

وزارة التعليم العالي والبحث العلمي

BADJI MOKHTAR -ANNABA
UNIVERSITY
UNIVERSITE BADJI MOKHTAR
ANNABA



جامعة باجي مختار
- عنابة -

Faculté des Sciences

Département de Mathématiques

THESE

Présentée
Pour l'obtention du grade de
Doctorat en Statistiques

Par
M. Nouredine SAAIDIA

Titre:

**La Famille des lois Gaussiennes Inverses et
ses Applications en Fiabilité et Analyse de
Survie.**

Soutenue le : 16/06/2011

Devant le jury composé de:

M. Boutabia H.

Mme Seddik-Ameur N.

M. Nikulin M.

Mme Huber C.

Mme Chadli A.

Mme Sadki O.

Prof.

Prof.

Prof.

Prof.

MC.

MC.

Univ. Badji Mokhtar-Annaba.

Univ. Badji Mokhtar-Annaba.

Univ. de Bordeaux 2, France.

Univ. René Descartes, Paris 5, France.

Univ. Badji Mokhtar-Annaba.

Univ. USTHB. Alger.

Président

Directrice de thèse

Co-directeur de thèse

Examinatrice

Examinatrice

Examinatrice

ملخص: في تحليل الموثوقية والبقاء، التوزيعات ذات دالة نسبة الخطر على شكل \cap ليست كثيرة جدا، وهي تشمل : توزيع معكوس جاوس ، توزيع الوغاريتم الطبيعي، توزيع الوغاريتم اللوجستيكي و توزيع ويبل المعمم. في هذه الرسالة نقوم بتطوير اختبار **Rao-Robson-Nikulin** لتوزيع معكوس جاوس ، ونعطي دراسة مقارنة بين توزيع معكوس جاوس والتوزيعات المذكورة أعلاه. نقوم بعدها بتنفيذ المحاكاة. النماذج ذات المتغيرات التفسيرية على أساس توزيع الغاوسي معكوس جاوس أيضا أعطيت في هذه الدراسة.

المفاتيح: توزيع معكوس جاوس ، اختبار χ^2 ، اختبار Rao-Robson-Nikulin ، توزيع الوغاريتم الطبيعي، توزيع الوغاريتم اللوجستيكي و توزيع ويبل المعمم، الموثوقية، تحليل البقاء، نموذج Sedyakin، نموذج AFT ، النظام المكرر ، مجالات الثقة، وقت العطل، المعقولة القصوى ، تقدير، فئات Neyman-Pearson.

Résumé: En fiabilité, et analyse de survie, les distributions qui ont une fonction de hasard unimodale ne sont pas trop nombreuses, qu'on peut citer: la distribution Gaussienne inverse, log-normale, log-logistique et la distribution de Weibull généralisée. Dans cette thèse, nous développons le test de Rao-Robson-Nikulin pour la distribution Gaussienne inverse, et nous donnons une étude comparative entre la distribution Gaussienne inverse et les distributions citées ci-dessus, puis nous réalisons des simulations. Ensuite nous construisons les modèles avec covariables basés sur la distribution Gaussienne inverse.

Mots-clés : Distribution Gaussienne inverse, Test de Chi-deux, Test de Rao-Robson-Nikulin, Distribution log-normale, Distribution log-logistique, Distribution de Weibull généralisée, 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 hazard function are not too many, they include: the inverse Gaussian distribution, log-normal, log-logistic and power generalized Weibull distribution. In this thesis, we develop the test of Rao-Robson-Nikulin for the inverse Gaussian distribution, and we give a comparative study between the inverse Gaussian distribution and the distributions listed above, then we realize simulations. We also construct models with covariates based on the inverse Gaussian distribution.

Key-words: Inverse Gaussian distribution, Chi-square test, Test of Rao-Robson-Nikulin, Log-normal distribution, Log-logistic distribution, Power generalized Weibull distribution, Reliability, Survival analysis, Sedyakin's model, AFT model, Redundant system, Confidence interval, Failure time, Maximum likelihood , Estimation, Neyman-Pearson classes.

DEDICACE

A mes très chers parents.

A ma femme et mes enfants: Oualaa Mayssoune, Med Zakaria et le petit poussin Raïd.

A toute la famille.

A Mikhail Nikulin.

Remerciements

Je commence d'abord par remercier le ministère de l'enseignement supérieur et de la recherche scientifique. Ce travail a été réalisé dans le cadre d'une cotutelle entre l'université Badji Mokhtar- Annaba, Algérie sous la direction de professeur N. Seddik-Ameur et l'université Victor Segalen, Bordeaux, sous la direction de professeur M.S. Nikulin.

J'adresse mes plus vifs remerciements à mon directeur de thèse Madame N. Seddik-Ameur, professeur au département de mathématiques à l'université Badji Mokhtar-Annaba, pour sa confiance, sa patience sa générosité, ses conseils constructifs, ses suggestions importantes et les corrections nécessaires. J'apprécie la pleine confiance et la liberté scientifique qu'elle m'a accordées pour la réalisation de cette thèse. Elle l'a dirigée avec une grande efficacité. Qu'elle trouve ici mon admiration et mes profonds respects.

J'exprime toute ma gratitude et mes vifs remerciements à mon co-directeur de thèse Monsieur M. S. Nikulin professeur à l'université de Bordeaux¹. Je voudrais notamment le remercier pour son accueil, le temps et la disponibilité dont il a fait preuve pendant ces années, ainsi que d'avoir cru en mes capacités et en me donnant une véritable liberté d'action et ce dans d'excellentes conditions. Il a dirigé mon travail avec grande compétence, pleine patience, et efficacité. Je pense avoir reçu son message sur une vision de la recherche et j'espère pouvoir l'honorer dans le futur. Qu'il trouve ici tous mes profonds respects, et mon admiration considérable.

Je remercie également Monsieur H. Boutabia professeur à l'université Badji Mokhtar- Annaba, pour l'intérêt qu'il a accordé à ce travail et pour la spontanéité avec laquelle il a accepté de présider le jury de ma soutenance. Qu'il puisse croire à ma grande estime et à ma très haute considération.

Je tiens à exprimer ma profonde reconnaissance et toute ma gratitude à Madame C. Huber-Carol professeur à l'université René Descartes, Paris5. Je suis très sensible à l'honneur qu'elle m'a fait en acceptant de juger cette thèse.

Je tiens à remercier très fort Mesdames A. Chadli, maître de conférences au département de mathématiques, université Badji Mokhtar-Annaba et O. Sadki, maître de conférences à l'université USTHB Alger, d'avoir accepté de juger ce travail. Qu'elles trouvent ici, l'expression de ma très haute considération.

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

J'aimerais aussi exprimer mes profonds respects à ma femme pour son encouragement et sa patience.

Je remercie très fort mes beaux parents, mes beaux frères et mes belles sœurs.

Je remercie chaleureusement et particulièrement M. Vincent Couallier pour son aide, sa patience et sa gentillesse, M. Léo Gerville-Réache, M. Pierrick Legrand, Pr. Badr- Eddine Ainseba, M. ramzan Tahir, C. Osmane, Melle Ingrid Rochelle, et toute l'équipe de

« Fiabilité et Statistiques Mathématiques », Bordeaux 2, pour leur aide et leur soutien.

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

J'exprime tous mes remerciements aux enseignants du département de mathématiques.

Mes remerciements vont plus particulièrement aux Pr. Sissaoui H., Pr. Makhlouf, Pr. Khodja, Bouchahed L., Kouche M., Chemam R. (départ. de physique), Marzoug M. (départ. de génie civil), (Départ. de mines), Pr. Hamdaoui W. (départ. De chimie de procédés), Pr. Hadadi S. (univ. de Guelma)...

Mes vifs remerciements vont aussi à Yacine Ait Ammar, Redha Souakri, N. Bouadjila, B. Trouzine, H. Trouzine, Chakri H., etc.

Mes vifs remerciements vont aussi à tous mes amis que je ne peux pas compter...

Mes vifs remerciements vont à Nadjib., Foued, Djaber, Taib, Adel, Farouk M,

Table de Matières

Liste des figures.....	1
Liste des tableaux.....	2
Liste des abréviations.....	3
Introduction.....	4
Chapitre 1. La famille des lois Gaussiennes inverses.....	8
1. Introduction.....	8
2. Propriétés de la famille des lois Gaussiennes inverses.....	8
3. Caractéristiques.....	13
4. Propriétés statistiques.....	14
5. Quelques applications.....	15
Chapitre 2. Test d'ajustement pour la famille des lois Gaussiennes inverses.....	20
1. Introduction.....	20
2. La théorie de Chi-deux.....	20
3. Tests d'ajustement pour la loi GI.....	28
4. Test d'ajustement de la loi GI contre la loi log-normale.....	39
5. Test d'ajustement de la loi GI contre la loi de Weibull généralisée.....	43
6. Test d'ajustement de Chi-deux pour la famille des lois GI pour les données censurées.....	46
Chapitre 3. Modèles avec covariables basés sur la famille des lois Gaussiennes inverses.....	58
1. Introduction.....	58
2. Le modèle AFT.....	58
3. Analyse des systèmes redondants.....	69
4. Autre modèles.....	79
Perspectives.....	82
Références.....	83
Liste des publications et communications.....	89
Annexe.....	92

Liste des figures

- 1.1. Fonctions de densité de la distribution GI.
- 1.2. Fonctions de hasard de la distribution GI.
- 2.1. Comparaison des fonctions de densité.
- 2.2. Comparaison des fonctions de survie.
- 2.3. Histogrammes de Y_2^n pour $(\mu, \lambda) = (1, 1)$, et $n = 20, 30, 50, 80, 100, 150, 200, 300, 400, 500$.
- 2.4. Comparaison des densités de GI et LN pour les données d'ERCA.
- 2.5. Graphes de la puissance pour la distribution LN.
- 2.6. Comparaison des densités de GI et de WG pour les données d'ERCA.
- 2.7. Graphes de la puissance pour la distribution WG.
- 2.8. Graphes des fonctions de densité, de survie et de hasard pour les données d'Efron.
- 3.1. Graphes des estimateurs de fonctions de distributions $\hat{F}_1, \hat{K}_2, \hat{K}_3$ et \hat{K}_4 ,
(La distribution Gaussienne inverse).

Liste des tableaux

- 1.1. Les quatre premiers moments de la distribution GI.
- 1.2. Les estimateurs de MVUE de quelques paramètres
- 2.1. Airborne communication transceiver data.
- 2.2. Niveau empirique pour $\mu = 0.25$ et $\lambda = (0.5, 1, 5, 10)$.
- 2.3. Niveau empirique pour $\mu = 1$ et $\lambda = (0.5, 1, 5, 10)$.
- 2.4. Niveau empirique pour $\mu = 4$ et $\lambda = (0.5, 1, 5, 10)$.
- 2.5. Niveau empirique pour $\mu = 10$ et $\lambda = (0.5, 1, 5, 10)$.
- 2.6. Etude de la puissance de la loi GI contre la loi LN.
- 2.7. Etude de la puissance de la loi GI contre la loi WG.
- 2.8. Quelques Tests d'ajustement pour les données censurées aléatoires à droite.
- 2.9. Données d'Efron.
- 3.1. Graphes des estimateurs de fonctions de distributions F_1 , K_2 , K_3 et K_4 .

Liste des abréviations

$\mathcal{L}(\cdot)$: la convergence en loi.

GI : Gaussienne inverse.

BS : Birnbaum–Saunders.

LN : Log-normale.

LL : Log-logistique.

WG : Weibull généralisée.

MM : Méthode des moments.

MV : Maximum de vraisemblance.

RRN : Rao-Robson-Nikulin.

DN : Dzhaparidze-Nikulin.

HRM : Hsuan-Robson-Mirvaliev.

MVUE : Estimateur non biaisés à variance minimale (Minimum Variance Unbiased Estimator).

Introduction

La thèse est consacrée aux études des modèles statistiques appliqués en fiabilité et analyse de survie, en particulier aux études de la famille paramétrique des lois Gaussiennes inverses. Il faut noter que la monographie de Bagdonavičius et Nikulion (2002) a été faite 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 ces propriétés, et c'est pour cela qu'elle souvent apparaisse comme un modèle compétitif parmi toutes les lois dont les fonctions de hasard sont unimodales. Puisque cette famille n'a pas été bien étudiée, nous commençons ici par la construction du test d'ajustement de Chi-deux pour cette famille elle même, et par présentation des résultats d'estimation statistique liés avec cette famille.

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 de chi-deux et dépend même à la fois des paramètres en question et de la méthode d'estimation.

Fisher (1924) 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 de 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), Nikulin (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.

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 (2011).

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 la famille des distributions Gaussiennes inverses.

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 les chercheurs de différents domaines, notamment en fiabilité, actuariat, histomorphométrie⁽¹⁾, réseaux électriques, hydrologie, management, démographie, tests de vie accélérés, météorologie, la 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 intéressant (voir Chhikara et Folks (1989)). La fonction de hasard de la loi Gaussienne inverse a une forme unimodale (croissante puis décroissante), de même que les distributions log-normale, log-logistique et Weibull généralisée [Bagdonavičius et Nikulin (2002)]. Sachant que les distributions présentant une fonction de hasard de forme 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 ces distributions, la fonction de hasard croit de zéro à son maximum puis diminue jusqu'à zéro. Par contre, pour la distribution Gaussienne inverse, la fonction de hasard croit 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.

Pour étudier la validité de ce modèle, nous avons développé la statistique de Rao-Robson-Nikulin proposée par [Nikulin (1973)], connue sous le nom de la statistique RRN (voir aussi : Drost (1988), Van Der Vaart (1998)), pour la distribution Gaussienne inverse avec estimation des paramètres par la méthode du maximum de vrai-

semblance directe. 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.

(1) : Histomorphométrie : l'étude quantitative de l'organisation et la structure microscopique d'un tissu (comme les os), en particulier par l'analyse assistée par ordinateur des images formées par un microscope.

D'autre part, nous avons fait une étude comparative [Lemeshko et al. (2010)] entre le test de Rao-Robson-Nikulin et les tests non-paramétriques de Kolmogorov-smirnov, de 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 trois chapitres.

Dans le premier, nous donnons une description générale de la distribution Gaussienne inverse et quelques propriétés et caractéristiques de cette distribution, puis nous donnons le tableau des estimateurs non biaisés à variance minimale des paramètres.

Dans le chapitre deux, nous exposons la théorie du test du chi-deux en donnant les méthodes d'estimation, puis nous construisons le test de Rao-Robson-Nikulin pour la loi Gaussienne inverse. Ensuite, nous construisons 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. Nous avons calculé la puissance de ces deux tests.

Dans le dernier chapitre, nous développons quelques applications de la loi Gaussienne inverse en fiabilité et en analyse de survie en construisant des modèles avec covariables basés sur celle-ci.

La famille des lois Gaussiennes inverses

1 Introduction

Dans ce chapitre, nous illustrons brièvement quelques caractéristiques et quelques propriétés statistiques de la distribution Gaussienne inverse (GI). Pour plus de détails, on peut voir Chhikara et Folks (1989), Seshadri (1993) et Seshadri (1999).

2 Propriétés de la famille des lois Gaussiennes inverses

2.1 Densité de probabilité

Définition 1.1. On dit qu'une variable aléatoire X suit une loi *Gaussienne inverse*, et on note $X \sim IG(\mu, \lambda)$ si sa fonction de densité s'écrit :

$$f(x, \mu, \lambda) = \left(\frac{\lambda}{2\pi x^3} \right)^{\frac{1}{2}} \exp \left\{ -\frac{\lambda(x - \mu)^2}{2\mu^2 x} \right\}, \quad x \geq 0, \quad \mu, \lambda > 0. \quad (1.1)$$

Où μ et λ sont deux paramètres strictement positifs, et μ représente la moyenne de X . La fonction de densité (1.1) 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 :

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

Le paramètre $\phi = \frac{\lambda}{\mu}$ est souvent appelé le paramètre de forme (shape parameter). Plus ϕ prend des valeurs modérées plus la densité (1.1) est fortement dissymétrique, et plus ϕ est grand, plus la densité (1.1) est proche d'une densité normale. Dans la figure 1.1, nous représentons la densité de $IG(\mu, \lambda)$ en fixant un paramètre et en variant l'autre.

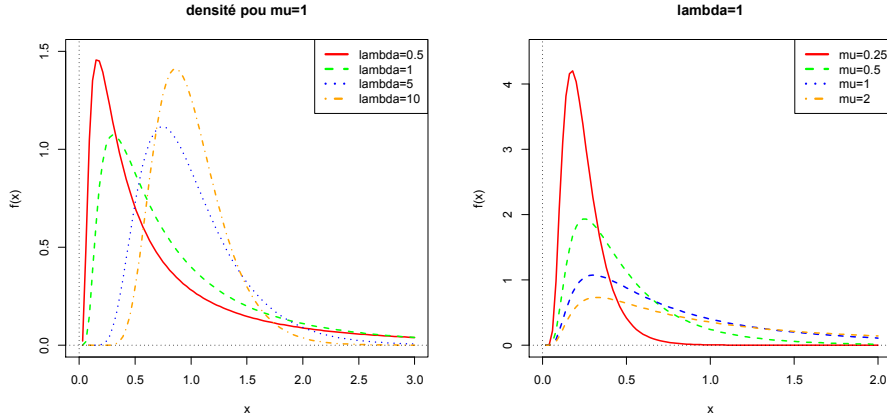


Fig. 1.1. Fonctions de densité de la distribution GI.

La fonction (1.1) 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 propriétés analogues à la distribution normale.

2.2 Fonction caractéristique et moments

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

$$M_X(t) = E(e^{tX}) = \int_0^{+\infty} e^{tx} f(x, \mu, \lambda) dx = \exp \left\{ \frac{\lambda}{\mu} \left(1 - \left(1 - \frac{2\mu^2 t}{\lambda} \right)^{\frac{1}{2}} \right) \right\}, \quad (1.2)$$

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

$$C_X(t) = E(e^{itX}) = \exp \left\{ \frac{\lambda}{\mu} \left(1 - \left(1 - \frac{2i\mu^2 t}{\lambda} \right)^{\frac{1}{2}} \right) \right\}. \quad (1.3)$$

ordre	moment simple	moment centré	cumulants
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	$\mu^4 + 6\mu^5/\lambda + 15\mu^6/\lambda^2 + 15\mu^7/\lambda^3$	$3\mu^6/\lambda^2 + 15\mu^7/\lambda^3$	$15\mu^7/\lambda^3$

Tab. 1.1. Les quatre premiers moments de la distribution GI.

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 (1.2) 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 : $E(X) = \mu$, $E(\frac{1}{X}) = \frac{E(X^2)}{\mu^3} = \frac{1}{\mu} + \frac{1}{\lambda}$.

Définition 1.2. 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.

$$g(t) = \ln E(e^{tX}) = \sum_{r=1}^{\infty} \kappa_r \frac{t^r}{r!} = \mu t + \sigma^2 \frac{t^2}{2} + \dots, \quad (1.4)$$

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 (1.4) au point $t = 0$, i.e. $\kappa_r = g^{(r)}(0)$.

En utilisant l'équation (1.2), la fonction génératrice des cumulants de $IG(\mu, \lambda)$ est donnée par :

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

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

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

et les cumulants κ_r sont donnés par la relation suivante :

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

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_{GI} , 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.

2.3 La fonction de répartition

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

$$\begin{aligned} F(x, \mu, \lambda) &= \int_0^x \left(\frac{\lambda}{2\pi t^3} \right)^{\frac{1}{2}} \exp \left\{ -\frac{\lambda(t-\mu)^2}{2\mu^2 t} \right\} dt \\ &= \Phi \left(\sqrt{\frac{\lambda}{x}} \left(\frac{x}{\mu} - 1 \right) \right) + \exp \left(\frac{2\lambda}{\mu} \right) \Phi \left(-\sqrt{\frac{\lambda}{x}} \left(\frac{x}{\mu} + 1 \right) \right), \end{aligned} \quad (1.5)$$

où $\Phi(\cdot)$ est la fonction de répartition de la loi normale standard $N(0, 1)$. La fonction de survie est :

$$S(x, \mu, \lambda) = 1 - F(x, \mu, \lambda) = \Phi \left(-\sqrt{\frac{\lambda}{x}} \left(\frac{x}{\mu} - 1 \right) \right) - \exp \left(\frac{2\lambda}{\mu} \right) \Phi \left(-\sqrt{\frac{\lambda}{x}} \left(\frac{x}{\mu} + 1 \right) \right).$$

Théorème 1.1 Si le rapport $\frac{\lambda}{\mu} \rightarrow \infty$ alors la fonction de répartition de X devient :

$$F(x, \mu, \lambda) = \Phi \left(\sqrt{\frac{\lambda}{x}} \left(\frac{x}{\mu} - 1 \right) \right).$$

Preuve :

Il suffit de montrer que :

$$\lim_{\frac{\lambda}{\mu} \rightarrow \infty} \exp \left(\frac{2\lambda}{\mu} \right) \Phi \left(-\sqrt{\frac{\lambda}{x}} \left(\frac{x}{\mu} + 1 \right) \right) = 0, \quad (1.6)$$

soit $y = \sqrt{\frac{\lambda}{x}} \left(\frac{x}{\mu} - 1 \right)$ alors (1.5) devient :

$$F(x, \mu, \lambda) = \Phi(y) + \exp\left(\frac{2\lambda}{\mu}\right) \Phi\left(-\sqrt{y^2 + \frac{4\lambda}{\mu}}\right), \quad (1.7)$$

d'où suivant (1.7) :

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

Ce dernier résultat est obtenu en appliquant la règle de l'Hôpital.

2.4 La fonction de hasard

La fonction de *hasard* de la loi Gaussienne inverse est donnée par :

$$h(x, \mu, \lambda) = \frac{f(x, \mu, \lambda)}{S(x, \mu, \lambda)} = \frac{\left(\frac{\lambda}{2\pi x^3}\right)^{\frac{1}{2}} \exp\left\{-\frac{\lambda(x-\mu)^2}{2\mu^2 x}\right\}}{\Phi\left(-\sqrt{\frac{\lambda}{x}} \left(\frac{x}{\mu} - 1\right)\right) - \exp\left(\frac{2\lambda}{\mu}\right) \Phi\left(-\sqrt{\frac{\lambda}{x}} \left(\frac{x}{\mu} + 1\right)\right)}, \quad x \geq 0. \quad (1.8)$$

On peut facilement vérifier que cette fonction est unimodale (ou de forme \cap). Cette forme est très appliquée en fiabilité.

On peut facilement démontrer que :

$$\lim_{x \rightarrow \infty} h(x, \mu, \lambda) = \frac{\lambda}{2\mu^2}. \quad (1.9)$$

On note que les familles de distributions possédant une fonction de hasard unimodale ne sont pas nombreuses, qu'on peut citer : la famille de distributions log-normales, log-logistiques, de Weibull généralisées [Bagdonavičius et Nikulin (2002)], de Birnbaum-Saunders [Birnbaum et Saunders (1969a,1969b)] et Gaussiennes inverses. Il s'avère que la distribution Gaussienne inverse est très concurrente à ces distributions et elle offre certains avantages par rapport à celles-ci,

parce que pour les distributions log-normale, log-logistique et de Weibull généralisée, la fonction de hasard croit de 0 à son maximum puis diminue jusqu'à 0. Par contre pour la distribution Gaussienne inverse, La fonction de hasard (1.8) croit de 0 à son maximum

puis tend asymptotiquement vers une constante non nulle égale à $\lambda/2\mu^2$. 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 distributions log-normale, log-logistique et de Weibull généralisée escamotent la panne ce qui n'est pas possible pour la plus part des systèmes.

Dans la figure 1.2 nous représentons la fonction de hasard de la distribution Gaussienne inverse en fixant un paramètre et en variant l'autre.

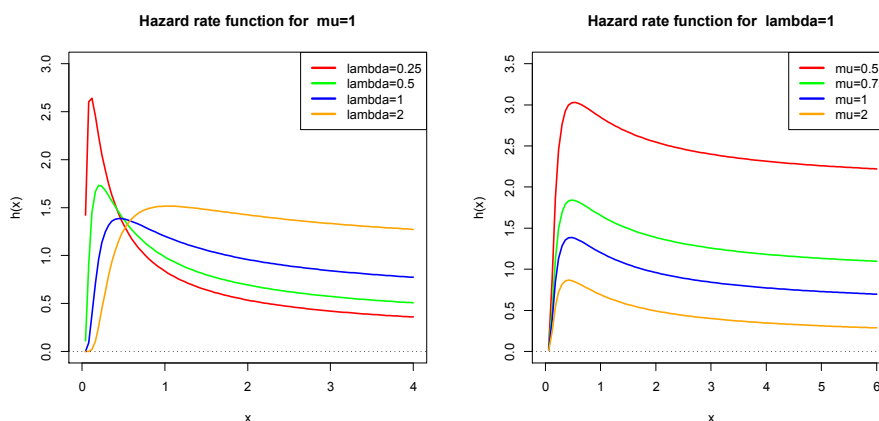


Fig.1.2. Fonctions de hasard de la distribution GI.

2.5 Analogies avec la loi Normale

Proposition 1.2. [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 1.3. [Seshadri (1993)]. Si $X_i \sim IG(\mu, \lambda)$, alors $X_i^{-1} \sim IG(1/\mu, \lambda/\mu^2) * \Gamma(1/2, \lambda/2)$.

3 Caractéristiques

Théorème 1.2. [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 Gaussienne inverse.

4 Propriétés statistiques

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

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

la fonction de log-vraisemblance est :

$$\ell_n(\theta) = \frac{n}{2} \ln \lambda - \frac{n}{2} \ln(2\pi) - \frac{3}{2} \sum_{i=1}^n \ln X_i - \sum_{i=1}^n \frac{\lambda(X_i - \mu)^2}{2\mu^2 X_i}, \quad (1.10)$$

les fonctions de score sont :

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

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

Les estimateurs du maximum de vraisemblance 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 1.4. [Seshadri (1993)]. Pour la famille $IG(\mu, \lambda)$, la statistique $T = (\bar{X}, V)^T$ est exhaustive minimale et complète pour θ .

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.

5 Quelques applications

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

5.1 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 en Berg (1994) ont préconisé l'utilisation de la distribution Gaussienne inverse pour la modélisation des coûts d'indemnisation.

5.2 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.

5.3 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^{ème} 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, télédétection (photogrammétrie), l'intensité du bruit du trafic, écologie, entomologie, histomorphométrie, réseaux électriques et en cardiologie.

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 (1993)).

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

$\hat{f}(x, \mu, \lambda)$ μ connu	$\hat{f}(x, \mu, \lambda) = \begin{cases} 0, & 0 < V < \frac{(x-\mu)^2}{\mu^2 x} \\ \frac{\Gamma(\frac{n-1}{2}) \left[V - \frac{(x-\mu)^2}{\mu^2 x} \right]^{\frac{n-4}{2}}}{\sqrt{nx^3} V^{\frac{n-3}{2}} \Gamma(\frac{n-2}{2})}, & V > \frac{(x-\mu)^2}{\mu^2 x} \end{cases},$ $n > 2$
$\hat{f}(x, \mu, \lambda)$ λ connu	$\hat{f}(x, \mu, \lambda) = \begin{cases} 0, & \bar{X}n < x \\ \frac{n(n-1)\sqrt{\lambda\bar{X}^3}}{\sqrt{2\pi nx^3(n\bar{X}-x)^3}} \exp\left\{-\frac{n\lambda(\bar{X}-x)^2}{2\bar{X}x(n\bar{X}-x)}\right\}, & \bar{X}n \geq x \end{cases},$
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>

Tab. 1.2. Les estimateurs de MVUE de quelques fonctions de paramètres pour la distribution GI.

Test d'ajustement pour la famille des lois Gaussiennes inverses

1 Introduction

Dans ce chapitre, nous donnons un aperçu du test de Chi-deux de Pearson ainsi que les méthodes d'estimation et les propriétés des estimateurs. Ensuite nous construisons le test de Chi-deux modifié pour la distribution Gaussienne inverse basé sur la statistique de Rao-Robson-Nikulin. Comme on a mentionné que la distribution Gaussienne inverse est concurrente aux distributions log-normale et de Weibulle généralisée, nous construisons alors 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 et ceci en utilisant la statistique de Rao-Robson-Nikulin (RNN) avec les classes de Neyman-Pearson.

2 La théorie du test de Chi-deux.

2.1 Test de Chi-deux de Pearson.

Soit X_1, X_2, \dots, X_n un n -échantillon. On considère le problème de tester l'hypothèse H_0 suivant :

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

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 :

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

Soit $p(\theta) = (p_1(\theta), p_2(\theta), \dots, p_r(\theta))^T$, avec $p_i(\theta) = \int_{a_{i-1}}^{a_i} f(x, \theta) dx$, $i = 1, 2, \dots, r$.

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 les données groupées 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 de 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)

$$\lim_{n \rightarrow \infty} \mathbf{P} (X_n^2(\theta) \geq x) = \mathbf{P} (\chi_{r-1}^2 \geq x),$$

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^* , il a été montré que la limite de distribution de la statistique de Pearson $X_n^2(\theta)$ ne sera 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'estimateurs 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'estimateurs 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 Le-Cam (1956) de la normalité asymptotique locale (local asymptotic normality) (LAN) et différentiabilité asymptotique de L_n au point θ sont satisfaites :

- 1) $L_n(\theta + \frac{1}{\sqrt{n}}h) - L_n(\theta) = \frac{1}{\sqrt{n}}h^T \Lambda_n(\theta) - \frac{1}{2}h^T I^{-1}(\theta)h + o_p(1)$, $h \in \mathbf{R}^s$,
- 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 \sqrt{n} - consistants estimateurs θ_n^* de θ , i.e. si $\theta_n^* \in C^*$, alors :

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

où $U = U(\theta)$ est une matrice de covariances non dégénérée . On remarque que les conditions de Lecam sont satisfaites. On considère la suite des \sqrt{n} - consistants estimateurs $\{\hat{\theta}_n\}$ 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 et d'après l'équation (2.5) la séquence des estimateurs [LeCam (1956), LeCam et al. (1983)] :

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

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 :

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

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 :

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

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

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

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

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

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) est 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 montre que l'équation (2.6) est équivalent à l'équation (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 alors 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 \sqrt{n} -consistant estimateur $\bar{\theta}_n$ de θ alors on obtient :

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

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

où χ_i^2 sont des variables aléatoires indépendantes obéissant à une loi du 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 Test d'ajustement du Chi-deux modifié

1) En 1973, Nikulin (1973a, 1973b) a proposé une modification du test de Pearson de Chi-deux pour la famille de distributions continues avec des paramètres de changement 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 Rao-Robson-Nikulin (notée RNN)(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 :

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

Pour calculer la statistique RNN, 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 :

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

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

$$U_n^2(\theta_n^*) = X_n^T(\theta_n^*) (I_r - B(\theta_n^*) I^{-1}(\theta_n^*) B^T(\theta_n^*)) X_n(\theta_n^*), \quad \theta_n^* \in C^*.$$

Le comportement asymptotique de la statistique de DN $U_n^2(\theta_n^*)$ est donné par ce théorème.

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

$$\lim_{n \rightarrow \infty} \mathbf{P}(U_n^2(\theta_n^*) > x) = \mathbf{P}(\chi_{r-s-1}^2 > x), \quad r > s + 2.$$

3) En 1976, Hsuan et Robson (1976) ont proposé une modification Y_n^2 du test standard de 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

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

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

Théorème 2.7. [Hsuan et Robson (1976)] Sous les conditions de régularité :

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

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

3 Tests d'ajustement pour la loi Gaussienne inverse

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 :

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

où $F_{GI}(x, \theta)$ est la fonction de répartition de la loi Gaussienne inverse définie par l'équation (1.5).

Soit $f_{GI}(x, \theta)$ la fonction de densité correspondante de $F_{GI}(x, \theta)$.

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 $\nu = (\nu_1, \nu_2, \dots, \nu_r)^T$ obtenu en regroupant l'échantillon X_1, X_2, \dots, X_n dans les sous-intervalles I_j .

Supposons que : $p_i(\theta) = \int_{a_{i-1}}^{a_i} f_{GI}(x, \theta) dx = F_{GI}(a_i) - F_{GI}(a_{i-1})$, $i = 1, 2, \dots, r$.

et désignons par $X_n(\theta)$ le vecteur :

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

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

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

Pour la calculer, nous prenons l'espérance des dérivées secondes suivantes de l'équation (1.10) :

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

$$\frac{\partial^2 \ell_n}{\partial \lambda^2} = -\frac{n}{2\lambda^2},$$

$$\frac{\partial^2 \ell_n}{\partial \mu \partial \lambda} = 0.$$

Pour tester l'hypothèse H_0 , nous pouvons utiliser la statistique de Rao-Robson-Nikulin donnée par l'équation (2.11) :

$$Y_n^2(\hat{\theta}_n) = X_n^2(\hat{\theta}_n) + \frac{1}{n} L^T(\hat{\theta}_n) (I(\hat{\theta}_n) - J(\hat{\theta}_n))^{-1} L(\hat{\theta}_n), \quad (2.12)$$

où

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

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

$$L(\theta) = (L_1(\theta), L_2(\theta))^T; \quad L_1(\theta) = \sum_{i=1}^r \frac{\nu_i}{p_i} \frac{\partial p_i}{\partial \mu}(\theta), \quad L_2(\theta) = \sum_{i=1}^r \frac{\nu_i}{p_i} \frac{\partial p_i}{\partial \lambda}(\theta).$$

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 :

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

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

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

Calculons les dérivées partielles de $p_i(\theta)$:

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

$$\frac{\partial p_i(\theta)}{\partial \lambda} = \frac{1}{\sqrt{2\lambda}} \left(\frac{1}{\sqrt{a_i}} \left(\frac{a_i}{\mu} - 1 \right) \varphi(\alpha_i) - \frac{1}{\sqrt{a_{i-1}}} \left(\frac{a_{i-1}}{\mu} - 1 \right) \varphi(\alpha_{i-1}) \right) + \frac{2}{\mu} \exp\left(\frac{2\lambda}{\mu}\right) (\Phi(\beta_i) - \Phi(\beta_{i-1})) +$$

$$\frac{1}{\sqrt{2\lambda}} \exp\left(\frac{2\lambda}{\mu}\right) \left(-\frac{1}{\sqrt{a_i}} \left(\frac{a_i}{\mu} + 1\right) \varphi(\beta_i) + \frac{1}{\sqrt{a_{i-1}}} \left(\frac{a_{i-1}}{\mu} + 1\right) \varphi(\beta_{i-1}) \right), \quad (2.15)$$

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 suit la distribution χ_{r-1}^2 à $r-1$ degrés de liberté (voir théorème 2.5).

La statistique (2.12) a été calculée d'une manière explicite [Saaidia et Seddik-Ameur (2010)] pour les différentes situations, selon que les paramètres soient connus ou non.

1. Cas où $\theta = (\mu, \lambda)^T$ est connu.

Dans ce cas, la statistique (2.12) devient :

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

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

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

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.

3. Cas où θ est inconnu.

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.
--

Tab. 2.1. Données d'entretien d'ERCA .

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

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

où

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

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

Au seuil α , si on choisit la valeur critique

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

l'hypothèse H_0 est acceptée si $Y_n^2(\hat{\theta}_n) \leq C_\alpha$. Dans le cas contraire, H_0 est rejetée.

Pour monter que ce test est maniable et pratique, nous l'appliquons à des observations réelle qui représentent les données d'entretien (en heures) pour un émetteur-récepteur de communication aéroporté (ERCA) "airborne communication transceiver" [Von Alven (1964)].

Pour ces données, et pour un nombre de classes équiprobables $r = 5$ (r peut être donné par la formule $r = \text{round}(1 + \ln n)$, voir Greenwood et Nikulin (1996)), 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.

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

$$\begin{cases} \sum_{i=1}^r \frac{\nu_i}{p_i(\theta)} \frac{\partial p_i(\theta)}{\partial \mu} = 0 \\ \sum_{i=1}^r \frac{\nu_i}{p_i(\theta)} \frac{\partial p_i(\theta)}{\partial \lambda} = 0 \end{cases} .$$

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

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

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

où

$p_i(\theta)$ est donné par la formule (2.13).

Pour résoudre ce système, nous pouvons utiliser la méthode de BFGS.

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.

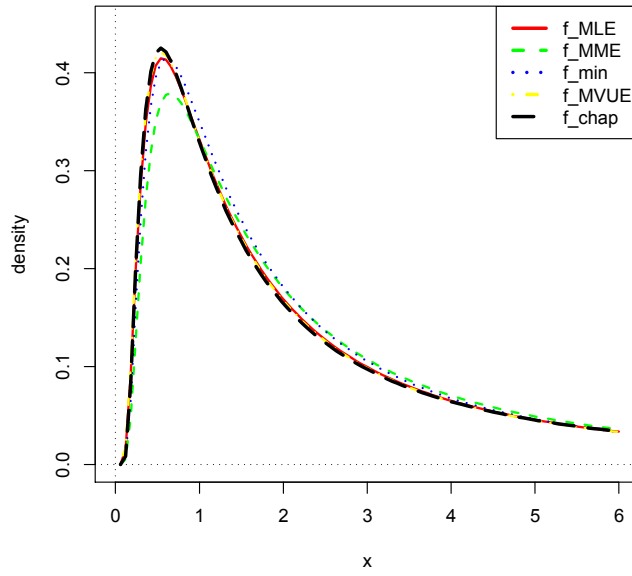


Fig. 2.1. Comparaison des fonctions de densité .

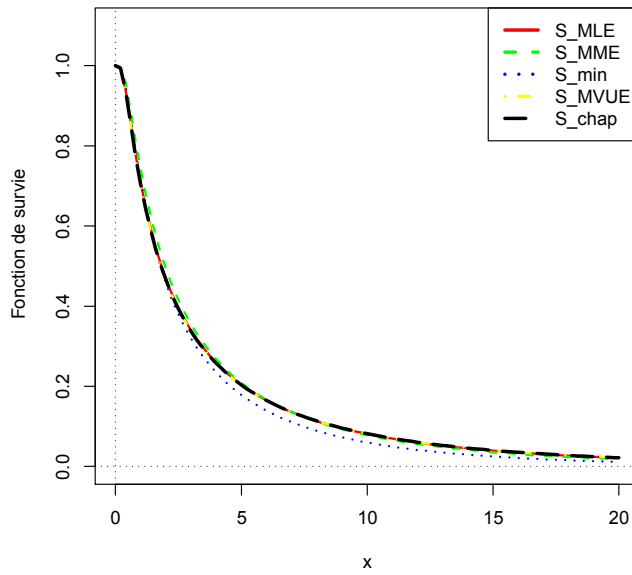


Fig. 2.2. Comparaison des fonctions de survie .

En appliquant la statistique de RNN, nous trouvons alors que $Y_{GI}^2 = 1.196187$. Pour un seuil $\alpha = 0.05$, la valeur critique $C_\alpha = \chi_{r-1,1-\alpha}^2 = 9.4877$, et nous concluons que la distribution Gaussienne inverse 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 statiques de Kolmogorov-Smirnov pour monter que la distribution Gaussienne inverse donne un bon ajustement pour ces données.

Calculons maintenant les valeurs de la statistique de RNN en considérant les distributions log-normale et de Weibull généralisée. Nous trouvons $Y_{LN}^2 = 5.5388$ et $Y_{WG}^2 = 0.8914$, d'où pour un seuil $\alpha = 0.05$, ces deux distributions donnent aussi un bon ajustement pour ces données.

3.1 Simulation

Pour étudier empiriquement le comportement de la statistique de RNN Y_n^2 , 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 RNN Y_n^2 , puis nous calculons le niveau empirique ($N.E.$) qui compte le nombre de fois où $Y_n^2 \leq C_\alpha = \chi_{r-1,1-\alpha}^2$ divisé par 5000 .

Comme le seuil $\alpha = 0.05$, alors nous avons théoriquement $N.E.=0.95$. Les résultats sont groupés dans les tableaux 2.2 à 2.5. Dans la figure (2.3), 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 en histogrammes.

Une lecture simple des tableaux nous indique que les valeurs du niveau empirique $N.E.$ sont très proches de la valeur théorique 0.5 . Dans la figure 2.3, le graphe reliant les milieux des histogrammes est confondu avec celui de la distribution χ_{r-1}^2 , ce qui confirme le théorème 2.5.

$(\mu, \lambda) = (0.25, 0.5)$										
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

Tab. 2.2. 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

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

$(\mu, \lambda) = (4, 0.5)$										
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

Tab. 2.4. 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

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

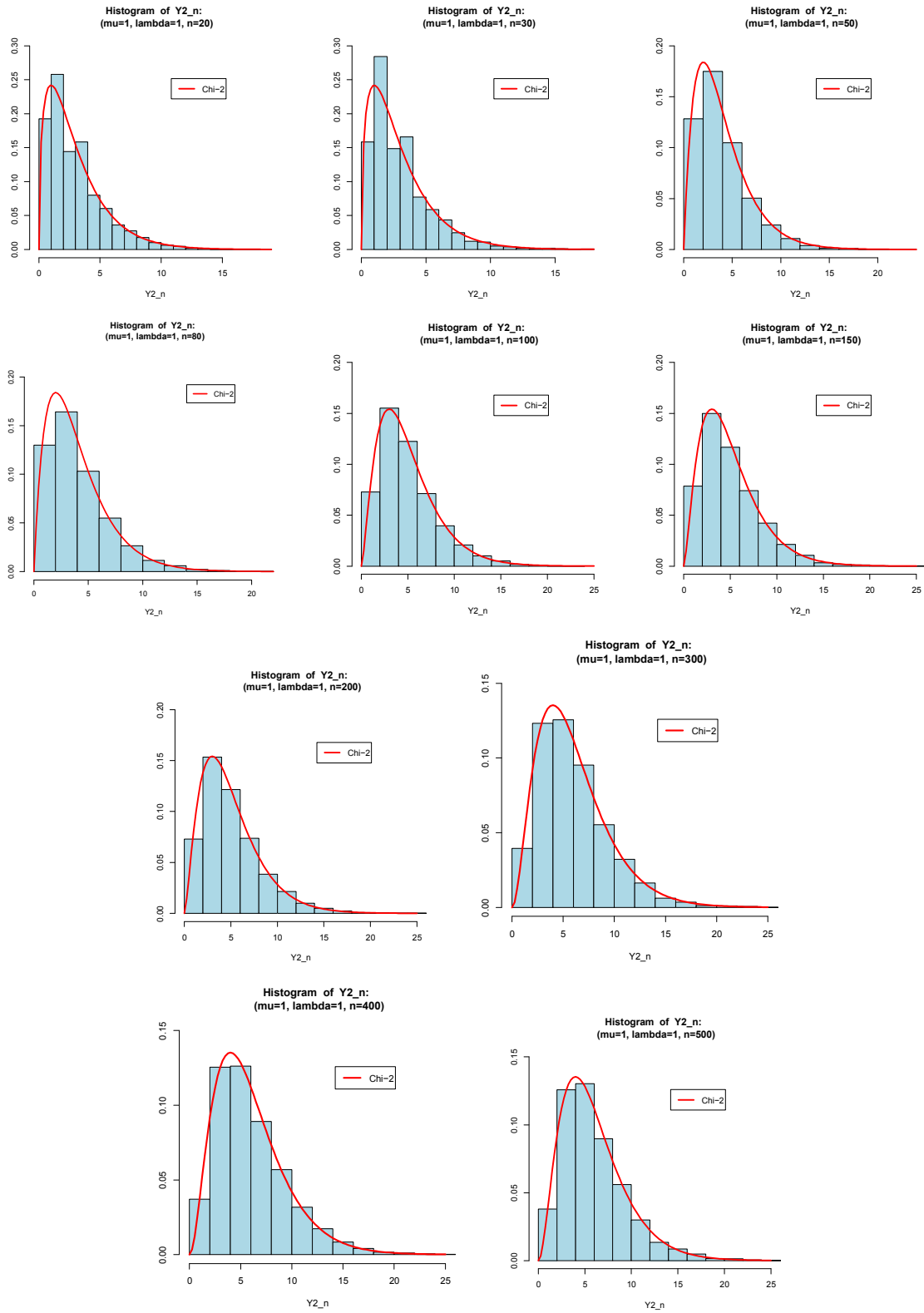


Fig. 2.3. Histogrammes de Y_2^n pour $(\mu, \lambda) = (1, 1)$, et $n = 20, 30, 50, 80, 100, 150, 200, 300, 400, 500$.

3.2 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 :

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

- **Test de Kolmogorov**

Soit la statistique :

$$D_n = \sup_{|n| < \infty} |F_n(x) - F(x, \theta)|,$$

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 $D_n = \max(D_n^+, D_n^-)$,

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

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

Si θ 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 :

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

Si θ 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 :

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

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

Remarque 2.5. Si θ est inconnu, la distribution de ces statistiques dépend d'un certain nombre de facteurs : la forme de la loi $F(x, \theta)$ 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 RRN.

4 Test du type du Chi-deux de la loi Gaussienne inverse contre la loi log-normale

Considérons l'hypothèse nulle H_0 telle que :

$$H_0 : P(X_i \leq x) = F_{GI}(x, \theta_{GI}).$$

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

$$H_1 : P(X_i \leq x) = F_{LN}(x, \theta_{LN}),$$

où

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

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

La fonction de densité de $LN(m, \sigma)$ est définit 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 RRN, il est raisonnable d'utiliser les classes de *Neyman-Pearson* (voir Greenwood et Nikulin (1996)) pour grouper les données :

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

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

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

Pour trouver les racines de (2.16), nous pouvons utiliser une méthode numérique. Pour cela nous avons réalisé un algorithme en constatant que l'équation (2.16) possède quatre point

d'intersection distincts quels que soient les paramètres $\hat{\theta}_{GI}$ 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 2.4, nous représentons les fonctions de densités $f_{GI}(x, \hat{\theta}_{GI})$ 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 $V = (V_1, V_2)$ et le vecteur de probabilités $P = (P_1, P_2)$ qui correspondent aux intervalles I_1 et I_2 tels que :

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

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

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

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

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 :

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

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

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

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

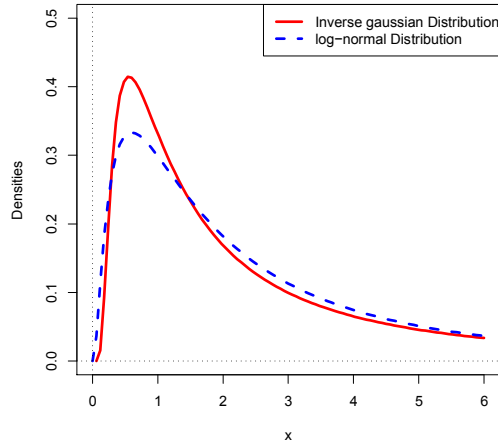


Fig. 2.4. Comparaison des densités de GI et de LN pour les données d'ERCA.

La valeur de la statistique de RRN est $Y_n^2 = 0.01699$. 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.

4.1 Etude de la puissance du test

Pour calculer la puissance du test, nous avons généré à partir de la loi log-normale avec les paramètres $m = 0$, $\sigma = 1$ des échantillons de taille n varie de 50, 60, 70, 80, 100, 150, 200, 250, 300, 350, 400, 450 à 500, et un seuil α varie de 0.01, 0.05, 0.1 à 0.2.

Pour chaque couple (n, α) nous avons procédé comme suit :

- (1) l'échantillon de taille n est généré à partir de la distribution alternative,
- (2) on calcule la valeur de la statistique de RNN,
- (3) pour le seuil en question, H_0 est rejetée si la valeur de la statistique RNN 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 le tableau 2.6, nous calculons la puissance pour chaque couple (n, α) , et nous voyons clairement que le test est puissant. Cette puissance est représentée graphiquement dans la figure (2.5).

$\alpha \backslash n$	50	60	70	80	90	100	150
0.01	0.063	0.079	0.088	0.103	0.120	0.139	0.208
0.05	0.152	0.163	0.177	0.229	0.234	0.283	0.379
0.1	0.250	0.247	0.284	0.308	0.346	0.361	0.500
0.2	0.367	0.381	0.416	0.447	0.464	0.509	0.612
$\alpha \backslash n$	200	250	300	350	400	450	500
0.01	0.252	0.387	0.413	0.549	0.554	0.616	0.686
0.05	0.477	0.544	0.626	0.702	0.774	0.805	0.825
0.1	0.578	0.675	0.754	0.788	0.827	0.878	0.888
0.2	0.702	0.754	0.802	0.847	0.891	0.917	0.952

Tab. 2.6. Etude de la puissance de la loi GI contre la loi LN.

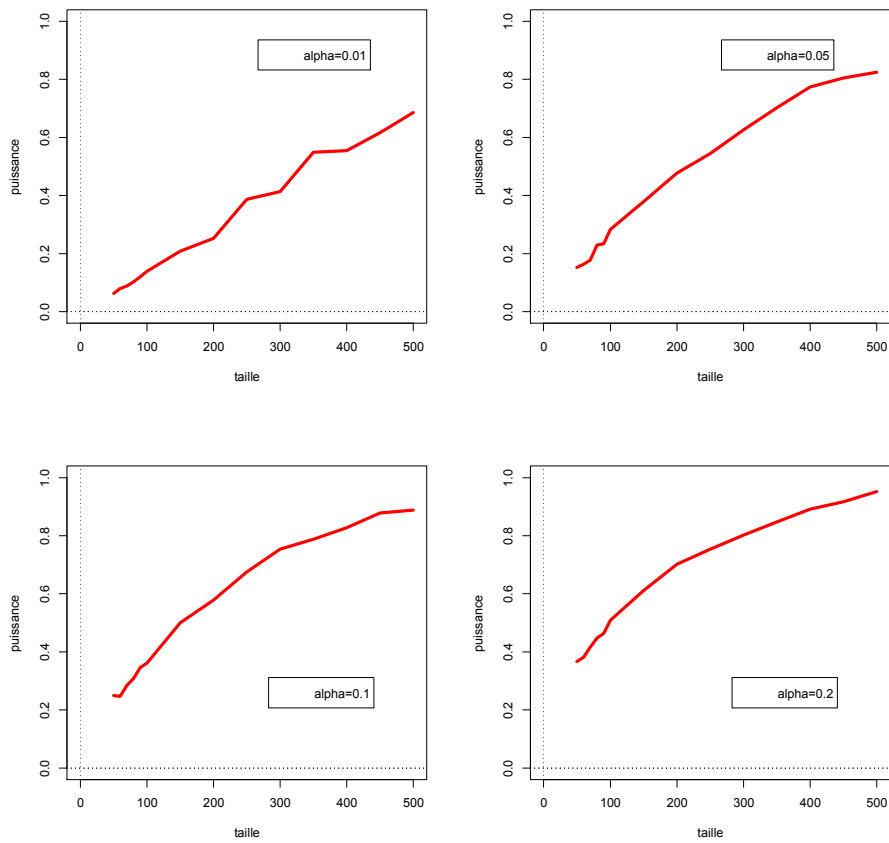


Fig. 2.5. Graphes de la puissance pour la distribution LN.

5 Test du type 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 :

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

nous voulons tester H_0 contre l'hypothèse alternative H_1 , de sorte que la distribution de X_i appartient à la famille des lois de Weibull généralisées (WG) :

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

La distribution de Weibull généralisée 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, forme \cap ou forme \cup (décroissante puis croissante), ceci explique pourquoi cette loi est la plus utilisée dans les modèles de vie accélérés. Dans le cas où la fonction de hasard est de forme \cap ou unimodale, nous avons nécessairement $\gamma > \nu > 1$.

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

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

et la fonction de survie est donnée par :

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

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

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

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

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

Cette équation possède aussi quatre points d'intersection distincts c_1^*, c_2^*, c_3^* et c_4^* , quels que soient les paramètres θ_{GI} et θ_{WG} .

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

$$\begin{aligned} \hat{\theta}_{GI} &= (\hat{\mu}_n, \hat{\lambda}_n) = (3.6756, 1.7148), \\ \hat{\theta}_{WG} &= (\hat{\sigma}_n, \hat{\nu}_n, \hat{\gamma}_n) = (0.5092, 3.0901, 7.3346), \end{aligned}$$

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

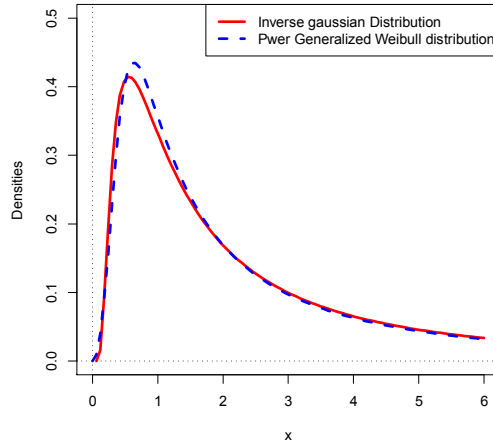


Fig. 2.6. Comparaison des densités de GI et de WG pour les données d'ERCA.

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 densités $f_{GI}(x, \hat{\theta}_{GI})$ et $f_{WG}(x, \hat{\theta}_{WG})$. Nous remarquons aussi qu'elles sont trop proches.

Après calculs, nous trouvons que la valeur de la statistique de RRN est $Y_n^2 = 0.009994682$. 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.

5.1 Etude de la puissance du test

En procédant de la même façon que dans le sous-paragraphe 4.1. Nous générons 5000 échantillons à partir de la distribution de Weibull généralisée avec les paramètres $\sigma = 2$, $\nu = 2$ et $\gamma = 4$ pour chaque taille n .

Les résultats sont résumés dans le tableau 2.7, et nous constatons que le test est puissant.

$\alpha \backslash n$	50	60	70	80	90	100	150
0.01	0.161	0.188	0.205	0.276	0.286	0.341	0.416
0.05	0.277	0.393	0.304	0.384	0.407	0.468	0.579
0.1	0.428	0.407	0.456	0.498	0.503	0.561	0.558
0.2	0.438	0.433	0.479	0.502	0.590	0.634	0.712
$\alpha \backslash n$	200	250	300	350	400	450	500
0.01	0.563	0.607	0.703	0.765	0.809	0.876	0.902
0.05	0.674	0.752	0.838	0.846	0.902	0.940	0.987
0.1	0.712	0.756	0.843	0.902	0.912	0.929	0.946
0.2	0.799	0.884	0.898	0.943	0.964	0.961	0.973

Tab. 2.7. Etude de la puissance de la loi GI contre la loi WG.

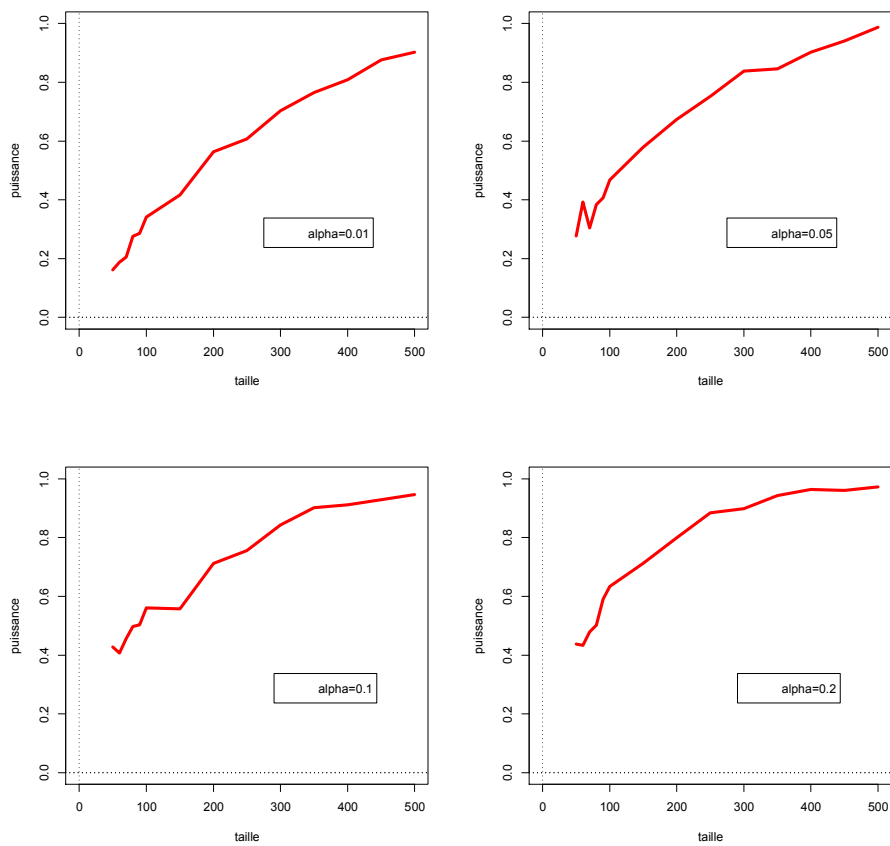


Fig. 2.7. Graphes de la puissance pour la distribution WG.

6 Test d'ajustement de Chi-deux pour la famille des loi Gaussiennes inverses en cas de données censurées

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 de Chi-deux ont été développés pour les hypothèses composites lorsque les paramètres sont estimés, les tests déjà publiés sont résumés dans le tableau 2.8.

Plusieurs tests d'ajustement ont été proposés pour des 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 de Chi-deux pour les données censurées doubles. Enfin Bagdonavičius et Nikulin (2011) ont construit un test basé sur la statistique de RRN \hat{Y}_n^2 .

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.

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

$$\delta_i = \begin{cases} 1 & , \text{ if } T_i \leq C_i \\ 0 & , \text{ otherwise} \end{cases} . \quad (2.18)$$

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 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 :

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

Tab. 2.8. Quelques Tests d'ajustement pour les données censurés aléatoires à droite.

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

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

$$L(\theta) = \prod_{i=1}^n f^{\delta_i}(X_i, \theta) S^{1-\delta_i}(X_i, \theta), \quad (2.19)$$

la fonction de log-vraisemblance est :

$$\ell(\theta) = \sum_{i=1}^n \delta_i \ln f(X_i, \theta) + \sum_{i=1}^n (1 - \delta_i) \ln S(X_i, \theta). \quad (2.20)$$

Si on pose :

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

alors l'équation (2.20) devient :

$$\ell(\theta) = \sum_{i=1}^n \delta_i \ln h(X_i, \theta) + \sum_{i=1}^n \ln S(X_i, \theta). \quad (2.21)$$

6.1 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 :

$$\ell(\mu, \lambda) = \sum_{i=1}^n \delta_i \left\{ \frac{1}{2} \ln \lambda - \frac{3}{2} \ln X_i - \lambda \frac{(X_i - \mu)^2}{2\mu^2 X_i} \right\} +$$

$$\sum_{i=1}^n (1 - \delta_i) \ln \left\{ \Phi \left(A(X_i) \right) - \exp\left(\frac{2\lambda}{\mu}\right) \Phi \left(B(X_i) \right) \right\}, \quad (2.22)$$

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 :

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

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

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.

Habib et Thomas [Habib and Thomas 1986] ont démontré que $\sqrt{n} \left(\hat{S}_n(t) - S(t, \hat{\theta}_n) \right)$ converge vers un processus Gaussien sous l'hypothèse H_0 .

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

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

Considérons le vecteur :

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

où

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

Sous les suppositions suivantes :

- 1) $f(t, \theta)$ and $F(t, \theta)$ sont deux fois continûment dérivable 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

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

où

$$Cov(Z(t_i), Z(t_j)) = S(t_i, \theta) S(t_j, \theta) \int_0^{t_i \wedge t_j} \frac{f(t, \theta)}{H(t) F^2(t, \theta)} dt.$$

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

$$\hat{Y}_n^2(\hat{\theta}_n) = \hat{Z}_n^T \hat{\Sigma}^{-1}(\hat{\theta}_n) \hat{Z}_n,$$

où la matrice $\hat{\Sigma}$ est l'estimateur de la matrice de covariance Σ et Σ^{-1} 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 :

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

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

6.2 Test de Bagdonavičius et Nikulin

Soit l'échantillon

$$(X_1, \delta_1), (X_2, \delta_2), \dots, (X_n, \delta_n), \quad (2.23)$$

où

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

Supposons que les temps de panne T_1, T_2, \dots, T_n sont n variables aléatoire 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 :

$$\ell(\theta) = \sum_{i=1}^n \delta_i \ln h(X_i, \theta) + \sum_{i=1}^n \ln S(X_i, \theta). \quad (2.24)$$

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

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

où

$$\dot{\ell}(\theta) = \frac{\partial}{\partial \theta} \ell(\theta) = \left(\frac{\partial}{\partial \theta_1} \ell(\theta), \dots, \frac{\partial}{\partial \theta_m} \ell(\theta) \right)^T. \quad (2.25)$$

La matrice d'information de Fisher est :

$$I(\theta) = -E_{\theta} \ddot{\ell}(\theta), \quad (2.26)$$

où

$$\ddot{\ell}(\theta) = \sum_{i=1}^n \delta_i \frac{\partial^2}{\partial \theta^2} \ln h(X_i, \theta) - \sum_{i=1}^n \frac{\partial^2}{\partial \theta^2} \Lambda(X_i, \theta). \quad (2.27)$$

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 (2.23) peut être représenté en termes de processus de comptage comme suit :

$$(N_1(t), Y_1(t), t \geq 0), \dots, (N_n(t), Y_n(t), t \geq 0), \quad (2.28)$$

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 :

$$N(t) = \sum_{i=1}^n N_i(t) \quad \text{et} \quad Y(t) = \sum_{i=1}^n Y_i(t). \quad (2.29)$$

Ecrivons les expressions des équations (2.24), (2.25), (2.26) et (2.27) en termes des processus stochastiques N_i et Y_i . Les trajectoires du processus de comptage $N_i(t)$ ont la forme :

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

Si $\delta_i = 1$, et $N_i(t) \equiv 0$, 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 :

$$\int_0^\infty \ln h(u, \theta) dN_i(u) = \begin{cases} \ln h(X_i, \theta), & \delta_i = 1, \\ 0, & \delta_i = 0. \end{cases} = \delta_i \ln h(X_i, \theta), \quad (2.30)$$

et

$$\int_0^\infty Y_i(u) h(u) du = \int_0^{X_i} h(u) du = -\ln S(X_i, \theta). \quad (2.31)$$

Alors les équations (2.24), (2.25), (2.26) et (2.27) s'écrivent :

$$\ell(\theta) = \int_0^\tau \{\ln h(u, \theta) dN(u) - Y(u)h(u, \theta)\} du, \quad (2.34)$$

$$\dot{\ell}(\theta) = \int_0^\tau \frac{\partial}{\partial \theta} \ln h(u, \theta) \{dN(u) - Y(u)h(u, \theta)\} du, \quad (2.35)$$

$$\ddot{\ell}(\theta) = \int_0^\tau \frac{\partial^2}{\partial \theta^2} \ln h(u, \theta) dN(u) - \int_0^\tau \frac{\partial^2}{\partial \theta^2} h(u, \theta) Y(u) du. \quad (2.36)$$

La matrice d'information de Fisher est :

$$I(\theta) = -E_\theta \ddot{\ell}(\theta) = \mathbf{E}_\theta \sum_{i=1}^n \int_0^\infty \frac{\partial}{\partial \theta} \ln h(u, \theta) \left(\frac{\partial}{\partial \theta} \ln h(u, \theta) \right)^T h(u, \theta) Y_i(u) du. \quad (2.37)$$

Sous les conditions suivantes [Hjort (1990)] :

1. Ils existent un voisinage Θ_0 de θ_0 la vraie valeur de θ tel que pour tout n et $\theta \in \Theta_0$, et pour tout $t \in [0, \tau]$, les dérivées partielles d'ordre 1, 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 :

$$Y_n^2 = Z^T \hat{V}^{-1} Z.$$

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} ,

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

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

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

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

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 .

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

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

Alors :

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

la fonction de hasard est :

$$h(t, \theta) = \frac{\left(\frac{\lambda}{2\pi t^3}\right)^{\frac{1}{2}} \exp\left\{-\frac{\lambda(t-\mu)^2}{2\mu^2 t}\right\}}{\Phi\left(-\sqrt{\frac{\lambda}{t}}\left(\frac{t}{\mu}-1\right)\right) - \exp\left(\frac{2\lambda}{\mu}\right)\Phi\left(-\sqrt{\frac{\lambda}{t}}\left(\frac{t}{\mu}+1\right)\right)}, \quad \theta = (\mu, \lambda)^T, \quad \mu, \lambda > 0.$$

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

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

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

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

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

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

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.
--

Tab. 2.9. Données d'Efron.

Les éléments de la matrice de Fisher :

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

se calculent de la même façon.

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 (2.9), 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 Haghghi (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 (2.24) est :

$$\hat{\theta} = (\hat{\mu}, \hat{\lambda})^T = (456.8747, 3.8889)^T, \quad \log\text{-vrais} = -5.52968.$$

Pour un nombre de classes $k = 5$, les points \hat{a}_j sont :

$$\hat{a}_1 = 1.689404, \quad \hat{a}_2 = 2.922234, \quad \hat{a}_3 = 4.755015, \quad \hat{a}_4 = 9.441967.$$

Après calcul de la statistique de RNN, nous trouvons $Y_n^2 = 9.823176$. Au seuil $\alpha = 0.05$ par exemple, la valeur critique $\chi_k^2 = 11.07050$, d'où la loi Gaussienne inverse donne aussi un bon ajustement pour les données d'Efron. Dans la figure (2.8), nous représentons la fonction de densité, la fonction de survie et la fonction de hasard pour les données d'Efron.

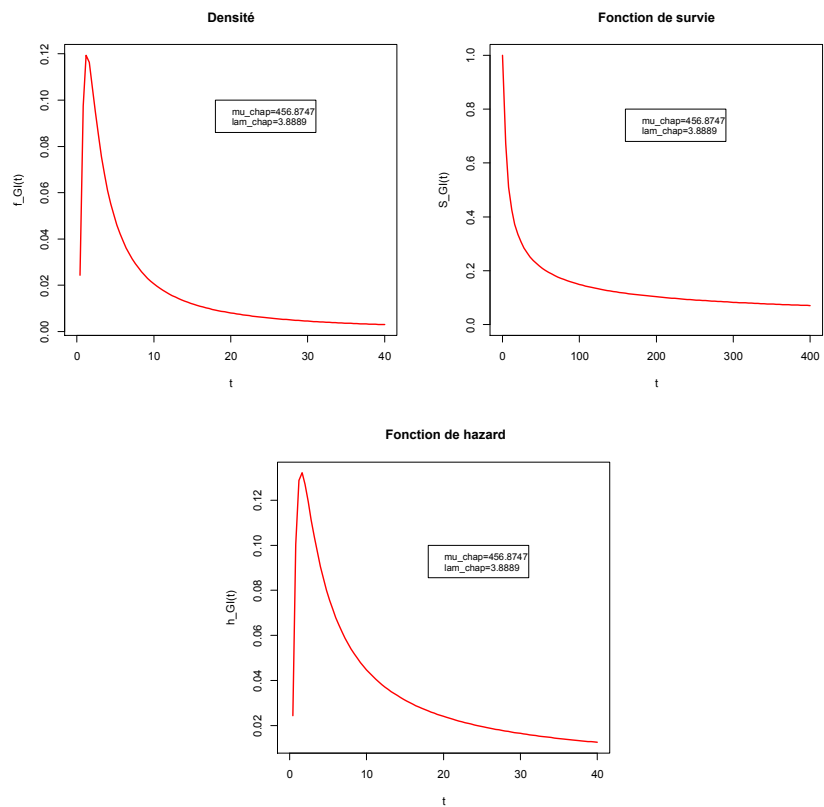


Fig. 2.8. Graphes des fonctions de densité, de survie et de hazard pour les données d'Efron.

Modèles avec covariables basés sur la famille des lois Gaussiennes inverses

1 Introduction

Puisque la distribution du temps de premier passage d'un processus du mouvement Brownien avec une dérive positive suit une loi Gaussienne inverse, l'utilisation de cette loi dans les modèles de durées de vie est particulièrement intéressant [Chhikara et Folks (1989)]. Dans ce chapitre nous verrons éventuellement quelques modèles avec covariables (ou stress) basés sur la loi Gaussienne inverse. 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.

Notre but est d'étudier le modèle AFT et les systèmes redondants basés sur la distribution Gaussienne inverse.

2 Le modèle AFT

2.1 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 Nikukin (2002), Meeker (1998), Nikulin et al. (2007)] ou le modèle de temps de panne accéléré.

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 :

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

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 :

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

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

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

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 mis 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 augmentée. 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 :

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

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

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

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, par exemple 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 (WG), log-normale (LN), log-logistique (LL), Gaussienne inverse (GI), Birnbaum-Saunders (BS), etc, et dans ce cas on obtient respectivement les modèles AFT paramétrique : AFT-Weibull, AFT-WG, AFT-LN, AFT-LL, AFT-GI, 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

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

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

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

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

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

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-GI

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

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

la fonction de hasard est :

$$h(t, \theta) = \frac{\left(\frac{\lambda}{2\pi t^3}\right)^{\frac{1}{2}} \exp\left\{-\frac{\lambda(t-\mu)^2}{2\mu^2 t}\right\}}{\Phi \left(-\sqrt{\frac{\lambda}{t}} \left(\frac{t}{\mu} - 1 \right) \right) - \exp\left(\frac{2\lambda}{\mu}\right) \Phi \left(-\sqrt{\frac{\lambda}{t}} \left(\frac{t}{\mu} + 1 \right) \right)}, \quad \theta = (\mu, \lambda)^T, \quad \mu, \lambda > 0.$$

Pour estimer les paramètres du modèle AFT-GI, 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 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)] :

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

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

alors La fonction de vraisemblance devient :

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

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

Le log-vraisemblance $\ell(\beta, \mu, \lambda)$ est :

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

Calculons les dérivées partielles de la fonction de log-vraisemblance $\ell(\beta, \mu, \lambda)$ 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)] :

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

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

$$\begin{aligned} U_{m+2}(\beta, \mu, \lambda) &= \frac{\partial \ell(\beta, \mu, \lambda)}{\partial \lambda} = \sum_{i=1}^k \sum_{j=1}^{n_i} \delta_{ij} \left(\frac{1}{2\lambda} - \frac{(K_{ij} - \mu)^2}{2\mu^2 K_{ij}} \right) + \\ &\sum_{i=1}^k \sum_{j=1}^{n_i} (1 - \delta_{ij}) \frac{\frac{1}{2\lambda} A_{ij} \varphi(A_{ij}) - e^{\frac{2\lambda}{\mu}} \left(\frac{2}{\mu} \Phi(B_{ij}) + \frac{1}{2\lambda} B_{ij} \varphi(B_{ij}) \right)}{S_{ij}}, \end{aligned} \quad (3.9)$$

où

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

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

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

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

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

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

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 :

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

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 [Saaidia 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

$$R_{ij} = \frac{1}{2\lambda} A_{ij} \varphi(A_{ij}) - e^{\frac{2\lambda}{\mu}} \left(\frac{2}{\mu} \Phi(B_{ij}) + \frac{1}{2\lambda} \varphi(B_{ij}) \varphi(B_{ij}) \right),$$

$$(R_{ij})'_\mu = \frac{\partial R_{ij}}{\partial \mu} = \frac{\sqrt{\lambda K_{ij}}}{2\lambda \mu^2} (1 - A_{ij}) \varphi(A_{ij}) +$$

$$\frac{e^{\frac{2\lambda}{\mu}}}{\mu^2} \left\{ 2 \left(1 + \frac{2\lambda}{\mu} \right) \Phi(B_{ij}) + \left(B_{ij} - \frac{2\sqrt{\lambda K_{ij}}}{\mu} \right) \varphi(B_{ij}) - \frac{\sqrt{\lambda K_{ij}}}{2\lambda} (1 - B_{ij}^2) \varphi(B_{ij}) \right\},$$

$$(S_{ij})'_\mu = \frac{\partial(S_{ij})}{\partial \mu} = \frac{1}{\mu^2} \left(\sqrt{\lambda K_{ij}} \left(\varphi(A_{ij}) - e^{\frac{2\lambda}{\mu}} \varphi(B_{ij}) \right) + 2\lambda \Phi(B_{ij}) \right).$$

$$I_{m+2, m+2} = -\frac{\partial^2 \ell(\beta, \mu, \lambda)}{\partial \lambda^2} = \frac{1}{2\lambda} \sum_{i=1}^k \sum_{j=1}^{n_i} \delta_{ij} - \sum_{i=1}^k \sum_{j=1}^{n_i} (1 - \delta_{ij}) \frac{(G_{ij})'_\lambda S_{ij} - G_{ij} (S_{ij})'_\lambda}{S_{ij}^2},$$

où

$$G_{ij} = \frac{1}{2\lambda} A_{ij} \varphi(A_{ij}) - e^{\frac{2\lambda}{\mu}} \left(\frac{2}{\mu} \Phi(B_{ij}) + \frac{1}{2\lambda} \varphi(B_{ij}) \varphi(B_{ij}) \right),$$

$$(G_{ij})'_\lambda = \frac{\partial R_{ij}}{\partial \lambda} = -\frac{1}{4\lambda^2} A_{ij} (1 + A_{ij}^2) \varphi(A_{ij}) -$$

$$e^{\frac{2\lambda}{\mu}} \left\{ \frac{4}{\mu^2} \Phi(B_{ij}) + \frac{2}{\mu \lambda} B_{ij} \varphi(B_{ij}) - \frac{1}{4\lambda^2} B_{ij} (1 + B_{ij}^2) \varphi(B_{ij}) \right\},$$

$$(S_{ij})'_\lambda = \frac{\partial(S_{ij})}{\partial \lambda} = \frac{1}{2\lambda} A_{ij} \varphi(A_{ij}) - e^{\frac{2\lambda}{\mu}} \left(\frac{2}{\mu} \Phi(B_{ij}) + \frac{1}{2\lambda} B_{ij} \varphi(B_{ij}) \right).$$

Ainsi la matrice d'information de Fisher est complètement calculée.

3 Analyse des systèmes redondants

3.1 Introduction

On considère un système de redondance avec une unité principale qui fonctionne aux conditions 'hot' et $(m-1)$ unités de réserve (stand-by units) fonctionnant aux conditions 'warm'. Si l'unité principale tombe en panne, une unité de réserve 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 distribution cumulative** 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.

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 distribution, 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 distribution, la fonction de survie et la fonction de densité des unités de réserve sont respectivement T_i, F_2, S_2 et $f_2, i = 2, \dots, m$. En utilisant le principe du modèle de Sedyakin, Bagdonavičius et al. (2008a, 2008b) ont donné la formulation mathématique basée sur le modèle AFT, et ils ont proposé des tests pour valider ces modèles.

Si une unité de réserve est passée du mode "warm" au mode "hot", sa fonction de distribution est différente de F_1 et F_2 . Le temps de panne de 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 distribution 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 Bagdonavicius et al. (2008a) 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.

$$F_2(t) = F_1(rt), \quad \forall t \geq 0, \quad r > 0. \quad (3.6)$$

La fonction de distribution du système redondant $S(1, m - 1)$ avec unité principale et $j - 1$ unités de réserve est donnée par

$$K_j(t) = \int_0^t F_1(t + ry - y) dK_{j-1}(y), \quad j = 2, \dots, m, \quad K_1(t) = F_1(t), \quad (3.7)$$

il s'ensuit que la fonction de distribution $K_m(t)$ du système $S(1, m - 1)$ est définie de façon récurrente à l'aide de l'équation (3.7). En pratique, souvent les fonctions de distributions des unités appartiennent à une famille paramétrique.

Pour $j = 2$, nous obtenons la formule de $K_2(t)$:

$$K_2(t) = \int_0^t F_1(t + ry - y)f_1(y)dy. \quad (3.8)$$

Supposons maintenant que les fonctions de distribution des unités appartiennent à la famille des distributions Gaussiennes inverses, i.e.

$$S_1(t, \mu, \lambda) = 1 - F_1(t, \mu, \lambda) = \Phi\left(-\sqrt{\frac{\lambda}{t}}\left(\frac{t}{\mu} - 1\right)\right) - \exp\left(\frac{2\lambda}{\mu}\right)\Phi\left(-\sqrt{\frac{\lambda}{t}}\left(\frac{t}{\mu} + 1\right)\right).$$

3.2 L'estimation de K_j

On suppose que ces données sont disponibles :

- a) T_{11}, \dots, T_{1n_1} de taille n_1 , T_{1i} est le temps de panne des unités testées en mode 'hot' ;
- b) T_{21}, \dots, T_{2n_2} de taille n_2 , T_{2i} est le temps de panne des unités testées en mode 'warm' .

Soit $\gamma = (r, \mu, \lambda)^T$, l'estimateur du maximum de vraisemblance $\hat{\gamma} = (\hat{r}, \hat{\mu}, \hat{\lambda})^T$ du paramètre γ maximisant la fonction de log-vraisemblance est donnée par (voir Lemeshko et al. (2010), Nikulin et Saaidia (2009)) :

$$\begin{aligned} \ell(r, \mu, \lambda) &= \frac{n}{2} \ln \lambda - \frac{n_2}{2} \ln r + \frac{\lambda n}{\mu} - \frac{3}{2} \sum_{i=1}^{n_1} \ln(T_{1i}) - \frac{3}{2} \sum_{i=1}^{n_2} \ln(T_{2i}) - \\ &\frac{\lambda}{2\mu^2} \sum_{i=1}^{n_1} T_{1i} - \frac{\lambda}{2} \sum_{i=1}^{n_1} T_{1i}^{-1} - \frac{\lambda r}{2\mu^2} \sum_{i=1}^{n_2} T_{2i} - \frac{\lambda}{2r} \sum_{i=1}^{n_2} T_{2i}^{-1}, \end{aligned} \quad (3.9)$$

where $n = n_1 + n_2$.

En calculant les dérivées de $\ell(r, \mu, \lambda)$ par rapport à r, μ, λ , Nous trouvons les fonction de score (le gradient) :

$$\begin{aligned} \frac{\partial \ell}{\partial r} &= -\frac{n_2}{2r} - \frac{\lambda}{2\mu^2} \sum_{i=1}^{n_2} T_{2i} + \frac{\lambda}{2r^2} \sum_{i=1}^{n_2} T_{2i}^{-1}, \\ \frac{\partial \ell}{\partial \mu} &= \frac{\lambda}{\mu^3} \sum_{i=1}^{n_1} T_{1i} + \frac{\lambda r}{\mu^3} \sum_{i=1}^{n_2} T_{2i} - \frac{\lambda n}{\mu^2}, \\ \frac{\partial \ell}{\partial \lambda} &= n\left(\frac{1}{2\lambda} + \frac{1}{\mu}\right) - \frac{1}{2\mu^2} \sum_{i=1}^{n_1} T_{1i} - \frac{1}{2} \sum_{i=1}^{n_1} T_{1i}^{-1} - \frac{r}{2\mu^2} \sum_{i=1}^{n_2} T_{2i} - \frac{1}{2r} \sum_{i=1}^{n_2} T_{2i}^{-1}. \end{aligned}$$

Pour déterminer l'estimateur $\hat{\gamma}$ on peut résoudre le système non linéaire des fonctions de score en égalisant celles-ci à zéro.

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 :

$$I_{11} = -E \left(\frac{\partial^2 \ell}{\partial r^2} \right) = -E \left(\frac{n_2}{2r^2} - \frac{\lambda}{r^3} \sum_{i=1}^{n_2} T_{2i}^{-1} \right) = -\frac{n_2}{2r^2} + \frac{\lambda}{r^3} \sum_{i=1}^{n_2} E(T_{2i}^{-1}).$$

Comme $F_2(t) = F_1(rt)$, alors en dérivant les deux membres, nous trouvons

$$f_2(t) = r f_1(rt),$$

alors :

$$\begin{aligned} E(T_{2i}^{-1}) &= E \left(\frac{1}{T_{2i}} \right) = \int_0^{+\infty} \frac{1}{t} f_2(t) dt = r \int_0^{+\infty} \frac{1}{t} f_1(rt) dt = r \int_0^{+\infty} \frac{1}{y} f_1(y) dy = \\ &= r E \left(\frac{1}{T_{1i}} \right) = r \left(\frac{1}{\mu} + \frac{1}{\lambda} \right), \end{aligned}$$

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 :

$$I^{-1}(\gamma) = \begin{pmatrix} \frac{2nr^2\mu}{n_1 n_2 (\mu+2\lambda)} & \frac{2r\mu^2}{n_1 (\mu+2\lambda)} & \frac{2r\mu\lambda}{n_1 (\mu+2\lambda)} \\ \frac{2r\mu^2}{n_1 (\mu+2\lambda)} & \frac{(n_1\mu+2n\lambda)\mu^3}{n_1 n (\mu+2\lambda)\lambda} & \frac{2n_2\mu^2\lambda}{n_1 n (\mu+2\lambda)} \\ \frac{2r\mu\lambda}{n_1 (\mu+2\lambda)} & \frac{2n_2\mu^2\lambda}{n_1 n (\mu+2\lambda)} & \frac{2(n\mu+2n_1\lambda)\lambda^2}{n_1 n (\mu+2\lambda)} \end{pmatrix}.$$

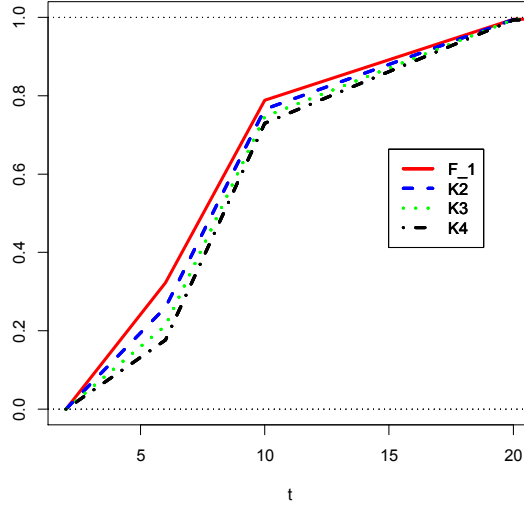


Fig. 3.1. Graphes des estimateurs de fonctions de distributions \hat{F}_1 , \hat{K}_2 , \hat{K}_3 et \hat{K}_4 , (La distribution Gaussienne inverse).

Si nous prenons $j = 2$, la fonction de distribution $K_2(t)$ 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 (3.1), nous représentons les estimateurs des fonctions \hat{F}_1 , \hat{K}_2 , \hat{K}_3 et \hat{K}_4 .

3.3 Les intervalles de confiance Asymptotiques de $K_2(t)$

Pour calculer les intervalles de confiance asymptotiques $(\underline{K}_2(t), \overline{K}_2(t))$ de $K_2(t)$ au seuil α , nous utilisons les résultats de Bagdonavicius et al. (2008a), Bagdonavicius et al. (2008b) et Bagdonavicius et al. Nikulin (2009).

$$\underline{K}_2(t) = \left(1 + \frac{1 - \hat{K}_2(t)}{\hat{K}_2(t)} \exp\left\{\frac{\hat{\sigma}_{\hat{K}_2} z_{1-\alpha/2}}{\sqrt{\hat{K}_2(t)(1 - \hat{K}_2(t))}}\right\}\right)^{-1}, \quad (3.10)$$

$$\bar{K}_2(t) = \left(1 + \frac{1 - \hat{K}_2(t)}{\hat{K}_2(t)} \exp \left\{ -\frac{\hat{\sigma}_{\hat{K}_2} z_{1-\alpha/2}}{\sqrt{\hat{K}_2(t)(1 - \hat{K}_2(t))}} \right\} \right)^{-1}, \quad (3.11)$$

où

$$\hat{\sigma}_{\hat{K}_2(t)}^2 = C_2^T(t, \hat{\gamma}) I^{-1}(\hat{\gamma}) C_2(t, \hat{\gamma}),$$

$$C_2(t, \gamma) = (C_{21}(t, \gamma), C_{22}(t, \gamma), C_{23}(t, \gamma))^T.$$

Les composantes du vecteur $C_2(t, \gamma)$ sont données par :

$$C_{21}(t, \gamma) = \int_0^t \frac{\partial F_1}{\partial r}(t + ry - y, \mu, \lambda) dF_1(y, \mu, \lambda),$$

$$C_{22}(t, \gamma) = \int_0^t \frac{\partial F_1}{\partial \mu}(t + ry - y, \mu, \lambda) dF_1(y, \mu, \lambda) + F_1(t + ry - y, \mu, \lambda) d\left(\frac{\partial F_1}{\partial \mu}(y, \mu, \lambda)\right),$$

$$C_{23}(t, \gamma) = \int_0^t \frac{\partial F_1}{\partial \lambda}(t + ry - y, \mu, \lambda) dF_1(y, \mu, \lambda) + F_1(t + ry - y, \mu, \lambda) d\left(\frac{\partial F_1}{\partial \lambda}(y, \mu, \lambda)\right).$$

Après des calculs laborieux, nous obtenons les expressions des dérivées partielles à l'intérieur des intégrales :

$$\frac{\partial F_1}{\partial r}(t + ry - y, \mu, \lambda) = \frac{y}{2} \sqrt{\frac{\lambda}{t + ry - y}} \left(\frac{1}{t + ry - y} + \frac{1}{\mu} \right) \varphi(A(t + ry - y)) -$$

$$\frac{y}{2} \exp\left(\frac{2\lambda}{\mu}\right) \sqrt{\frac{\lambda}{t + ry - y}} \left(-\frac{1}{t + ry - y} + \frac{1}{\mu} \right) \varphi(B(t + ry - y)),$$

$$\frac{\partial F_1}{\partial \mu}(t + ry - y, \mu, \lambda) = \frac{\sqrt{\lambda(t + ry - y)}}{\mu^2} \left(-\varphi(A(t + ry - y)) + \varphi(B(t + ry - y)) e^{2\lambda/\mu} \right) -$$

$$\exp\left(\frac{2\lambda}{\mu}\right) \frac{2\lambda}{\mu^2} \Phi(-B(t + ry - y)),$$

$$\begin{aligned} \frac{\partial F_1}{\partial \lambda}(t + ry - y; \mu, \lambda) &= \frac{1}{2\lambda} A(t + ry - y) \varphi(A(t + ry - y)) + \\ \frac{2}{\mu} \exp\left(\frac{2\lambda}{\mu}\right) \Phi(-B(t + ry - y)) &- \frac{1}{2\lambda} \exp\left(\frac{2\lambda}{\mu}\right) B(t + ry - y) \varphi(B(t + ry - y)), \end{aligned}$$

où

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

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

$$\begin{aligned} \left[\frac{\partial F_1}{\partial \mu}(y, \mu, \lambda) \right]'_y &= \frac{1}{2\mu^2} \sqrt{\frac{\lambda}{y}} [1 - A(y, \mu, \lambda) B(y, \mu, \lambda)] \left(-\varphi(A(y, \mu, \lambda)) + \exp\left(\frac{2\lambda}{\mu}\right) \varphi(B(y, \mu, \lambda)) \right) + \\ &\quad \frac{\lambda}{\mu^2 y} \exp\left(\frac{2\lambda}{\mu}\right) A(y, \mu, \lambda) \varphi(B(y, \mu, \lambda)), \end{aligned}$$

$$\left[\frac{\partial F_1}{\partial \lambda}(y, \mu, \lambda) \right]'_y = \frac{1}{4\lambda y} B(y, \mu, \lambda) (1 - A^2(y, \mu, \lambda)) \varphi(A(y, \mu, \lambda)) -$$

$$\frac{1}{\mu y} \exp\left(\frac{2\lambda}{\mu}\right) A(y, \mu, \lambda) \varphi(B(y, \mu, \lambda)) - \frac{1}{4\lambda y} A(y, \mu, \lambda) (1 - B^2(y, \mu, \lambda)) \varphi(B(y, \mu, \lambda)).$$

avec $A(t) = \sqrt{\frac{\lambda}{t}} \left(\frac{t}{\mu} - 1 \right)$, $B(t) = \sqrt{\frac{\lambda}{t}} \left(\frac{t}{\mu} + 1 \right)$ et $\varphi(\cdot)$ est la fonction de densité de $N(0, 1)$.

et $z_{1-\alpha/2}$ est le $(1 - \alpha/2)$ -quantile de $N(0, 1)$.

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

Tab. 3.1. Niveau de Confiance pour $K_2(t)$.

3.4 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 (3.10) et (3.11), 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 distribution $K_2(t)$ donnée par la formule (3.8) puis en divisant par 5000. Les résultats sont résumés dans le tableau 3.1.

3.5 Systèmes redondants avec données censurées

Dans le cas où les données sont censurées, nous pouvons les traiter de la même manière que pour les données complètes.

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 $d_1 = \sum_{i=1}^{n_1} \delta_{1i}$, $d_2 = \sum_{j=1}^{n_2} \delta_{2j}$.

Soit $\gamma = (r, \mu, \lambda)^T$, l'estimateur du maximum de vraisemblance $\hat{\gamma} = (\hat{r}, \hat{\mu}, \hat{\lambda})^T$ du paramètre γ maximisant la fonction de log-vraisemblance

$$\begin{aligned} \ell(r, \mu, \lambda) &= \frac{d_1 + d_2}{2} \ln \lambda - \frac{d_2}{2} \ln r + \frac{(d_1 + d_2)\lambda}{\mu} - \frac{3}{2} \sum_{i=1}^{n_1} \delta_{1i} \ln(X_{1i}) - \\ &\frac{\lambda}{2\mu^2} \sum_{i=1}^{n_1} \delta_{1i} X_{1i} - \frac{\lambda}{2} \sum_{i=1}^{n_1} \delta_{1i} X_{1i}^{-1} - \frac{3}{2} \sum_{j=1}^{n_2} \delta_{2j} \ln(X_{2j}) - \frac{\lambda r}{2\mu^2} \sum_{j=1}^{n_2} \delta_{2j} X_{2j} - \frac{\lambda}{2r} \sum_{j=1}^{n_2} \delta_{2j} X_{2j}^{-1} + \\ &\sum_{i=1}^{n_1} (1 - \delta_{1i}) \ln \left\{ \Phi \left(-\sqrt{\frac{\lambda}{X_{1i}}} \left(\frac{X_{1i}}{\mu} - 1 \right) \right) - \exp\left(\frac{2\lambda}{\mu}\right) \Phi \left(-\sqrt{\frac{\lambda}{X_{1i}}} \left(\frac{X_{1i}}{\mu} + 1 \right) \right) \right\} + \\ &\sum_{j=1}^{n_2} (1 - \delta_{2j}) \ln \left\{ \Phi \left(-\sqrt{\frac{\lambda}{rX_{2j}}} \left(\frac{rX_{2j}}{\mu} - 1 \right) \right) - \exp\left(\frac{2\lambda}{\mu}\right) \Phi \left(-\sqrt{\frac{\lambda}{rX_{2j}}} \left(\frac{rX_{2j}}{\mu} + 1 \right) \right) \right\}. \end{aligned}$$

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

alors

$$S_1(t, \mu, \lambda) = \Phi \left(-\sqrt{\frac{\lambda}{t}} \left(\frac{t}{\mu} - 1 \right) \right) - \exp\left(\frac{2\lambda}{\mu}\right) \Phi \left(-\sqrt{\frac{\lambda}{t}} \left(\frac{t}{\mu} + 1 \right) \right) = \Phi(A(t)) - \exp\left(\frac{2\lambda}{\mu}\right) \Phi(B(t)).$$

En prenons les dérivées de $\ell(r, \mu, \lambda)$ par rapport à r, μ, λ , nous trouvons les fonction de score (gradient) suivantes :

$$\begin{aligned} \frac{\partial \ell}{\partial r} &= -\frac{d_2}{2r} - \frac{\lambda}{2\mu^2} \sum_{i=1}^{n_2} \delta_{2j} X_{2i} + \frac{\lambda}{2r^2} \sum_{i=1}^{n_2} \delta_{2j} X_{2i}^{-1} + \\ &\frac{1}{2r} \sum_{i=1}^{n_2} (1 - \delta_{2j}) \frac{B(rX_{2j})\varphi(A(rX_{2j})) - e^{\frac{2\lambda}{\mu}} A(rX_{2j})\varphi(B(rX_{2j}))}{S_1(rX_{2j}, \mu, \lambda)}, \\ \frac{\partial \ell}{\partial \mu} &= -\frac{\lambda}{\mu^2} (d_1 + d_2) + \frac{\lambda}{\mu^3} \sum_{i=1}^{n_1} \delta_{1i} X_{1i} + \frac{\lambda r}{\mu^3} \sum_{i=1}^{n_2} \delta_{2j} X_{2i} + \\ &\frac{1}{\mu^2} \sum_{i=1}^{n_1} (1 - \delta_{1i}) \frac{\sqrt{\lambda X_{1i}} \left(\varphi(A(X_{1i})) - e^{\frac{2\lambda}{\mu}} \varphi(B(X_{1i})) \right) + 2\lambda e^{\frac{2\lambda}{\mu}} \Phi(B(X_{1i}))}{S_1(X_{1i}, \mu, \lambda)} + \end{aligned}$$

$$\frac{1}{\mu^2} \sum_{j=1}^{n_2} (1 - \delta_{2j}) \frac{\sqrt{\lambda r X_{2j}} \left(\varphi(A(rX_{2j})) - e^{\frac{2\lambda}{\mu}} \varphi(B(rX_{2j})) \right) + 2\lambda e^{\frac{2\lambda}{\mu}} \Phi(B(rX_{2j}))}{S_1(rX_{2j}, \mu, \lambda)},$$

$$\begin{aligned} \frac{\partial \ell}{\partial \lambda} &= \left(\frac{1}{2\lambda} + \frac{1}{\mu} \right) (d_1 + d_2) - \frac{1}{2\mu^2} \sum_{i=1}^{n_1} \delta_{1i} X_{1i} - \frac{1}{2} \sum_{i=1}^{n_1} \delta_{1i} X_{1i}^{-1} - \frac{r}{2\mu^2} \sum_{j=1}^{n_2} \delta_{2j} X_{2j} - \frac{1}{2r} \sum_{j=1}^{n_2} \delta_{2j} X_{2j}^{-1} + \\ &\sum_{i=1}^{n_1} (1 - \delta_{1i}) \frac{\frac{1}{2\lambda} \left[A(X_{1i}) \varphi(A(X_{1i})) - e^{\frac{2\lambda}{\mu}} B(X_{1i}) \varphi(B(X_{1i})) \right] - \frac{2}{\mu} e^{\frac{2\lambda}{\mu}} \Phi(B(X_{1i}))}{S_1(X_{1i}, \mu, \lambda)} + \\ &\sum_{j=1}^{n_2} (1 - \delta_{2j}) \frac{\frac{1}{2\lambda} \left[A(rX_{2j}) \varphi(A(rX_{2j})) - e^{\frac{2\lambda}{\mu}} B(rX_{2j}) \varphi(B(rX_{2j})) \right] - \frac{2}{\mu} e^{\frac{2\lambda}{\mu}} \Phi(B(rX_{2j}))}{S_1(rX_{2j}, \mu, \lambda)}. \end{aligned}$$

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.

Le calcul de l'information de Fisher, l'estimateur de la fonction de distribution du système Redondant $K_m(t)$ et les intervalles de confiance de cette dernière feront l'objet de futurs travaux.

4 Autres modèles

4.1 Le modèle de Cox

Le modèle de Cox sur E est le plus populaire en analyse de survie. il est donné en terme de la fonction de hasard :

$$\lambda_{x(\cdot)} = r(x(t)) \lambda_0(t), \quad x(\cdot) \in E,$$

où $\lambda_0(t)$ est la fonction de hasard de base (généralement inconnue) et r est une fonction positive souvent paramétrisée comme suit :

$$r(x) = e^{\beta^T x}, \quad \beta = (\beta_1, \dots, \beta_m)^T.$$

On suppose que

$$\lambda_0(t) \in \left\{ \frac{\left(\frac{\lambda}{2\pi t^3}\right)^{\frac{1}{2}} \exp\left\{-\frac{\lambda(t-\mu)^2}{2\mu^2 t}\right\}}{\Phi\left(-\sqrt{\frac{\lambda}{t}}\left(\frac{t}{\mu}-1\right)\right) - \exp\left(\frac{2\lambda}{\mu}\right)\Phi\left(-\sqrt{\frac{\lambda}{t}}\left(\frac{t}{\mu}+1\right)\right)}, \quad t \geq 0, \quad \mu > 0, \quad \lambda > 0 \right\}.$$

tel que nous avons ce qu'on appelle le modèle de Cox paramétrique-inverse Gaussien, où le vecteur β exprime l'effet de covariables $x(\cdot)$ sur la distribution de la durée de vie $T_{x(\cdot)}$.

Considérons l'hypothèse

$$H_0 : \beta = 0.$$

Sous H_0 il n'y a pas d'influence des covariables sur la distribution de la durée de vie $T_{x(\cdot)} = T$, et dans ce cas (sous H_0) la durée de vie T suit une distribution Gaussienne inverse. Donc, pour tester H_0 on peut utiliser, par exemple, le test du Chi-deux exposé précédemment.

4.2 Modèle de Fragilité

Nous parlons d'un modèle de fragilité (*frailty model*) si la fonction de hasard d'un individu est influencé par une variable aléatoire positive *non-observable* Z , appelée variable de fragilité (*frailty variable*), de telle manière :

$$\lambda_{x(\cdot)}(t|Z = z) = z\lambda_0(t)r\{x(t)\}, \quad x(\cdot) \in E,$$

où $\lambda_0(\cdot)$ est la fonction de hasard de base.

Si on suppose que la densité de la variable de fragilité Z appartient à la famille de lois Gaussiennes inverses, alors nous obtenons ce que l'on appelle *inverse Gaussian frailty model* E .

4.3 Modèle de dégradation

Dans les domaine d'engineering , les unités formant un système dégradant au fil du temps. Le niveau de détérioration d'un élément est représenté par un processus de dégradation, si ce dernier est régi par un processus de Wiener alors la distribution de premier passage de temps suit une loi Gaussienne inverse.

Définition. Soit $W(t), t > 0$ un processus. On dit que W est un processus de Wiener de tendance linéaire m et de variance σ^2 si :

- • $W(0)=0$,
- • W est un processus stochastique à accroissements indépendants à trajectoires sûrement continues,
- • $\forall t > 0, h > 0$, la loi de l'accroissement $W(t+h) - W(t)$ est une loi normale $N(mh, \sigma^2 h)$.

Propriétés.

- • $E(W(t)) = mt, \text{var}(W(t)) = \sigma^2 t,$
- • si W_0 est un processus de Wiener standard, i.e. $m = 0, \sigma^2 = 1,$ alors $W(t) = mt + \sigma W_0(t)$ est un processus de Wiener de tendance linéaire m et de variance $\sigma^2.$

Soit $W(t), t > 0$ un processus de Wiener de tendance linéaire m et de variance $\sigma^2.$

Si la dégradation est régi par un processus de Wiener de tendance linéaire m et de variance $\sigma^2,$ la panne survient si le processus de dégradation $W(t)$ atteint un niveau critique $z_0.$

Le premier temps du passage T_0 avec :

$$T_0 = \inf \{t : W(t) \geq z_0\},$$

suit une distribution Gaussienne inverse $IG(\frac{z_0}{m}, \frac{z_0^2}{\sigma^2})$ de fonction de densité :

$$f(t, z_0, m, \sigma) = \frac{z_0}{\sigma \sqrt{2\pi t^3}} \exp\left\{-\frac{(z_0 - mt)^2}{2\sigma^2 t}\right\}.$$

Conclusion et Perspectives

Les tests d'ajustement du type Chi-deux modifié sont appliqués à un modèle paramétrique de fonction de hasard unimodale couramment utilisé 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. 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 système redondants basés sur cette famille sont aussi étudiés.

Nous espérons réaliser ces taches dans le futur

- Réaliser une étude comparative des tests de Rao-Robson-Nikulin, Dzhaparidze-Nikulin et Hsuan-Robson-Mirvaliev pour la distribution Gaussienne inverse.
- Réaliser le test de type du Chi-deux de la loi Gaussienne inverse contre la loi log-logistique basé sur la statistique de Rao-Robson-Nikulin.
- Réaliser une étude comparative des tests de Rao-Robson-Nikulin, Dzhaparidze-Nikulin et Hsuan-Robson-Mirvaliev pour la distribution Birnbaum-Saunders.(En collaboration avec R.Tahir)
- Réaliser le test du type du Chi-deux la loi Gaussienne inverse contre la loi Birnbaum-Saunders basé sur la statistique de Rao-Robson-Nikulin. (En collaboration avec R.Tahir)
- Réaliser la simulation des modèles AFT-GI, AFT-LN, AFT-LL, AFT-BS et AFT-WG. (En collaboration avec R.Tahir)
- Faire estimation paramétriques et étudier le test statistique, proposé par Bagdonavicius and Nikulin, pour les systèmes redondants et faire les simulations pour ceux-ci.(En collaboration avec R.Tahir).

- Faire des logiciels permettant d'appliquer les modèles AFT semi-paramétriques 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 Rao-Blackwell-Nikulin, Kolmogorov, Cramér-von Mises, Anderson-Darling pour la distribution de Birnbaum-Saunders.
- Estimer la fonction de distribution \hat{K}_m pour les systèmes redondants et faire les simulations pour celle-ci.

Références

- [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., Messbah M., Nikulin M. (Eds)), ISTE & Wiley : London, 203–215.
- [5] Bagdonavičius, V., Kruopis, J., Nikulin, M.(2011) *Non-parametric Tests for Censored Data*, John Wiley & Sons Inc.
- [6] 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*.
- [7] Bagdonavičius, V., Masiulaityte, I., Nikukin, M. (2008a) Statistical analysis of redundant system "warm" stand-by units, *Stochastics : An International Journal of Probability and Stochastic Processes*, 80 #2-3, 115-128.
- [8] 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* (Huber C., Limnios N., Balakrishnan N., Messbah M., Nikulin M. (Eds)), ISTE & Wiley : London, 189–202 .
- [9] Bagdonavičius, V., Masiulaityte, I., Nikulin, M. (2009) *Advances in Degradation Modeling : Applications to Reliability, Survival Analysis, and Finance* (Nikulin, M., Limnios N., Balakrishnan N., Kahle W., Huber C. (Eds)), Birkhäuser Boston, 293–310.
- [10] Bagdonavičius, V., Masiulaityte, I., Nikulin, M. (2010) Parametric estimation of redundant system reliability from censored data. In : *Mathematical and Statistical Methods in Reliability* (Balakrishnan, N., Nikulin, M., Rykov, V. (Eds)), Springer Verlag : Boston.
- [11] Bagdonavičius V., Nikukin, M. (2002) *Accelerated Life Models : Modeling and Statistical Analysis*, Chapman and Hall.
- [12] Bagdonavičius, V., Nikulin, M. (2011) Chi-squared tests for general composite hypotheses from censored samples *Comptes Rendus de l'Académie des Sciences de Paris, Mathématiques*, V. 349, N° 3-4, 219-223.
- [13] Berg, P.T. (1994) Deductibles and the Inverse Gaussian Distributions, *Astin Bulletin*, 24, 319-23.

- [14] Birnbaum, Z. W. and Saunders, S. C. (1969a). A new family of life distribution, *Journal of Applied Probability*, vol. 6, 319-327.
- [15] 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.
- [16] Bolshev, L. N. (1987) Probability theory and mathematical statistics, in : Selected Works (edited by Yu. V. Prokhorov) [in Russian], Nauka, Moscow.
- [17] Bolshev, L.N., Smirnov, N.V. (1983) Tables of Mathematical Statistics. Science, Moscow,[in Russian].
- [18] Chen, J. (1975) Goodness of Fit Tests Under Random Censorship, PhD dissertation, Oregon State University.
- [19] Chernoff, H., Lehmann, E. L. (1954) The use of maximum likelihood estimates in tests for goodness of fit, *Ann. Math. Statist.*, 25 , 579–586.
- [20] Chhikara, R.S., Folks, J.L., (1977). The Inverse Gaussian Distribution as a Life Time Model. *Technometrics* , 19 , 461-468.
- [21] Chhikara, R.S., Folks, J.L. (1989) The Inverse Gaussian Distribution, Marcel Dekker, New York.
- [22] Cramer, U. (1946). *Mathematical Methods of statistics*, Princeton University Press, New York.
- [23] Drost, F. (1988) Asymptotics for Generalized Chi-squared Goodness-of-fit Tests, amsterdam : Centre for Mathematics and Computer Sciences, CWI Tracs, 48.
- [24] Dzhaparidze, K. O. , Nikulin, M. S. (1974) On a modification of the standard statistic of Pearson, *Theory of Probability and Its Applications*, 19, 851–852.
- [25] 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.
- [26] 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.
- [27] Efron, B. (1988) Logistic Regression, Survival Analysis, and the Kaplan-Meier Curve. *Journal of the American Statistical Association* 83, 414-425.
- [28] 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.
- [29] Greenwood, P. S., Nikulin, M. (1996) *A guide to Chi-squared Testing*, John Wiley and Sons, New York.
- [30] 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.
- [31] Habib, M. G. H. (1981) A Chi-Square Goodness-of Fit Test for Censored Data, PhD dissertation, Oregon State University.

- [32] 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.
- [33] Hadwiger, H. (1940) "Eine analytische reproductions-funktion fur biologische Gesamtheiten.", *Skandinavisk Aktuarietidskrift*, 23, 101-113.
- [34] Haghghi, F. (2004) Fiabilité et Analyse de Survie : Le Modèle semi-paramétrique de Dégradation. Une généralisation de la famille Weibull et ses application, Thèse de Doctorat, Université Victor Segalen Bordeaux 2.
- [35] 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.
- [36] Henze, Norbert, Klar, Bernhard (2002) Goodness-of-fit tests for the inverse gaussian distribution based on the empirical Laplace transform, *journal of Annals of the Institute of Statistical Mathematics*, vol. 54, no2, pp. 425-444
- [37] 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.
- [38] Holton, R.E. (1913) Frequency of Recurrence of Hudson River Floods, *U.S. Weather Bureau Bulletin*, Z, 109-12.
- [39] 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.
- [40] Huber, C., Limnios, N., Mesbah M., Nikulin, M. (Eds) (2008) *Mathematical Methods in Survival analysis, Reliability and Quality of Life*, Wiley-ISTE.
- [41] Ionescu, D.C, Limnios, N. (Eds) (1999) *Statistical and Probabilistic Models in Reliability*, Birkhauser Boston.
- [42] 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.
- [43] 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.
- [44] Lancaster, H.O. (1969) *The Chi-squared Distribution*, John Wiley, New York.
- [45] Lawless, J.F. (2003) *Statistical Models and Methods for Lifetime Data*, 2nd ed., New York : John Wiley.
- [46] Lawless, J.F. (2003) *Statistical Models and Methods for Lifetime Data*, 2nd ed., New York : John Wiley.
- [47] LeCam, L. (1956) Locally asymptotically normal families of distributions, In *University of California, Publication in Statistics* , vol.3, p. 37-98.
- [48] 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.
- [49] 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. (ISSN 0005-1179)

- [50] 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.
- [51] 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.
- [52] 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.
- [53] McCulloch, C. E. (1985) Relationships among some chi-squared goodness of fit statistics. Communications in Statistics-Theory and Methods, 14 :593–603. 30
- [54] Meeker, W.Q., Escobar, L.A. (1998) Statistical Methods for reliability Data, John Wiley and Sons, INC.
- [55] Mirvaliev, M. (2001) An investigation of generalized chi-squared type statistics, Ph.D. Thesis, Academy of Science, Republic of Uzbekistan, Tashkent.
- [56] 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.
- [57] 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.
- [58] 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.
- [59] Nadarajah, S., (2009) An alternative inverse Gaussian distribution, Mathematics and Computers in Simulation, V.79 , Issue 5, p : 1721-1729.
- [60] Nelson, W. (1990) Accelerated Testing : Statistical Models, Test Plans, and Data Analysis, John Wiley and Sons : New York .
- [61] 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.
- [62] Nikulin, M. Haghghi, F. (2009) On the power generalizedWeibull family : model for cancer censored data, METRON-International Journal of Statistics, vol. LXVII, n.1, pp.75-86.
- [63] 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.
- [64] Nikulin, M.S. (1973a) Chi-square Test For Continuous Distributions with Shift and Scale Parameters, teor. Veroyatn. Primen., 18, No. 3, 559-568.
- [65] Nikulin, M. S. (1973b) On a chi-square test for continuous distributions, Theory of Probability and its Applications, 18, 638–639.

- [66] 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.
- [67] 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.
- [68] Nikulin, M.S., Gerville-Réache, L., Couallier, V. (2007) Statistiques des Essais Accélérés, Hermes : London.
- [69] Onara, A., Padgett, W.J. (2000) Accelerated test models with the inverse Gaussian distribution, Journal of Statistical Planning and Inference 89, 119-133.
- [70] O'Reilly, Federico, Rueda, Raúl (1992) Goodness of fit for the inverse Gaussian distribution, Canadian Journal of Statistics, V., Issue 4, 387–397.
- [71] 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.
- [72] 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.
- [73] 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.
- [74] Saaidia, N. (2009) Sur les applications de la famille des lois Gaussiennes inverses en fiabilité, 41èmes Journées de Statistique, Bordeaux, 19-22 mai 2009.
- [75] 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.
- [76] Saunders, S.C., Lawless, J.F. (2007) Reliability, life testing and the prediction of service lives : for engineers and Scientists, Springer New York, 310p.
- [77] Seal, H.L. (1969) Stochastic Theory of a Risk Business, New York :John Wiley & Sons Inc
- [78] Seal, H.L. (1978) From Aggregate Claim Distributions to Probability of Ruin, Astin. Bulletin, 10,47-53.
- [79] Seshadri, V., (1993) The Inverse Gaussian Distribution : A Case Study in Exponential Families, Clarendon Press :Oxford,
- [80] Seshadri, V., (1999) The inverse Gaussian Distribution : statistical theory and applications, Springer,New York.
- [81] 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.
- [82] 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.

- [83] 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.
- [84] Van Der Vaart, A.W., (1998) Asymptotic Statistics. Cambridge Series in Statistics and probabilistic Mathematics, Cambridge : Cambridge University Press.
- [85] 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.
- [86] 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.
- [87] Voinov, V., Nikulin, M. (1993) Unbiased Estimators and Their Applications, V.1 Univariate case, Dordrecht : Kluwer Academic Publishers.
- [88] Xiao, Wang, Dihua, Xu. (2010) An Inverse Gaussian Process Model for Degradation Data, Technometrics, 2010, V. 52, No 2, p 188-197.
- [89] 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.
- [90] Von Alven, W. H. (ed.) (1964) Reliability Engineering by ARINC, Englewood Cliffs, N.J.Prentice-Hall, Inc.
- [91] 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

1) Lemeshko, B. Yu., Lemeshko, S.B., Nikulin, M.S. and Saaidia, N. (2010) 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. (ISSN 0005-1179).
(www.springerlink.com/content/96q81327036j54jx/).

2) 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. (ISSN 1626-1607).
(<http://www.lsta.upmc.fr/docs/Couv.ISUP%2054fasc3.pdf>).

Articles soumis

1) Nikulin, M., saaidia, N., Seddik-Ameur, N. (2011) An Application of the AFT Model based on the Inverse Gaussian Distribution for Reliability Analysis of Redundant Systems. soumis au journal ” Probability and Statistics ”.

2) M. Nikulin, N. Saaidia, R. Tahir, ” Reliability Analysis of Redundant Systems with Unimodal Hazard Rate Functions”.
soumis au journal ” Mathematics in Engineering, Science and Aerospace”.

Chapitre dans un livre

Boris Yu. Lemeshko, Stanislav B. Lemeshko, Kseniya A. Akushkina, Mikhail S. Nikulin, and Nouredine Saaidia, ”Inverse Gaussian Model and Its Applications in Reliability and Survival Analysis”? chapter 13. In : Mathematical and Statistical Models and Methods in Reliability (Eds. Rykov, V.V. ; Balakrishnan, N. ; Nikulin, M.S.) Birkhäuser Boston, p 458. Novembre 2010.
(<http://www.springer.com/birkhauser/applied+probability+and+statistics/book/978-0-8176-4970-8>)

Articles en préparation

1) Modified Chi-squared goodness-of-fit tests for the inverse Gaussian distribution. (les annales de l'ISUP).

2) Modified Chi-squared Goodness-of-Fit Tests for the Birnbaum-Saunders Distribution. (les annales de l'ISUP).

Communications internationales

1) N. Saaidia.

Sur les applications de la famille des lois Gaussiennes inverse en Fiabilité. Bordeaux, France, le 25 - 29 Mai 2009 : 41ème Journées de Statistique. (hal.inria.fr/docs/00/38/66/80/PDF/p119.pdf).

2) 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 St Petersburg , 28 juin- 3 juillet 2009, volume 1, p : 323-334. ([http : //pws.math.spbu.ru/proceeding1.pdf](http://pws.math.spbu.ru/proceeding1.pdf))

3) Nikulin M., Saaidia N. Inverse Gaussian Family and its Applications in Reliability.

Study by simulation. St Petersburg , 28 juin- 3 juillet 2009, volume 2, p : 657-661. ([http : //pws.math.spbu.ru/proceeding2.pdf](http://pws.math.spbu.ru/proceeding2.pdf)).

4) N. Saaidia, R. Tahir, N. Seddik-Ameur.

The Inverse Gaussian Distribution and its Applications in Accelerated Life Testing when the Hazard Rate Function has -shape. Clermont Ferrant 19-21 Mai 2010. ([http : //polytech.univ - bpclermont.fr/ALT2010/ALT2010Programme.pdf](http://polytech.univ-bpclermont.fr/ALT2010/ALT2010Programme.pdf))

5) Ramzan Tahir, N. Saaidia.

Bivariate Models in Survival Analysis and Reliability. Clermont Ferrant 19-21 Mai 2010. ([http : //polytech.univ - bpclermont.fr/ALT2010/ALT2010Programme.pdf](http://polytech.univ-bpclermont.fr/ALT2010/ALT2010Programme.pdf))

5) Ramzan Tahir, N. Saaidia.

Bivariate Models in Survival Analysis and Reliability. Clermont Ferrant 19-21 Mai 2010.
(<http://polytech.univ-bpclermont.fr/ALT2010/ALT2010programme.pdf>)

6) Nikulin M., Saaidia N., Ramzan Tahir.
An Application of the AFT Model based on the Log-normal Distribution for Reliability Analysis of Redundant Systems. Qualita angers 2011, 23-25 mars.
(<http://www.rufereq.asso.fr/QUALITA/program.html>)

7) Saaidia N., Nikulin M., Ramzan Tahir.
An Application of the AFT Model Based on the Generalized Weibull Distribution for Reliability Analysis of Redundant Systems. Tunis, Tunisie, le 23 - 27 Mai 2011 : 43ème Journées de Statistique. (*Communication Acceptée*).

ملخص: في تحليل الموثوقية والبقاء، التوزيعات ذات دالة نسبة الخطر على شكل \cap ليست كثيرة جدا، وهي تشمل: توزيع معكوس جاوس، توزيع الوغاريتم الطبيعي، توزيع الوغاريتم اللوجستيكي و توزيع ويبل المعمم. في هذه الرسالة نقوم بتطوير اختبار **Rao-Robson-Nikulin** لتوزيع معكوس جاوس، ونعطي دراسة مقارنة بين توزيع معكوس جاوس والتوزيعات المذكورة أعلاه. نقوم بعدها بتنفيذ المحاكاة. النماذج ذات المتغيرات التفسيرية على أساس توزيع الغاوسي معكوس جاوس أيضا أعطيت في هذه الدراسة.

المفاتيح: توزيع معكوس جاوس، اختبار χ^2 ، اختبار Rao-Robson-Nikulin، توزيع الوغاريتم الطبيعي، توزيع الوغاريتم اللوجستيكي و توزيع ويبل المعمم، الموثوقية، تحليل البقاء، نموذج Sedyakin، نموذج AFT، النظام المكرر، مجالات الثقة، وقت العطل، المعقولة القصوى، تقدير، فئات Neyman-Pearson.

Résumé: En fiabilité, et analyse de survie, les distributions qui ont une fonction de hasard unimodale ne sont pas trop nombreuses, qu'on peut citer: la distribution Gaussienne inverse, log-normale, log-logistique et la distribution de Weibull généralisée. Dans cette thèse, nous développons le test de Rao-Robson-Nikulin pour la distribution Gaussienne inverse, et nous donnons une étude comparative entre la distribution Gaussienne inverse et les distributions citées ci-dessus, puis nous réalisons des simulations. Ensuite nous construisons les modèles avec covariables basés sur la distribution Gaussienne inverse.

Mots-clés : Distribution Gaussienne inverse, Test de Chi-deux, Test de Rao-Robson-Nikulin, Distribution log-normale, Distribution log-logistique, Distribution de Weibull généralisée, 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 hazard function are not too many, they include: the inverse Gaussian distribution, log-normal, log-logistic and power generalized Weibull distribution. In this thesis, we develop the test of Rao-Robson-Nikulin for the inverse Gaussian distribution, and we give a comparative study between the inverse Gaussian distribution and the distributions listed above, then we realize simulations. We also construct models with covariates based on the inverse Gaussian distribution.

Key-words: Inverse Gaussian distribution, Chi-square test, Test of Rao-Robson-Nikulin, Log-normal distribution, Log-logistic distribution, Power generalized Weibull distribution, Reliability, Survival analysis, Sedyakin's model, AFT model, Redundant system, Confidence interval, Failure time, Maximum likelihood, Estimation, Neyman-Pearson classes.
