theta}_{LL})$ suit la distribution χ_{r-1}^2 à $r - 1$ degrés de liberté.
- • La statistique $U_n^2(\hat{\theta}_{LL})$ suit la distribution χ_{r-3}^2 à $r - 3$ degrés de liberté.
- • La statistique $M_n^2(\hat{\theta}_{LL})$ suit la distribution χ_2^2 à 2 degrés de liberté.

Pour un seuil α , H_0 est acceptée si :

- • $Y_n^2(\hat{\theta}_{LL}) \leq C_\alpha$, $C_\alpha = \chi_{r-1,1-\alpha}^2$, sinon H_0 est refusée.
- • $U_n^2(\hat{\theta}_{LL}) \leq C_\alpha$, $C_\alpha = \chi_{r-3,1-\alpha}^2$, sinon H_0 est refusée.
- • $M_n^2(\hat{\theta}_{LL}) \leq C_\alpha$, $C_\alpha = \chi_{2,1-\alpha}^2$, sinon H_0 est refusée.

5 Tests d'ajustement pour la loi de Weibull exponentielle

Nous voulons tester l'hypothèse composite H_0 selon laquelle la distribution de l'échantillon $\mathbb{X} = (X_1, X_2, \dots, X_n)^T$ vérifie :

$$H_0 : P(X_i \leq x) = F_{EW}(x, \theta_{EW}), \quad x \geq 0, \quad \theta_{EW} = (\sigma, \nu, \gamma)^T,$$

où $F_{EW}(x, \theta_{EW})$ est la fonction de répartition de la loi EW donnée par l'équation (1.8).

La fonction de log-vraisemblance est :

$$\ell(\theta_{EW}) = (\ln \nu - \nu \ln \sigma - \ln \gamma) n + (\nu - 1) \sum_{i=1}^n \ln X_i - \sum_{i=1}^n \left(\frac{X_i}{\sigma} \right)^\nu + \left(\frac{1}{\gamma} - 1 \right) \sum_{i=1}^n \ln \left(1 - e^{-\left(\frac{X_i}{\sigma} \right)^\nu} \right),$$

ainsi les fonctions de scores sont

$$\frac{\partial \ell(\theta_{EW})}{\partial \sigma} = -\frac{n\nu}{\sigma} + \frac{\nu}{\sigma} \sum_{i=1}^n K_i - \frac{\nu}{\sigma} \left(\frac{1}{\gamma} - 1 \right) \sum_{i=1}^n \frac{K_i e^{-K_i}}{1 - e^{-K_i}},$$

$$\frac{\partial \ell(\theta_{EW})}{\partial \nu} = \left(\frac{1}{\nu} - \ln \sigma \right) n + \sum_{i=1}^n \ln X_i - \frac{1}{\nu} \sum_{i=1}^n K_i \ln(K_i) + \frac{1}{\nu} \left(\frac{1}{\gamma} - 1 \right) \sum_{i=1}^n \frac{K_i e^{-K_i} \ln(K_i)}{1 - e^{-K_i}},$$

$$\frac{\partial \ell(\theta_{EW})}{\partial \gamma} = -\frac{n}{\gamma} - \frac{1}{\gamma^2} \sum_{i=1}^n \ln \left(1 - e^{-K_i} \right).$$

Posons $K_i = \left(\frac{X_i}{\sigma} \right)^\nu$. Les dérivées partielles secondes de $\ell(\theta_{EW})$ sont

$$\frac{\partial^2 \ell(\theta_{EW})}{\partial \sigma^2} = \frac{n\nu}{\sigma^2} - \frac{\nu(\nu+1)}{\sigma^2} \sum_{i=1}^n K_i + \frac{\nu}{\sigma^2} \left(\frac{1}{\gamma} - 1 \right) \sum_{i=1}^n \frac{K_i e^{-K_i}}{1 - e^{-K_i}} -$$

$$\frac{\nu^2}{\sigma^2} \left(\frac{1}{\gamma} - 1 \right) \sum_{i=1}^n \frac{K_i e^{-K_i} (K_i + e^{-K_i} - 1)}{(1 - e^{-K_i})^2},$$

$$\frac{\partial^2 \ell(\theta_{EW})}{\partial \sigma \partial \nu} = -\frac{n}{\sigma} + \frac{1}{\sigma} \sum_{i=1}^n K_i (1 + \ln K_i) + \frac{1}{\sigma} \left(\frac{1}{\gamma} - 1 \right) \sum_{i=1}^n \frac{K_i e^{-K_i}}{1 - e^{-K_i}} \left\{ -1 + \ln K_i \left(\frac{K_i}{1 - e^{-K_i}} - 1 \right) \right\},$$

$$\frac{\partial^2 \ell(\theta_{EW})}{\partial \sigma \partial \gamma} = \frac{\nu}{\sigma \gamma^2} \sum_{i=1}^n \frac{K_i e^{-K_i}}{1 - e^{-K_i}},$$

$$\frac{\partial^2 \ell(\theta_{EW})}{\partial \nu^2} = -\frac{n}{\nu^2} - \frac{1}{\nu^2} \sum_{i=1}^n K_i (\ln K_i)^2 + \frac{1}{\nu^2} \left(\frac{1}{\gamma} - 1 \right) \sum_{i=1}^n \frac{K_i (\ln K_i)^2 e^{-K_i} (1 - e^{-K_i} - K_i)}{(1 - e^{-K_i})^2},$$

$$\frac{\partial^2 \ell(\theta_{EW})}{\partial \nu \partial \gamma} = -\frac{1}{\nu \gamma^2} \sum_{i=1}^n \frac{K_i e^{-K_i} \ln K_i}{1 - e^{-K_i}},$$

$$\frac{\partial^2 \ell(\theta_{EW})}{\partial \gamma^2} = \frac{n}{\gamma^2} + \frac{2}{\gamma^3} \sum_{i=1}^n \ln(1 - e^{-K_i}),$$

Pour tester l'hypothèse H_0 , nous pouvons utiliser les 3 tests statistiques $Y_n^2(\hat{\theta}_{EW})$, $U_n^2(\hat{\theta}_{EW})$ et $M_n^2(\hat{\theta}_{EW})$,

avec

$$J(\theta_{EW}) = B^T(\theta_{EW})B(\theta_{EW}), \quad B(\theta_{EW}) = \begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ \vdots & \vdots & \vdots \\ b_{r1} & b_{r2} & b_{r3} \end{pmatrix},$$

où

$$b_{i1}(\theta_{EW}) = \frac{1}{\sqrt{p_i(\theta_{EW})}} \frac{\partial p_i(\theta_{EW})}{\partial \sigma}, \quad b_{i2}(\theta_{EW}) = \frac{1}{\sqrt{p_i(\theta_{EW})}} \frac{\partial p_i(\theta_{EW})}{\partial \nu},$$

et

$$b_{i3}(\theta_{EW}) = \frac{1}{\sqrt{p_i(\theta_{EW})}} \frac{\partial p_i(\theta_{EW})}{\partial \gamma}, \quad i = 1, 2, \dots, r,$$

avec

$$p_i(\theta_{EW}) = \left\{ 1 - e^{-\left(\frac{a_i}{\sigma}\right)^\nu} \right\}^{\frac{1}{\gamma}} - \left\{ 1 - e^{-\left(\frac{a_{i-1}}{\sigma}\right)^\nu} \right\}^{\frac{1}{\gamma}},$$

$$\frac{\partial p_i(\theta_{EW})}{\partial \sigma} = -\frac{\nu}{\sigma \gamma} \left(\frac{a_i}{\sigma}\right)^\nu e^{-\left(\frac{a_i}{\sigma}\right)^\nu} \left\{ 1 - e^{-\left(\frac{a_i}{\sigma}\right)^\nu} \right\}^{\frac{1}{\gamma}-1} + \frac{\nu}{\sigma \gamma} \left(\frac{a_{i-1}}{\sigma}\right)^\nu e^{-\left(\frac{a_{i-1}}{\sigma}\right)^\nu} \left\{ 1 - e^{-\left(\frac{a_{i-1}}{\sigma}\right)^\nu} \right\}^{\frac{1}{\gamma}-1},$$

$$\frac{\partial p_i(\theta_{EW})}{\partial \nu} = \frac{1}{\gamma} \left(\frac{a_i}{\sigma}\right)^\nu \ln\left(\frac{a_i}{\sigma}\right) e^{-\left(\frac{a_i}{\sigma}\right)^\nu} \left\{ 1 - e^{-\left(\frac{a_i}{\sigma}\right)^\nu} \right\}^{\frac{1}{\gamma}-1} -$$

$$\frac{1}{\gamma} \left(\frac{a_{i-1}}{\sigma}\right)^\nu \ln\left(\frac{a_{i-1}}{\sigma}\right) e^{-\left(\frac{a_{i-1}}{\sigma}\right)^\nu} \left\{ 1 - e^{-\left(\frac{a_{i-1}}{\sigma}\right)^\nu} \right\}^{\frac{1}{\gamma}-1},$$

$$\frac{\partial p_i(\theta_{EW})}{\partial \gamma} = -\frac{1}{\gamma^2} \ln\left(1 - e^{-\left(\frac{a_i}{\sigma}\right)^\nu}\right) \left\{ 1 - e^{-\left(\frac{a_i}{\sigma}\right)^\nu} \right\}^{\frac{1}{\gamma}} + \frac{1}{\gamma^2} \ln\left(1 - e^{-\left(\frac{a_{i-1}}{\sigma}\right)^\nu}\right) \left\{ 1 - e^{-\left(\frac{a_{i-1}}{\sigma}\right)^\nu} \right\}^{\frac{1}{\gamma}},$$

et la matrice d'information de Fisher $I(\theta_{EW})$ peut être remplacée par

$$\hat{I}(\hat{\theta}_{EW}) = -\frac{1}{n} \begin{pmatrix} \frac{\partial^2 \ell(\hat{\theta}_{EW})}{\partial \sigma^2} & \frac{\partial^2 \ell(\hat{\theta}_{EW})}{\partial \sigma \partial \nu} & \frac{\partial^2 \ell(\hat{\theta}_{EW})}{\partial \sigma \partial \gamma} \\ \frac{\partial^2 \ell(\hat{\theta}_{EW})}{\partial \sigma \partial \nu} & \frac{\partial^2 \ell(\hat{\theta}_{EW})}{\partial \nu^2} & \frac{\partial^2 \ell(\hat{\theta}_{EW})}{\partial \nu \partial \gamma} \\ \frac{\partial^2 \ell(\hat{\theta}_{EW})}{\partial \sigma \partial \gamma} & \frac{\partial^2 \ell(\hat{\theta}_{EW})}{\partial \nu \partial \gamma} & \frac{\partial^2 \ell(\hat{\theta}_{EW})}{\partial \gamma^2} \end{pmatrix}.$$

Sous H_0 , et pour n suffisamment grand :

- • La statistique $Y_n^2(\hat{\theta}_{EW})$ suit la distribution χ_{r-1}^2 à $r - 1$ degrés de liberté.
- • La statistique $U_n^2(\hat{\theta}_{EW})$ suit la distribution χ_{r-4}^2 à $r - 4$ degrés de liberté.
- • La statistique $M_n^2(\hat{\theta}_{EW})$ suit la distribution χ_3^2 à 3 degrés de liberté.

Pour un seuil α , H_0 est acceptée si :

- • $Y_n^2(\hat{\theta}_{EW}) \leq C_\alpha$, $C_\alpha = \chi_{r-1, 1-\alpha}^2$, sinon H_0 est refusée.
- • $U_n^2(\hat{\theta}_{EW}) \leq C_\alpha$, $C_\alpha = \chi_{r-4, 1-\alpha}^2$, sinon H_0 est refusée.
- • $M_n^2(\hat{\theta}_{EW}) \leq C_\alpha$, $C_\alpha = \chi_{3, 1-\alpha}^2$, sinon H_0 est refusée.

6 Tests d'ajustement pour la loi de Weibull généralisée

Le test d'ajustement du Chi-deux pour cette distribution basé sur la statistique NRR est étudié par Haghghi (2004).

Considérons l'hypothèse composite H_0 selon laquelle la distribution de l'échantillon $\mathbb{X} = (X_1, X_2, \dots, X_n)^T$ vérifie :

$$H_0 : P(X_i \leq x) = F_{GW}(x, \theta_{GW}), \quad x \geq 0, \quad \theta_{GW} = (\sigma, \nu, \gamma)^T,$$

où $F_{GW}(x, \theta_{GW})$ est la fonction de répartition de la loi GW donnée par l'équation (1.10).

La fonction de log-vraisemblance est :

$$\ell(\theta_{GW}) = (\ln \nu - \nu \ln \sigma - \ln \gamma) n + (\nu - 1) \sum_{i=1}^n \ln X_i + \left(\frac{1}{\gamma} - 1\right) \sum_{i=1}^n \ln \left(1 + \left(\frac{x}{\sigma}\right)^\nu\right) - \sum_{i=1}^n \left(1 + \left(\frac{x}{\sigma}\right)^\nu\right)^{\frac{1}{\gamma}},$$

en posant $K_i = \left(\frac{X_i}{\sigma}\right)^\nu$, les fonctions de scores sont

$$\frac{\partial \ell(\theta_{GW})}{\partial \sigma} = -\frac{n\nu}{\sigma} - \frac{\nu}{\sigma} \left(\frac{1}{\gamma} - 1 \right) \sum_{i=1}^n \frac{K_i}{1+K_i} + \frac{\nu}{\sigma\gamma} \sum_{i=1}^n K_i (1+K_i)^{\frac{1}{\gamma}-1},$$

$$\frac{\partial \ell(\theta_{GW})}{\partial \nu} = \left(\frac{1}{\nu} - \ln \sigma \right) n + \sum_{i=1}^n \ln X_i + \frac{1}{\nu} \left(\frac{1}{\gamma} - 1 \right) \sum_{i=1}^n \frac{K_i \ln(K_i)}{1+K_i} - \frac{1}{\nu\gamma} \sum_{i=1}^n K_i (1+K_i)^{\frac{1}{\gamma}-1} \ln(K_i),$$

$$\frac{\partial \ell(\theta_{GW})}{\partial \gamma} = -\frac{n}{\gamma} - \frac{1}{\gamma^2} \sum_{i=1}^n \ln(1+K_i) + \frac{1}{\gamma^2} \sum_{i=1}^n (1+K_i)^{\frac{1}{\gamma}} \ln(1+K_i).$$

Les dérivées partielles secondes de $\ell(\theta_{GW})$ sont

$$\frac{\partial^2 \ell(\theta_{GW})}{\partial \sigma^2} = \frac{n\nu}{\sigma^2} + \frac{\nu}{\sigma^2} \left(\frac{1}{\gamma} - 1 \right) \sum_{i=1}^n \frac{K_i}{1+K_i} + \frac{\nu^2}{\sigma^2} \left(\frac{1}{\gamma} - 1 \right) \sum_{i=1}^n \frac{K_i}{(1+K_i)^2} -$$

$$\frac{\nu}{\sigma^2\gamma} \sum_{i=1}^n K_i (1+K_i)^{\frac{1}{\gamma}-2} \left(1 + \nu + K_i + \frac{\nu}{\gamma} K_i \right),$$

$$\frac{\partial^2 \ell(\theta_{GW})}{\partial \sigma \partial \nu} = -\frac{n}{\sigma} - \frac{1}{\sigma} \left(\frac{1}{\gamma} - 1 \right) \sum_{i=1}^n \frac{K_i(1+K_i + \ln K_i)}{(1+K_i)^2} +$$

$$\frac{1}{\sigma\gamma} \sum_{i=1}^n K_i (1+K_i)^{\frac{1}{\gamma}-2} \left(1 + K_i + \ln K_i + \frac{1}{\gamma} K_i \ln K_i \right),$$

$$\frac{\partial^2 \ell(\theta_{GW})}{\partial \sigma \partial \gamma} = \frac{\nu}{\sigma\gamma^2} \sum_{i=1}^n \frac{K_i}{1+K_i} - \frac{\nu}{\sigma\gamma^2} \sum_{i=1}^n K_i (1+K_i)^{\frac{1}{\gamma}-1} \left(1 + \frac{1}{\gamma} \ln(1+K_i) \right),$$

$$\frac{\partial^2 \ell(\theta_{GW})}{\partial \nu^2} = -\frac{n}{\nu^2} + \frac{1}{\nu^2} \left(\frac{1}{\gamma} - 1 \right) \sum_{i=1}^n \frac{K_i (\ln K_i)^2}{(1+K_i)^2} - \frac{1}{\nu^2\gamma} \sum_{i=1}^n K_i (\ln K_i)^2 (1+K_i)^{\frac{1}{\gamma}-2} \left(1 + \frac{1}{\gamma} K_i \right),$$

$$\frac{\partial^2 \ell(\theta_{GW})}{\partial \nu \partial \gamma} = -\frac{1}{\nu\gamma^2} \sum_{i=1}^n \frac{K_i \ln K_i}{1+K_i} + \frac{1}{\nu\gamma^2} \sum_{i=1}^n K_i \ln K_i (1+K_i)^{\frac{1}{\gamma}-1} \left(1 + \frac{1}{\gamma} \ln(1+K_i) \right),$$

$$\frac{\partial^2 \ell(\theta_{GW})}{\partial \gamma^2} = \frac{n}{\gamma^2} + \frac{2}{\gamma^3} \sum_{i=1}^n \ln(1 + K_i) - \frac{1}{\gamma^3} \sum_{i=1}^n (1 + K_i)^{\frac{1}{\gamma}} \ln(1 + K_i) \left(2 + \frac{1}{\gamma} \ln(1 + K_i) \right).$$

Pour tester l'hypothèse H_0 , nous pouvons utiliser les 3 tests statistiques $Y_n^2(\hat{\theta}_{GW})$, $U_n^2(\hat{\theta}_{GW})$ et $M_n^2(\hat{\theta}_{GW})$,
avec

$$J(\theta_{GW}) = B^T(\theta_{GW})B(\theta_{GW}), \quad B(\theta_{GW}) = \begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ \vdots & \vdots & \vdots \\ b_{r1} & b_{r2} & b_{r3} \end{pmatrix},$$

où

$$b_{i1}(\theta_{GW}) = \frac{1}{\sqrt{p_i(\theta_{GW})}} \frac{\partial p_i(\theta_{GW})}{\partial \sigma}, \quad b_{i2}(\theta_{GW}) = \frac{1}{\sqrt{p_i(\theta_{GW})}} \frac{\partial p_i(\theta_{GW})}{\partial \nu},$$

et

$$b_{i3}(\theta_{GW}) = \frac{1}{\sqrt{p_i(\theta_{GW})}} \frac{\partial p_i(\theta_{GW})}{\partial \gamma}, \quad i = 1, 2, \dots, r,$$

avec

$$p_i(\theta_{GW}) = e^{1-(1+(\frac{a_{i-1}}{\sigma})^\nu)^{\frac{1}{\gamma}}} - e^{1-(1+(\frac{a_i}{\sigma})^\nu)^{\frac{1}{\gamma}}},$$

$$\frac{\partial p_i(\theta_{GW})}{\partial \sigma} = \frac{\nu}{\sigma^{\nu+1}\gamma} \left\{ a_{i-1}^\nu \left(1 + \left(\frac{a_{i-1}}{\sigma} \right)^\nu \right)^{\frac{1}{\gamma}-1} e^{1-(1+(\frac{a_{i-1}}{\sigma})^\nu)^{\frac{1}{\gamma}}} - a_i^\nu \left(1 + \left(\frac{a_i}{\sigma} \right)^\nu \right)^{\frac{1}{\gamma}-1} e^{1-(1+(\frac{a_i}{\sigma})^\nu)^{\frac{1}{\gamma}}} \right\},$$

$$\frac{\partial p_i(\theta_{GW})}{\partial \nu} = -\frac{1}{\gamma} \left(\frac{a_{i-1}}{\sigma} \right)^\nu \ln \left(\frac{a_{i-1}}{\sigma} \right) \left(1 + \left(\frac{a_{i-1}}{\sigma} \right)^\nu \right)^{\frac{1}{\gamma}-1} e^{1-(1+(\frac{a_{i-1}}{\sigma})^\nu)^{\frac{1}{\gamma}}} +$$

$$\frac{1}{\gamma} \left(\frac{a_i}{\sigma} \right)^\nu \ln \left(\frac{a_i}{\sigma} \right) \left(1 + \left(\frac{a_i}{\sigma} \right)^\nu \right)^{\frac{1}{\gamma}-1} e^{1-(1+(\frac{a_i}{\sigma})^\nu)^{\frac{1}{\gamma}}},$$

$$\begin{aligned} \frac{\partial p_i(\theta_{GW})}{\partial \gamma} &= \frac{1}{\gamma^2} \ln \left(1 + \left(\frac{a_{i-1}}{\sigma} \right)^\nu \right) e^{1 - \left(1 + \left(\frac{a_{i-1}}{\sigma} \right)^\nu \right)^{\frac{1}{\gamma}}} - \\ &\quad \frac{1}{\gamma^2} \ln \left(1 + \left(\frac{a_i}{\sigma} \right)^\nu \right) e^{1 - \left(1 + \left(\frac{a_i}{\sigma} \right)^\nu \right)^{\frac{1}{\gamma}}}, \end{aligned}$$

et la matrice d'information de Fisher $I(\theta_{GW})$ peut être remplacée par

$$\hat{I}(\hat{\theta}_{GW}) = -\frac{1}{n} \begin{pmatrix} \frac{\partial^2 \ell(\hat{\theta}_{GW})}{\partial \sigma^2} & \frac{\partial^2 \ell(\hat{\theta}_{GW})}{\partial \sigma \partial \nu} & \frac{\partial^2 \ell(\hat{\theta}_{GW})}{\partial \sigma \partial \gamma} \\ \frac{\partial^2 \ell(\hat{\theta}_{GW})}{\partial \sigma \partial \nu} & \frac{\partial^2 \ell(\hat{\theta}_{GW})}{\partial \nu^2} & \frac{\partial^2 \ell(\hat{\theta}_{GW})}{\partial \nu \partial \gamma} \\ \frac{\partial^2 \ell(\hat{\theta}_{GW})}{\partial \sigma \partial \gamma} & \frac{\partial^2 \ell(\hat{\theta}_{GW})}{\partial \nu \partial \gamma} & \frac{\partial^2 \ell(\hat{\theta}_{GW})}{\partial \gamma^2} \end{pmatrix}.$$

Sous H_0 , et pour n suffisamment grand :

- • La statistique $Y_n^2(\hat{\theta}_{GW})$ suit la distribution χ_{r-1}^2 à $r - 1$ degrés de liberté.
- • La statistique $U_n^2(\hat{\theta}_{GW})$ suit la distribution χ_{r-4}^2 à $r - 4$ degrés de liberté.
- • La statistique $M_n^2(\hat{\theta}_{GW})$ suit la distribution χ_3^2 à 3 degrés de liberté.

Pour un seuil α , H_0 est acceptée si :

- • $Y_n^2(\hat{\theta}_{GW}) \leq C_\alpha$, $C_\alpha = \chi_{r-1, 1-\alpha}^2$, sinon H_0 est refusée.
- • $U_n^2(\hat{\theta}_{GW}) \leq C_\alpha$, $C_\alpha = \chi_{r-4, 1-\alpha}^2$, sinon H_0 est refusée.
- • $M_n^2(\hat{\theta}_{GW}) \leq C_\alpha$, $C_\alpha = \chi_{3, 1-\alpha}^2$, sinon H_0 est refusée.

7 Tests d'ajustement pour la loi de Birnbaum-Saunders

Le test d'ajustement du Chi-deux est initialement étudié par Tahir et Saaidia (2012) et tahir (2012).

Considérons à tester l'hypothèse composite H_0 selon laquelle la distribution de l'échantillon $\mathbb{X} = (X_1, X_2, \dots, X_n)^T$ vérifie :

$$H_0 : P(X_i \leq x) = F_{BS}(x, \theta_{BS}), \quad x \geq 0, \quad \theta_{BS} = (\alpha, \beta)^T,$$

2.7 Tests d'ajustement pour la loi de Birnbaum-Saunders

où $F_{BS}(x, \theta_{BS})$ est la fonction de répartition de la loi BS donnée par l'équation (1.13).

La fonction de log-vraisemblance est :

$$\ell(\theta_{BS}) = \left(\frac{1}{\alpha^2} - \ln \alpha - \ln \beta \right) n - \frac{1}{2\alpha^2\beta} \sum_{i=1}^n X_i - \frac{\beta}{2\alpha^2} \sum_{i=1}^n X_i^{-1} + \sum_{i=1}^n \ln \left[\left(\frac{\beta}{X_i} \right)^{\frac{1}{2}} + \left(\frac{\beta}{X_i} \right)^{\frac{3}{2}} \right],$$

l'estimateur de MV $\hat{\theta}$ peut se déterminer par la résolution du système non linéaire suivant :

$$\begin{aligned} \frac{\partial \ell(\theta_{BS})}{\partial \alpha} &= - \left(\frac{1}{\alpha} + \frac{2}{\alpha^3} \right) n + \frac{1}{\alpha^3\beta} \sum_{i=1}^n X_i + \frac{\beta}{\alpha^3} \sum_{i=1}^n X_i^{-1} = 0, \\ \frac{\partial \ell(\theta_{BS})}{\partial \beta} &= - \frac{n}{2\beta} + \frac{1}{2\alpha^2\beta^2} \sum_{i=1}^n X_i - \frac{1}{2\alpha^2} \sum_{i=1}^n X_i^{-1} + \sum_{i=1}^n \frac{1}{X_i + \beta} = 0. \end{aligned}$$

Les dérivées partielles secondes de $\ell(\theta)$ sont données par

$$\begin{aligned} \frac{\partial^2 \ell(\theta_{BS})}{\partial \alpha^2} &= \frac{n}{\alpha^2} \left(1 + \frac{6}{\alpha^2} \right) - \frac{3}{\alpha^4\beta} \sum_{i=1}^n X_i - \frac{3\beta}{\alpha^4} \sum_{i=1}^n X_i^{-1}, \\ \frac{\partial^2 \ell(\theta_{BS})}{\partial \alpha \partial \beta} &= - \frac{1}{\alpha^3\beta^2} \sum_{i=1}^n X_i + \frac{1}{\alpha^3} \sum_{i=1}^n X_i^{-1}, \\ \frac{\partial^2 \ell(\theta_{BS})}{\partial \beta^2} &= \frac{n}{2\beta^2} - \frac{1}{\alpha^2\beta^3} \sum_{i=1}^n X_i - \sum_{i=1}^n \frac{1}{(X_i + \beta)^2}, \end{aligned}$$

Pour tester l'hypothèse H_0 , nous pouvons utiliser les 3 tests statistiques $Y_n^2(\hat{\theta}_{BS})$, $U_n^2(\hat{\theta}_{BS})$ et $M_n^2(\hat{\theta}_{BS})$,

avec

$$J(\theta_{BS}) = B^T(\theta_{BS})B(\theta_{BS}), \quad B(\theta_{BS}) = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ \vdots & \vdots \\ b_{r1} & b_{r2} \end{pmatrix},$$

$$b_{i1}(\theta_{BS}) = \frac{1}{\sqrt{p_i(\theta_{BS})}} \frac{\partial p_i(\theta_{BS})}{\partial \alpha}, \quad b_{i2}(\theta_{BS}) = \frac{1}{\sqrt{p_i(\theta_{BS})}} \frac{\partial p_i(\theta_{BS})}{\partial \beta}, \quad i = 1, 2, \dots, r,$$

avec

$$p_i(\theta_{BS}) = \Phi \left[\frac{1}{\alpha} \left\{ \left(\frac{\beta}{a_{i-1}} \right)^{\frac{1}{2}} - \left(\frac{a_{i-1}}{\beta} \right)^{\frac{1}{2}} \right\} \right] - \Phi \left[\frac{1}{\alpha} \left\{ \left(\frac{\beta}{a_i} \right)^{\frac{1}{2}} - \left(\frac{a_i}{\beta} \right)^{\frac{1}{2}} \right\} \right],$$

$$\frac{\partial p_i(\theta_{BS})}{\partial \alpha} = -\frac{1}{\alpha} A(a_i, \theta_{BS}) \varphi(A(a_i, \theta_{BS})) + \frac{1}{\alpha} A(a_{i-1}, \theta_{BS}) \varphi(A(a_{i-1}, \theta_{BS})),$$

$$\frac{\partial p_i(\theta_{BS})}{\partial \beta} = -\frac{1}{2\alpha} \left[\frac{\sqrt{\beta a_i}}{\beta^2} + \frac{1}{\sqrt{\beta a_i}} \right] \varphi(A(a_i, \theta_{BS})) + \frac{1}{2\alpha} \left[\frac{\sqrt{\beta a_{i-1}}}{\beta^2} + \frac{1}{\sqrt{\beta a_{i-1}}} \right] \varphi(A(a_{i-1}, \theta_{BS})),$$

$$\text{où } A(x, \theta_{BS}) = \frac{1}{\alpha} \left\{ \left(\frac{\beta}{x} \right)^{\frac{1}{2}} - \left(\frac{x}{\beta} \right)^{\frac{1}{2}} \right\}.$$

La matrice d'information de Fisher $I(\theta_{BS})$ peut être remplacée par

$$\hat{I}(\hat{\theta}_{BS}) = -\frac{1}{n} \begin{pmatrix} \frac{\partial^2 \ell^2(\hat{\theta}_{BS})}{\partial \alpha^2} & \frac{\partial^2 \ell^2(\hat{\theta}_{BS})}{\partial \alpha \partial \beta} \\ \frac{\partial^2 \ell^2(\hat{\theta}_{BS})}{\partial \alpha \partial \beta} & \frac{\partial^2 \ell^2(\hat{\theta}_{BS})}{\partial \beta^2} \end{pmatrix}.$$

Sous H_0 , et pour n suffisamment grand :

- • La statistique $Y_n^2(\hat{\theta}_{BS})$ suit la distribution χ_{r-1}^2 à $r - 1$ degrés de liberté.
- • La statistique $U_n^2(\hat{\theta}_{BS})$ suit la distribution χ_{r-3}^2 à $r - 3$ degrés de liberté.
- • La statistique $M_n^2(\hat{\theta}_{BS})$ suit la distribution χ_2^2 à 2 degrés de liberté.

Pour un seuil α , H_0 est acceptée si :

- • $Y_n^2(\hat{\theta}_{BS}) \leq C_\alpha$, $C_\alpha = \chi_{r-1, 1-\alpha}^2$, sinon H_0 est refusée.
- • $U_n^2(\hat{\theta}_{BS}) \leq C_\alpha$, $C_\alpha = \chi_{r-3, 1-\alpha}^2$, sinon H_0 est refusée.
- • $M_n^2(\hat{\theta}_{BS}) \leq C_\alpha$, $C_\alpha = \chi_{2, 1-\alpha}^2$, sinon H_0 est refusée.

8 Tests d'ajustement pour la loi Gaussienne inverse

Le test d'ajustement du Chi-deux pour la distribution IG est bien étudié par saaidia (2009), Lemeshko (2010a), Nikulin et Saaidia (2009), Saaidia et Seddik-Ameur (2010), Saaidia et Ramzan Tahir (2012).

Nous voulons tester l'hypothèse composite H_0 selon laquelle la distribution de l'échantillon $\mathbb{X} = (X_1, X_2, \dots, X_n)^T$ vérifie :

$$H_0 : P(X_i \leq x) = F_{IG}(x, \theta_{IG}), \quad x \geq 0, \quad \theta_{IG} = (\mu, \lambda)^T,$$

où $F_{IG}(x, \theta_{IG})$ est la fonction de répartition de la loi IG définie par l'équation (1.20).

Partageons la droite réelle en r sous-intervalles I_1, I_2, \dots, I_r mutuellement disjoints, par les points :

$$0 = a_0 < a_1 < \dots < a_{r-1} < a_r = +\infty.$$

Soit ν le vecteur de fréquences $\nu = (\nu_1, \nu_2, \dots, \nu_r)^T$, et le vecteur de probabilités $p(\theta_{IG}) = (p_1(\theta_{IG}), p_2(\theta_{IG}), \dots, p_r(\theta_{IG}))^T$ obtenus en regroupant l'échantillon X_1, X_2, \dots, X_n dans les sous-intervalles I_j .

Avec :

$$p_i(\theta_{IG}) = \int_{a_{i-1}}^{a_i} f_{IG}(x, \theta_{IG}) dx = F_{IG}(a_i, \theta_{IG}) - F_{IG}(a_{i-1}, \theta_{IG}), \quad i = 1, 2, \dots, r.$$

La matrice d'information de Fisher $I_n(\theta)$ pour les données non groupées est :

$$I_n(\theta_{IG}) = nI(\theta_{IG}) = \begin{pmatrix} \frac{n\lambda}{\mu^3} & 0 \\ 0 & \frac{n}{2\lambda^2} \end{pmatrix}.$$

Pour la calculer, nous prenons l'espérance des dérivées secondes suivantes de l'équation (1.26) :

$$\frac{\partial^2 \ell(\theta_{IG})}{\partial \mu^2} = n \frac{\lambda}{\mu^3} (2\mu - 3\bar{X}),$$

$$\frac{\partial^2 \ell(\theta_{IG})}{\partial \lambda^2} = -\frac{n}{2\lambda^2},$$

$$\frac{\partial^2 \ell(\theta_{IG})}{\partial \mu \partial \lambda} = 0.$$

Pour tester l'hypothèse H_0 , nous pouvons utiliser les 3 tests statistiques $Y_n^2(\hat{\theta}_{IG})$, $U_n^2(\hat{\theta}_{IG})$ et $M_n^2(\hat{\theta}_{IG})$,

où

$$J(\theta_{IG}) = B^T(\theta_{IG})B(\theta_{IG}), \quad B(\theta_{IG}) = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ \vdots & \vdots \\ b_{r1} & b_{r2} \end{pmatrix},$$

$$b_{i1}(\theta_{IG}) = \frac{1}{\sqrt{p_i(\theta_{IG})}} \frac{\partial p_i(\theta_{IG})}{\partial \mu}, \quad b_{i2}(\theta_{IG}) = \frac{1}{\sqrt{p_i(\theta_{IG})}} \frac{\partial p_i(\theta_{IG})}{\partial \lambda}, \quad i = 1, 2, \dots, r,$$

Dans le cas d'équiprobabilité (i.e. tous les sous-intervalles I_j ont des probabilités égales : $p_1 = p_2 = \dots = p_r$), les bornes des sous-intervalles a_i sont données par :

$$a_i = F_{IG}^{-1}(p_1 + p_2 + \dots + p_i) = F_{IG}^{-1}\left(\frac{i}{r}\right), \quad i = 1, \dots, r - 1.$$

Comme $p_i(\theta_{IG}) = F_{IG}(a_i) - F_{IG}(a_{i-1})$, $i = 1, 2, \dots, r$, alors nous pouvons écrire :

$$\begin{aligned} p_i(\theta_{IG}) &= \left[\Phi\left(\sqrt{\frac{\lambda}{a_i}}\left(\frac{a_i}{\mu} - 1\right)\right) + \exp\left(\frac{2\lambda}{\mu}\right) \Phi\left(-\sqrt{\frac{\lambda}{a_i}}\left(\frac{a_i}{\mu} + 1\right)\right) \right] - \\ &\left[\Phi\left(\sqrt{\frac{\lambda}{a_{i-1}}}\left(\frac{a_{i-1}}{\mu} - 1\right)\right) + \exp\left(\frac{2\lambda}{\mu}\right) \Phi\left(-\sqrt{\frac{\lambda}{a_{i-1}}}\left(\frac{a_{i-1}}{\mu} + 1\right)\right) \right]. \end{aligned} \quad (2.13)$$

Calculons les dérivées partielles de $p_i(\theta_{IG})$:

$$\begin{aligned} \frac{\partial p_i(\theta_{IG})}{\partial \mu} &= \frac{\sqrt{\lambda}}{\mu^2} (-\sqrt{a_i}\varphi(\alpha_i) + \sqrt{a_{i-1}}\varphi(\alpha_{i-1})) + \frac{2\lambda}{\mu^2} \exp\left(\frac{2\lambda}{\mu}\right) (-\Phi(\beta_i) + \Phi(\beta_{i-1})) + \\ &\frac{\sqrt{\lambda}}{\mu^2} \exp\left(\frac{2\lambda}{\mu}\right) (\sqrt{a_i}\varphi(\beta_i) - \sqrt{a_{i-1}}\varphi(\beta_{i-1})), \end{aligned} \quad (2.14)$$

$$\begin{aligned} \frac{\partial p_i(\theta_{IG})}{\partial \lambda} &= \frac{1}{\sqrt{2\lambda}} \left(\frac{1}{\sqrt{a_i}} \left(\frac{a_i}{\mu} - 1 \right) \varphi(\alpha_i) - \frac{1}{\sqrt{a_{i-1}}} \left(\frac{a_{i-1}}{\mu} - 1 \right) \varphi(\alpha_{i-1}) \right) + \frac{2}{\mu} \exp\left(\frac{2\lambda}{\mu}\right) (\Phi(\beta_i) - \Phi(\beta_{i-1})) + \\ &\quad \frac{1}{\sqrt{2\lambda}} \exp\left(\frac{2\lambda}{\mu}\right) \left(-\frac{1}{\sqrt{a_i}} \left(\frac{a_i}{\mu} + 1 \right) \varphi(\beta_i) + \frac{1}{\sqrt{a_{i-1}}} \left(\frac{a_{i-1}}{\mu} + 1 \right) \varphi(\beta_{i-1}) \right), \end{aligned} \quad (2.15)$$

où

$$\alpha_i = \sqrt{\frac{\lambda}{a_i}} \left(\frac{a_i}{\mu} - 1 \right) \quad \text{et} \quad \beta_i = -\sqrt{\frac{\lambda}{a_i}} \left(\frac{a_i}{\mu} + 1 \right), \quad i = 1, \dots, r.$$

Sous H_0 , et pour n suffisamment grand :

- • La statistique $Y_n^2(\hat{\theta}_{IG})$ suit la distribution χ_{r-1}^2 à $r - 1$ degrés de liberté.
- • La statistique $U_n^2(\hat{\theta}_{IG})$ suit la distribution χ_{r-3}^2 à $r - 3$ degrés de liberté.
- • La statistique $M_n^2(\hat{\theta}_{IG})$ suit la distribution χ_2^2 à 2 degrés de liberté.

Pour un seuil α , H_0 est acceptée si :

- • $Y_n^2(\hat{\theta}_{IG}) \leq C_\alpha$, $C_\alpha = \chi_{r-1, 1-\alpha}^2$, sinon H_0 est refusée.
- • $U_n^2(\hat{\theta}_{IG}) \leq C_\alpha$, $C_\alpha = \chi_{r-3, 1-\alpha}^2$, sinon H_0 est refusée.
- • $M_n^2(\hat{\theta}_{IG}) \leq C_\alpha$, $C_\alpha = \chi_{2, 1-\alpha}^2$, sinon H_0 est refusée.

La statistique $Y_n^2(\hat{\theta}_{IG})$ a été calculée d'une manière explicite [Saaidia (2009), Saaidia et Seddik-Ameur (2010)] pour les différentes situations, selon que les paramètres soient connus ou non.

1. Cas où $\theta_{IG} = (\mu, \lambda)^T$ est connu :

Dans ce cas, la statistique $Y_n^2(\hat{\theta}_{IG})$ devient :

$$Y_n^2 = X_n^T(\theta_{IG}) X_n(\theta_{IG}) = X_n^2(\theta) = \sum_{i=1}^r \frac{(\nu_i - np_i)^2}{np_i}.$$

Au seuil α , si on choisit la valeur critique :

$$C_\alpha = \chi_{r-1, 1-\alpha}^2,$$

l'hypothèse H_0 est acceptée si $Y_n^2 \leq C_\alpha$. Dans le cas contraire, H_0 est rejetée.

2. Cas où μ ou λ est inconnu :

On a alors

$$Y_n^2(\hat{\theta}_{IG}) = \begin{cases} X_n^2(\hat{\lambda}) + \frac{1}{n} \frac{\left(\sum_{i=1}^r \frac{\nu_i}{p_i} \frac{\partial p_i}{\partial \lambda}(\hat{\lambda})\right)^2}{\frac{1}{2\hat{\lambda}} - \sum_{i=1}^r \frac{1}{p_i} \left(\frac{\partial p_i}{\partial \lambda}(\hat{\lambda})\right)^2}, & \text{si } \mu \text{ est connu et } \lambda \text{ est inconnu} \\ X_n^2(\hat{\mu}) + \frac{1}{n} \frac{\left(\sum_{i=1}^r \frac{\nu_i}{p_i} \frac{\partial p_i}{\partial \mu}(\hat{\mu})\right)^2}{\frac{\lambda}{\hat{\mu}^3} - \sum_{i=1}^r \frac{1}{p_i} \left(\frac{\partial p_i}{\partial \mu}(\hat{\mu})\right)^2}, & \text{si } \mu \text{ est inconnu et } \lambda \text{ est connu} \end{cases}.$$

3. Cas où θ est inconnu :

Par des calculs longs et difficiles, nous obtenons la forme de la statistique de NRR (2.12) comme suit :

$$\begin{aligned} Y_n^2(\hat{\theta}_{IG}) &= X_n^2(\hat{\theta}_{IG}) + \frac{1}{n|M|} \left[\left(\frac{1}{2\hat{\lambda}_n^2} - \sum_{i=1}^r \frac{1}{p_i} \omega_{i2}^2(\hat{\theta}_{IG}) \right) \left(\sum_{i=1}^r \frac{\nu_i}{p_i} \omega_{i1}(\hat{\theta}_{IG}) \right)^2 \right] + \\ &2 \left[\left(\sum_{i=1}^r \frac{1}{p_i} \omega_{i1}(\hat{\theta}_{IG}) \omega_{i2}(\hat{\theta}_{IG}) \right) \left(\sum_{i=1}^r \frac{\nu_i}{p_i} \omega_{i1}(\hat{\theta}_{IG}) \right) \left(\sum_{i=1}^r \frac{\nu_i}{p_i} \omega_{i2}(\hat{\theta}_{IG}) \right) \right] + \\ &\left[\left(\frac{\hat{\lambda}_n}{\hat{\mu}_n^3} - \sum_{i=1}^r \frac{1}{p_i} \omega_{i1}^2(\hat{\theta}_{IG}) \right) \left(\sum_{i=1}^r \frac{\nu_i}{p_i} \omega_{i2}(\hat{\theta}_{IG}) \right)^2 \right], \end{aligned}$$

où

$$\begin{aligned} M &= i(\hat{\theta}_{IG}) - J(\hat{\theta}_{IG}) = \begin{pmatrix} \frac{\hat{\lambda}_n}{\hat{\mu}_n^3} - \sum_{i=1}^r b_{i1}^2 & -\sum_{i=1}^r b_{i1} b_{i2} \\ -\sum_{i=1}^r b_{i1} b_{i2} & \frac{1}{2\hat{\lambda}_n^2} - \sum_{i=1}^r b_{i2}^2 \end{pmatrix}, \\ \omega_{i1}(\hat{\theta}_{IG}) &= \frac{\partial p_i}{\partial \mu}(\hat{\theta}_{IG}) \quad \text{et} \quad \omega_{i2}(\hat{\theta}_{IG}) = \frac{\partial p_i}{\partial \lambda}(\hat{\theta}_{IG}), \end{aligned}$$

et $|M|$ est le déterminant de la matrice M .

0.2, 0.3, 0.5, 0.5, 0.5, 0.6, 0.6, 0.7, 0.7, 0.7, 0.8, 0.8, 1.0, 1.0, 1.0, 1.0, 1.1, 1.3, 1.5, 1.5, 1.5, 1.5, 2.0, 2.0, 2.2, 2.5, 2.7, 3.0, 3.0, 3.3, 3.3, 4.0, 4.0, 4.5, 4.7, 5.0, 5.4, 5.4, 7.0, 7.5, 8.8, 9.0, 10.3, 22.0, 24.5.
--

Tableau 2.1 – Données d'entretien d'ERCA.

Exemple :

Pour montrer que ces tests sont maniables et pratiques, nous les appliquons à des observations réelles qui représentent les données d'entretien (en heures) pour un émetteur-récepteur de communication aéroporté (ERCA) "airborne communication transceiver" [Von Alven (1964), p156].

Pour ces données, et pour un nombre de classes équiprobables $r = 5$, nous avons calculé pour les différentes méthodes les estimateurs $\hat{\theta}_{IG}$, $\hat{\theta}_{IG}$, $\tilde{\theta}_{IG}$ et $\bar{\theta}_{IG}$ de θ qui sont respectivement :

- • $\hat{\theta}_{IG} = (3.6756, 1.6005)$ est l'estimateur de MVUE,
- • $\hat{\theta}_{IG} = (3.6756, 1.7148)$ est l'estimateur du maximum de vraisemblance,
- • $\tilde{\theta}_{IG} = (3.2228, 1.8945)$ est l'estimateur de la méthode du minimum du Chi-deux,
- • $\bar{\theta}_{IG} = (3.6756, 2.0042)$ est l'estimateur de la méthode des moments.

Les estimateurs sans biais de variance minimale de $\hat{\theta}_{IG}$ et de la densité \hat{f} sont calculés en utilisant le tableau 1.2. Il est clair que les deux estimateurs $\hat{\theta}_{IG}$ et $\hat{\theta}_{IG}$ sont proches.

Remarque 2.3. Pour calculer l'estimateur $\tilde{\theta}_{IG}$, nous résolvons un système non linéaire formé par ces deux équations non linéaires :

$$\left\{ \begin{array}{l} \sum_{i=1}^r \frac{\nu_i}{p_i(\theta_{IG})} \frac{\partial p_i(\theta_{IG})}{\partial \mu} = 0 \\ \sum_{i=1}^r \frac{\nu_i}{p_i(\theta_{IG})} \frac{\partial p_i(\theta_{IG})}{\partial \lambda} = 0 \end{array} \right.$$

Chapitre 2. Test d'ajustement du Chi-deux pour les familles de fonction de hasard unimodale

En tenant compte des formules (2.14) et (2.15), les deux équations de ce système s'écrivent :

$$\sum_{i=1}^r \frac{\nu_i}{p_i(\theta_{IG})} \left\{ \frac{\sqrt{\lambda}}{\mu^2} (-\sqrt{a_i}\varphi(\alpha_i) + \sqrt{a_{i-1}}\varphi(\alpha_{i-1})) + \frac{2\lambda}{\mu^2} \exp\left(\frac{2\lambda}{\mu}\right) (-\Phi(\beta_i) + \Phi(\beta_{i-1})) + \frac{\sqrt{\lambda}}{\mu^2} \exp\left(\frac{2\lambda}{\mu}\right) (\sqrt{a_i}\varphi(\beta_i) - \sqrt{a_{i-1}}\varphi(\beta_{i-1})) \right\} = 0,$$

$$\sum_{i=1}^r \frac{\nu_i}{p_i(\theta_{IG})} \left\{ \frac{\sqrt{\lambda}}{\mu^2} (-\sqrt{a_i}\varphi(\alpha_i) + \sqrt{a_{i-1}}\varphi(\alpha_{i-1})) + \frac{2\lambda}{\mu^2} \exp\left(\frac{2\lambda}{\mu}\right) (-\Phi(\beta_i) + \Phi(\beta_{i-1})) + \frac{\sqrt{\lambda}}{\mu^2} \exp\left(\frac{2\lambda}{\mu}\right) (\sqrt{a_i}\varphi(\beta_i) - \sqrt{a_{i-1}}\varphi(\beta_{i-1})) \right\} = 0,$$

où $p_i(\theta_{IG})$ est donné par la formule (2.13).

Dans la figure 2.1 (figure 2.2), nous représentons graphiquement les fonctions de densité (de survie) de la distribution Gaussienne inverse correspondantes à ces estimateurs, ainsi l'estimateur de MVUE de la fonction de densité (de survie) \hat{f} (\hat{S}) (voir tableau 1.2). Nous remarquons que les fonctions de densités sont très proches ainsi que les fonctions de survie.

En appliquant les 3 tests, nous trouvons alors :

$$Y_n^2(\hat{\theta}_{IG}) = 1.196187, U_n^2(\hat{\theta}_{IG}) = 1.018421 \text{ et } M_n^2(\hat{\theta}_{IG}) = 0.177766.$$

Pour un seuil $\alpha = 0.05$, les valeurs critiques pour les 3 tests sont respectivement : $C_\alpha = \chi_{r-1,1-\alpha}^2 = 9.487729$, $C_\alpha = \chi_{r-3,1-\alpha}^2 = 5.991465$ et $C_\alpha = \chi_{2,1-\alpha}^2 = 5.991465$,

et nous concluons alors que la distribution IG donne une bonne adéquation pour ces données.

Remarque 2.4. Notons que Von Alven (1964) a proposé la distribution log-normale pour ajuster les données d'ERCA. Chhikara et Folks (1977) ont proposé la statistique de Kolmogorov-Smirnov pour montrer que la distribution IG donne un bon ajustement pour ces données.

Calculons maintenant les valeurs des 3 tests en considérant les distributions LN , LL , BS , EW et

2.8 Tests d'ajustement pour la loi Gaussienne inverse

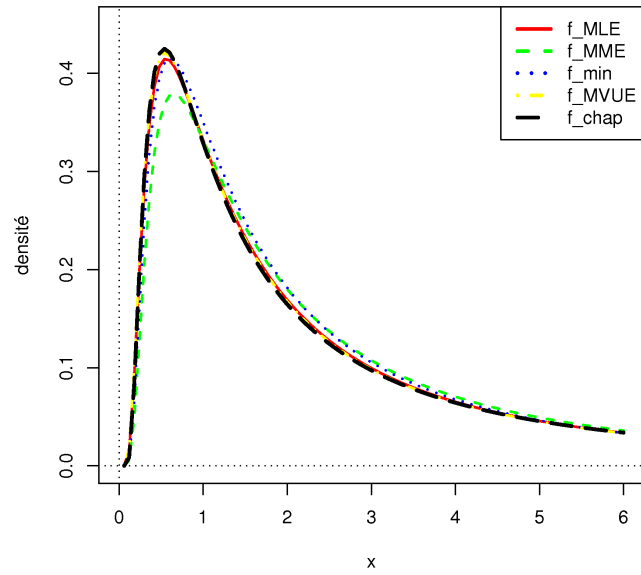


Figure 2.1 – Comparaison des fonctions de densité pour IG .

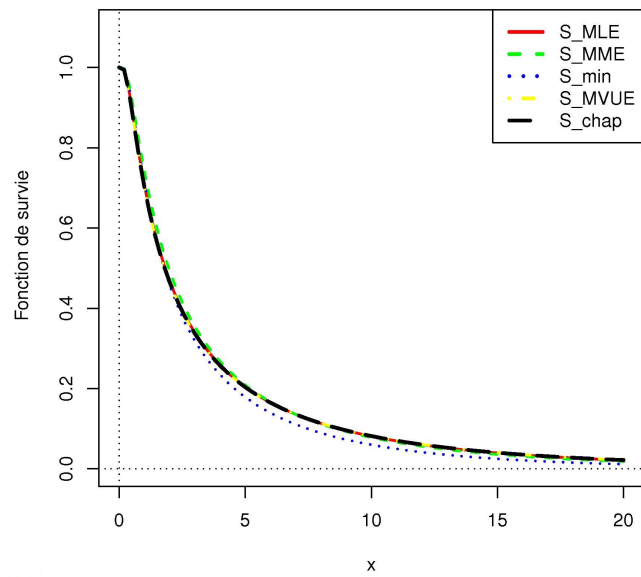


Figure 2.2 – Comparaison des fonctions de survie pour IG .

Loi	$Y_n^2(\hat{\theta})$	$U_n^2(\hat{\theta})$	$M_n^2(\hat{\theta})$
<i>LN</i>	5.538831	3.588526	1.950305
<i>LL</i>	14.90841	0.04543832	14.86297
<i>BS</i>	5.216523	3.214698	2.001825
<i>EW</i>	0.9898445	0.2338761	0.7559684
<i>GW</i>	1.352059	0.3482281	1.003831

Tableau 2.2 – Valeurs des 3 tests pour les Données d'entretien d'ERCA.

GW. Les résultats sont résumés au tableau 2.2.

Pour un seuil $\alpha = 0.05$, les valeurs critiques pour les 3 tests pour les distributions *EW* et *GW* sont respectivement : $C_\alpha = \chi_{r-1,1-\alpha}^2 = 9.487729$, $C_\alpha = \chi_{r-4,1-\alpha}^2 = 3.841459$ et $C_\alpha = \chi_{3,1-\alpha}^2 = 7.814728$.

Il s'ensuit que toutes les distributions donnent un bon ajustement pour les données d'entretien d'ERCA, sauf la distribution *LL* en tenant compte des statistiques $Y_n^2(\hat{\theta}_{LL})$ et $M_n^2(\hat{\theta}_{LL})$, H_0 n'est pas acceptée.

8.1 Simulation

Pour étudier empiriquement le comportement de la statistique de NRR $Y_n^2(\hat{\theta}_{IG})$, nous avons généré à la base de la distribution Gaussienne inverse des échantillons de taille $n = 20, 30, 50, 80, 100, 150, 200, 300, 400$ et 500 , et pour diverses combinaisons des paramètres μ et λ .

Dans cette étude, nous avons choisi le seuil $\alpha = 0.05$. Le paramètre μ varie de : $0,25, 1, 4$ à 10 , et le paramètre λ varie de : $0,5, 1, 5$ à 10 .

Considérons le cas d'équiprobabilité. Chaque échantillon (n, μ, λ) est répété 5000 fois. Pour chaque opération, nous calculons la statistique de NRR $Y_n^2(\hat{\theta}_{IG})$, puis nous calculons le niveau empirique (*N.E.*) qui compte le nombre de fois où $Y_n^2(\hat{\theta}_{IG}) \leq C_\alpha = \chi_{r-1,1-\alpha}^2$ divisé par 5000 .

Comme le seuil $\alpha = 0.05$, alors nous avons théoriquement *N.E.* = 0.95 . Les résultats sont groupés dans les tableaux 2.3 à 2.6. Dans les figures 2.3 et 2.4, nous représentons pour $(\mu, \lambda) = (1, 1)$ et $n = 20, 30, 50, 80, 100, 150, 200, 300, 400$ et 500 , les valeurs de $Y_n^2(\hat{\theta}_{IG})$ en histogrammes.

Une lecture simple des tableaux nous indique que les valeurs du niveau empirique *N.E.* sont très proches de la valeur théorique 0.95 . Dans les figures 2.3 et 2.4, le graphe reliant les milieux des histogrammes est étroitement confondu avec celui de la distribution χ_{r-1}^2 , ce qui confirme le théorème 2.5.

2.8 Tests d'ajustement pour la loi Gaussienne inverse

$(\mu, \lambda) = (0.25, 0.5)$										
n	20	30	50	80	100	150	200	300	400	500
$N.E.$	0.9520	0.9540	0.9486	0.9532	0.9516	0.9516	0.9418	0.9558	0.9512	0.9502
$(\mu, \lambda) = (0.25, 1)$										
n	20	30	50	80	100	150	200	300	400	500
$N.E.$	0.9532	0.9534	0.9520	0.9448	0.9454	0.9474	0.9482	0.9520	0.9512	0.9516
$(\mu, \lambda) = (0.25, 5)$										
n	20	30	50	80	100	150	200	300	400	500
$N.E.$	0.9558	0.9562	0.9540	0.9500	0.9552	0.9518	0.9562	0.9490	0.9464	0.9526
$(\mu, \lambda) = (0.25, 10)$										
n	20	30	50	80	100	150	200	300	400	500
$N.E.$	0.9500	0.9552	0.9468	0.9494	0.9534	0.9482	0.9492	0.9524	0.9574	0.9526

Tableau 2.3 – Niveau empirique pour $\mu = 0.25$ et $\lambda = (0.5, 1, 5, 10)$.

$(\mu, \lambda) = (1, 0.5)$										
n	20	30	50	80	100	150	200	300	400	500
$N.E.$	0.9586	0.9498	0.9484	0.9506	0.9536	0.9468	0.9532	0.9480	0.9506	0.9468
$(\mu, \lambda) = (1, 1)$										
n	20	30	50	80	100	150	200	300	400	500
$N.E.$	0.9522	0.9574	0.9584	0.9478	0.9486	0.9494	0.9484	0.9552	0.9468	0.9490
$(\mu, \lambda) = (1, 5)$										
n	20	30	50	80	100	150	200	300	400	500
$N.E.$	0.9492	0.9544	0.9618	0.9472	0.9540	0.9496	0.9482	0.9476	0.9520	0.9486
$(\mu, \lambda) = (1, 10)$										
n	20	30	50	80	100	150	200	300	400	500
$N.E.$	0.9534	0.9467	0.9534	0.9498	0.9465	0.9522	0.9534	0.9468	0.9478	0.9521

Tableau 2.4 – Niveau empirique pour $\mu = 1$ et $\lambda = (0.5, 1, 5, 10)$.

Chapitre 2. Test d'ajustement du Chi-deux pour les familles de fonction de hasard unimodale

$(\mu, \lambda) = (4, 0.5)$										
n	20	30	50	80	100	150	200	300	400	500
$N.E.$	0.9592	0.9518	0.9564	0.9506	0.9468	0.9444	0.9528	0.9492	0.9434	0.9489
$(\mu, \lambda) = (4, 1)$										
n	20	30	50	80	100	150	200	300	400	500
$N.E.$	0.9516	0.9524	0.9576	0.9423	0.9498	0.9454	0.9494	0.9558	0.9478	0.9523
$(\mu, \lambda) = (4, 5)$										
n	20	30	50	80	100	150	200	300	400	500
$N.E.$	0.9565	0.9512	0.9467	0.9487	0.9545	0.9534	0.9521	0.9540	0.9498	0.9512
$(\mu, \lambda) = (4, 10)$										
n	20	30	50	80	100	150	200	300	400	500
$N.E.$	0.9528	0.9553	0.9590	0.9487	0.9477	0.9489	0.9487	0.9543	0.9469	0.9493

Tableau 2.5 – Niveau empirique pour $\mu = 4$ et $\lambda = (0.5, 1, 5, 10)$.

$(\mu, \lambda) = (10, 0.5)$										
n	20	30	50	80	100	150	200	300	400	500
$N.E.$	0.9419	0.9527	0.9520	0.9489	0.9482	0.9485	0.9460	0.9476	0.9551	0.9498
$(\mu, \lambda) = (10, 1)$										
n	20	30	50	80	100	150	200	300	400	500
$N.E.$	0.9521	0.9556	0.9498	0.9456	0.9478	0.9490	0.9467	0.9572	0.9499	0.9499
$(\mu, \lambda) = (10, 5)$										
n	20	30	50	80	100	150	200	300	400	500
$N.E.$	0.9487	0.9534	0.9507	0.9454	0.9460	0.9498	0.9478	0.9532	0.9455	0.9489
$(\mu, \lambda) = (10, 10)$										
n	20	30	50	80	100	150	200	300	400	500
$N.E.$	0.9521	0.9543	0.9520	0.9488	0.9477	0.9489	0.9488	0.9523	0.9432	0.9498

Tableau 2.6 – Niveau empirique pour $\mu = 10$ et $\lambda = (0.5, 1, 5, 10)$.

2.8 Tests d'ajustement pour la loi Gaussienne inverse

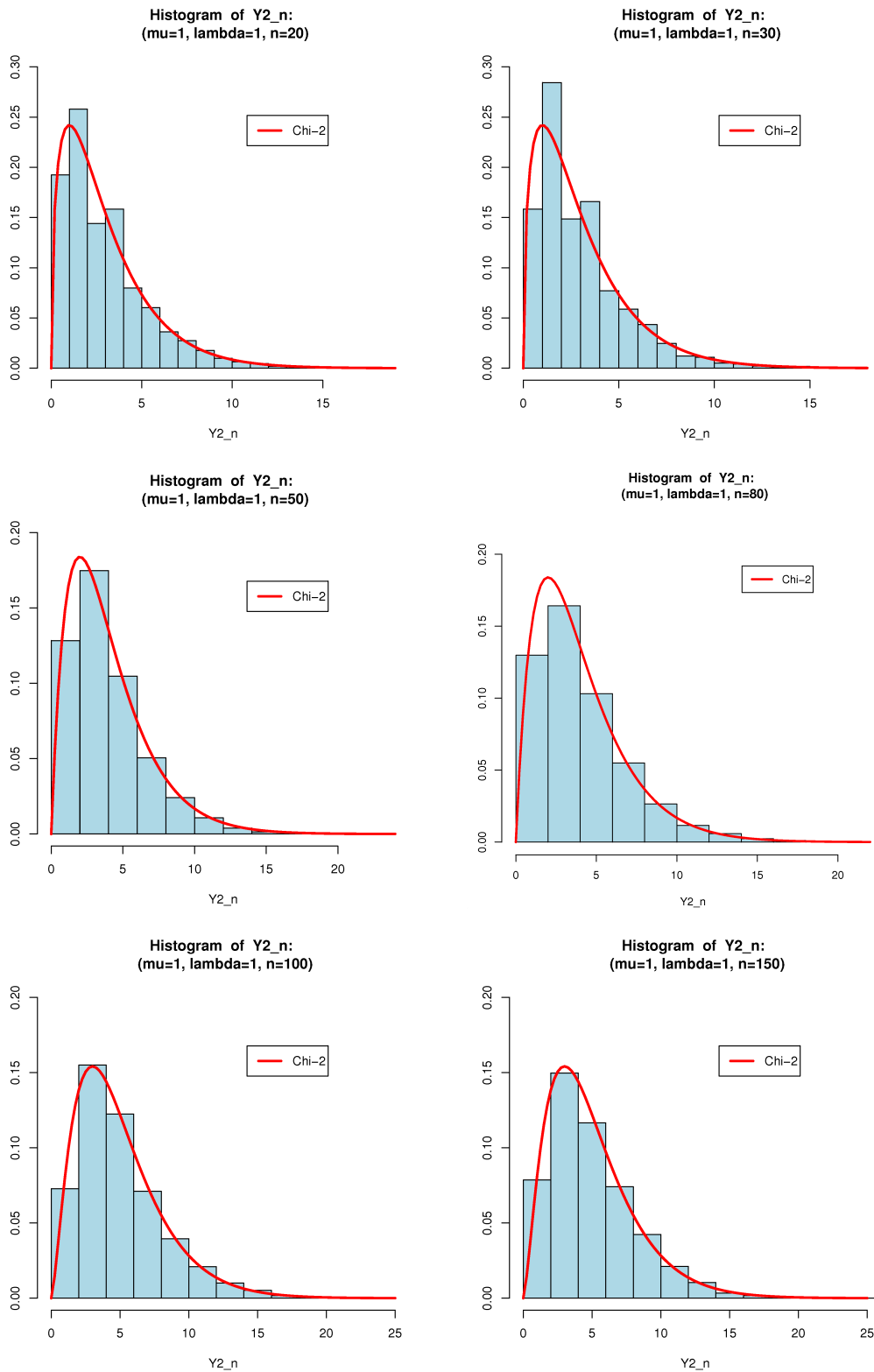


Figure 2.3 – Histogrammes de $Y2_n$ pour $(\mu, \lambda) = (1, 1)$, et $n = 20, 30, 50, 80, 100, 150$.

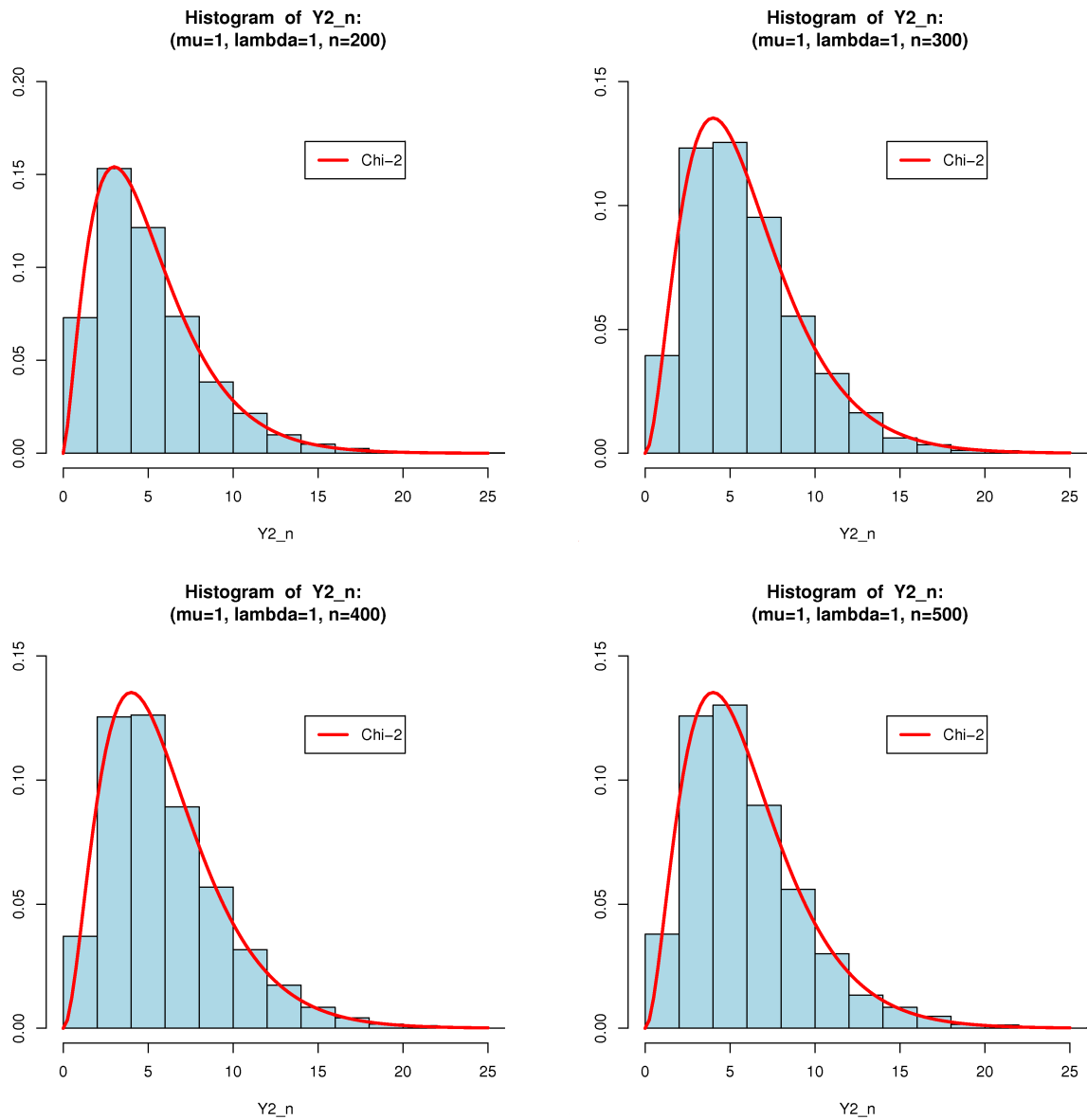


Figure 2.4 – Histogrammes de Y_2^n pour $(\mu, \lambda) = (1, 1)$, et $n = 200, 300, 400, 500$.

8.2 Autres tests d'ajustement pour la loi Gaussienne inverse

Soit $\mathbb{X} = (X_1, X_2, \dots, X_n)^T$ un n -échantillon, nous voulons tester l'hypothèse composite H_0 selon laquelle la distribution de l'échantillon \mathbb{X} vérifie :

$$H_0 : P(X_i \leq x) = F_{IG}(x, \theta_{IG}), \quad x \geq 0, \quad \theta_{IG} = (\mu, \lambda)^T.$$

• Test de Kolmogorov

Soit la statistique :

$$D_n = \sup_{|n| < \infty} |F_n(x) - F_{IG}(x, \theta_{IG})|,$$

où $F_n(x)$ est la fonction empirique. En pratique, il est préférable d'utiliser le test basé sur D_n avec correction de Bolshev [Bolshev (1987)] de la forme [Bolshev et Smirnov (1983)] :

$$S_K = \frac{6nD_n + 1}{6\sqrt{n}},$$

où $D_n = \max(D_n^+, D_n^-)$,

$$D_n^+ = \max_{1 \leq i \leq n} \left\{ \frac{i}{n} - F_{IG}(x_i, \theta_{IG}) \right\}, \quad D_n^- = \max_{1 \leq i \leq n} \left\{ F_{IG}(x_i, \theta_{IG}) - \frac{i-1}{n} \right\},$$

avec x_1, x_2, \dots, x_n sont en ordre croissant.

Si θ_{IG} est connu, la distribution de la statistique de Bolshev obéit à la distribution de Kolmogorov $K(S)$ [Bolshev et Smirnov (1983)].

• Test de Cramér-Von Mises

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

$$S_\omega = \frac{1}{12n} + \sum_{i=1}^n \left\{ F_{IG}(x_i, \theta_{IG}) - \frac{2i-1}{2n} \right\}^2.$$

Si θ_{IG} est connu, la statistique de Cramér-Von Mises à la distribution $a_1(S)$ [Bolshev et Smirnov (1983)].

• Test d'Anderson-Darling

La statistique de Ω^2 de Anderson-Darling peut s'écrire sous la forme :

$$S_{\Omega} = -n - 2 \sum_{i=1}^n \left\{ \frac{2i-1}{2n} \ln F_{IG}(x_i, \theta_{IG}) + \left(1 - \frac{2i-1}{2n}\right) \ln(1 - F_{IG}(x_i, \theta_{IG})) \right\}.$$

Si θ_{IG} est connu, la distribution de la statistique d'Anderson-Darling est $a_2(S)$ [Bolshev et Smirnov (1983)].

Remarque 2.5. Si θ_{IG} est inconnu, la distribution de ces statistiques dépend d'un certain nombre de facteurs : la forme de la loi $F_{IG}(x, \theta_{IG})$ correspondant à l'hypothèse H_0 , la méthode d'estimation des paramètres et le nombre de paramètres estimés, (voir Lemeshko et Lemeshko (2009), Lemeshko et al. (2010a)).

Dans [Lemeshko et al. (2010b)], nous donnons la comparaison entre ces tests et le test de NRR $Y_n^2(\hat{\theta}_{IG})$.

9 Test de Chi-deux de la loi Gaussienne inverse contre la loi log-normale

Considérons l'hypothèse nulle H_0 telle que :

$$H_0 : P(X_i \leq x) = F_{IG}(x, \theta_{IG}), \quad \theta_{IG} = (\mu, \sigma)^T.$$

Nous voulons tester l'hypothèse H_0 contre l'hypothèse alternative H_1 :

$$H_1 : P(X_i \leq x) = F_{LN}(x, \theta_{LN}), \quad \theta_{LN} = (m, \sigma)^T,$$

où

F_{LN} est la fonction de répartition de log-normale $LN(m, \sigma)$ définie par :

$$F_{LN}(x, \theta_{LN}) = \Phi \left(\frac{\ln x - m}{\sigma} \right).$$

La fonction de densité de $LN(m, \sigma)$ est définie par :

$$f_{LN}(x, \theta_{LN}) = \frac{1}{\sigma \sqrt{2\pi} x} \exp \left\{ -\frac{1}{2} \left(\frac{\ln x - m}{\sigma} \right)^2 \right\}.$$

En appliquant la statistique de NRR, il est raisonnable d'utiliser les classes de *Neyman-Pearson*

2.9 Test de Chi-deux de la loi Gaussienne inverse contre la loi log-normale

(voir Greenwood et Nikulin (1996)) pour grouper les données :

$$I_1 = \{x : f_{IG}(x, \theta_{IG}) \geq f_{LN}(x, \theta_{LN})\}, \quad I_2 = \{x : f_{IG}(x, \theta_{IG}) < f_{LN}(x, \theta_{LN})\},$$

les intervalles I_1 et I_2 sont déterminés par la résolution de l'équation non linéaire :

$$f_{IG}(x, \hat{\theta}_{IG}) = f_{LN}(x, \hat{\theta}_{LN}), \quad (2.16)$$

Pour trouver les racines de l'équation (2.16), nous pouvons utiliser une méthode numérique. Pour cela nous avons réalisé un algorithme en constatant que l'équation (2.16) possède quatre points d'intersection distincts quels que soient les paramètres $\hat{\theta}_{IG}$ et $\hat{\theta}_{LN}$. Pour les données d'ERCA, nous trouvons :

$$a_1^* = 0.2839932, \quad a_2^* = 1.4492008, \quad a_3^* = 8.6505096, \quad a_4^* = 50.6410849.$$

Notons que par l'égalisation des premiers moments des lois Gaussienne inverse et log-normale, il résulte que l'équation (2.16) possède aussi quatre points d'intersection distincts, et nous constatons que les deux distributions sont très proches l'une de l'autre "close each other".

Dans la figure 2.5, nous représentons les fonctions de densité $f_{IG}(x, \hat{\theta}_{IG})$ et $f_{LN}(x, \hat{\theta}_{LN})$. Nous remarquons qu'elles sont très proches.

Considérons la partition de la droite réelle par les points :

$$0 = a_0 < a_1^* < a_2^* < a_3^* < a_4^* < a_5 = +\infty,$$

nous obtenons le vecteur de fréquences $\nu = (\nu_1, \nu_2, \dots, \nu_5)^T$, et le vecteur de probabilités $p = (p_1, p_2, \dots, p_5)^T$. Ensuite nous calculons le vecteur de fréquences $V = (V_1, V_2)^T$ et le vecteur de probabilités $P = (P_1, P_2)^T$ qui correspondent aux intervalles I_1 et I_2 tels que :

$$P_1 = \sum_{f_{IG} \geq f_{LN}} p_i = \int_{I_1} f_{IG}(x, \hat{\theta}_{IG}) dx, \quad P_2 = \sum_{f_{IG} < f_{LN}} p_i = \int_{I_2} f_{IG}(x, \hat{\theta}_{IG}) dx,$$

$$V_1 = \sum_{f_{IG} \geq f_{LN}} \nu_i, \quad V_2 = \sum_{f_{IG} < f_{LN}} \nu_i.$$

Alors, pour tester H_0 contre H_1 , nous pouvons utiliser la statistique de NRR avec les classes I_1 et I_2 :

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

avec

$$X_n(\hat{\theta}_{IG}) = \left(\frac{V_1 - nP_1(\hat{\theta}_{IG})}{\sqrt{nP_1(\hat{\theta}_{IG})}}, \frac{V_2 - nP_2(\hat{\theta}_{IG})}{\sqrt{nP_2(\hat{\theta}_{IG})}} \right)^T.$$

D'après le théorème 2.5, nous savons que :

$$\lim_{n \rightarrow \infty} \mathbf{P} \left(Y_n^2(\hat{\theta}_{IG}) > x | H_0 \right) = \mathbf{P}(\chi_1^2 > x),$$

Considérons les données d'ERCA, les estimateurs $\hat{\theta}_{IG}$ et $\hat{\theta}_{LN}$ du maximum de vraisemblance sont :

$$\hat{\theta}_{IG} = (\hat{\mu}, \hat{\lambda})^T = (3.6756, 1.7148)^T,$$

$$\hat{\theta}_{LN} = (\hat{m}, \hat{\sigma})^T = (0.68843, 1.09518)^T.$$

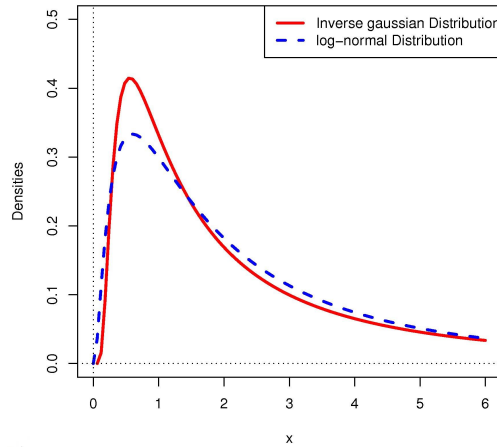


Figure 2.5 – Comparaison des densités de *IG* et de *LN* pour les données d'ERCA.

La valeur de la statistique de NRR est $Y_n^2(\hat{\theta}_{IG}) = 0.01699$. Pour un seuil $\alpha = 0.05$, la valeur critique est $C_\alpha = \chi_{1,1-\alpha}^2 = 3.841459$, alors nous acceptons H_0 , et nous concluons que, pour les données d'ERCA, la distribution Gaussienne inverse donne un meilleur ajustement par rapport à loi log-normale.

10 Test du Chi-deux de la loi Gaussienne inverse contre la loi de Weibull généralisée

Considérons l'hypothèse nulle H_0 pour laquelle la distribution de X_i appartient à la famille de lois Gaussiennes inverses :

$$H_0 : P(X \leq x) = F_{IG}(x, \theta_{IG}), \quad \theta_{IG} = (\mu, \lambda)^T, \quad x \geq 0,$$

nous voulons tester H_0 contre l'hypothèse alternative H_1 :

$$H_1 : P(T \leq x) = F_{GW}(x, \theta_{GW}), \quad \theta_{GW} = (\sigma, \nu, \gamma)^T \in \mathbf{R}_+^3, \quad x \geq 0.$$

La fonction de densité de la distribution de Weibull généralisée est donnée par :

$$f_{GW}(x, \theta_{GW}) = \frac{\nu}{\gamma \sigma^\nu} x^{\nu-1} \left(1 + \left(\frac{x}{\sigma}\right)^\nu\right)^{\frac{1}{\gamma}-1} \exp\left\{1 - \left(1 + \left(\frac{x}{\sigma}\right)^\nu\right)^{\frac{1}{\gamma}}\right\}, \quad x \geq 0,$$

et la fonction de survie est donnée par :

$$S_{GW}(x, \theta_{GW}) = 1 - F_{GW}(x, \theta_{GW}) = \exp\left\{1 - \left(1 + \left(\frac{x}{\sigma}\right)^\nu\right)^{\frac{1}{\gamma}}\right\}, \quad x \geq 0.$$

Pour tester H_0 contre H_1 , nous pouvons utiliser la statistique de NRR avec les classes de Neyman-Pearson :

$$I_1 = \{x : f_{IG}(x, \theta_{GI}) \geq f_{GW}(x, \theta_{GW})\}, \quad I_2 = \{x : f_{IG}(x, \theta_{IG}) < f_{GW}(x, \theta_{GW})\}.$$

Comme auparavant, les classes I_1 et I_2 s'obtiennent en résolvant l'équation :

$$f_{IG}(x, \hat{\theta}_{IG}) = f_{GW}(x, \hat{\theta}_{GW}). \quad (2.17)$$

Cette équation (2.17) possède aussi quatre points d'intersection distincts c_1^* , c_2^* , c_3^* et c_4^* , quels que soient les paramètres θ_{IG} et θ_{GW} .

Prenons les données d'ERCA, nous calculons les estimateurs $\hat{\theta}_{IG}$ et $\hat{\theta}_{GW}$ du maximum de vraisemblance :

$$\begin{aligned} \hat{\theta}_{IG} &= (\hat{\mu}, \hat{\lambda})^T = (3.6756, 1.7148)^T, \\ \hat{\theta}_{GW} &= (\hat{\sigma}, \hat{\nu}, \hat{\gamma})^T = (0.5092, 3.0901, 7.3346)^T, \end{aligned}$$

comme $\hat{\gamma} > \hat{\nu} > 1$, la fonction de hazard h_{GW} est unimodale.

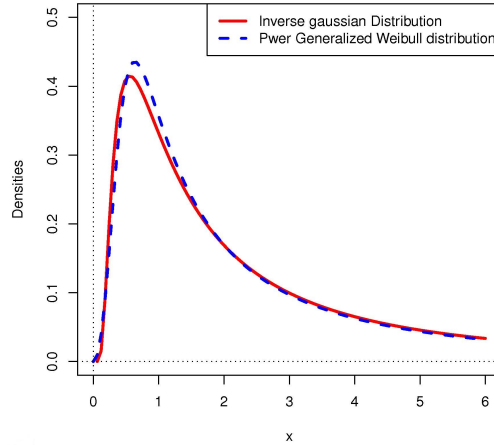


Figure 2.6 – Comparaison des densités de IG et de GW pour les données d'ERCA.

Les points d'intersection sont

$$c_1^* = 0.1776, c_2^* = 0.5106, c_3^* = 2.0326, \text{ et } c_4^* = 22.1466.$$

Dans la figure 2.6, nous représentons les fonctions de densité $f_{IG}(x, \hat{\theta}_{IG})$ et $f_{GW}(x, \hat{\theta}_{WG})$. Nous remarquons aussi qu'elles sont très proches.

Après calculs, nous trouvons que la valeur de la statistique de NRR est $Y_n^2(\hat{\theta}_{IG}) = 0.009994682$. Pour un seuil $\alpha = 0.05$, la valeur critique est $C_\alpha = \chi_{1,1-\alpha}^2 = 3.841459$, et par conséquent, nous acceptons H_0 . Ce résultat nous indique que la loi Gaussienne inverse donne un meilleur ajustement pour les données d'ERCA par rapport à la loi de Weibull généralisée.

11 Etude de la puissance des tests : Simulation

Dans ce paragraphe, nous étudions la puissance des 3 tests vu ci-dessus $Y_n^2(\hat{\theta}_{IG})$, $U_n^2(\hat{\theta}_{IG})$ et $M_n^2(\hat{\theta}_{IG})$ pour la distribution Gaussienne inverse comme hypothèse nulle contre les distributions alternatives LN , LL , EW , GW et BS .

Pour calculer la puissance du test, nous avons généré à partir de la distribution alternative des échantillons de taille $n = 200$ (par exemple) et pour un seuil $\alpha = 0.2$.

Nous avons procédé comme suit :

- (1) l'échantillon de taille n est généré à partir de la distribution alternative,
- (2) on calcule les valeurs des statistiques $Y_n^2(\hat{\theta}_{IG})$, $U_n^2(\hat{\theta}_{IG})$ et $M_n^2(\hat{\theta}_{IG})$,
- (3) pour le seuil en question, H_0 est rejetée si la valeur de la statistique dépasse la valeur critique,
- (4) les étapes 1-3 sont répétées 5000 fois,
- (5) la puissance est obtenue par compter le nombre de rejets de H_0 divisé par 5000.

Dans les figures 2.7, 2.8, 2.9, 2.10 et 2.11 nous représentons la puissance des 3 tests. Dans le cas des intervalles équiprobables, on voit clairement que le test $U_n^2(\hat{\theta}_{IG})$ possède une puissance faible pour toutes les distributions alternatives en question. En revanche, les tests $Y_n^2(\hat{\theta}_{IG})$ et $M_n^2(\hat{\theta}_{IG})$ sont les plus puissants pour toutes les distributions alternatives considérées en variant le nombre d'intervalles r . Il sera très intéressant si nous considérons dans cette étude les classes de *Neyman-Pearson* (Greenwood et Nikulin (1996)) et d'évaluer la puissance de ces tests.

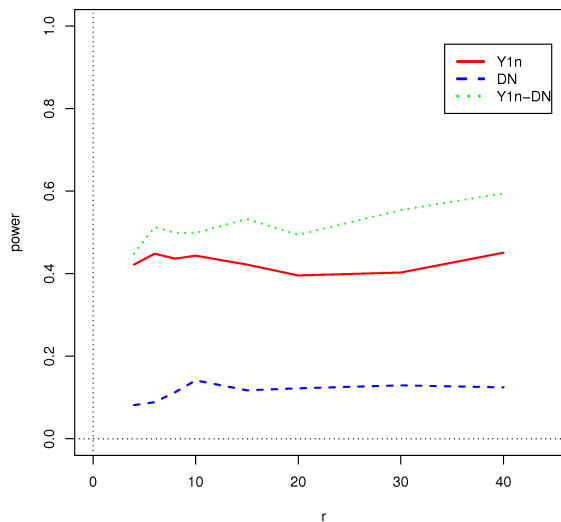


Figure 2.7 – Puissances estimées des 3 tests $Y_n^2(\hat{\theta}_n)$, $U_n^2(\hat{\theta}_n)$ et $M_n^2(\hat{\theta}_n)$ en fonction de r contre la distribution lognormale, taille $n = 200$, seuil $\alpha = 0.2$.

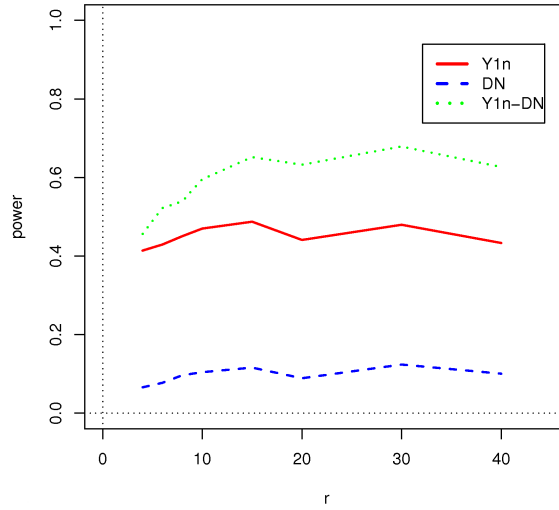


Figure 2.8 – Puissances estimées des 3 tests $Y_n^2(\hat{\theta}_n)$, $U_n^2(\hat{\theta}_n)$ et $M_n^2(\hat{\theta}_n)$ en fonction de r contre la distribution loglogistique , taille $n = 200$, seuil $\alpha = 0.2$.

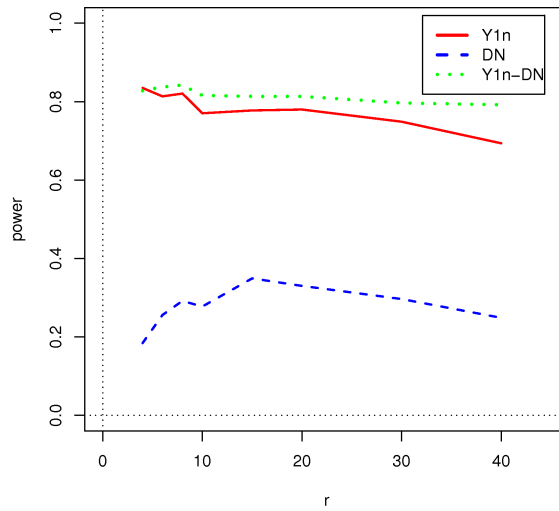


Figure 2.9 – Puissances estimées des 3 tests $Y_n^2(\hat{\theta}_n)$, $U_n^2(\hat{\theta}_n)$ et $M_n^2(\hat{\theta}_n)$ en fonction de r contre la distribution de Weibull exponentielle , taille $n = 200$, seuil $\alpha = 0.2$.

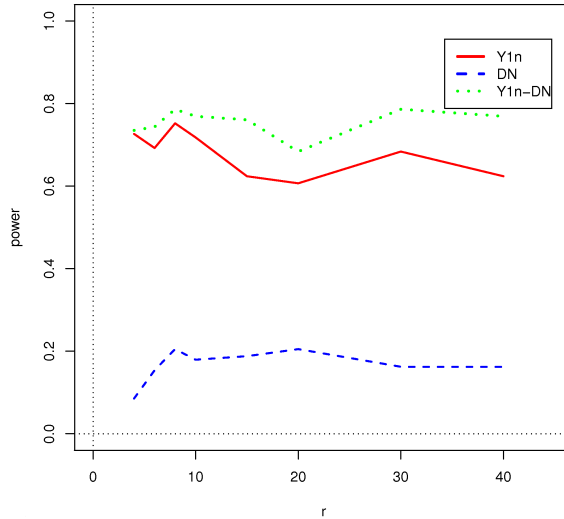


Figure 2.10 – Puissances estimées des 3 tests $Y_n^2(\hat{\theta}_n)$, $U_n^2(\hat{\theta}_n)$ et $M_n^2(\hat{\theta}_n)$ en fonction de r contre la distribution de Weibull généralisée, taille $n = 200$, seuil $\alpha = 0.2$.

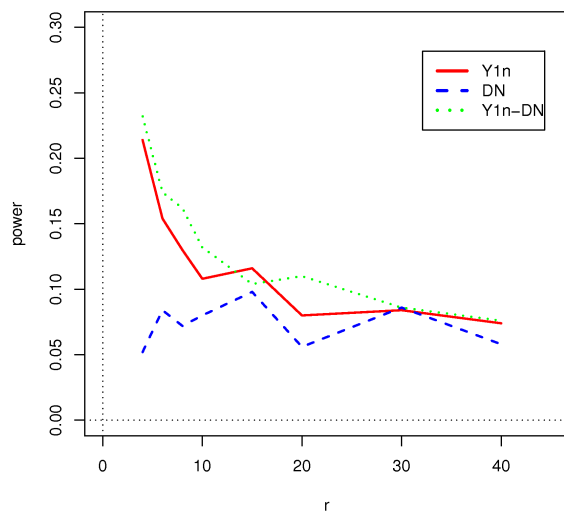


Figure 2.11 – Puissances estimées des 3 tests $Y_n^2(\hat{\theta}_n)$, $U_n^2(\hat{\theta}_n)$ et $M_n^2(\hat{\theta}_n)$ en fonction de r contre la distribution de Birnbaum-Saunders, taille $n = 200$, seuil $\alpha = 0.2$.

12 Test d'ajustement du Chi-deux pour les données censurées

12.1 Introduction

En fiabilité et analyse de survie, nous rencontrons souvent des observations incomplètes, et dans cette situation les méthodes habituelles ne sont plus valables. Dans le cas de censure aléatoire à droite,

plusieurs tests d'ajustement du type du Chi-deux ont été développés pour les hypothèses composites lorsque les paramètres sont estimés, les tests déjà publiés sont résumés dans le tableau 2.7.

Plusieurs tests d'ajustement ont été proposés pour les données censurées. Dans sa thèse, Chen (1975) avait construit un test du type Chi-deux. Habib et Thomas (1986) ont développé une statistique du type Chi-deux où l'estimateur de Kaplan-Meier $\hat{S}_n(x)$ est comparée à l'estimateur paramétrique $S(x, \hat{\theta}_n)$, où $\hat{\theta}_n$ est l'estimateur du maximum de vraisemblance de θ . Akritas (1988) a proposé une statistique du type Chi-deux basée sur l'idée de comparer le nombre de pannes observé et attendu dans chaque classe. Hjort (1990) a élaboré une statistique du type de Chi-deux pour tester la validité du modèle paramétrique pour les données de vie basée sur le processus du hasard cumulatif. Kim (1993) a également proposé un test du type de Chi-deux basé sur l'estimateur non paramétrique de type produit-limite. Nikulin et Solev (1999) ont construit un test de type du Chi-deux pour les données censurées doubles. Enfin, Bagdonavičius et Nikulin (2011) ont construit un test basé sur la statistique de NRR \hat{Y}_n^2 .

Pour les familles loglogistique, lognormale, de Weibull exponentielle, de Weibull généralisée et de Birnbaum-Saunders, les tests d'ajustement du Chi-deux en cas de données censurées ont été étudiés par Tahir (2012)

12.2 Test d'ajustement du Chi-deux pour la famille des lois Gaussiennes inverses

Considérons le problème de tester l'hypothèse composite H_0 pour laquelle les données proviennent de la distribution Gaussienne inverse. Nous développons en détail les deux tests d'ajustement de Habib et Thomas (1986) et de Bagdonavičius et Nikulin (2011).

Selon le modèle de censure aléatoire, on suppose que les temps de panne T_1, T_2, \dots, T_n sont non négatifs et indépendants, et les variables de censure C_1, C_2, \dots, C_n sont aussi non négatives et sont supposées aléatoires. On suppose aussi que C_i et T_i sont indépendantes.

Nous observons $X_i = \min(T_i, C_i)$ et les fonctions indicatrices δ_i définies par :

$$\delta_i = \begin{cases} 1 & , \text{ si } T_i \leq C_i \\ 0 & , \text{ sinon} \end{cases} .$$

Soit $S(t, \theta) = 1 - F(t, \theta)$, $\theta = (\mu, \lambda)^T$ est la fonction de survie de la distribution Gaussienne inverse, $f(t, \theta)$ est la fonction de densité correspondante de $F(t, \theta)$, $G(t)$ est la fonction de survie

2.12 Test d'ajustement du Chi-deux pour les données censurées

Année	Auteur(s)	Type
1986	M.G. Habib et D.R. Thomas	Chi-deux
1988	M. Akritas	Chi-deux
1990	N.L. Hjort	Chi-deux
1993	J.H. Kim	Chi-deux
1999	M. Nikulin et V.N. Solev	Chi-deux
2011	V. Bagdonavičius et M. Nikulin	Chi-deux

Tableau 2.7 – Quelques tests d'ajustement pour les données censurées aléatoires à droite.

inconnue et $g(t)$ est la fonction de densité correspondante de $G(t)$.

Nous supposons que la censure est non-informative, cela signifie que les fonctions $G(t)$ et $g(t)$ ne dépendent pas du paramètre θ .

la fonction de vraisemblance est :

$$L(\theta) = \prod_{i=1}^n f^{\delta_i}(X_i, \theta) S^{1-\delta_i}(X_i, \theta) G^{\delta_i}(X_i) g^{1-\delta_i}(X_i).$$

En faisant sortir les coefficients qui ne dépendent pas du paramètre θ , la fonction de vraisemblance pourra être présentée sous la forme suivante :

$$L(\theta) = \prod_{i=1}^n f^{\delta_i}(X_i, \theta) S^{1-\delta_i}(X_i, \theta),$$

la fonction de log-vraisemblance est :

$$\ell(\theta) = \sum_{i=1}^n \delta_i \ln f(X_i, \theta) + \sum_{i=1}^n (1 - \delta_i) \ln S(X_i, \theta). \quad (2.18)$$

Si on pose :

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

alors l'équation (2.18) devient :

$$\ell(\theta) = \sum_{i=1}^n \delta_i \ln h(X_i, \theta) + \sum_{i=1}^n \ln S(X_i, \theta).$$

12.3 Test de Habib et Thomas

Ce test est bien adapté pour les données censurées aléatoires à droite où l'estimateur de Kaplan-Meier $\hat{S}_n(x)$ est comparé à l'estimateur paramétrique $S(x, \hat{\theta}_n)$, où $\hat{\theta}_n$ est l'estimateur du maximum de vraisemblance de θ .

La fonction de log-vraisemblance est :

$$\begin{aligned} \ell(\mu, \lambda) = & \sum_{i=1}^n \delta_i \left\{ \frac{1}{2} \ln \lambda - \frac{3}{2} \ln X_i - \lambda \frac{(X_i - \mu)^2}{2\mu^2 X_i} \right\} + \\ & \sum_{i=1}^n (1 - \delta_i) \ln \left\{ \Phi(A(X_i)) - \exp\left(\frac{2\lambda}{\mu}\right) \Phi(B(X_i)) \right\}, \end{aligned} \quad (2.19)$$

où

$$A(X_i) = -\sqrt{\frac{\lambda}{X_i}} \left(\frac{X_i}{\mu} - 1 \right) \quad \text{et} \quad B(X_i) = -\sqrt{\frac{\lambda}{X_i}} \left(\frac{X_i}{\mu} + 1 \right).$$

Les fonctions de score $U_l(\mu, \lambda)$, $l = 1, 2$ sont :

$$\begin{aligned} U_1(\mu, \lambda) = & \frac{\partial \ell(\mu, \lambda)}{\partial \mu} = \frac{\lambda}{\mu^3} \sum_{i=1}^n \delta_i (X_i - \mu) + \\ & \frac{1}{\mu^2} \sum_{i=1}^n (1 - \delta_i) \frac{\sqrt{\lambda X_i} \varphi(A(X_i)) + \exp\left(\frac{2\lambda}{\mu}\right) (2\lambda \Phi(B(X_i)) - \sqrt{\lambda X_i} \varphi(B(X_i)))}{S(X_i, \mu, \lambda)}, \end{aligned}$$

$$\begin{aligned} U_2(\mu, \lambda) = & \frac{\partial \ell(\mu, \lambda)}{\partial \lambda} = \sum_{i=1}^n \delta_i \left(\frac{1}{2\lambda} - \frac{(X_i - \mu)^2}{2\mu^2 X_i} \right) + \\ & \sum_{i=1}^n (1 - \delta_i) \frac{\frac{1}{2\lambda} A(X_i) \varphi(A(X_i)) - \exp\left(\frac{2\lambda}{\mu}\right) \left(\frac{2}{\mu} \Phi(B(X_i)) + \frac{1}{2\lambda} B(X_i) \varphi(B(X_i)) \right)}{S(X_i, \mu, \lambda)}, \end{aligned}$$

Pour déterminer l'estimateur du maximum de vraisemblance $\hat{\theta}_n$ de θ , nous pouvons résoudre le système non linéaire formé par les fonctions de score en égalisant celles-ci à zéro.

2.12 Test d'ajustement du Chi-deux pour les données censurées

Habib et Thomas (1986) ont démontré que $Z_n(t) = \sqrt{n}(\hat{S}_n(t) - S(t, \theta))$ défini dans l'intervalle $[0, T]$ où $H(T)S(T, \theta) > 0$ converge faiblement vers un processus Gaussien $Z(t)$ de moyenne 0 avec

$$\text{Cov}(Z(t), Z(s)) = S(t, \theta)S(s, \theta) \int_0^s \frac{f(u, \theta)}{H(u)S^2(u, \theta)} du, \quad \text{pour tout } 0 \leq s \leq t \leq T.$$

On partage l'intervalle $[0, T]$ en r intervalles : I_1, I_2, \dots, I_r mutuellement disjoints par les points :

$$0 = t_0 < t_1 < \dots < t_{r-1} < t_r = T < +\infty.$$

Considérons le vecteur :

$$\hat{Z}_n = \sqrt{n}(\hat{S}_n - S_{\hat{\theta}_n}),$$

où

$$\hat{S}_n = (\hat{S}_n(t_1), \hat{S}_n(t_2), \dots, \hat{S}_n(t_{r-1}))^T \quad \text{and} \quad S_{\hat{\theta}_n} = (S(t_1, \hat{\theta}_1), S(t_2, \hat{\theta}_2), \dots, S(t_{r-1}, \hat{\theta}_n))^T.$$

Pour montrer la convergence faible de $\hat{Z}_n(t)$, Habib et Thomas (1986) ont fait les suppositions suivantes :

- 1) $f(t, \theta)$ and $F(t, \theta)$ sont deux fois continûment dérivables par rapport à θ ,
- 2) la matrice d'information de Fisher $I(\theta)$ est définie positive, et continue par rapport à θ , où

$$I_{ij} = - \int \frac{\partial^2 \ln f(t, \theta)}{\partial \theta_i \partial \theta_j} H(t) f(t, \theta) dt - \int \frac{\partial^2 \ln S(t, \theta)}{\partial \theta_i \partial \theta_j} h(t) S(t, \theta) dt, \quad i, j = 1, 2,$$

- 3) l'estimateur du Maximum de vraisemblance $\hat{\theta}_n$ de θ existe et \sqrt{n} -consistant avec

$$\sqrt{n}(\hat{\theta}_n - \theta) = I^{-1}W_n + o_p(1),$$

où

$$W_n = n^{-1/2} \sum_{i=1}^n \frac{\partial \ln g(X_i, \delta_i, \theta)}{\partial \theta},$$

et g est la densité jointe de la distribution de (X, δ) .

Soit

$$B = B(\theta) = \left[\frac{\partial F(t_i, \theta)}{\partial \theta_j} \right]_{(r-1) \times 2},$$

et

$$V = V(\theta) = [Cov(Z(t_i), Z(t_j))]_{(r-1) \times (r-1)},$$

Pour tester H_0 , Habib et Thomas ont construit le test de Pearson du Chi-deux modifié qui a la forme quadratique suivante :

$$\hat{Y}_n^2(\hat{\theta}_n) = \hat{Z}_n^T \hat{\Sigma}^{-}(\hat{\theta}_n) \hat{Z}_n, \quad (2.20)$$

où la matrice $\hat{\Sigma}$ est l'estimateur de la matrice de covariance Σ et Σ^{-} est son inverse généralisée, de sorte que :

$$\Sigma(\theta) = V(\theta) - B(\theta)I^{-1}(\theta)B^T(\theta), \quad \text{rang}\Sigma = r - 1.$$

Le comportement asymptotique de la statistique $\hat{Y}_n^2(\hat{\theta}_n)$ est donné par le théorème suivant :

Théorème 2.8. [Habib et Thomas (1986)] Sous les conditions de régularité ci-dessus, et pour n suffisamment grand :

$$\lim_{n \rightarrow \infty} \mathbf{P} \left(\hat{Y}_n^2(\hat{\theta}_n) \leq x | H_0 \right) = \mathbf{P}(\chi_r^2 \leq x).$$

Notons qu'à l'absence de la censure, la statistique \hat{Y}_n^2 se réduit à celle de NRR considérée précédemment. Nous pouvons aussi considérer le cas de données avec censure double (voir Nikulin et Solev (1999)).

12.4 Test de Bagdonavičius et Nikulin

Soit l'échantillon

$$(X_1, \delta_1), (X_2, \delta_2), \dots, (X_n, \delta_n), \quad (2.21)$$

où

$$X_i = T_i \wedge C_i, \quad \delta_i = \mathbf{1}_{\{T_i \leq C_i\}}.$$

2.12 Test d'ajustement du Chi-deux pour les données censurées

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

Notons par :

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

la fonction de hasard cumulative.

La fonction de log-vraisemblance est :

$$\ell(\theta) = \sum_{i=1}^n \delta_i \ln h(X_i, \theta) + \sum_{i=1}^n \ln S(X_i, \theta). \quad (2.22)$$

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

$$\dot{\ell}(\hat{\theta}) = \mathbf{0},$$

où

$$\dot{\ell}(\theta) = \frac{\partial}{\partial \theta} \ell(\theta) = \left(\frac{\partial}{\partial \theta_1} \ell(\theta), \dots, \frac{\partial}{\partial \theta_m} \ell(\theta) \right)^T. \quad (2.23)$$

La matrice d'information de Fisher est :

$$I(\theta) = -E_{\theta} \ddot{\ell}(\theta), \quad (2.24)$$

où

$$\ddot{\ell}(\theta) = \sum_{i=1}^n \delta_i \frac{\partial^2}{\partial \theta^2} \ln h(X_i, \theta) - \sum_{i=1}^n \frac{\partial^2}{\partial \theta^2} \Lambda(X_i, \theta). \quad (2.25)$$

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

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

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

L'échantillon (2.21) peut être représenté en termes de processus de comptage comme suit :

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

où

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

Posons :

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

Ecrivons les expressions des équations (2.22), (2.23), (2.24) et (2.25) en fonction des processus stochastiques N_i et Y_i . Les trajectoires du processus de comptage $N_i(t)$ ont la forme :

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

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

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

En utilisant ces processus, on obtient les relations suivantes :

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

et

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

Alors les équations (2.22), (2.23), (2.24) et (2.25) s'écrivent :

$$\ell(\theta) = \int_0^\tau \{\ln h(u, \theta) dN(u) - Y(u) h(u, \theta)\} du,$$

$$\dot{\ell}(\theta) = \int_0^\tau \frac{\partial}{\partial \theta} \ln h(u, \theta) \{dN(u) - Y(u) h(u, \theta)\} du,$$

$$\ddot{\ell}(\theta) = \int_0^\tau \frac{\partial^2}{\partial \theta^2} \ln h(u, \theta) dN(u) - \int_0^\tau \frac{\partial^2}{\partial \theta^2} h(u, \theta) Y(u) du.$$

La matrice d'information de Fisher est :

$$I(\theta) = -E_{\theta} \ddot{\ell}(\theta) = \mathbf{E}_{\theta} \sum_{i=1}^n \int_0^{\infty} \frac{\partial}{\partial \theta} \ln h(u, \theta) \left(\frac{\partial}{\partial \theta} \ln h(u, \theta) \right)^T h(u, \theta) Y_i(u) du.$$

Sous les conditions suivantes [Hjort (1990)] :

1. Il existe un voisinage Θ_0 de θ_0 la vraie valeur de θ telle que pour tout n et $\theta \in \Theta_0$, et pour tout $t \in [0, \tau]$, les dérivées partielles d'ordre 1, 2 et 3 de $\lambda(t, \theta)$ par rapport à θ existent et sont continues et bornées. On suppose aussi que la fonction de log-vraisemblance est trois fois différentiables par rapport à $\theta \in \theta_0$.
2. $\lambda(t, \theta)$ is bornée dans $[0, \tau] \times \Theta_0$.
3. Il existe une fonction positive $y(t)$ telle que :

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

4. La matrice $i(\theta_0) = \lim_{n \rightarrow \infty} \frac{I(\theta_0)}{n}$ (la limite existe sous les conditions 1-3) est définie positive.

Partageons l'intervalle $[0, \tau]$ en k sous-intervalles $I_j = (a_{j-1}, a_j]$, $a_0 = 0$, $a_k = \tau$, et notons par :

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

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

où $a \wedge b = \min(a, b)$.

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

$$Y_n^2 = Z^T \hat{V}^- Z. \tag{2.26}$$

où

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

est l'inverse généralisée de la matrice \hat{V} ,

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

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

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

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

Cette statistique peut s'écrire :

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

où

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

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

Choix de \hat{a}_j : Soit

$$b_i = (n - i)\Lambda(X_{(i)}, \hat{\theta}) + \sum_{l=1}^i \Lambda(X_{(l)}, \hat{\theta}).$$

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

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

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

d'où

$$\hat{a}_j = \Lambda^{-1} \left([E_j - \sum_{l=1}^{i-1} \Lambda(X_{(l)}, \hat{\theta})] / (n - i + 1), \hat{\theta} \right), \quad \hat{a}_k = \max(X_{(n)}, \tau),$$

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

Remarque 2.6. Nous voyons que les deux statistiques (2.20) et (2.26) sont semblables.

12.4.1 Application à la loi Gaussienne inverse

Supposons que sous H_0 , la distribution de T_1, T_2, \dots, T_n suit une loi Gaussienne inverse, de fonction de survie :

$$S(t, \theta) = 1 - F(t, \theta) = \Phi \left(-\sqrt{\frac{\lambda}{t}} \left(\frac{t}{\mu} - 1 \right) \right) - \exp\left(\frac{2\lambda}{\mu}\right) \Phi \left(-\sqrt{\frac{\lambda}{t}} \left(\frac{t}{\mu} + 1 \right) \right), \quad \theta = (\mu, \lambda)^T,$$

$$S(t, \theta) = 1 - F(t, \theta) = \Phi(A(t, \mu, \lambda)) - e^{\frac{2\lambda}{\mu}} \Phi(B(t, \mu, \lambda)),$$

où

$$A(t, \theta) = -\sqrt{\frac{\lambda}{t}} \left(\frac{t}{\mu} - 1 \right) \quad \text{et} \quad B(t, \theta) = -\sqrt{\frac{\lambda}{t}} \left(\frac{t}{\mu} + 1 \right).$$

Alors :

$$\begin{aligned}
 E_j &= \sum_{i: X_i > a_{j-1}} \ln \frac{S(a_{j-1}, \hat{\theta})}{S(a_j \wedge X_i, \hat{\theta})} \\
 &= \sum_{i: X_i > a_{j-1}} \ln \frac{\Phi\left(A(a_{j-1}, \hat{\theta})\right) - e^{\frac{2\lambda}{\mu}} \Phi\left(B(a_{j-1}, \hat{\theta})\right)}{\Phi\left(A(a_j \wedge X_i, \hat{\theta})\right) - e^{\frac{2\lambda}{\mu}} \Phi\left(B(a_j \wedge X_i, \hat{\theta})\right)},
 \end{aligned}$$

la fonction de hasard est :

$$h(t, \theta) = \frac{\left(\frac{\lambda}{2\pi t^3}\right)^{\frac{1}{2}} \exp\left\{-\frac{\lambda(t-\mu)^2}{2\mu^2 t}\right\}}{\Phi\left(-\sqrt{\frac{\lambda}{t}}\left(\frac{t}{\mu} - 1\right)\right) - \exp\left(\frac{2\lambda}{\mu}\right) \Phi\left(-\sqrt{\frac{\lambda}{t}}\left(\frac{t}{\mu} + 1\right)\right)},$$

Calculons les dérivées partielles de la fonction $\ln h$ par rapport à μ et λ :

$$\frac{\partial \ln h(t, \theta)}{\partial \mu} = \frac{\lambda(t-\mu)}{\mu^3} - \frac{1}{S(t, \theta)} \left\{ \frac{\sqrt{\lambda t}}{\mu^2} \varphi(A(t, \theta)) + \left(\frac{2\lambda}{\mu^2} \Phi(B(t, \theta)) - \frac{\sqrt{\lambda t}}{\mu^2} \varphi(B(t, \theta)) \right) \exp\left(\frac{2\lambda}{\mu}\right) \right\},$$

$$\begin{aligned}
 \frac{\partial \ln h(t, \theta)}{\partial \lambda} &= \frac{1}{2\lambda} - \frac{(t-\mu)^2}{2\mu^2 t} - \frac{1}{S(t, \theta)} \left\{ \frac{1}{2\lambda} A(t, \theta) \varphi(A(t, \theta)) - \right. \\
 &\quad \left. \left(\frac{2}{\mu} \Phi(B(t, \theta)) + \frac{1}{2\lambda} B(t, \mu, \lambda) \varphi(B(t, \theta)) \right) \exp\left(\frac{2\lambda}{\mu}\right) \right\}.
 \end{aligned}$$

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

$$\begin{aligned}
 \hat{C}_{1j} &= \frac{\hat{\lambda}}{n\hat{\mu}^3} \sum_{i: X_i \in I_j} \delta_i (X_i - \hat{\mu}) - \frac{\sqrt{\hat{\lambda}}}{n\hat{\mu}^2} \sum_{i: X_i \in I_j} \delta_i \frac{\sqrt{X_i} \varphi(A(X_i, \hat{\theta}))}{S(X_i, \hat{\theta})} - \\
 &\quad \frac{e^{\frac{2\hat{\lambda}}{\mu}}}{n\hat{\mu}^2} \sum_{i: X_i \in I_j} \delta_i \frac{2\hat{\lambda} \Phi(B(X_i, \hat{\theta})) - \sqrt{\hat{\lambda} X_i} \varphi(B(X_i, \hat{\theta}))}{S(X_i, \hat{\theta})},
 \end{aligned}$$

$$\hat{C}_{2j} = \frac{1}{2n\hat{\lambda}} \sum_{i: X_i \in I_j} \delta_i - \frac{1}{2n\hat{\mu}^2} \sum_{i: X_i \in I_j} \delta_i \frac{(X_i - \hat{\mu})^2}{X_i} - \frac{1}{2n\hat{\lambda}} \sum_{i: X_i \in I_j} \delta_i \frac{A(X_i, \hat{\theta})\varphi(A(X_i, \hat{\theta}))}{S(X_i, \hat{\theta})} +$$

$$\frac{e^{\frac{2\hat{\lambda}}{\hat{\mu}}}}{n} \sum_{i: X_i \in I_j} \delta_i \frac{\frac{2}{\hat{\mu}}\Phi(B(X_i, \hat{\theta})) + \frac{1}{2\hat{\lambda}}B(X_i, \hat{\theta})\varphi(B(X_i, \hat{\theta}))}{S(X_i, \hat{\theta})}.$$

Les éléments de la matrice de Fisher sont :

$$\hat{i} = \begin{pmatrix} \hat{i}_{11} & \hat{i}_{12} \\ \hat{i}_{21} & \hat{i}_{22} \end{pmatrix},$$

avec

$$\hat{i}_{11} = \frac{1}{n} \sum_{i=1}^n \delta_i \left[\frac{\hat{\lambda}(X_i - \hat{\mu})}{\hat{\mu}^3} - \frac{\sqrt{\hat{\lambda}X_i}}{\hat{\mu}^2 S(X_i, \hat{\theta})} \varphi(A(X_i, \hat{\theta})) - \frac{e^{\frac{2\hat{\lambda}}{\hat{\mu}}}}{S(X_i, \hat{\theta})} \left(\frac{2\hat{\lambda}}{\hat{\mu}^2} \Phi(B(X_i, \hat{\theta})) - \frac{\sqrt{\hat{\lambda}X_i}}{\hat{\mu}^2} \varphi(B(X_i, \hat{\theta})) \right) \right]^2,$$

$$\hat{i}_{12} = \frac{1}{n} \sum_{i=1}^n \delta_i \left[\frac{\hat{\lambda}(X_i - \hat{\mu})}{\hat{\mu}^3} - \frac{\sqrt{\hat{\lambda}X_i}}{S(X_i, \hat{\theta})\hat{\mu}^2} \varphi(A(X_i, \hat{\theta})) - \frac{e^{\frac{2\hat{\lambda}}{\hat{\mu}}}}{S(X_i, \hat{\theta})} \left(\frac{2\hat{\lambda}}{\hat{\mu}^2} \Phi(B(X_i, \hat{\theta})) - \frac{\sqrt{\hat{\lambda}X_i}}{\hat{\mu}^2} \varphi(B(X_i, \hat{\theta})) \right) \right] \times$$

$$\left[\frac{1}{2\hat{\lambda}} - \frac{(X_i - \hat{\mu})^2}{2\hat{\mu}^2 X_i} - \frac{1}{2\hat{\lambda}S(X_i, \hat{\theta})} A(X_i, \hat{\theta})\varphi(A(X_i, \hat{\theta})) + \frac{e^{\frac{2\hat{\lambda}}{\hat{\mu}}}}{S(X_i, \hat{\theta})} \left(\frac{2}{\hat{\mu}}\Phi(B(X_i, \hat{\theta})) + \frac{1}{2\hat{\lambda}}B(X_i, \hat{\theta})\varphi(B(X_i, \hat{\theta})) \right) \right] = \hat{i}_{21},$$

7, 34, 42, 63, 64, 74*, 83, 84, 91, 108, 112, 129, 133, 133, 139, 140, 140, 146, 149, 154, 157, 160, 160, 165, 173, 176, 185*, 218, 225, 241, 248, 273, 277, 279*, 297, 319*, 405, 417, 420, 440, 523*, 523, 583, 594, 1101, 1116*, 1146, 1226*, 1349*, 1412*, 1417.

Tableau 2.8 – Données d'Efron.

$$\hat{i}_{22} = \frac{1}{n} \sum_{i=1}^n \delta_i \left[\frac{1}{2\hat{\lambda}} - \frac{(X_i - \hat{\mu})^2}{2\hat{\mu}^2 t} - \frac{1}{2\hat{\lambda} S(X_i, \hat{\theta})} A(X_i, \hat{\theta}) \varphi(A(X_i, \hat{\theta})) + \frac{e^{\frac{2\hat{\lambda}}{\hat{\mu}}}}{S(X_i, \hat{\theta})} \left(\frac{2}{\hat{\mu}} \Phi(B(X_i, \hat{\theta})) + \frac{1}{2\hat{\lambda}} B(X_i, \hat{\theta}) \varphi(B(X_i, \hat{\theta})) \right) \right]^2.$$

12.4.2 Réanalyse des données de "Arm A" du cancer de la tête et du cou [Efron (1988)].

Les durées de survie (en jours) des ($n = 51$) patients sont données au tableau (2.8), où * représente la censure.

En transformant ces données en mois, Efron (1988) a estimé la fonction de hasard pour ces données en montrant qu'elle est unimodale. Mudholkar et al. (1995) ont montré que distribution de Weibull Exponentielle peut ajuster ces données. Nikulin et Haghighi (2004,2009) ont montré en utilisant le test d'Akritis que la distribution de Weibull généralisée donne un bon ajustement pour ces données.

Pour la loi Gaussienne inverse, l'estimateur $\hat{\theta}_n$ de θ maximisant la fonction de log-vraisemblance (2.19) est :

$$\hat{\theta} = (\hat{\mu}, \hat{\lambda})^T = (19.29642, 4.758374)^T, \quad \log\text{-vrais} = -3.035701.$$

Pour un nombre de classes $k = 5$, Nous trouvons alors :

j	1	2	3	4	5
\hat{a}_j	2.074719	3.606261	6.120298	12.584717	τ
U_j	4	5	16	7	10
e_j	8.778149	8.778149	8.778149	8.778149	8.778149

2.12 Test d'ajustement du Chi-deux pour les données censurées

Après calcul de la statistique de NRR, nous trouvons $Y_n^2 = 17.29910$. Au seuil $\alpha = 0.05$, et pour $k = 5$, la valeur critique $\chi_k^2 = 11.07050$, d'où H_0 n'est acceptée.

Chapitre 3

Le modèle de Régression AFT.

1 Introduction

Dans le domaine industriel (ou médical), l'étude de la fiabilité de composants et de systèmes nécessite l'observation des temps de panne des unités qui fonctionnent sous différentes variables explicatives, appelées souvent stress, telles que la température, le voltage, la pression, l'humidité, etc. Pour les systèmes hautement fiables, l'observation des pannes des unités sont rares et peuvent durer très longtemps. Pour estimer la fonction de fiabilité sous les stress usuels, il devient nécessaire d'effectuer des *expériences accélérées* qui permettent de raccourcir la durée de vie du système en lui appliquant des stress élevés afin d'observer ses pannes dans un temps raisonnable, ce qui constitue l'objet des méthodes de tests de vie accélérés (*Accelerated Life Testing (ALT)*) dont le modèle le plus fréquemment utilisé est le modèle AFT (*Accelerated Failure Time*) [Bagdonavičius et Nikulin (2002), Meeker (1998), Nikulin et al. (2007)] ou le modèle de temps de panne accéléré.

Plusieurs modèles AFT paramétriques ont été étudiés : (voir par exemple : Bagdonavičius et Nikulin (2002)). Le modèle AFT basé sur la distribution de Birnbaum-Saunders a été étudié par Tahir (2012). Puisque la distribution du temps de premier passage d'un processus Brownien avec une dérive positive suit une loi Gaussienne inverse, l'utilisation de cette loi dans les modèles de durée de vie est particulièrement intéressant [Chhikara et Folks (1989)]. Ces modèles sont utilisés aujourd'hui non seulement en fiabilité et analyse de survie, mais aussi en démographie, dynamique des populations, gérontologie, biologie, génétique, radiobiologie, biophysique, problèmes de la longévité, vieillissement, dégradation, etc.

Dans ce chapitre nous nous intéressons à étudier le modèle AFT basé sur la loi Gaussienne inverse et son application pour la construction de systèmes redondants.

2 Le modèle AFT

2.1 Introduction

Considérons le cas où la fonction de hasard présente la forme \cap (ou unimodale) et l'application de la distribution Gaussienne inverse comme une fonction de survie de base dans le modèle AFT.

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

$$E = \left\{ x(\cdot) = (x_1(\cdot), x_2(\cdot), \dots, x_m(\cdot))^T : (0, \infty) \rightarrow \mathbf{R}^m, x(\cdot) \text{ est déterministe} \right\}$$

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

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

$$S_{x(\cdot)}(t) = \mathbf{P} \left(T_{x(\cdot)} \geq t \right) = 1 - F_{x(\cdot)}(t), \quad x(\cdot) \in E.$$

La fonction de hasard de $T_{x(\cdot)}$ sous $x(\cdot)$ est :

$$\lambda_{x(\cdot)}(t) = \lim_{h \rightarrow 0} \frac{1}{h} \mathbf{P} \left(t \leq T_{x(\cdot)} < t + h \mid T_{x(\cdot)} \geq t \right) = - \frac{S'_{x(\cdot)}(t)}{S_{x(\cdot)}(t)}.$$

Notons par :

$$\Lambda_{x(\cdot)}(t) = \int_0^t \lambda_{x(\cdot)}(u) du = - \ln(S_{x(\cdot)}(t)),$$

la fonction de hasard cummulative de $T_{x(\cdot)}$.

Nous disons que le stress $x(\cdot)$ est supérieur à un stress $y(\cdot)$ et nous écrivons $x(\cdot) > y(\cdot)$, si pour tout $t \geq 0$ l'inégalité $S_{y(\cdot)}(t) \geq S_{x(\cdot)}(t)$ est vérifiée et il existe $t_0 > 0$ tel que $S_{y(\cdot)}(t_0) > S_{x(\cdot)}(t_0)$. En ALT les types de stress les plus utilisés sont les stress constants, stress en escalier, stress croissants (monotones), stress cycliques et les stress aléatoires (voir e.g. Nelson (1990), Meeker et Escobar (1998), Lawless (2003)). Le cas le plus commun est lorsque le stress est unidimensionnel. Les stress variant dans le temps les plus fréquemment utilisés sont les stress en escalier : les unités sont mises au banc d'essai à un stress initial faible et si elles ne tombent pas en panne dans un temps prédéterminé t_1 , le stress est augmenté. Si les unités ne tombent pas encore en panne dans un temps prédéterminé $t_2 > t_1$, le stress est augmenté à nouveau, et ainsi de suite.

Notons par E_m le sous-ensemble de E des stress en escalier à m niveaux de la forme :

$$x(t) = \begin{cases} x_1, & 0 \leq t < t_1 \\ x_2, & t_1 \leq t < t_2 \\ \vdots & \\ x_m, & t_{m-1} \leq t < t_m = +\infty \end{cases}, \quad (3.1)$$

où x_1, x_2, \dots, x_k appartiennent à E_1 . Le sous-ensemble E_2 de E prend la forme :

$$x(t) = \begin{cases} x_1, & 0 \leq t < t_1 \\ x_2, & t \geq t_1 \end{cases}. \quad (3.2)$$

2.2 Le principe du modèle de Sedyakin

En 1966, Sedyakin a formulé son fameux principe physique de la fiabilité, qui suppose que pour deux populations d'unités fonctionnant sous différents stress constants x_1 et x_2 , les deux instants t_1 et t_2 sont équivalents si les probabilités de survie jusqu'à ces moments sont égales :

$$\mathbf{P}(T_{x_1} \geq t_1) = S_{x_1}(t_1) = S_{x_2}(t_2) = \mathbf{P}(T_{x_2} \geq t_2), \quad x_1, x_2 \in E_1.$$

Si après ces moments équivalents les unités de deux groupes sont observées sous le même stress x_2 , i.e. la première population est observée sous le stress en escalier $x(\cdot) \in E_2$ de la forme (3.2) et la seconde population est observée tout le temps sous le stress constant x_2 , alors pour tout s , on a :

$$\lambda_{x(\cdot)}(t_1 + s) = \lambda_{x_2}(t_2 + s).$$

En utilisant cette idée de Sedyakin, il a été proposé dans Bagdonavičius et Nikulin (2002) une généralisation du modèle de Sedyakin (*GSM*) au cas des stress variant dans le temps en supposant que la fonction de hasard $\lambda_{x(\cdot)}(t)$ à tout instant t est une fonction du stress à cet instant et de la probabilité de survie jusqu'à cet instant :

$$\lambda_{x(\cdot)}(t) = g(x(t), \Lambda_{x(\cdot)}(t)), \quad x(\cdot) \in E,$$

telle que g est une fonction positive sur $E \times \mathbf{R}^+$.

Il est facile de montrer que si le modèle *GSM* est réalisé sur E_2 , alors la fonction de hasard sous le stress $x(\cdot) \in E_2$, satisfait l'égalité :

$$\lambda_{x(\cdot)}(t) = \begin{cases} \lambda_{x_1}(t), & 0 \leq t < t_1 \\ \lambda_{x_2}(t - t_1 + t_1^*), & t \geq t_1 \end{cases},$$

où le moment t_1^* est déterminé par la relation $S_{x_1}(t_1) = S_{x_2}(t_1^*)$. Cette dernière relation implique que :

$$\lambda_{x(\cdot)}(t_1 + s) = \lambda_{x_2}(t_1^* + s), \quad \forall x(\cdot) \in E, \quad \forall s \geq 0.$$

La fonction de survie sous le stress $x(\cdot) \in E_2$ satisfait :

$$S_{x(\cdot)}(t) = \begin{cases} S_{x_1}(t), & 0 \leq t < t_1 \\ S_{x_2}(t - t_1 + t_1^*), & t \geq t_1 \end{cases}.$$

Nous utilisons par exemple le modèle de Sedyakin pour construire les systèmes redondants.

2.3 Définition du modèle AFT

Le modèle AFT est défini sur E , si et seulement si il existe sur E une fonction positive r et sur $[0, +\infty)$ une fonction de survie de base S_0 telles que :

$$S_{x(\cdot)}(t) = S_0 \left(\int_0^t r(x(u)) du \right), \quad \forall x(\cdot) \in E. \quad (3.3)$$

Sur l'ensemble E_1 , l'équation (3.3) devient :

$$S_x(t) = S_0(r(x)t), \quad \forall x \in E_1,$$

et sur E_m , l'équation (3.3) s'écrit [Bagdonavičius et Nikulin (2002)] :

$$S_x(t) = S_0 \left\{ \sum_{j=1}^{i-1} r(x_j)(t_j - t_{j-1}) + r(x_i)(t - t_{i-1}) \right\} =$$

$$S_{x_i} \left\{ t - t_{i-1} + \frac{1}{r(x_i)} \sum_{j=1}^{i-1} r(x_j)(t_j - t_{j-1}) \right\}, \quad t \in [t_{i-1}, t_i], \quad i = 1, 2, \dots, m.$$

Si la fonction $r(x)$ et S_0 sont inconnues, on trouve le modèle AFT non-paramétrique. Si la fonction S_0 est complètement inconnue, on aura le modèle AFT semi-paramétrique. Si la fonction S_0 est connue et que $r(x)$ peut être paramétrisée, dans ce cas on aura le modèle AFT paramétrique. Pour l'analyse semi-paramétrique et paramétrique du modèle AFT, on peut voir Bagdonavičius et Nikulin (2002),

Bagdonavičius et al. (2011).

Pour les modèles AFT paramétrique, les estimateurs des paramètres peuvent être obtenus quasiment par toutes les méthodes standards pour tous les plans d'expérience. Dans la pratique la fonction de survie de base S_0 est tirée de certaines familles de distributions paramétriques, telles que Weibull (Weibull), Weibull généralisée (GW), log-normale (LN), log-logistique (LL), Gaussienne inverse (IG), Birnbaum-Saunders (BS), etc, et dans ce cas on obtient respectivement les modèles AFT paramétrique : AFT-Weibull, AFT-GW, AFT-LN, AFT-LL, AFT-IG, AFT-BS, etc.

La fonction $r(x)$ est souvent paramétrisée comme suit :

$$r(x) = e^{-\beta_0 - \beta_1 \psi(x)},$$

où

ψ est une fonction donnée de x , cette fonction peut prendre plusieurs formes :

- • $r(x) = e^{-\beta_0 - \beta_1 x}$, $\psi(x) = x$, c'est le modèle *log-linear*.
- • $r(x) = e^{-\beta_0 - \beta_1 \ln x}$, $\psi(x) = \ln(x)$, c'est le modèle *power-rule*.
- • $r(x) = e^{-\beta_0 - \beta_1/x}$, $\psi(x) = 1/x$, c'est le modèle d'*Arrhenius*.
- • $r(x) = e^{-\beta_0 - \beta_1 \ln \frac{x}{1-x}}$, $0 < x < 1$, $\psi(x) = \ln \frac{x}{1-x}$, c'est le modèle de *Meeker-Luvalle*.

Généralement, nous prenons :

$$r(x) = e^{-\beta^T x},$$

où $\beta = (\beta_0, \dots, \beta_m)^T$ est le vecteur de régression inconnu. Dans ce cas, le modèle AFT paramétrique (3.3) est donné par la formule :

$$S_{x(\cdot)}(t) = S_0 \left(\int_0^t e^{-\beta^T x(u)} du \right), \quad x(\cdot) \in E. \quad (3.4)$$

2.4 Plan d'expériences

Comme on a mentionné, le but des tests de vies accélérés (ALT) est de donner l'estimateur de la fonction de fiabilité $S_{x(0)}$ sous le stress usuel $x^{(0)}$, en utilisant les données d'expériences accélérées lorsque les unités sont testées sous stress supérieur au stress usuel.

Plusieurs plans d'expériences sont utilisés en ALT. Nous pouvons citer :

- *Premier plan d'expériences*

Notons par $x^{(0)} = (x_{00}, x_{01}, \dots, x_{0m})$, $x_{00} = 1$, le stress usuel. Généralement nous prenons $m = 1$ (stress unidimensionnel), ou $m = 2$ (stress bidimensionnel).

Soient $x^{(1)}, \dots, x^{(k)}$ des stress accélérés constants dans le temps tels que :

$$x^{(0)} < x^{(1)} < \dots < x^{(k)},$$

avec $x_i = (x_{i0}, x_{i1}, \dots, x_{im})^T \in E_m$, $x_{i0} = 1$. Le stress usuel $x^{(0)}$ n'est pas utilisée durant l'expérience. Selon le premier plan d'expériences k groupes d'unités sont testés. Le i^{eme} groupe de n_i unités, $\sum_{i=1}^k n_i = n$, est testé sous le stress $x^{(i)}$. Les données peuvent être complètes ou censurées à droite.

- *Deuxième plan d'expériences*

n unités sont testées sous des stress en escalier

$$x(t) = \begin{cases} x^{(1)}, & 0 \leq t < t_1 \\ x^{(2)}, & t_1 \leq t < t_2 \\ \vdots \\ x^{(m)}, & t_{m-1} \leq t < t_m = +\infty \end{cases}, \quad (3.5)$$

où

$$x^{(j)} = (x_{j0}, x_{j1}, \dots, x_{jm})^T \in E_m, x_{j0} = 1.$$

Si la fonction $r(x)$ est paramétrisée comme suit :

$$r(x) = e^{-\beta^T x},$$

la fonction de survie de l'équation (3.4) sous le stress $x(\cdot)$ de la forme (3.5) peut s'écrire :

$$S_{x(\cdot)}(t) = S_0 \left\{ 1_{\{i>1\}} \sum_{j=1}^{i-1} e^{-\beta^T x_j (t_j - t_{j-1})} + e^{-\beta^T x_i (t - t_{i-1})} \right\}, \quad t \in [t_{i-1}, t_i[, \quad i = 1, \dots, m.$$

2.5 Estimation dans les modèles AFT-LN et AFT-WG

Les modèles AFT-LN et AFT-GW ont été bien étudiés par Bagdonavičius et Nikulin (2002). Bagdonavičius et al. (2008) ont réalisé une étude en comparant les modèles AFT-Weibull, AFT-LN et AF-GW, et ils ont constaté que dans le cas où la fonction de hasard est en forme de baignoire, le modèle AFT-GW est nettement plus approprié pour l'estimation de la fonction de survie sous le stress usuel en faisant des expériences accélérées.

Pour les modèles AFT-LN et AFT-GW, les fonctions de survie de base S_0 sont respectivement :

$$S_0(t) = 1 - \Phi(\nu \ln t), \quad \nu > 0, \text{ (LN)}$$

$$S_0(t) = \exp\{1 - (1 + t^\nu)^{\frac{1}{\gamma}}\}, \quad \nu, \gamma > 0, \text{ (GW)}.$$

Pour notre situation, la fonction de hasard est unimodale, il suit nécessairement que $\gamma > \nu > 1$ pour la distribution de Weibull généralisée.

2.6 Estimation dans le modèle AFT-IG

Nous proposons l'étude du modèle paramétrique AFT-IG, et dans ce cas la fonction de survie de base $S_0(t)$ appartient à la famille des lois Gaussiennes inverses, i.e. :

$$S_0(t) = \Phi\left(-\sqrt{\frac{\lambda}{t}}\left(\frac{t}{\mu} - 1\right)\right) - \exp\left(\frac{2\lambda}{\mu}\right)\Phi\left(-\sqrt{\frac{\lambda}{t}}\left(\frac{t}{\mu} + 1\right)\right), \quad t > 0, \quad \mu, \lambda > 0,$$

la fonction de hasard est :

$$h(t, \theta) = \frac{\left(\frac{\lambda}{2\pi t^3}\right)^{\frac{1}{2}} \exp\left\{-\frac{\lambda(t-\mu)^2}{2\mu^2 t}\right\}}{\Phi\left(-\sqrt{\frac{\lambda}{t}}\left(\frac{t}{\mu} - 1\right)\right) - \exp\left(\frac{2\lambda}{\mu}\right)\Phi\left(-\sqrt{\frac{\lambda}{t}}\left(\frac{t}{\mu} + 1\right)\right)}, \quad \theta = (\mu, \lambda)^T.$$

Pour estimer les paramètres du modèle AFT-IG, nous considérons le premier plan d'expérience. Soit t_i est la durée maximale d'expérience pour le i^{eme} groupe. Soient $\beta = (\beta_0, \beta_1, \dots, \beta_m)^T$ les

Chapitre 3. Le modèle de Régression AFT.

paramètres de régression. la durée de vie de la j^{eme} unité du i^{eme} groupe est notée par T_{ij} . Soit $X_{ij} = T_{ij} \wedge t_i$ et δ_{ij} est la fonction indicatrice définie par :

$$\delta_{ij} = \begin{cases} 1 & , \text{ si } T_{ij} < t_i \\ 0 & , \text{ sinon} \end{cases} .$$

Après calcul, nous obtenons la fonction de vraisemblance $L(\beta, \mu, \lambda)$ [Saaidia et al. (2010)] :

$$\begin{aligned} L &= \prod_{i=1}^k \prod_{j=1}^{n_i} \left\{ \frac{e^{-\beta T_{ij}} \left(\frac{\lambda}{2\pi X_{ij}^3 e^{-3\beta T_{ij}}} \right)^{\frac{1}{2}} \exp\left\{-\frac{\lambda(X_{ij}e^{-\beta T_{ij}} - \mu)^2}{2\mu^2 X_{ij} e^{-\beta T_{ij}}}\right\}}{\Phi\left(-\sqrt{\frac{\lambda}{X_{ij}e^{-\beta T_{ij}}}} \left(\frac{1}{\mu} X_{ij} e^{-\beta T_{ij}} - 1\right)\right) - e^{\frac{2\lambda}{\mu}} \Phi\left(-\sqrt{\frac{\lambda}{X_{ij}e^{-\beta T_{ij}}}} \left(\frac{1}{\mu} X_{ij} e^{-\beta T_{ij}} + 1\right)\right)} \right\}^{\delta_{ij}} \times \\ &\left\{ \Phi\left(-\sqrt{\frac{\lambda}{X_{ij}e^{-\beta T_{ij}}}} \left(\frac{1}{\mu} X_{ij} e^{-\beta T_{ij}} - 1\right)\right) - e^{\frac{2\lambda}{\mu}} \Phi\left(-\sqrt{\frac{\lambda}{X_{ij}e^{-\beta T_{ij}}}} \left(\frac{1}{\mu} X_{ij} e^{-\beta T_{ij}} + 1\right)\right) \right\} \\ &= \prod_{i=1}^k \prod_{j=1}^{n_i} \left\{ e^{-\beta T_{ij}} \left(\frac{\lambda}{2\pi X_{ij}^3 e^{-3\beta T_{ij}}} \right)^{\frac{1}{2}} \exp\left\{-\frac{\lambda(X_{ij}e^{-\beta T_{ij}} - \mu)^2}{2\mu^2 X_{ij} e^{-\beta T_{ij}}}\right\} \right\}^{\delta_{ij}} \times \\ &\left\{ \Phi\left(-\sqrt{\frac{\lambda}{X_{ij}e^{-\beta T_{ij}}}} \left(\frac{1}{\mu} X_{ij} e^{-\beta T_{ij}} - 1\right)\right) - e^{\frac{2\lambda}{\mu}} \Phi\left(-\sqrt{\frac{\lambda}{X_{ij}e^{-\beta T_{ij}}}} \left(\frac{1}{\mu} X_{ij} e^{-\beta T_{ij}} + 1\right)\right) \right\}^{1-\delta_{ij}} . \end{aligned}$$

Notons $K_{ij} = X_{ij}e^{-\beta T_{ij}}$, $A_{ij} = -\sqrt{\frac{\lambda}{K_{ij}}} \left(\frac{K_{ij}}{\mu} - 1\right)$ et $B_{ij} = -\sqrt{\frac{\lambda}{K_{ij}}} \left(\frac{K_{ij}}{\mu} + 1\right)$,

alors la fonction de vraisemblance devient :

$$L(\beta, \mu, \lambda) = \prod_{i=1}^k \prod_{j=1}^{n_i} \left\{ \frac{e^{-\beta T_{ij}} \left(\frac{\lambda}{2\pi K_{ij}^3} \right)^{\frac{1}{2}} \exp\left\{-\frac{\lambda(K_{ij}-\mu)^2}{2\mu^2 K_{ij}}\right\}}{\Phi(A_{ij}) - e^{\frac{2\lambda}{\mu}} \Phi(B_{ij})} \right\}^{\delta_{ij}} \left\{ \Phi(A_{ij}) - e^{\frac{2\lambda}{\mu}} \Phi(B_{ij}) \right\},$$

$$= \prod_{i=1}^k \prod_{j=1}^{n_i} \left\{ e^{-\beta^T x^{(i)}} \left(\frac{\lambda}{2\pi K_{ij}^3} \right)^{\frac{1}{2}} \exp \left\{ -\frac{\lambda(K_{ij} - \mu)^2}{2\mu^2 K_{ij}} \right\} \right\}^{\delta_{ij}} \left\{ \Phi(A_{ij}) - e^{\frac{2\lambda}{\mu}} \Phi(B_{ij}) \right\}^{1-\delta_{ij}},$$

et la fonction de log-vraisemblance $\ell(\beta, \mu, \lambda)$ est :

$$\begin{aligned} \ell(\beta, \mu, \lambda) = & \sum_{i=1}^k \sum_{j=1}^{n_i} \delta_{ij} \left\{ -\beta^T x^{(i)} + \frac{1}{2} \ln \lambda - \frac{3}{2} \ln K_{ij} - \lambda \frac{(K_{ij} - \mu)^2}{2\mu^2 K_{ij}} \right\} + \\ & \sum_{i=1}^k \sum_{j=1}^{n_i} (1 - \delta_{ij}) \ln \left\{ \Phi(A_{ij}) - e^{\frac{2\lambda}{\mu}} \Phi(B_{ij}) \right\}. \end{aligned} \quad (3.6)$$

Nous remarquons que la fonction de log-vraisemblance est très complexe.

Calculons les dérivées partielles de la fonction de log-vraisemblance $\ell(\beta, \mu, \lambda)$ (3.6) par rapport à β_l ($l = 0, \dots, m$), μ et λ . Nous obtenons les fonctions de score (ou gradient) $U_l(\beta, \mu, \lambda)$, $l = 0, \dots, m + 2$ [Saaidia et al. (2010)] :

$$\begin{aligned} U_l(\beta, \mu, \lambda) = & \frac{\partial \ell(\beta, \mu, \lambda)}{\partial \beta_l} = \frac{1}{2} \sum_{i=1}^k \sum_{j=1}^{n_i} \delta_{ij} x_{il} \left\{ 1 + \frac{\lambda(K_{ij}^2 - \mu^2)}{\mu^2 K_{ij}} \right\} - \\ & \frac{1}{2} \sum_{i=1}^k \sum_{j=1}^{n_i} (1 - \delta_{ij}) x_{il} \left\{ \frac{B_{ij} \varphi(A_{ij}) - e^{\frac{2\lambda}{\mu}} A_{ij} \varphi(B_{ij})}{S_{ij}} \right\}, \quad l = 0, \dots, m, \end{aligned} \quad (3.7)$$

$$\begin{aligned} U_{m+1}(\beta, \mu, \lambda) = & \frac{\partial \ell(\beta, \mu, \lambda)}{\partial \mu} = \frac{\lambda}{\mu^3} \sum_{i=1}^k \sum_{j=1}^{n_i} \delta_{ij} (K_{ij} - \mu) + \\ & \frac{1}{\mu^2} \sum_{i=1}^k \sum_{j=1}^{n_i} (1 - \delta_{ij}) \frac{\sqrt{\lambda K_{ij}} \varphi(A_{ij}) + e^{\frac{2\lambda}{\mu}} (2\lambda \Phi(B_{ij}) - \sqrt{\lambda K_{ij}} \varphi(B_{ij}))}{S_{ij}}, \end{aligned} \quad (3.8)$$

$$U_{m+2}(\beta, \mu, \lambda) = \frac{\partial \ell(\beta, \mu, \lambda)}{\partial \lambda} = \sum_{i=1}^k \sum_{j=1}^{n_i} \delta_{ij} \left(\frac{1}{2\lambda} - \frac{(K_{ij} - \mu)^2}{2\mu^2 K_{ij}} \right) +$$

$$\sum_{i=1}^k \sum_{j=1}^{n_i} (1 - \delta_{ij}) \frac{\frac{1}{2\lambda} A_{ij} \varphi(A_{ij}) - e^{\frac{2\lambda}{\mu}} \left(\frac{2}{\mu} \Phi(B_{ij}) + \frac{1}{2\lambda} B_{ij} \varphi(B_{ij}) \right)}{S_{ij}}, \quad (3.9)$$

où

$$S_{ij} = \Phi(A_{ij}) - e^{\frac{2\lambda}{\mu}} \Phi(B_{ij}).$$

Pour calculer l'estimateur du maximum de vraisemblance $(\hat{\beta}, \hat{\mu}, \hat{\lambda})$ de (β, μ, λ) , nous pouvons résoudre le système non linéaire formé en égalisant les fonctions de score (3.7), (3.8) et (3.9) à zéro.

L'estimateur de la fonction de survie sous le stress usuel $x^{(0)}$ est

$$\hat{S}_{x^{(0)}}(t) = \Phi \left(-\sqrt{\frac{\hat{\lambda}}{te^{-\hat{\beta}T x^{(0)}}}} \left(\frac{te^{-\hat{\beta}T x^{(0)}}}{\hat{\mu}} - 1 \right) \right) - e^{\frac{2\hat{\lambda}}{\hat{\mu}}} \Phi \left(-\sqrt{\frac{\hat{\lambda}}{te^{-\hat{\beta}T x^{(0)}}}} \left(\frac{te^{-\hat{\beta}T x^{(0)}}}{\hat{\mu}} + 1 \right) \right).$$

Les intervalles de confiance de $S_{x^{(0)}}(t)$ au seuil α sont

$$\left(1 + \frac{1 - \hat{S}_{x^{(0)}}(t)}{\hat{S}_{x^{(0)}}(t)} e^{\pm \hat{\sigma}_{Q_{x^{(0)}}} \omega_{1-\frac{\alpha}{2}}} \right)^{-1},$$

où

$$\hat{\sigma}_{Q_{x^{(0)}}} = \frac{1}{(\hat{S}_{x^{(0)}}(t))^2 (1 - \hat{S}_{x^{(0)}}(t))^2} \sum_{k=0}^{m+2} \sum_{j=0}^{m+2} a_k(t, \hat{\beta}, \hat{\mu}, \hat{\lambda}) I^{kj}(\hat{\beta}, \hat{\mu}, \hat{\lambda}) a_j(t, \hat{\beta}, \hat{\mu}, \hat{\lambda}),$$

$$a_k(t, \hat{\beta}, \hat{\mu}, \hat{\lambda}) = \frac{1}{2} x_k \left(\hat{B}_t \varphi(\hat{A}_t) + e^{2\hat{\lambda}/\hat{\mu}} \hat{A}_t \varphi(\hat{B}_t) \right), \quad k = 0, \dots, m,$$

$$a_{m+1}(t, \hat{\beta}, \hat{\mu}, \hat{\lambda}) = \frac{1}{\hat{\mu}^2} \left\{ \sqrt{\hat{\lambda} \hat{K}_t} \left(\varphi(\hat{A}_t) - e^{2\hat{\lambda}/\hat{\mu}} \varphi(\hat{B}_t) \right) + e^{2\hat{\lambda}/\hat{\mu}} \Phi(\hat{A}_t) \right\},$$

$$a_{m+2}(t, \hat{\beta}, \hat{\mu}, \hat{\lambda}) = \frac{1}{2\hat{\lambda}} \left(\hat{A}_t \varphi(\hat{A}_t) - e^{2\hat{\lambda}/\hat{\mu}} \hat{B}_t \varphi(\hat{B}_t) \right) - \frac{2}{\hat{\mu}} e^{2\hat{\lambda}/\hat{\mu}} \Phi(\hat{B}_t),$$

avec

$$\hat{A}_t = -\sqrt{\frac{\hat{\lambda}}{te^{-\hat{\beta}T x^{(0)}}}} \left(\frac{te^{-\hat{\beta}T x^{(0)}}}{\hat{\mu}} - 1 \right) \quad \text{et} \quad \hat{B}_t = -\sqrt{\frac{\hat{\lambda}}{te^{-\hat{\beta}T x^{(0)}}}} \left(\frac{te^{-\hat{\beta}T x^{(0)}}}{\hat{\mu}} + 1 \right),$$

$\omega_{1-\frac{\alpha}{2}}$ est le $(1 - \frac{\alpha}{2})$ -quantile de $N(0, 1)$ et $I^{kj}(\hat{\beta}, \hat{\mu}, \hat{\lambda})$ sont les éléments de la matrice $I^{-1}(\hat{\beta}, \hat{\mu}, \hat{\lambda})$.

La matrice d'information de Fisher est :

$$I(\beta, \mu, \lambda) = -E \left(\frac{\partial^2 \ell(\beta, \mu, \lambda)}{\partial \theta_i \theta_j} \right), \quad \theta = (\beta_0, \dots, \beta_m, \mu, \lambda)^T,$$

ou encore

$$I(\beta, \mu, \lambda) = \begin{pmatrix} I_{00} & I_{01} & \cdots & I_{0m} & I_{0(m+1)} & I_{0(m+2)} \\ I_{10} & I_{11} & \cdots & I_{1m} & I_{1(m+1)} & I_{1(m+2)} \\ \vdots & \vdots & \cdots & \cdot & \cdot & \cdot \\ I_{(m+2)0} & I_{(m+2)1} & \cdots & I_{(m+2)m} & I_{(m+2)(m+1)} & I_{(m+2)(m+2)} \end{pmatrix}.$$

La matrice $I(\beta, \mu, \lambda)$ peut être remplacée par :

$$\hat{I}(\hat{\beta}, \hat{\mu}, \hat{\lambda}) = -\frac{\partial^2 \ell(\hat{\beta}, \hat{\mu}, \hat{\lambda})}{\partial \theta_i \theta_j}, \quad \theta = (\beta_0, \dots, \beta_m, \mu, \lambda)^T,$$

les éléments de la matrice d'information de Fisher $\hat{I}(\beta, \mu, \lambda) = (I_{ls}(\beta, \mu, \lambda))_{(m+3) \times (m+3)}$ sont [Saadia et al. (2010)] :

$$I_{ls} = -\frac{\partial^2 \ell(\beta, \mu, \lambda)}{\partial \beta_l \beta_s} = \frac{\lambda}{2\mu^2} \sum_{i=1}^k \sum_{j=1}^{n_i} \delta_{ij} x_{il} x_{is} \frac{K_{ij}^2 + \mu^2}{K_{ij}} + \frac{1}{2} \sum_{i=1}^k \sum_{j=1}^{n_i} (1 - \delta_{ij}) x_{is} \frac{(M_{ij})'_{\beta_l} S_{ij} - M_{ij} (S_{ij})'_{\beta_l}}{S_{ij}^2},$$

où

$$S_{ij} = \Phi(A_{ij}) - \exp\left(\frac{2\lambda}{\mu}\right) \Phi(B_{ij}),$$

$$M_{ij} = \left(B_{ij} \varphi(A_{ij}) - e^{\frac{2\lambda}{\mu}} A_{ij} \varphi(B_{ij}) \right),$$

$$(M_{ij})'_{\beta_l} = \frac{\partial M_{ij}}{\partial \beta_l} = \frac{1}{2} x_{il} \left[A_{ij} (B_{ij}^2 - 1) \varphi(A_{ij}) - e^{\frac{2\lambda}{\mu}} B_{ij} (A_{ij}^2 - 1) \varphi(B_{ij}) \right],$$

$$(S_{ij})'_{\beta_l} = \frac{\partial S_{ij}}{\partial \beta_l} = -\frac{1}{2} x_{il} \left(B_{ij} \varphi(A_{ij}) - e^{\frac{2\lambda}{\mu}} A_{ij} \varphi(B_{ij}) \right).$$

$$I_{l,m+1} = -\frac{\partial^2 \ell(\beta, \mu, \lambda)}{\partial \beta_l \partial \mu} = \frac{\lambda}{\mu^3} \sum_{i=1}^k \sum_{j=1}^{n_i} \delta_{ij} x_{il} K_{ij} - \frac{1}{\mu^2} \sum_{i=1}^k \sum_{j=1}^{n_i} (1 - \delta_{ij}) \frac{(N_{ij})'_{\beta_l} S_{ij} - N_{ij} (S_{ij})'_{\beta_l}}{S_{ij}^2},$$

où

$$N_{ij} = \sqrt{\lambda K_{ij}} \varphi(A_{ij}) + e^{\frac{2\lambda}{\mu}} \left(2\lambda \Phi(B_{ij}) - \sqrt{\lambda K_{ij}} \varphi(B_{ij}) \right),$$

$$(N_{ij})'_{\beta_l} = \frac{\partial N_{ij}}{\partial \beta_l} = \frac{1}{2} x_{il} \sqrt{\lambda K_{ij}} (-1 + A_{ij} B_{ij}) \left(\varphi(A_{ij}) - e^{\frac{2\lambda}{\mu}} \varphi(B_{ij}) \right) - x_{il} \lambda A_{ij} e^{\frac{2\lambda}{\mu}} \varphi(B_{ij}),$$

$$(S_{ij})'_{\beta_l} = \frac{\partial S_{ij}}{\partial \beta_l} = -\frac{1}{2} x_{il} \left(B_{ij} \varphi(A_{ij}) - e^{\frac{2\lambda}{\mu}} A_{ij} \varphi(B_{ij}) \right).$$

$$I_{l,m+2} = -\frac{\partial^2 \ell(\beta, \mu, \lambda)}{\partial \beta_l \partial \lambda} = -\frac{1}{2\mu^2} \sum_{i=1}^k \sum_{j=1}^{n_i} \delta_{ij} x_{il} \frac{K_{ij}^2 - \mu^2}{K_{ij}} - \sum_{i=1}^k \sum_{j=1}^{n_i} (1 - \delta_{ij}) \frac{(D_{ij})'_{\beta_l} S_{ij} - D_{ij} (S_{ij})'_{\beta_l}}{S_{ij}^2},$$

où

$$D_{ij} = \frac{1}{2\lambda} A_{ij} \varphi(A_{ij}) - e^{\frac{2\lambda}{\mu}} \left(\frac{2}{\mu} \Phi(B_{ij}) + \frac{1}{2\lambda} B_{ij} \varphi(B_{ij}) \right),$$

$$(D_{ij})'_{\beta_l} = \frac{\partial D_{ij}}{\partial \beta_l} = \frac{1}{4\lambda} x_{il} B_{ij} \varphi(A_{ij}) (-1 + A_{ij}^2) - x_{il} e^{\frac{2\lambda}{\mu}} A_{ij} \varphi(B_{ij}) \left(-\frac{1}{\mu} + \frac{1}{4\lambda} (-1 + B_{ij}^2) \right),$$

$$(S_{ij})'_{\beta_l} = \frac{\partial S_{ij}}{\partial \beta_l} = -\frac{1}{2} x_{il} \left(B_{ij} \varphi(A_{ij}) - e^{\frac{2\lambda}{\mu}} A_{ij} \varphi(B_{ij}) \right).$$

$$I_{m+1,m+1} = -\frac{\partial^2 \ell(\beta, \mu, \lambda)}{\partial \mu^2} = \lambda \sum_{i=1}^k \sum_{j=1}^{n_i} \delta_{ij} \frac{3K_{ij} - 2\mu}{\mu^4} - \sum_{i=1}^k \sum_{j=1}^{n_i} (1 - \delta_{ij}) \frac{\mu((F_{ij})'_{\mu} S_{ij} - F_{ij} (S_{ij})'_{\mu}) - 2F_{ij} S_{ij}}{\mu^3 S_{ij}^2},$$

où

$$F_{ij} = \sqrt{\lambda K_{ij}} \varphi(A_{ij}) + e^{\frac{2\lambda}{\mu}} \left(2\lambda \Phi(B_{ij}) - \sqrt{\lambda K_{ij}} \varphi(B_{ij}) \right),$$

$$(F_{ij})'_\mu = \frac{\partial F_{ij}}{\partial \mu} = -\frac{\lambda K_{ij}}{\mu^2} A_{ij} \varphi(A_{ij}) + e^{\frac{2\lambda}{\mu}} \left(-\frac{4\lambda^2}{\mu^2} \Phi(B_{ij}) + \frac{4\lambda \sqrt{\lambda K_{ij}}}{\mu^2} \varphi(B_{ij}) + \frac{\lambda K_{ij}}{\mu^2} B_{ij} \varphi(B_{ij}) \right),$$

$$(S_{ij})'_\mu = \frac{\partial(S_{ij})}{\partial \mu} = \frac{1}{\mu^2} \left\{ \sqrt{\lambda K_{ij}} \left(\varphi(A_{ij}) - e^{\frac{2\lambda}{\mu}} \varphi(B_{ij}) \right) + 2\lambda e^{\frac{2\lambda}{\mu}} \Phi(B_{ij}) \right\}.$$

$$I_{m+1,m+2} = -\frac{\partial^2 \ell(\beta, \mu, \lambda)}{\partial \mu \partial \lambda} = -\frac{1}{\mu^3} \sum_{i=1}^k \sum_{j=1}^{n_i} \delta_{ij} (K_{ij} - \mu) - \sum_{i=1}^k \sum_{j=1}^{n_i} (1 - \delta_{ij}) \frac{(R_{ij})'_\mu S_{ij} - R_{ij} (S_{ij})'_\mu}{S_{ij}^2},$$

où

$$R_{ij} = \frac{1}{2\lambda} A_{ij} \varphi(A_{ij}) - e^{\frac{2\lambda}{\mu}} \left(\frac{2}{\mu} \Phi(B_{ij}) + \frac{1}{2\lambda} \varphi(B_{ij}) \varphi(B_{ij}) \right),$$

$$(R_{ij})'_\mu = \frac{\partial R_{ij}}{\partial \mu} = \frac{\sqrt{\lambda K_{ij}}}{2\lambda \mu^2} (1 - A_{ij}) \varphi(A_{ij}) +$$

$$\frac{e^{\frac{2\lambda}{\mu}}}{\mu^2} \left\{ 2 \left(1 + \frac{2\lambda}{\mu} \right) \Phi(B_{ij}) + (B_{ij} - \frac{2\sqrt{\lambda K_{ij}}}{\mu}) \varphi(B_{ij}) - \frac{\sqrt{\lambda K_{ij}}}{2\lambda} (1 - B_{ij}^2) \varphi(B_{ij}) \right\},$$

$$(S_{ij})'_\mu = \frac{\partial(S_{ij})}{\partial \mu} = \frac{1}{\mu^2} \left(\sqrt{\lambda K_{ij}} \left(\varphi(A_{ij}) - e^{\frac{2\lambda}{\mu}} \varphi(B_{ij}) \right) + 2\lambda \Phi(B_{ij}) \right).$$

$$I_{m+2,m+2} = -\frac{\partial^2 \ell(\beta, \mu, \lambda)}{\partial \lambda^2} = \frac{1}{2\lambda} \sum_{i=1}^k \sum_{j=1}^{n_i} \delta_{ij} - \sum_{i=1}^k \sum_{j=1}^{n_i} (1 - \delta_{ij}) \frac{(G_{ij})'_\lambda S_{ij} - G_{ij} (S_{ij})'_\lambda}{S_{ij}^2},$$

où

$$G_{ij} = \frac{1}{2\lambda} A_{ij} \varphi(A_{ij}) - e^{\frac{2\lambda}{\mu}} \left(\frac{2}{\mu} \Phi(B_{ij}) + \frac{1}{2\lambda} \varphi(B_{ij}) \varphi(B_{ij}) \right),$$

$$(G_{ij})'_\lambda = \frac{\partial G_{ij}}{\partial \lambda} = -\frac{1}{4\lambda^2} A_{ij} (1 + A_{ij}^2) \varphi(A_{ij}) -$$

$$e^{\frac{2\lambda}{\mu}} \left\{ \frac{4}{\mu^2} \Phi(B_{ij}) + \frac{2}{\mu\lambda} B_{ij} \varphi(B_{ij}) - \frac{1}{4\lambda^2} B_{ij} (1 + B_{ij}^2) \varphi(B_{ij}) \right\},$$

$$(S_{ij})'_\lambda = \frac{\partial(S_{ij})}{\partial\lambda} = \frac{1}{2\lambda} A_{ij} \varphi(A_{ij}) - e^{\frac{2\lambda}{\mu}} \left(\frac{2}{\mu} \Phi(B_{ij}) + \frac{1}{2\lambda} B_{ij} \varphi(B_{ij}) \right).$$

Ainsi la matrice d'information de Fisher est complètement calculée. L'étude par simulation des modèles *AFT-IG*, *AFT-BS*, *AFT-LN*, *AFT-LL*, *AFT-EW* et *AFT-GW* fait l'objet de futur travaux.

2.7 Test d'ajustement pour le modèle AFT paramétrique

Dans cette section nous donnons le test d'ajustement basé sur la statistique NRR pour le modèle AFT paramétrique proposée par Bagdonavicius et Nikulin (2012). Ils ont proposé ce test d'ajustement pour différents modèles (AFT) paramétriques.

Supposons la donnée de n variables de pannes indépendantes. On considère l'hypothèse $H_0 z(t) = (z_0(t), z_1(t), \dots, z_m(t))^T$, $z_0 = 1$

$$S(t|z) = S_0 \left(\int_0^t e^{-\beta^T z(u)} du; \gamma \right),$$

$$\beta = (\beta_0, \dots, \beta_m)^T,$$

$$\theta = (\beta^T, \gamma^T)^T.$$

2.7.1 Construction du test

Nous construisons maintenant le test de Chi-deux pour l'hypothèse H_0 à partir des temps de pannes (failure time) censurées à droite :

$$(X_1, \delta_1, z_1), (X_2, \delta_2, z_2), \dots, (X_n, \delta_n, z_n),$$

où

$$X_i = T_i \wedge C_i, \quad \delta_i = \mathbf{1}_{\{T_i \leq C_i\}}, \quad z_i(t) = (1, z_{i1}(t), \dots, z_{im}(t))^T,$$

T_i sont les temps de pannes, C_i les temps de censure et z_i le vecteur des covariables.

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

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

où

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

Posons :

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

Supposons que les processus N_i, Y_i, z_i sont observés dans un temps fini τ et que la censure est non-informative.

On divise l'intervalle $[0, \tau]$ au k sous-intervalles $I_j =]a_{j-1}, a_j]$, $a_0 = 0, a_k = \tau$, et on note par :

$$U_j = N(a_j) - N(a_{j-1})$$

le nombre de pannes observé dans le sous-intervalle $I_j, j = 1, 2, \dots, k$. Pour estimer le nombre prévu de pannes dans l'intervalle I_j sous l'hypothèse H_0 , nous considérons l'égalité :

$$EN_i(t) = E \int_0^t h_i(u, \theta) Y_i(u) du,$$

et nous pouvons compter

$$e_j = \sum_{i=1}^n \int_{I_j} h_i(u, \hat{\theta}) Y_i(u) du,$$

pannes, et $h_i(u, \theta)$ est la fonction de hasard de T_i sous z_i , et $\hat{\theta}$ est l'estimateur de MV de θ sous H_0 .

Alors le test statistique est basé sur le vecteur

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

2.7.2 La distribution asymptotique de Z

Les propriétés de la statistique Z peuvent être étudiées à travers le processus stochastique

$$H_n(t) = \frac{1}{\sqrt{n}} \left(N(t) - \sum_{i=1}^n \int_0^t h_i(u, \hat{\theta}) Y_i(u) du \right).$$

Chapitre 3. Le modèle de Régression AFT.

Les propriétés des estimateurs MV se résument dans les conditions A suivantes :

Conditions A :

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

et

$$\sqrt{n}(\hat{\theta} - \theta_0) = i^{-1}(\theta_0) \frac{1}{\sqrt{n}} \dot{\ell}(\theta_0) + o_p(1);$$

où

$$\dot{\ell}(\theta) = \sum_{i=1}^n \int_0^\tau \frac{\partial}{\partial \theta} \ln h_i(u, \theta) \{dN_i(u) - Y_i(u)h_i(u, \theta)du\},$$

avec

$$\dot{\ell}(\theta) = \frac{\partial}{\partial \theta} \ell(\theta) = \left(\frac{\partial}{\partial \beta_0} \ell(\theta), \dots, \frac{\partial}{\partial \gamma_q} \ell(\theta) \right)^T, \quad \theta = (\beta^T, \gamma^T)^T,$$

est le vecteur des fonctions de scores.

Soient

$$S^{(0)}(t, \theta) = \sum_{i=1}^n Y_i(t)h_i(t, \theta), \quad S^{(1)}(t, \theta) = \sum_{i=1}^n Y_i(t) \frac{\partial \ln h_i(t, \theta)}{\partial \theta} h_i(t, \theta),$$

$$S^{(2)}(t, \theta) = \sum_{i=1}^n Y_i(t) \frac{\partial^2 \ln h_i(t, \theta)}{\partial \theta^2} h_i(t, \theta).$$

Conditions B : Ils existent un voisinage Θ_0 de θ_0 et des fonctions continues et bornées sur $[0, \tau] \times \Theta_0$:

$$s^{(0)}(t, \theta), \quad s^{(1)}(t, \theta) = \frac{\partial s^{(0)}(t, \theta)}{\partial \theta}, \quad s^{(2)}(t, \theta) = \frac{\partial^2 s^{(0)}(t, \theta)}{\partial \theta^2},$$

telles que pour $j = 0, 1, 2$:

$$\sup_{t \in [0, \tau], \theta \in \Theta} \left\| \frac{1}{n} S^j(t, \theta) - s^j(t, \theta) \right\| \xrightarrow{P} 0 \quad \text{as } n \rightarrow \infty.$$

Les conditions B impliquent que uniformément sur $[0, \tau]$:

$$\frac{1}{n} \sum_{i=1}^n \int_0^t h_i(u, \theta_0) Y_i(u) du \xrightarrow{P} A(t), \quad \frac{1}{n} \sum_{i=1}^n \int_0^t \frac{\partial h_i(t, \theta_0)}{\partial \theta} Y_i(u) du \xrightarrow{P} C(t),$$

où A et C sont des fonctions finies.

Lemme 1. [Bagdonavicius et al. (2012)] Sous les conditions A et B nous aurons :

$$H_n \xrightarrow{P} V \quad \text{sur} \quad D[0, \tau],$$

où

$D[0, \tau]$ est l'espace de Skorokhod des fonctions càdlàg, V est une martingale Gaussienne de moyenne 0, telle que pour tous $0 \leq u \leq v \leq \tau$:

$$\text{cov}(V(u), V(v)) = A(u) - C^T(u)i^{-1}(\theta_0)C(v).$$

Pour $i = 0, \dots, m + q = s$; $j, j' = 1, \dots, k$ on pose :

$$V_j = V(a_j) - V(a_{j-1}); \quad \sigma_{jj'} = \text{cov}(V_j, V_{j'}),$$

$$A_j = A(a_j) - A(a_{j-1}), C_{ij} = C_i(a_j) - C_i(a_{j-1}), \quad \mathcal{C}_j = (C_{0j}, \dots, C_{sj})^T,$$

$$\Sigma = [\sigma_{jj'}]_{k \times k}; \quad \mathcal{C} = [C_{ij}]_{(s+1) \times k},$$

notons par \mathcal{A} la matrice diagonale de dimension $k \times k$ avec ces éléments diagonaux A_1, \dots, A_k .

Théorème 3.1. [Bagdonavicius et al. (2012)] Sous les conditions A et B

$$Z \xrightarrow{P} Y \sim N_k(0, \Sigma) \quad \text{as} \quad n \rightarrow \infty,$$

où

$$\Sigma = \mathcal{A} - C^T i^{-1}(\theta_0)C,$$

Remarque 3.1. Soit la matrice G définie par $G = i - \mathcal{C}\mathcal{A}^{-1}\mathcal{C}^T$, alors la formule :

$$\Sigma^{-1} = \mathcal{A}^{-1} + \mathcal{A}^{-1}\mathcal{C}^T G^{-1}\mathcal{C}\mathcal{A}^{-1},$$

implique qu'on a besoin d'inverser la matrice \mathcal{A} et de trouver l'inverse généralisé de la matrice G de dimensions $s \times s$.

Théorème 3.2. [Bagdonavicius et al. (2012)] Sous les conditions A et B, les estimateurs de A_j , C_j , $i(\theta_0)$ et Σ suivant sont consistants :

$$\hat{A}_j = U_j/n, \quad \hat{C}_j = \frac{1}{n} \sum_{i=1}^n \int_{I_j} \frac{\partial \ln h_i(u, \hat{\theta})}{\partial \theta} dN_i(u),$$

et

$$\hat{i} = \frac{1}{n} \sum_{i=1}^n \int_{I_j} \frac{\partial \ln h_i(u, \hat{\theta})}{\partial \theta} \left(\frac{\partial \ln h_i(u, \hat{\theta})}{\partial \theta} \right)^T dN_i(u).$$

2.7.3 Le test Statistique

A la base des théorèmes 1 et 2, pour tester l'hypothèse H_0 nous pouvons utiliser le test de type Chi-deux Y_n^2 tel que :

$$Y_n^2 = Z^T \hat{\Sigma} Z,$$

avec

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

Ce test peut s'écrire sous la forme :

$$Y_n^2 = \sum_{i=1}^n \frac{(U_j - e_j)^2}{U_j} + Q,$$

où

$$U_j = \sum_{i: X_i \in I_j} \delta_i, \quad e_j = \sum_{i=1}^n \int_{I_j} h_i(u, \hat{\theta}) Y_i(u) du, \quad Q = W^T \hat{G}^- W,$$

$$W = (W_0, \dots, W_s)^T, \quad \hat{G} = [\hat{g}_{ll'}]_{s \times s}, \quad \hat{g}_{ll'} = \hat{i}_{ll'} - \sum_{i=j}^k \hat{C}_{lj} \hat{C}_{l'j} \hat{A}_j^{-1},$$

$$\hat{C}_{lj} = \frac{1}{n} \sum_{i: X_i \in I_j} \delta_i \frac{\partial}{\partial \theta_l} \ln h_i(X_i, \hat{\theta}), \quad \hat{i}_{ll'} = \frac{1}{n} \sum_{i=1}^n \delta_i \frac{\partial \ln h_i(X_i, \hat{\theta})}{\partial \theta_l} \frac{\partial \ln h_i(X_i, \hat{\theta})}{\partial \theta_{l'}},$$

$$W_l = \sum_{i=j}^k \hat{C}_{lj} \hat{A}_j^{-1} Z_j, \quad l, l' = 1, 2, \dots, s.$$

Sous l'hypothèse H_0 et pour n suffisamment grand, la statistique Y_n^2 suit asymptotiquement une loi de χ_r^2 à $r = \text{rang}(\Sigma^-) = \text{tr}(\Sigma^- \Sigma)$ degrés de liberté. Si la matrice G est non-dégénérée alors $r = k$. Pour un seuil α , l'hypothèse H_0 est rejetée si $Y_n^2 > C_\alpha = \chi_{r, 1-\alpha}^2$.

2.7.4 Choix de \hat{a}_j :

Soit

$$E_k = \sum_{i=1}^n \int_0^\tau h_i(u, \hat{\theta}) Y_i(u) du = \sum_{i=1}^n h_i(X_i, \hat{\theta}), \quad E_j = \frac{j}{k} E_k, \quad j = 1, \dots, k.$$

Nous choisissons alors \hat{a}_j d'une manière à avoir des nombres égaux de pannes prévues dans tous les sous-intervalles. Anisi les \hat{a}_j vérifient les égalités :

$$g(\hat{a}_j) = E_j, \quad g(a) = \sum_{i=1}^n \int_0^a h_i(u, \hat{\theta}) Y_i(u) du.$$

Si on considère la statistique d'ordre $(X_{(1)}, \dots, X_{(n)})$, il vient que :

$$g(a) = \sum_{l=1}^n h_l(X_l \wedge a, \hat{\theta}) = \sum_{i=1}^n \left[\sum_{l=i}^n h_{(l)}(a, \hat{\theta}) + \sum_{l=1}^{i-1} h_{(l)}(X_{(l)}, \hat{\theta}) \right] \mathbf{1}_{[X_{(i-1)}, X_{(i)}]}(a)$$

est une fonction continue et croissante sur $[0, \tau]$.

Soit

$$b_i = \sum_{l=i+1}^n h_{(l)}(X_{(i)}, \hat{\theta}) + \sum_{l=1}^i h_{(l)}(X_{(l)}, \hat{\theta}).$$

Si $E_j \in [b_{i-1}, b_i]$ alors \hat{a}_j est l'unique solution de l'équation

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

et on a : $0 < \hat{a}_1 < \hat{a}_2 < \dots < \hat{a}_k = \tau$, et dans ce cas : $e_j = E_k/k$ pour tout j .

2.7.5 Application à la loi Gaussienne inverse

Considérons le modèle AFT-IG :

$$H_0 : S_i(t) = S_0 \left(\int_0^t e^{-\beta^T z_i(u)} du \right), \quad i = 1, \dots, n$$

où

$$S_0(t) = \Phi \left(-\sqrt{\frac{\lambda}{t}} \left(\frac{t}{\mu} - 1 \right) \right) - \exp\left(\frac{2\lambda}{\mu}\right) \Phi \left(-\sqrt{\frac{\lambda}{t}} \left(\frac{t}{\mu} + 1 \right) \right), \quad \mu, \lambda > 0$$

avec $\beta = (\beta_0, \dots, \beta_m)^T$, $z(u) = (z_0(u), \dots, z_m(u))^T$, $z_0(u) \equiv 0$. Si $z_i(u)$ sont constants, alors :

$$H_0 : S_i(t) = S_0 \left(e^{-\beta^T z_i} t \right), \quad i = 1, \dots, n.$$

La fonction de vraisemblance $L(\beta, \mu, \lambda)$ est :

$$L = \prod_{i=1}^n \left\{ \frac{e^{-\beta^T z_i} \left(\frac{\lambda}{2\pi X_i^3 e^{-3\beta^T z_i}} \right)^{\frac{1}{2}} \exp\left\{-\frac{\lambda(X_i e^{-\beta^T z_i} - \mu)^2}{2\mu^2 X_i e^{-\beta^T z_i}}\right\}}{\Phi\left(-\sqrt{\frac{\lambda}{X_i e^{-\beta^T z_i}}}\left(\frac{1}{\mu} X_i e^{-\beta^T z_i} - 1\right)\right) - e^{\frac{2\lambda}{\mu}} \Phi\left(-\sqrt{\frac{\lambda}{X_i e^{-\beta^T z_i}}}\left(\frac{1}{\mu} X_i e^{-\beta^T z_i} + 1\right)\right)} \right\}^{\delta_i} \times \left\{ \Phi\left(-\sqrt{\frac{\lambda}{X_i e^{-\beta^T z_i}}}\left(\frac{1}{\mu} X_i e^{-\beta^T z_i} - 1\right)\right) - e^{\frac{2\lambda}{\mu}} \Phi\left(-\sqrt{\frac{\lambda}{X_i e^{-\beta^T z_i}}}\left(\frac{1}{\mu} X_i e^{-\beta^T z_i} + 1\right)\right) \right\},$$

Avec $K_i = X_i e^{-\beta^T z_i}$, $A_i = -\sqrt{\frac{\lambda}{K_i}} \left(\frac{K_i}{\mu} - 1 \right)$ et $B_i = -\sqrt{\frac{\lambda}{K_i}} \left(\frac{K_i}{\mu} + 1 \right)$,

Le log-vraisemblance $\ell(\beta, \mu, \lambda)$ est :

$$\ell(\beta, \mu, \lambda) = \sum_{i=1}^n \delta_i \left\{ -\beta^T z_i + \frac{1}{2} \ln \lambda - \frac{3}{2} \ln K_i - \lambda \frac{(K_i - \mu)^2}{2\mu^2 K_i} \right\} + \sum_{i=1}^n (1 - \delta_i) \ln \left\{ \Phi(A_i) - e^{\frac{2\lambda}{\mu}} \Phi(B_i) \right\}.$$

La fonction de hasard pour le modèle AFT-IG est :

$$h(X_i, \theta) = \frac{e^{-\beta^T z_i} \left(\frac{\lambda}{2\pi X_i^3 e^{-3\beta^T z_i}} \right)^{\frac{1}{2}} \exp\left\{-\frac{\lambda(X_i e^{-\beta^T z_i} - \mu)^2}{2\mu^2 X_i e^{-\beta^T z_i}}\right\}}{\Phi\left(-\sqrt{\frac{\lambda}{X_i e^{-\beta^T z_i}}}\left(\frac{1}{\mu} X_i e^{-\beta^T z_i} - 1\right)\right) - e^{\frac{2\lambda}{\mu}} \Phi\left(-\sqrt{\frac{\lambda}{X_i e^{-\beta^T z_i}}}\left(\frac{1}{\mu} X_i e^{-\beta^T z_i} + 1\right)\right)},$$

d'où

$$\ln h(X_i, \theta) = \left\{ -\beta^T z_i + \frac{1}{2} \ln \lambda - \frac{3}{2} \ln K_i - \lambda \frac{(K_i - \mu)^2}{2\mu^2 K_i} \right\} - \ln \left\{ \Phi(A_i) - e^{\frac{2\lambda}{\mu}} \Phi(B_i) \right\}.$$

et les dérivées partielles de cette fonctions sont :

$$\frac{\partial \ln h(X_i, \theta)}{\partial \theta_l} = \frac{x_{il}}{2} \left\{ 1 + \frac{\lambda(K_i^2 - \mu^2)}{\mu^2 K_i} + \frac{B_i \varphi(A_i) - e^{\frac{2\lambda}{\mu}} A_i \varphi(B_i)}{S_i} \right\}, \quad l = 1, \dots, m;$$

$$\frac{\partial \ln h(X_i, \theta)}{\partial \mu} = \left\{ \frac{\lambda}{\mu^3} (K_i - \mu) - \frac{1}{\mu^2} \frac{\sqrt{\lambda K_i} \varphi(A_i) + e^{\frac{2\lambda}{\mu}} (2\lambda \Phi(B_i) - \sqrt{\lambda K_i} \varphi(B_i))}{S_i} \right\};$$

$$\frac{\partial \ln h(X_i, \theta)}{\partial \lambda} = \left\{ \left(\frac{1}{2\lambda} - \frac{(K_i - \mu)^2}{2\mu^2 K_i} \right) - \frac{\frac{1}{2\lambda} A_i \varphi(A_i) - e^{\frac{2\lambda}{\mu}} \left(\frac{2}{\mu} \Phi(B_i) + \frac{1}{2\lambda} B_i \varphi(B_i) \right)}{S_i} \right\}.$$

Les éléments de la matrice \mathcal{C} sont :

$$\hat{C}_{lj} = \frac{1}{n} \sum_{i: X_i \in I_j} \frac{x_{il}}{2} \delta_i \left\{ 1 + \frac{\lambda(K_i^2 - \mu^2)}{\mu^2 K_i} + \frac{B_i \varphi(A_i) - e^{\frac{2\lambda}{\mu}} A_i \varphi(B_i)}{S_i} \right\},$$

$$\hat{C}_{m+1,j} = \frac{1}{n} \sum_{i: X_i \in I_j} \delta_i \left\{ \frac{\lambda}{\mu^3} (K_i - \mu) - \frac{1}{\mu^2} \frac{\sqrt{\lambda K_i} \varphi(A_i) + e^{\frac{2\lambda}{\mu}} (2\lambda \Phi(B_i) - \sqrt{\lambda K_i} \varphi(B_i))}{S_i} \right\},$$

$$\hat{C}_{m+2,j} = \frac{1}{n} \sum_{i: X_i \in I_j} \delta_i \left\{ \left(\frac{1}{2\lambda} - \frac{(K_i - \mu)^2}{2\mu^2 K_i} \right) - \frac{\frac{1}{2\lambda} A_i \varphi(A_i) - e^{\frac{2\lambda}{\mu}} \left(\frac{2}{\mu} \Phi(B_i) + \frac{1}{2\lambda} B_i \varphi(B_i) \right)}{S_i} \right\},$$

où

$$l, j = 1, \dots, k, \quad S_i = \Phi(A_i) - e^{\frac{2\lambda}{\mu}} \Phi(B_i).$$

Les éléments de la matrice de Fisher sont :

$$\hat{i}_{ll'} = \frac{1}{4n} \sum_{i: X_i \in I_j} \delta_i x_{il} x_{il'} \left\{ 1 + \frac{\lambda(K_i^2 - \mu^2)}{\mu^2 K_i} + \frac{B_i \varphi(A_i) - e^{\frac{2\lambda}{\mu}} A_i \varphi(B_i)}{S_i} \right\}^2, \quad l, l' = 1, \dots, m,$$

$$\begin{aligned} \hat{i}_{l,m+1} &= \frac{1}{n} \sum_{i: X_i \in I_j} \delta_i \frac{x_{il}}{2} \left\{ 1 + \frac{\lambda(K_i^2 - \mu^2)}{\mu^2 K_i} + \frac{B_i \varphi(A_i) - e^{\frac{2\lambda}{\mu}} A_i \varphi(B_i)}{S_i} \right\} \times \\ &\quad \left\{ \frac{\lambda}{\mu^3} (K_i - \mu) - \frac{1}{\mu^2} \frac{\sqrt{\lambda K_i} \varphi(A_i) + e^{\frac{2\lambda}{\mu}} (2\lambda \Phi(B_i) - \sqrt{\lambda K_i} \varphi(B_i))}{S_i} \right\}, \end{aligned}$$

$$\begin{aligned} \hat{i}_{m+1,m+1} &= \frac{1}{n} \sum_{i: X_i \in I_j} \delta_i \left\{ \frac{\lambda}{\mu^3} (K_i - \mu) - \frac{1}{\mu^2} \frac{\sqrt{\lambda K_i} \varphi(A_i) + e^{\frac{2\lambda}{\mu}} (2\lambda \Phi(B_i) - \sqrt{\lambda K_i} \varphi(B_i))}{S_i} \right\}^2, \\ \hat{i}_{l,m+2} &= \frac{1}{n} \sum_{i: X_i \in I_j} \delta_i \frac{x_{il}}{2} \left\{ 1 + \frac{\lambda(K_i^2 - \mu^2)}{\mu^2 K_i} + \frac{B_i \varphi(A_i) - e^{\frac{2\lambda}{\mu}} A_i \varphi(B_i)}{S_i} \right\} \times \\ &\quad \left\{ \left(\frac{1}{2\lambda} - \frac{(K_i - \mu)^2}{2\mu^2 K_i} \right) - \frac{\frac{1}{2\lambda} A_i \varphi(A_i) - e^{\frac{2\lambda}{\mu}} \left(\frac{2}{\mu} \Phi(B_i) + \frac{1}{2\lambda} B_i \varphi(B_i) \right)}{S_i} \right\}, \\ \hat{i}_{m+1,m+2} &= \frac{1}{n} \sum_{i: X_i \in I_j} \delta_i \left\{ \frac{\lambda}{\mu^3} (K_i - \mu) - \frac{1}{\mu^2} \frac{\sqrt{\lambda K_i} \varphi(A_i) + e^{\frac{2\lambda}{\mu}} (2\lambda \Phi(B_i) - \sqrt{\lambda K_i} \varphi(B_i))}{S_i} \right\} \times \\ &\quad \left\{ \left(\frac{1}{2\lambda} - \frac{(K_i - \mu)^2}{2\mu^2 K_i} \right) - \frac{\frac{1}{2\lambda} A_i \varphi(A_i) - e^{\frac{2\lambda}{\mu}} \left(\frac{2}{\mu} \Phi(B_i) + \frac{1}{2\lambda} B_i \varphi(B_i) \right)}{S_i} \right\}, \\ \hat{i}_{m+2,m+2} &= \frac{1}{n} \sum_{i: X_i \in I_j} \delta_i \left\{ \left(\frac{1}{2\lambda} - \frac{(K_i - \mu)^2}{2\mu^2 K_i} \right) - \frac{\frac{1}{2\lambda} A_i \varphi(A_i) - e^{\frac{2\lambda}{\mu}} \left(\frac{2}{\mu} \Phi(B_i) + \frac{1}{2\lambda} B_i \varphi(B_i) \right)}{S_i} \right\}^2. \end{aligned}$$

A cause de la difficulté de maximiser la fonction de log-vraisemblance, l'application de ce test à des exemples en fiabilité et analyse de survie est soumise à la recherche des algorithmes d'optimisation plus performants.

Chapitre 4

Analyse des systèmes redondants

1 Introduction

La redondance est une approche visant à améliorer la fiabilité des systèmes. La conception de redondance augmente le coût et la complexité du système avec une haute fiabilité des composants électriques et mécaniques modernes. De nombreuses applications n'ont pas besoin de redondance afin de réussir, mais si le coût de la panne est assez élevé, la redondance peut être une option intéressante.

Considérons un système de redondance avec une unité principale qui fonctionne aux conditions "hot" et $(m - 1)$ unités en veille (stand-by units) fonctionnant aux conditions "warm". Si l'unité principale tombe en panne, une unité en veille va la remplacer automatiquement et ainsi de suite. Notons $S(1, m - 1)$ un tel système. **Le problème est d'obtenir les intervalles de confiance pour la fonction de répartition** du système de redondance en utilisant des données de panne de deux groupes d'unités. On suppose que la commutation du mode "warm" en mode "hot" ne provoque aucun choc et aucun dommage pour les unités du système.

Dans ce chapitre, nous commençons par donner la formulation mathématique du système $S(1, 1)$ avec une unité principale et une unité en veille, et le système $S(1, m - 1)$, puis nous construisons les systèmes redondants basés sur les familles dont la fonction de hasard est unimodale.

2 Système redondant $S(1, 1)$

En utilisant le principe du modèle de Sedyakin, Bagdonavičius et Masiulaityle et Nikulin (2008a,2008b) ont donné la formulation mathématique basée sur le modèle AFT, et ils ont proposé des tests pour valider ces modèles.

Supposons que le temps de panne T_1 de l'unité principale a la fonction de répartition F_1 et la densité de probabilité f_1 , le temps de panne T_2 de l'unité en veille a la fonction de répartition F_2 et

la densité de probabilité f_2 .

Le temps de panne du système est $T = \max(T_1, T_2)$.

Notons par $f_2^{(y)}(x)$ la densité conditionnelle de l'unité en veille sachant que l'unité principale tombe en panne à l'instant y . Il est clair que $f_2^{(y)}(x) = f_2(x)$ si $0 \leq x \leq y$.

La fonction de répartition de T s'écrit

$$F(t) = P(T_1 \leq t, T_2 \leq t) = \int_0^t \left\{ \int_0^y f_2(x) dx + \int_y^t f_2^{(y)}(x) dx \right\} f_1(y) dy. \quad (4.1)$$

Si l'unité en veille est en mode "warm" alors $f_2(x) = 0$, $f_2^{(y)}(x) = f_1(x - y)$, $x > y$, et on a :

$$F(t) = \int_0^t \left\{ \int_y^t f_1(x - y) dx \right\} f_1(y) dy = \int_0^t F_1(t - y) dF_1(y).$$

Dans le mode "hot", $f_2^{(y)}(x) = f_2(x) = f_1(x)$, et alors $F(t) = [F_1(t)]^2$.

Considérons l'hypothèse H_0 suivante :

$$H_0 : f_2^{(y)}(x) = f_1(x + g(y) - y), \quad \text{pour tout } x \geq y \geq 0,$$

où $g(y)$ est l'instant pour lequel en mode "hot" correspond à l'instant y en mode "warm" avec :

$$F_1(g(y)) = P(T_1 \leq g(y)) = P(T_2 \leq y) = F_2(y),$$

alors

$$g(y) = F_1^{-1}(F_2(y)).$$

Conditionnellement (sachant $T_1 = y$), l'hypothèse correspond au modèle de Sedyakin. Dans ce cas la formule (4.1) implique sous H_0

$$F(t) = \int_0^t F_1(t + g(y) - y) dF_1(y). \quad (4.2)$$

En particulier, si l'on suppose que la distribution des unités fonctionnant en "warm" et "hot" ne diffère que par l'échelle, i.e.

$$F_2(t) = F_1(rt), \quad (4.3)$$

pour $r > 0$, alors $g(y) = ry$. Dans cette situation l'hypothèse suivante est à vérifier

$$H_0^* : f_2^{(y)}(x) = f_1(x + ry - y), \quad \forall \quad x \geq y \geq 0, \quad (4.4)$$

Conditionnellement (sachant $T_1 = y$), l'hypothèse (4.4) correspond au modèle AFT (Bagdonavičius et Masiulaityle et Nikulin (2008a, 2008b)).

La fonction de répartition du système redondant $S(1, 1)$ s'écrit :

$$K_2(t) = \int_0^t F_1(t + ry - y) dF_1(y).$$

3 Système redondant $S(1, m - 1)$

Selon les résultats de Bagdonavičius et Masiulaityle et Nikulin (2008a, 2008b, 2009, 2010), on suppose qu'en mode "hot", le temps de panne, la fonction de répartition, la fonction de survie et la fonction de densité de l'unité principale sont respectivement T_1, F_1, S_1 et f_1 , et en mode 'warm' le temps de panne, la fonction de répartition, la fonction de survie et la fonction de densité des unités en veille sont respectivement T_i, F_2, S_2 et $f_2, i = 2, \dots, m$.

Si une unité en veille est passée du mode "warm" au mode "hot", sa fonction de répartition est différente de F_1 et F_2 . Le temps de panne du système $S(1, m - 1)$ est :

$$T^{(m)} = T_1 \vee T_2 \vee \dots \vee T_{m-1} \vee T_m = (T_1 \vee T_2 \vee \dots \vee T_{m-1}) \vee T_m.$$

Notons par K_j la fonction de répartition de $T^{(j)}$, $j = 2, \dots, m$. Nous savons que :

$$K_j(t) = \mathbf{P}\{T^{(j)} \leq t\} = \mathbf{P}\{T^{(j-1)} \leq t, T_j \leq t\} = \int_0^t \mathbf{P}\{T_j \leq t | T^{(j-1)} = y\} dK_{j-1}(y).$$

Selon Bagdonavičius et Masiulaityle et Nikulin (2009), la densité conditionnelle de T_j sachant $T^{(j-1)} = y$ est :

$$f_{T_j | T^{(j-1)} = y}(t) = \begin{cases} f_2(t), & \text{if } t \leq y \\ f_1(t + ry - y), & \text{if } t > y \end{cases},$$

pour $r > 0$. Cela signifie que les distributions des unités fonctionnant aux modes "warm" et "hot" ne diffèrent que par une échelle, i.e.

$$F_2(t) = F_1(rt) \quad , \forall t \geq 0, \quad r > 0.$$

La fonction de répartition du système redondant $S(1, m - 1)$ avec unité principale et $j - 1$ unités

en veille est donnée par

$$K_j(t) = \int_0^t F_1(t + ry - y) dK_{j-1}(y), \quad j = 2, \dots, m, \quad K_1(t) = F_1(t), \quad (4.5)$$

il s'ensuit que la fonction de répartition $K_m(t)$ du système $S(1, m-1)$ est définie de façon récurrente à l'aide de l'équation (4.5). En pratique, souvent les fonctions de répartition des unités appartiennent à une famille paramétrique.

Pour $j = 2$, nous obtenons la formule de $K_2(t)$:

$$K_2(t) = \int_0^t F_1(t + ry - y) f_1(y) dy. \quad (4.6)$$

4 L'estimation de $K_j(t)$

Supposons que ces données sont disponibles :

a) données censurées à droite

$$(X_{11}, \delta_{11})^T, \dots, (X_{1n_1}, \delta_{1n_1})^T$$

de taille n_1 , où $X_{1i} = T_{1i} \wedge C_{1i}$, $\delta_{1i} = 1_{\{T_{1i} \leq C_{1i}\}}$.

T_{1i} est le temps de panne des unités testées en mode "hot", C_{1i} est la censure.

b) données censurées à droite

$$(X_{21}, \delta_{21})^T, \dots, (X_{2n_2}, \delta_{2n_2})^T$$

de taille n_2 , où $X_{2j} = T_{2j} \wedge C_{2j}$, $\delta_{2j} = 1_{\{T_{2j} \leq C_{2j}\}}$.

T_{2i} est le temps de panne des unités testées en mode "warm", C_{2i} est la censure.

Soient $m_1 = \sum_{i=1}^{n_1} \delta_{1i}$, $m_2 = \sum_{j=1}^{n_2} \delta_{2j}$, $n = n_1 + n_2$.

Supposons que, dans des conditions "hot" la fonction de répartition $F_1(t; \theta)$ est absolument continue et dépend d'un paramètre de dimension finie $\theta \in \Theta \subset \mathbf{R}^k$.

Posons $\gamma = (r, \theta)^T$, l'estimateur du maximum de vraisemblance $\hat{\gamma} = (\hat{r}, \hat{\theta}^T)^T$ du paramètre γ maximisant la fonction de log-vraisemblance :

$$\begin{aligned} \ell(\gamma) &= \sum_{i=1}^{n_1} \delta_{1i} \ln f_1(X_{1i}, \theta) + m_2 \ln r + \sum_{j=1}^{n_2} \delta_{2j} \ln f_1(rX_{2j}, \theta) + \\ &\sum_{i=1}^{n_1} (1 - \delta_{1i}) \ln S_1(X_{1i}, \theta) + \sum_{j=1}^{n_2} (1 - \delta_{2j}) \ln S_1(rX_{2j}, \theta). \end{aligned}$$

La fonction de répartition $K_j(t)$ avec unité principale et $j - 1$ unités en veille est estimée par

$$\hat{K}_j(t) = \int_0^t F_1(t + \hat{r}y - y, \hat{\theta}) d\hat{K}_{j-1}(y), \quad j = 2, \dots, m, \quad \hat{K}_1(t) = F_1(t, \hat{\theta}).$$

5 L'intervalle de confiance asymptotique de $K_j(t)$

En utilisant les résultats de Bagdonavičius et Masiulaityle et Nikulin (2008a, 2008b, 2009, 2010), l'intervalle de confiance de $K_j(t)$ au seuil α est

$$\hat{K}_j(t) \pm \hat{\sigma}_{\hat{K}_j(t)}^2 z_{1-\alpha/2},$$

ou encore l'intervalle $(\underline{K}_j(t), \overline{K}_j(t))$ de $K_j(t)$ au seuil α est donné par

$$\underline{K}_j(t) = \left(1 + \frac{1 - \hat{K}_j(t)}{\hat{K}_j(t)} \exp \left\{ \frac{\hat{\sigma}_{\hat{K}_j(t)} z_{1-\alpha/2}}{\sqrt{\hat{K}_j(t)(1 - \hat{K}_j(t))}} \right\} \right)^{-1}, \quad (4.7)$$

$$\overline{K}_j(t) = \left(1 + \frac{1 - \hat{K}_j(t)}{\hat{K}_j(t)} \exp \left\{ -\frac{\hat{\sigma}_{\hat{K}_j(t)} z_{1-\alpha/2}}{\sqrt{\hat{K}_j(t)(1 - \hat{K}_j(t))}} \right\} \right)^{-1}, \quad (4.8)$$

où

$$\hat{\sigma}_{\hat{K}_j(t)}^2 = C_2^T(t, \hat{\gamma}) I^{-1}(\hat{\gamma}) C_2(t, \hat{\gamma}),$$

$$C_2(t, \gamma) = (C_{21}(t, \gamma), C_{22}^T(t, \gamma))^T.$$

Les composantes du vecteur $C_2(t, \gamma)$ sont données par :

$$C_{21}(t, \gamma) = \int_0^t \frac{\partial F_1}{\partial r}(t + ry - y, \theta) dF_1(y, \theta),$$

$$C_{22}(t, \gamma) = \int_0^t \frac{\partial F_1}{\partial \theta}(t + ry - y, \theta) dF_1(y, \theta) + F_1(t + ry - y, \theta) d\left(\frac{\partial F_1}{\partial \theta}(y, \theta)\right),$$

avec $z_{1-\alpha/2}$ est le $(1 - \alpha/2)$ -quantile de $N(0, 1)$ et I est l'information de Fisher.

Rappelons que Bagdonavičius et Masiulaityle et Nikulin (2008a) ont donné l'estimation non paramétrique de la fonction de répartition des systèmes redondants. Bagdonavičius et Masiulaityle et Nikulin (2009) ont traité l'estimation de la fonction de répartition des systèmes redondants basés sur les familles paramétriques : Weibull, exponentielle et log-logistique. Le système redondant basé sur la famille de Birnbaum-Saunders est bien étudié par Tahir (2012).

Construisons maintenant les systèmes redondants basés sur les familles paramétriques de fonction de hasard unimodale vu précédemment.

6 La distribution Gaussienne inverse

Supposons maintenant que les fonctions de répartition des unités appartiennent à la famille des distributions Gaussiennes inverses [Nikulin et Saaidia (2009), Lemeshko et al. (2010b), Nikulin et Saaidia et Tahir (2011a)], i.e.

$$S_1(t, \mu, \lambda) = 1 - F_1(t, \mu, \lambda) = \Phi\left(-\sqrt{\frac{\lambda}{t}}\left(\frac{t}{\mu} - 1\right)\right) - \exp\left(\frac{2\lambda}{\mu}\right)\Phi\left(-\sqrt{\frac{\lambda}{t}}\left(\frac{t}{\mu} + 1\right)\right).$$

l'estimateur du maximum de vraisemblance $\hat{\gamma} = (\hat{r}, \hat{\mu}, \hat{\lambda})^T$ du paramètre γ maximisant la fonction de log-vraisemblance

$$\begin{aligned} \ell(r, \mu, \lambda) = & \frac{m_1 + m_2}{2} \ln \lambda - \frac{m_2}{2} \ln r + \frac{(m_1 + m_2)\lambda}{\mu} - \frac{3}{2} \sum_{i=1}^{n_1} \delta_{1i} \ln(X_{1i}) - \\ & \frac{\lambda}{2\mu^2} \sum_{i=1}^{n_1} \delta_{1i} X_{1i} - \frac{\lambda}{2} \sum_{i=1}^{n_1} \delta_{1i} X_{1i}^{-1} - \frac{3}{2} \sum_{j=1}^{n_2} \delta_{2j} \ln(X_{2j}) - \frac{\lambda r}{2\mu^2} \sum_{j=1}^{n_2} \delta_{2j} X_{2j} - \frac{\lambda}{2r} \sum_{j=1}^{n_2} \delta_{2j} X_{2j}^{-1} + \\ & \sum_{i=1}^{n_1} (1 - \delta_{1i}) \ln \left\{ \Phi\left(-\sqrt{\frac{\lambda}{X_{1i}}}\left(\frac{X_{1i}}{\mu} - 1\right)\right) - \exp\left(\frac{2\lambda}{\mu}\right)\Phi\left(-\sqrt{\frac{\lambda}{X_{1i}}}\left(\frac{X_{1i}}{\mu} + 1\right)\right) \right\} + \\ & \sum_{j=1}^{n_2} (1 - \delta_{2j}) \ln \left\{ \Phi\left(-\sqrt{\frac{\lambda}{rX_{2j}}}\left(\frac{rX_{2j}}{\mu} - 1\right)\right) - \exp\left(\frac{2\lambda}{\mu}\right)\Phi\left(-\sqrt{\frac{\lambda}{rX_{2j}}}\left(\frac{rX_{2j}}{\mu} + 1\right)\right) \right\}. \end{aligned}$$

En calculant les dérivées de $\ell(r, \mu, \lambda)$ par rapport à r, μ, λ , nous trouvons les fonction de score (le gradient) :

$$\begin{aligned} \frac{\partial \ell}{\partial r} &= -\frac{m_2}{2r} - \frac{\lambda}{2\mu^2} \sum_{i=1}^{n_2} \delta_{2j} X_{2i} + \frac{\lambda}{2r^2} \sum_{i=1}^{n_2} \delta_{2j} X_{2i}^{-1} + \\ &\frac{1}{2r} \sum_{i=1}^{n_2} (1 - \delta_{2j}) \frac{B(rX_{2j})\varphi(A(rX_{2j})) - e^{\frac{2\lambda}{\mu}} A(rX_{2j})\varphi(B(rX_{2j}))}{S_1(rX_{2j}, \mu, \lambda)}, \\ \\ \frac{\partial \ell}{\partial \mu} &= -\frac{\lambda}{\mu^2} (m_1 + m_2) + \frac{\lambda}{\mu^3} \sum_{i=1}^{n_1} \delta_{1i} X_{1i} + \frac{\lambda r}{\mu^3} \sum_{i=1}^{n_2} \delta_{2j} X_{2i} + \\ &\frac{1}{\mu^2} \sum_{i=1}^{n_1} (1 - \delta_{1i}) \frac{\sqrt{\lambda X_{1i}} \left(\varphi(A(X_{1i})) - e^{\frac{2\lambda}{\mu}} \varphi(B(X_{1i})) \right) + 2\lambda e^{\frac{2\lambda}{\mu}} \Phi(B(X_{1i}))}{S_1(X_{1i}, \mu, \lambda)} + \\ &\frac{1}{\mu^2} \sum_{j=1}^{n_2} (1 - \delta_{2j}) \frac{\sqrt{\lambda r X_{2j}} \left(\varphi(A(rX_{2j})) - e^{\frac{2\lambda}{\mu}} \varphi(B(rX_{2j})) \right) + 2\lambda e^{\frac{2\lambda}{\mu}} \Phi(B(rX_{2j}))}{S_1(rX_{2j}, \mu, \lambda)}, \end{aligned}$$

$$\begin{aligned} \frac{\partial \ell}{\partial \lambda} &= \left(\frac{1}{2\lambda} + \frac{1}{\mu} \right) (m_1 + m_2) - \frac{1}{2\mu^2} \sum_{i=1}^{n_1} \delta_{1i} X_{1i} - \frac{1}{2} \sum_{i=1}^{n_1} \delta_{1i} X_{1i}^{-1} - \frac{r}{2\mu^2} \sum_{j=1}^{n_2} \delta_{2j} X_{2j} - \frac{1}{2r} \sum_{j=1}^{n_2} \delta_{2j} X_{2i}^{-1} + \\ &\sum_{i=1}^{n_1} (1 - \delta_{1i}) \frac{\frac{1}{2\lambda} \left[A(X_{1i})\varphi(A(X_{1i})) - e^{\frac{2\lambda}{\mu}} B(X_{1i})\varphi(B(X_{1i})) \right] - \frac{2}{\mu} e^{\frac{2\lambda}{\mu}} \Phi(B(X_{1i}))}{S_1(X_{1i}, \mu, \lambda)} + \\ &\sum_{j=1}^{n_2} (1 - \delta_{2j}) \frac{\frac{1}{2\lambda} \left[A(rX_{2j})\varphi(A(rX_{2j})) - e^{\frac{2\lambda}{\mu}} B(rX_{2j})\varphi(B(rX_{2j})) \right] - \frac{2}{\mu} e^{\frac{2\lambda}{\mu}} \Phi(B(rX_{2j}))}{S_1(rX_{2j}, \mu, \lambda)}, \end{aligned}$$

avec

$$A(t) = -\sqrt{\frac{\lambda}{t}} \left(\frac{t}{\mu} - 1 \right), \quad B(t) = -\sqrt{\frac{\lambda}{t}} \left(\frac{t}{\mu} + 1 \right).$$

Pour déterminer l'estimateur du maximum de vraisemblance $\hat{\gamma}$ de γ , nous pouvons résoudre le système non linéaire formé par les fonctions de score en égalisant celles-ci à zéro.

Afin de simplifier les calculs, supposons maintenant que les données sont complètes. Les dérivées partielles secondes de la fonction de log-vraisemblance sont :

$$\frac{\partial^2 \ell}{\partial r^2} = +\frac{n_2}{2r^2} - \frac{\lambda}{r^3} \sum_{i=1}^{n_2} T_{2i}^{-1},$$

$$\frac{\partial^2 \ell}{\partial r \partial \mu} = \frac{\lambda}{\mu^3} \sum_{i=1}^{n_2} T_{2i},$$

$$\frac{\partial^2 \ell}{\partial r \partial \lambda} = -\frac{1}{2\mu^2} \sum_{i=1}^{n_2} T_{2i} + \frac{1}{2r^2} \sum_{i=1}^{n_2} T_{2i}^{-1},$$

$$\frac{\partial^2 \ell}{\partial \mu^2} = -\frac{3\lambda}{\mu^4} \sum_{i=1}^{n_1} T_{1i} - \frac{3\lambda r}{\mu^4} \sum_{i=1}^{n_2} T_{2i} + \frac{2\lambda n}{\mu^3},$$

$$\frac{\partial^2 \ell}{\partial \mu \partial \lambda} = \frac{1}{\mu^3} \sum_{i=1}^{n_1} T_{1i} + \frac{r}{\mu^3} \sum_{i=1}^{n_2} T_{2i} - \frac{n}{\mu^2},$$

$$\frac{\partial^2 \ell}{\partial \lambda^2} = -\frac{n}{2\lambda^2}.$$

Calculons la matrice d'information de Fisher :

$$I(\gamma) = -E \left(\frac{\partial^2 \ell(\gamma)}{\partial \gamma_i \partial \gamma_j} \right)_{3 \times 3}, \quad \gamma = (r, \mu, \lambda)^T,$$

Posons

$$I(\gamma) = \begin{pmatrix} I_{11} & I_{12} & I_{13} \\ I_{21} & I_{22} & I_{23} \\ I_{31} & I_{32} & I_{33} \end{pmatrix}.$$

L'élément I_{11} est :

$$I_{11} = -E \left(\frac{\partial^2 \ell}{\partial r^2} \right) = -E \left(\frac{n_2}{2r^2} - \frac{\lambda}{r^3} \sum_{i=1}^{n_2} T_{2i}^{-1} \right) = -\frac{n_2}{2r^2} + \frac{\lambda}{r^3} \sum_{i=1}^{n_2} E \left(T_{2i}^{-1} \right).$$

Comme $F_2(t) = F_1(rt)$, alors en dérivant les deux membres, nous trouvons

$$f_2(t) = r f_1(rt),$$

alors :

$$\begin{aligned} E \left(T_{2i}^{-1} \right) &= E \left(\frac{1}{T_{2i}} \right) = \int_0^{+\infty} \frac{1}{t} f_2(t) dt = r \int_0^{+\infty} \frac{1}{t} f_1(rt) dt = r \int_0^{+\infty} \frac{1}{y} f_1(y) dy = \\ &= r E \left(\frac{1}{T_{1i}} \right) = r \left(\frac{1}{\mu} + \frac{1}{\lambda} \right), \end{aligned}$$

d'après la proposition 1.1 (chapitre 1).

Donc

$$I_{11} = \left(\frac{1}{2} + \frac{\lambda}{\mu} \right) \frac{n_2}{r^2}.$$

De la même manière nous calculons les autres éléments de la matrice $I(\gamma)$. Nous trouvons alors :

$$I(\gamma) = \begin{pmatrix} \frac{(\mu+2\lambda)n_2}{2\mu r^2} & -\frac{\lambda n_2}{r\mu^2} & -\frac{n_2}{2r\lambda} \\ -\frac{\lambda n_2}{r\mu^2} & \frac{\lambda n}{\mu^3} & 0 \\ -\frac{n_2}{2r\lambda} & 0 & \frac{n}{2\lambda^2} \end{pmatrix}.$$

La matrice inverse de $I(\gamma)$ est donnée par :

$$I^{-1}(\gamma) = \begin{pmatrix} \frac{2nr^2\mu}{n_1n_2(\mu+2\lambda)} & \frac{2r\mu^2}{n_1(\mu+2\lambda)} & \frac{2r\mu\lambda}{n_1(\mu+2\lambda)} \\ \frac{2r\mu^2}{n_1(\mu+2\lambda)} & \frac{(n_1\mu+2n\lambda)\mu^3}{n_1n(\mu+2\lambda)\lambda} & \frac{2n_2\mu^2\lambda}{n_1n(\mu+2\lambda)} \\ \frac{2r\mu\lambda}{n_1(\mu+2\lambda)} & \frac{2n_2\mu^2\lambda}{n_1n(\mu+2\lambda)} & \frac{2(n\mu+2n_1\lambda)\lambda^2}{n_1n(\mu+2\lambda)} \end{pmatrix}.$$

Si nous prenons $j = 2$, la fonction de répartition $K_2(t)$ est estimée par :

$$\hat{K}_2(t) = \int_0^t \sqrt{\frac{\hat{\lambda}}{2\pi y^3}} \exp\left\{-\frac{\hat{\lambda}(y - \hat{\mu})^2}{2\hat{\mu}^2 y}\right\} \Phi\left(\sqrt{\frac{\hat{\lambda}}{t + \hat{r}y - y}} \left(\frac{t + \hat{r}y - y}{\hat{\mu}} - 1\right)\right) dy +$$

$$\int_0^t \sqrt{\frac{\hat{\lambda}}{2\pi y^3}} \exp\left\{-\frac{\hat{\lambda}(y - \hat{\mu})^2}{2\hat{\mu}^2 y} + \frac{2\hat{\lambda}}{\hat{\mu}}\right\} \Phi\left(-\sqrt{\frac{\hat{\lambda}}{t + \hat{r}y - y}} \left(\frac{t + \hat{r}y - y}{\hat{\mu}} + 1\right)\right) dy.$$

Pour calculer \hat{K}_2 , nous utilisons des méthodes numériques d'intégration.

Dans la figure (4.1), nous représentons les estimateurs des fonctions \hat{F}_1 , \hat{K}_2 , \hat{K}_3 et \hat{K}_4 .

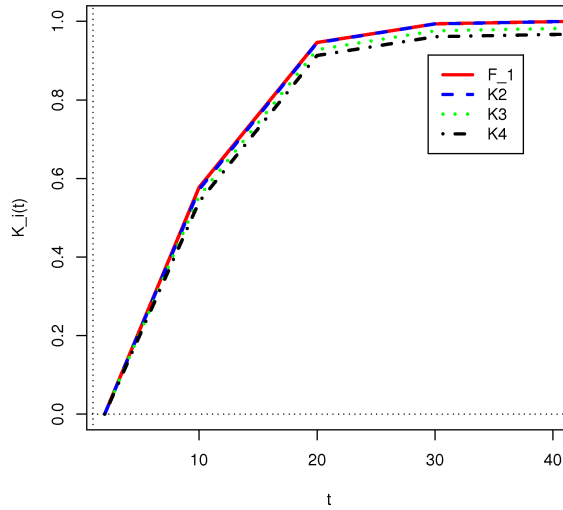


Figure 4.1 – Graphes des estimateurs des fonctions de répartition \hat{F}_1 , \hat{K}_2 , \hat{K}_3 et \hat{K}_4 , (la distribution Gaussienne inverse).

Au seuil α , l'intervalle de confiance $(\underline{K}_j(t), \overline{K}_j(t))$ de $K_j(t)$ est

$$\underline{K}_j(t) = \left(1 + \frac{1 - \hat{K}_j(t)}{\hat{K}_j(t)} \exp \left\{ \frac{\hat{\sigma}_{\hat{K}_j} z_{1-\alpha/2}}{\sqrt{\hat{K}_j(t)(1 - \hat{K}_j(t))}} \right\} \right)^{-1},$$

$$\overline{K}_j(t) = \left(1 + \frac{1 - \hat{K}_j(t)}{\hat{K}_j(t)} \exp \left\{ -\frac{\hat{\sigma}_{\hat{K}_j} z_{1-\alpha/2}}{\sqrt{\hat{K}_j(t)(1 - \hat{K}_j(t))}} \right\} \right)^{-1},$$

où

$$\hat{\sigma}_{\hat{K}_j(t)}^2 = C_2^T(t, \hat{\gamma}) I^{-1}(\hat{\gamma}) C_2(t, \hat{\gamma}),$$

$$C_2(t, \gamma) = (C_{21}(t, \gamma), C_{22}(t, \gamma), C_{23}(t, \gamma))^T.$$

Les composantes du vecteur $C_2(t, \gamma)$ sont :

$$C_{21}(t, \gamma) = \int_0^t \frac{\partial F_1}{\partial r}(t + ry - y, \mu, \lambda) dF_1(y, \mu, \lambda),$$

$$C_{22}(t, \gamma) = \int_0^t \frac{\partial F_1}{\partial \mu}(t + ry - y, \mu, \lambda) dF_1(y, \mu, \lambda) + F_1(t + ry - y, \mu, \lambda) d\left(\frac{\partial F_1}{\partial \mu}(y, \mu, \lambda)\right),$$

$$C_{23}(t, \gamma) = \int_0^t \frac{\partial F_1}{\partial \lambda}(t + ry - y, \mu, \lambda) dF_1(y, \mu, \lambda) + F_1(t + ry - y, \mu, \lambda) d\left(\frac{\partial F_1}{\partial \lambda}(y, \mu, \lambda)\right),$$

avec

$$\frac{\partial F_1}{\partial r}(t + ry - y, \mu, \lambda) = \frac{y}{2} \sqrt{\frac{\lambda}{t + ry - y}} \left(\frac{1}{t + ry - y} + \frac{1}{\mu} \right) \varphi(A_1(t + ry - y)) -$$

$$\frac{y}{2} \exp\left(\frac{2\lambda}{\mu}\right) \sqrt{\frac{\lambda}{t + ry - y}} \left(-\frac{1}{t + ry - y} + \frac{1}{\mu} \right) \varphi(B_1(t + ry - y)),$$

$$\frac{\partial F_1}{\partial \mu}(t + ry - y, \mu, \lambda) = \frac{\sqrt{\lambda(t + ry - y)}}{\mu^2} \left(-\varphi(A_1(t + ry - y)) + \varphi(B_1(t + ry - y)) e^{2\lambda/\mu} \right) -$$

$$\exp\left(\frac{2\lambda}{\mu}\right)\frac{2\lambda}{\mu^2}\Phi(-B_1(t+ry-y)),$$

$$\begin{aligned} \frac{\partial F_1}{\partial \lambda}(t+ry-y; \mu, \lambda) &= \frac{1}{2\lambda}A_1(t+ry-y)\varphi(A_1(t+ry-y))+ \\ \frac{2}{\mu}\exp\left(\frac{2\lambda}{\mu}\right)\Phi(-B_1(t+ry-y)) &- \frac{1}{2\lambda}\exp\left(\frac{2\lambda}{\mu}\right)B_1(t+ry-y)\varphi(B_1(t+ry-y)), \end{aligned}$$

on peut aussi écrire

$$C_{21}(t, \gamma) = \int_0^t \frac{\partial F_1}{\partial r}(t+ry-y, \mu, \lambda) f_1(y, \mu, \lambda) dy,$$

$$C_{22}(t, \gamma) = \int_0^t \left(\frac{\partial F_1}{\partial \mu}(t+ry-y, \mu, \lambda) f_1(y, \mu, \lambda) + F_1(t+ry-y, \mu, \lambda) \left[\frac{\partial F_1}{\partial \mu}(y, \mu, \lambda) \right]'_y \right) dy,$$

$$C_{23}(t, \gamma) = \int_0^t \left(\frac{\partial F_1}{\partial \lambda}(t+ry-y, \mu, \lambda) f_1(y, \mu, \lambda) + F_1(t+ry-y, \mu, \lambda) \left[\frac{\partial F_1}{\partial \lambda}(y, \mu, \lambda) \right]'_y \right) dy,$$

où

$$\left[\frac{\partial F_1}{\partial \mu}(y, \mu, \lambda) \right]'_y = \frac{1}{2\mu^2} \sqrt{\frac{\lambda}{y}} [1 - A_1(y)B_1(y)] \left(-\varphi(A_1(y)) + \exp\left(\frac{2\lambda}{\mu}\right)\varphi(B_1(y)) \right) +$$

$$\frac{\lambda}{\mu^2 y} \exp\left(\frac{2\lambda}{\mu}\right) A_1(y) \varphi(B_1(y)),$$

$$\left[\frac{\partial F_1}{\partial \lambda}(y, \mu, \lambda) \right]'_y = \frac{1}{4\lambda y} B_1(y) (1 - A_1^2(y)) \varphi(A_1(y)) -$$

$$\frac{1}{\mu y} \exp\left(\frac{2\lambda}{\mu}\right) A_1(y) \varphi(B_1(y)) - \frac{1}{4\lambda y} A_1(y) (1 - B_1^2(y)) \varphi(B_1(y)).$$

avec

$$A_1(t) = \sqrt{\frac{\lambda}{t}} \left(\frac{t}{\mu} - 1 \right), \quad B_1(t) = \sqrt{\frac{\lambda}{t}} \left(\frac{t}{\mu} + 1 \right)$$

t	4	8	10	20	30	40	50
$K_2(t)$	0.05099	0.55988	0.75297	0.99049	0.99966	0.99998	0.99999
N.C.	0.9367	0.8887	0.8355	0.8660	0.8904	0.9104	0.8722

Tableau 4.1 – Niveau de confiance pour $K_2(t)$,
(la distribution Gaussienne inverse).

6.1 Simulation

Considérons le cas où les données sont complètes. Nous générons des échantillons de taille $n_1 = n_2 = 100$. Nous simulons les temps de pannes T_{1i} et T_{2j} à partir de la distribution Gaussienne inverse avec les paramètres :

$$T_{1i} \sim IG(\mu_1, \lambda_1), \quad T_{2j} \sim IG(\mu_2, \lambda_2),$$

$$\mu_1 = 8, \quad \mu_2 = 4, \quad \lambda_1 = 40, \quad \lambda_2 = 20.$$

Chaque échantillon est répété 5000 fois. Pour différentes valeurs de t , et pour chaque échantillon nous calculons par simulation les intervalles de confiance au seuil (risque) $\alpha = 0.1$ en appliquant les formules (4.7) et (4.8), puis nous calculons le niveau de confiance (N.C.) qui compte le nombre de fois où les intervalles de confiance couvrant la valeur réelle de la fonction de répartition $K_2(t)$ donnée par la formule (4.6) puis en divisant par 5000. Les résultats sont résumés dans le tableau 4.1.

7 La distribution log-normale

Supposons maintenant que les fonctions de distribution des unités appartiennent à la famille des distributions log-normales [Nikulin et Saaidia et Tahir (2011a,2011b)], i.e.

$$S_1(t) = 1 - F_1(t) = 1 - \Phi\left(\frac{\ln t - m}{\sigma}\right).$$

la fonction de log-vraisemblance est

$$\ell(r, m, \sigma) = -(m_1 + m_2) \ln \sigma - \sum_{i=1}^{n_1} \delta_{1i} \ln X_{1i} - \sum_{j=1}^{n_2} \delta_{2j} \ln X_{2j} +$$

$$-\frac{1}{2\sigma^2} \sum_{i=1}^{n_1} \delta_{1i} (\ln X_{1i} - m)^2 - \frac{1}{2\sigma^2} \sum_{j=1}^{n_2} \delta_{2j} (\ln r + \ln X_{2j} - m)^2 +$$

$$\sum_{i=1}^{n_1} (1 - \delta_{1i}) \ln \left\{ 1 - \Phi \left(\frac{\ln X_{1i} - m}{\sigma} \right) \right\} + \sum_{j=1}^{n_2} (1 - \delta_{2j}) \ln \left\{ 1 - \Phi \left(\frac{\ln r + \ln X_{2j} - m}{\sigma} \right) \right\}.$$

Supposons maintenant que les données sont complètes, les fonctions de scores sont :

$$\frac{\partial \ell}{\partial r} = \frac{n_2}{r\sigma^2} (m - \ln r) - \frac{1}{r\sigma^2} \sum_{j=1}^{n_2} \ln(T_{2j}),$$

$$\frac{\partial \ell}{\partial m} = -\frac{nm}{\sigma^2} + \frac{n_2 \ln r}{\sigma^2} + \frac{1}{\sigma^2} \sum_{i=1}^{n_1} \ln(T_{1i}) + \frac{1}{\sigma^2} \sum_{j=1}^{n_2} \ln(T_{2j}),$$

$$\frac{\partial \ell}{\partial \sigma} = -n \left(\frac{1}{\sigma} - \frac{m^2}{\sigma^3} \right) - \frac{n_2 \ln r}{\sigma^3} (2m - \ln r) - \frac{2m}{\sigma^3} \sum_{i=1}^{n_1} \ln(T_{1i}) +$$

$$\frac{2}{\sigma^3} (-m + \ln r) \sum_{j=1}^{n_2} \ln(T_{2j}) + \frac{1}{\sigma^3} \sum_{i=1}^{n_1} \ln^2(T_{1i}) + \frac{1}{\sigma^3} \sum_{j=1}^{n_2} \ln^2(T_{2j}).$$

Pour déterminer l'estimateur du maximum de vraisemblance $\hat{\gamma}$ de γ , nous pouvons résoudre le système non linéaire formé par les fonctions de score en égalisant celles-ci à zéro.

Les dérivées secondes de la fonction log-vraisemblance sont :

$$\frac{\partial^2 \ell}{\partial r^2} = -\frac{n_2}{r^2\sigma^2} (m + \ln r + 1) + \frac{1}{r^2\sigma^2} \sum_{j=1}^{n_2} \ln(T_{2j}); \quad \frac{\partial^2 \ell}{\partial r \partial m} = \frac{n_2}{r\sigma^2};$$

$$\frac{\partial^2 \ell}{\partial r \partial \sigma} = \frac{2n_2}{r\sigma^3} (-m + \ln r) + \frac{2}{r\sigma^3} \sum_{j=1}^{n_2} \ln(T_{2j}); \quad \frac{\partial^2 \ell}{\partial m^2} = -\frac{n}{\sigma^2};$$

$$\frac{\partial^2 \ell}{\partial m \partial \sigma} = \frac{2nm}{\sigma^3} - 2n_2 \frac{\ln r}{\sigma^3} - \frac{2}{\sigma^3} \sum_{i=1}^{n_1} \ln(T_{1i}) - \frac{2}{\sigma^3} \sum_{j=1}^{n_2} \ln(T_{2j});$$

$$\frac{\partial^2 \ell}{\partial \sigma^2} = -n \left(-\frac{1}{\sigma^2} + \frac{3m^2}{\sigma^4} \right) + \frac{3n_2}{\sigma^4} (2m \ln r - \ln^2 r) + \frac{6m}{\sigma^4} \sum_{i=1}^{n_1} \ln(T_{1i}) +$$

$$\frac{6}{\sigma^4} (m - \ln r) \sum_{j=1}^{n_2} \ln(T_{2j}) - \frac{3}{\sigma^4} \sum_{i=1}^{n_1} \ln^2(T_{1i}) - \frac{3}{\sigma^4} \sum_{j=1}^{n_2} \ln^2(T_{2j}).$$

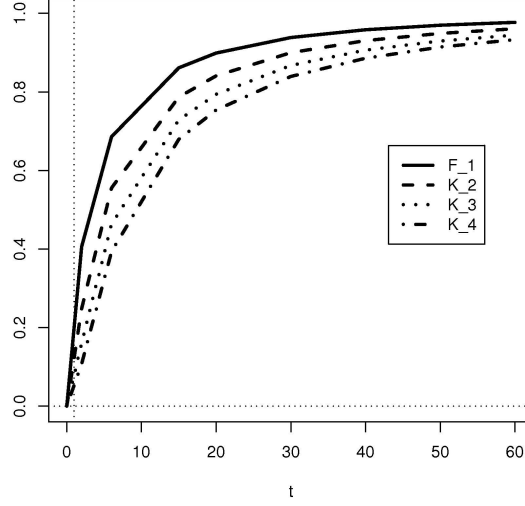


Figure 4.2 – Graphes des estimateurs des fonctions de répartition $\hat{F}_1, \hat{K}_2, \hat{K}_3$ et \hat{K}_4 , (distribution log-normale).

Alors la matrice d'information de Fisher est

$$I(\gamma) = \begin{pmatrix} \frac{n_2}{r^2\sigma^2} & -\frac{n_2}{r\sigma^2} & 0 \\ -\frac{n_2}{r\sigma^2} & \frac{n}{\sigma^2} & 0 \\ 0 & 0 & \frac{2n}{\sigma^2} \end{pmatrix}.$$

L'inverse de cette matrice est

$$I^{-1}(\gamma) = \begin{pmatrix} \frac{nr^2\sigma^2}{n_1n_2} & \frac{r\sigma^2}{n_1} & 0 \\ \frac{r\sigma^2}{n_1} & \frac{\sigma^2}{n_1} & 0 \\ 0 & 0 & \frac{\sigma^2}{2n} \end{pmatrix}.$$

En prenant $j = 2$, the c.d.f. $K_2(t)$ est estimée par

$$\hat{K}_2(t) = \frac{1}{\sqrt{2\pi}\hat{\sigma}} \int_0^t \frac{1}{y} \Phi \left(\frac{\ln(t + \hat{r}y - y) - \hat{m}}{\hat{\sigma}} \right) \exp \left\{ -\frac{\ln y - \hat{m}}{2\hat{\sigma}^2} \right\} dy.$$

Dans la figure (4.2) nous représentons les graphes des fonctions de répartition $\hat{F}_1, \hat{K}_2, \hat{K}_3$ et \hat{K}_4 .

Au seuil α , l'intervalle de confiance $(\underline{K}_j(t), \overline{K}_j(t))$ de $K_j(t)$ est

$$\underline{K}_j(t) = \left(1 + \frac{1 - \hat{K}_j(t)}{\hat{K}_j(t)} \exp \left\{ \frac{\hat{\sigma}_{\hat{K}_j} z_{1-\alpha/2}}{\sqrt{\hat{K}_j(t)(1 - \hat{K}_j(t))}} \right\} \right)^{-1},$$

$$\bar{K}_j(t) = \left(1 + \frac{1 - \hat{K}_j(t)}{\hat{K}_j(t)} \exp \left\{ -\frac{\hat{\sigma}_{\hat{K}_j} z_{1-\alpha/2}}{\sqrt{\hat{K}_j(t)(1 - \hat{K}_j(t))}} \right\} \right)^{-1},$$

où

$$\hat{\sigma}_{\hat{K}_j(t)}^2 = C_2^T(t, \hat{\gamma}) I^{-1}(\hat{\gamma}) C_2(t, \hat{\gamma}),$$

$$\mathbf{C}_2(t, \gamma) = (\mathbf{C}_{21}(t, \gamma), \mathbf{C}_{22}(t, \gamma), \mathbf{C}_{23}(t, \gamma))^T,$$

$$\mathbf{C}_{21}(t, \gamma) = \int_0^t \frac{\partial F_1}{\partial r}(t + ry - y, m, \sigma) dF_1(y, m, \sigma),$$

$$\mathbf{C}_{22}(t, \gamma) = \int_0^t \frac{\partial F_1}{\partial m}(t + ry - y, m, \sigma) dF_1(y, m, \sigma) + F_1(t + ry - y, m, \sigma) d\left(\frac{\partial F_1}{\partial m}(y, m, \sigma)\right),$$

$$\mathbf{C}_{23}(t, \gamma) = \int_0^t \frac{\partial F_1}{\partial \sigma}(t + ry - y, m, \sigma) dF_1(y, \mu, \lambda) + F_1(t + ry - y, m, \sigma) d\left(\frac{\partial F_1}{\partial \sigma}(y, m, \sigma)\right).$$

les dérivées partielles sont

$$\frac{\partial F_1}{\partial r}(t + ry - y, m, \sigma) = \frac{y}{\sigma(t + ry - y)} \varphi\left(\frac{\ln(t + ry - y) - m}{\sigma}\right),$$

$$\frac{\partial F_1}{\partial m}(t + ry - y; m, \sigma) = -\frac{1}{\sigma} \varphi\left(\frac{\ln(t + ry - y) - m}{\sigma}\right),$$

$$\frac{\partial F_1}{\partial \sigma}(t + ry - y; m, \sigma) = -\frac{\ln(t + ry - y) - m}{\sigma^2} \varphi\left(\frac{\ln(t + ry - y) - m}{\sigma}\right),$$

nous pouvons aussi écrire :

$$\mathbf{C}_{21}(t, \gamma) = \frac{1}{\sqrt{2\pi}} \int_0^t \frac{1}{\sigma^2(t + ry - y)} \exp\left\{-\frac{(\ln y - m)^2}{2\sigma^2}\right\} \varphi\left(\frac{\ln(t + ry - y) - m}{\sigma}\right) dy,$$

$$\mathbf{C}_{22}(t, \gamma) = -\frac{1}{\sigma^2 \sqrt{2\pi}} \int_0^t \frac{1}{y} \exp\left\{-\frac{(\ln y - m)^2}{2\sigma^2}\right\} \varphi\left(\frac{\ln(t + ry - y) - m}{\sigma}\right) dy +$$

$$\frac{1}{\sigma^3} \int_0^t \Phi\left(\frac{\ln(t + ry - y) - m}{\sigma}\right) \left(\frac{\ln y - m}{y}\right) \varphi\left(\frac{\ln y - m}{\sigma}\right) dy,$$

$$\mathbf{C}_{23}(t, \gamma) = -\frac{1}{\sigma^3 \sqrt{2\pi}} \int_0^t \frac{\ln(t + ry - y) - m}{\sigma^2} \exp\left\{-\frac{(\ln y - m)^2}{2\sigma^2}\right\} \varphi\left(\frac{\ln(t + ry - y) - m}{\sigma}\right) dy +$$

t	10	50	100	200	300
$K_2(t)$	0.7119	0.9556	0.9855	0.9961	0.9983
N.C. (%)	82.25	88.87	83.55	86.60	89.04

Tableau 4.2 – Niveau de confiance pour $K_2(t)$,
(la distribution log-normale).

$$\frac{1}{\sigma^2} \int_0^t \frac{1}{y} \Phi \left(\frac{\ln(t + ry - y) - m}{\sigma} \right) \left(\frac{(\ln y - m)^2}{\sigma^2} - 1 \right) \varphi \left(\frac{\ln y - m}{\sigma} \right) dy.$$

7.1 Simulation

Considérons le cas des données complètes de taille $n_1 = n_2 = 100$, chaque échantillon est répété 5000 fois. Nous trouvons par simulation les intervalles de confiance donnée par les formules (4.7) et (4.8) au seuil $\alpha = 0.1$. Nous simulons les temps de pannes T_{1i} et T_{2j} à partir de la distribution log-normale avec les paramètres :

$$T_{1i} \sim LN(m_1, \sigma_1), \quad T_{2j} \sim LN(m_2, \sigma_2)$$

$$m_1 = 1, \quad m_2 = -0.3862944, \quad \sigma_1 = \sigma_2 = 1.5.$$

Pour différentes valeurs de t , les proportions des intervalles de confiance (N.C.) couvrant la vraie valeur de la fonction de répartition $K_2(t)$ (4.6) sont données dans le tableau (4.2).

8 La distribution log-logistique

Cette distribution est bien étudiée par Bagdonavičius et Masiulaityle et Nikulin (2010). Supposons que les distribution des temps de pannes en mode "hot" et "warm" suivent la loi log-logistique, i.e.

$$S_1(t) = 1 - F_1(t) = \frac{1}{1 + \left(\frac{t}{\mu}\right)^\nu}.$$

la fonction de log-vraisemblance est

$$\ell(r, \mu, \nu) = (m_1 + m_2) \ln \nu - (m_1 + m_2) \nu \ln \mu + m_2 \nu \ln r + (\nu - 1) \left(\sum_{i=1}^{n_1} \delta_{1i} \ln X_{1i} + \sum_{j=1}^{n_2} \delta_{2j} \ln X_{2j} \right) -$$

$$\sum_{i=1}^{n_1} (1 + \delta_{1i}) \ln \left(1 + \left(\frac{X_{1i}}{\mu} \right)^\nu \right) - \sum_{j=1}^{n_2} (1 + \delta_{2j}) \ln \left(1 + \left(\frac{r X_{2j}}{\mu} \right)^\nu \right).$$

Les dérivées partielles sont données par

$$\begin{aligned}\frac{\partial \ell}{\partial r} &= \frac{\nu m_2}{r} - \frac{\nu}{r} \sum_{j=1}^{n_2} (1 + \delta_{2j}) \frac{\left(\frac{rX_{2j}}{\mu}\right)^\nu}{1 + \left(\frac{rX_{2j}}{\mu}\right)^\nu}, \\ \frac{\partial \ell}{\partial \mu} &= -\frac{\nu m}{\mu} + \frac{\nu}{\mu} \sum_{i=1}^{n_1} (1 + \delta_{1i}) \frac{\left(\frac{X_{1i}}{\mu}\right)^\nu}{1 + \left(\frac{X_{1i}}{\mu}\right)^\nu} + \frac{\nu}{\mu} \sum_{j=1}^{n_2} (1 + \delta_{2j}) \frac{\left(\frac{rX_{2j}}{\mu}\right)^\nu}{1 + \left(\frac{rX_{2j}}{\mu}\right)^\nu}, \\ \frac{\partial \ell}{\partial \nu} &= \frac{m_1 + m_2}{\nu} - (m_1 + m_2) \ln \mu + m_2 \ln r + \sum_{i=1}^{n_1} \delta_{1i} \ln X_{1i} + \sum_{j=1}^{n_2} \delta_{2j} \ln X_{2j} - \\ &\quad \sum_{i=1}^{n_1} (1 + \delta_{1i}) \frac{\left(\frac{X_{1i}}{\mu}\right)^\nu \ln \left(\frac{X_{1i}}{\mu}\right)}{1 + \left(\frac{X_{1i}}{\mu}\right)^\nu} - \sum_{j=1}^{n_2} (1 + \delta_{2j}) \frac{\left(\frac{rX_{2j}}{\mu}\right)^\nu \ln \left(\frac{rX_{2j}}{\mu}\right)}{1 + \left(\frac{rX_{2j}}{\mu}\right)^\nu}.\end{aligned}$$

Dans le cas des données complètes, la matrice d'information de Fisher devient :

$$I(r, \mu, \nu) = \begin{pmatrix} \frac{n_2 \nu^2}{3r^2} & -\frac{n_2 \nu^2}{3r\mu} & 0 \\ -\frac{n_2 \nu^2}{3r\mu} & \frac{n_2 \nu^2}{3\mu^2} & 0 \\ 0 & 0 & \frac{n\{3+2\Gamma''(2)-2[\Gamma'(2)]^2\}}{3\nu^2} \end{pmatrix}.$$

Son inverse est

$$I^{-1}(r, \mu, \nu) = \begin{pmatrix} \frac{3nr^2}{n_1 n_2 \nu^2} & \frac{3r\mu}{n_1 \nu^2} & 0 \\ \frac{3r\nu}{n_1 \nu^2} & \frac{3\mu^2}{n_1 \nu^2} & 0 \\ 0 & 0 & \frac{3\nu^2}{n\{3+2\Gamma''(2)-2[\Gamma'(2)]^2\}} \end{pmatrix},$$

où $\Gamma'(\cdot)$ et $\Gamma''(\cdot)$ sont respectivement les dérivées d'ordre 1 et 2 de la fonction gamma.

La fonction de répartition $K_2(t)$ est estimée par

$$\hat{K}_2(t) = 1 - \frac{1}{1 + \left(\frac{t}{\hat{\mu}}\right)^{\hat{\nu}}} - \frac{\hat{\nu}}{\hat{\mu}} \int_0^t \left(\frac{y}{\hat{\mu}}\right)^{\hat{\nu}-1} \frac{1}{1 + \left(\frac{t+\hat{r}y-y}{\hat{\mu}}\right)^{\hat{\nu}}} \frac{1}{\left(1 + \left(\frac{y}{\hat{\mu}}\right)^{\hat{\nu}}\right)^2} dy,$$

et les fonctions C_{2i} sont données par :

$$C_{21}(t; r, \mu, \nu) = \nu^2 \int_0^t \left(\frac{y}{\mu}\right)^\nu \left(\frac{t+ry-y}{\mu}\right)^{\nu-1} \left(1 + \left(\frac{y}{\mu}\right)^\nu\right)^{-2} \left(1 + \left(\frac{t+ry-y}{\mu}\right)^\nu\right)^{-2} dy,$$

t	50	100	200	300	400	500
$K_2(t)$	0.016	0.138	0.517	0.743	0.851	0.905
N.C. (%)	89.0	88.8	90.4	89.6	89.5	90.5

Tableau 4.3 – Niveau de confiance pour $K_2(t)$,
(la distribution log-logistique).

$$C_{22}(t; r, \mu, \nu) = -\frac{\nu}{\mu} \left(\frac{t}{\mu}\right)^\nu \left(1 + \left(\frac{t}{\mu}\right)^\nu\right)^{-2} - \int_0^t \frac{\nu^2 \left(\frac{y}{\mu}\right)^{\nu-1}}{\mu^2 \left(1 + \left(\frac{t+ry-y}{\mu}\right)^\nu\right) \left(1 + \left(\frac{y}{\mu}\right)^\nu\right)^2} \times$$

$$\frac{1 - \left(\frac{y}{\mu}\right)^\nu - 2 \left(\frac{y}{\mu}\right)^\nu \left(\frac{t+ry-y}{\mu}\right)^\nu}{\left(1 + \left(\frac{y}{\mu}\right)^\nu\right) \left(1 + \left(\frac{t+ry-y}{\mu}\right)^\nu\right)} dy,$$

$$C_{23}(t; r, \mu, \nu) = \left(\frac{t}{\mu}\right)^\nu \left(1 + \left(\frac{t}{\mu}\right)^\nu\right)^{-2} \ln\left(\frac{t}{\mu}\right) -$$

$$\frac{1}{\mu} \int_0^t \left(\frac{y}{\mu}\right)^{\nu-1} \left(1 + \left(\frac{y}{\mu}\right)^\nu\right)^{-2} \left(1 + \left(\frac{t+ry-y}{\mu}\right)^\nu\right)^{-1} \times$$

$$\left[1 + \left(\frac{y}{\mu}\right)^\nu - \nu \ln\left(\frac{y}{\mu}\right) \left(1 - \left(\frac{y}{\mu}\right)^\nu\right) \left(1 + \left(\frac{y}{\mu}\right)^\nu\right)^{-1} -$$

$$\nu \left(\frac{t+ry-y}{\mu}\right)^\nu \left(1 + \left(\frac{t+ry-y}{\mu}\right)^\nu\right)^{-1} \ln\left(\frac{t+ry-y}{\mu}\right)\right] dy.$$

8.1 Simulation

Considérons le cas des données complètes de taille $n_1 = n_2 = 100$, chaque échantillon est répété 5000 fois. Nous trouvons par simulation les intervalles de confiance donnée par les formules (4.7) et (4.8) au seuil $\alpha = 0.1$. Nous simulons les temps de pannes T_{1i} et T_{2j} à partir de la loi log-logistique avec les paramètres :

$$T_{1i} \sim LL(\alpha_1, \beta_1), \quad T_{2j} \sim LL(\alpha_2, \beta_2)$$

$$\alpha_1 = \alpha_2 = 2, \quad \beta_1 = 100, \quad \beta_2 = 300.$$

Pour différentes valeurs de t , les proportions des intervalles de confiance (N.C.) couvrant la vraie valeur de la fonction de répartition $K_2(t)$ (4.6) sont données dans le tableau (4.3).

9 La distribution de Weibull généralisée

Supposons que les distributions des temps de pannes en mode "hot" et "warm" suivent la loi de Weibull généralisée [saaidia et al. (2011)], i.e.

$$F_1(t, \theta) = 1 - S_1(t, \theta) = 1 - \exp \left\{ 1 - \left(1 + \left(\frac{t}{\sigma} \right)^\nu \right)^{\frac{1}{\gamma}} \right\}, \quad t \geq 0, \quad \theta = (\sigma, \nu, \gamma)^T \in \mathbf{R}_+^* \times \mathbf{R}_+^* \times \mathbf{R}_+^*,$$

la densité correspondante est :

$$f_1(t, \theta) = \frac{\nu}{\gamma \sigma^\nu} t^{\nu-1} \left(1 + \left(\frac{t}{\sigma} \right)^\nu \right)^{\frac{1}{\gamma}-1} \exp \left\{ 1 - \left(1 + \left(\frac{t}{\sigma} \right)^\nu \right)^{\frac{1}{\gamma}} \right\}, \quad t \geq 0,$$

Posons $\gamma_1 = (r, \theta^T)^T$, la fonction de log-vraisemblance est :

$$\begin{aligned} \ell(\gamma_1) &= (n_1 + n_2) + (m_1 + m_2) (\ln \nu - \ln \gamma - \nu \ln \sigma) + \nu m_2 \ln r + \\ &\quad (\nu - 1) \sum_{i=1}^{n_1} \delta_{1i} \ln(X_{1i}) + (\nu - 1) \sum_{j=1}^{n_2} \delta_{2j} \ln(X_{2j}) + \\ &\quad \left(\frac{1}{\gamma} - 1 \right) \sum_{i=1}^{n_1} \delta_{1i} \ln \left(1 + \left(\frac{X_{1i}}{\sigma} \right)^\nu \right) + \left(\frac{1}{\gamma} - 1 \right) \sum_{j=1}^{n_2} \delta_{2j} \ln \left(1 + \left(\frac{r X_{2j}}{\sigma} \right)^\nu \right) - \\ &\quad \sum_{i=1}^{n_1} \left(1 + \left(\frac{X_{1i}}{\sigma} \right)^\nu \right)^{\frac{1}{\gamma}} - \sum_{j=1}^{n_2} \left(1 + \left(\frac{r X_{2j}}{\sigma} \right)^\nu \right)^{\frac{1}{\gamma}}. \end{aligned}$$

Posons : $V_i = \left(\frac{X_{1i}}{\sigma} \right)^\nu$ et $U_j = \left(\frac{r X_{2j}}{\sigma} \right)^\nu$, alors la fonction de log-vraisemblance devient :

$$\begin{aligned} \ell(\gamma_1) &= (n_1 + n_2) + (m_1 + m_2) (\ln \nu - \ln \gamma - \nu \ln \sigma) + \nu m_2 \ln r + \\ &\quad (\nu - 1) \sum_{i=1}^{n_1} \delta_{1i} \ln(X_{1i}) + (\nu - 1) \sum_{j=1}^{n_2} \delta_{2j} \ln(X_{2j}) + \\ &\quad \left(\frac{1}{\gamma} - 1 \right) \sum_{i=1}^{n_1} \delta_{1i} \ln(1 + V_i) + \left(\frac{1}{\gamma} - 1 \right) \sum_{j=1}^{n_2} \delta_{2j} \ln(1 + U_j) - \\ &\quad \sum_{i=1}^{n_1} (1 + V_i)^{\frac{1}{\gamma}} - \sum_{j=1}^{n_2} (1 + U_j)^{\frac{1}{\gamma}}. \end{aligned}$$

Les fonctions de scores sont :

$$\frac{\partial \ell}{\partial r} = \frac{\nu m_2}{r} + \frac{\nu}{r} \left(\frac{1}{\gamma} - 1 \right) \sum_{j=1}^{n_2} \delta_{2j} \frac{U_j}{1 + U_j} - \frac{\nu}{r \gamma} \sum_{j=1}^{n_2} U_j (1 + U_j)^{\frac{1}{\gamma}-1};$$

$$\begin{aligned} \frac{\partial \ell}{\partial \sigma} &= -\frac{\nu}{\sigma}(m_1 + m_2) - \frac{\nu}{\sigma} \left(\frac{1}{\gamma} - 1 \right) \sum_{i=1}^{n_1} \delta_{1i} \frac{V_i}{1 + V_i} - \frac{\nu}{\sigma} \left(\frac{1}{\gamma} - 1 \right) \sum_{j=1}^{n_2} \delta_{2j} \frac{U_j}{1 + U_j} + \\ &\quad \frac{\nu}{\sigma \gamma} \sum_{i=1}^{n_1} V_i (1 + V_i)^{\frac{1}{\gamma} - 1} + \frac{\nu}{\sigma \gamma} \sum_{j=1}^{n_2} U_j (1 + U_j)^{\frac{1}{\gamma} - 1}; \end{aligned}$$

$$\begin{aligned} \frac{\partial \ell}{\partial \nu} &= (m_1 + m_2) \left(\frac{1}{\nu} - \ln \sigma \right) + m_2 \ln r + \sum_{i=1}^{n_1} \delta_{1i} \ln X_{1i} + \sum_{j=1}^{n_2} \delta_{2j} \ln X_{2j} + \\ &\quad \frac{1}{\nu} \left(\frac{1}{\gamma} - 1 \right) \sum_{i=1}^{n_1} \delta_{1i} \frac{V_i \ln V_i}{1 + V_i} + \frac{1}{\nu} \left(\frac{1}{\gamma} - 1 \right) \sum_{j=1}^{n_2} \delta_{2j} \frac{U_j \ln U_j}{1 + U_j} - \\ &\quad \frac{1}{\nu \gamma} \sum_{i=1}^{n_1} V_i \ln V_i (1 + V_i)^{\frac{1}{\gamma} - 1} - \frac{1}{\nu \gamma} \sum_{j=1}^{n_2} U_j \ln U_j (1 + U_j)^{\frac{1}{\gamma} - 1}; \end{aligned}$$

$$\begin{aligned} \frac{\partial \ell}{\partial \gamma} &= -\frac{1}{\gamma}(m_1 + m_2) - \frac{1}{\gamma^2} \sum_{i=1}^{n_1} \delta_{1i} \ln(1 + V_i) - \frac{1}{\gamma^2} \sum_{j=1}^{n_2} \delta_{2j} \ln(1 + U_j) + \\ &\quad \frac{1}{\gamma^2} \sum_{i=1}^{n_1} (1 + V_i)^{\frac{1}{\gamma}} \ln(1 + V_i) + \frac{1}{\gamma^2} \sum_{j=1}^{n_2} (1 + U_j)^{\frac{1}{\gamma}} \ln(1 + U_j). \end{aligned}$$

Pour trouver l'estimateur $\hat{\gamma}_1$ de γ_1 , on peut résoudre le système formé par les fonctions de score en égalisant celles-ci à zéro.

Les dérivées partielles secondes $\ddot{\ell}(\gamma_1)$ sont données par :

$$\begin{aligned} \frac{\partial^2 \ell}{\partial r^2} &= -\frac{\nu m_2}{r^2} - \frac{\nu}{r^2} \left(\frac{1}{\gamma} - 1 \right) \sum_{j=1}^{n_2} \delta_{2j} \frac{U_j}{1 + U_j} + \frac{\nu^2}{r^2} \left(\frac{1}{\gamma} - 1 \right) \sum_{j=1}^{n_2} \delta_{2j} \frac{U_j}{(1 + U_j)^2} + \\ &\quad + \frac{\nu}{r^2 \gamma} \sum_{j=1}^{n_2} U_j (1 + U_j)^{\frac{1}{\gamma} - 1} - \frac{\nu^2}{r^2 \gamma} \sum_{j=1}^{n_2} U_j \left(1 + \frac{1}{\gamma} U_j \right) (1 + U_j)^{\frac{1}{\gamma} - 2}; \end{aligned}$$

$$\frac{\partial^2 \ell}{\partial r \partial \sigma} = -\frac{\nu^2}{r \sigma} \left(\frac{1}{\gamma} - 1 \right) \sum_{j=1}^{n_2} \delta_{2j} \frac{U_j}{(1 + U_j)^2} + \frac{\nu^2}{r \sigma \gamma} \sum_{j=1}^{n_2} U_j \left(1 + \frac{1}{\gamma} U_j \right) (1 + U_j)^{\frac{1}{\gamma} - 2};$$

$$\frac{\partial^2 \ell}{\partial r \partial \nu} = \frac{m_2}{r} + \frac{1}{r} \left(\frac{1}{\gamma} - 1 \right) \sum_{j=1}^{n_2} \delta_{2j} \frac{U_j \ln U_j}{(1 + U_j)^2} + \frac{1}{r} \left(\frac{1}{\gamma} - 1 \right) \sum_{j=1}^{n_2} \delta_{2j} \frac{U_j}{1 + U_j} -$$

$$\frac{1}{r\gamma} \sum_{j=1}^{n_2} U_j \ln U_j \left(1 + \frac{1}{\gamma} U_j\right) (1 + U_j)^{\frac{1}{\gamma}-2} - \frac{1}{r\gamma} \sum_{j=1}^{n_2} U_j (1 + U_j)^{\frac{1}{\gamma}-1};$$

$$\frac{\partial^2 \ell}{\partial r \partial \gamma} = -\frac{\nu}{r\gamma^2} \sum_{j=1}^{n_2} \frac{U_j}{1 + U_j} + \frac{\nu}{r\gamma^2} \sum_{j=1}^{n_2} U_j \left(1 + \frac{1}{\gamma} \ln(1 + U_j)\right) (1 + U_j)^{\frac{1}{\gamma}-1};$$

$$\frac{\partial^2 \ell}{\partial \sigma^2} = \frac{\nu}{\sigma^2} (m_1 + m_2) + \frac{\nu}{\sigma^2} \left(\frac{1}{\gamma} - 1\right) \sum_{i=1}^{n_1} \delta_{1i} \frac{V_i (1 + \nu + V_i)}{(1 + V_i)^2} + \frac{\nu}{\sigma^2} \left(\frac{1}{\gamma} - 1\right) \sum_{j=1}^{n_2} \delta_{2j} \frac{U_j (1 + \nu + U_j)}{(1 + U_j)^2} -$$

$$\frac{\nu}{\sigma^2 \gamma} \sum_{i=1}^{n_1} V_i \left(2 + \left(\frac{1}{\gamma} + 1\right) V_i\right) (1 + V_i)^{\frac{1}{\gamma}-2} - \frac{\nu}{\sigma^2 \gamma} \sum_{j=1}^{n_2} U_j \left(2 + \left(\frac{1}{\gamma} + 1\right) U_j\right) (1 + U_j)^{\frac{1}{\gamma}-2};$$

$$\frac{\partial^2 \ell}{\partial \sigma \partial \nu} = -\frac{1}{\sigma} (m_1 + m_2) - \frac{1}{\sigma} \left(\frac{1}{\gamma} - 1\right) \sum_{i=1}^{n_1} \delta_{1i} \frac{V_i (1 + V_i + \ln V_i)}{(1 + V_i)^2} - \frac{1}{\sigma} \left(\frac{1}{\gamma} - 1\right) \sum_{j=1}^{n_2} \delta_{2j} \frac{U_j (1 + U_j + \ln U_j)}{(1 + U_j)^2} +$$

$$\frac{1}{\sigma \gamma} \sum_{i=1}^{n_1} V_i \left[1 + V_i + \ln V_i + \frac{1}{\gamma} V_i \ln V_i\right] (1 + V_i)^{\frac{1}{\gamma}-2} + \frac{1}{\sigma \gamma} \sum_{j=1}^{n_2} U_j \left[1 + U_j + \ln U_j + \frac{1}{\gamma} U_j \ln U_j\right] (1 + U_j)^{\frac{1}{\gamma}-2};$$

$$\frac{\partial^2 \ell}{\partial \sigma \partial \gamma} = \frac{\nu}{\sigma \gamma^2} \sum_{i=1}^{n_1} \delta_{1i} \frac{V_i}{1 + V_i} + \frac{\nu}{\sigma \gamma^2} \sum_{j=1}^{n_2} \delta_{2j} \frac{U_j}{1 + U_j} - \frac{\nu}{\sigma \gamma^2} \sum_{i=1}^{n_1} V_i \left[1 + \frac{1}{\gamma} \ln(1 + V_i)\right] (1 + V_i)^{\frac{1}{\gamma}-1} -$$

$$\frac{\nu}{\sigma \gamma^2} \sum_{j=1}^{n_2} U_j \left[1 + \frac{1}{\gamma} \ln(1 + U_j)\right] (1 + U_j)^{\frac{1}{\gamma}-1};$$

$$\frac{\partial^2 \ell}{\partial \nu^2} = -\frac{1}{\nu^2} (m_1 + m_2) + \frac{1}{\nu^2} \left(\frac{1}{\gamma} - 1\right) \sum_{i=1}^{n_1} \delta_{1i} \frac{V_i \ln^2 V_i}{(1 + V_i)^2} + \frac{1}{\nu^2} \left(\frac{1}{\gamma} - 1\right) \sum_{j=1}^{n_2} \delta_{2j} \frac{U_j \ln^2 U_j}{(1 + U_j)^2} -$$

$$\frac{1}{\nu^2 \gamma^2} \sum_{i=1}^{n_1} V_i \ln^2 V_i \left(1 + \frac{1}{\gamma} V_i\right) (1 + V_i)^{\frac{1}{\gamma}-2} - \frac{1}{\nu^2 \gamma^2} \sum_{j=1}^{n_2} U_j \ln^2 U_j \left(1 + \frac{1}{\gamma} U_j\right) (1 + U_j)^{\frac{1}{\gamma}-2};$$

$$\frac{\partial^2 \ell}{\partial \nu \partial \gamma} = -\frac{1}{\nu \gamma^2} \sum_{i=1}^{n_1} \delta_{1i} \frac{V_i \ln V_i}{1 + V_i} - \frac{1}{\nu \gamma^2} \sum_{j=1}^{n_2} \delta_{2j} \frac{U_j \ln U_j}{1 + U_j} +$$

$$\frac{1}{\nu\gamma^2} \sum_{i=1}^{n_1} V_i \ln V_i \left[1 + \frac{1}{\gamma} \ln(1 + V_i) \right] (1 + V_i)^{\frac{1}{\gamma}-1} + \frac{1}{\nu\gamma^2} \sum_{j=1}^{n_2} U_j \ln U_j \left[1 + \frac{1}{\gamma} \ln(1 + U_j) \right] (1 + U_j)^{\frac{1}{\gamma}-1};$$

$$\begin{aligned} \frac{\partial^2 \ell}{\partial \gamma^2} &= \frac{1}{\gamma^2} (m_1 + m_2) + \frac{2}{\gamma^3} \sum_{i=1}^{n_1} \delta_{1i} \ln(1 + V_i) + \frac{2}{\gamma^3} \sum_{j=1}^{n_2} \delta_{2j} \ln(1 + U_j) \\ &\quad - \frac{2}{\gamma^3} \sum_{i=1}^{n_1} \ln(1 + V_i) (1 + V_i)^{\frac{1}{\gamma}} - \frac{1}{\gamma^4} \sum_{i=1}^{n_1} \ln^2(1 + V_i) (1 + V_i)^{\frac{1}{\gamma}} \\ &\quad - \frac{2}{\gamma^3} \sum_{j=1}^{n_2} \ln(1 + U_j) (1 + U_j)^{\frac{1}{\gamma}} - \frac{1}{\gamma^4} \sum_{j=1}^{n_2} \ln^2(1 + U_j) (1 + U_j)^{\frac{1}{\gamma}}. \end{aligned}$$

La matrice d'information de Fisher est :

$$I(\gamma_1) = -E\ddot{\ell}(\gamma_1),$$

et elle peut être remplacée par $\hat{I}(\hat{\gamma}_1) = -\ddot{\ell}(\hat{\gamma}_1)$.

La fonction de répartition $K_2(t)$ is estimée par :

$$\hat{K}_2(t) = \frac{\hat{\nu}}{\hat{\gamma}\hat{\sigma}^{\hat{\nu}}} \int_0^t \left(1 - e^{1 - \left(1 + \left(\frac{t + \hat{\nu}y - y}{\hat{\sigma}} \right)^{\hat{\nu}} \right)^{\frac{1}{\hat{\gamma}}}} \right) y^{\hat{\nu}-1} \left(1 + \left(\frac{y}{\hat{\sigma}} \right)^{\hat{\nu}} \right)^{\frac{1}{\hat{\gamma}}-1} e^{1 - \left(1 + \left(\frac{y}{\hat{\sigma}} \right)^{\hat{\nu}} \right)^{\frac{1}{\hat{\gamma}}}} dy.$$

Dans la figure (4.3) nous représentons les graphes des fonctions de répartition \hat{F}_1 , \hat{K}_2 , \hat{K}_3 et \hat{K}_4 .

Nous pouvons alors calculer l'intervalle de confiance $(\underline{K}_2(t), \overline{K}_2(t))$ pour $K_2(t)$ au seuil α

$$\begin{aligned} \underline{K}_2(t) &= \left(1 + \frac{1 - \hat{K}_2(t)}{\hat{K}_2(t)} \exp \left\{ \frac{\hat{\sigma}_{\hat{K}_2} z_{1-\alpha/2}}{\sqrt{\hat{K}_2(t)(1 - \hat{K}_2(t))}} \right\} \right)^{-1}, \\ \overline{K}_2(t) &= \left(1 + \frac{1 - \hat{K}_2(t)}{\hat{K}_2(t)} \exp \left\{ -\frac{\hat{\sigma}_{\hat{K}_2} z_{1-\alpha/2}}{\sqrt{\hat{K}_2(t)(1 - \hat{K}_2(t))}} \right\} \right)^{-1}, \end{aligned}$$

avec

$$\hat{\sigma}_{\hat{K}_2(t)}^2 = C_2^T(t, \hat{\gamma}_1) \hat{I}^{-1}(\hat{\gamma}_1) C_2(t, \hat{\gamma}_1),$$

où

$$C_2(t, \gamma_1) = (C_{21}(t, \gamma_1), C_{22}(t, \gamma_1), C_{23}(t, \gamma_1), C_{24}(t, \gamma_1))^T,$$

$$C_{21}(t, \gamma_1) = \int_0^t \frac{\partial F_1}{\partial r}(t + ry - y; \sigma, \nu, \gamma) dF_1(y; \sigma, \nu, \gamma),$$

$$C_{22}(t, \gamma_1) = \int_0^t \frac{\partial F_1}{\partial \sigma}(t + ry - y; \sigma, \nu, \gamma) dF_1(y, m, \sigma, \nu, \gamma) + F_1(t + ry - y; \sigma, \nu, \gamma) d\left(\frac{\partial F_1}{\partial \sigma}(y; \sigma, \nu, \gamma)\right),$$

$$C_{22}(t, \gamma_1) = \int_0^t \frac{\partial F_1}{\partial \nu}(t + ry - y; \sigma, \nu, \gamma) dF_1(y, m, \sigma, \nu, \gamma) + F_1(t + ry - y; \sigma, \nu, \gamma) d\left(\frac{\partial F_1}{\partial \nu}(y; \sigma, \nu, \gamma)\right),$$

$$C_{22}(t, \gamma_1) = \int_0^t \frac{\partial F_1}{\partial \gamma}(t + ry - y; \sigma, \nu, \gamma) dF_1(y, m, \sigma, \nu, \gamma) + F_1(t + ry - y; \sigma, \nu, \gamma) d\left(\frac{\partial F_1}{\partial \gamma}(y; \sigma, \nu, \gamma)\right).$$

Remarque. Dans le cas où la fonction du hasard est \cap - shape, nous aurons nécessairement $\gamma > \nu > 1$.

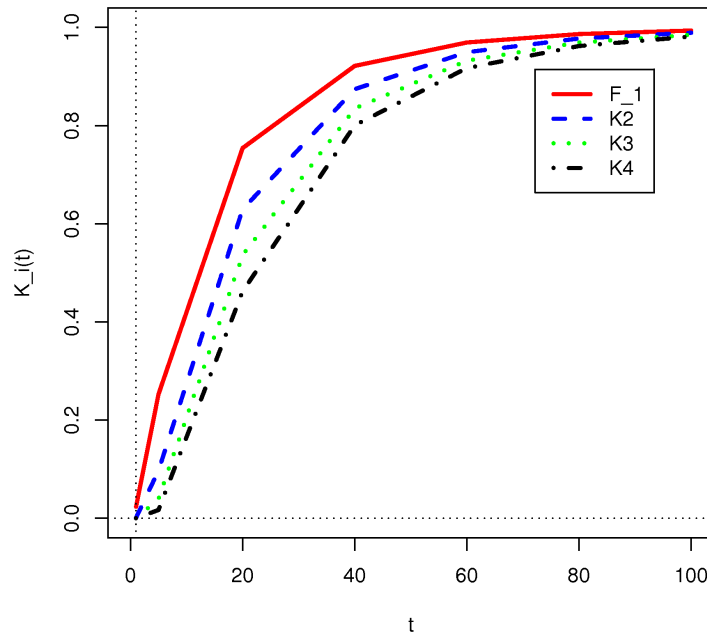


Figure 4.3 – Graphes des estimateurs de fonctions de répartition \hat{F}_1 , \hat{K}_2 , \hat{K}_3 et \hat{K}_4 , (la distribution de Weibull généralisée).

t	1	5	20	40	60	80	100
$K_2(t)$	0.0034	0.18036	0.7070	0.8941	0.9529	0.9765	0.9874
N.C. (%)	78.76	79.08	80.54	78.32	81.39	78.89	82.21

Tableau 4.4 – Niveau de confiance pour $K_2(t)$,
(la distribution de Weibull généralisée).

9.1 Simulation

Considérons le cas des données complètes de taille $n_1 = n_2 = 100$, chaque échantillon est répété 5000 fois. Nous trouvons par simulation les intervalles de confiance donnée par la formule (4.7) et (4.8) au seuil $\alpha = 0.2$. Nous simulons les temps de pannes T_{1i} et T_{2j} à partir de la loi de Weibull généralisée avec les paramètres :

$$T_{1i} \sim PGW(\sigma_1, \nu_1, \gamma_1), \quad T_{2j} \sim PGW(\sigma_2, \nu_2, \gamma_2)$$

$$\sigma_1 = 4, \quad \sigma_2 = 4/5, \quad \nu_1 = \nu_2 = 2, \quad \gamma_1 = \gamma_2 = 4$$

Pour différentes valeurs de t , les proportions des intervalles de confiance (N.C.) couvrant la vraie valeur de la fonction de répartition $K_2(t)$ (4.6) sont données dans le tableau (4.4).

Conclusion générale et perspectives

Les tests d'ajustement du type Chi-deux modifié sont appliqués aux modèles paramétriques de fonction de hasard unimodale couramment utilisés en fiabilité et en analyse de survie. L'exemple traité de données d'entretien d'ERCA montre que la distribution Gaussienne inverse présente un meilleur ajustement par rapport aux autres distributions citées dans cette thèse. La puissance des tests est aussi étudiée en remarquant que la statistique de McCulloch $M_n^2(\hat{\theta})$ possède une grande puissance. Le test d'ajustement du type Chi-deux pour la famille des distributions Gaussiennes inverses avec censure est aussi étudié.

Le modèle AFT basé sur la famille des distributions Gaussiennes inverses avec censure est bien étudié. Les systèmes redondants basés sur les familles paramétriques de fonction de hasard unimodale sont aussi étudiés.

Nous espérons réaliser ces tâches dans le futur

- • Réaliser le test de type du Chi-deux de la loi Gaussienne inverse contre la loi log-logistique et la loi de Weibull exponentielle basé sur la statistique de Rao-Robson-Nikulin.
- • Réaliser les simulations pour les systèmes redondants basés sur les familles des lois de fonction de hasard unimodale dans le cas des données censurées.
- • Réaliser la simulation des modèles AFT-IG, AFT-LN, AFT-LL, AFT-BS, AFT-EW et AFT-WG.

- • Faire des logiciels permettant d'appliquer les modèles AFT semi-paraétriques et avec des cross-effects des fonctions de survie aux études des données biomédicales. (En collaboration avec R.Tahir).

- • Réaliser une étude comparative des tests de NRR, DN, McCulloch, Kolmogorov, Cramér-von Mises et Anderson-Darling pour la distribution Gaussienne inverse.

- Réaliser une étude comparative des tests de NRR, DN, McCulloch, Kolmogorov, Cramér-von Mises et Anderson-Darling pour la distribution de Birnbaum-Saunders. (En collaboration avec R.Tahir).

- Réaliser une étude comparative des tests de NRR, DN, McCulloch, Kolmogorov, Cramér-von Mises et Anderson-Darling pour la distribution de Kumaraswamy-Log-Logistique.

- • Etude comparative des distributions LN, LL, EW, GW, BS, IG, et Kumaraswamy-Log-Logistique.

- • Estimer la fonction de distribution \hat{K}_m , $m \geq 3$ pour les systèmes redondants et faire avec censure les simulations pour celle-ci.

Bibliographie

1. Aguirre, N., Nikulin, M. (1994) Chi-Squared Goodness-of-fit Test for the Family of Logistic Distribution, *Kybernetika*, 30 : 214-222.
2. Akritas, Michael G. (1988) Pearson-Type Goodness-of-Fit Tests : The Univariate Case. *Journal of the American Statistical Association*, 83 (401) :222-230.
3. Bagdonavičius, V., Cheminade, O., Nikulin, M. (2004) Statistical planning and inference in accelerated life testing using the CHSS model, *Journal of the Statistical Planning and Inference*, 2 : 535-551.
4. Bagdonavičius, V., Clerjaud, L., Nikulin, M. (2008) Accelerated Life Testing when the Hazard Rate Function has Cup Shape, In : *Mathematical methods in Survival analysis, reliability and quality of life* (Huber C., Limnios N., Balakrishnan N., Mesbah M., Nikulin M. (Eds)), ISTE & Wiley : London, 203-215.
5. Bagdonavicius, V., Levulienė, R., and Nikulin, M. (2012). Chi-squared goodness-of-fit tests for parametric accelerated failure time models. *Communications in Statistics - Theory and Methods*.
6. Bagdonavičius, V., Nikulin, M. (2011) Chi-squared Goodness-of-fit Test for Right Censored Data. *International Journal of Applied Mathematics and Statistics*, vol. 24, p.30-50
7. Bagdonavičius, V., Kruopis, J., Nikulin, M.(2011a) *Non-parametric Tests for Censored Data*, ISTE & J. Wiley.
8. Bagdonavicius, V., Kruopis, J., and Nikulin, M. (2011b). *Nonparametric tests for complete Data*. ISTE & J. Wiley.

9. Bagdonavicius, V., Levulienė, R., and Nikulin, M. (2012). Chi-squared goodness-of-fit tests for parametric accelerated failure time models. *Communications in Statistics - Theory and Methods*.
10. Bagdonavičius, V., Masiulaityte, I., Nikulin, M. (2008a) Statistical analysis of redundant system with "warm" stand-by units, *Stochastics : An International Journal of Probability and Stochastic Processes*, 80 #2-3, 115–128.
11. Bagdonavičius, V., Masiulaityte, I., Nikulin, M. (2008b) Statistical analysis of redundant system with one stand-by unit. In : *Mathematical methods in Survival analysis, reliability and quality of life* (Eds. Huber C., Limnios N., Balakrishnan N., Messbah M., Nikulin M.), ISTE & Wiley : London, 189–202.
12. Bagdonavičius, V., Masiulaityte, I., Nikulin, M. (2009) Asymptotic Properties of Redundant Systems Reliability Estimators. In : *Advances in Degradation Modeling. Applications to Reliability, Survival Analysis, and Finance* (Eds. Nikulin, M., Limnios N., Balakrishnan N., Kahle W., Huber C.), Birkhäuser Boston, 293–310.
13. Bagdonavičius, V., Masiulaityte, I., Nikulin, M. (2010) Parametric estimation of redundant system reliability from censored data. In : *Mathematical and Statistical Methods in Reliability* (Eds. Balakrishnan, N., Nikulin, M., Rykov, V.), Birkhäuser Boston, p 177–191.
14. Bagdonavičius V., Nikulin, M. (2002) *Accelerated Life Models : Modeling and Statistical Analysis*, Chapman and Hall.
15. Bagdonavičius, V., Nikulin, M. (2011) Chi-squared tests for general composite hypotheses from censored samples *Comptes Rendus de Academie des Sciences de Paris, Mathématiques*, V. 349, N°. 3-4, 219-223.
16. Berg, P.T. (1994) Deductibles and the Inverse Gaussian Distributions, *Astin Bulletin*, 24, 319-23.
17. Birnbaum, Z. W. and Saunders, S. C. (1969a). A new family of life distribution, *Journal of Applied Probability*, vol. 6, 319-327.
18. Birnbaum, Z. W. and Saunders, S. C. (1969b). Estimation for a family of life distributions with applications to fatigue", *Journal of Applied Probability*, vol. 6, 328-347.

-
19. Bolshev, L. N. (1987) Probability theory and mathematical statistics, in : Selected Works (edited by Yu. V. Prokhorov) [in Russian], Nauka, Moscow.
 20. Bolshev, L.N., Smirnov, N.V. (1983) Tables of Mathematical Statistics. Science, Moscow,(in Russian).
 21. Chen, J. (1975) Goodness of Fit Tests Under Random Censorship, PhD dissertation, Oregon State University.
 22. Chernoff, H., Lehmann, E. L. (1954) The use of maximum likelihood estimates in tests for goodness of fit, *Ann. Math. Statist.*, 25,578-586.
 23. Chhikara, R.S., Folks, J.L., (1977). The Inverse Gaussian Distribution as a Life Time Model. *Technometrics* , 19 , 461-468.
 24. Chhikara, R.S., Folks, J.L. (1989) The Inverse Gaussian Distribution, Marcel Dekker, New York.
 25. Cramer, U. (1946). *Mathematical Methods of statistics*, Princeton University Press, New York.
 26. Drost, F. (1988) Asymptotics for Generalized Chi-squared Goodness-of-fit Tests, amsterdam : Centre for Mathematics and Computer Sciences, CWI Tracs, 48.
 27. Dzhaparidze, K. O. , Nikulin, M. S. (1974) On a modification of the standard statistic of Pearson, *Theory of Probability and Its Applications*, 19, 851-862.
 28. Dzhaparidze, K. O. and Nikulin, M. S. (1992). On calculating chi-squared type statistics, *Problems of the theory of probability distributions*, 12, 1-17, Nauka, St. Petersburg, Russia.
 29. Edgeman, R.L., Scott, R.C., Pavur, R.J. (1988) A modified Kolmogorov-Smimov test for the inverse Gaussian density with unknown parameters, *Comm. Statist. Simulation*, 17, 1203-1212.
 30. Efron, B. (1988) Logistic Regression, Survival Analysis, and the Kaplan-Meier Curve. *Journal of the American Statistical Association* 83, 414-425.
 31. Fisher, R. (1928) On a property connecting the chi-square measure of discrepancy with the method of maximum likelihood. *Atti de Congresso Internazionale di Matematici*, Bologna, 6,

- 95-100.
32. Greenwood, P. S., Nikulin, M. (1996) A guide to Chi-squared Testing, John Wiley and Sons, New York.
 33. Gunes, H., Dietz, D.C., Auclai, P.F., Moor, A.H. (1997) Modified goodness-of-fit tests for the inverse Gaussian distribution Computational statistics and Data analysis, 24, p.63-77.
 34. Habib, M. G. H. (1981) A Chi-Square Goodness-of Fit Test for Censored Data, PhD dissertation, Oregon State University.
 35. Habib, M.G., Thomas, D.R. (1986) Chi-squared Goodness-of-Fit Tests For Randomly Censored Data, Annals of Statistics, V.14, N°2, p 759-765.
 36. Hadwiger, H. (1940) 'Eine analytische reproductions-funktion fur biologische Gesamtheiten.', Skandinavisk Aktuarietidskrift, 23, 101-113.
 37. Haghighi, F. (2004) Fiabilité et Analyse de Survie : Le Modèle semi-paramétrique de Dégrada-tion. Une généralisation de la famille Weibull et ses application, Thèse de Doctorat, Université Victor Segalen Bordeaux 2.
 38. Hazen, A. (1914) Storage to be Provided in Impounding Reservoirs for Municipal Water Supply, transactions of the American Society of Civil Engineering, Pap. 1308, 77, 1547-50.
 39. Henze, Norbert, Klar, Bernhard (2002) Goodness-of-fit tests for the inverse gaussian distribu-tion based on the empirical Laplace transform, journal of Annals of the Institute of Statistical Mathematics,, vol. 54, no2, pp. 425-444.
 40. Hjort, N.L. (1990) Goodness of Fit Test in Models for Life History Data Based on Cumulative Hazard Rates, The annals of statistics, V.18, N°3, p1221-1258.
 41. Holton, R.E. (1913) Frequency of Recurrence of Hudson River Floods, U.S. Weither Bureau Bulletin, Z, 109-12.
 42. Hsuan, T.A, Robson, D.S. (1976) The χ^2 - Goodness-of-fit Tests with Moment Type Estimator, Communications in Statistics - Theory and Methods, 16 , 1509-1519.

-
43. Huber, C., Limnios, N., Mesbah M., Nikulin, M. (Eds) (2008) *Mathematical Methods in Survival analysis, Reliability and Quality of Life*, Wily-ISTE.
 44. Ionescu, D.C, Lmmnios, N. (Eds) (1999) *Statistical and Probabilistic Models in Reliability*, Birkhauser Boston.
 45. Kahle, W., Lehmann, A. (2010) *The Wiener Process as a Degradation Model : Modeling and Parameter Estimation*, In : *Advances in Degradation Modeling : Applications to Reliability, Survival Analysis and Finance* (Eds : Nikulin M.S., Limnios N., Balakrishnan N., Kahle W., Huber C.), Birkhäuser Boston , 416p.
 46. Kim, J.H. (1993) *Chi-squared Goodness-of-Fit Tests For Randomly Censored Data*, *Annals of Statistics*, the annals of statistics, V.21, N°3, p 1621-1639.
 47. Lancaster, H.O. (1969) *The Chi-squared Distribution*, John Wiley, New York.
 48. Lawless, J.F. (2003) *Statistical Models and Methods for Lifetime Data*, 2nd ed., New York : John Wiley.
 49. LeCam, L. (1956) *Locally asymptotically normal families of distributions*, In *University of California, Publication in Statistics* , vol.3, p. 37-98.
 50. LeCam, L., Mahan,C., Singh, A. (1983) *An extension of a Theorem of H.Chernoff and E.L.Lehmann*. In : *Recent advances in statistics*, Academic Press, Orlando, 303-332.
 51. Lemeshko, B. Yu., Lemeshko, S.B., Nikulin, M..S., Saaidia N. (2010a) *Modeling Statistic Distributions for Nonparametric Goodness-of-Fit Criteria for Testing Complex Hypotheses with Respect to the Inverse Gaussian Law*, *Automation and Remote Control*, Vol. 71, No. 7, pp. 1358-1373.
 52. Lemeshko, B.Y., Lemeshko, S.B., Akushkina, K.A., Nikulin, M., Saaidia, N., (2010b) *Inverse Gaussian Model and Its Applications in Reliability and Survival Analysis*, In : *Mathematical and Statistical Models and Methods in Reliability* (Eds. Rykov, V.V., Balakrishnan, N., Nikulin, M.S.), Birkhäuser Boston, p 293-315.

53. Lemeshko, B.Yu., Lemeshko, S.B.(2009) Distribution models for nonparametric tests for fit in verifying complicated hypotheses and maximum-likelihood estimators. Part 1. Measurement Techniques, V. 52, N 6, 555-565.
54. Lemeshko, S.B., Nikulin, M.S., Saaidia, N. (2009) Simulation of the statistics distributions and power of the goodness-of-fit tests in composite hypotheses testing rather inverse Gaussian distribution, Proceedings of the 6th St. Petersburg Workshop on Simulation, St. Petersburg, June 28–July , VVM comm. Ltd., St. Petersburg, V. 1, 323-328.
55. McCulloch, C. E. (1985) Relationships among some chi-squared goodness of fit statistics. Communications in Statistics-Theory and Methods, 14 ,593-603.
56. Meeker, W.Q., Escobar, L.A. (1998) Statistical Methods for reliability Data, John Wiley and Sons, INC.
57. Mirvaliev, M. (2001) An investigation of generalized chi-squared type statistics, Ph.D. Thesis, Academy of Science, Republic of Uzbekistan, Tashkent.
58. Moore, D. S.n Spruill, M. C. (1975) Unified large-sample theory of general chi-squared statistics for tests of fit, Annals of Statistics, 3, 599-616.
59. Mudholkar, G. S., Natarajan, R., (2001) Chaubey, Y. P. A Goodness-of-Fit Test for The Inverse Gaussian Distribution using its Independence Characterization, Sankhya : The Indian Journal of Statistics, Volume 63, Series B, pt. 3, 362-374.
60. Mudholkar, G., Srivastava D., Freimer M. (1995) The exponentiated Weibull family : a reanalysis of the bus-motor-failure data, Technometrics, vol. 37, p. 436-445.
61. Mudholkar, G.S., Srivastava, D. K. and Kollia, G. D. (1996). A generalization of the Weibull distribution with application to the analysis of survival data. Journal of the American Statistical Association, 91, 1575-1583.
62. Nadarajah, S., (2009) An alternative inverse Gaussian distribution, Mathematics and Computers in Simulation, V.79 , Issue 5, p : 1721-1729.

-
63. Nelson, W. (1990) *Accelerated Testing : Statistical Models, Test Plans, and Data Analysis*, John Wiley and Sons : New York.
 64. Nikulin, M.S., Haghghi, F. (2004) A Chi-Squared Test for the Generalized Power Weibull Family for the Head-and-Neck Cancer Censored Data, *Journal of Mathematical Sciences* Volume 133, Number 3, 1333-1341.
 65. Nikulin, M. Haghghi, F. (2009) On the power generalized Weibull family : model for cancer censored data, *METRON-International Journal of Statistics*, vol. LXVII, n.1, pp.75-86.
 66. Nikulin, M.S., Limnios, N., Balakrishnan, N., Kahle, W., Huber, C. (Eds.) (2009) *Advances in Degradation Modeling : Applications to Reliability, Survival Analysis and Finance*, Birkhäuser Boston, 454p.
 67. Nikulin, M.S. (1973a) Chi-square Test For Continuous Distributions with Shift and Scale Parameters, *teor. Veroyatn. Primen.*, 18, No. 3, 559-568.
 68. Nikulin, M.S. (1973b) Chi-square test for normality. In *International Vilnius Conference on Probability Theory and Mathematical Statistics*, V.2, 119–122.
 69. M. Nikulin, N. Balakrishnan , R. Tahir, N. Saaidia. "Modified Chi-Squared Goodness-of-Fit Test for Birnbaum-Saunders Distribution". *The International Workshop on Applied Methods of Statistical analysis. Simulations and Statistical Inference (AMSA'2011)*, 20-22 September 2011, Novosibirsk, Russia.
 70. Nikulin, M.S., Saaidia, N. (2009) Inverse Gaussian family and its applications in reliability, *Study by simulation*, St. Petersburg, June 28–July 4. VVM comm. Ltd., St. Petersburg **V. 2**, 657-661.
 71. Nikulin, M., Saaidia, N., Tahir R. (2011a) Reliability analysis of redundant systems by simulation for data with unimodal hazard rate functions. *Journal of Mathematics in Engineering, Science and Aerospace*, Vol. 2, No. 3, pp. 277-286.
 72. Nikulin M., Saaidia N., Tahir, R. (2011b) An Application of the AFT Model based on the Log-normal Distribution for Reliability Analysis of Redundant Systems. *9th International Pulridisciplinary Congress on Quality and Dependability, Qualita 2011*, 22-25 March 2011, Angers, France.

73. Nikulin, M.S., Solev, V.N. (1999) Chi-squares Goodness-of-fit Test for Doubly Censored Data with Applications in Survival Analysis and Reliability, In : Statistical and Probabilistic Models in Reliability, D.C.Ionescu, N.Limnios (eds), Birkhauser, Boston, 101-112.
74. Nikulin, M.S., Gerville-Réache, L., Couallier, V. (2007) Statistiques des Essais Accélérés, Hermes : London.
75. Onara, A., Padgett, W.J. (2000) Accelerated test models with the inverse Gaussian distribution, Journal of Statistical Planning and Inference 89, 119-133.
76. O'Reilly, Federico, Rueda, Raúl (1992) Goodness of fit for the inverse Gaussian distribution, Canadian Journal of Statistics, V., Issue 4, 387-397.
77. Rao, K.C., Robson, D.S. (1974) A chi-square statistic for goodness-of-fit tests within the exponential family, Communications in Statistics, 3, 1139-1153.
78. Rykov, V.V., Balakrishnan, N., Nikulin, M.S. (Eds.) (2010) Mathematical and Statistical Models and Methods in Reliability, Applications to Medicine, Finance, and Quality Control Series : Statistics for Industry and Technology, Birkhäuser Boston, 458 p.
79. Saaidia, N. Seddik-Ameur, N. (2010) Chi-Squared Type Test for the Inverse Gaussian Distribution, les annales de l'ISUP, Volume LIV - Fascicule 3, pp. 67-84.
80. Saaidia, N. (2009) Sur les applications de la famille des lois Gaussiennes inverses en fiabilité, 41èmes Journées de Statistique, Bordeaux, 19-22 mai 2009.
81. Saaidia N., Nikulin M., Ramzan Tahir (2011) An Application of the AFT Model Based on the Generalized Weibull Distribution for Reliability Analysis of Redundant Systems. 43ème Journées de Statistique, 23-27 May , Tunis, Tunisia.
82. Saaidia, N., Tahir, R., Seddik-Ameur, N. (2010) The Inverse Gaussian Distribution and its Applications in Accelerated Life Testing when the Hazard Rate Function has \cap -shape, International Conference on Accelerated Life Testing, Reliability-based Analysis and Design 19-21 Mai 2010, Clermont-Ferrand.

-
83. Saaidia N., Ramzan Tahir. "A Modified Chi-squared Goodness-of-Fit Test for the Inverse Gaussian Distributions and its Applications in Reliability". International conference in statistics organized (S2MRSA) in honor of Prof Mikhail Nikulin, Bordeaux, July 4-6th 2012, France.
 84. Santana, T.V.F., Ortega, E.M.M., Cordeiro, G.M., Silva, G.O. (2012) The Kumaraswamy-Log-Logistic Distribution, *Journal of Statistical Theory and Applications* , V. 11, N° 3, 2012, pp. 265-291.
 85. Saunders, S.C., Lawless , J.F. (2007) *Reliability, life testing and the prediction of service lives : for engineers and Scientists*, Springer New York, 310p.
 86. Seal, H.L. (1969) *Stochastic Theory of a Risk Business*, New York :John Wiley & Sons Inc
 87. Seal, H.L. (1978) From Aggregate Claim Distributions to Probability of Ruin, *Astin. Bulletin*, 10,47-53.
 88. Seshadri, V., (1993) *The Inverse Gaussian Distribution : A Case Study in Exponential Families*, Clarendon Press :Oxford,
 89. Seshadri, V., (1999) *The inverse Gaussian Distribution : statistical theory and applications*, Springer,New York.
 90. Shunsuke, Sato, Junko, Inoue, (2007). Inverse gaussian distribution and its application. *Electronics and Communications in Japan (Part III : Fundamental Electronic Science)* Volume 77, Issue 1, p : 32-42.
 91. Silva, R.S., Lopes, H.F., Migonn, H.S., (2006). The extended generalized inverse Gaussian distribution for log-linear and stochastic volatility models. *Brazilian Journal of Probability and Statistics*. V.20, p. 67-91.
 92. Tahir, M.R. (2012) *On Validation Of Parametric Models Applied In Survival Analysis And Reliability*, Thèse de Doctorat, Université de Bordeaux, France.
 93. Ramzan Tahir, Saaidia N. "A Modified Chi-squared Goodness-of-Fit Test for the Birnbaum-Saunders Distributions and its Applications in Reliability". 44ème Journées de Statistique, 21-25 May 2012, Brussels, Belgium

94. Takagi, K., Kumagai, S., Matsunagat, I, Kusakal, Y.(1997) Application of Inverse Gaussian Distribution to Occupational Exposure Data , Ann. occur. Hyg., Vol. 41, No 5, pp 505-514.
95. Van Der Vaart, A.W., (1998) Asymptotic Statistics. Cambridge Series in Statistics and probabilistic Mathematics, Cambridge : Cambridge University Press.
96. Voinov, V., Alloyarova, R., Pya, N. (2008) Recent Achievemets in Modified Chi-squared Goodness-of-fit Testing, In :Statistical Models and Methods for Biomedical and Technical Systems,(Eds. F. Vonta, M. Nikulin, N. Limnios, C. Huber.), Birkhäuser, Boston, p 241-258.
97. Voinov, V., Pya, N., Alloyarova, R. (2009) A Comparative Study of Some Modified Chi-squared Tests, Communi. in Stat.-Simul. and Comput., 38 :2, 355-367.
98. Voinov, V., Nikulin M. (1989) Unbiased estimators and their applications. Moscow, Nauka, 440 p, (in Russian).
99. Voinov, V., Nikulin, M. (1993) Unbiased Estimators and Their Applications, V.1 Univariate case, Dordrecht : Kluwer Academic Publishers.
100. Voinov, V., Alloyarova, R., Pya, N. (2008) A Modified Chi-squared Goodness-of-fit Test for the Three parameter Weibull Distribution and its Applications in Reliability, In : Mathematical Methods in Survival Analysis, Reliability and Quality of Life (Eds. Huber C., Limnios N., Mesbah M., Nikulin M.), Wily-ISTE.
101. Voinov, V., Nikulin, M., Balakrishnan, N. (2013). Chi-Squared goodness-of-fit tests with applications. Academic Press. Amsterdam.
102. Von Alven, W. H. (ed.) (1964) Reliability Engineering by ARINC, Englewood Cliffs, N.J.Prentice-Hall, Inc.
103. Xiao, Wang, Dihua, Xu. (2010) An Inverse Gaussian Process Model for Degradation Data, Technometrics, 2010, V. 52, No 2, p 188-197.
104. Zhang, B. (1999) A chi-squared goodness-of-fit test for logistic regression models based on case-control data, Biometrika, 86, 531-539.

Liste de Publications et de Communications

Publications

Journal Articles

- B. Yu. Lemesenko, S. B. Lemesenko, M. S. Nikulin, and N. Saaidia. "Modeling Statistic Distributions for Nonparametric Goodness-of-Fit Criteria for Testing Complex Hypotheses with Respect to the Inverse Gaussian Law ". Automation and Remote Control, 2010, Vol. 71, No. 7, pp. 1358-1373.
- N. Saaidia et N. Seddik-Ameur. " Chi-Squared Type Test for the Inverse Gaussian Distribution", les Annales de l'ISUP, Volume LIV - Fascicule 3, 2010, pp. 67-84.
- M. Nikulin, N. Saaidia, R. Tahir. " Reliability analysis of redundant systems by simulation for data with unimodal hazard rate functions". Journal of Mathematics in Engineering, Science and Aerospace, Vol. 2, No. 3, 2011, pp. 277-286.

Chapters in Books

- Boris Yu. Lemesenko, Stanislav B. Lemesenko, Kseniya A. Akushkina, Mikhail S. Nikulin, and Nouredine Saaidia, " The Inverse Gaussian Model and Its Applications in Reliability and Survival Analysis". In : Mathematical and Statistical Models and Methods in Reliability (Eds. Rykov, V.V. ; Balakrishnan, N. ; Nikulin, M.S.) Birkhäuser Boston, p 458. Novembre 2010.
- Mikhail Nikulin, Nouredine Saaidia, Ramzan Tahir , "Recent Results in the Analysis of Redundant Systems", In : Recent Advances in System Reliability Signatures, Multi-state Systems and Statistical Inference (Eds. Anatoly Lisnianski, Ilia Frenkel) Springer, p 334, September 2011.

submitted papers

- Saaidia N. , " A Goodness of Fit Tests for the Inverse Gaussian Distributions and Its Applications in Reliability".
 - Saaidia N., Tahir R., Tran Q.X. " A Goodness of Fit Tests for the Birnbaum-Saunders Distributions and Its Applications in Reliability".
 - Saaidia N. Tran Q.X. " A Chi-squared Goodness of Fit Tests for the Kumaraswamy-Log- Logistic Distribution".
-

Conference and Seminar Presentations

- N. Saaidia. "On the Application of the Inverse Gaussian Family in Reliability". 41ème Journées de Statistique, 25-29 May 2009, Bordeaux, France.
- Lemeshko S.B., Nikulin M.S., Saaidia N. "Simulation of the statistics distributions and power of the goodness-of-fit tests in composite hypotheses testing rather Inverse Gaussian distribution". 6th Workshop on Simulation June 28-July 4, 2009, St Petersburg, Russia.
- Nikulin M., Saaidia N. "Inverse Gaussian Family and its Applications in reliability. Study by simulation". 6th Workshop on Simulation June 28-July 4, 2009, St Petersburg, Russia.
- Saaidia, R. Tahir, N. Seddik-Ameur. "The Inverse Gaussian Distribution and its Applications in Accelerated Life Testing when the Hazard Rate Function has \cap -shape". 3rd international conference on Accelerated Life Testing, Reliability-based Analysis and Design (ALT'2010), 19-21 May 2010, Clermont Ferrant, France.
- Ramzan Tahir, N. Saaidia. "Bivariate Models in Survival Analysis and Reliability". 3rd international conference on Accelerated Life Testing, Reliability-based Analysis and Design (ALT'2010). 19-21 May 2010, Clermont Ferrant, France.
- Nikulin M., Saaidia N., Ramzan Tahir. "An Application of the AFT Model based on the Log-normal Distribution for Reliability Analysis of Redundant Systems". 9th International Pulridisciplinary Congress on Quality and Dependability, Qualita 2011, 22-25 March 2011, Angers, France.
- Saaidia N., Nikulin M., Ramzan Tahir. "An Application of the AFT Model Based on the Gene-

- ralized Weibull Distribution for Reliability Analysis of Redundant Systems". 43ème Journées de Statistique, 23-27 May 2011, Tunis, Tunisia.
- M. Nikulin, N. Balakrishnan , R. Tahir, N. Saaidia. "Modified Chi-Squared Goodness-of-Fit Test for Birnbaum-Saunders Distribution". The International Workshop on Applied Methods of Statistical analysis. Simulations and Statistical Inference (AMSA'2011), 20-22 September 2011, Novosibirsk, Russia.
 - N. Saaidia, "A Chi-Squared Test for the Family of Inverse Gaussian Distributions for Censored Data". The International Workshop on Applied Methods of Statistical analysis. Simulations and Statistical Inference (AMSA'2011), 20-22 September 2011, Novosibirsk, Russia.
 - Ramzan Tahir, Saaidia N. "A Modified Chi-squared Goodness-of-Fit Test for the Birnbaum-Saunders Distributions and its Applications in Reliability". 44ème Journées de Statistique, 21-25 May 2012, Brussels, Belgium.
 - Saaidia N., Ramzan Tahir. "A Modified Chi-squared Goodness-of-Fit Test for the Inverse Gaussian Distributions and its Applications in Reliability". International conference in statistics organized (S2MRSA) in honor of Prof Mikhail Nikulin, Bordeaux, July 4-6th 2012, France.
 - Saaidia N., Ramzan Tahir, Tran X.Q.. "A Modified Chi-squared Goodness-of-Fit Test for the Kumaraswamy-Log-Logistic Distributions and its Applications in Reliability". 45ème Journées de Statistique, 21-25 May 2013, Toulouse, France. (Accepted).
 - Saaidia N., Ramzan Tahir, Tran X.Q.. "A Chi-squared Goodness-of-Fit Test for the Kumaraswamy-Log-Logistic Distributions with censored data". 7th International Workshop on Simulation, May 21-25, 2013, Italy. (submitted).

Résumé : En fiabilité et en analyse de survie, les distributions qui ont une fonction de hasard unimodale ne sont pas nombreuses, qu'on peut citer : Gaussienne inverse , log-normale, log-logistique, de Birnbaum-Saunders, de Weibull exponentielle et de Weibull généralisée. Dans cette thèse, nous développons les tests modifiés du Chi-deux pour ces distributions tout en comparant la distribution Gaussienne inverse avec les autres. Ensuite nous construisons le modèle AFT basé sur la distribution Gaussienne inverse et les systèmes redondants basés sur les distributions de fonction de hasard unimodale.

Mots clés : Distribution Gaussienne inverse, Distribution log-normale, Distribution log-logistique, Distribution de Birnbaum-Saunders, Distribution de Weibull exponentielle, Distribution de Weibull généralisée, Test du Chi-deux, Test modifié du Chi-deux, Fiabilité, Analyse de survie, Modèle de Sedyakin, Modèle AFT, Systèmes redondants, Intervalle de confiance, Temps de panne, Maximum de vraisemblance, Estimation, Classes de Neyman-Pearson.

Abstract : In reliability and survival analysis, distributions that have a unimodal or \cap -shape hazard rate function are not too many, they include : the inverse Gaussian, log-normal, log-logistic, Birnbaum-Saunders, exponential Weibull and power generalized Weibull distributions. In this thesis, we develop the modified Chi-squared tests for these distributions, and we give a comparative study between the inverse Gaussian distribution and the other distributions, then we realize simulations. We also construct the AFT model based on the inverse Gaussian distribution and redundant systems based on distributions having a unimodal hazard rate function.

Key-words : Inverse Gaussian distribution, log-normal distribution, log-logistic distribution, Birnbaum-Saunders Distribution , exponential Weibull distribution, power generalized Weibull distribution, Chi-square test, modified Chi-squared, Reliability, Survival analysis, Sedyakin's model, AFT model, Redundant system, Confidence interval, Failure time, Maximum likelihood , Estimation, Neyman-Pearson classes.