Modelling and characterization of extreme fatigue risks for metallic materials

Emilie Miranda

Ecole doctorale de Sciences Mathématiques de Paris Centre

THÈSE DE DOCTORAT DE SORBONNE UNIVERSITÉ

Discipline : Mathématiques Appliquées

Spécialité : Statistique

Présentée par

Emilie Miranda

Modélisation et caractérisation des risques extrêmes en fatigue des matériaux

Dirigée par M. Michel Broniatowski

Au vu des rapports établis par
Mme Anne-Laure Fougères et M. Jan Beirlant

Soutenue publiquement le 14 octobre 2020 devant le jury composé de :

M. Broniatowski Michel	Professeur Sorbonne Université	Directeur
Mme Biret Maëva	Ingénieure Safran Aircraft Engines	Co-encadrante
Mme Fougères Anne-Laure	Professeur Université Claude Bernard Lyon 1	Rapporteur
M. Lopez Olivier	Professeur Sorbonne Université	Examinateur
M. Stummer Wolfgang	Professeur University of Erlangen–Nürnberg	Examinateur
Mme Duveau Catherine	Ingénieure Safran Aircraft Engines	Examinatrice

Remerciements

Mes remerciements s'adressent à toutes celles et ceux qui m'ont aidée et soutenue dans la réalisation de cette thèse. Leur présence a été pour moi bien plus précieuse qu'ils ne peuvent l'imaginer.

Mes premiers remerciements vont à mon directeur de thèse, M. Broniatowski, pour son implication et ses encouragements au cours de ces années. Son soutien constant et sa grande disponibilité ont permis à cette thèse d'aboutir. L'intérêt qu'il a porté à ces sujets de recherche ont été une grande source de motivation et m'ont permis de regagner confiance dans les moments plus difficiles.

Je remercie chaleureusement Anne-Laure Fougères et Jan Beirlant, rapporteurs de cette thèse, d'avoir pris le temps de lire attentivement ce mémoire. Leurs rapports bienveillants et constructifs ont permis d'enrichir ce texte. Merci également à Olivier Lopez, Wolfang Stummer, Catherine Duveau et Maëva Biret d'avoir accepté de faire partie du jury.

Je remercie l'ensemble du service Matériaux et Procédés de Safran Aircraft Engines pour m'avoir accueillie ces trois années. En particulier, un grand merci à Maëva, ma responsable de thèse, pour m'avoir accompagnée et soutenue dans mes travaux. Ses commentaires et ses idées ont largement contribué à nourrir ce manuscrit. Merci également à Catherine et à Fabien pour m'avoir accueillie dans leur service. Je pense également à Florine, Joël, Amélie, Patrick, Kévin, Nar, et tous ceux que j'ai eu la chance de côtoyer.

Merci aux doctorants du LPSM, et en particulier à Qi Min et Eric, avec qui j'ai partagé des temps de pause bienvenus autour d'un café.

Merci également à ma famille. Ma mère et mon frère, Alexis, ont été et sont toujours les deux pilliers sur lesquels je peux me reposer en temps de doute. Enfin, merci à mon grand-père. Le souvenir de son regard empreint de fierté m'a donné la motivation de poursuivre mes efforts même en son absence.

Résumé

Ces travaux ont pour fil conducteur une application industrielle en fiabilité des matériaux : on s'intéresse à la gestion des risques extrêmes associés à un endommagement en fatigue. Cette problématique industrielle soulève une série de questions qui s'articulent autour de deux axes. Le premier porte sur l'estimation d'un quantile de défaillance extrême à partir de données dichotomiques de dépassements de seuils. Un plan d'expériences séquentiel est développé afin de cibler progressivement la queue de distribution et d'échantillonner sous des distributions tronquées, sur le modèle du Splitting. Des modèles de type GEV et Weibull sont considérés et estimés séquentiellement à travers une procédure de maximum de vraisemblance adaptée aux données binaires.
Le deuxième axe de recherche concerne le développement d'outils méthodologiques permettant de déterminer la modélisation de la durée de vie la plus adaptée aux données de fatigue. Dans ce cadre, une première méthode de test d'hypothèses composites sur des données affectées par un bruit additif est proposée. La statistique de test est construite à partir d'indicateurs de divergence et généralise le test du rapport de vraisemblance. La perte de puissance liée à la présence de données bruitées est mesurée par simulations à travers des comparaisons avec le test de Neyman Pearson sur les hypothèses les moins favorables.
Une deuxième procédure vise à tester le nombre de composantes d'un mélange dans un cadre paramétrique. La statistique du test est basée sur des estimateurs de divergences exprimées sous leur forme duale dans le cadre de modèles paramétriques. La distribution limite obtenue pour la statistique de test sous l'hypothèse nulle s'applique également aux mélanges d'un nombre quelconque de composantes $k \geq 2$.

Mots-clé : Quantiles extrêmes ; Plans d'expériences sur information binaire ; Procédures de tests ; Nombre de composantes d'un mélange ; Test d'hypothèses composites ; Estimation de divergences ; Application industrielle.

Abstract

This work is motivated by a series of questions raised by an industrial issue in material reliability; more specifically, it focuses on extreme risks associated with fatigue damage. This study is divided into two parts. The first one consists in estimating an extreme failure quantile from trials whose outcomes are reduced to indicators of whether the specimen have failed at the tested stress levels. Making use of a splitting approach, we propose a sequential design method which decomposes the target probability level into a product of probabilities of conditional events of higher order. The method consists in gradually targeting the tail of the distribution and sampling under truncated distributions. The model is GEV or Weibull, and sequential estimation of its parameters involves an improved maximum likelihood procedure for binary data.

The second axis aims at developing methodological tools to model fatigue life. To this end, we propose a first test method on composite hypotheses for data affected by additive noise. We handle the problem of maximal decrease of the power for tests on this kind of corrupted data. Comparisons of such tests are considered based on their performances with respect to the Neyman Pearson test between least favourable hypotheses. It is shown that statistics based on divergence type indicators may perform better than natural generalizations of the Likelihood Ratio Test.

The second test procedure aims at testing for the number of components of a mixture distribution in a parametric setting. The test statistic is based on divergence estimators derived through the dual form of the divergence in parametric models. We provide a standard limit distribution for the test statistic under the null hypothesis, that holds for mixtures of any number of components $k \geq 2$.

Keywords : Extreme quantiles; Design of experiments on binary information; Test procedure; Number of components of a mixture model; Composite hypothesis testing; Divergence based estimation; Industrial application.

Table des matières

Table des figures

Liste des tableaux

Chapitre 1

Introduction générale

Les travaux présentés dans ce mémoire sont issus d'une collaboration entre l'université Paris VI et l'entreprise Safran Aircraft Engines, filiale du groupe industriel Safran.

Safran Aircraft Engines conçoit, développe, produit et commercialise, seul ou en coopération, des moteurs pour avions civils et militaires et pour satellites. La société propose également une gamme de services pour l'entretien de leurs moteurs, la gestion de leurs flottes et l'optimisation de leurs opérations.

Dans un contexte extrêmement compétitif, l'entreprise cherche à se démarquer en proposant des moteurs innovants et d'une fiabilité irréprochable. La maîtrise des risques techniques auxquels un moteur peut être confronté est donc un enjeu majeur. Le constructeur est soumis à une règlementation stricte : la commercialisation des pièces produites est conditionnée au respect d'un cahier des charges exigeant concernant la caractérisation des propriétés de durée de vie et de résistance. Afin de garantir la sûreté des matériels, il est crucial de caractériser le comportement du moteur et son usure selon les différentes conditions d'utilisation et les différents environnements auxquels il peut être confronté. Ces études permettent de développer des stratégies de maintenance optimales et assurent que l'entretien et le remplacement des pièces du moteur répondent à des contraintes de sécurité, de performances, mais également à des contraintes budgétaires.

La robustesse d'une pièce est affectée par de nombreux paramètres liés aux méthodes et procédés de fabrication ainsi qu'à ses conditions d'utilisation. La conception et la spécification d'un moteur doit donc être établie en fonction de ces contraintes opérationnelles et de façon à les optimiser. Au sein de Safran Aircraft Engines, la division Matériaux et Procédés intervient en amont de la construction des moteurs et est chargée de définir des stratégies de caractérisation, de développer des expertises établissant leurs conditions d'utilisation, d'élaborer des programmes d'industrialisation, de réduction des coûts et d'assurer la qualité des pièces fabriquées.

Les travaux présentés dans le manuscrit s'inscrivent dans ces problématiques. Ils se

concentrent sur les questions de fiabilité concernant, en amont de la construction des pièces de moteurs, la caractérisation des matériaux métalliques ou composites qui sont utilisés dans la production. En particulier, une question majeure en fiabilité est la résistance des pièces à différents types de sollicitations et d'endommagement. Dans cette étude, on s'intéresse à un processus d'endommagement spécifique subi en cours de vol : l'endommagement en fatigue.

1.1 Cadre industriel : endommagement en fatigue

Le phénomène de fatigue correspond à la modification des propriétés d'un matériau sous l'effet de l'application répétée d'une charge, pouvant mener à la rupture. Les chargements sont appliqués de manière cyclique, comme représenté sur la figure 1.1, et caractérisés par :

- les niveaux de contraintes maximale et minimale, $\sigma_{\min}$ et $\sigma_{\max}$;
- le niveau moyen appliqué σ_m ;
- la contrainte alternée, définie comme la demi amplitude de la variation du chargement $\sigma_a = \frac{\sigma_{\max} - \sigma_{\min}}{2}$;
- le rapport de charge $\frac{\sigma_{\min}}{\sigma_{\max}}$;
- la température T.

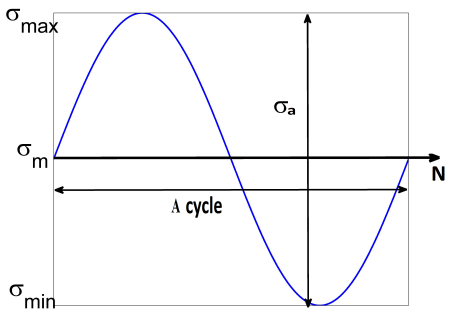


FIGURE 1.1 – Application d'un cycle d'effort sur un matériau

Dans le cas des pièces de turboréacteur, les vibrations subies durant le vol sont à l'origine de ce type d'endommagement.

Afin d'étudier la résistance d'un matériau, des études expérimentales sont menées en contrôlant les variables définies ci-dessus pour reproduire les conditions de vol. Les essais réalisés consistent à appliquer sur une éprouvette [1] des cycles de contraintes jusqu'à

1. Une éprouvette est un échantillon standardisé de la matière ou de la pièce avant usinage.

rupture ou date de fin d'essai. On mesure le nombre de cycles à rupture N (généralement censuré à droite) pour chaque niveau de sollicitation σ.

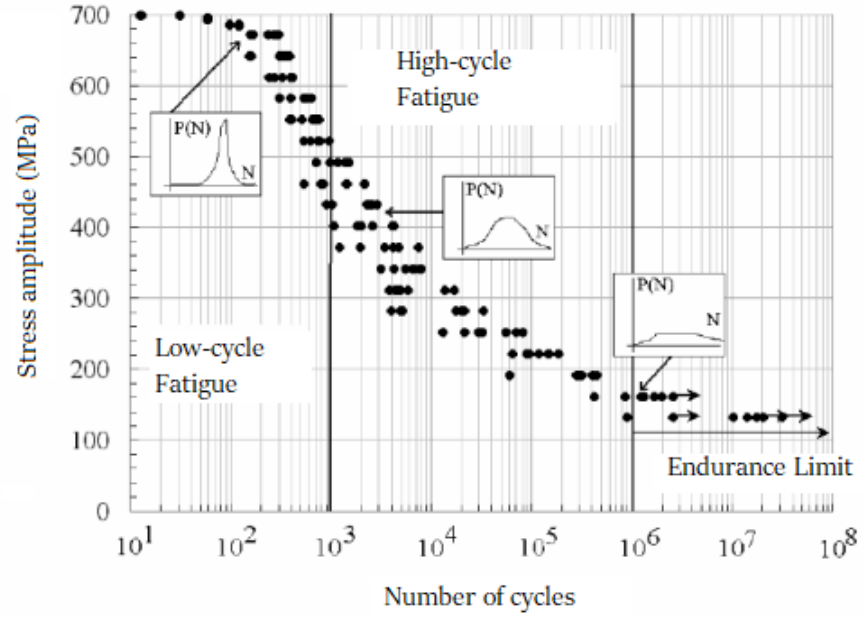


FIGURE 1.2 – Schéma d'une courbe de Wöhler

Les résultats de ces plans d'essai permettent de caractériser le comportement en fatigue d'un matériau et sont généralement représentés sous la forme d'une courbe de Wöhler, aussi appelée courbe S-N (cf. Figure 1.2). Celle-ci met en évidence trois régimes de fatigue :

- La fatigue oligocyclique ou *Low Cycle Fatigue* (LCF) est caractérisée par des durées de vie courtes, associées à des niveaux de contrainte élevés. A des niveaux voisins de la limite d'élasticité [2], la rupture est quasi-immédiate. Pour ces hauts niveaux de sollicitation, la rupture est généralement engendrée par la propagation d'une fissure amorcée en surface du matériau.
- La fatigue à grands nombres de cycles ou domaine d'endurance limitée correspond à des durées de vie inférieures à 10^6 - 10^7 cycles, liées à l'application de niveaux de contraintes environ deux fois inférieurs à la limite d'élasticité. Dans ce régime, la durée de vie augmente log-linéairement à mesure que ces derniers diminuent, jusqu'à tendre vers une asymptote horizontale. La rupture peut être engendrée par la propagation d'une fissure ayant amorcé en surface ou au cœur de l'éprouvette pour un même niveau de chargement. La co-existence des deux modalités de rupture engendre une dispersion des essais plus importante.

2. La limite d'élasticité est la contrainte à laquelle le matériau se déforme de manière irréversible et va donc rompre rapidement.

- La fatigue à très grands nombres de cycles ou *High Cycle Fatigue* (HCF), aussi appelée domaine d'endurance illimitée, intervient lorsque la durée de vie devient très grande, voire illimitée, soit parce qu'aucun défaut ne s'est amorcé, soit parce que les fissures cessent de se propager. La courbe atteint alors une asymptote horizontale et les résultats observés sont alors très dispersés. Cette asymptote est plus ou moins marquée selon les matériaux et on définit également une limite de fatigue conventionnelle qui correspond à la résistance du matériau en fatigue à 10^7 cycles. En dessous de la limite d'endurance ou de la limite de fatigue conventionnelle, la durée de vie du matériau est supposée être infinie. Ces limites servent donc pour dimensionner en fatigue les pièces mécaniques.

La tenue en fatigue d'un matériau s'étudie sous des angles différents selon le type de régime considéré. En effet, afin de caractériser la tenue en fatigue oligocyclique, l'étude se porte généralement sur la distribution de la durée de vie pour un niveau de chargement donné. En revanche, l'approche est différente concernant le régime d'endurance. Il s'agit plutôt d'estimer la distribution de la contrainte à un nombre de cycles fixés. Dans les deux cas, on peut s'intéresser essentiellement à certains indicateurs des distributions des variables d'intérêt. Cela peut être leurs moyennes ou bien leurs quantiles. Dans les travaux présentés dans la suite, l'accent est mis sur la caractérisation des comportements extrêmes des matériaux en fatigue. C'est pourquoi il s'agira d'étudier d'une part la durée de vie dite minimale, c'est-à-dire le quantile d'ordre 0.1% de la durée de vie en fatigue et, de l'autre, pour le régime à très grands nombres de cycles, la contrainte admissible minimale correspondant à un quantile de défaillance à 0.1% à un nombre de l'ordre de 10^7 cycles.

La caractérisation de ces deux quantités (durée de vie minimale et contrainte admissible) est utilisée pour dimensionner la pièce, en fixer le prix, en fournir une durée d'utilisation ainsi que pour en contrôler la qualité.

1.2 Contributions

Les travaux présentés dans ce manuscrit s'articulent autour des deux axes exposés ci-dessus, à savoir :

- la modélisation et l'estimation des risques minimaux en fatigue HCF, correspondant à l'estimation d'un quantile de défaillance extrême à une durée de vie n_0 grande fixée ;
- la recherche d'outils méthodologiques permettant de modéliser la durée de vie à un niveau de contrainte fixé en fatigue oligocyclique, ce qui constitue un prérequis

nécessaire à l'estimation de la durée de vie minimale.

La partie I se concentre sur la caractérisation de la résistance de matériaux métalliques ou composites en fatigue à très grands nombres de cycles. L'enjeu industriel consiste à proposer un plan d'essai couplé à une méthode d'estimation afin d'estimer la contrainte admissible minimale à un nombre de cycles n_0 fixé. En effet, la méthodologie existante ne permet pas d'estimer un quantile extrême et est plutôt conçue pour cibler la contrainte admissible moyenne. La contrainte minimale est dont pour l'instant obtenue en imposant un abattement forfaitaire sur la contrainte admissible moyenne.
D'un point de vue théorique, la problématique consiste à proposer un plan d'expériences en vue d'estimer un quantile extrême à partir d'un échantillon d'observations ne donnant qu'une information binaire (de type dépassement de seuil) sur la variable d'intérêt. L'absence de méthodologie adaptée à cette question dans le cadre de données dichotomiques a conduit au développement d'une procédure originale mettant en relation les méthodes d'échantillonnage préférentiel (ici, le splitting) avec les résultats en valeurs extrêmes sur les lois limites des probabilités de dépassements de seuils.
La méthodologie étudiée a été publiée dans les actes du congrès dans la base CNRS I-revues (2018 [43]) et notamment présentée au Congrès Lambda Mu 21 et à la conférence internationale EVA 2019.

La partie II porte sur la modélisation des courbes S-N en fatigue oligocyclique. Elle s'inscrit dans la continuité des travaux de R. Fouchereau (2014 [35]) et vise à proposer une nouvelle modélisation de la distribution de la durée de vie à un niveau de contrainte donné. L'objectif est de mettre en place un modèle générique s'appuyant sur les résultats théoriques en mécanique de la rupture et applicable à tous types de matériaux (métalliques et composites). Ce changement d'approche doit également permettre de modéliser et estimer plus efficacement les durées de vie minimales. Les travaux réalisés dans cette partie visent à fournir les outils méthodologiques afin de sélectionner la modélisation la plus adaptée aux données de fatigue, à travers deux procédures de tests statistiques.

La première, présentée dans le chapitre 4, permet de tester la loi d'une variable affectée d'un bruit additif. Il s'agit d'un test d'hypothèses composites correspondant à une généralisation du test du rapport de vraisemblance à partir de statistiques de type divergence. Présentée dans un cadre plus général de métrologie, elle est facilement adaptable au cas des données en fatigue et fournit un outil puissant d'aide à la décision. Ces travaux sont le fruit d'un travail joint avec M. Broniatowski, J. Jureckova et A. Kumar et ont été publiés dans la revue *entropy* (Broniatowski et al. 2019 [10]).

La seconde procédure est présentée dans le chapitre 5 et porte sur le test du nombre de composantes d'un mélange. Les composantes sont supposées appartenir à des familles

paramétriques pouvant être ou non identiques et de paramètres inconnus. Ces travaux sont une extension de Broniatowski et Keziou (2006 [12]). La statistique de test proposée est fondée sur des estimateurs de divergences sous leur forme duale. Un article a été publié en collaboration avec M. Broniatowski et W. Stummer dans les actes de la conférence Geometric Science of Information (Broniatowski et al. 2019 [14]).

Première partie

A sequential design for the estimation of minimal allowable stress in material fatigue

Chapitre 2

A sequential design for extreme quantiles estimation under binary sampling

2.1 Objectives

2.1.1 Theoretical challenge

Consider a non negative random variable X with distribution function G. Let $X_1, .., X_n$ be n independent copies of X. The aim of this paper is to estimate $q_{1-\alpha}$, the $(1-\alpha)$-quantile of G when α is much smaller than $1/n$. We therefore aim at the estimation of extreme quantiles. This question has been handled by various authors, and we will review their results later. The approach which we develop is quite different since we do not assume that the X_i's can be observed. For any threshold x, we define the r.v.

$$Y = \begin{cases} 1 \text{ if } X \leq x \\ 0 \text{ if } X > x \end{cases}$$

which therefore has a Bernoulli distribution with parameter $G(x)$. We do choose the threshold x, however we do not observe X, but merely Y. Therefore any inference on G suffers from a severe loss of information. This kind of setting is common in industrial statistics: When exploring the strength of a material, or of a bundle, we may set a constraint x, and observe whether the bundle breaks or not when subjected at this level of constraint.

In the following, we will denote R the resistance of this material, we observe Y. Inference on G can be performed for large n making use of many thresholds x. Unfortunately such a procedure will not be of any help for extreme quantiles. To address this issue, we will

consider a design of experiment enabling to progressively characterize the tail of the distribution by sampling at each step in a more extreme region of the density. It will thus be assumed in the following that we are able to observe Y not only when R follows G but also when R follows the conditional distribution of R given $\{R > x\}$. In such a case we will be able to estimate $q_{1-\alpha}$ even when $\alpha < 1/n$ where n designates the total number of trials. In material sciences, this amounts to consider trials based on artificially modified materials. When aiming at estimation of extreme upper quantiles, this amounts to strengthen the material. We would consider a family of increasing thresholds $x_1, .., x_m$ and for each of them realize $K_1, .., K_m$ trials. Each block of iid realizations Y's is therefore a function of the corresponding unobserved R's with distribution G conditioned upon $\{R > x_l\}$, $1 \leq l \leq m$. design which allows for the estimation of extreme quantiles.

The present setting is therefore quite different from that usually considered for similar problems under complete information. As sketched above it is specifically suited for industrial statistics and reliability studies in materials science.

From a strictly statistical standpoint, the question described above can be solved by considering that the distribution G is of some special form, namely that the conditional distribution of R given $\{R > x\}$ has a functional form which differs from that of G only through some changes of the parameters. In this case, simulation under these conditional distributions can be performed for adaptive choice of the thresholds x_l's, substituting the above sequence of trials. This sequential procedure allows to estimate iteratively the initial parameters of G and to obtain $q_{1-\alpha}$ combining corresponding quantiles of the conditional distributions above thresholds, a method named splitting. In this method, we will choose sequentially the x_l's in a way that $q_{1-\alpha}$ will be obtained easily from the last distribution of x conditioned upon $\{R > x_m\}$.

In safety issues or in pharmaceutical control, the focus is usually set on the control of minimal risks and therefore on the behavior of a variable of interest (strength, maximum tolerated dose) for small or even very small levels. In these settings the above considerations turn to be equivalently stated through a clear change of variable, considering the inverse of the variable of interest. In the following and to make this approach more intuitive, we choose as the main thread of this study an example in material fatigue. We look at a safety property, namely thresholds x which specify very rare events, typically failures under very small solicitations.

As stated above, the problem at hand in this study is the estimation of very small quantiles. Since classical techniques in risk theory, for example the modelling of extreme risks and exceedances over thresholds by the Generalized Pareto Distribution, pertain to large quantiles estimation, we will reduce this question to the more common setting. Denoting R the variable of interest and $\widetilde{R} := 1/R$, then obviously, for $x > 0$, $\{R < x\}$ is equivalent to

$\{\widetilde{R} > u\}$ with $u = 1/x$. In this paper we will therefore make use of this simple duality, stating formulas for R, starting with classical results pertaining to $\widetilde{R}$ when necessary. Note that when q_α designates the $\alpha-$quantile of R and respectively $\widetilde{q}_{1-\alpha}$ the $(1-\alpha)-$quantile of $\widetilde{R}$, it holds $q_\alpha = 1/\widetilde{q}_{1-\alpha}$.Those notations may be a bit cumbersome; however they result in a more familiar framework.
In this framework, we are focusing on extreme minimal risk. The critical quantities that are used to characterize minimal risk linked to fatigue damage are failure quantiles, called in this framework allowable stresses at a given number of cycles and for a fixed level of probability. Those quantiles are of great importance since they intervene in decisions to dimension engine pieces, in pricing decisions as well as in maintenance policies.
This chapter is organized as follows. Paragraph 2.1.2 formalizes the problem in the framework of the industrial application of Safran Aircraft Engines. In Section 2.2, a short survey of extreme quantiles estimation and of existing designs of experiment are studied as well as their applicability to extreme quantiles estimation. Then, a new procedure is proposed in Section 2.3 and elaborated for a Generalized Pareto model. An estimation procedure is detailed and evaluated in Section 2.4. An alternative Weibull model for the design proposed is also presented in Section 2.5. Lastly, Sections 2.6 and 2.7 provide a few ideas discussing model selection and behavior under misspecification as well as hints about extensions of the models studied beforehand.

2.1.2 Formalization of the industrial problem

The aim of this study is to propose a new design method for the characterization of allowable stress in very high cycle fatigue (HCF), for a very low risk α of order 10^{-3}. We are willing to obtain a precise estimation method of the $\alpha-$failure quantile based on a minimal number of trials.
Denote N the lifetime of a material in terms of number of cycles to failure and S the stress amplitude of the loading, in MPa. Let n_0 be the targeted time span of order $10^6 - 10^7$.
The allowable stress s_α at n_0 cycles and level of probability $\alpha = 10^{-3}$ is the level of stress that guarantee that the risk of failure before n_0 does not exceed α and is defined by:

$$s_\alpha = \sup\{s : \mathbb{P}_s(N \leq n_0) \leq \alpha\} \tag{2.1}$$

where $\mathbb{P}_s(N \leq n_0) = \mathbb{P}(N \leq n_0 | S = s)$.

We will now introduce a positive r.v. $R = R_{n_0}$ modelling the resistance of the material at n_0 cycles and homogeneous to the stress. R is the variable of interest in this study and

its distribution is defined with respect to the conditional lifetime distribution by:

$$\mathbb{P}(R \leq s) = \mathbb{P}_s(N \leq n_0) \tag{2.2}$$

Thus, the allowable stress can be rewritten as the $\alpha-$quantile of the distribution of R.

$$s_\alpha = \sup\{s : \mathbb{P}(R \leq s) \leq \alpha\} \tag{2.3}$$

However, R is not directly observed. Indeed, the usable data collected at the end of a test campaign consists in couples of censored fatigue life - stress levels $(\min(N, n_0), s)$ where s is part of the design of the experiment. The relevant information that can be drawn from those observations to characterize R is restricted to indicators of whether or not the tested specimen has failed at s before n_0. Therefore, the relevant observations obtained through a campaign of n trials are formed by a sample of variables $Y_1, ..., Y_n$ with for $1 \leq i \leq n$,

$$Y_i = \begin{cases} 1 \text{ if } R_i \leq s_i \\ 0 \text{ if } R_i > s_i \end{cases}$$

where s_i is the stress applied on specimen i.

Note that the number of observations is constrained by industrial and financial considerations; thus α is way lower than $1/n$ and we are considering a quantile lying outside the sample range.

While we motivate this paper with the above industrial application, note that this kind of problem is of interest in other domains, such as broader reliability issues or medical trials through the estimation of the maximal dose of tolerated toxicity for a given drug.

2.2 Extreme quantile estimation, a short survey

As seen above estimating the minimal admissible constraint raises two issues: on one hand the estimation of an extreme quantile, and on the other hand the need to proceed to inference based on exceedances under thresholds. We present a short exposition of these two areas, keeping in mind that the literature on extreme quantile estimation deals with complete data or data under right censoring.

2.2.1 Extreme quantiles estimation methods

Extreme quantile estimation in the univariate setting is widely covered in the literature when the variable of interest X is either completely or partially observed.

The usual framework is to study the $(1-\alpha)-$quantile of a r.v X, denoted $x_{1-\alpha}$, with very small α.

The most classical case corresponds to the setting where $x_{1-\alpha}$ is drawn from a n sample of observations $X_1,\dots X_n$. We can distinguish estimation of high quantiles, where $x_{1-\alpha}$ lies inside the sample range, see Weissman 1978 [56] and Dekkers and al. 1989 [25], and the estimation of extreme quantiles outside the boundary of the sample, see for instance De Haan and Rootzén 1993 [24]. It is assumed that X belongs to the domain of attraction of an extreme value distribution. The tail index of the latter is then estimated through maximum likelihood (Weissman 1978 [56]) or through an extension of Hill's estimator (see the moment estimator by Dekkers and al. 1989 [25]). Lastly, the estimator of the quantile is deduced from the inverse function of the distribution of the k largest observations. Note that all the above references assume that the distribution has a Pareto tail. An alternative modelling has been proposed by De Valk 2016 [26] and De Valk and Cai 2018 [27] and consists in assuming a Weibull type tail, which enables to release some second order hypotheses on the tail. This last work deals with the estimation of extreme quantiles lying way outside the sample range and will be used as a benchmark method in the following sections.

Recent studies have also tackled the issue of censoring. For instance, Beirlant and al. 2007 [9] and Einmahl and al. 2008 [33] proposed a generalization of the peak-over-threshold method when the data are subjected to random right censoring and an estimator for extreme quantiles. The idea is to consider a consistent estimator of the tail index on the censored data and divide it by the proportion of censored observations in the tail. Worms and Worms 2014 [57] studied estimators of the extremal index based on Kaplan Meier integration and censored regression.

However the literature does not cover the case of complete truncation, i.e when only exceedances over given thresholds are observed. Indeed, all of the above are based on estimations of the tail index over weighed sums of the higher order statistics of the sample, which are not available in the problem of interest in this study. Classical estimation methods of extreme quantiles are thus not suited to the present issue.

In the following, we study designs of experiment at use in industrial contexts and their possible application to extreme quantiles estimation.

2.2.2 Sequential design based on dichotomous data

In this section we review two standard methods in the industry and in biostatistics, which are the closest to our purpose. Up to our knowledge, no technique specifically addresses inference for extreme quantiles in this setting.

We address the estimation of small quantiles, hence the events of interest are of the form

$\{R < s\}$ and the quantile is q_α for small α.
The first method is the *staircase*, which is the present tool used to characterize material fatigue strength.

The second one is the *Continual Reassessment Method (CRM)* which is adapted for assessing the admissible toxicity level of a drug in Phase 1 clinical trials.
Both methods rely on a parametric model for the distribution of the strength variable R. We have considered two specifications, which allow for simple comparisons of performance, and do not aim at an accurate modelling in safety.

The Staircase method

Assume that R belongs to a parametric family with parameter θ_0. Devised by Dixon and Mood (1948 [29]), this technique aims at the estimation of the parameter θ_0 through sequential search based on data of exceedances under thresholds. The procedure is as follows.

Procedure

Fix

- The initial value for the constraint, S_{ini},
- The step $\delta > 0$,
- The number of cycles n_0 to perform before concluding a trial,
- The total number of items to be tested, K.

The first item is tested at level $s_{(1)} = S_{ini}$.The second item is then tested at level $s_{(2)} = S_{ini} - \delta$ in case of failure and $s_{(2)} = S_{ini} + \delta$ otherwise. The levels at which the $K-2$ remaining specimen are to be sequentially tested are determined by the results of the previous trials: they are increased by a step δ in case of survival and decreased by δ in case of failure. The procedure is illustrated in Figure 2.1.
Note that the proper conduct of the Staircase method relies on strong assumptions on the choice of the design parameters. Firstly, S_{ini} has to be sufficiently close to the expectation of R and secondly, δ has to lay between 0.5σ and 2σ, where σ designates the standard deviation of the distribution of R.
Denote $\mathbb{P}(R \leq s) = \phi(s, \theta_0)$ and Y the variable associated to the issue of the trial: Y_i, $1 \leq i \leq K$, takes value 1 under failure and 0 otherwise. Thus $Y_i = \mathbb{1}_{N_a \leq n_0} \sim \mathscr{B}(\phi(s_i, \theta_0))$.

Relative error			
On the parameter		**On s_α**	
Mean	Std	Mean	Std
-0.252	0.178	0.4064874	0.304

TABLE 2.1 – Results obtained using the *Staircase* method through simulations under the exponential model.

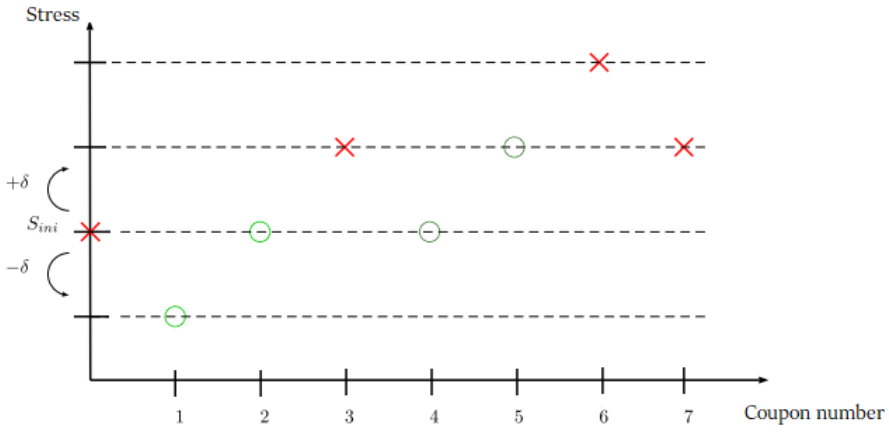


FIGURE 2.1 – Staircase procedure

Estimation

After the K trials, the parameter θ_0 is estimated through maximization of the likelihood, namely

$$\widehat{\theta} = \underset{\theta}{\operatorname{argmax}} \prod_{i=1}^{K} \phi(s_i,\theta)^{y_i} (1-\phi(s_i,\theta))^{(1-y_i)}. \tag{2.4}$$

Numerical results

The accuracy of the procedure has been evaluated on the two models presented below on a batch of 1000 replications, each with $K = 100$.

Exponential case

Let $R \sim \mathscr{E}(\lambda)$ with $\lambda = 0.2$. The input parameters are $S_{\text{ini}} = 5$ and $\delta = 15 \in \left[0.5 \times \frac{1}{\lambda^2}, 2 \times \frac{1}{\lambda^2}\right]$.

As shown in Table 2.1, the relative error pertaining to the parameter λ is roughly 25%, although the input parameters are somehow optimal for the method. The resulting relative error on the 10^{-3} quantile is 40%. Indeed the parameter λ is underestimated, which results in an overestimation of the variance $1/\lambda^2$, which induces an overestimation of the 10^{-3} quantile.

Gaussian case

Relative error					
On μ		**On σ**		**On s_α**	
Mean	Std	Mean	Std	Mean	Std
-0.059	0.034	1.544	0.903	-1.753	0.983

TABLE 2.2 – Results obtained using the *Staircase* method through simulations under the Gaussian model.

We now choose $R \sim \mathcal{N}(\mu, \sigma)$ with $\mu = 60$ and $\sigma = 10$. The value of S_{ini} is set to the expectation and $\delta = 7$ belongs to the interval $\left[\frac{\sigma}{2}, 2\sigma\right]$. The same procedure as above is performed and yields the results in Table 2.2.
The expectation of R is recovered rather accurately, whereas the estimation of the standard deviation suffers a loss in accuracy, which in turn yields a relative error of 180 % on the 10^{-3} quantile.

Drawback of the Staircase method

A major advantage of the Staircase lies in the fact that the number of trials to be performed in order to get a reasonable estimator of the mean is small. However, as shown by the simulations, this method is not adequate for the estimation of extreme quantiles. Indeed, the latter follows from an extrapolation based on estimated parameters, which furthermore may suffer of bias. Also, reparametrization of the distribution making use of the theoretical extreme quantile would not help, since the estimator would inherit of a large lack of accuracy.

The Continuous Reassesment Method (CRM)

General principle

The CRM (O'Quigley, Pepe and Fisher, 1990[45]) has been designed for clinical trials and aims at the estimation of q_α among J stress levels $s_1, ..., s_J$, when α is of order 20%.
Denote $\mathbb{P}(R \leq s) = \psi(s, \beta_0)$. The estimator of q_α is

$$s^* := \underset{s \in \{s_1, ..., s_J\}}{\operatorname{arginf}} \, |\psi(s, \beta_0) - \alpha|.$$

This optimization is performed iteratively and K trials are performed at each iteration. Start with an initial estimator $\widehat{\beta_1}$ of β_0, for example through a Bayesian choice as proposed in [45]. Define

$$s_1^* := \underset{s \in \{s_1, ..., s_J\}}{\operatorname{arginf}} \, |\psi(s, \widehat{\beta_1}) - \alpha|.$$

Every iteration follows a two-step procedure:

Step 1. Perform K trials under $\psi(.,\beta_0)$, say $R_{1,1},..,R_{1,K}$ and observe only their value under threshold, say $Y_{1,k}:=1_{R_{1,k}<s_1^*}, 1\le k\le K$.
Step i. Iteration $i\le 2$ consists in two steps :

- Firstly an estimate $\widehat{\beta_i}$ of β_0 is produced on the basis of the information beared by the trials performed in all the preceding iterations through maximum likelihood under $\psi(.,\beta_0)$ (or by maximizing the posterior distribution of the parameter).
-
$$s_i^* := \underset{s\in\{s_1,...,s_J\}}{\operatorname{arginf}} |\psi(s,\widehat{\beta_i})-\alpha|;$$
This stress level s_i^* is the one under which the next K trials $Y_{i,1},\dots,Y_{i,K}$ will be performed in the Bernoulli scheme $\mathscr{B}\left(\psi(s_i^*,\beta_0)\right)$.

The stopping rule depends on the context: it happens either when a maximum number of trials is reached or when the results are stabilized.
Note that the bayesian inference is useful in the cases where there is no diversity in the observations at some iterations of the procedure, i.e when, at a given level of test s_i^*, only failures or survivals are observed.

Application to fatigue data
The application to the estimation of the minimal allowable stress is treated in a bayesian setting. We do not directly put a prior on the parameter β_0, but rather on the probability of failure. We consider a prior information of the form: *at a given stress level s', we can expect l failures out of n trials.* Denote $\pi_{s'}$ the prior indexed on the stress level s'. $\pi_{s'}$ models the failure probability at level s' and has a Beta distribution given by

$$\pi_{s'}\sim\beta(l,n-l+1). \tag{2.5}$$

Let R follow an exponential distribution: $\forall s\ge 0, \psi(s,\beta_0)=p_s=1-\exp(-\beta_0 s)$.
It follows $\forall s,\ \beta_0=-\frac{1}{s}\log(1-p_s)$.
Define the random variable $\Lambda_s=-\frac{1}{s}\log(1-\pi_s)$. By definition of $\pi_{s'}$, $\Lambda_{s'}$ is distributed as an l-order statistic of a uniform distribution $U_{l,n}$.
The estimation procedure of the CRM is obtained as follows:
Step 1. Compute an initial estimator of the parameter

$$\Lambda_{s'}=\frac{1}{N}\sum_{i=1}^{N}-\frac{1}{s'}\log(1-\pi_{s'}^i)$$

with $\pi_{s'}^i\sim\beta(l,n-l+1), 1\le i\le N$. Define

$$s_1^* := \underset{s\in\{s_1,...,s_J\}}{\operatorname{arginf}} |(1-\exp(-\Lambda_{s'}s))-\alpha|.$$

Relative error			
On the $0.1-$quantile		**On the $10^{-3}-$quantile**	
Mean	Std	Mean	Std
0.129	0.48	-0.799	0.606

TABLE 2.3 – Results obtained through *CRM* on simulations for the exponential model

and perform K trials at level s_1^*. Denote the observations $Y_{1,k} := 1_{R_{1,k}<s_1^*}, 1 \leq k \leq K$.
Step i. At iteration i, compute the posterior distribution of the parameter:

$$\pi_{s_i}^* \sim \beta\left(l + \sum_{j=1}^{i}\sum_{k=1}^{K} Y_{j,k}\ ,\ n + (K \times i) - \left(l + \sum_{j=1}^{i}\sum_{k=1}^{K} Y_{j,k}\right) + 1\right) \tag{2.6}$$

The above distribution also corresponds an order statistic of the uniform distribution $U_{l+\sum_{j=1}^{i}\sum_{k=1}^{K} Y_{j,k}\ ,\ n+(K\times i)}$. We then obtain an estimate $\Lambda_{s_1^*}$.
The next stress level s_{i+1}^* to be tested in the procedure is then given by

$$s_{i+1}^* := \underset{s\in\{s_1,\ldots,s_J\}}{\operatorname{arginf}} |(1-\exp(-\Lambda_{s_1^*} s)) - \alpha|.$$

Numerical simulation for the CRM

Under the exponential model with parameter $\lambda = 0.2$ and through $N = 10$ iterations of the procedure, and $J = 10$, with equally distributed thresholds $s_1, .., s_J$, and performing $K = 50$ trials at each iteration, the results in Table 2.3 are obtained.
The $10^{-3}-$quantile is poorly estimated on a fairly simple model. Indeed for thresholds close to the expected quantile, nearly no failure is observed. So, for acceptable K, the method is not valid; Figure 2.2 shows the increase of accuracy with respect to K.
Both the Staircase and the CRM have the same drawback in the context of extreme quantile estimation, since the former targets the central tendency of the variable of interest and the latter aims at the estimation of quantiles of order 0.2 or so, far from the target $\alpha = 10^{-3}$. Therefore, we propose an original procedure designed for the estimation of extreme quantiles under binary information.

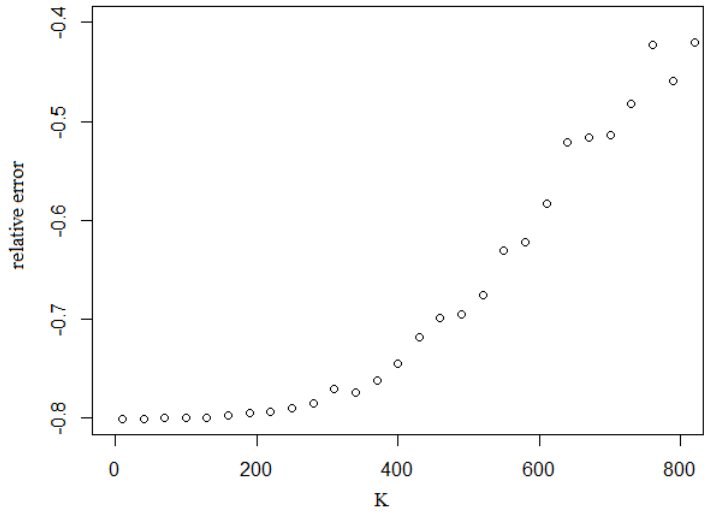


FIGURE 2.2 – Relative error on the 10^{-3}-quantile with respect to the number of trials for each stress level

2.3 A new design for the estimation of extreme quantiles

2.3.1 Splitting

The design we propose is directly inspired by the general principle of Splitting methods used in the domain of rare events simulation and introduced by Kahn and Harris (1951 [40]).

The idea is to overcome the difficulty of targeting an extreme event by decomposing the initial problems into a sequence of less complex estimation problem. This is enabled by the splitting methodology which decompose a small probability into the product of higher order probabilities.

Denote $\mathbb{P}$ the distribution of the r.v. R. The event $\{R \leq s_\alpha\}$ can be expressed as the intersection of inclusive events for $s_\alpha = s_{m+1} < s_m < s_{m-1} < \ldots < s_1$ it holds:

$$\{R \leq s_\alpha\} = \{R \leq s_{m+1}\} \subset \cdots \subset \{R \leq s_1\}.$$

It follows that

$$\mathbb{P}(R \leq s_\alpha) = \mathbb{P}(R \leq s_1) \prod_{j=1}^{m} \mathbb{P}(R \leq s_{j+1} \mid R \leq s_j) \tag{2.7}$$

The thresholds $(s_j)_{j=1,\dots,m+1}$ should be chosen such that all $\mathbb{P}(R \leq s_{j+1} \mid R \leq s_j)_{j=1,\dots,m}$ be of order $p = 0.2$ or 0.3, in such a way that $\{R \leq s_{j+1}\}$ is observed in experiments performed under the conditional distribution of R given $\{R \leq s_j\}$, and in a way which makes α recoverable by a rather small number of such probabilities $\mathbb{P}(R \leq s_{j+1} \mid R \leq s_j)$ making use of (2.7).
From the formal decomposition in (2.7), a practical experimental scheme can be deduced. Its form is given in algorithm 1.

Procedure 1 Splitting procedure

Initialization
Fix

- the number m of iterations to be performed (and of levels to be tested);
- the level of conditional probabilities p (laying between 20 and 30 %);

} such that $p^{m+1} \approx \alpha$

- the first tested level s_1 (ideally the $p-$quantile of the distribution of R);
- the number K of trials to be performed at each iteration.

First step

- K trials are performed at level s_1. The observations are the indicators of failure $Y_{1,1}, \dots, Y_{1,K}$, where $Y_{1,i} = \mathbb{1}(R_{1,i} < s_1)$ of distribution $\mathscr{B}(\mathbb{P}(R \leq s_1))$.
- Determination of s_2, $p-$quantile of the truncated distribution $R \mid R \leq s_1$.

Iteration $j = 2$ **to** m

- K trials are performed at level s_j under the truncated distribution of $R \mid R \leq s_{j-1}$ resulting to observations $Y_{j,1}, \dots, Y_{j,k} \sim \mathscr{B}\left(\mathbb{P}(R \leq s_j \mid R \leq s_{j-1})\right)$.
- Determination of s_{j+1}, the $p-$quantile of $R \mid R \leq s_j$.

The last estimated quantile s_{m+1} provides the estimate of s_α.

2.3.2 Sampling under the conditional probability

In practice batches of specimen are put under trial, each of them with a decreasing strength; this allows to target the tail of the distribution $\mathbb{P}$ iteratively.

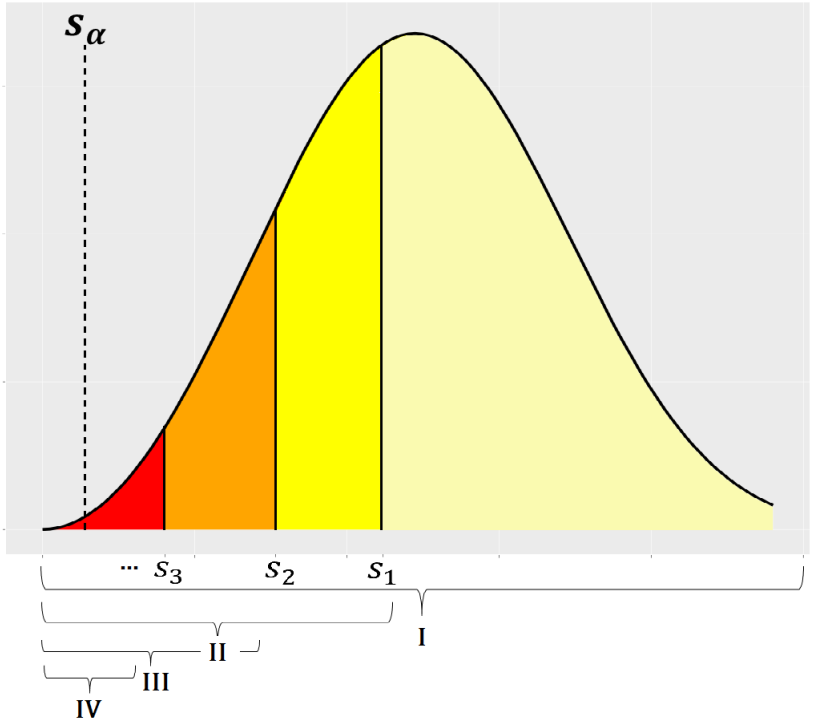


FIGURE 2.3 – Sampling under the strengh density at n_0 cycles

In other words, in the first step, points are sampled in zone (I). Then in the following step, only specimen with strength in zone II are considered, and so on. In the final step, the specimen are sampled in zone IV. At level s_m, they have a very small probability to fail before n_0 cycles under $\mathbb{P}$, however under their own law of failure, which is $\mathbb{P}(. \mid R \leq s_{m-1})$, they have a probability of failure of order 0.2.

In practice, sampling in the tail of the distribution is achieved by introducing flaws in the batches of specimens. The idea is that the strength of the material varies inversely with respect to the size of the incorporated flaws. The flaws are spherical and located inside the specimen (not on its surface). Thus, as the procedure moves on, the trials are performed on samples of materials incorporating flaws of greater diameter. This procedure is based on the hypothesis that there is a correspondence between the strength of the material with flaw of diameter θ and the truncated strength of this same material without flaw under level of stress s^*, i.e. we assume that noting R_θ the strength of the specimen with flaw of size θ, it holds that there exists s^* such that

$$\mathscr{L}(R_\theta) \approx \mathscr{L}(R \mid R \leq s^*).$$

Before launching a validation campaign for this procedure, a batch of 27 specimen has been machined including spherical defects whose sizes vary between 0 and 1.8mm (see Figure 2.4). These first trials aim at estimating the decreasing relation between mean

allowable stress and defects diameter θ. This preliminary study enabled to draw the abatement fatigue curve as a function of θ, as shown in Figure 2.5.

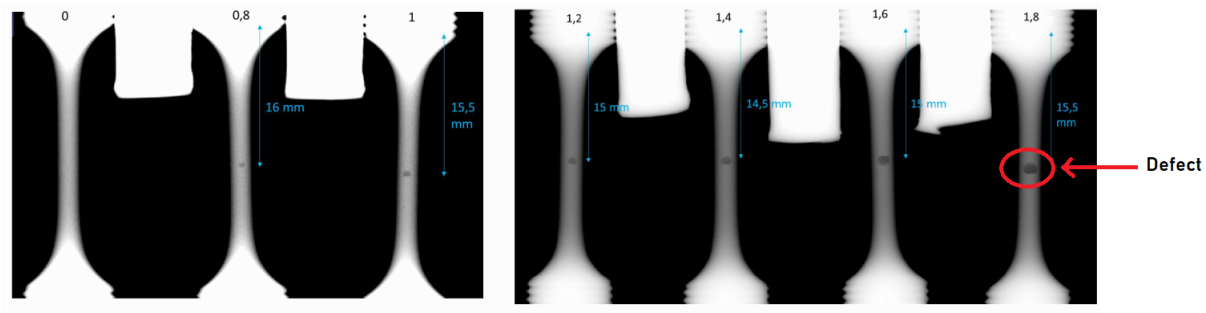


FIGURE 2.4 – Coupons incorporating spherical defects of size varying from 0 mm (on the left) to 1.8 mm (on the right)

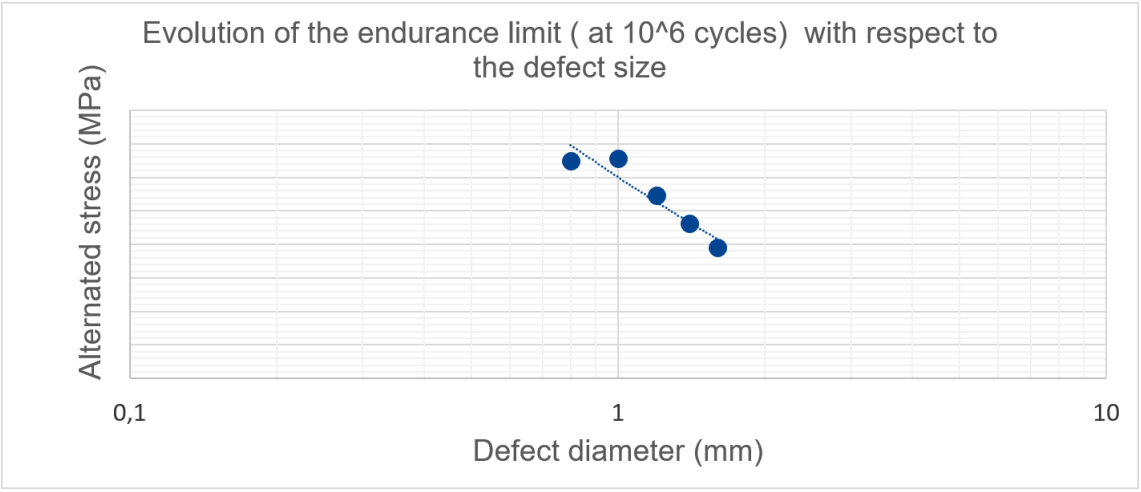


FIGURE 2.5 – Mean allowable stress with respect to the defect size

Results in Figure 2.5 will be used during the splitting procedure to select the diameter θ to be incorporated in the batch of specimens tested at the current iteration as reflecting the sub-population of material of smaller resistance.

Practically, this amounts to add at each step j of the procedure the determination of the mean or median s_{j+1}^m of the distribution of the estimated conditional distribution $R \mid R \leq s_j$ on top of its $p-$quantile. The flaw size θ_{j+1} that will be introduced in the batch of specimen tested at step $j+1$ is such that

$$\mathbb{E}(R_{\theta_{j+1}}) \approx s_{j+1}^m$$

or

$$\sup\left\{s : \mathbb{P}(R_{\theta_{j+1}} \leq s) \leq 0.5\right\} \approx s_{j+1}^m.$$

2.3.3 Modelling the distribution of the strength, Pareto model

The events under consideration have small probability under $\mathbb{P}$. By (2.7) we are led to consider the limit behavior of conditional distributions under smaller and smaller thresholds, for which we make use of classical approximations due to Balkema and de Haan (1974[4]) which stands as follows, firstly in the commonly known setting of exceedances over increasing thresholds.
Denote $\widetilde{R} := 1/R$.

Theorem 1. For $\widetilde{R}$ of distribution F belonging to the maximum domain of attraction of an extreme value distribution with tail index c, i.e. $F \in MDA(c)$, it holds that : There exists $a = a(s) > 0$, such that:

$$\lim_{s\to\infty} \sup_{0\le x<\infty} \left| \frac{1-F(x+s)}{1-F(s)} - \left(1 - G_{(c,a)}(x)\right) \right| = 0$$

where $G_{(c,a)}$ is defined through

$$G_{(c,a)}(x) = 1 - \exp\left\{ -\int_0^{\frac{x}{a}} [(1+ct)_+]^{-1}\, dt \right\}$$

with $a > 0$ and $c \in \mathbb{R}$.

The distribution G is the Generalized Pareto distribution $GPD(c,a)$ which is defined explicitly through

$$1 - G(x) = \begin{cases} (1+\frac{c}{a}x)^{-1/c} \text{ when } c \neq 0 \\ \exp(-\frac{x}{a}) \text{ when } c = 0 \end{cases}$$

where $x \ge 0$ for $c \ge 0$ and $0 \le x \le -\frac{a}{c}$ if $c < 0$.
Generalized Pareto distributions enjoy invariance through threshold conditioning, an important property for our sake. Indeed it holds, for $\widetilde{R} \sim GDP(c,a)$ and $x > s$,

$$\mathbb{P}\left(\widetilde{R} > x \mid \widetilde{R} > s\right) = \left(1 + \frac{c(x-s)}{a+cs}\right)^{-1/c} \tag{2.8}$$

We therefore state:

Proposition 2. When $\widetilde{R} \sim GPD(c,a)$ then, given $\left(\widetilde{R} > s\right)$, the r.v. $\widetilde{R} - s$ follows a $GPD(c, a+cs)$.

The GPD's are on the one hand stable under thresholding and on the other appear as the limit distribution for thresholding operations. This chain of arguments is quite usual in statistics, motivating the recourse to the normal or stable laws for additive models. This

plays in favor of GPD's as modelling the distribution of $\widetilde{R}$ for excess probability inference. Due to the lack of memory property, the exponential distribution which appears as a possible limit distribution for excess probabilities in Theorem 1 do not qualify for modelling. Moreover since we handle variables R which can approach 0 arbitrarily (i.e. unbounded $\widetilde{R}$) the parameter c is assumed positive.
Turning to the context of the minimal admissible constraint, we make use of the r.v. $R = 1/\widetilde{R}$ and proceed to the corresponding change of variable.
When $c > 0$, the distribution function of the r.v. R writes for nonnegative x:

$$F_{c,a}(x) = \left(1 + \frac{c}{ax}\right)^{-1/c}. \tag{2.9}$$

For $0 < x < u$, the conditional distribution of R given $\{R < u\}$ is

$$\mathbb{P}(R < x \mid R < u) = \left(1 - \frac{c(\frac{1}{x} - \frac{1}{u})}{a + \frac{c}{u}}\right)^{-1/c}$$

which proves that the distribution of R is stable under threshold conditioning with parameter (a_u, c) with

$$a_u = a + \frac{c}{u}. \tag{2.10}$$

In practice at each step j in the procedure the stress level s_j equals the corresponding threshold $1/\widetilde{s}_j$, a right quantile of the conditional distribution of $\widetilde{R}$ given $\{\widetilde{R} > \widetilde{s}_{j-1}\}$. Therefore the observations take the form $Y_i = \mathbb{1}_{R_i < s_{j-1}} = \mathbb{1}_{\widetilde{R}_i > \widetilde{s}_{j-1}}$, $i = 1, \ldots, K_j$.
A convenient feature of model (2.9) lies in the fact that the conditional distributions are completely determined by the initial distribution of R , therefore by a and c. The parameters a_j of the conditional distributions are determined from these initial parameters and by the corresponding stress level s_j; see (2.10).

2.3.4 Notations

The distribution function of the r.v. $\widetilde{R}$ is a $GPD(c_T, a_T)$ of distribution function $G_{(c_T,a_T)}$. Note $\overline{G}_{(c_T,a_T)} = 1 - G_{(c_T,a_T)}$.
Our proposal relies on iterations. We make use of a set of thresholds $(\widetilde{s}_1, ..., \widetilde{s}_m)$ and define for any $j \in \{1, ..., m\}$

$$\overline{G}_{(c_j,a_j)}(x - \widetilde{s}_j) = \mathbb{P}(\widetilde{R} > x \big| \widetilde{R} > \widetilde{s}_j)$$

with $c_j = c_T$ and $a_j = a_T + c_T\widetilde{s}_j$ where we used Proposition 2.
At iteration j, denote $(\widehat{c}, \widehat{a})_j$ the estimators of (c_j, a_j).Therefore $1 - G_{(\widehat{c},\widehat{a})_j}(x - \widetilde{s}_j)$ estimates $\mathbb{P}(\widetilde{R} > x \big| \widetilde{R} > \widetilde{s}_j)$. Clearly, estimators of (c_T, a_T) can be recovered from $(\widehat{c}, \widehat{a})_j$

through $\widehat{c}_T = \widehat{c}$ and $\widehat{a}_T = \widehat{a} - \widehat{c}\,\widetilde{s}_j$.

2.3.5 Sequential design for the extreme quantile estimation

Fix m and p , where m denotes the number of stress levels under which the trials will be performed, and p is such that $p^m = \alpha$.

Set a first level of stress, say s_1 large enough (i.e. $\widetilde{s}_1 = 1/s_1$ small enough) so that $p_1 = \mathbb{P}(R < s_1)$ is large enough and perform trials at this level. The optimal value of s_1 should satisfy $p_1 = p$, which cannot be secured. This choice is based on expert advice.

Turn to $\widetilde{R} := 1/R$. Estimate c_T and a_T, for the GPD (c_T, a_T) model describing $\widetilde{R}$, say $(\widehat{c}, \widehat{a})_1$, based on the observations above $\widetilde{s}_1$ (note that under s_1 the outcomes of R are easy to obtain, since the specimen is tested under medium stress).

Define

$$\widetilde{s}_2 := \sup\left\{s : \overline{G}_{(\widehat{c},\widehat{a})_1}(s - \widetilde{s}_1) < p\right\} \tag{2.11}$$

the $(1-p)-$quantile of $G_{(\widehat{c},\widehat{a})_1}$. $\widetilde{s}_2$ is the level of stress to be tested at the following iteration.

Iterating from step $j = 2$ to m, perform K trials under $G_{(c_1,a_1)}$ say $\widetilde{R}_{j,1}, .., \widetilde{R}_{j,K}$ and consider the observable variables $Y_{j,i} := 1_{\widetilde{R}_{j,i} > \widetilde{s}_j}$. Therefore the K iid replications $Y_{j,1}, .., Y_{j,K}$ follow a Bernoulli $\mathscr{B}(\overline{G}_{(c_{j-1},a_{j-1})}(\widetilde{s}_j - \widetilde{s}_{j-1}))$, where $\widetilde{s}_j$ has been determined at the previous step of the procedure. Estimate (c_j, a_j) in the resulting Bernoulli scheme, say $(\widehat{c}, \widehat{a})_j$. Then define

$$\begin{aligned} \widetilde{s}_{j+1} &:= \sup\left\{s : \overline{G}_{(\widehat{c},\widehat{a})_j}(s - \widetilde{s}_j) < p\right\} \\ &= G^{-1}_{(\widehat{c},\widehat{a})_j}(1-p) + \widetilde{s}_j, \end{aligned} \tag{2.12}$$

which is the $(1-p)-$quantile of the estimated conditional distribution of $\widetilde{R}$ given $\{\widetilde{R} > \widetilde{s}_j\}$, i.e. $G_{(\widehat{c},\widehat{a})_j}$. It is also the next level to be tested, except at iteration m, where s_{m+1} is the last estimated quantile and, by its very definition, a proxy of $\widetilde{q}_{1-\alpha}$.

In practice a conservative choice for the number of sequences of trials m is given by $m = \left\lceil \frac{log\alpha}{logp} \right\rceil - 1$, where $\lceil . \rceil$ denotes the ceiling function. This implies that the attained probability $\widetilde{\alpha}$ is less than or equal to α.

The $m+1$ stress levels $\widetilde{s}_1 < \cdots < \widetilde{s}_m < \widetilde{s}_{m+1} = \widetilde{q}_{1-\alpha}$ satisfy

$$\begin{aligned} \widetilde{\alpha} &= \overline{G}(\widetilde{s}_1) \prod_{j=1}^{m} \overline{G}_{(\widehat{c},\widehat{a})_j}(\widetilde{s}_{j+1} - \widetilde{s}_j) \\ &= p_1 p^m \end{aligned}$$

Although quite simple in its definition, this method bears a number of drawbacks, mainly

in the definition of $(\widehat{c},\widehat{a})_j$. The next section addresses this question.

2.4 Sequential enhanced design and estimation method in the Pareto model

In this section we focus on the estimation of the parameters (c_T, a_T) in the $GPD(c_T, a_T)$ distribution of $\widetilde{R}$. One of the main difficulties lies in the fact that the available information does not consist of replications of the r.v. $\widetilde{R}$ under the current conditional distribution $G_{(c_j,a_j)}$ of $\widetilde{R}$ given $\left(\widetilde{R} > \widetilde{s_j}\right)$ but merely on very downgraded functions of those.
At step j, we are given $G_{(\widehat{c},\widehat{a})_{j-1}}$ and define $\widetilde{s}_j$ as its $(1-p)$−quantile. Simulating K r.v. $\widetilde{R}_{j,i}$ with distribution $G_{(c_{j-1},a_{j-1})}$, the observable outcomes are the Bernoulli (p) r.v.'s $Y_{j,i} := 1_{\widetilde{R}_{j,i}>\widetilde{s}_j}$. This loss of information with respect to the $\widetilde{R}_{j,i}$'s makes the estimation step for the coefficients $(\widehat{c},\widehat{a})_j$ quite complex; indeed $(\widehat{c},\widehat{a})_j$ is obtained through the $Y_{j,i}$'s, $1 \le i \le K$.

2.4.1 Estimation procedure based on classical optimization criteria

The first approach consists in analyzing the results obtained through standard Maximum Likelihood Estimation of the parameters $(\widehat{c},\widehat{a})_j$ at each step j of the procedure. In this case, the estimation at each step $j \ge 1$ is made only on the basis of current data, i.e the current observations $Y_{j,1},\dots,Y_{j,K}$ obtained under the Bernoulli scheme.
At the first iteration, the estimates are given by:

$$(\widehat{c},\widehat{a})_T = \underset{(c,a)}{\operatorname{argmax}} \sum_{i=1}^{K} Y_{1,i} \log\left(\bar{G}_{(c,a)}(\widetilde{s}_1)\right) + (1-Y_{1,i}) \log\left(G_{(c,a)}(\widetilde{s}_1)\right)$$

Given the estimates of the parameters of the distribution $G_{(c_T,a_T)}$, it follows the parameters of the conditional law of $\widetilde{R}$ given $\widetilde{R} > \widetilde{s}_1$:

$$(\widehat{c},\widehat{a})_1 = (\widehat{c}_T, \widehat{a}_T + \widehat{c}_T\widetilde{s}_1)$$

and the following level to be tested is obtained through (2.11). Similarly, for steps $j > 1$, the parameters of the current conditional distribution of $\widetilde{R}$ given $\widetilde{R} > \widetilde{s}_{j-1}$ are estimated through Maximum Likelihood based on the observations $Y_{j,1},\dots,Y_{j,K}$ resulting from trials at level $\widetilde{s}_j$:

$$(\widehat{c},\widehat{a})_{j-1} = \underset{(c,a)}{\operatorname{argmax}} \sum_{i=1}^{K} Y_{j,i} \log\left(\bar{G}_{(c,a)}(\widetilde{s}_j)\right) + (1-Y_{j,i}) \log\left(G_{(c,a)}(\widetilde{s}_j)\right)$$

Minimum	Q25	Q50	Mean	Q75	Maximum
67.07	226.50	327.40	441.60	498.90	10 320.00

TABLE 2.4 – Estimation of the $(1-\alpha)-$quantile, $\widetilde{s}_\alpha = 469.103$, through procedure 2.3.5 with $K = 50$

	$\widetilde{s}_m$ for $K = 30$		$\widetilde{s}_m$ for $K = 50$	
$\widetilde{s}_\alpha$	Mean	Std	Mean	Std
469.103	1 276.00	12 576.98	441.643	562.757

TABLE 2.5 – Estimation of the $(1-\alpha)-$quantile, $\widetilde{s}_\alpha = 469.103$, through procedure 2.3.5 for different values of K

Estimates for the next conditional distribution and next level of stress are obtained as stated above.

The quantile $\widetilde{q}_{1-\alpha}$ obtained with this Maximum Likelihood Estimation based procedure is loosely estimated for small α : As measured on 1000 simulation runs, large standard deviation of $\widehat{\widetilde{q}}_{1-\alpha}$ is due to poor estimation of the iterative parameters $(\widehat{c}, \widehat{a})_{j+1}$. We have simulated $n = 200$ realizations of r.v.'s Y_i with common Bernoulli distribution with parameter $\overline{G}_{(c_T,a_T)}(\widetilde{s}_1)$. Figure 2.6 shows the log likelihood function of this sample as the parameter of the Bernoulli $\overline{G}_{(c',a')}(\widetilde{s}_0)$ varies according to (c', a'). As expected this function is nearly flat in a very large range of (c', a').
This explains the poor results in Table 2.5 obtained through the Splitting procedure when the parameters at each step are estimated by maximum likelihood, especially in terms of dispersion of the estimations. Moreover, the accuracy of the estimator of $\widetilde{q}_{1-\alpha}$ quickly decreases with the number K of replications $Y_{j,i}$, $1 \le i \le K$.
Changing the estimation criterion by some alternative method does not improve significantly; Figure 2.7 shows the distribution of the resulting estimators of $\widetilde{q}_{1-\alpha}$ for various estimation methods (minimum Kullback Leibler, minimum Hellinger and minimum L1 distances - see their definitions in Appendix 2.9.1) of (c_T, a_T).
This motivates the need for an enhanced estimation procedure.

2.4.2 An enhanced sequential criterion for estimation

We consider an additional criterion which makes use of the iterative nature of the procedure. We will impose some control on the stability of the estimators of the conditional quantiles through the sequential procedure. At iteration 1, the estimation is performed using only Maximum Likelihood and remains as described in 2.4.1. Note that in practice, some additional information may also be available, related to the mean or median

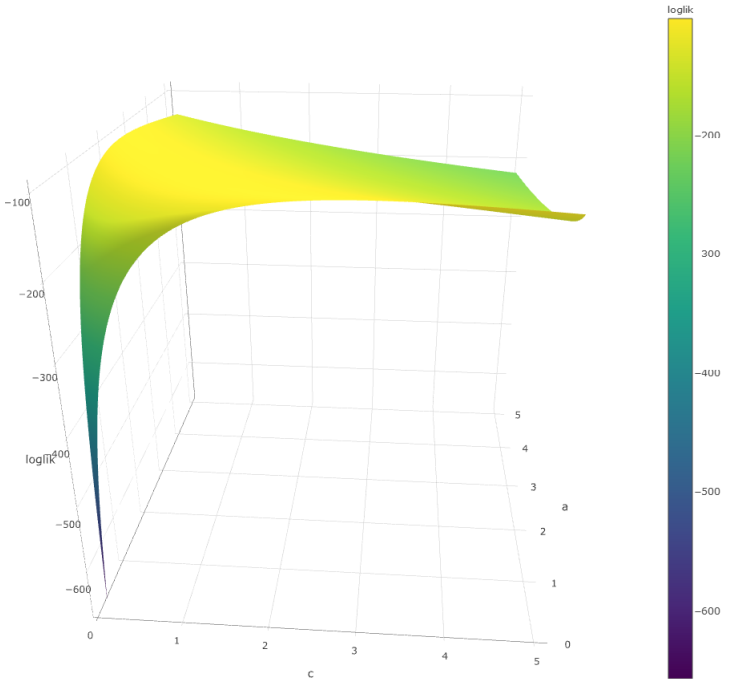


FIGURE 2.6 – Log-likelihood of the Pareto model with binary data

of the distribution if the studied material has already been the subject of some testing campaign.

At iteration $j-1$, the sample $Y_{j-1,i}$, $1 \leq i \leq K$ has been generated under $G_{(c_{j-2},a_{j-2})}$ and provides an estimate of p through

$$\widehat{p}_{j-1} := \frac{1}{K} \sum_{i=1}^{K} Y_{j-1,i}. \tag{2.13}$$

The above $\widehat{p}_{j-1}$ estimates $\mathbb{P}\left(\widetilde{R} > \widetilde{s}_{j-1} \mid \widetilde{R} > \widetilde{s}_{j-2}\right)$ conditionally on $\widetilde{s}_{j-1}$ and $\widetilde{s}_{j-2}$. We write this latter expression $\mathbb{P}\left(\widetilde{R} > \widetilde{s}_{j-1} \mid \widetilde{R} > \widetilde{s}_{j-2}\right)$ as a function of the parameters obtained at iteration j, namely $(\widehat{c}, \widehat{a})_j$.The above r.v's $Y_{j-1,i}$ stem from variables $\widetilde{R}_{j-1,i}$ greater than $\widetilde{s}_{j-2}$. At step j, estimate then $\mathbb{P}\left(\widetilde{R} > \widetilde{s}_{j-1} \mid \widetilde{R} > \widetilde{s}_{j-2}\right)$ making use of $G_{(\widehat{c},\widehat{a})_j}$. Denote $G_{(\widehat{c},\widehat{a})^j_{j-2}}$ the updated estimation of $G_{(c_{j-2},a_{j-2})}$, defined by $\widehat{c}^j_{j-2} = \widehat{c}_j$ and $\widehat{a}^j_{j-2} = \widehat{a}_j + \widehat{c}_j\left(\widetilde{s}_{j-2} - \widetilde{s}_j\right)$. This backward estimator writes

$$\frac{\overline{G}_{(\widehat{c},\widehat{a})^j_{j-2}}(\widetilde{s}_{j-1})}{\overline{G}_{(\widehat{c},\widehat{a})^j_{j-2}}(\widetilde{s}_{j-2})} = 1 - G_{(\widehat{c},\widehat{a})^j_{j-2}}(\widetilde{s}_{j-1} - \widetilde{s}_{j-2}).$$

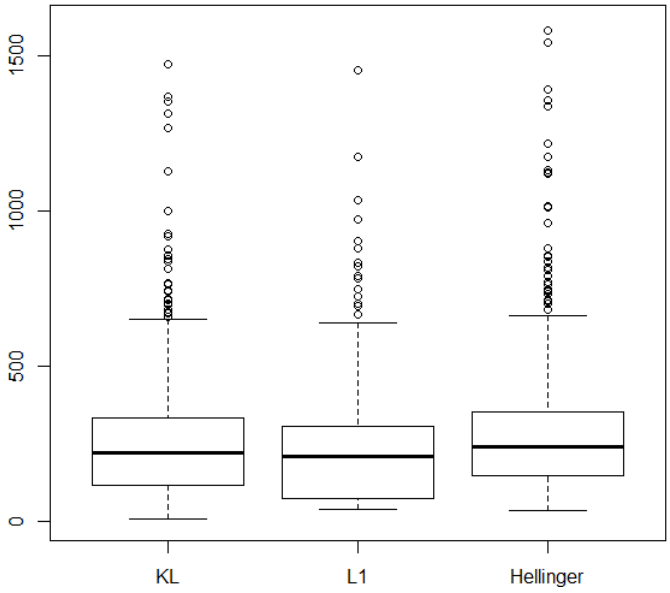


FIGURE 2.7 – Estimations of the $\alpha-$quantile based on the Kullback-Leibler, L1 distance and Hellinger distance criterion

The distance

$$\left|\left(\overline{G}_{(\widehat{c},\widehat{a})_{j-2}^{j}}(\widetilde{s}_{j-1}-\widetilde{s}_{j-2})\right)-\widehat{p}_{j-1}\right| \tag{2.14}$$

should be small, since both $\overline{G}_{(\widehat{c},\widehat{a})_{j-2}^{j}}(\widetilde{s}_{j-1}-\widetilde{s}_{j-2})$ and $\widehat{p}_{j-1}$ should approximate p.
Consider the distance between quantiles

$$\left|(\widetilde{s}_{j-1}-\widetilde{s}_{j-2})-G^{-1}_{(\widehat{c},\widehat{a})_{j-2}^{j}}(1-\widehat{p}_{j-1})\right|. \tag{2.15}$$

An estimate $(\widehat{c},\widehat{a})_j$ can be proposed as the minimizer of the above expression for $(\widetilde{s}_{j-1}-\widetilde{s}_{j-2})$ for all j. This backward estimation provides coherence with respect to the unknown initial distribution $G_{(c_T,a_T)}$. Would we have started with a good guess $(\widehat{c},\widehat{a})=(c_T,a_T)$ then the successive $(\widehat{c},\widehat{a})_j$, $\widetilde{s}_{j-1}$, etc...would make (2.15) small, since $\widetilde{s}_{j-1}$ (resp. $\widetilde{s}_{j-2}$) would estimate the $p-$conditional quantile of $\mathbb{P}\left(.|\widetilde{R}>\widetilde{s}_{j-2}\right)$ (resp. $\mathbb{P}\left(.|\widetilde{R}>\widetilde{s}_{j-3}\right)$).
It remains to argue on the set of plausible values where the quantity in (2.15) should be minimized.
We suggest to consider a confidence region for the parameter (c_T,a_T). With $\widehat{p}_j$ defined in (2.13) and $\gamma\in(0,1)$ define the $\gamma-$confidence region for p by

$$I_\gamma=\left[\widehat{p}_j-z_{1-\gamma/2}\sqrt{\frac{\widehat{p}_j(1-\widehat{p}_j)}{K-1}};\widehat{p}_j+z_{1-\gamma/2}\sqrt{\frac{\widehat{p}_j(1-\widehat{p}_j)}{K-1}}\right]$$

Parameters	Relative error on $\widetilde{s}_\alpha$	
	Mean	Std
$c = 0.8$, $a_0 = 1.5$ and $\widetilde{s}_\alpha = 469.103$	-0.222	0.554
$c = 1.5$, $a_0 = 1.5$ and $\widetilde{s}_\alpha = 31621.777$	-0.504	0.720
$c = 1.5$, $a_0 = 3$ and $\widetilde{s}_\alpha = 63243.550$	0.310	0.590

TABLE 2.6 – Mean and std of relative errors on the $(1-\alpha)-$quantile of GPD calculated through 400 replicas of procedure 2.4.2.

where z_τ is the $\tau-$quantile of the standard normal distribution. Define

$$\mathscr{S}_j = \left\{(c,a) : \left(1 - G_{(c,a)}(\widetilde{s}_j - \widetilde{s}_{j-1})\right) \in I_\gamma\right\}.$$

Therefore $\mathscr{S}_j$ is a plausible set for $(\widehat{c}_T, \widehat{a}_T)$.

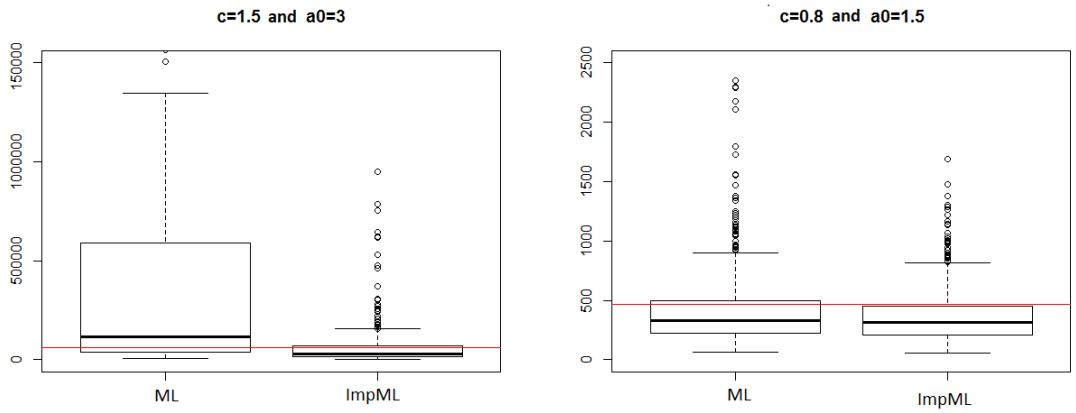
The above discussion is summarized as follows:
At iteration j, the estimator of (c_T, a_T) is a solution of the minimization problem

$$\min_{(c,a)\in\mathscr{S}_j} \left| (\widetilde{s}_{j-1} - \widetilde{s}_{j-2}) - G^{-1}_{(c,a+c\widetilde{s}_{j-2})}(1-\widehat{p}_{j-1}) \right|.$$

The optimization method used is the Safip algorithm (Biret and Broniatowski, 2016 [8], see Appendix 2.9.2).
As seen hereunder, this heuristics provides good performance.

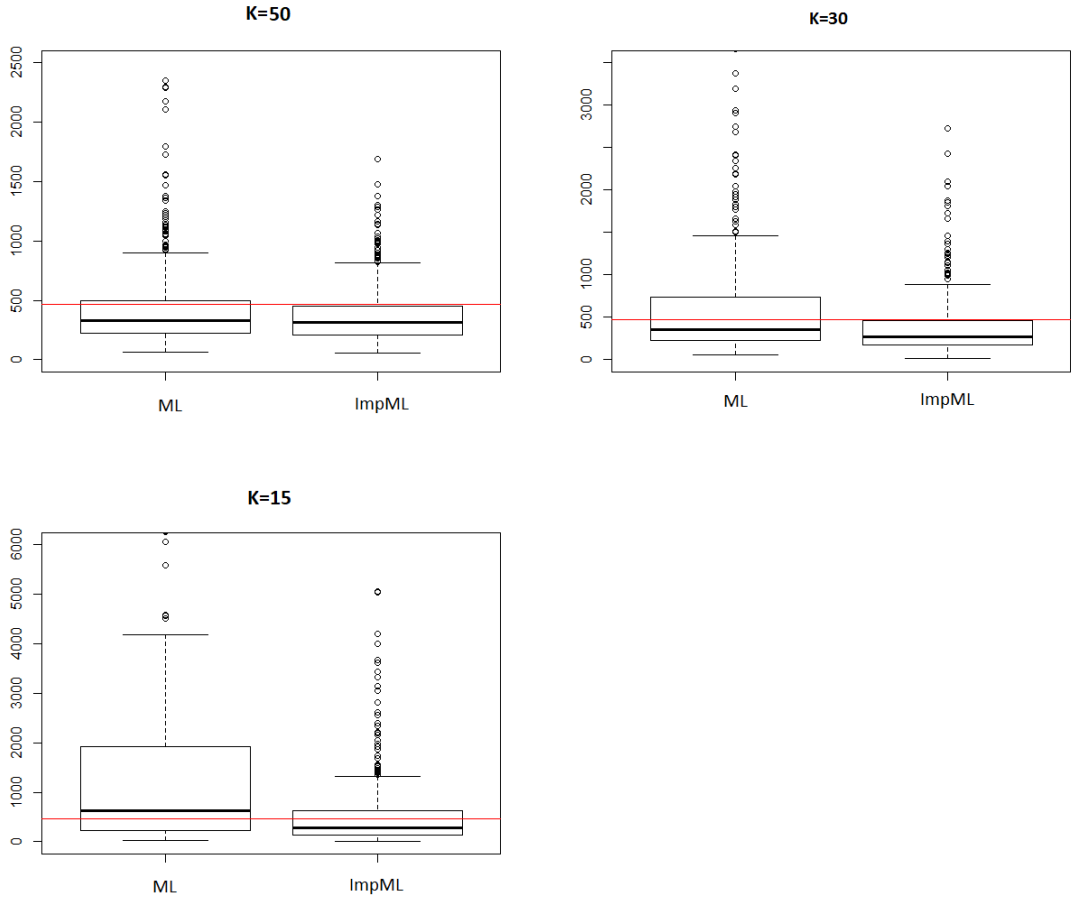
2.4.3 Simulation based numerical results

This procedure has been applied in three cases. A case considered as reference is $(c_T, a_T) = (1.5, 1.5)$; secondly the case when $(c_T, a_T) = (0.8, 1.5)$ describes a light tail with respect to the reference. Thirdly, a case $(c_T, a_T) = (1.5, 3)$ defines a distribution with same tail index as the reference, but with a larger dispersion index.
Table 2.6 shows that the estimation of $\widetilde{q}_{1-\alpha}$ deteriorates as the tail of the distribution gets heavier; also the procedure tends to underestimate $\widetilde{q}_{1-\alpha}$.
Despite these drawbacks, we observe an improvement with respect to the simple Maximum Likelihood estimation; this is even more clear, when the tail of the distribution is heavy. Also, in contrast with the ML estimation, the sensitivity with respect to the number K of replications at each of the iterations plays in favor of this new method: As K decreases, the gain with respect to Maximum Likelihood estimation increases notably, see Figure 2.9.



The red line stands stands for the real value of s_α

FIGURE 2.8 – Estimations of the $(1-\alpha)-$quantile of two GPD obtained by Maximum Likelihood and by the improved Maximum Likelihood method



The red line stands stands for the real value of s_α

FIGURE 2.9 – Estimations of the $(1-\alpha)-$quantile of a $GPD(0.8, 1.5)$ obtained by Maximum Likelihood and by the improved Maximum Likelihood method for different values of K.

2.4.4 Performance of the sequential estimation

As stated in chapter 2.2, there is to our knowledge no method dealing with similar question available in the literature. Therefore we compare the results of our method, based on observed exceedances over thresholds, with the results that could be obtained by classical extreme quantiles estimation methods assuming we have complete data at our disposal; those may be seen as benchmarks for an upper bound of the performance of our method.

Estimation of an extreme quantile based on complete data, de Valk's estimator

In order to provide an upper bound for the performance of the estimator, we make use of the estimator proposed by De Valk and Cai (2016). This work aims at the estimation of a quantile of order $p_n \in [n^{-\tau_1}; n^{-\tau_2}]$, with $\tau_2 > \tau_1 > 1$, where n is the sample size. This question is in accordance with the industrial context which motivated the present paper. De Valk's proposal is a modified Hill estimator adapted to log-Weibull tailed models. De Valk's estimator is consistent, asymptotically normally distributed, but is biased for finite sample size.We briefly recall some of the hypotheses which set the context of de Valk's approach.
Let $X_1, \ldots, X_n$ be n iid r.v's with distribution F, and denote $X_{k:n}$ the $k-$ order statistics. A tail regularity assumption is needed in order to estimate a quantile with order greater than $1 - 1/n$.
Denote $U(t) = F^{-1}(1 - 1/t)$, and let the function q be defined by

$$q(y) = U(e^y) = F^{-1}\left(1 - e^{-y}\right)$$

for $y > 0$.
Assume that

$$\lim_{y \to \infty} \frac{\log q(y\lambda) - \log q(y)}{g(y)} = h_\theta(\lambda) \quad \lambda > 0 \tag{2.16}$$

where g is a regularly varying function and

$$h_\theta(\lambda) = \begin{cases} \frac{\lambda^\theta - 1}{\theta} \text{ if } \theta \neq 0 \\ \log \lambda \text{ if } \theta = 0 \end{cases}$$

de Valk writes condition 2.16 as $\log q \in ERV_\theta(g)$.

Remark : Despite its naming of log-Generalized tails, this condition also holds for Pareto tailed distributions, as can be checked, providing $\theta = 1$.
We now introduce de Valk's extreme quantile estimator.

Let

$$\vartheta_{k,n} := \sum_{j=k}^{n} \frac{1}{j}.$$

Let $q(z)$ be the quantile of order $e^{-z} = p_n$ of the distribution F. The estimator makes use of $X_{n-l_n:n}$, an intermediate order statistics of $X_1, .., X_n$, where l_n tends to infinity as $n \to \infty$ and $l_n \ / n \to 0$.
de Valk's estimator writes

$$\widehat{q}(z) = X_{n-l_n:n} \exp\left\{ g(\vartheta_{l_n,n}) h_\theta\left(\frac{z}{\vartheta_{l_{n+1},n}} \right) \right\}. \tag{2.17}$$

When the support of F overlaps $\mathbb{R}^-$ then the sample size n should be large; see de Valk ([27]) for details.
Note that, in the case of a $GPD(c, a)$, parameter θ is known and equal to 1 and the normalizing function g is defined by $g(x) = cx$ for $x > 0$.

Loss in accurracy due to binary sampling

In Table 2.7 we compare the performance of de Valk's method with ours on the model, making use of complete data in de Valk's estimation, and of dichotomous ones in our approach. Clearly de Valk's results cannot be attained by the present sequential method, due to the loss of information induced by thresholding and dichotomy. Despite this, the results can be compared, since even if the bias of the estimator clearly exceeds the corresponding bias of de Valk's, its dispersion is of the same order of magnitude, when handling heavy tailed GPD models. Note also that given the binary nature of the data considered, the average relative error is quite honorable. We can assess that a large part of the volatility of the estimator produced by our sequential methodology is due to the nature of the GPD model as well as to the sample size.

2.5 Sequential design for the Weibull model

The main property which led to the GPD model is the stability through threshold conditioning. However the conditional distribution of $\widetilde{R}$ given $\{\widetilde{R} > s\}$ also takes a rather simple form which allows for some variation of the sequential design method under the Weibull hypothesis.

2.5.1 The Weibull model

Denote $\widetilde{R} \sim W(\alpha, \beta)$, with $\alpha, \beta > 0$ a Weibull r.v. with scale parameter α and shape parameter β. Let G denote the distribution function of $\widetilde{R}$, g its density function and G^{-1} its

Parameters	**Relative error on the** $(1-\alpha)-$**quantile**			
	On complete data		**On binary data**	
	Mean	Std	Mean	Std
$c = 0.8$, $a_0 = 1.5$ and $s_\alpha = 469.103$	0.052	0.257	-0.222	0.554
$c = 1.5$, $a_0 = 1.5$ and $s_\alpha = 31621.777$	0.086	0.530	-0.504	0.720
$c = 1.5$, $a_0 = 3$ and $s_\alpha = 63243.550$	0.116	0.625	0.310	0.590

TABLE 2.7 – Mean and std of the relative errors on the $(1-\alpha)-$quantile of GPD on complete and binary data for samples of size $n = 250$ computed through 400 replicas of both estimation procedures.
Estimations on complete data are obtained with de Valk's method; estimations on binary data are provided by the sequential design.

quantile function. We thus write for non negative x

$$G(x) = 1 - \exp\left(-\left(\frac{x}{\alpha}\right)^{\beta}\right)$$
$$\text{for } 0 < u < 1, \quad G^{-1}(u) = \alpha(-\log(1-u))^{1/\beta}$$

The conditional distribution of $\widetilde{R}$ is a truncated Weibull distribution

$$\text{for } \widetilde{s}_2 > \widetilde{s}_1, \quad \mathbb{P}(\widetilde{R} > \widetilde{s}_2 \mid \widetilde{R} > \widetilde{s}_1) = \frac{\mathbb{P}(\widetilde{R} > \widetilde{s}_2)}{\mathbb{P}(\widetilde{R} > \widetilde{s}_1)} = \exp\left\{\left(-\left(\frac{s_2}{\alpha}\right)^{\beta} + \left(\frac{s_1}{\alpha}\right)^{\beta}\right)\right\}$$

Denote G_{s_2} the distribution function of $\widetilde{R}$ given $\{\widetilde{R} > \widetilde{s}_2\}$.

The following result helps. For $\widetilde{s}_2 > \widetilde{s}_1$,

$$\log\mathbb{P}(\widetilde{R} > \widetilde{s}_2 \mid \widetilde{R} > \widetilde{s}_1) = \left[\left(\frac{\widetilde{s}_2}{\widetilde{s}_1}\right)^{\beta} - 1\right] \log\mathbb{P}(\widetilde{R} > \widetilde{s}_1) \tag{2.18}$$

Assuming $\mathbb{P}(\widetilde{R} > \widetilde{s}_1) = p$, and given $\widetilde{s}_1$ we may find $\widetilde{s}_2$ the conditional quantile of order $1-p$ of the distribution of $\widetilde{R}$ given $\{\widetilde{R} > \widetilde{s}_1\}$. This solves the first iteration of the sequential estimation procedure through

$$\log p = \left[\left(\frac{\widetilde{s}_2}{\widetilde{s}_1}\right)^{\beta} - 1\right] \log p$$

where the parameter β has to be estimated on the first run of trials.

The same type of transitions holds for the iterative procedure; indeed for $\widetilde{s}_{j+1} > \widetilde{s}_j > \widetilde{s}_{j-1}$

$$\begin{aligned}\log\mathbb{P}(\widetilde{R} > \widetilde{s}_{j+1} \mid \widetilde{R} > \widetilde{s}_j) &= \left[\frac{\log\mathbb{P}(\widetilde{R} > \widetilde{s}_{j+1} \mid \widetilde{R} > \widetilde{s}_{j-1})}{\log\mathbb{P}(\widetilde{R} > \widetilde{s}_j \mid \widetilde{R} > \widetilde{s}_{j-1})} - 1\right] \log\mathbb{P}(\widetilde{R} > \widetilde{s}_j \mid \widetilde{R} > \widetilde{s}_{j-1}) \\ &= \left[\frac{\widetilde{s}_{j-1}^{\beta} - \widetilde{s}_{j+1}^{\beta}}{\widetilde{s}_{j-1}^{\beta} - \widetilde{s}_{j}^{\beta}} - 1\right] \log\mathbb{P}(\widetilde{R} > \widetilde{s}_j \mid \widetilde{R} > \widetilde{s}_{j-1})\end{aligned} \tag{2.19}$$

At iteration j the thresholds $\widetilde{s}_j$ and $\widetilde{s}_{j-1}$ are known; the threshold $\widetilde{s}_{j+1}$ is the $(1-p)-$ quantile of the conditional distribution, $\mathbb{P}(\widetilde{R} > \widetilde{s}_{j+1} \mid \widetilde{R} > \widetilde{s}_j) = p$, hence solving

$$\log p = \left[\frac{\widetilde{s}_{j-1}^{\beta} - \widetilde{s}_{j+1}^{\beta}}{\widetilde{s}_{j-1}^{\beta} - \widetilde{s}_{j}^{\beta}} - 1\right] \log p$$

where the estimate of β is updated from the data collected at iteration j.

2.5.2 Numerical results

Similarly as in Sections 2.4.3 and 2.4.4 we explore the performance of the sequential design estimation on the Weibull model. We estimate the $(1-\alpha)-$ quantile of the Weibull distribution in three cases. In the first one, the scale parameter a and the shape parameter b satisfy $(a, b) = (3, 0.9)$. This corresponds to a strictly decreasing density function, with heavy tail. In the second case, the distribution is skewed since $(a, b) = (3, 1.5)$ and the third case is $(a, b) = (2, 1.5)$ and describes a less dispersed distribution with lighter tail. Table 2.8 shows that the performance of our procedure here again largely depends on the shape of the distribution. The estimators are less accurate in case 1, corresponding to a heavier tail. Those results are compared to the estimation errors on complete data through de Valk's methodology. As expected, the loss of accuracy linked to data deterioration is similar to what was observed under the Pareto model, although a little more important. This can be explained by the fact that the Weibull distribution is less adapted to the splitting structure than the GPD.

2.6 Model selection and misspecification

In the above sections, we considered two models whose presentation was mainly motivated by theoretical properties. As it has already been stated in paragraph 2.3.3, the modelling of $\widetilde{R}$ by a GPD with c strictly positive is justified by the assumption that the support of the original variable R may be bounded by 0. However, note that the GPD

Parameters	Relative error on the $(1-\alpha)-$quantile			
	On binary data		On complete data	
	Mean	Std	Mean	Std
$a_0 = 3$, $b_0 = 0.9$ et $s_\alpha = 25.69$	0.282	0.520	0.127	0.197
$a_0 = 3$, $b_0 = 1.5$ et $s_\alpha = 10.88$	-0.260	0.490	0.084	0.122
$a_0 = 2$, $b_0 = 1.5$ et $s_\alpha = 7.25$	-0.241	0.450	0.088	0.140

TABLE 2.8 – Mean and std of relative errors on the $(1-\alpha)-$quantile of Weibull distributions on complete and binary data for samples of size $n = 250$ computed through 400 replicas.
Estimations on complete data are obtained with de Valk's method; estimations on binary data are provided by the sequential design.

model can be easily extended to the case where $c = 0$. It then becomes the trivial case of the estimation of an exponential distribution.

Though we did exclude the exponential case while modelling the excess probabilities of $\widetilde{R}$ by a GPD, we still considered the Weibull model in section 2.5, which belongs to the max domain of attraction for $c = 0$. On top of being exploitable in the splitting structure, the Weibull distribution is a classical tool when modelling reliability issues, it thus seemed natural to propose an adaptation of the sequential method for it.

In this section, we discuss the modelling decisions and give some hints on how to deal with misspecification.

2.6.1 Model selection

The decision between the Pareto model with tail index strictly positive and the Weibull model has been covered in the literature. There exists a variety of tests on the domain of attraction of a distribution.
Dietrich and al. (2002 [28]) and Drees and al. (2006 [31]) both propose a test for extreme value conditions related to Cramer-von Mises tests. Let X be a r.v of distribution function G. The null hypothesis is

$$H_0 : G \in MDA(c_0).$$

In our case, the theoretical value for the tail index is $c_0 = 0$. The former test provides a testing procedure based on the tail empirical quantile function, while the latter uses a weighted approximation of the tail empirical distribution. Choulakian and Stephens (2001 [20]) proposes a goodness of fit test in the fashion of Cramer-von Mises tests in which the unknown parameters are replaced by maximum likelihood estimators. The test consists in two steps: firstly the estimation of the unknown parameters, and secondly

the computation of the Cramer-von Mises W^2 or Anderson-Darling A^2 statistics. Let $X_1, \ldots, X_n$ be a random sample of distribution G. The hypothesis to be tested is:

$$H_0 : \text{The sample is coming from a } GPD(c_0, \widehat{a}).$$

The associated test statistics are given by:

$$W^2 = \sum_{i=1}^{n} \left(\widehat{G}(x_{(i)}) - \frac{2i-1}{2n} \right)^2 + \frac{1}{12n};$$
$$A^2 = -n - \frac{1}{n} \sum_{i=1}^{n} (2i-1) \left\{ \log(\widehat{G}(x_{(i)})) + \log(1 - \widehat{G}(x_(n+1-i)) \right\},$$

where $x_{(i)}$ denotes the $i-$th order statistic of the sample. The authors provide the corresponding tables of critical points.
Jurečková and Picek (2001 [39]) designed a non-parametric test for determining whether a distribution G is light or heavy tailed. The null hypothesis is defined by :

$$H_{c_0} : x^{1/c_0}(1 - G(x)) \leq 1 \quad \forall x > x_0 \text{ for some } x_0 > 0$$

with fixed hypothetical c_0. The test procedure consists in splitting the data set in N samples and computing the empirical distribution of the extrema of each sample.

The evaluation of the suitability of each model for fatigue data is precarious. The main difficulty here is that it is not possible to perform goodness-of-fit type tests, since firstly, we collect the data sequentially during the procedure and do not have a sample of available observations beforehand, and secondly, we do not observe the variable of interest R but only peaks over chosen thresholds. The existing tests procedures are not compatible with the reliability problem we are dealing with. On the first hand, they assume that the variable of interest is fully observed and are mainly semi-parametric or non-parametric tests based on order statistics. On the other hand, their performances rely on the availability of a large volume of data. This is not possible in the design we consider since fatigue trial are both time consuming and extremely expensive.

Another option consists of validating the model *a posteriori*, once the procedure is completed using expert advices to confirm or not the results. For that matter, a procedure following the design presented in 2.3.2 is currently being carried out. Its results should be available in a few months and will give hints on the most relevant model.

2.6.2 Handling misspecification under the Pareto model

In paragraph 2.3.3, we assumed that $\widetilde{R}$ initially follows a GPD. In practice, the distribution may have its excess probabilities converge towards it as the thresholds increase but differ

from a GPD. In the following, let us assume that $\widetilde{R}$ does not follow a GPD (of distribution function F) but another distribution G whose tail gets closer and closer to a GPD.

In this case, the issue is to control the distance between G and the theoretical GPD and to determine from which thresholding level it becomes negligible. One way to deal with this problem is to restrict the model to a class of distributions that are not so distant from F: Assume that the distribution function G of the variable of interest $\widetilde{R}$ belongs to a neighborhood of the $GPD(c,a)$ of distribution function F, defined by:

$$V_\epsilon(F) = \left\{ G : \sup_x |\bar{F}(x) - \bar{G}(x)| w(x) \leq \epsilon \right\}, \tag{2.20}$$

where $\epsilon \geq 0$ and w an increasing weight function such that $\lim_{x\to\infty} w(x) = \infty$.
$V_\epsilon(F)$ defines a neighborhood which does not tolerate large departures from F in the right tail of the distribution.

Let $x \geq s$, it follows from (2.20) a bound for the conditional probability of x given $R > s$:

$$\frac{\bar{F}(x) - \epsilon / w(x)}{\bar{F}(s) + \epsilon / w(s)} \leq \frac{\bar{G}(x)}{\bar{G}(s)} \leq \frac{\bar{F}(x+) + \epsilon / w(x)}{\bar{F}(s) - \epsilon / w(s)}. \tag{2.21}$$

When $\epsilon = 0$, the bounds of (2.21) match the conditional probabilities of the theoretical Pareto distribution.
In order to control the distance between F and G, the bound above may be rewritten in terms of relative error with respect to the Pareto distribution. Using a Taylor expansion of the right and left bounds when ϵ is close to 0, it becomes:

$$1 - u(s,x).\epsilon \leq \frac{\frac{\bar{G}(x)}{\bar{G}(s)}}{\frac{\bar{F}(x)}{\bar{F}(s)}} \leq 1 + u(x,s).\epsilon, \tag{2.22}$$

where

$$u(s,x) = \frac{\left(1 + \frac{cs}{a}\right)^{1/c}}{w(s)} + \frac{\left(1 + \frac{cx}{a}\right)^{1/c}}{w(x)}.$$

For a given ϵ close to 0, the relative error on the conditional probabilities can be controlled upon s. Indeed, then the relative error is bounded by a fixed level $\delta > 0$ whenever:

$$\frac{\left(1 + \frac{cs}{a}\right)^{1/c}}{w(s)} \leq \frac{\delta}{\epsilon} \frac{\left(1 + \frac{cx}{a}\right)^{1/c}}{w(x)}.$$

2.7 Perspectives, generalization of the two models

In this work, we have considered two models for $\widetilde{R}$ that exploits the thresholding operations used in the splitting method. This is a limit of this procedure as the lack of relevant information provided by the trials do not enable a flexible modelling of the distribution of the resistance. In the following, we present ideas of extensions and generalizations of those models, based on common properties of the GPD and Weibull models.

2.7.1 Variations around mixture forms

When the tail index is positive, the GPD is completely monotone, and thus can be written as the Laplace transform of a probability distribution. Thyrion (1964[52]) and Thorin (1977[51]) established that a $GPD(a_T, c_T)$, with $c_T > 0$, can be written as the Laplace transform of a Gamma r.v V whose parameters are functions of a_T and c_T : $V \sim \Gamma\left(\frac{1}{c_T}, \frac{a_T}{c_T}\right)$. Denote v the density of V,

$$\begin{aligned} \forall x \geq 0, \quad \bar{G}(x) &= \int_0^\infty \exp(-xy)\, v(y)\, dy \\ \text{where} \quad v(y) &= \frac{(a_T/c_T)^{1/c}}{\Gamma(1/c_T)} y^{1/c_T - 1} \exp\left(-\frac{a_T y}{c_T}\right). \end{aligned} \tag{2.23}$$

It follows that the conditional survival function of $\widetilde{R}$, $\bar{G}_{s_j}$, is given by:

$$\begin{aligned} \mathbb{P}(\widetilde{R} > \widetilde{s}_{j+1} \mid \widetilde{R}_j > \widetilde{s}_j) &= \bar{G}_{\widetilde{s}_j}(\widetilde{s}_{j+1} - \widetilde{s}_j) \\ &= \int_0^\infty \exp\left\{-(\widetilde{s}_{j+1} - \widetilde{s}_j) y\right\} v_j(y)\, dy, \\ &\quad \text{where } V_j \text{ is a r.v of distribution } \Gamma\left(\frac{1}{c_j}, \frac{a_j}{c_j}\right). \end{aligned}$$

with $c_j = c_T$ and $a_j = a_{j-1} + c_T(\widetilde{s}_j - \widetilde{s}_{j-1})$.

Expression (2.23) gives room to an extension of the Pareto model. Indeed, we could consider distributions of $\widetilde{R}$ that share the same mixture form with a mixing variable W that possesses some common characteristics with the Gamma distributed r.v. V.

Similarly, the Weibull distribution $W(\alpha, \beta)$ can also be written as the Laplace transform of a stable law of density g whenever $\beta \leq 1$. Indeed, it holds from Feller 1971[34]) (p. 450, Theorem 1) that:

$$\forall x \geq 0, \quad \exp\left\{-x^\beta\right\} = \int_0^\infty \exp(-xy)\, g(y)\, dy \tag{2.24}$$

where g is the density of an infinitely divisible probability distribution.
It follows, for $s_j < s_{j+1}$

$$\begin{aligned}\mathbb{P}(\widetilde{R} > \widetilde{s}_{j+1} \mid \widetilde{R}_j > \widetilde{s}_j) &= \frac{\exp\left\{-(\widetilde{s}_{j+1}/\alpha)^\beta\right\}}{\exp\left\{-(\widetilde{s}_j/\alpha)^\beta\right\}} \\ &= \frac{\int_0^\infty \exp\left\{(-(\widetilde{s}_{j+1}/\alpha)y\right\} g(y)dy}{\int_0^\infty \exp\left\{-(\widetilde{s}_j/\alpha)y\right\} g(y)dy} = \frac{\int_0^\infty \exp\left\{-(\widetilde{s}_{j+1}/\alpha)y\right\} g(y)dy}{K(s_j)} \\ &= \frac{1}{K(s_j)} \int_0^\infty \exp\left\{-\widetilde{s}_{j+1}u\right\} g_\alpha(u))du \\ &\qquad \text{with } u = y/\alpha \text{ and } g_\alpha(u) = \alpha g(\alpha u)\end{aligned} \tag{2.25}$$

Thus an alternative modelling of $\widetilde{R}$ could consist in any distribution that can be written as a Laplace transform of a stable law of density $w_{\alpha,\beta}$ defined on $\mathbb{R}_+$ and parametrized by (α, β), that complies to the following condition: For any $s > 0$, the distribution function of the conditional distribution of $\widetilde{R}$ given $\widetilde{R} > s$ can be written as the Laplace transform of $w_{\alpha,\beta}^{(\alpha,s)}(.)$ where

$$x > s, w_{\alpha,\beta}^{(\alpha,s)}(x) = \frac{\alpha w_{\alpha,\beta}(\alpha x)}{K(s)},$$

where $K(.)$ is defined in (2.25).

2.7.2 Variation around the GPD

Another approach, inspired by Naveau et al. (2016[44]), consists in modifying the model so that the distribution of $\widetilde{R}$ tends to a GPD as x tends to infinity and it takes a more flexible form near 0.
$\widetilde{R}$ is generated through $G^{-1}_{(c_T,a_T)}(U)$ with $U \sim \mathcal{U}[0,1]$. Let us consider now a deformation of the uniform variable $V = L^{-1}(U)$ defined on $[0,1]$, and the transform W of the GPD: $W^{-1}(U) = G^{-1}_{(c_T,a_T)}(L^{-1}(U))$.
The survival function of the GPD being completely monotone, we can choose W so that the distribution of $\widetilde{R}$ keeps this property.

Proposition 3. If ϕ: $[0,\infty[\to \mathbb{R}$ is completely monotone and let ψ be a positive function, such that its derivative is completely monotone, then $\phi(\psi)$ est completely monotone.

The transformation of the GPD has cumulative distribution function $W = L(G_(c_T, a_T))$ and survival function $\bar{W} = \bar{L}(G_(c_T, a_T))$. $G(c_T, a_T)$ is a Berstein function, thus $\bar{W}$ is completely monotone if $\bar{L}$ is also.

Examples of admissible functions:

(1) Exponential form :

$$\begin{aligned} &L(0) = 0 \\ &L(x) = \frac{1-\exp(-\lambda x^{\alpha})}{1-\exp(-\lambda)} \quad \text{avec } 0 \leq \alpha \leq 1 \text{ et } \lambda > 0 \\ &L(1) = 1 \end{aligned}$$

The obtained transformation is: $\forall x > 0$,

$$\bar{W}_{(\lambda,c_T,a_T)}(x) = \bar{L}(G(x)) = \frac{\exp\left(-\lambda\left[1-(1+\frac{c_T}{a_T})^{-1/c_T}\right]^{\alpha}\right)-\exp(-\lambda)}{1-\exp(-\lambda)}$$

with $\bar{W}_{(\lambda,c_T,a_T)}(x)$ completely monotone.

(2) Logarithmic form:

$$\begin{aligned} &L(0) = 0 \\ &L(x) = \frac{\log(x+1)}{\log 2} \quad \left(\text{ or more generally } \frac{\log(\alpha x+1)}{\log 2},\ \alpha > 0\right) \\ &L(1) = 1 \end{aligned}$$

and $\forall x > 0$,

$$\bar{W}_{(c_T,a_T)}(x) = 1 - \frac{\log\left(2-(1+\frac{c_T}{a_T})^{-1/c_T}\right)}{\log 2}$$

(3) Root form:

$$\begin{aligned} &L(0) = 0 \\ &L(x) = \frac{\sqrt{x+1}-1}{\sqrt{2}-1} \\ &L(1) = 1 \end{aligned}$$

and

$$\bar{W}_{(c_T,a_T)}x) = 1 - \frac{\sqrt{2-(1+\frac{c_T x}{a_T})^{-1/c_T}}-1}{\sqrt{2}}$$

(4) Fraction form:

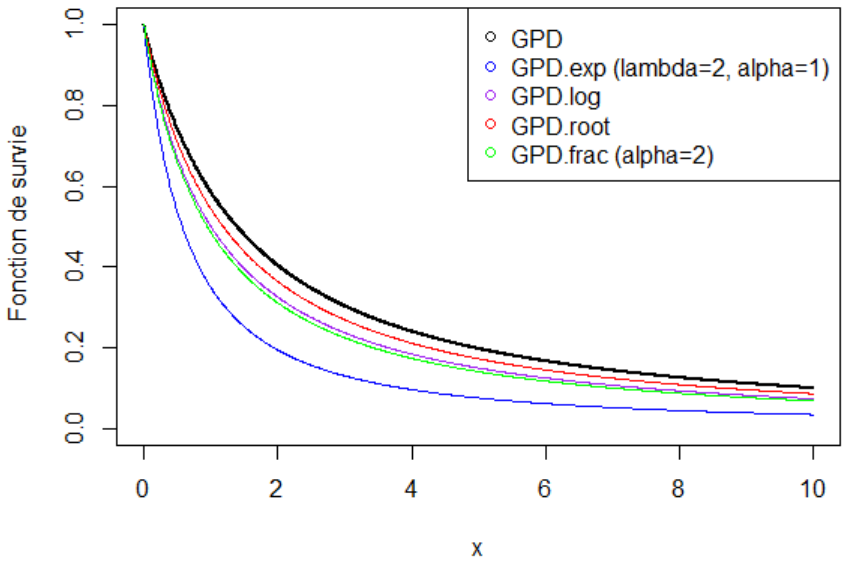


FIGURE 2.10 – Survival functions associated with transformations of the GPD$(0.8, 1.5)$

$$
\begin{aligned}
L(0) &= 0 \\
L(x) &= \frac{(\alpha+1)x}{x+\alpha}, \quad \alpha > 0 \\
L(1) &= 1
\end{aligned}
$$

and

$$
\bar{W}_{(\alpha,c_T,a_T)}(x) = 1 - \frac{(\alpha+1)\left(1-(1+\frac{c_T x}{a_T})^{-1/c}\right)}{1-(1+\frac{c_T x}{a_T})^{-1/c_T}+\alpha}
$$

The shapes of the above transformations of the GPD are shown in Figure 2.10.

However those transformations do not conserve the stability through thresholding of the Pareto distribution. Thus, their implementation does not give stable results. Still they give some insight on a simple generalization of the proposed models usable under additional information on the variable of interest.

2.8 Conclusion

The splitting induced procedure presented in this chapter proposes an innovative experimental plan to estimate an extreme quantile failure. Its development has been motivated by on the one hand major industrial stakes, and on the other hand the lack of relevance of the used methodology. The main difficulty in this setting is the nature of the information at hand, since the variable of interest is latent, therefore only peaks over thresholds may be observed. Indeed, this study is directly driven from an application in material fatigue strength: when performing a fatigue trial, the strength of the specimen obviously can not be observed; only the indicator of whether or not the strength was greater than the tested level is available.

Among the methodologies dealing with such a framework, none is adapted to the estimation of extreme quantiles. We therefore proposed a plan based on splitting methods in order to decompose the initial problem into less complex ones. The splitting formula introduces a formal decomposition which has been adapted into a practical sampling strategy targeting progressively the tail of the distribution of interest.

The structure of the splitting equation has motivated the parametric hypothesis on the distribution of the resistance. Two models exploiting a stability property have been presented: one assuming a Generalized Pareto Distribution and the other a Weibull distribution.

The associated estimation procedure has been designed to use the iterative and stable structure of the model by combining a classical maximum likelihood criterion with a consistency criterion on the sequentially estimated quantiles. The quality of the estimates obtained through this procedure have been evaluated numerically. Though constrained by the quantity and quality of information, those results can still be compared to what would be obtained ideally if the variable of interest was observed.

On a practical note, while the GPD is the most adapted to the splitting structure, the Weibull distribution has the benefit of being particularly suitable for reliability issues. The experimental campaign launched to validate the method will contribute to select a model.

2.9 Appendix

2.9.1 Alternative estimation criterion, divergence minimization

Denote $\mathbf{P}_j$ the Bernoulli distribution of the observations $Y_{j,1}, \ldots, Y_{j,n}$ obtained at iteration j of the procedure and $\mathbf{P}_n$ their empirical distribution. Let p_j be the probability

of failure under the Pareto model defined by $p_j = \mathbb{P}_{(\widehat{c},\widehat{a})_j}\left(\widetilde{R} > s_j - s_{j-1} \mid \widetilde{R} > s_{j-1}\right)$ and denote $\bar{p}_j = 1 - p_j$. Let n_1 and n_0 be the number of observed failures and survivals : $n_1 = \sum_i Y_{j,i}$ and $n_0 = n - n_1$. Alternative estimation methods of the maximum likelihood procedure consist in minimizing the following distances over the parameters of the model :

- $\mathscr{L}_1$ distance :

$$d_1 = \left|\frac{n_1}{n} - p_j\right| + \left|\frac{n_0}{n} - p_j\right| \tag{2.26}$$

- Kullback-Leibler distance :

$$D_{KL}\left(\mathbf{P}_j \| \mathbf{P}_n\right) = p_j \log\left(\frac{p_j}{n_1/n}\right) + \bar{p}_j \log\left(\frac{\bar{p}_j}{n_0/n}\right) \tag{2.27}$$

- Helliger distance :

$$H\left(\mathbf{P}_j, \mathbf{P}_n\right) = \frac{1}{\sqrt{2}}\sqrt{\left(\sqrt{\frac{n_1}{n}} - \sqrt{p_j}\right)^2 + \left(\sqrt{\frac{n_0}{n}} - \sqrt{\bar{p}_j}\right)^2} \tag{2.28}$$

- khi-square distance :

$$D^2 = \frac{\left(\frac{n_1}{n} - p_j\right)^2}{p_j} + \frac{\left(\frac{n_0}{n} - p_j\right)^2}{\bar{p}_j} \tag{2.29}$$

The khi-square distance is not represented in the results in Figure 2.7 because it gives extremely scattered results.

2.9.2 Algorithm for global optimization

The optimizations at each estimation step of the splitting procedure are performed using a global optimization method, called the SAFIP algorithm and introduced by Biret and Broniatowski (2016 [8]).
The algorithm aims at solving equations of the form $f(x) = 0$, where f is a real valued function defined on $\mathscr{X}$. Without regularity assumptions on f, it returns a set of solutions

$$\mathscr{S} = \left\{x \in \mathscr{X} : f(x) = 0\right\}.$$

The procedure consists in generating sequences $((z_i)_{i\in\mathbb{N}}), z_i \in \mathscr{X}$. Converging sequences are conserved while the others are discarded, until a fixed number of solutions N of the optimization problem is found.
The sequences are defined iteratively such that

$$z_{i+1} = z_i + \frac{z_{i-1} - z_i}{2} + k f(z_i).$$

Denoting $R_i = |z_i - z_{i-1}|$, it follows that

$$R_{i+1} \leq \frac{R_i}{2} + k|f(z_i)|.$$

Practically, each sequence $(z_i)_i$ is initiated by z_0 and z_1 uniformly drawn from $\mathscr{X}$. Let $R_1 = |z_1 - z_0|$. The following rule determines whether the sequence is continued or discarded:
For $i \leq 1$, if $|f(z_i) \leq C|f(z_{i-1}|$, then define

$$z_{i+1} := z_i + u_i,$$

where u_i is drawn from the ball $\mathscr{B}\left(0, R_i/2 + k|f(z_{i-1})|\right)$.
Otherwise, the sequence is stopped. The stopping rule depends on a fixed tolerance parameter.
Biret and Broniatowski proved that sequences generated this way converge almost surely to a limit in $\mathscr{S}$.
Note that k and C are tuning parameters.

Deuxième partie

Modelling tools for S-N curves

Chapitre 3

Modelling mean and quantile S-N curves

3.1 Objectives

The second part of the thesis focuses on pursuing works from R. Fouchereau (2014 [35]), on modelling the S-N curve as a whole.

The methodology used for characterizing the lifetime of a material remains the same than for studying the minimal allowable stress. Experimental campaigns are conducted over a batch of n specimens loaded at J different levels of constraint $(\sigma_{a,1},\dots,\sigma_{a,J})$ or of deformation $(\epsilon_{a,1},\dots,\epsilon_{a,J})$, with $J \leq n$. Here σ_a and ϵ_a stand respectively for alternated constraint and alternated deformation, i.e the half amplitude of the applied stress or of the observed deformation.
Results consists in couples stress-lifetime : $\left((N_k,\sigma_{a,k})_{0\leq k\leq n}\right)$ or $\left((N_k,\epsilon_{a,k})_{0\leq k\leq n}\right)$

Usually the controlled variable for low-cycle fatigue is the degree of deformation, while for high-cycle fatigue, it is the level of stress. Thus, σ_a and ϵ_a can be both used to refer to this variable.

A major challenge in aircraft industry is the characterization of medium and extreme behavior of a material under fatigue damage at any level of solicitation, by estimating :

- Mean S-N curves ;
- Quantile S-N curves, especially the minimal curve at failure probability 0.1%.

Those objectives have to take into account operational constraints related to the costs of a campaign, namely :

- The amount of information varies according to the fatigue regime studied ;

— The amount of information varies according to the different trial conditions ;

— The amount of information is related to the critical role of the engine part of interest.

The models considered must be compatible with restricted sample sizes, namely between 30 and 50 observations at a given level σ.

Thus the work has a two-fold objective. Firstly, it consists in proposing a modelling of fatigue life relevant with mecanical and physical knowledged pertaining to failure. Secondly, it aims at proposing statistical tools to specify the model and address shortcomings in the methodologies currently implemented by Safran Aircraft Engines for the estimation of 0.1%−quantile fatigue curves. Indeed, since the quantity of interest is located far in the left tail of the distribution, its estimation is highly dependent on the parametric hypothesis made on the fatigue life distribution at a given level of stress.

In the following, we will present in Section 3.2 how are presently constructed the mean and minimal S-N curves and the limits of this methodology. An alternative modelling based on the mechanics of failure is then proposed. The following chapters are devoted to the presentation of two test methodologies that will enable to define the most relevant model for fatigue life in any fatigue regime. Indeed, Chapter 4 provides a test procedure to determine whether the observations are drawn from a unique random variable or from a sum of random variables. Lastly Chapter 5 introduces a test for the number of components of a mixture distribution.

3.2 State of the art on the estimation of S-N curves

3.2.1 Physical models

There are a few mathematical models aiming at characterizing the different fatigue regimes.

Those models are usually expressed either as functions of the strain or of the lifetime, taking either the form :

$$N = f_\theta(\sigma_a) + u \tag{3.1}$$

or

$$\sigma_a = f_\theta(N) + u \tag{3.2}$$

where u is the error term that is usually assumed to follow a Gaussian distribution.
The most commonly used models are given in table 3.1.

Model	Fatigue regime	Equation
Basquin	High cycle fatigue	$\log(N) = a + b \times \log(\sigma_a)$
Stromeyer	High cycle fatigue and endurance limit	$\sigma_a = \sigma_e + \left(\frac{a}{N}\right)^b$
Bastenaire	Low and high cycle fatigue and endurance limit	$N = \frac{A \exp\left\{-\left(\frac{\sigma_a - \sigma_e}{B}\right)^C\right\}}{\sigma_a - \sigma_e}$

Note: σ_e refers to the endurance limit of the material.

TABLE 3.1 – Main fatigue models

Among those, Basquin equation is the most widely used. One of the main reason for the popularity of this model lies in its simplicity. Yet it allows for some extension: The parameters a and b of the equation may depend on some other controlled variable, such as the type of trial or of material or the temperature. For some materials, it has been shown that the slope in the Basquin equation is independent to the temperature. This result allows to estimate a model on all trials as a whole rather than segmenting the data base according to the temperature. More generally, the integration of a qualitative or discrete variable X amounts to estimate:

$$\log(N) = a + b \times \log(\sigma_a) + u \tag{3.3}$$

where the parameters a and b are function of X: $a = a(X), b = b(X)$, where X stands for the temperature and/or the stress-ratio.

Though its log-linear form is a powerful argument in its favour, the model only fits partially the data. Indeed, as illustrated in Figure 3.1, lifetime fatigue data display a bimodal distribution.

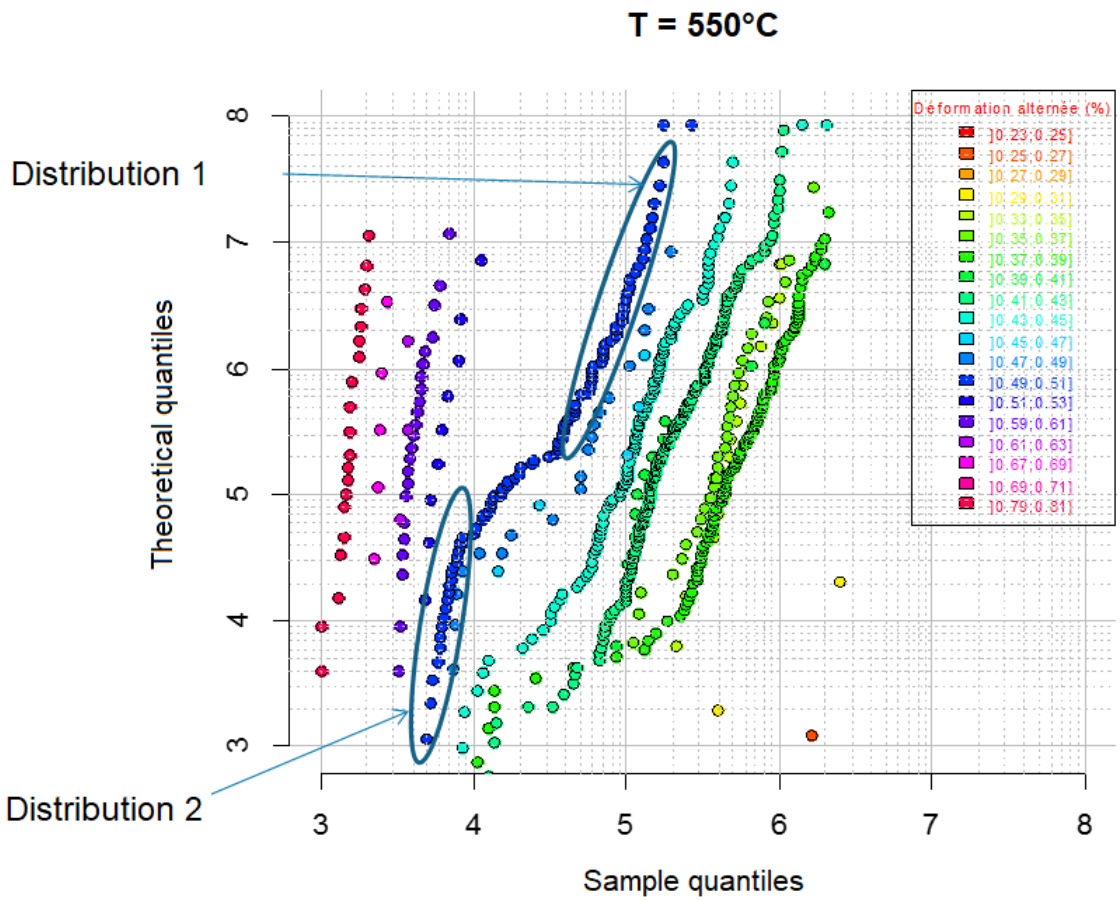


FIGURE 3.1 – Quantile-quantile Gaussian plot of the logarithm of the numbers of cycles to failure for different levels of strain, nickel based material, T = 550°C

This bimodal feature finds its origin in the different mechanisms leading to failure. The inclusion of defects and their position in the material play a major part in the failure of a specimen.
The present methodology used by Safran Aircraft Engines emphasizes on the following dichotomy :

- Crack initiations at grain boundaries, due to inclusions at the surface of the material. They occur quickly and the mean fatigue life and its dispersion are small.
- Internal crack initiations lead to more high and scattered fatigue lives.

Thus, as shown in Figure 3.2, low number of cycles to failure are predominantly due to fatigue crack initiations at surface while high number of cycles corresponds mostly to internal initiations. However in the central part of the curve, both types coexists in variable proportions as the fatigue life increases and the distribution of the latter is bimodal. This observations pleads in favor of the modelling of the fatigue life at a given deformation level as a mixture distribution.

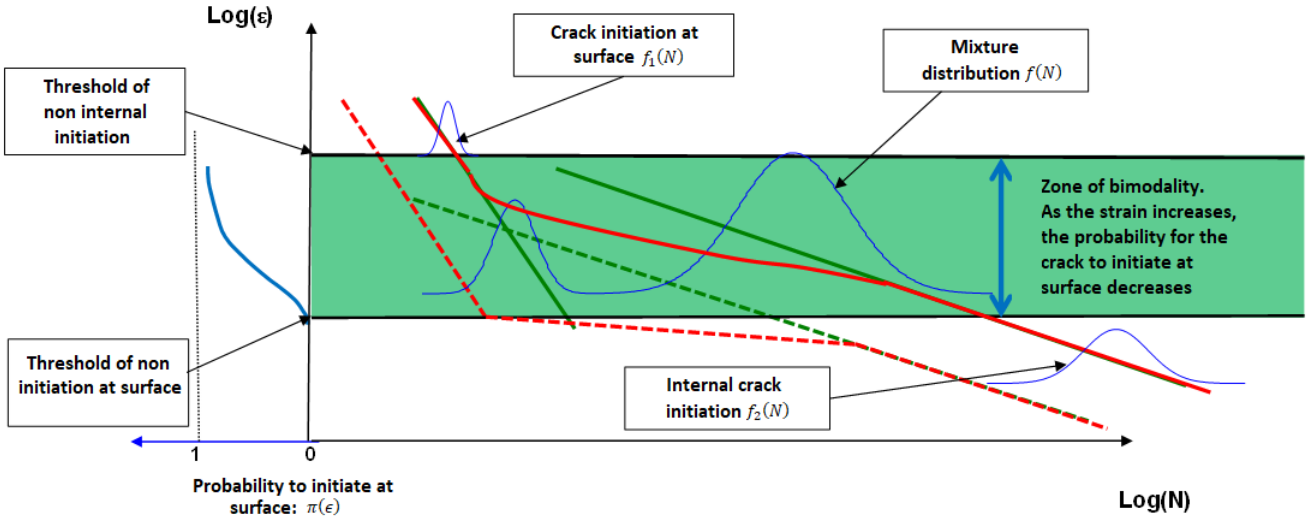


FIGURE 3.2 – Illustration of the mixture distribution

3.2.2 Representation of the mean S-N curve as a two-component mixture

The current modelling relies on the above dichotomy: The tested specimen is either subjected to the first failure mechanism with probability π or the second, with probability $1-\pi$, where $\pi = \pi(\epsilon)$. The fatigue life N is equal to the number of cycles to the first mode of failure or to the second failure mode. The density of N therefore writes:

$$f(N \mid \epsilon_a) = \pi(\epsilon_a) f_1(N \mid \epsilon_a) + (1 - \pi(\epsilon_a)) f_2(N \mid \epsilon_a). \tag{3.4}$$

Each of the failure modes is associated with a Basquin model. Thus the mean curve equation is the following:

$$\log(N) = \pi(\epsilon_a)\left(a_1 + b_1 \log(\epsilon_a)\right) + (1 - \pi(\epsilon_a))\left(a_2 + b_2 \log(\epsilon_a)\right) \tag{3.5}$$

where $\pi(\epsilon_a)$ is usually modeled by a logistic distribution with parameters m and s :

$$\pi(\epsilon_a) = \frac{1}{1 + \exp\left(\frac{-\epsilon_a - m}{s}\right)}. \tag{3.6}$$

Parameters estimation

The model estimation follows a two-step procedure:

1. The first step consists in attributing each observation to a failure mode. It is performed through an Expectation-Maximization algorithm. The problem can be formalized as follows: Denote

   - $(N_i, \epsilon_{a,i})_{i=1\dots n}$, the set of n observations.
   - $(N_i, Z_i, \epsilon_{a,i})_{i=1\dots n}$, the unobserved complete data. Here Z is the latent variable, namely the indicator of initiation at surface. Note that Z can be observed in specific trials. In those cases, it is used to check the results.
   - The parameter vector: $\theta = (a_1, b_1, a_2, b_2, m, s)$.

   The complete likelihood of the above model is given by:

$$\begin{aligned}\log L(N, Z, \epsilon_a \mid \theta) = \sum_{i=1}^{n} & z_i \log(\pi(\epsilon_{a,i} \mid s, m) f_1(N_a, \epsilon_{a,i} \mid a_1, b_1)) \\ & + (1 - z_i) \log((1 - \pi(\epsilon_{a,i} \mid s, m)) f_2(N_a, \epsilon_{a,i} \mid a_2, b_2).\end{aligned}$$

   The estimation of the parameter vector θ is obtained iterating the following steps: At iteration k,

   (a) **E step:** Compute the expectation of the log-likelihood associated with the current estimate of θ.

$$Q\left(\theta, \theta^{(k)}\right) = \mathbb{E}_{\theta^{(k)}}\left(\log L(N, Z) \mid N\right).$$

   (b) **M step:** Update the estimate by maximizing the expectation of the log-likelihood.

$$\theta^{(k+1)} = \underset{\theta \in \Theta}{\operatorname{argmax}}\, Q\left(\theta, \theta^{(k)}\right).$$

   Once the estimations obtained, each observation is then allocated to a class based on the estimated probability of belonging to each of them: denote C_1 and C_2 the two classes of observations, for $j = 1, 2$, observation i is assigned to C_j if

$$\widehat{\mathbb{P}}(N_i \in C_j) = \max_{l=1,2} \widehat{\mathbb{P}}(N_i \in C_l),$$

   where

$$\widehat{\mathbb{P}}(N_i \in C_j) = \frac{\widehat{\pi}(\epsilon_{a,i}) \widehat{F}_1(N_i)}{\widehat{\pi}(\epsilon_{a,i}) \widehat{F}_1(N_i) + \widehat{\pi}(\epsilon_{a,i}) \widehat{F}_2(N_i)}.$$

2. Two Basquin model are then fitted on the two classes of observations determined through the EM algorithm.

3.2.3 Construction of the minimal S-N curves

The minimal curves can be obtained through two different procedures. Firstly by making use of the unimodal models. The quantile curve is obtained by translating the mean curve up to a flat-rate abatement factor k. Then the quantile curve is equal to the mean curve minus k times the standard deviation of the Basquin model, where k is fixed to 3 when the number of observations is sufficient. The second estimation method consists in exploiting the mixture feature, but is not used in practice because the obtained results are extremely sensitive to the estimations of the proportion of each failure mode, especially in zones where π is close to 0 or 1.

This is why the minimal curves are presently constructed on the basis of the estimations of the Basquin models on each sub-population. They are obtained by applying the following rules:

- **For $\epsilon > \epsilon_0$:** take the 0.1%−quantile of the population subjected to crack initiation at surface by applying an abatement factor to the associated Basquin model.
- **For $N > 10^6$:** similarly, take the 0.1%−quantile of the population of internal crack initiation.
- **For $\epsilon \leq \epsilon_0$ and $N \leq 10^6$:** the quantile curve is the result of a simple linear interpolation in a log−log diagram between the two unimodal models as illustrated by Figure 3.3.

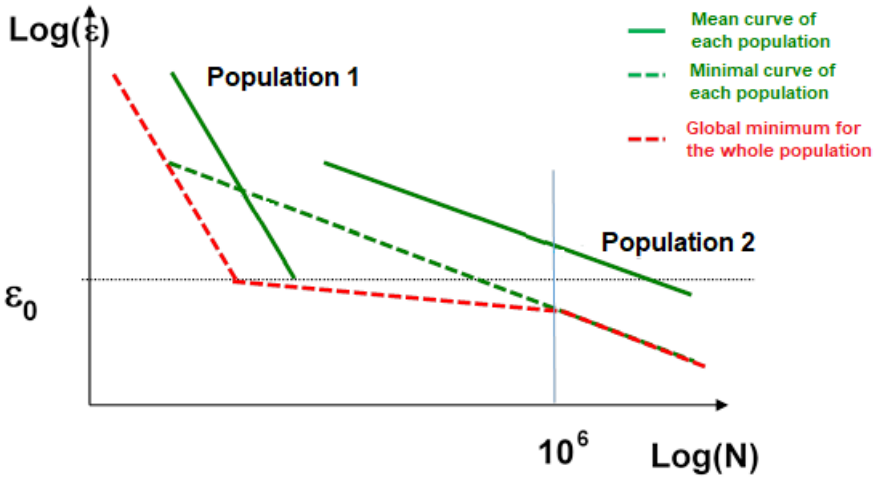


FIGURE 3.3 – Estimation method of the minimal S-N curves used by Safran Aircraft Engines

This estimation method is unsatisfactory in several respects. Firstly, the estimation of the mean curve is not direct and is derived from a two step estimation procedure performed on the same data set. Secondly the quantile curves are only obtained through a simple

translation of the mean curve. Furthermore, the mixture feature of the fatigue life is only used to classify the observations. Those methodological drawbacks motivates the seek of an alternative modelling and construction method for the fatigue curves that directly exploit the mixture structure. In the following section, we consider a new model for fatigue life more fitted to the data structure.

3.3 An alternative modelling to fatigue life, initiation-propagation model

3.3.1 Fracture mechanics

In order to propose an alternative to the mixture of Basquin models, we consider another form for the mixture distribution whose component are more directly linked to fracture mechanics results. The application of a level of stress during a sufficiently long time will cause the deformation of the material, followed by the formation of a crack, whose propagation will lead to rupture. Thus failure occurs after:

- the crack initiation period N_i,
- the crack propagation period N_p.

The duration of each component is highly dependent to the level of stress considered. Indeed, the more stress is applied, the more quickly the crack will initiate and the material will break. On the other hand, for low levels of strength, the initiation period is very long (millions of cycles) and the propagation is quite negligible in comparison.
The fatigue life is the sum of those two times, i.e.:

$$\underbrace{N}_{\text{measured}} = \underbrace{N_i}_{\text{unknown}} + \underbrace{N_p}_{\text{partially measured or calculated}} \tag{3.7}$$

where N is observed and might be subject to right censoring, N_p can be measured through specific propagation trials and N_a is completely unknown. This dichotomy has also been used by Fouchereau (2014 [35]).

The proposed modelling replaces the distinction between surface and internal initiation by the one between fast and slow initiation. The latter has several perks with respect to the previous one. It exploits mechanical properties of fracture and contrary to the present model, do not necessitate costly fractographic data. Moreover, the distinction fast vs. slow initiations is valid for any metallic material, while the previous was not. The two dichotomies are not completely equivalent in that initiation at surface does not recover all fast crack initiations and reciprocally.

As stated above, it is usually considered that when the applied stress is sufficiently high, a crack will immediately initiate in the material, that is to say from the first cycle of application. In this case, the life to failure N could be reduced to the propagation time. A possible model could therefore be the following: let

- F_i be the distribution function of N_i;
- G_p be the distribution function of N_p;
- $H_{N_i+N_p}$ be the distribution function the sum of N_i and N_p;
- Z, the indicator of short initiation, distributed according to $\mathscr{B}(\pi(\epsilon_a))$.

The number of cycles to failure writes

$$N = \begin{cases} N_p & \text{if } Z = 1 \\ N_i + N_p & \text{else.} \end{cases}$$

and its distribution is the following:

$$F(x) = \pi(\epsilon_a) G_p(x) + (1 - \pi(\epsilon_a)) H_{N_i+N_p}(x).$$

Usually, the initiation and propagation periods are modeled by a Lognormal distribution. While the propagation life remains largely unknown, the propagation time can be measured through fractographic studies. These data enable to obtain a more precise information on N_p.

3.3.2 Modelling the propagation period

Crack propagation is studied through fractography, i.e. the analysis of fracture surfaces. The trials performed in this framework, a specimen is given an initial crack of determined size. It is then loaded at a given stress and the evolution of the crack size is measured regularly until it reaches a maximum length or until failure. Thus since the specimen is already cut at the beginning of the experiment, the measured time corresponds only to the propagation period N_p.
Those trials can be exploited in order to represent precisely the propagation N_p. This is done by making use of a classical model of fatigue crack propagation rate: Paris law, given by:

$$\frac{da}{dN_p} = C.\Delta K^m \tag{3.8}$$

where

- a denotes the crack size;
- da/dN_p is the fatigue crack growth;

- $\Delta K = \Delta\sigma_a\sqrt{\pi a}F$ is the stress intensity range;
- C et m are experimentally determined material constants which also depend on environmental effects, stress ratio.

An explicit form of propagation life can be obtained through integration of (3.8),

$$N_p = \frac{a_f^{1-m/2} - a_i^{1-m/2}}{(1-m/2)\,C\pi^{m/2}\,(\sigma_a F)^m}, \tag{3.9}$$

where a_i and a_f denotes respectively the initial and final crack size.
The random elements are the model parameters m and C which are derived from the following linear regression:

$$\log\left(\frac{da}{dN_p}\right) = \log C + m\log\Delta K + \epsilon, \tag{3.10}$$

m and $\log C$ are thus assumed to be Gaussian, which has been verified through trials. They also are linearly dependent: there exists α and β such that

$$\log C = \alpha + \beta m + \epsilon_2, \tag{3.11}$$

with

$$\begin{aligned} m &\sim \mathcal{N}\left(\mu, s^2\right), \\ \epsilon_2 &\sim \mathcal{N}\left(0, \sigma^2\right), \\ \log C &\sim \mathcal{N}\left(\mu, b^2 s^2 + \sigma^2\right). \end{aligned}$$

The resulting distribution of N_p is also Lognormal. Figure 3.4 shows that simulated law of N_p based on the estimated parameters on fractographic data fits perfectly a Lognormal.

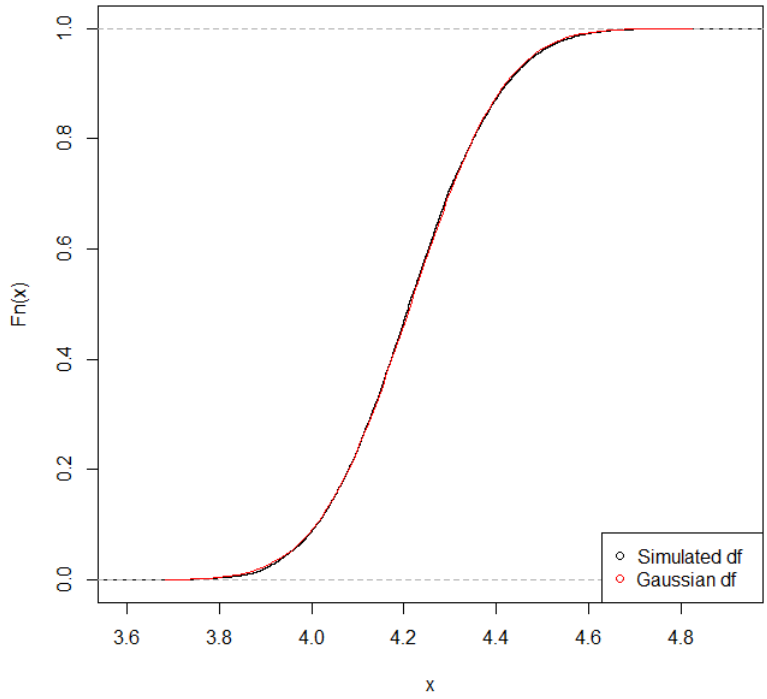


FIGURE 3.4 – Empirical cumulative distribution function of $\log N_p$ and Gaussian distribution function

This model also highlights connections between the parameters and the test conditions. For instance, Figure 3.5 illustrates how N_p depends on to the loading level in expectation and in variance as well.

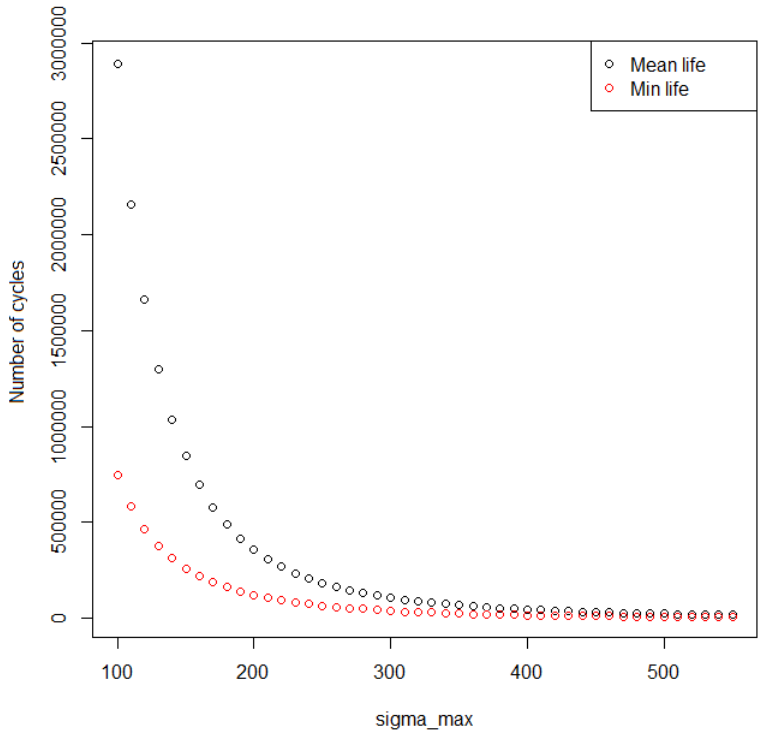


FIGURE 3.5 – Evolution of the mean and minimum propagation time N_p with respect to the maximum loading $\sigma_{\max}$

3.3.3 Number of components

As stated in the above sections, data representation highlights the existence of at least two components in the distribution of the number of cycles to failure (see Figure 3.1). However, no further study has investigated the exact number of components, as well as the nature of those components. A number of questions arises:

1. Under a high cyclic loading, a crack initiate extremely quickly and the resulting fatigue life is low. It is therefore considered that $N \approx N_p$, which essentially means that the life of the material mainly consists in the propagation period. However, this assumption isn't supported by studies results and is mostly a simplicity argument. But it should be investigated whether or not the distribution of short initiation failure mode may be reduced to the distribution of the propagation period N_p.

2. Analogously under very low stress, failure occurs at very high number of cycles. The life to failure then consists mainly in the initiation period and the propagation period is way shorter, i.e. $N_p \ll N_i$. However then again, assuming that the propagation time is negligible when studying very low stresses should be investigated in order to provide an adequate modelling.

3. Another issue pertains to the mixture zone. It is difficult to evaluate from which level of stress or deformation the population becomes heterogeneous, i.e. we go from a single component to a mixture distribution. Moreover, we do observe that there are at least two components in this region, but there also might be more fracture modes at stake. We need to be able to determine the number of components of the distribution of N at any stress or deformation level.

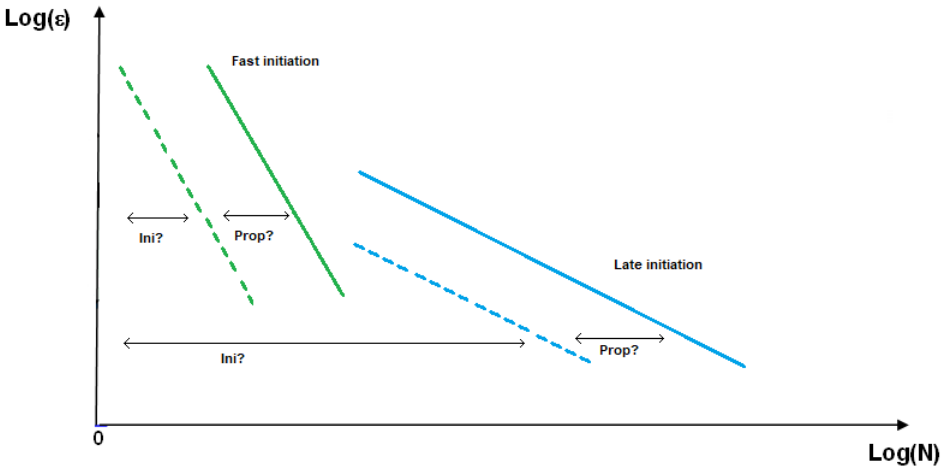


FIGURE 3.6 – Initiation and propagation mechanisms according to the failure mode

It follows that several modelling are possible, depending on how the above questions are addressed:

$$F(x) = \pi(\epsilon_a) G_p(x) + (1 - \pi(\epsilon_a)) H_{N_i+N_p}(x), \tag{3.12}$$

$$F(x) = \pi(\epsilon_a) H_{N_i+N_p}(x) + (1 - \pi(\epsilon_a)) F_i(x), \tag{3.13}$$

$$F(x) = \pi(\epsilon_a) H_{N_i+N_p^{(1)}}(x) + (1 - \pi(\epsilon_a)) H_{N_i+N_p^{(2)}}(x), \tag{3.14}$$

$$F(x) = \pi_1(\epsilon_a) G_p(x) + \pi_2(\epsilon_a) H_{N_i+N_p}(x) + (1 - \pi_1(\epsilon_a) - \pi_2(\epsilon_a)) F_i(x). \tag{3.15}$$

The first challenge to construct adequate S-N curves is to provide an adequate model which evolves as the stress decreases. Therefore the next chapters focus on proposing modelling tools to discriminate between models (3.12) to (3.15). Two major issues are tackled: the determination of the distribution of the components in the regions of the curve where the population is homogeneous and the number of components in the mixture zone.

Chapitre 4

Composite Tests Under Corrupted Data

4.1 Introduction

A situation which is commonly met in quality control is the following: some characteristic Z of an item is supposed to be random, and a decision about its distribution has to be done based on a sample of such items. However this variable is affected by a measurement error, a random noise V_δ, such that the observations do not consist in realizations of Z but of a variable $X := Z + V_\delta$.
Denote F_0 and G_0 the competing distribution functions (of respective densities f_0 and g_0) for Z, and H_δ the distribution function of the error V_δ with density h_δ. V_δ is assumed to be a transformation of a random variable V of distribution H, typically $V_\delta := \sqrt{\delta}V$. Its realizations are assumed to be mutually independent and independent on the item.
Therefore the density of the measurement X is either $f_\delta := f_0 * h_\delta$ or $g_\delta := g_0 * h_\delta$ where $*$ denotes the convolution operation. We denote F_δ (resp. G_δ) the distribution function with density f_δ (resp. g_δ).
The problem of interest studied in Broniatowski and al. (2018 [11]) is how the measurement errors can affect the conclusion of the likelihood ratio test with statistics

$$L_n := \frac{1}{n}\sum \log \frac{g_0}{f_0}(X_i).$$

For small δ, the result of Guo (2009 [37]) enables to estimate the true log-likelihood ratio (true Kullback-Leibler divergence) even when we only dispose of locally perturbed data by additive measurement error. The distribution function H of measurement errors is considered unknown, up to zero expectation and unit variance. When we use the likelihood ratio test while ignoring the possible measurement errors, we can incur a loss in

both errors of the first and second kind. However, it is shown in [11] that for small δ the original likelihood ratio test (LRT) is still most powerful, only on a slightly changed significance level. The test problem leads to a composite null and alternative classes $\mathbf{H}_0$ or $\mathbf{H}_1$ of distributions of random variables $Z + V_\delta$ with $V_\delta := \sqrt{\delta} V$. If those families are bounded by alternating Choquet capacities of order 2, then the minimax test is based on the likelihood ratio of the pair of the least favorable distributions of $\mathbf{H}_0$ and $\mathbf{H}_1$, respectively (see Huber and Strassen, 1973 [38]). Moreover, Eguchi and Copas (2005 [32]) showed that the overall loss of power caused by a misspecified alternative equals to the Kullback-Leibler divergence between the original and the corrupted alternatives. Surprisingly, the value of the overall loss is independent of the choice of null hypothesis. The arguments of Guo [37] and of Narayanan and Srinivasa (2007 [50]) enable to approximate the loss of power locally for a broad set of alternatives. The asymptotic behavior of the loss of power of the test based on sampled data is considered in [11], and supplemented with numerical illustration.

4.2 Statement of the test problem

The aim is to propose a class of statistics for testing the composite hypotheses $\mathbf{H}_0$ and $\mathbf{H}_1$, extending the optimal Neyman-Pearson LRT between f_0 and g_0. Unlike in [11], the scaling parameter δ is not supposed to be small, but merely to belong to some interval bounded away from 0.

We assume that the distribution H of the random variable (r.v.) V is known. In a metrology setting, this is not such a strong assumption, since in the tuning of the offset of a measurement device, it is customary to perform a large number of observations on the noise under controlled environment.

Therefore this first step produces a good basis for the modelling of the distribution of the density h. Although the distribution of V is known, under operational conditions the distribution of the noise is modified: for given δ in $[\delta_{\min}, \delta_{\max}]$ with $\delta_{\min} > 0$, denote V_δ a r.v. whose distribution is obtained through some transformation from the distribution of V which quantifies the level of the random noise. Some classical example is when $V_\delta = \sqrt{\delta} V$, but at times we can have some weaker assumption which amounts to some decomposability property with respect to δ: for instance, in the Gaussian case we assume that for all δ, η, there exists some r.v. $W_{\delta,\eta}$ such that $V_{\delta+\eta} =_d V_\delta + W_{\delta,\eta}$, where V_δ and $W_{\delta,\eta}$ are independent.

The test problem can be stated as follows: a batch of n iid measurements $X_i := Z_i + V_{\delta,i}$

is performed, where $\delta > 0$ is unknown, and we consider the family of tests of

$$\begin{aligned} \mathbf{H}_0(\delta) := &\quad \{X \text{ has density } f_\delta\} \\ &\quad \text{vs} \\ \mathbf{H}_1(\delta) := &\quad \{X \text{ has density } g_\delta\} \end{aligned}$$

with $\delta \in \Delta = [\delta_{\min}, \delta_{\max}]$. A class of combined test of $\mathbf{H}_0$ vs $\mathbf{H}_1$ is proposed, in the spirit of Bahadur (1960 [2]), Bahadur (1971 [3], Tusnády (1987 [55]) and Birgé (1981 [7]).
Under every fixed n, we assume that δ is allowed to run over a finite p_n components of the vector $\Delta_n := [\delta_{\min} = \delta_{0,n}, .., \delta_{p_n,n} = \delta_{\max}]$. The present construction is essentially non asymptotic, neither on n nor on δ, in contrast with [11] where δ was supposed to lie in a small neighborhood of 0. However, with increasing n it would be useful to consider that the array $\left(\delta_{j,n}\right)_{j=1}^{p_n}$ is getting dense in $\Delta = [\delta_{\min}, \delta_{\max}]$ and that

$$\lim_{n\to\infty} \frac{\log p_n}{n} = 0. \tag{4.1}$$

For the sake of notational brevity, we denote Δ by the above grid Δ_n and all suprema or infima over Δ are supposed to be over Δ_n. For any event B and any δ in Δ , $F_\delta(B)$ (resp. $G_\delta(B)$) designates the probability of B under distribution F_δ (resp. G_δ). Given a sequence of levels α_n, we consider a sequence of test criteria $T_n := T_n(X_1, .., X_n)$ of $\mathbf{H}_0(\delta)$, and the pertaining critical regions

$$T_n(X_1, .., X_n) > A_n \tag{4.2}$$

such that

$$F_\delta(T_n(X_1, .., X_n) > A_n) \le \alpha_n \quad \forall \delta \in \Delta,$$

leading to rejection of $\mathbf{H}_0(\delta)$ for at least some $\delta \in \Delta$.
In an asymptotic context, it is natural to assume that α_n converges to 0 as n increases, since an increase in the sample size allows for a smaller first kind risk. For example in [7], α_n takes the form $\alpha_n := \exp\{-na_n\}$ for some sequence $a_n \to \infty$.
In the sequel the Kullback-Leibler discrepancy between probability measures Q and P with respective densities p and q with respect to the Lebesgue measure on $\mathbb{R}$ is denoted

$$K(Q, P) := \int \log \frac{q(x)}{p(x)} q(x) dx$$

whenever defined, and takes value $+\infty$ otherwise.
This chapter handles some issues with respect to this context. In Section 4.3 we consider some test procedure based on the supremum of Likelihood Ratios (LR) for various values of δ, and define T_n. The threshold for such a test is obtained for any level α_n and some

lower bound for its power is provided. In Section 4.4 we develop an asymptotic approach to the least favorable hypotheses (LFH) for these tests and prove that asymptotically least favorable hypotheses are obtained through minimization of the Kullback-Leibler divergence between the two composite classes **H0** and **H1** independently upon the level of the test.
Section 4.4.3 considers the performances of the test numerically; indeed the numerical power of the test under the least favorable couple of hypotheses is compared with the theoretical lower bound as obtained in Section 4.3. We show on several examples that the minimal power measured under the LFH is indeed larger than the theoretical lower bound; this result shows that simulation results overperform theoretical bounds.
Since no argument plays in favor of any type of optimality for the test based on the supremum of Likelihood Ratios for composite testing, we consider to substitute those ratios by some other kinds of scores, in the family of divergence based concepts, extending the Likelihood Ratio in a natural way. Such an approach has already been extensively treated, starting with Liese and Vajda (1987 [42]). Extensions of the Kullback-Leibler based criterions (such as the Likelihood Ratio) to power type criterions have been proposed for many applications in Physics and in Statistics; see e.g. Tsallis (1987 [54]). We explore the properties of those new tests under the couple of hypotheses minimizing the Kullback-Leibler divergence between the two composite classes **H0** and **H1**. We show that in some cases we can build a test procedure whose properties overperform the above supremum of the LRTs, and we provide some explanation for this fact. This is the scope of Section 4.5.
Lastly, in Section 4.6, the test procedure is adapted to testing a simple hypothesis against a composite one. It corresponds to an industrial application to fatigue life data.

4.3 An extension of the Likelihood Ratio test

For any δ in Δ, let

$$T_{n,\delta} := \frac{1}{n}\sum_{i=1}^{n} \log \frac{g_\delta}{f_\delta}(X_i), \tag{4.3}$$

and define

$$T_n := \sup_{\delta\in\Delta} T_{n,\delta}.$$

Consider for fixed δ the Likelihood Ratio Test with statistics $T_{n,\delta}$ which is uniformly most powerful (UMP) within all tests of $\mathbf{H0}(\delta) := p_T = f_\delta$ vs $\mathbf{H1}(\delta) := p_T = g_\delta$ where p_T designates the distribution of the generic r.v. X . The test procedure to be discussed aims at solving the question : does there exist some δ for which $\mathbf{H0}(\delta)$ would be rejected vs

H1(δ), for some prescribed value of the first kind risk?
Whenever **H0**(δ) is rejected in favor of **H1**(δ) for some δ we reject **H0**:=$f_0 = g_0$ in favor of **H1**:=$f_0 \neq g_0$. A critical region for this test with level α_n is defined through $\{T_n > A_n\}$ with

$$\begin{aligned} P_{\mathbf{H0}}(\mathbf{H1}) &= \sup_{\delta\in\Delta} F_\delta\left(T_n > A_n\right) \\ &= \sup_{\delta\in\Delta} F_\delta\left(\bigcup_{\delta'} T_{n,\delta'} > A_n\right) \leq \alpha_n. \end{aligned}$$

Since for any sequence of events $B_1, \ldots, B_{p_n}$,

$$F_\delta\left(\bigcup_{k=1}^{p_n} B_k\right) \leq p_n \max_{1\leq k\leq p_n} F_\delta\left(B_k\right),$$

it holds

$$P_{\mathbf{H0}}(\mathbf{H1}) \leq p_n \max_{\delta\in\Delta}\max_{\delta'\in\Delta} F_\delta\left(T_{n,\delta'} > A_n\right). \tag{4.4}$$

An upper bound for $P_{\mathbf{H0}}(\mathbf{H1})$ can be obtained making use of the Chernoff inequality for the right side of (4.4), providing an upper bound for the risk of first kind for a given A_n. The correspondence between A_n and this risk allows to define the threshold A_n accordingly.
Turning to the power of this test we define the risk of second kind through the crude bound

$$\begin{aligned} P_{H_1}(\mathbf{H0}) &:= \sup_{\eta\in\Delta} G_\eta\left(T_n \leq A_n\right) \\ &= \sup_{\eta\in\Delta} G_\eta\left(\sup_{\delta\in\Delta} T_{n,\delta} \leq A_n\right) \\ &= \sup_{\eta\in\Delta} G_\eta\left(\bigcap_{\delta\in\Delta} T_{n,\delta} \leq A_n\right) \\ &\leq \sup_{\eta\in\Delta} G_\eta\left(T_{n,\eta} \leq A_n\right). \end{aligned} \tag{4.5}$$

The last term in (4.5) can be bounded from above through the Chernoff inequality, which yields a lower bound for the minimal power of the test under any hypothesis g_η in **H1**.
Let α_n denote a sequence of levels such that

$$\limsup_{n\to\infty} \alpha_n < 1.$$

We make use of the following hypothesis:

$$\inf_{\delta\in\Delta}\inf_{\delta'\in\Delta}\int \log\frac{f_{\delta'}}{g_{\delta'}} f_\delta > 0. \tag{4.6}$$

Remark 4. Since

$$\int \log\frac{f_{\delta'}}{g_{\delta'}} f_\delta = K(F_\delta, G_{\delta'}) - K(F_\delta, F_{\delta'}),$$

hypothesis (4.6) means that the classes of distributions $(F_\delta)_\delta$ and $(G_\delta)_\delta$ are well separated in the sense of Kullback-Leibler discrepancy. Making use of the Chernoff-Stein Lemma (see Theorem 8 in the Appendix), hypothesis (4.6) entails that any LRT with H0: $p_T = f_\delta$ vs H1: $p_T = g_{\delta'}$ is asymptotically more powerful than any LRT with H0: $p_T = f_\delta$ vs H1: $p_T = f_{\delta'}$.

Both hypotheses (4.7) and (4.8) hereunder are used to provide the critical region and the power of the test.
For all δ, δ' define

$$Z_{\delta'} := \log\frac{g_{\delta'}}{f_{\delta'}}(X)$$

and let

$$\varphi_{\delta,\delta'}(t) := \log E_{F_\delta}\left(\exp(tZ_{\delta'})\right) = \log\int\left(\frac{g_{\delta'}(x)}{f_{\delta'}(x)}\right)^t f_\delta(x)dx.$$

With $\mathcal{N}_{\delta,\delta'}$, the set of all t such that $\varphi_{\delta,\delta'}(t)$ is finite, we assume

$$\mathcal{N}_{\delta,\delta'} \text{ is a non void open neighborhood of } 0. \tag{4.7}$$

Define further

$$J_{\delta,\delta'}(x) := \sup_t\left(tx - \varphi_{\delta,\delta'}(t)\right)$$

and let

$$J(x) := \min_{(\delta,\delta')\in\Delta\times\Delta} J_{\delta,\delta'}(x).$$

For any η, let

$$W_\eta := -\log\frac{g_\eta}{f_\eta}(x)$$

and let

$$\psi_\eta(t) := \log E_{G_\eta}\left(\exp\left(tW_\eta\right)\right).$$

Let $\mathcal{M}_\eta$ be the set of all t such that $\psi_\eta(t)$ is finite. Assume

$$\mathcal{M}_\eta \text{ is a non void neighborhood of } 0. \tag{4.8}$$

Let

$$I_\eta(x) := \sup_t tx - \log E_{G_\eta}\left(\exp\left(tW_\eta\right)\right) \tag{4.9}$$

and

$$I(x) := \inf_\eta I_\eta(x).$$

We also assume some accessory condition on the support of $Z_{\delta'}$ and W_η respectively under F_δ and under G_η; see (4.17) and (4.20) in the proof of Theorem 8. Suppose the regularity assumptions (4.7) and (4.8) fulfilled for all δ, δ' and η. Assume further that p_n fulfills (4.1).
The following result holds.

Proposition 5. Whenever (4.6) holds, for any sequence of levels α_n bounded away from 1, defining

$$A_n := J^{-1}\left(-\frac{1}{n}\log\frac{\alpha_n}{p_n}\right),$$

it holds, for large n,

$$P_{\mathbf{H0}}(\mathbf{H1}) = \sup_{\delta\in\Delta} F_\delta\left(T_n > A_n\right) \le \alpha_n$$

and

$$P_{\mathbf{H1}}(\mathbf{H1}) = \sup_{\delta\in\Delta} G_\delta\left(T_n > A_n\right) \ge 1 - \exp\left(-nI\left(A_n\right)\right).$$

4.4 Minimax tests under noisy data, least favorable hypotheses

4.4.1 An asymptotic definition for the least favorable hypotheses

We prove that the above procedure is asymptotically minimax for testing the composite hypothesis $\mathbf{H}_0$ against the composite alternative $\mathbf{H}_1$; indeed we identify the least favorable hypotheses, say $F_{\delta_*} \in \mathbf{H}_0$ and $G_{\delta_*} \in \mathbf{H}_1$, which lead to minimal power and maximal first kind risk for these tests. This requires a discussion on the definition and existence of such least favourable couple of hypotheses in an asymptotic context; indeed for fixed sample size the usual definition only leads to an explicit definition in very specific cases. Unlike in [11], the minimax tests will not be in the sense of Huber and Strassen. Indeed, on one hand, hypotheses $\mathbf{H}_0$ and $\mathbf{H}_1$ are not defined in topological neighbourhoods of F_0 and G_0, but rather through a convolution under a parametric setting; on the other hand, the specific test of $\{\mathbf{H}_0(\delta),\ \delta \in \Delta\}$ against $\{\mathbf{H}_1(\delta),\ \delta \in \Delta\}$ does not require capacities dominating the corresponding probability measures.

Throughout the subsequent text we shall assume that there exists δ_* such that

$$\min_{\delta\in\Delta} K\left(F_\delta, G_\delta\right) = K\left(F_{\delta_*}, G_{\delta_*}\right). \tag{4.10}$$

We shall call the pair of distributions $\left(F_{\underline{\delta}},\, G_{\underline{\delta}}\right)$ a least favorable for the sequence of tests of critical region $\mathbb{1}\{T_n > A_n\}$ if it satisfies

$$\begin{aligned} F_\delta\left(T_n \le A_n\right) &\ge F_{\underline{\delta}}\left(T_n \le A_n\right) \\ &\ge G_{\underline{\delta}}\left(T_n \le A_n\right) \ge G_\delta\left(T_n \le A_n\right) \end{aligned} \tag{4.11}$$

for all $\delta \in \Delta$. The condition of unbiasedness of the test is captured with the central inequality in (4.11).

Because under a finite n such a pair can be constructed only in few cases, we should take a recourse of (4.11) to the asymptotics $n \to \infty$. We shall show that any pair of distributions $\left(F_{\delta_*} G_{\delta_*}\right)$ achieving (4.10) be named least favorable. Indeed, it satisfies the inequality (4.11) asymptotically on the logarithmic scale.

Specifically, we say that $\left(F_{\underline{\delta}}, G_{\underline{\delta}}\right)$ is a least favorable pair of distributions when for any $\delta \in \Delta$

$$\begin{aligned} \lim\inf_{n\to\infty} \frac{1}{n} \log F_{\underline{\delta}}\left(T_n \le A_n\right) &\ge \lim_{n\to\infty} \frac{1}{n} \log G_{\underline{\delta}}\left(T_n \le A_n\right) \\ &\ge \lim_{n\to\infty} \sup \frac{1}{n} \log G_\delta\left(T_n \le A_n\right). \end{aligned} \tag{4.12}$$

Define the total variation distance

$$d_{TV}\left(F_\delta, G_\delta\right) := \sup_B \left|F_\delta(B) - G_\delta(B)\right|$$

where the supremum is over all Borel sets B of $\mathbb{R}$. We will assume that for all n

$$\alpha_n < 1 - \sup_{\delta\in\Delta} d_{TV}\left(F_\delta, G_\delta\right). \tag{4.13}$$

We state our main result, whose proof is deferred to the Appendix.

Theorem 6. For any level α_n satisfying (4.13) the couple $\left(F_{\delta_*}, G_{\delta_*}\right)$ is a least favorable couple of hypotheses for the family of tests $1\{T_n \ge A_n\}$ in the sense of (4.12).

4.4.2 Identifying the least favorable hypotheses

We now concentrate on (4.10).
The following results state that the Kullback-Leibler discrepancy $K(F_\delta, G_\delta)$ reaches its minimal value when the noise V_δ is "maximal", under some additivity property with respect to δ. This result is not surprising: adding noise deteriorates the ability to discriminate between the two distributions F_0 and G_0; this effect is captured in $K(F_\delta, G_\delta)$, which takes its minimal value for the maximal δ.

Proposition 7. Assume that for all δ, η, there exists some r.v $W_{\delta,\eta}$ such that $V_{\delta+\eta} =_d V_\delta + W_{\delta,\eta}$ where V_δ and $W_{\delta,\eta}$ are independent. Then

$$\delta_* = \delta_{\max}.$$

This result holds as a consequence of Lemma 12 in the Appendix.
In the Gaussian case, when h is the standard normal density, Proposition 7 holds since $h_{\delta+\eta} = h_\delta * h_{\eta-\delta}$ with $h_\varepsilon(x) := \left(1/\sqrt{\varepsilon}\right) h\left(x/\sqrt{\varepsilon}\right)$. In order to model a symmetric noise we may consider a symmetrized Gamma density as follows: set

$$h_\delta(x) := (1/2)\, 1^+(1,\delta)(x) + (1/2)\, 1^-(1,\delta)(x)$$

where $1^+(1,\delta)$ designates the Gamma density with scale parameter 1 and shape parameter δ and $1^-(1,\delta)$, the Gamma density on $\mathbb{R}^-$ with same parameter. Hence a r.v. with density h_δ is symmetrically distributed and has variance 2δ. In this case $h_{\delta+\eta}(x) = h_\delta * h_\eta(x)$ and thus 7 also holds. Note that for values of δ less than or equal to 1, the density h_δ is bimodal, which does not play in favour of such densities for modelling the uncertainty due to the noise; in contrast with the Gaussian case, h_δ cannot be obtained from h_1 by any scaling. The centered Cauchy distribution may help as a description of heavy tailed symmetric noise and keeps uni-modality through convolution; it satisfies the requirements of Proposition 7 since $f_\delta * f_\eta(x) = f_{\delta+\eta}(x)$ where $f_\varepsilon(x) := \varepsilon/\pi\left(x^2+\varepsilon^2\right)$. In this case δ acts as a scaling since f_δ is the density of δX where X has density f_1.
In practice the interesting case is when δ is the variance of the noise and corresponds to a scaling of a generic density, as occurs for the Gaussian case or for the Cauchy case. In the examples which will be used hereunder we also consider symmetric exponentially distributed densities (Laplace densities) or symmetric Weibull densities with given shape parameter. The Weibull distribution also fulfills the condition in Proposition 7, being infinitely divisible (see [36]).

4.4.3 Numerical performance of the minimax test

As frequently observed numerical results deduced from theoretical bounds are of poor interest, which is due to the sub-optimality of the involved inequalities, they may be sharpened on specific cases. This motivates the need for simulations. We consider two cases which can be considered as benchmarks.

A. In the first case f_0 is a normal density with expectation 0 and variance 1, whereas g_0 is a normal density with expectation 0.3 and variance 1.

B. The second case handles a situation where f_0 and g_0 belong to different models: f_0 is a lognormal density with location parameter 1 and scale parameter 0.2, whereas g_0 is a Weibull density on $\mathbb{R}^+$ with shape parameter 5 and scale parameter 3. Those two densities differ strongly in terms of asymptotic decay. They are however very close one to the other in terms of their symmetrized Kullback-Leibler divergence (so-called Jeffrey distance). Indeed centering on the log normal distribution f_0, the closest among all Weibull densities is at distance 0.10; the density g_0 is at distance 0.12 from f_0.

Both cases are treated considering four types of distribution for the noise:

a. the noise h_δ is a centered normal density with variance δ^2.

b. the noise h_δ is a centered Laplace density with parameter $\lambda(\delta)$

c. the noise h_δ is a symmetrized Weibull density with shape parameter 1.5 and variable scale parameter $\beta(\delta)$

d. the noise h_δ is Cauchy with density $h_\delta(x) = 1(\delta)/\pi\left(1(\delta)^2 + x^2\right)$.

In order to compare the performance of the test under those four distributions, we have adopted the following rule: the parameter of the distribution of the noise is tuned such that for each value $\underline{\delta}$, it holds $P\left(\left|V_{\underline{\delta}}\right| > \underline{\delta}\right) = \Phi(1) - \Phi(-1) \sim 0.65$, where Φ stands for the standard Gaussian cumulative function. Thus, distributions b to d are scaled with respect to the Gaussian noise with variance δ^2.

In both cases A and B the range of δ is $\Delta = (\delta_{\min} = 0.1, \delta_{\max})$ and we have selected a number of possibilities for $\delta_{\max}$, ranging from 0.2 to 0.7.

In case A we selected $\delta^2_{\max} = 0.5$ which is a signal-to-noise ratio equal to 0.7, a commonly chosen bound in quality control tests.

In case B the variance of f_0 is roughly 0.6 and the variance of g_0 is roughly 0.4. The maximal value of $\delta^2_{\max}$ is roughly 0.5. This is thus a maximal upper bound for a practical modelling.

We present some power functions making use of the theoretical bounds together with the corresponding ones based on simulation runs. As seen, the performance of the theo-

retical approach is weak; we have focused on simulation, after some comparison with the theoretical bounds.

Case A: the shift problem

In this subsection we evaluate the quality of the theoretical power bound defined in the previous sections. Thus we compare the theoretical formula to the empirical lower performance obtained through simulations under the least favorable hypotheses.

Theoretical power bound

While supposedly valid at finite n, the theoretical power bound given by (4.23) still assumes some sort of asymptotics, since a good approximation of the bound entails a fine discretization of Δ to compute $I(A_n) = \inf_{\eta\in\Delta_n} I_\eta(A_n)$. Thus, by condition (4.1), n has to be large. Therefore, in the following, we will compute this lower bound for n sufficiently large, that is, at least 100 observations, which is also consistent with industrial applications.

Numerical power bound

In order to obtain a minimal bound of power for the composite test, we compute the power of the test $\mathbf{H}_0(\delta_*)$ against $\mathbf{H}_1(\delta_*)$ where δ^* defines the couple of LFH's $\left(F_{\delta_*}, G_{\delta_*}\right)$. Following Proposition 7, the LF hypotheses for the test defined by T_n^1 when the noise follows a Gaussian, a Cauchy or a symmetrized Weibull distribution is achieved for $\left(F_{\delta_{\max}}, G_{\delta_{\max}}\right)$.
When the noise follows a Laplace distribution, the couple of LF hypotheses is the one that satisfies:

$$\left(F_{\delta_*}, G_{\delta_*}\right) = \arg \min_{(F_\delta, G_\delta), \delta\in\Delta_n} K(F_\delta, G_\delta) \tag{4.14}$$

In both cases A and B, this condition is also satisfied for $\delta^* = \delta_{\max}$

Comparison of the two power curves

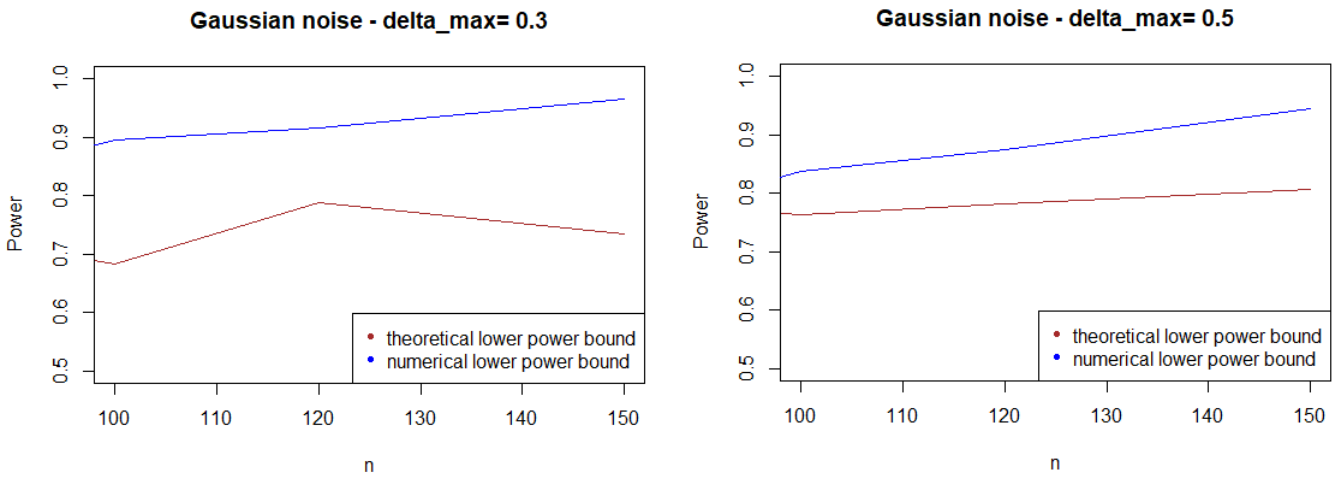


FIGURE 4.1 – Theoretical and numerical power bound of the test of case A under Gaussian noise with respect to n for a first kind risk $\alpha = 0.05$

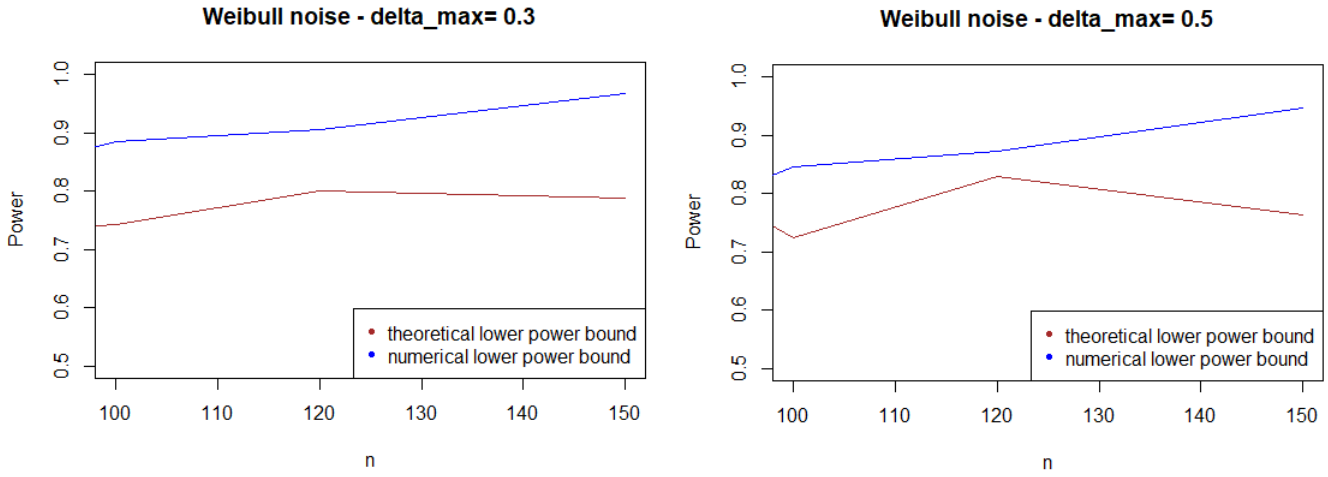


FIGURE 4.2 – Theoretical and numerical power bound of the test of case A under symmetrized Weibull noise with respect to n for a first kind risk $\alpha = 0.05$

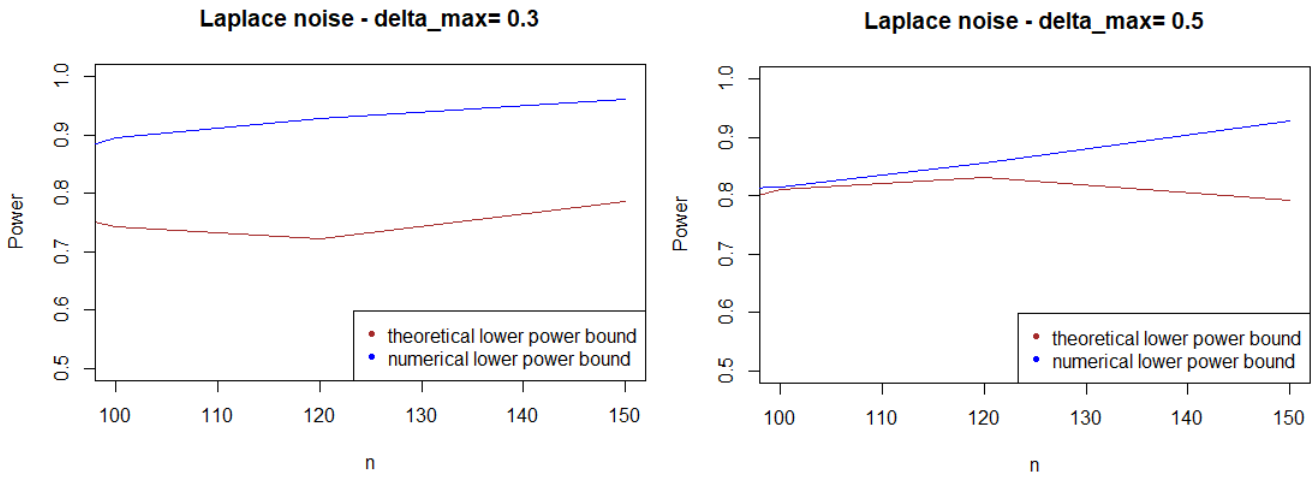


FIGURE 4.3 – Theoretical and numerical power bound of the test of case A under a symmetrized Laplacian noise with respect to n for a first kind risk $\alpha = 0.05$

As expected, Figures 4.1 to 4.3 show that the theoretical lower bound is always under the empirical lower bound when n is high enough to provide a good approximation of $I(A_n)$. This is also true when the noise follows a Cauchy distribution, but for a bigger sample size than in the figures above ($n > 250$).

In most cases, the theoretical bound tends to largely underestimate the power of the test, when compared to its minimal performance given by simulations under the least favorable hypotheses. The gap between the two also tends to increase as n grows. This result may be explained by the large bound provided by (4.5), while the numerical performance are obtained with respect to the least favorable hypotheses.

On a computational perspective, the computational cost of the theoretical bound is way higher than its numeric counterpart.

Case B: the tail thickness problem

The calculation of the moment generating function appearing in the formula of $I_\eta(x)$ in (4.9) is numerically unstable, which renders the computation of the theoretical bound impossible. Thus, in the following sections, the performance of the test will be evaluated numerically through Monte Carlo replications.

4.5 Some alternative statistics for testing

4.5.1 A family of composite tests based on divergence distances

This Section provides a similar treatment as above, dealing now with some extension of the LRT test to the same composite setting. The class of tests is related to the divergence

based approach to testing, and it includes the cases considered so far. For reasons developed in Section 4.4.3 we argue through simulation and do not develop the corresponding Large Deviation approach.

The statistics T_n can be generalized in a natural way, defining a family of tests depending on some parameter γ. For $\gamma \neq 0, 1$, let

$$\phi_\gamma(x) := \frac{x^\gamma - \gamma x + \gamma - 1}{\gamma(\gamma - 1)}$$

a function defined on $(0, \infty)$ with values in $(0, \infty)$, setting

$$\phi_0(x) := -\log x + x - 1$$

and

$$\phi_1(x) := x \log x - x + 1.$$

For $\gamma \leq 2$ this class of functions is instrumental in order to define the so-called power divergences between probability measures, a class of pseudo-distances widely used in statistical inference; see for example [6].

Associated to this class consider the function

$$\begin{aligned} \varphi_\gamma(x) &:= -\frac{d}{dx}\phi_\gamma(x) \\ &= \frac{1 - x^{\gamma-1}}{\gamma - 1} \text{ for } \gamma \neq 0, 1 \end{aligned}$$

and we also consider

$$\varphi_1(x) := -\log x$$
$$\varphi_0(x) := \frac{1}{x} - 1$$

from which the statistics

$$T^\gamma_{n,\delta} := \frac{1}{n} \sum_{i=1}^{n} \varphi_\gamma(X_i)$$

and

$$T^\gamma_n := \sup_\delta T^\gamma_{n,\delta}$$

are well defined for all $\gamma \leq 2$. Figure 4.4 illustrates the functions φ_γ according to γ.

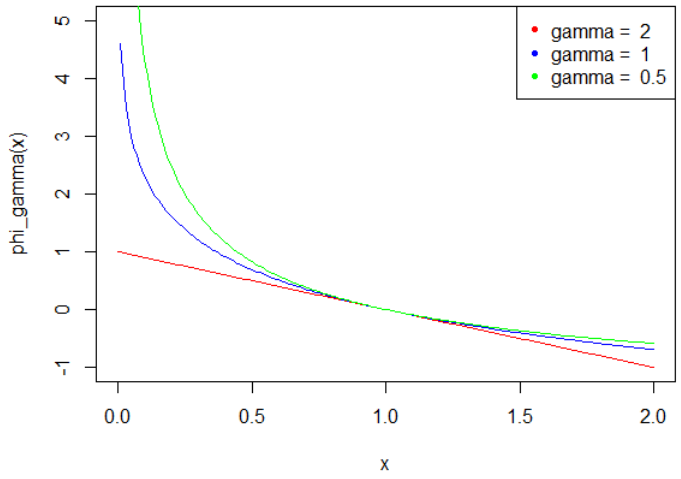


FIGURE 4.4 – φ_γ for $\gamma = 0.5, 1$ and 2

Fix a risk of first kind α and the corresponding power of the LRT pertaining to $H0(\delta_*)$ vs $H1(\delta_*)$ through

$$1-\beta := G_{\delta_*}\left(T^1_{n,\delta_*} > s_\alpha\right)$$

with

$$s_\alpha := \inf\left\{s : F_{\delta_*}\left(T^1_{n,\delta_*} > s\right) \le \alpha\right\}.$$

Define accordingly the power of the test based on T^γ_n under the same hypotheses,

$$s^\gamma_\alpha := \inf\left\{s : F_{\delta_*}\left(T^\gamma_n > s\right) \le \alpha\right\}$$

and

$$1-\beta' := G_{\delta_*}\left(T^\gamma_n > s^\gamma_\alpha\right).$$

Firstly δ_* defines the couple of hypotheses $\left(F_{\delta_*}, G_{\delta_*}\right)$ such that the LRT with statistics T^1_{n,δ_*} has maximal power among all tests $H0(\delta_*)$ vs $H1(\delta_*)$. Furthermore, by Theorem 8 it has minimal power on the logarithmic scale among all tests $H0(\delta)$ vs $H1(\delta)$.
On the other hand $\left(F_{\delta_*}, G_{\delta_*}\right)$ is the LF couple for the test with statistics T^1_n among all couples (F_δ, G_δ).
These two facts allow for the definition of the loss of power making use of T^1_n instead of T^1_{n,δ_*} for testing $H0(\delta_*)$ vs $H1(\delta_*)$. This amounts to consider the price of aggregating the local tests $T^1_{n,\delta}$, a necessity since the true value of δ is unknown. A natural indicator

for this loss consists in the difference

$$\Delta_n^1 := G_{\delta_*}\left(T^1_{n,\delta_*} > s_\alpha\right) - G_{\delta_*}\left(T^1_n > s^1_\alpha\right) \geq 0.$$

Consider now an aggregated test statistics T_n^γ. We do not have at hand a similar result as in Proposition 5. We thus consider the behavior of the test $H0(\delta_*)$ vs $H1(\delta_*)$ although $\left(F_{\delta_*}, G_{\delta_*}\right)$ may not be a LFH for the test statistics T_n^γ. The heuristics which we propose makes use of the corresponding loss of power with respect to the LRT through

$$\Delta_n^\gamma := G_{\delta_*}\left(T^1_{n,\delta_*} > s_\alpha\right) - G_{\delta_*}\left(T^\gamma_n > s^\gamma_\alpha\right).$$

We will see that it may happen that Δ_n^γ improves over Δ_n^1. We define the optimal value of γ , say γ^*, such that

$$\Delta_n^{\gamma^*} \leq \Delta_n^\gamma$$

for all γ.

In the various figures hereunder, NP corresponds to the LRT defined between the LFH's $\left(F_{\delta_*}, G_{\delta_*}\right)$, KL to the test with statistics T_n^1 (hence as presented Section 4.3), HELL corresponds to $T_n^{1/2}$ which is associated to the Hellinger power divergence, and G=2 corresponds to $\gamma = 2$.

4.5.2 A practical choice for composite tests based on simulation

We consider the same cases A and B as described in Section 4.4.3.

As stated in the previous section, the performances of the different test statistics are compared considering the test of $\mathbf{H}_0(\delta_*)$ against $\mathbf{H}_1(\delta_*)$ where δ^* is defined as explained in section 4.4.3 as the LF hypotheses for the test T_n^1. In both cases A and B, this corresponds to $\delta^* = \delta_{\max}$.

Case A: the shift problem

Overall, the aggregated tests perform well when the problem consists in identifying a shift in a distribution. Indeed, for the three values of γ (0.5, 1 and 2), the power remains above 0.7 for any kind of noise and any value of δ_*. Moreover, the power curves associated to T_n^γ mainly overlap with the optimal test T^1_{n,δ_*}.

a. Under Gaussian noise, the power remains mostly stable over the values of δ_*, as shown by Figure 4.5. The tests with statistics T_n^1 and T_n^2 are equivalently powerful for large values of δ_*, while the first one achieves higher power when δ_* is small.

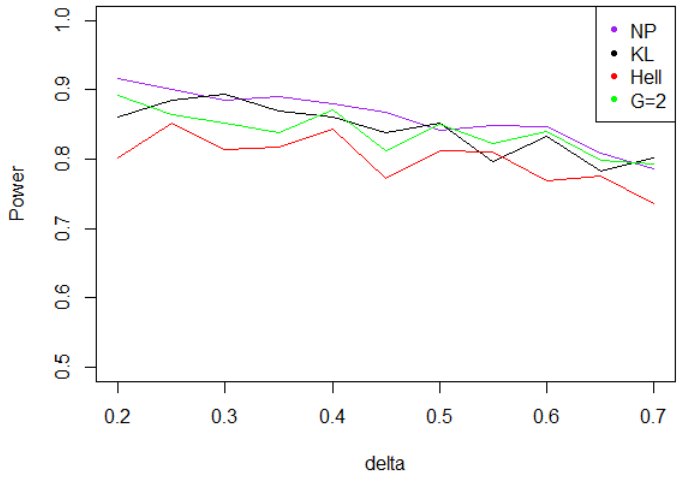


FIGURE 4.5 – Power of the test of case A under Gaussian noise with respect to $\delta_{\max}$ for a first kind risk $\alpha = 0.05$ and a sample size $n = 100$

b. When the noise follows a Laplace distribution, the three power curves overlap the NP power curve, and the different test statistics can be indifferently used. Under such a noise, the alternate hypotheses are extremely well distinguished by the class of tests considered, and this remains true as δ_* increases (cf. Figure 4.6).

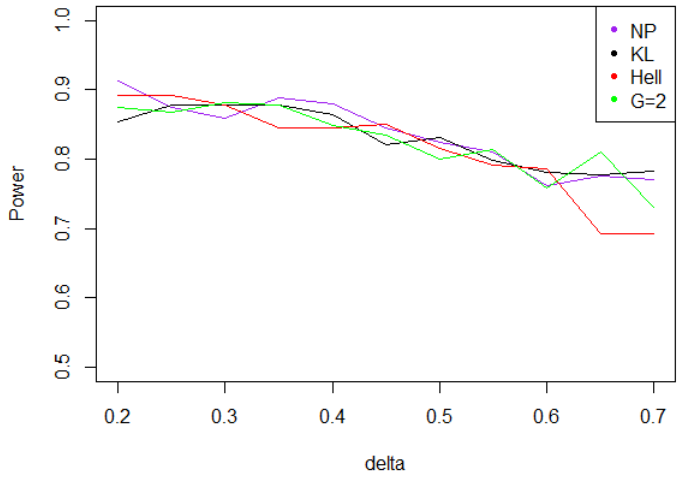


FIGURE 4.6 – Power of the test of case A under Laplacian noise with respect to $\delta_{\max}$ for a first kind risk $\alpha = 0.05$ and a sample size $n = 100$

c. Under the Weibull hypothesis, T_n^1 and T_n^2 perform similarly well and almost always as well as T_{n,δ_*}^1, while the power curve associated to $T_n^{1/2}$ remains below. Figure 4.7 illustrates that, as $\delta_{\max}$ increases, the power does not decrease much.

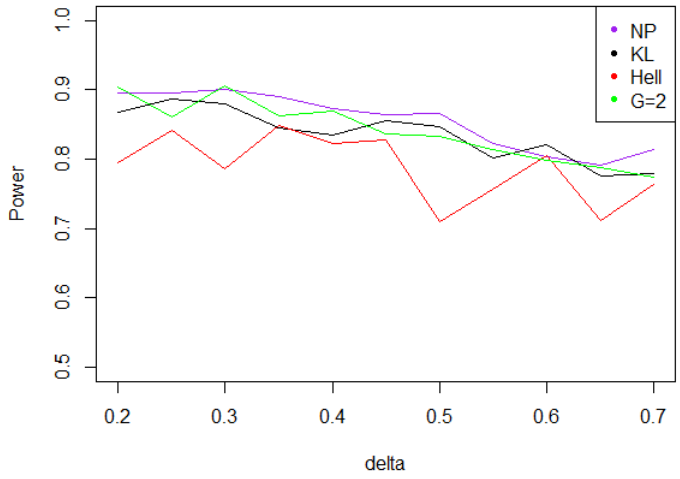


FIGURE 4.7 – Power of the test of case A under symmetrized Weibull noise with respect to $\delta_{\max}$ for a first kind risk $\alpha = 0.05$ and a sample size $n = 100$

d. Under a Cauchy assumption, the alternate hypotheses are less distinguishable than under any other parametric hypothesis on the noise, since the maximal power is about 0.84, while it exceeds 0.9 in cases a, b and c (cf. Figures 4.5 to 4.8). The capacity of the tests to discriminate between $\mathbf{H0}(\delta_{\max})$ and $\mathbf{H1}(\delta_{\max})$ is almost independent of the value of $\delta_{\max}$ and the power curves are mainly flat.

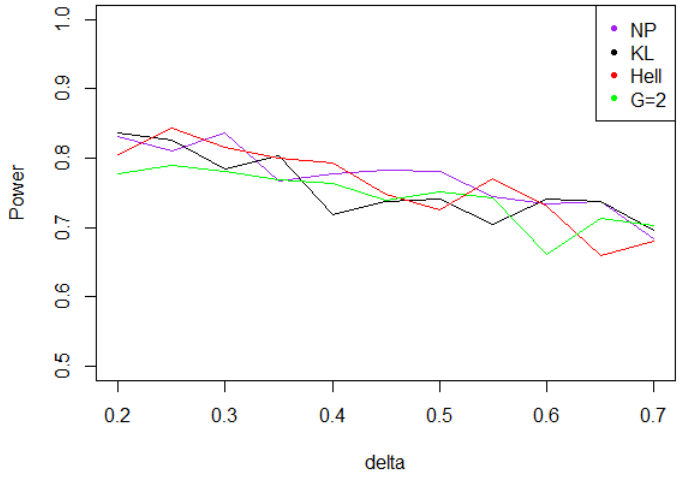


FIGURE 4.8 – Power of the test of case A under a noise following a Cauchy distribution with respect to $\delta_{\max}$ for a first kind risk $\alpha = 0.05$ and a sample size $n = 100$

Case B: the tail thickness problem

a. With the noise defined by case A (Gaussian noise), for KL ($\gamma = 1$), $\delta_* = \delta_{\max}$ due to Proposition 7 and statistics T_n^1 provides the best power uniformly upon $\delta_{\max}$. Figure 4.9 shows a net decrease of the power as $\delta_{\max}$ increases (recall that the power is evaluated under the least favorable alternative $G_{\delta_{\max}}$).

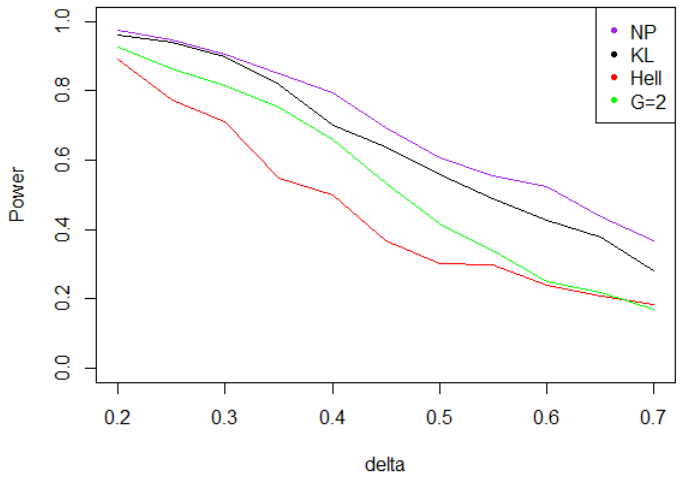


FIGURE 4.9 – Power of the test of case B under Gaussian noise with respect to $\delta_{\max}$ for a first kind risk $\alpha = 0.05$ and a sample size $n = 100$. The NP curve corresponds to the optimal Neyman Pearson test under $\delta_{\max}$. The KL, Hellinger and $G = 2$ curves stand respectively for $\gamma = 1, \gamma = 0.5$ and $\gamma = 2$ cases.

b. When the noise follows a Laplace distribution the situation is quite peculiar. For any value of δ in Δ , the modes $M_{G_{\delta_{\max}}}$ and $M_{F_{\delta_{\max}}}$ of the distributions of $\left(f_\delta / g_\delta\right)(X)$ under $G_{\delta_{\max}}$ and under $F_{\delta_{\max}}$ are quite separated, both larger than 1. Also for δ all the values of $\left|\phi_\gamma\left(M_{G_{\delta_{\max}}}\right) - \phi_\gamma\left(M_{F_{\delta_{\max}}}\right)\right|$ are quite large for large values of γ. We may infer that the distributions of $\phi_\gamma\left(\left(f_\delta / g_\delta\right)(X)\right)$ under $G_{\delta_{\max}}$ and under $F_{\delta_{\max}}$ are quite distinct for all δ, which in turn imply that the same fact holds for the distributions of T_n^γ for large γ. Indeed simulations presented in Figure 4.10 show that the maximal power of the test tends to be achieved when $\gamma = 2$.

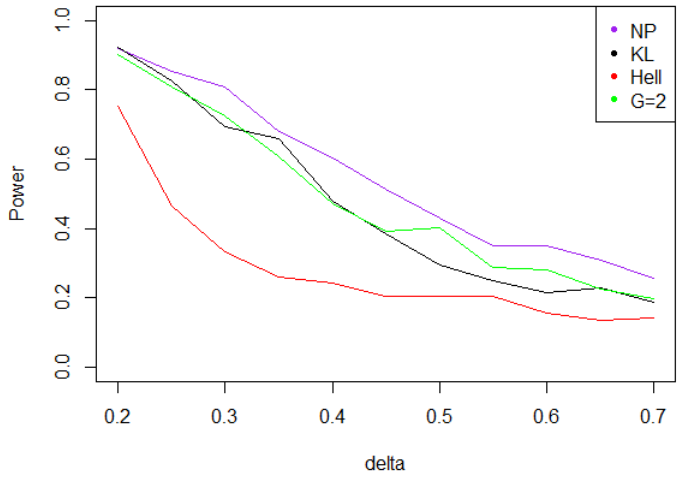


FIGURE 4.10 – Power of the test of case B under Laplacian noise with respect to $\delta_{\max}$ for a first kind risk $\alpha = 0.05$ and a sample size $n = 100$

c. When the noise follows a symmetric Weibull distribution the power function when $\gamma = 1$ is very close to the power of the LRT between $F_{\delta_{\max}}$ and $G_{\delta_{\max}}$ (cf. Figure 4.11). Indeed uniformly on δ and on x the ratio $\left(f_\delta / g_\delta\right)(x)$ is close to 1. Therefore the distribution of T_n is close to that of $T_{n,\delta_{\max}}$ which plays in favor of the KL composite test.

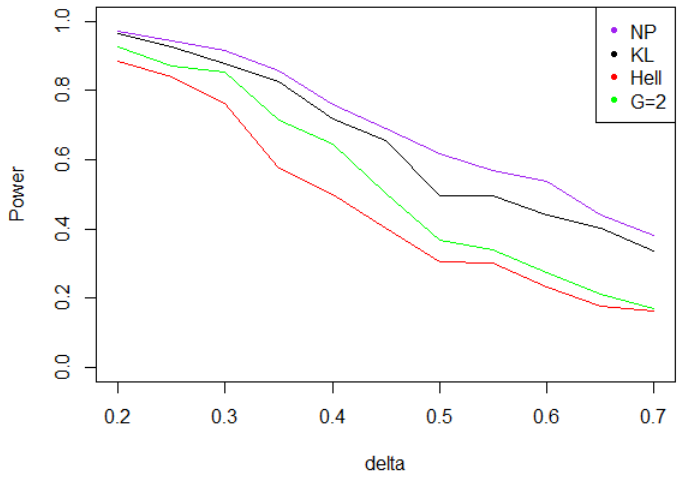


FIGURE 4.11 – Power of the test of case B under symmetrized Weibull noise with respect to $\delta_{\max}$ for a first kind risk $\alpha = 0.05$ and a sample size $n = 100$

d. Under a Cauchy distribution, similarly to case A, Figure 4.12 shows that T_n^γ achieves the maximal power for $\gamma = 1$ and 2, closely followed by $\gamma = 0.5$.

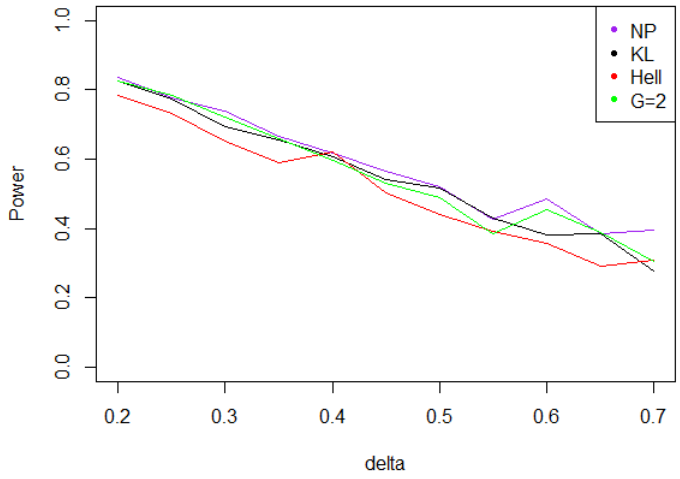


FIGURE 4.12 – Power of the test of case B under a noise following a Cauchy distribution with respect to $\delta_{\max}$ for a first kind risk $\alpha = 0.05$ and a sample size $n = 100$

4.6 Application to fatigue life data

The above procedure can be easily adapted to the industrial case that we are interested in.

4.6.1 Testing the existence of a convolution

As stated in Section 3.3, we denote N_i the crack initiation period and N_p the crack propagation period.

Depending on the level of stress applied, the duration of N_i relative to that of N_p largely differs. Indeed as the level of solicitation increases, the crack tend to initiate earlier and earlier. At extreme cases, N_i may even be negligible with respect to the propagation period. Similarly, when the stress levels are extremely low, the initiation period may be so long (of the order of millions of cycles) that the propagation time is quite negligible in comparison. However, there is no physical evidence that for σ high, $N_i = 0$ and reciprocally, that for σ low, $N_p = 0$.

We propose an adaptation of the composite test to identify at a given loading σ very high or very low, i.e. for values of σ or ϵ for which the life to failure population is homogeneous, whether the lifetime of the material should be modeled by only one period or by the sum of both.

In the following, we will focus on the case where stress σ is high. Thus the crack initiation time is treated as a potential noise on the data and the test consists in determining

whether this noise is negligible or not.
Note

- f_0 the known density of N_p and F_0 its distribution function;
- g_δ the density of $N_p + N_i$ with distribution G_δ where $\delta > 0$ is an unknown scaling parameter of some known distribution H . Therefore

$$g_\delta := f_0 * h_\delta.$$

  As above $\delta \in \Delta_n = (\delta_1 = \delta_{\min}, \ldots, \delta_{\max})$.

We consider here the testing for high σ of the following hypotheses:

H_0 : X has distribution F_0, i.e. $X = N_p$ vs H_1 : X has distribution G_δ for some $\delta \in \Delta$ i.e. $X = N_i + N_p$ for some δ

The test statistics writes

$$T^1_{n,\delta} = \frac{1}{n}\sum_{i=1}^{n} \varphi^1\left(\frac{f_0 * h_\delta}{f_0}(X_i)\right),$$

and

$$T^1_n(\mathbf{X}_1^n) = \sup_{\delta\in\Delta_n} T_{n,\delta} = \sup_{\delta\in\Delta_n} \sum_{i=1}^{n} \varphi^1\left(\frac{f_0 * h_\delta}{f_0}(X_i)\right),$$

with $1 \leq 2$.
In this setting, the least favorable hypotheses is the couple (F_0, G_{δ_*}) which minimizes $K(G_\delta, F_0)$ over δ.
For a given level $\alpha \in (0,1)$ the threshold A_n that defines the critical region of the test is obtained through simulations such that

$$P_{\mathbf{H}_0}(\mathbf{H}_1) := F_0(T^1_n > A_n) \leq \alpha_n.$$

The power of the test is also computed numerically. But note that in the case of Kullback-Leibler based test statistic, the same kind of power bound than (4.23) can be obtained (see Appendix 4.8.4)

4.6.2 Simulation results

The crack propagation period is usually modelled by a lognormal distribution, thus we will assume that f is a lognormal density of location parameter μ_p and scale parameter σ_p completely defined.
Two types of parametric hypotheses on the noise N_i are treated, corresponding to very localized densities since the order of magnitude of N_i is smaller than N_p:

1. Uniformly distributed noise: $N_i \sim \mathscr{U}[0,\delta]$
2. Gamma distributed noise: $N_i \sim \mathscr{G}amma(\delta,1)$

The following tests have been performed under simulations for a fixed first kind risk $\alpha = 0.05$. The minimal power of the test is evaluated numerically under the least favorable hypotheses $(H_0, H_1(\delta_*))$ which satisfy:

$$K\left(F, F * G_{\delta_*}\right) = \min_{\delta\in\Delta} K\left(F, G_\delta\right). \tag{4.15}$$

They are the least distinguishable couple of hypotheses in terms of Kullback-Leibler divergence. In the following examples, they correspond to the case $\delta = \delta_{\min}$, $(H_0, H_1(\delta_{\min}))$.

Case 1: Uniformly distributed noise

In this case, the noise is a uniform variable, whose amplitude is controlled by δ. Note that, in this application, the least favorable hypotheses correspond to the minimal Signal-to-noise ratio, while they are achieved for the maximal SNR in the previous sections. Indeed, we want to be able to detect the presence of a small signal that should then be taken into account to adapt the lifetime modelling. Thus the least favorable case is achieved for $\delta = \delta_{\min}$.

As shown in Figure 4.13, the distance between $\mathbf{H}_0$ and $\mathbf{H}_1(\delta)$ grows as δ increases. Thus, the alternate hypotheses are all the more dictinct when δ gets bigger. The least favorable hypotheses are therefore the pair $\mathbf{H}_0$ vs $\mathbf{H}_1(\delta_{\min})$.

The power of the test is evaluated under the least favorable hypotheses. The power increases quickly as $\delta_{\min}$ increases, as shown in Figure 4.14. How fast the power reaches 100% depends on the sample size n. In industrial applications, a reasonable sample size would be around 50 (red curve on 4.14). Thus, the convolution would be detected correctly with probability 0.9 for $\delta \geq 0.4$.

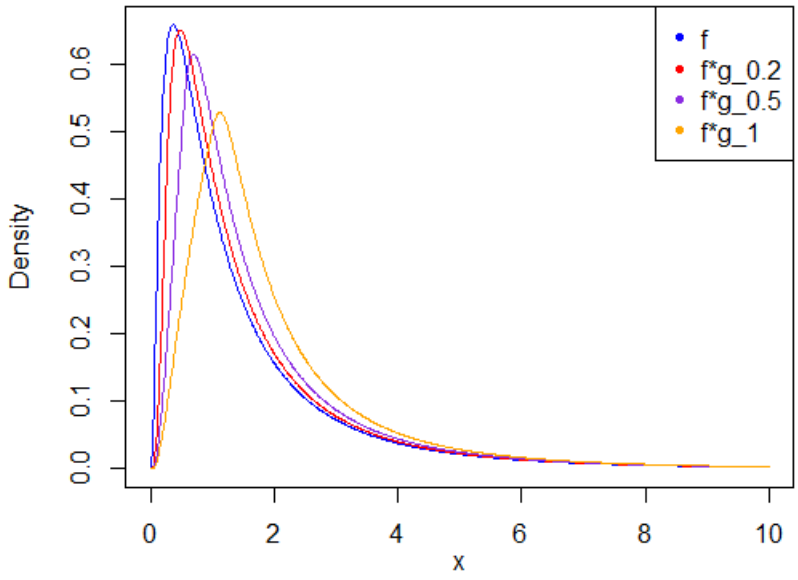


FIGURE 4.13 – Density of the lognormal distributed variable N_p and of the convolutions of N_p and uniformly distributed N_i for a range of values of δ

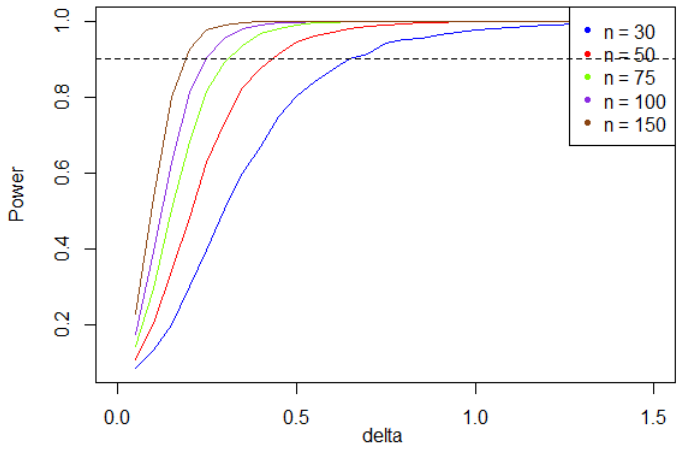


FIGURE 4.14 – Power of the test of case 1 with respect to δ for a first kind risk $\alpha = 0.05$ and different sample sizes

Case 2: Gamma distribution

In the second example, we consider the following case:

- N_p follows a Lognormal distribution of parameters $(1, 1)$;
- N_i follows a Gamma distribution $\mathcal{G}amma(\delta, 1)$ where $\Delta = [\delta_{\min}, 2.5]$

Figure 4.15 shows how f is distorted by the convolution as δ increases. In this case, the power of the test increases more slowly as δ grows (see Figure 4.16). Small values of

$\delta_{\min}$ make the alternate hypotheses too close to be distinguishable from a small sample of observations. Thus when $n = 50$, the power reaches 90% only when $\delta_{\min} = 0.8$. This situation corresponds to a SNR of 0.23, while the same power level could be reached under the Uniform hypothesis for a SNR of 0.003. The performances of the test are highly dependent on the type and intensity of distortion of the distribution under the null hypothesis.

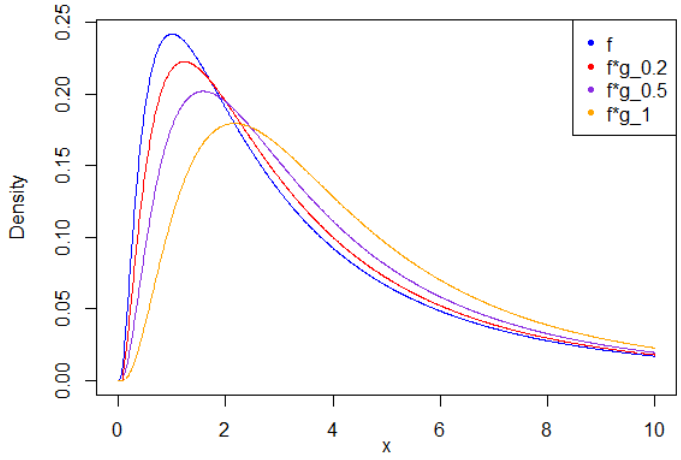


FIGURE 4.15 – Density of the lognormal distributed variable N_p and of the convolutions of N_p and Gamma distributed N_i for a range of values of δ

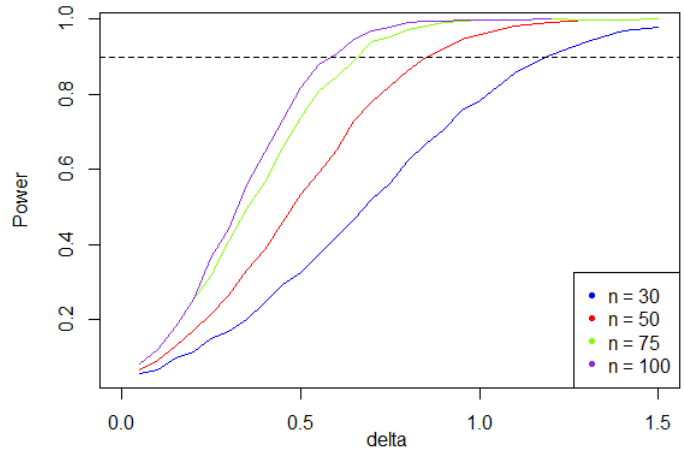


FIGURE 4.16 – Power of the test of case 2 with respect to δ for a first kind risk $\alpha = 0.05$ and different sample sizes

The above simulations were performed in order to give the order of magnitude of the power of the test applied on real data according to the number of observations available and the type of distribution to be tested. The hypotheses on N_i distribution were oriented toward laws defined on small ranges of values.

4.7 Conclusion

We have considered a composite testing problem where simple hypotheses in either **H0** and **H1** are paired, due to corruption in the data. The test statistics are defined through aggregation of simple Likelihood Ratio Tests. The critical region for this test and a lower bound of its power is produced. We have shown that this test is minimax, evidencing the least favorable hypotheses. We have considered the minimal power of the test under such a least favorable hypothesis, both theoretically and by simulation, for a number of cases, including corruption by Gaussian, Laplacian, Weibull and Cauchy noise. Whatever the chosen distribution of the noise, the actual minimal power as measured through simulation is quite higher than obtained through analytic developments. Least favorable hypotheses are defined in an asymptotic sense, and are proved to be the couple of simple hypotheses in **H0** and **H1** which are the closest in terms of the Kullback-Leibler divergence, as a consequence of the Chernoff-Stein Lemma. We next consider aggregation of tests where the Likelihood Ratio is substituted by a divergence-based statistics. This choice extends the former one, and may produce aggregate tests with higher power than obtained through aggregation of the LRTs, as examplified and analysed. Open questions are related to possible extensions of the Chernoff-Stein Lemma for divergence-based statistics.

This test procedure can be easily adapted to testing between a simple and a composite hypothesis. This other formulation is suited for the fatigue life application that we are interested in. It can thus be applied to fatigue life data in order to determinate whether for high level of solicitation, the life to failure can be reduced to propagation or not, and reciprocally for very small levels of stress, whether the number of cycles to failure can be reduced to crack initiation time or not.

4.8 Appendix

4.8.1 Proof of Proposition 5

The critical region of the test

Define

$$Z_{\delta'} := \log \frac{g_{\delta'}}{f_{\delta'}}(X)$$

which satisfies

$$\begin{aligned} E_{F_\delta}(Z_{\delta'}) &= \int \log \frac{g_{\delta'}}{f_{\delta'}}(x) f_\delta(x) dx \\ &= \int \log \frac{g_{\delta'}}{f_\delta}(x) f_\delta(x) dx + \int \log \frac{f_\delta}{f_{\delta'}}(x) f_\delta(x) dx \\ &= K(F_\delta, F_{\delta'}) - K(F_\delta, G_{\delta'}). \end{aligned}$$

Note that for all δ,

$$K(F_\delta, F_{\delta'}) - K(F_\delta, G_{\delta'}) = \int \log \frac{g_{\delta'}}{f_{\delta'}} f_\delta$$

is negative for δ' close to δ, assuming that

$$\delta' \mapsto \int \log \frac{g_{\delta'}}{f_{\delta'}} f_\delta$$

is a continuous mapping. Assume therefore that (4.6) holds which means that the classes of distributions (G_δ) and (F_δ) are somehow well separated. This implies that $E_{F_\delta}(Z_{\delta'}) < 0$ for all δ and δ'.

In order to obtain an upper bound for $F_\delta\left(T_{n,\delta'}(\mathbf{X}_n) > A_n\right)$ for all δ, δ' in Δ through the Chernoff Inequality, consider

$$\varphi_{\delta,\delta'}(t) := \log E_{F_\delta}\left(\exp(tZ_{\delta'})\right) = \log \int \left(\frac{g_{\delta'}(x)}{f_{\delta'}(x)}\right)^t f_\delta(x) dx.$$

Let

$$t^+\left(\mathcal{N}_{\delta,\delta'}\right) := \sup\left\{t \in \mathcal{N}_{\delta,\delta'} : \varphi_{\delta,\delta'}(t) < \infty\right\}$$

The function $\left(\delta, \delta', x\right) \mapsto J_{\delta,\delta'}(x)$ is continuous on its domain, and since $t \mapsto \varphi_{\delta,\delta'}(t)$ is a strictly convex function which tends to infinity as t tends to $t^+\left(\mathcal{N}_{\delta,\delta'}\right)$ it holds that

$$\lim_{x\to\infty} J_{\delta,\delta'}(x) = +\infty$$

for all δ, δ' in Δ_n.

We now consider an upper bound for the risk of first kind on a logarithmic scale.

We consider

$$A_n > E_{F_\delta}(Z_{\delta'}) \tag{4.16}$$

for all δ, δ' . Then by Chernoff inequality

$$\frac{1}{n} \log F_\delta\left(T_{n,\delta'}(\mathbf{X}_n) > A_n\right) \le -J_{\delta,\delta'}(A_n).$$

Since A_n should satisfy

$$\exp\left(-nJ_{\delta,\delta'}(A_n)\right) \leq \alpha_n$$

with α_n bounded away from 1, A_n surely satisfies (4.16) for large n.
The mapping $m_{\delta,\delta'}(t) := (d/dt)\,\varphi_{\delta,\delta'}(t)$ is a homeomorphism from $\mathcal{N}_{\delta,\delta'}$ onto the closure of the convex hull of the support of the distribution of $Z_{\delta'}$ under F_δ (see e.g. [5]). Denote

$$\operatorname*{ess\,sup}_{\delta} Z_{\delta'} := \sup\{x : \text{for all } \epsilon > 0,\ F_\delta\left(Z_{\delta'} \in (x-\epsilon, x) > 0\right)\}$$

We assume that

$$\operatorname*{ess\,sup}_{\delta} Z_{\delta'} = +\infty \tag{4.17}$$

which is convenient for our task and quite common in practical industrial modelling. This assumption may be weakened, at notational cost mostly. It follows that

$$\lim_{t \to t^+\left(\mathcal{N}_{\delta,\delta'}\right)} m_{\delta,\delta'}(t) = +\infty.$$

It holds

$$J_{\delta,\delta'}\left(E_{F_\delta}(Z_{\delta'})\right) = 0$$

and, as seen previously

$$\lim_{x \to \infty} J_{\delta,\delta'}(x) = +\infty.$$

On the other hand

$$m_{\delta,\delta'}(0) = E_{F_\delta}(Z_{\delta'}) = K(F_\delta, F_{\delta'}) - K(F_\delta, G_{\delta'}) < 0.$$

Let

$$\begin{aligned}\mathcal{I} &:= \left(\sup_{\delta,\delta'} E_{F_\delta}(Z_{\delta'}), \infty\right) \\ &= \left(\sup_{\delta,\delta'} K(F_\delta, F_{\delta'}) - K(F_\delta, G_{\delta'}), \infty\right)\end{aligned}$$

By (4.17) the interval $\mathcal{I}$ is not void.
We now define A_n such that (4.4) holds, namely

$$P_{\mathbf{H0}}(\mathbf{H1}) \leq p_n \max_{\delta} \max_{\delta'} F_\delta\left(T_{n,\delta'} > A_n\right) \leq \alpha_n$$

holds for any α_n in $(0,1)$. Note that

$$A_n \geq \max_{\delta,\delta'} E_{F_\delta}(Z_{\delta'}) = \max_{(\delta,\delta')\in\Delta\times\Delta} K(F_\delta, F_{\delta'}) - K(F_\delta, G_{\delta'}) \tag{4.18}$$

for all n large enough since α_n is bounded away from 1.
The function

$$J(x) := \min_{(\delta,\delta')\in\Delta\times\Delta} J_{\delta,\delta'}(x)$$

is continuous and increasing as is the infimum of a finite collection of continuous increasing functions, all defined on $\mathscr{I}$.
Since

$$P_{\mathbf{H0}}(\mathbf{H1}) \leq p_n \exp(-nJ(A_n)),$$

given α_n, define

$$A_n := J^{-1}\left(-\frac{1}{n}\log\frac{\alpha_n}{p_n}\right). \tag{4.19}$$

This is well defined for $\alpha_n \in (0,1)$ since $\sup_{(\delta,\delta')\in\Delta\times\Delta} E_{F_\delta}(Z_{\delta'}) < 0$ and $-(1/n)\log(\alpha_n/p_n) > 0$.

The power function

We now evaluate a lower bound for the power of this test, making use of the Chernoff inequality to get an upper bound for the second risk.
Starting from (4.5)

$$P_{H_1}(\mathbf{H0}) \leq \sup_{\eta\in\Delta} G_\eta\left(T_{n,\eta} \leq A_n\right),$$

define

$$W_\eta := -\log\frac{g_\eta}{f_\eta}(x).$$

It holds

$$E_{G_\eta}(W_\eta) = \int \log\frac{f_\eta(x)}{g_\eta(x)} g_\eta(x)\,dx = -K(G_\eta, F_\eta).$$

and

$$m_\eta(t) := (d/dt)\log E_{G_\eta}(\exp tW_\eta)$$

which is an increasing homeomorphism from $\mathscr{M}_\eta$ onto the closure of the convex hull of the support of W_η under G_η. For any η, the mapping

$$x \mapsto I_\eta(x)$$

is a strictly increasing function of $\mathscr{K}_\eta := \left(E_{G_\eta}(W_\eta), \infty\right)$ onto $(0, +\infty)$, where the same

notation as above holds for $\operatorname{ess\,sup}_\eta W_\eta$, here under G_η, and where we assumed

$$\operatorname*{ess\,sup}_\eta W_\eta = \infty \tag{4.20}$$

for all η.

Assume that A_n satisfies

$$A_n \in \mathcal{K} := \bigcap_{\eta\in\Delta} \mathcal{K}_\eta \tag{4.21}$$

namely

$$A_n \geq \sup_{\eta\in\Delta} E_{G_\eta}\left(W_\eta\right) = -\inf_{\eta\in\Delta} K\left(G_\eta, F_\eta\right). \tag{4.22}$$

Making use of Chernoff inequality we get

$$P_{H_1}(\mathbf{H0}) \leq \exp\left(-n \inf_{\eta\in\Delta} I_\eta(A_n)\right).$$

Each function $x \mapsto I_\eta(x)$ is increasing on $(E_{G_\eta}\left(W_\eta\right), \infty)$. Therefore the function

$$x \mapsto I(x) := \inf_{\eta\in\Delta} I_\eta(x)$$

is continuous and increasing, as is the infimum of a finite number of continuous increasing functions on the same interval $\mathcal{K}$, which is not void, due to (4.20).

We have proved that whenever (4.22) holds a lower bound tor the test of **H0** vs **H1** is given by

$$\begin{aligned} P_{H_1}(\mathbf{H1}) &\geq 1 - \exp\left(-nI(A_n)\right) \\ &= 1 - \exp\left(-nI\left(J^{-1}\left(-\frac{1}{n}\log\frac{\alpha_n}{p_n}\right)\right)\right). \end{aligned} \tag{4.23}$$

We now collect the above discussion in order to complete the proof.

A synthetic result

The function J is one to one from I onto $K := \left(J\left(\sup_{(\delta,\delta')\in\Delta\times\Delta} E_\delta\left(Z_{\delta'}\right)\right), \infty\right)$. Since under $F_\delta, J_{\delta,\delta'}\left(E_\delta\left(Z_{\delta'}\right)\right) = 0$, it follows that $J\left(\sup_{(\delta,\delta')\in\Delta\times\Delta} E_\delta\left(Z_{\delta'}\right)\right) \geq 0$. Since $E_{F_\delta}\left(Z_{\delta'}\right) = K\left(F_\delta, F_{\delta'}\right) - K\left(F_\delta, G_{\delta'}\right) < 0$, whatever α_n in $(0,1)$ there exists a unique $A_n \in \left(-\inf_{(\delta,\delta')\in\Delta\times\Delta}\left(K\left(F_\delta, G_{\delta'}\right) - K\left(F_\delta, F_{\delta'}\right)\right), \infty\right)$ which defines the critical region with level α_n.

For the lower bound on the power of the test, we have assumed $A_n \in \mathcal{K} = \left(\sup_{\eta\in\Delta} E_\eta\left(W_\eta\right), \infty\right) = \left(-\inf_{\eta\in\Delta} K(G_\eta, F_\eta), \infty\right)$.

In order to collect our results in a unified setting it is useful to state some connection

between $\inf_{(\delta,\delta')\in\Delta\times\Delta}[K(F_\delta, G_{\delta'}) - K(F_\delta, F_{\delta'})]$ and $\inf_{\eta\in\Delta} K(G_\eta, F_\eta)$. See (4.18) and (4.22). Since $K(G_\delta, F_\delta)$ is positive it results from (4.6) that

$$\sup_{(\delta,\delta')\in\Delta\times\Delta} \int \log \frac{f_{\delta'}}{g_{\delta'}} f_\delta < \sup_{\delta\in\Delta} K(G_\delta, F_\delta) \tag{4.24}$$

which implies the following fact:
Let α_n be bounded away from 1. Then (4.18) is fulfilled for large n, and therefore there exists A_n such that

$$\sup_{\delta\in\Delta} F_\delta\left(T_n > A_n\right) \leq \alpha_n.$$

Furthermore by (4.24) , condition (4.22) holds, which yields the lower bound for the power of this test, as stated in (4.23).

4.8.2 Proof of Theorem 4.4.1

We will repeatedly make use of the following result (Theorem 3 in [41]), which is an extension of the Chernoff-Stein Lemma (see [19])

Theorem 8. [Krafft and Plachky] Let x_n be such that

$$F_\delta\left(T_{n,\delta} > x_n\right) \leq \alpha_n$$

with $limsup_{n\to\infty}\alpha_n < 1$. Then

$$\lim_{n\to\infty} \frac{1}{n} \log G_\delta\left(T_{n,\delta} \leq x_n\right) = -K\left(F_\delta, G_\delta\right).$$

Remark 9. The above result indicates that the power of the Neyman Pearson test only depends on its level on the second order on the logarithmic scale.

Define A_{n,δ_*} such that

$$F_{\delta_*}\left(T_n \leq A_n\right) = F_{\delta_*}\left(T_{n,\delta_*} \leq A_{n,\delta_*}\right).$$

This exists and is uniquely defined due to the regularity of the distribution of T_{n,δ_*} under F_{δ_*}. Since $\mathbf{1}\left[T_{n,\delta_*} > A_n\right]$ is the likelihood ratio test of $\mathbf{H}_0(\delta_*)$ against $\mathbf{H}_1(\delta_*)$ of the size α_n, it follows by unbiasedness of the LRT that

$$F_{\delta_*}\left(T_n \leq A_n\right) = F_{\delta_*}\left(T_{n,\delta_*} \leq A_{n,\delta_*}\right) \geq G_{\delta_*}\left(T_{n,\delta_*} \leq A_{n,\delta_*}\right).$$

We shall later verify the validity of the conditions of Theorem 8, namely that

$$\limsup_{n\to\infty} F_{\delta_*}\left(T_{n,\delta_*} \le A_{n,\delta_*}\right) < 1. \tag{4.25}$$

Assuming (4.25) we get by Theorem 8

$$\limsup_{n\to\infty} \frac{1}{n}\log F_{\delta_*}\left(T_n \le A_n\right) \ge \lim_{n\to\infty} \frac{1}{n}\log G_{\delta_*}\left(T_{n,\delta_*} \le A_{n,\delta_*}\right) = -K\left(F_{\delta_*}, G_{\delta_*}\right).$$

We shall now prove that

$$\lim_{n\to\infty} \frac{1}{n}\log G_{\delta_*}\left(T_{n,\delta_*} \le A_{n,\delta_*}\right) = \lim_{n\to\infty} \frac{1}{n}\log G_{\delta_*}\left(T_n \le A_n\right).$$

Let B_{n,δ_*} be such that

$$G_{\delta_*}\left(T_{n,\delta_*} \le B_{n,\delta_*}\right) = G_{\delta_*}\left(T_n \le A_n\right).$$

By regularity of the distribution of T_{n,δ_*} under G_{δ_*} such a B_{n,δ_*} is defined in a unique way. We will prove that the condition in Theorem 8 holds, namely

$$\limsup_{n\to\infty} F_{\delta_*}\left(T_{n,\delta_*} \le B_{n,\delta_*}\right) < 1. \tag{4.26}$$

$$\lim_{n\to\infty} \frac{1}{n}\log G_{\delta_*}\left(T_{n,\delta_*} \le A_{n,\delta_*}\right) = \lim_{n\to\infty} \frac{1}{n}\log G_{\delta_*}\left(T_n \le A_n\right) = -K\left(F_{\delta_*}, G_{\delta_*}\right).$$

Incidentally, we have obtained that $\lim_{n\to\infty} \frac{1}{n}\log G_{\delta_*}\left(T_n \le A_n\right)$ exists. Therefore we have proven that

$$\limsup_{n\to\infty} \frac{1}{n}\log F_{\delta_*}\left(T_n \le A_n\right) \ge \lim_{n\to\infty} \frac{1}{n}\log G_{\delta_*}\left(T_n \le A_n\right)$$

which is a form of unbiasedness. For $\delta \neq \delta_*$, let $B_{n,\delta}$ be defined by

$$G_{\delta}\left(T_{n,\delta} \le B_{n,\delta}\right) = G_{\delta}\left(T_n \le A_n\right).$$

As above, $B_{n,\delta}$ is well-defined. Assuming

$$\limsup_{n\to\infty} F_{\delta}\left(T_{n,\delta} \le B_{n,\delta}\right) < 1, \tag{4.27}$$

it follows from Theorem 8 that

$$\lim_{n\to\infty} \frac{1}{n}\log G_{\delta}\left(T_n \le A_n\right) = \lim_{n\to\infty} \frac{1}{n}\log G_{\delta}\left(T_{n,\delta} \le B_{n,\delta}\right) = -K\left(F_{\delta}, G_{\delta}\right).$$

Since $K\left(F_{\delta_*}, G_{\delta_*}\right) \leq K\left(F_\delta, G_\delta\right)$, we have proven

$$\limsup_{n\to\infty} \frac{1}{n} \log F_{\delta_*}\left(T_n \leq A_n\right) \geq \lim_{n\to\infty} \frac{1}{n} \log G_{\delta_*}\left(T_n \leq A_n\right) \geq \lim_{n\to\infty} \frac{1}{n} \log G_\delta\left(T_n \leq A_n\right).$$

It remains to verify the conditions (4.25), (4.26) and (4.27). We will only verify (4.27) since the two other conditions differ only by notation. We have

$$\begin{aligned} G_\delta\left(T_{n,\delta} > B_{n,\delta}\right) = G_\delta\left(T_n > A_n\right) &\leq F_\delta\left(T_n > A_n\right) + d_{TV}\left(F_\delta, G_\delta\right) \\ &\leq \alpha_n + d_{TV}\left(F_\delta, G_\delta\right) < 1 \end{aligned}$$

by hypothesis (4.13). By the law of large numbers, under G_δ

$$\lim_{n\to\infty} T_{n,\delta} = K(G_\delta, F_\delta) \qquad G_\delta - a.s$$

Therefore, for large n,

$$\liminf_{n\to\infty} B_{n,\delta} \geq K(G_\delta, F_\delta) \qquad G_\delta - a.s$$

Since under F_δ

$$\lim_{n\to\infty} T_{n,\delta} = -K(F_\delta, G_\delta) \qquad F_\delta - a.s$$

this implies that

$$\lim_{n\to\infty} F_\delta\left(T_{n,\delta} > B_{n,\delta}\right) < 1.$$

4.8.3 Proof of Proposition 7

We now prove the three lemmas that we used.

Lemma 10. Let P, Q and R denote three distributions with respective continuous and bounded densities p, q and r. Then

$$K(P * R, Q * R) \leq K(P, Q). \tag{4.28}$$

Démonstration. Let $\mathscr{P} := (A_1, .., A_K)$ be a partition of $\mathbb{R}$ and $p := \left(p_1, .., p_K\right)$ denote the probabilities of $A_1, .., A_K$ under P. Set the same definition for $q_1, .., q_K$ and for $r_1, .., r_K$. Recall that the log-sum inequality writes

$$\left(\sum a_i\right) \log \frac{\sum b_i}{\sum c_i} \leq \sum a_i \log \frac{b_i}{c_i}$$

for positive vectors $(a_i)_i$, $(b_i)_i$ and $(c_i)_i$. By the above inequality for any $i \in \{1, \ldots, K\}$,

denoting $(p * r)$ the convolution of p and r,

$$(p*r)_j \log \frac{(p*r)_j}{(q*r)_j} \leq \sum_{i=1}^{K} p_j r_{i-j} \log \frac{p_j r_{i-j}}{q_j r_{i-j}}.$$

Summing upon $j \in \{1, \ldots, K\}$ yields

$$\sum_{j=1}^{K} (p*r)_j \log \frac{(p*r)_j}{(q*r)_j} \leq \sum_{j=1}^{K} p_j \log \frac{p_j}{q_j}.$$

which is

$$K_{\mathscr{P}}(P*R, Q*R) \leq K_{\mathscr{P}}(P,Q)$$

where $K_{\mathscr{P}}$ designates the Kullback-Leibler divergence defined on $\mathscr{P}$. Refine the partition and go to the limit (Riemann Integrals), getting (4.28). □

We now set a classical general result which states that when R_δ denotes a family of distributions with some decomposability property, then the Kullback-Leibler divergence between $P * R_\delta$ and $Q * R_\delta$ is a decreasing function of δ.

Lemma 11. Let P and Q satisfy the hypotheses of Lemma 10 and let $(R_\delta)_{\delta>0}$ denote a family of p.m's on $\mathbb{R}$ and denote accordingly V_δ a r.v. with distribution R_δ. Assume that for all δ and η there exists a r.v. $W_{\delta,\eta}$ independent upon V_δ such that

$$V_{\delta+\eta} =_d V_\delta + W_{\delta,\eta}.$$

Then the function $\delta \mapsto K(P * R_\delta, Q * R_\delta)$ is non increasing.

Démonstration. It holds, using Lemma 10, for positive η

$$\begin{aligned} K\left(P * R_{\delta+\eta}, Q * R_{\delta+\eta}\right) &= K\left((P * R_\delta) * W_{\delta,\eta}, (Q * R_\delta) * W_{\delta,\eta}\right) \\ &\leq K\left(P * R_\delta, Q * R_\delta\right) \end{aligned}$$

which proves the claim. □

Lemma 12. Let P , Q and R be three probability distributions with respective continuous and bounded densities p, q and r. Assume that

$$K(P,Q) \leq K(Q,P)$$

where all involved quantities are assumed to be finite. Then

$$K(P*R, Q*R) \leq K(Q*R, P*R).$$

Démonstration. We proceed as in Lemma 10 using partitions, denoting $p_1, .., p_K$ the induced probability of P on $\mathscr{P}$. Then

$$\begin{aligned} K_{\mathscr{P}}(P*R, Q*R) - K_{\mathscr{P}}(Q*R, P*R) &= \sum_i \sum_j \left(p_j r_{i-j} + q_j r_{i-j}\right) \log \frac{\sum_j p_j r_{i-j}}{\sum_j q_j r_{i-j}} \\ &\leq \sum_j \sum_i \left(p_j r_{i-j} + q_j r_{i-j}\right) \log \frac{p_j}{q_j} \\ &= \sum_j \left(p_j + q_j\right) \log \frac{p_j}{q_j} \\ &= K_{\mathscr{P}}(P, Q) - K_{\mathscr{P}}(Q, P) \leq 0 \end{aligned}$$

where we used the log-sum inequality and the fact that $K(P,Q) \leq K(Q,P)$ implies $K_{\mathscr{P}}(P,Q) \leq K_{\mathscr{P}}(Q,P)$ by the data processing inequality. □

4.8.4 Critical region and power of the test adapted to the industrial application

The critical region is defined by $\{T_n^1(N) > A_n(\alpha_n)\}$ where $A_n = A_n(\alpha_n)$ is such that:

$$\begin{aligned} P_{\mathbf{H}_0}(\mathbf{H}_1) &:= F_0(T_n^1 > A_n) \leq \alpha_n \\ &= F_0\left(\exists \delta' \in \Delta_n : \frac{1}{n} \sum_{i=1}^n \varphi^1\left(\frac{g_\delta}{f_0}(X_i)\right) > A_n\right) \\ &= F_0\left(\bigcup_{\delta' \in \Delta_n} \left\{\frac{1}{n} \sum_{i=1}^n \varphi^1\left(\frac{g_\delta}{f_0}(X_i)\right) > A_n\right\}\right) \\ &\leq p_n \sup_{\delta' \in \Delta_n} F_0\left(T_{n,\delta'}^1 > A_n\right). \end{aligned}$$

The theoretical power bound is obtained similarly to (4.23):

$$\begin{aligned} P_{H_1(\delta)}(H_0) &= P_{F_\delta}(T_n^1(N) \leq A_n) \\ &= P_{F_\delta}\left(\sup_{\delta'} \int \log \frac{f * g_{\delta'}}{f}(x)\, dx \leq A_n\right) \\ &= P_{F_\delta}\left(\bigcap_{\delta'} \left\{\int \log \frac{f * g_{\delta'}}{f}(x)\, dx \leq A_n\right\}\right) \\ &\leq P_{F_\delta}\left(\int \log \frac{f * g_\eta}{f}(x)\, dx \leq A_n\right) \quad \text{with } \eta \in \Delta_n \end{aligned}$$

$$P_{H_1}(H_0) = \sup_{\delta} P_{H_1(\delta)}(H_0)$$
$$\leq \sup_{\delta} P_{F_\delta}\left(\int \log \frac{f * g_\eta}{f}(x)\,dx \leq A_n\right) \quad \text{with } \eta \in \Delta_n.$$

It follows

$$P_{H_1}(H_0) \leq \sup_{\delta} P_{F_\delta}\left(\int \log \frac{f * g_\delta}{f}(x)\,dx \leq A_n\right)$$

and by Chernoff inequality

$$P_{H_1}(H_0) \leq \sup_{\delta} \exp\{-nI_\delta(A_n)\} \tag{4.29}$$

where $I_\delta(x) := \sup_t tx - \log E_{G_\delta}\left(\exp tW_\delta\right)$.

Chapitre 5

Testing the number and the nature of the components in a mixture distribution

5.1 Introduction

The test problem for the number of components of a finite mixture has been extensively treated when the total number of components k is equal to 2, leading to a satisfactory solution; the limit distribution of the generalized likelihood ratio statistic is non standard, since it is $0.5\delta_0 + 0.5\chi^2(1)$, a mixture of a Dirac mass at 0 and a $\chi^2(1)$ with weights equal to 1/2; see e.g. Titterington (1985 [53]) and Self an Liang (1987 [49]).
When $k > 2$, the problem is much more complicated. Self and Liang [49] obtained the limit distribution of the generalized likelihood ratio statistic, which is non standard and complex. This result yields significant numerical difficulties for the calculation of the critical value of the test. Those drawbacks motivate the search for an alternative testing procedure for a population homogeneity. In section 5.3, we propose a unified treatment for all these cases, with simple and standard limit distribution, that also holds for mixtures of $k > 2$ components, both when the parameter θ_T is an interior or a boundary point of the parameter space Θ. Moreover, confidence regions for the mixture parameter θ_T even when $k = 2$ are intractable through the generalized likelihood ratio statistic. Indeed, the limit law of the generalized likelihood ratio statistic depends heavily on the fact that θ is a boundary or an interior point of the parameter space. For example, when $k = 2$, the limit distribution of the generalized likelihood ratio statistic is $0.5\delta_0 + 0.5\chi^2(1)$ when $\theta = 0$ and $\chi^2(1)$ when $0 < \theta < 1$. Therefore, the confidence level is not defined uniquely. At the opposite, we will prove in section 5.3 that the proposed dual χ^2-statistic yields quite standard confidence regions even when $k > 2$. Section 5.4 proposes a few

simulations on two-component mixtures coming either from the same or from different parametric families.

5.1.1 Number of components of a parametric mixture model

Consider a k-component parametric mixture model P_θ ($k \geq 2$) defined as follows:

$$P_\theta := \sum_{i=1}^{k} w_i P_{a_i}^{(i)} \tag{5.1}$$

where $\left\{P_{a_1}^{(1)}; a_1 \in A_1\right\}, \ldots, \left\{P_{a_k}^{(k)}; a_k \in A_k\right\}$ are k parametric models and $A_1, \ldots, A_k$ are k sets in $\mathbb{R}^{d_1}, \ldots, \mathbb{R}^{d_k}$ with $d_1, \ldots, d_k \in \mathbb{N}^*$ and $0 \leq w_i \leq 1$, $\sum w_i = 1$. Note that we consider an nonstandard framework in which the weights w_i are allowed to be equal to 0. Note Θ the parameter space:

$$\theta \in \Theta := \left\{ (w_1, \ldots, w_k, a_1, \ldots, a_k)^T \in [0,1]^k \times A_1 \times \cdots \times A_k \text{ such that } \sum_{i=1}^{k} w_i = 1 \right\}, \tag{5.2}$$

and assume that the model is identifiable. Let $k_0 \in \{1, \ldots, k-1\}$.

We are willing to test if $(k - k_0)$ components in (5.1) have null coefficients. We assume that their labels are $k_0 + 1, ..., k$. Denote Θ_0 the subset of Θ defined by

$$\Theta_0 := \left\{\theta \in \Theta \text{ such that } w_{k_0+1} = \cdots = w_k = 0\right\}.$$

On the basis of an i.i.d sample $X_1, \ldots, X_n$ with distribution P_{θ_T}, $\theta_T \in \Theta$, we intend to perform tests of the hypothesis

$$\mathscr{H}_0 : \theta_T \in \Theta_0 \text{ against the alternative } \mathscr{H}_1 : \theta_T \in \Theta \setminus \Theta_0. \tag{5.3}$$

5.1.2 Motivations

When considering the test (5.3), it is known that the generalized likelihood ratio test, based on the statistic LR defined by

$$LR(X) := 2 \log \frac{\sup_{\theta \in \Theta} \prod_{i=1}^{n} p_\theta(X_i)}{\sup_{\theta \in \Theta_0} \prod_{i=1}^{n} p_\theta(X_i)}, \tag{5.4}$$

is not valid, since the asymptotic approximation by χ^2 distribution does not hold in this case; the problem is due to the fact that the null value of θ_T is not in the interior of the parameter space Θ. We clarify now this problem.

For simplicity, consider a mixture of two known densities p_0 and p_1 with $p_0 \neq p_1$:

$$p_\theta = (1-\theta)p_0 + \theta p_1 \text{ where } \theta \in \Theta := [0,1]. \tag{5.5}$$

Given data $X_1, \ldots, X_n$ with distribution P_{θ_T} and density p_{θ_T}, $\theta_T \in [0,1]$, consider the test problem

$$\mathscr{H}_0 : \theta_T = 0 \text{ against the alternative } \mathscr{H}_1 : \theta_T > 0. \tag{5.6}$$

The generalized likelihood ratio statistic for this test problem is

$$W_n(0) := 2\log \frac{L(\widehat{\theta})}{L(0)}, \tag{5.7}$$

where $\widehat{\theta}$ is the maximum likelihood estimator of θ_T.
Under suitable regularity conditions we can prove that the limit distribution of the statistic W_n in (5.7) is $0.5\delta_0 + 0.5\chi^2_1$, a mixture of the χ^2-distribution and the Dirac measure at zero; see e.g Titterington and al. [53], Self and Liang [49] and Ciuperca [21].
Moreover, in the case of more than two components and $k - k_0 \geq 2$, the limit distribution of the GLR statistic (5.4) under $\mathscr{H}_0$ is complicate and not standard (not a χ^2 distribution) which poses some difficulty in determining the critical value that will give correct asymptotic size; see Self and Liang [49]. Azais and al. [1] proposes for instance a likelihood ratio approach for mixtures and give the asymptotic properties of the test, but its numerical application is extremely complicated, especially under non-Gaussian mixtures. On the other hand, the likelihood ratio statistic

$$W_n(\theta) := 2log\frac{L(\widehat{\theta})}{L(\theta)} \tag{5.8}$$

can not be used to construct an asymptotic confidence region for the parameter θ_T since its limit law is not the same when $\theta_T = 0$ and $\theta_T > 0$.
The case where some parameter of the model belongs to the frontier of the domain is a special case of power models, see for instance Castillo and al. [17] for related statistical issues.

In the sequel, we propose a simple solution for testing the number of components of a parametric mixture model. This method consists in constructing a test statistic based on $\varphi-$divergences and their asymptotic properties. In the following section, we provide the general framework that will be used to construct the test procedure, i.e. the definitions, representation and properties of $\varphi-$divergences.

5.2 Some definition and notation in relation with minimum divergence inference

Let $\mathscr{P} := \{P_\theta, \theta \in \Theta\}$ be an identifiable parametric model on $\mathbb{R}^s$ where Θ is a subset of $\mathbb{R}^d$. All measures in $\mathscr{P}$ will be assumed to be measure equivalent sharing therefore the same support. The parameter space Θ does not need to be open in the present setting. It may even happen that the model includes measures which would not be probability distributions; cases of interest cover the present setting, namely models including unnormalized mixtures of probability distributions; see Broniatowski and Keziou [13].
The f-divergences were introduced by Csiszar [23] as convex non-negative dissimilarities between two probability distributions. Let f be a convex function on $\mathbb{R}_+$, that possibly takes infinite values at 0 and such that $f(1) = 0$. Denote by F the f-divergence between two probability distributions P and Q:

$$F(\alpha,\theta) := \int_{\mathbb{R}^s} f\left(\frac{dP_\alpha}{dP_\theta}(x)\right) dP_\theta(x).$$

Extensions to cases where Q is a finite signed measure and P a probability measure are called $\varphi-$divergences.
Let φ be a proper closed convex function from $]-\infty,+\infty[$ to $[0,+\infty]$ with $\varphi(1) = 0$ and such that its domain $\mathrm{dom}\varphi := \{x \in \mathbb{R} \text{ such that } \varphi(x) < \infty\}$ is an interval with endpoints $a_\varphi < 1 < b_\varphi$ (which may be finite or infinite). For two measures P_α and P_θ in $\mathscr{P}$ the φ-divergence between the two is defined by

$$\phi(\alpha,\theta) := \int_{\mathbb{R}^s} \varphi\left(\frac{dP_\alpha}{dP_\theta}(x)\right) dP_\theta(x).$$

The basic property of $\varphi-$ divergences states that when φ is strictly convex on a neighborhood of $x = 1$, then

$$\phi(\alpha,\theta) = 0 \text{ if and only if } \alpha = \theta.$$

We refer to Liese and Vajda [42] chapter 1 for a complete study of those properties. See also Pardo [46]. Let us simply quote that in general $\phi(\alpha,\theta)$ and $\phi(\theta,\alpha)$ are not equal. Hence, φ-divergences usually are not distances, but they merely measure some difference between two measures. A main feature of divergences between distributions of random variables X and Y is the invariance property with respect to common smooth change of variables.

5.2.1 Examples of φ-divergences

The Kullback-Leibler (KL), modified Kullback-Leibler (KL_m), χ^2, modified χ^2 (χ^2_m), Hellinger (H), and L_1 divergences are respectively associated to the convex functions

- $\varphi(x) = x\log x - x + 1$,
- $\varphi(x) = -\log x + x - 1$,
- $\varphi(x) = \frac{1}{2}(x-1)^2$,
- $\varphi(x) = \frac{1}{2}(x-1)^2/x$,
- $\varphi(x) = 2(\sqrt{x}-1)^2$ and
- $\varphi(x) = |x-1|$.

All these divergences except the L_1 one, belong to the class of the so called "power divergences" introduced in Cressie and Read [22] (see also Liese and Vajda [42] chapter 2), a class which takes its origin from Rényi [48]. They are defined through the class of convex functions

$$\begin{aligned} x \in]0,+\infty[\mapsto \varphi_\gamma(x) &:= \frac{x^\gamma - \gamma x + \gamma - 1}{\gamma(\gamma-1)} \quad \text{if } \gamma \in \mathbb{R}\setminus\{0,1\} \\ \varphi_0(x) &:= -\log x + x - 1, \\ \varphi_1(x) &:= x\log x - x + 1. \end{aligned} \tag{5.9}$$

So, the KL-divergence is associated to φ_1, the KL_m to φ_0, the χ^2 to φ_2, the χ^2_m to φ_{-1} and the Hellinger distance to $\varphi_{1/2}$.

Consider any φ-divergence except the likelihood divergence, with φ being a differentiable function. When θ_T in $int\Theta$ is defined as the true parameter of the distribution of the i.i.d. sample $(X_1,..,X_n)$, it is convenient to assume that

$$\begin{aligned} &\text{There exists a neighborhood } \mathscr{U} \text{ of } \theta_T \text{ for which} \\ &\phi(\theta,\theta') \text{ is finite whatever } \theta \text{ and } \theta' \text{ in } \mathscr{U}. \end{aligned} \tag{A}$$

We will only consider divergences defined through differentiable functions φ, which we assume to satisfy

(**RC**) There exists a positive δ such that for all c in $[1-\delta, 1+\delta]$, we can find numbers c_1, c_2, c_3 such that $\varphi(cx) \le c_1\varphi(x) + c_2|x| + c_3$, for all real x.

Condition (**RC**) holds for all power divergences including KL and KL_m divergences.

For all divergences considered in this paper it will be assumed that for any α and θ in $\mathscr{U}$

$$\int \left|\varphi'\left(\frac{dP_\theta}{dP_\alpha}\right)\right| dP_\theta < \infty. \tag{5.10}$$

We state the following Lemma covering nearly all classical divergences (see Liese and Vajda (1987) [42] and Broniatowski and Kéziou (2006) [12], Lemma 3.2).

Lemma 13. Assume that **RC** holds and $\phi(\theta,\alpha)$ is finite. Then (5.10) holds.

5.2.2 Dual form of the divergence and dual estimators in parametric models

The following representation is the cornerstone of parametric inference through divergence based methods.

Theorem 14. Let θ belong to Θ and let $\phi(\theta,\theta_T)$ be finite. Assume that **RC** holds together with Condition (A) . Then

$$\begin{aligned}\phi(\theta,\theta_T) &= \sup_{\alpha\in\mathscr{U}} \int \varphi'\left(\frac{dP_\theta}{dP_\alpha}\right) dP_\theta - \int \varphi^{\#}\left(\frac{dP_\theta}{dP_\alpha}\right) dP_{\theta_T} \\ &= \sup_{\alpha\in\mathscr{U}} \int h(\theta,\alpha,x) dP_{\theta_T}\end{aligned} \tag{5.11}$$

Furthermore the sup is reached at θ_T and uniqueness holds.

For the Cressie-Read family of divergences with $\gamma \neq 0,1$ this representation writes

$$\begin{aligned}\phi_\gamma(\theta,\theta_T) &= \sup_{\alpha\in\mathscr{U}} \left\{ \frac{1}{\gamma-1}\int \left(\frac{dP_\theta}{dP_\alpha}\right)^{\gamma-1} dP_\theta - \frac{1}{\gamma}\int \left(\frac{dP_\theta}{dP_\alpha}\right)^{\gamma} dP_{\theta_T} - \frac{1}{\gamma(\gamma-1)} \right\} \\ &= \sup_{\alpha\in\mathscr{U}} \int h(\theta,\alpha,x) dP_{\theta_T}\end{aligned}$$

Under the above notation and hypotheses define

$$T_\theta\left(P_{\theta_T}\right) := \arg\sup_{\alpha\in\mathscr{U}} \int h(\theta,\alpha,x) dP_{\theta_T}. \tag{5.12}$$

It then holds, for any θ such that $\phi(\theta,\theta_T)$ is finite

$$T_\theta\left(P_{\theta_T}\right) = \theta_T \quad \text{for all } \theta_T \in \Theta.$$

Also let

$$S\left(P_{\theta_T}\right) := \arg\inf_{\theta\in\Theta} \sup_{\alpha\in\mathscr{U}} \int h(\theta,\alpha,x) dP_{\theta_T}. \tag{5.13}$$

which also satisfies

$$S\left(P_{\theta_T}\right) = \theta_T$$

for all θ_T in Θ. We thus state : under the hypotheses of Theorem 14, both statistical functionals T_θ and S are Fisher consistent.

From (5.11), simple estimators for $\phi(\theta,\theta_T)$ and θ_T can be defined, plugging any convergent empirical measure in place of P_{θ_T} and taking the infimum in θ in the resulting estimator of $\phi(\theta,\theta_T)$.
In the context of simple i.i.d. sampling, introducing the empirical measure $P_n := \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$where the X_i's are i.i.d. r.v's with common unknown distribution P_{θ_T} in $\mathscr{P}$, the natural estimator of $\phi(\theta,\theta_T)$ is

$$\begin{aligned}\phi_n(\theta,\theta_T) &:= \sup_{\alpha\in\mathscr{U}}\left\{\int h(\theta,\alpha,x)\; dP_n(x)\right\}\\ &= \sup_{\alpha\in\mathscr{U}}\int \varphi'\left(\frac{dP_\theta}{dP_\alpha}\right) dP_\theta - \frac{1}{n}\sum_{i=1}^n \varphi^{\#}\left(\frac{dP_\theta}{dP_\alpha}(X_i)\right) \text{ when } (A) \text{ holds.}\end{aligned}$$

As stated in theorem 3.2 in Broniatowski and Keziou [13]:

Theorem 15. Under some derivability assumptions on $\varphi\left(\frac{dP_\theta}{dP_\alpha}\right)$ (conditions A.0 to A.2 in Broniatowski and Keziou [13]),

$$\text{If } \theta = \theta_T, \text{ then } \frac{2n}{\varphi''(1)}\phi_n(\theta,\theta_T) \xrightarrow{d} \chi^2_{(d)} \text{ for } d = \dim(\Theta). \tag{5.14}$$

This last result of convergence of the estimated $\varphi-$divergence is of great interest in the problem we are taking on and serves as the basis for the test procedure that we propose.

5.3 A simple solution for testing finite mixture models

5.3.1 Testing between mixtures of fully characterized components

Let us consider a set of signed measures defined by

$$p_\theta = (1-\theta)p_0 + \theta p_1, \quad \theta \in \mathbb{R}, \tag{5.15}$$

where p_0 and p_1 are two known densities (belonging or not to the same parametric family).

The mixture (5.5) is clearly contained in (5.15) and the case $\theta_T = 0$ is in this framework an interior point of the parameter space $\mathbb{R}$. In relation with (5.5), the case $\theta_T = 0$ is now an interior point of the parameter space.

We observe a random sample $X_1, \ldots, X_n$ of distribution p_T. We are willing to test:

$$H_0 : p_T = p_0 \quad \text{vs} \quad H_1 : p_T = p_\theta \neq p_0 \tag{5.16}$$

which can be reduced to

$$H_0 : \theta = 0 \quad \text{vs} \quad H_1 : \theta \neq 0 \tag{5.17}$$

whenever $p_0 \neq p_1$ is met. The latter condition ensures the identifiability of the model and enables to consider different parametric families for p_0 and p_1. Conversely, Chen and al. [18], for instance, assumes that $0 < \theta < 1$, and tests the equality of the parameters of p_0 and p_1 inside a unique family $\mathscr{F}$.
In the following, we thus assume that $p_0 \neq p_1$.

5.3.2 Test statistics

The choice of the test statistic is driven by the result given in Theorem 15. Accordingly, let ϕ be any divergence associated with convex finite functions and such that 0 is an interior point of the space parameter defined by:

$$\Theta := \left\{ \alpha \in \mathbb{R} : \int | \varphi' \left(\frac{dP_0}{dP_\alpha} \right) | \, dP_0 < \infty \right\} \tag{5.18}$$

Then the statistic $2n\phi_n(0, \theta_T)$ can be used as a test statistic for (5.17) and

$$2n\phi_n(0, \theta_T) \longrightarrow \chi^2_{(1)} \quad \text{when} \quad H_0 \text{ holds.} \tag{5.19}$$

Also, (5.19) holds when testing whether the true distribution is a k_0 component mixture or a k component mixture as in (5.3). In this case, the test statistic $2n\phi_n(\Theta_0, \theta_T)$ converges to a $\chi^2_{(k-k_0)}$ distribution when H_0 holds.
While many divergences meet the former properties, we restrict in the sequel ourselves to two generators.

Chi-square divergence

The first divergence that we consider is the χ^2-divergence. The corresponding φ function $\varphi_2(x) := \frac{1}{2}(x-1)^2$ is defined and convex on whole $\mathbb{R}$; an example when $\mathscr{P}$ may contain signed finite measures and not be restricted to probability measures is considered in

Broniatowski and Keziou [12] in relation with a two components mixture model defined in (5.15) and where θ is allowed to assume values in an open neighborhood Θ of 0, in order to provide a test for (5.17), with θ an interior point of Θ.

Extended Kullback-Leibler divergence

The second divergence that we retain is generated by a function described below, namely

$$\varphi_c(x) := (x+e^c-1)\cdot\log(x+e^c-1)+1-(x+e^c-1)+(1-c)\cdot(e^c-1)-c\cdot x \;\geq\; 0 \qquad x\in]1-e^c,\infty[,\quad c\in\mathbb{R}, \tag{5.20}$$

which has been derived within the recent general framework of Broniatowski and Stummer [16]. It is straightforward to see that φ_c is strictly convex and satisfies $\varphi_c(1) = 0 = \varphi_c'(1)$. For the special choice $c = 0$, (5.20) reduces to the omnipresent Kullback-Leibler divergence generator

$$\varphi_0(x) := x\log x - x + 1 \;\geq\; 0, \qquad x\in]0,\infty[.$$

According to (5.20), in case of $c > 0$ the domain $]1-e^c,\infty[$ of φ_c covers also negative numbers (see Broniatowski and Stummer [15] for insights on divergence-generators with general real-valued domain); thus, the same facts hold for the new generator than for the χ^2 and this opens the gate to considerable comfort in testing mixture-type hypotheses against corresponding marginal-type alternatives, as we derive in the following. We denote KL_c the corresponding divergence functional for which $KL_c(Q,P)$ is well defined whenever P is a probability measure and Q is a signed measure.
It can be noted that, depending on the type of model considered, the validity of the test can be subject to constraints over the parameters of the densities. Indeed, the convergence of $I = \int |\phi'\left(\frac{p_0}{p_\theta}\right)| dP_0$ is not always guaranteed. This kind of considerations may guide the choice of the test statistic. For instance, in some cases, including scaling models, conditions that are required for the χ^2-divergence, do not apply to the KL_c-divergence. For instance, consider a Gaussian mixture model with different variances:

$$p_0 \sim \mathcal{N}(\mu,\sigma_0^2), \qquad p_1 \sim \mathcal{N}(\mu,\sigma_1^2).$$

The convergence of I with the χ^2 requires either $\sigma_1^2 > \sigma_0^2$ or $\sigma_0^2 > \sigma_1^2 > \frac{1}{2}\sigma_0^2$. On the other hand, the convergence is always ensured with the KL_c-divergence.
The same observations can be made for lognormal, exponential and Weibull densities.

5.3.3 Generalization to parametric distributions with unknown parameters

In the previous section, the densities of each component were supposed to be known. We now generalize to the case where the components belong to parametric families with unknown parameter. We therefore deal with a way more complicate issue and consider a generalized test procedure which aggregates tests of simple hypotheses over the components densities parameter spaces.

We present the generalization for a two component mixture, but it is valid as well for k component mixtures with $k \geq 2$. Let us assume $p_0 \in \mathscr{F}_0 = \{p_0(. \mid \lambda_0) : \lambda_0 \in \Lambda_0\}$ and $p_1 \in \mathscr{F}_1 = \{p_1(. \mid \lambda_1) : \lambda_1 \in \Lambda_1\}$, with Λ_0 and Λ_1 being compact subsets of $\mathbb{R}^d$, $d \geq 1$.

We consider aggregated tests of composite hypotheses.

For λ_0 fixed, $H_0(\lambda_0)$ is accepted if $\forall \lambda_1 \in \Lambda_1$, $H_0(\lambda_0)$ is accepted against $H_1(\lambda_0, \lambda_1)$. The aggregated hypothesis $H_0(\Lambda_0)$ is accepted if $\forall \lambda_0 \in \Lambda_0, \lambda_1 \in \Lambda_1$, $H_0(\lambda_0)$ is accepted against $H_1(\lambda_0, \lambda_1)$.

Thus the null hypothesis of homogeneity of the population is rejected when there exists at least one couple of parameters $(\lambda_0^*, \lambda_1^*) \in \Lambda_0 \times \Lambda_1$ with $\lambda_1^* \neq \lambda_0^*$ for which the simple hypothesis $H_0(\lambda_0^*)$ is rejected in favor of $H_1(\lambda_0^*, \lambda_1^*)$. In other words, the possibility that the underlying distribution is a mixture is enough for us to reject that there is a unique component.

Another perspective would be to consider that the null hypothesis $H_0(\Lambda_0)$ is rejected when there is no $\lambda_0 \in \Lambda_0$ such that $\forall \lambda_1 \in \Lambda_1$, $H_0(\lambda_0)$ is accepted against $H_1(\lambda_0, \lambda_1)$.

Note the condition $\{\lambda_1^* \neq \lambda_0^*\}$ is only required when p_0 and p_1 belong to the same parametric family.

Let $2n\phi_n(0, \theta_T \mid \lambda_0, \lambda_1)$ be the test statistic of the test (5.16) of the simple hypotheses $H_0(\lambda_0)$ vs $H_1(\lambda_0, \lambda_1)$ when λ_0 and λ_1 are fixed. Recall that $\phi_n(0, \theta_T \mid \lambda_0, \lambda_1)$ is the estimated divergence between $p_0(. \mid \lambda_0)$ and $p_{\theta_T}(. \mid \lambda_0, \lambda_1)$. The test statistic for (5.17) is derived from:

$$\Phi_n(0, \theta_T) = \sup_{\lambda_0 \in \Lambda_0} \sup_{\lambda_1 \in \Lambda_1 \setminus \lambda_0} \phi_n(0, \lambda_T \mid \lambda_0, \lambda_1) \tag{5.21}$$

where the parameter spaces Λ_0 and Λ_1 can be discretized in $\Lambda_{0,n}$ and $\Lambda_{1,n}$ for the sake of computational complexity.

In order to facilitate the computation of the test statistic, the successive optimizations

have been rearranged as follows:

$$\Phi_n(0,\theta_T) = \sup_{\alpha\in\Theta}\left\{\sup_{\lambda_0\in\Lambda_0}\sup_{\lambda_1\in\Lambda_1\setminus\lambda_0}\int\varphi'\left(\frac{dP_{0,\lambda_0}}{dP_{\alpha,(\lambda_0,\lambda_1)}}\right)dP_{0,\lambda_0} - \frac{1}{n}\sum_{i=1}^{n}\varphi^{\#}\left(\frac{dP_{0,\lambda_0}}{dP_{\alpha,(\lambda_0,\lambda_1)}}(X_i)\right)\right\} \tag{5.22}$$

The critical region R_{Λ_0,Λ_1} associated with the aggregated test can be defined as follows:

$$\left\{\Phi_n(0,\theta_T)\in R_{\Lambda_0,\Lambda_1}\right\} = \cup_{\lambda_0\in\Lambda_0}\cup_{\lambda_1\in\Lambda_1}\left\{\phi_n(0,\theta_T\mid\lambda_0,\lambda_1)\in R_{\lambda_0,\lambda_1}(\alpha)\right\} \tag{5.23}$$

where $R_{\lambda_0,\lambda_1}(\alpha)$ is the critical region of risk α for the test of the simple hypotheses $H_0(\lambda_0)$ vs $H_0(\lambda_0,\lambda_1)$. α can then be tuned to ensure that the probability of (5.23) is of the wanted first kind level of risk $\alpha^\star$ for the global test.

Note that in this case, we do not have an equivalence to Theorem 15. Indeed, we do not directly estimate the true parameters of the densities p_0 and p_1, but rather aggregate the test over the parameter spaces Λ_0 and Λ_1. Thus, there is still no convergence result on the test statistic Φ_n. In the following, we evaluate the performance of the proposed test procedure through numerical simulations.

5.4 Numerical simulations

5.4.1 Mixture of fully characterized components

We here consider the simple case of a mixture between a lognormal and a Weibull distribution whose parameters are supposed to be known. Results in Table 5.1 show that the test procedure of simple hypotheses performs well when the two components are fully characterized.

	Lognormal and Weibull Mixture $lN(\lambda_0,0.2)$ vs $0.8lN(\lambda_0,0.2)+0.2\mathscr{W}(\lambda_1,2)$			
n=250 observations				
	χ^2 **test statistic**		$\mathbf{KL}_c$ **test statistic**	
First kind risk	0.05	0.10	0.05	0.10
Power	0.98	1	0.99	1

TABLE 5.1 – Power of the test for a lognormal and Weibull mixture with fully characterized components

5.4.2 Mixture of unknown components within a parametric family

The performances of the test procedure are evaluated numerically on three two-component mixtures. In the first two examples, both components belong to the same parametric family, while in the third, p_0 and p_1 are from different models. In each case, the distributions of the components are such that the resulting mixture is not bimodal. The following results are used on an illustrative basis. The critical regions are determined as to guarantee the value of the first kind risk $\alpha^\star$ at 0.05 and 0.10.

Lognormal mixture

We first consider a Lognormal mixture. The two components belonging to the same parametric family, we can compare the performance of the divergence based test with the modified likelihood ratio test proposed by Chen and al.[18].
The alternate hypotheses are the following:

$$H_0 : p_T = p_0 \sim lN(\lambda_0, 1) \quad vs \quad H_1 : p_T = p_\theta \sim (1-\theta)lN(\lambda_0, 1) + \theta lN(\lambda_1, 1),$$

where $\lambda_0 \in \Lambda_0 = [0.4, 1.6]$ and $\lambda_1 \in \Lambda_1 = [1.4, 2.6]$.
The critical region is computed numerically through Monte Carlo simulations under H_0. The power of the test is also computed numerically when the realizations are drawn from the mixture model with $\theta = 0.2, \lambda_0 = 1$ and $\lambda_1 = 2$ for the χ^2 and KL_c test statistics and Chen's modified likelihood ratio.
The results in table 5.2 show that both χ^2 and KL_c outperform the modified likelihood ratio test and the test based on the KL_c divergence achieves in this case the greatest power.

	Lognormal Mixture $lN(1,1)$ vs $0.8lN(1,1) + 0.2lN(2,1)$					
n=250 observations						
	χ^2 test statistic		**KL_c test statistic**		**Chen's modified lik ratio**	
First kind risk	0.05	0.10	0.05	0.10	0.05	0.10
Power	0.22	0.41	0.50	0.65	0.12	0.18

TABLE 5.2 – Power of the tests for three types of mixtures whose components are Lognormal

Gamma mixture

We test the following hypothesis

$$H_0 : p_T = p_0 \sim \mathcal{G}(\lambda_0, 1) \quad vs \quad H_1 : p_T = p_\theta \sim (1-\theta)\mathcal{G}(\lambda_0, 1) + \theta\mathcal{G}(\lambda_1, 2),$$

where $\lambda_0 \in \Lambda_0 = [1.4, 2.6]$ and $\lambda_1 \in \Lambda_1 = [4.4, 5.6]$. The realizations are drawn from the mixture model with $\theta = 0.2, \lambda_0 = 2$ and $\lambda_1 = 5$.
Here again, both divergence based statistics achieves higher power than the modified likelihood ratio test (cf. table 5.3).

	Gamma Mixture $\mathcal{G}(2,1)$ vs $0.8\mathcal{G}(2,1) + 0.2\mathcal{G}(5,2)$					
n=250 observations						
	χ^2 test statistic		**KL$_c$ test statistic**		**Chen's modified lik ratio**	
First kind risk	0.05	0.10	0.05	0.10	0.05	0.10
Power	0.31	0.46	0.35	0.45	0.13	0.22

TABLE 5.3 – Power of the tests for three types of mixtures whose components are Gamma distributed

Weibull and Lognormal mixture

We here consider the case where the two components are from different parametric families. We want to test

$$H_0 : p_T = p_0 \sim lN(\lambda_0, 0.2) \quad vs \quad H_1 : p_T = p_\theta \sim (1-\theta)lN(\lambda_0, 0.2) + \theta\mathcal{W}(\lambda_1, 2),$$

where $\lambda_0 \in \Lambda_0 = [0.4, 1.6]$ and $\lambda_1 \in \Lambda_1 = [2.4, 3.6]$. The realizations are drawn from the mixture model with $\theta = 1, \lambda_0 = 1$ and $\lambda_1 = 3$.
The results in table 5.4 show that the test based on the KL_c divergence performs better than the χ^2 statistic.

	Lognormal and Weibull Mixture $lN(\lambda_0, 0.2)$ vs $0.8lN(\lambda_0, 0.2) + 0.2\mathcal{W}(\lambda_1, 2)$			
n=250 observations				
	χ^2 test statistic		**KL$_c$ test statistic**	
First kind risk	0.05	0.10	0.05	0.10
Power	0.28	0.47	0.34	0.57

TABLE 5.4 – Power of the tests for three types of mixtures whose components belong to a different parametric family

Concerning the choice of the test statistic, we might note that the KL_c performs better when the two alternate distributions differ mainly in their central tendency, while the χ^2 might be prefered when the difference lays in the tails.

5.5 Concluding remarks

The test procedure proposed in this chapter enables to discriminate between a k-component and a $(k+p)$-component mixture, $k, p > 0$. Based on previous work from Broniatowski and Keziou [13], this chapter extends it to the case of components from parametric families with unknown parameters. The test statistic is an aggregation of divergence based statistics in their dual form.
Though theoretically applicable to any $k-$component mixture, the performances of the test have been studied on several two-component models, whose components either belong from the same parametric family or not. It has yet to be run on models of $k > 2$ components.

Chapitre 6

Conclusion générale et perspectives

Les travaux présentés dans cette thèse apportent des outils méthodologiques et de modélisation ; le fil conducteur étant la caractérisation des risques extrêmes en fatigue des matériaux.

Méthode séquentielle de planification d'essai

Apports

La première partie porte sur l'étude de la fatigue à très grands nombres de cycles. L'absence de méthodologie visant spécifiquement à estimer un quantile extrême sur des données binaires de dépassements de seuils a motivé le développement d'une nouvelle méthode de planification d'essais séquentielle inspirée du Splitting. Celle-ci se fonde sur un échantillonnage dans les régions de plus en plus extrêmes de la distribution de la résistance du matériau. La modélisation proposée exploite la structure du splitting en supposant la stabilité par seuillage des lois considérées à travers un modèle de Pareto généralisé et une adaptation d'un modèle de Weibull. A cette procédure est associée une méthode d'estimation tirant partie de la modélisation et de la dimension itérative pour pallier la nature dégradée de l'information disponible.

Perspectives

Les contraintes en termes de nature et de quantité de données rendent difficile d'envisager une modélisation plus flexible de la résistance du matériau. En effet, l'inflation de paramètres et la perte de la propriété de stabilité par seuillage compliquent largement la procédure d'estimation. Quelques pistes de généralisation exploitant la complète monotonie des modèles étudiés ont été ébauchées mais nécessiteraient une étude plus poussée.

Modélisation des courbes S-N

Apports

La seconde partie se concentre sur l'étude de la tenue en fatigue olygocyclique à travers la construction des courbes S-N. Les travaux réalisés dans le cadre de cette étude ont porté sur des questionnements autour de la modélisation de la durée de vie à niveaux de sollicitations fixés.

L'idée développée dans le chapitre 3.2 consiste à faire reposer la modélisation de la durée de vie sur des résultats en mécanique des matériaux. Ainsi l'hétérogénéité des données s'explique par la coexistence d'amorçages courts et d'amorçages lents, survenant en proportions inversement proportionnelles selon le niveau de sollicitation. Cependant la construction d'un modèle adapté aux données suppose de répondre au préalable à un certain nombre de questions : Quelle est la loi des durées de vie en amorçages courts et lents ? A partir de quel niveau de chargement y a-t-il hétérogénéité de la population de durées de vie ? etc...

Les chapitres suivants ont consisté à apporter des outils statistiques permettant d'affiner le choix de la modélisation. Le chapitre 4 a introduit une procédure de tests d'hypothèses composites visant à discriminer entre deux hypothèses portant sur la loi d'une variable bruitée. Elle repose sur l'agrégation de tests du rapport de vraisemblance. Cette procédure est directement adaptable à la question de l'étude de la loi de la durée de vie. Dans le cas des faibles niveaux de contrainte pris en exemple dans le chapitre, l'essentiel de la durée de vie correspond à la propagation d'une fissure amorcée très tôt. Dans ce cadre, il s'agit de tester si la durée de vie en propagation est affectée d'un bruit additif correspondant au temps d'amorçage.

Le chapitre 5 introduit une autre procédure de test visant à déterminer le nombre de composantes d'un mélange. La statistique de test est construite à partir d'estimateurs de divergence obtenus en exploitant la forme duale de la divergence dans un cadre paramétrique. Cette méthodologie fournit un outil permettant de déterminer les seuils de sollicitation au delà et en deçà desquels la population n'est plus homogène. Elle peut également être utilisée pour valider l'hypothèse selon laquelle seuls deux modes de rupture coexistent. Elle apporte des réponses complémentaires à la procédure portant sur la nature de chaque sous-population, laquelle vise à spécifier les hypothèses paramétriques les plus adaptées aux différentes composantes dans les plages de contraintes pour lesquelles la population est homogène.

Perspectives

Les travaux présentés portent sur le développement d'aide à la modélisation. Les procédures de test ont pour l'instant été évaluées par simulations afin d'en étudier la puissance. Concernant le test sur le nombre de composantes d'un mélange, les essais numériques portent seulement sur des modèles à deux composantes. Il serait donc intéressant de mener des simulations supplémentaires sur des mélanges à $k > 2$ composantes. Il reste maintenant à appliquer ces outils sur les données de fatigue.

Ces outils fournissent une base permettant d'envisager un autre axe de recherche : l'estimation de la distribution de la durée de vie sur le modèle de mélange sélectionné. Une piste envisageable consiste à adapter un algorithme de type EM à l'estimation de quantile. En effet, si ce type d'algorithme permet d'estimer convenablement la durée de vie moyenne, il en va différemment en ce qui concerne la durée de vie minimale.
Considérons le modèle de mélange suivant :

$$f_\theta(y) = \lambda_1 f_1(y \mid \theta_1) + \lambda_2 f_2(y \mid \theta_2)$$

où Y est la variable observée de densité f_θ, X la variable cachée de classe et $\theta = (\theta_1, \theta_2, \lambda_1, \lambda_2)$ le vecteur de paramètres.
La vraisemblance complète du modèle est donnée par :

$$\log L(X, Y \mid \theta) = \sum_{i=1}^{n} x_i \log(\lambda_1 f_1(y_i \mid \theta_1)) + (1 - x_i) \log(\lambda_2 f_2(y_i \mid \theta_2).$$

Les deux étapes de l'algorithme classique sont données par :
Étape E : Calcul de l'espérance de la log-vraisemblance complète :

$$Q\left(\theta, \theta^{(k)}\right) = \mathbb{E}_{\theta^{(k)}}\left(\log L(X, Y \mid \theta) \mid X\right).$$

Étape M : Estimation de θ en maximisant Q :

$$\theta^{(k+1)} = \underset{\theta \in \Theta}{\operatorname{argmax}}\, Q\left(\theta, \theta^{(k)}\right).$$

Le terme à maximiser Q peut se décomposer sous la forme de la somme de deux termes :

$$
\begin{aligned}
\underset{\theta\in\Theta}{\operatorname{argmax}}\,\mathbb{E}_{\theta^{(k)}}\left(\log L(X,Y\mid\theta)\mid X\right) &= \underset{\theta\in\Theta}{\operatorname{argmax}}\sum_i^n\int \log(f(y_i,x\mid\theta))h(x\mid\theta^{(k)})dx \\
&\qquad \text{avec} \quad h(x\mid\theta^{(k)}) = \frac{f(y_i,x\mid\theta)}{f(y\mid\theta^{(k)})} \\
&= \underset{\theta\in\Theta}{\operatorname{argmax}}\sum_i^n\int \log\left(h(x\mid\theta)f(y\mid\theta)\right)h(x\mid\theta^{(k)})dx \\
&\dots \\
&= \underset{\theta\in\Theta}{\operatorname{argmax}}\sum_i^n \log(f(y_i\mid\theta)) \\
&+ \sum_i^n\int \log\left(\frac{h(x\mid\theta)}{h(x\mid\theta^{(k)})}\right)h(x\mid\theta^{(k)}) \\
&= \underset{\theta\in\Theta}{\operatorname{argmin}}\,\widehat{D}_\varphi(f_\theta,f_T) + \frac{1}{n}D_\psi(\theta,\theta^{(k)}).
\end{aligned}
$$

Classiquement, $D_\varphi(f_\theta, f_T)$ correspond à la log-vraisemblance du modèle. Cependant la décomposition ainsi obtenue fait apparaître une généralisation possible de l'algorithme à toute divergence prise entre la vraie densité f_T et le modèle f_θ.
En particulier, s'agissant d'estimer un quantile, un choix possible est la divergence du χ^2

$$\chi(f_\theta,f_T) = \int \frac{1}{2}\left(\frac{f_\theta}{f_T}(x)-1\right)^2 f_T(x)dx,$$

dont la forme pénalise plus fortement les petits écarts, et qui pourrait permettre de gagner en précision sur l'estimation des queues de distribution.

Bibliographie

[1] Azais, J.M., Gassiat, E., Mercadier, C. : Asymptotic distribution and local power of the log-likelihood ratio test for mixtures : bounded and unbounded cases. *Bernoulli* 12(5), 775–799 (2006).

[2] Bahadur, R.R. Stochastic comparison of tests. *Ann. Math. Statist.* 31, 276–295 (1960).

[3] Bahadur, R.R. : *Some limit theorems in statistics*. Society for Industrial and Applied Mathematics : Philadelpha, PA, USA (1971).

[4] Balkema A. A. and De Haan L. : Residual Life Time at Great Age. *Ann. Prob.*, vol. 2 (5) pp. 762 - 804 (1974).

[5] Barndorff-Nielsen, O. *Information and exponential families in statistical theory*. John Wiley & Sons : New York, NY, USA (1978).

[6] Basu, A. ; Shioya, H. ; Park C. *Statistical inference : The minimum distance approach*. CRC Press : Boca Raton, FL, USA (2011).

[7] Birgé, L. Vitesses maximales de décroissance des erreurs et tests optimaux associés. *Z. Wahrsch. Verw. Gebiete*, *55*, 261–273 (1981).

[8] Biret, M. and Broniatowski, M.: SAFIP : a streaming algorithm for inverse problems. *Arxiv* (2016).

[9] Beirlant, J. ; Guillou, A. and Dierckx, G. and Fils-Villetard, A. : Estimation of the extreme value index and extreme quantiles under random censoring. *Extremes*, vol. 10 (3) pp. 151 - 174 (2007).

[10] Broniatowski, M. ; Jureckova, J. ; Kumar, A.M. and Miranda, E. : Composite Tests under Corrupted Data. *Entropy* 21(1), 63 (2019).

[11] Broniatowski, M. ; Jurečková, J. and Kalina, J. Likelihood ratio testing under measurement errors. *Entropy*, 20 (12), 966 (2018).

[12] Broniatowski, M. and Keziou, A. : Minimization of ϕ-divergences on sets of signed measures. *Studia Sci. Math. Hungar.*, 43(4), 403–442 (2006).

[13] Broniatowski, M. and Keziou, A : Parametric estimation and tests through divergences and the duality technique. *Journal of Multivariate Analysis* 100, 16–36 (2009).

[14] Broniatowski M. ; Miranda E. and Stummer W. : Testing the number and the nature of the components in a mixture distribution. *GSI 2019, Springer, Geometric Science of Information, lecture notes in computer science* (2019).

[15] Broniatowski, M. and Stummer, W. : Some universal insights on divergences for statistics, machine learning and artificial intelligence. In : Nielsen, F. (ed.) *Geometric Structures of Information*, pp. 149—211. Springer Nature, Switzerland (2019).

[16] Broniatowski, M. and Stummer, W. : A bare simulation approach to finding minimum distances. *Preprint* (2019).

[17] Castillo, N.O. ; Gallardo, D.I. ; Bolfarine,H. and Gomez,H.W. : Truncated Power-Normal Distribution with Application to Non-Negative Measurements. *Entropy*, 20(6), 433 (2018).

[18] Chen, H. ; Chen, J. and Kalbfleisch, J.D. : A modified likelihood ratio test for homogeneity in finite mixture models. *J.R. Statist. Soc.* B 63(1), 19–29 (2001).

[19] Chernoff, H. : Large-sample theory : Parametric case. *Ann. Math. Statist.* 27, 1–22 (1956).

[20] Choulakian, V., Stephens, M. A. : Goodness-of-fit tests for the generalized Pareto distribution. *Technometrics* 43(4), 4978–484 (2001)

[21] Ciuperca, G. : Likelihood Ratio Statistic for Exponential Mixtures. *Ann. Inst. Stat. Math.*, 54(3), 585-594 (2002).

[22] Cressie, N. and Read, T. R. C. : Multinomial goodness-of-fit tests. *J. Roy. Statist. Soc.* Ser. B, 46(3), 440–464 (1984).

[23] Csiszár, I. : Eine informationstheoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizität von Markoffschen Ketten. *Magyar Tud. Akad. Mat. Kutató Int. Közl.*, 8, 85–108 (1963).

[24] De Haan, L. and Rootzén, H. : On the estimation of high quantiles. *J. Statist. Plann. Inference*, vol. 35 (1) pp. 1 - 13 (1993).

[25] Dekkers, A.L.M., Einmahl, J. H. J. and De Haan L. : A moment estimator for the index of an extreme-value distribution. *Ann. Statist.*, vol. 17 (4) pp. 1833—1855 (1989).

[26] De Valk, C. : Approximation of high quantiles from intermediate quantiles. *Extremes*, vol. 4 pp. 661 - 684 (2016).

[27] De Valk, C. and Cai, J.J : A high quantile estimator based on the log-Generalised Weibull tail limit. *Econometrics and Statistics*, vol. 6 pp. 107 - 128 (2018).

[28] Dietrich, D., De Haan, L. and Hüsler, J. : Testing extreme value conditions. *Extremes* 5(1), 71–85 (2002).

[29] Dixon, W. J. and Mood, A. M.: A Method for Obtaining and Analyzing Sensitivity Data. *Journal of the American Statistical Association*, vol. 43 pp. 109 - 126 (1948).

[30] Dixon, W. J.: The Up-and-Down Method for Small Samples. *Journal of the American Statistical Association*, vol. 60 pp. 967 - 978 (1965).

[31] Drees, H., De Haan, L. and Li, D.: Approximations to the tail empirical distribution function with application to testing extreme valueconditions. *J. Statist. Plann. Inference* 136(10), 3498–3538 (2006).

[32] Eguchi, S. and Copas, J. : Interpreting Kullback-Leibler divergence with the Neyman-Pearson lemma. *J. Multivar. Anal.*, 97, 2034–2040 (2005).

[33] Einmahl, J.H.J, Fils-Villetard, A. and Guillou, A.: Statistics of extremes under random censoring, *Bernoulli*, vol 14(1) pp. 207–227 (2008).

[34] Feller, W.: An introduction to probability theory and its applications. Wiley, vol 2 (1971).

[35] Fouchereau, R.: Modélisation probabiliste des courbes S-N. Université Paris Sud (2014).

[36] Goldie, C.: A class of infinitely divisible random variables. *Proc. Cambridge Philos. Soc.*, 63, 1141–1143 (1967).

[37] Guo, D.: Relative entropy and score function: New information-estimation relationships through arbitrary additive perturbation. *ISIT 2009, Proceedings IEEE International Symposium on Information Theory*, June 28-July 3, 2009, Seoul, Korea, pp. 814–818 (2009).

[38] Huber, P. and Strassen, V.: Minimax tests and the Neyman-Pearson lemma for capacities. *Ann. Statist.*, 2, 251–273 (1973).

[39] Jurečková, J. and Picek, J.: A Class of Tests on the Tail Index. *Extremes*, 4, 165 – 183 (2001).

[40] Kahn, H. and E Harris, T.: Estimation of Particle Transmission by Random Sampling. *National Bureau of Standards Applied Mathematics Series*, vol. 12 (1951).

[41] Krafft, O. and Plachky, D.: Bounds for the power of likelihood ratio tests and their asymptotic properties. *Ann. Math. Statist.*, 41, 1646–1654 (1970).

[42] Liese, F. and Vajda, I.: Convex statistical distances. *BSB B. G. Teubner Verlagsgesellschaft vol 95.*, Leipzig (1987).

[43] Miranda, E., Broniatowski, M. and Biret, M.: Stratégie de planification d'essais pour la caractérisation de contrainte admissible en fatigue des matériaux. *I-Revues CNRS, Actes, Actes du 21ème congrès Lambda-Mu* (2018).

[44] Naveau, P. ; Huser, R. ; Ribereau, P. and Hannart A. : Modeling jointly low, moderate, and heavy rainfall intensities without a threshold selection. *Water Resources Research*, vol. 52 (4) pp. 2753 - 2769 (2016).

[45] O'Quigley, J. ; Pepe, M. and Fisher, L.l.: Continual Reassessment Method: A Practical Design for Phase 1 Clinical Trials in Cancer. *Biometrics*, vol. 46 pp. 33 - 48 (1990).

[46] Pardo, L. : Statistical inference based on divergence measures. *Statistics : Textbooks and Monographs*, 185. Chapman & Hall/CRC, Boca Raton, FL (2006).

[47] Pickands, J. : Statistical Inference using extreme order statistics. *Ann. Prob.*, vol. 3 (1) pp. 119 - 131 (1975).

[48] Rényi, A. : On measures of entropy and information. *Proc. 4th Berkeley Sympos. Math. Statist. and Prob.*, vol. 1 pp. 547–561 Univ. California Press, Berkeley, Calif (1961).

[49] Self, S. G. and Liang, K.-Y. : Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *J. Amer. Statist. Assoc.*, 82(398), 605–610 (1987).

[50] Narayanan, K.R.and Srinivasa, A.R. On the thermodynamic temperature of a general distribution. *Preprint available online : https://arxiv.org/abs/0711.1460* (2007).

[51] Thorin, O. : On the infinite divisibility of the Pareto distribution. *Scand. Actuarial J.*, vol. 1 pp. 31 - 40 (1977).

[52] Thyrion, P. : Les lois exponentielles composées. *Bulletin de l'Association Royale des Actuaires Belges*, vol. 62 pp. 35 - 44 (1964).

[53] Titterington, D. M., Smith, A. F. M., and Makov, U. E. : Statistical analysis of finite mixture distributions. *Wiley Series in Probability and Mathematical Statistics : Applied Probability and Statistics*. John Wiley & Sons Ltd., Chichester (1985).

[54] Tsallis, C. Possible generalization of BG statistics. *J. Stat. Phys.*, 52, 479–485 (1987).

[55] Tusnády, G. On asymptotically optimal tests. *Ann. Statist.*, 5, 385–393 (1987).

[56] Weissman, I. :Estimation of parameters and large quantiles based on the k largest observations. *J. Amer. Statist. Assoc.*, vol. 73 pp. 812 - 815 (1978).

[57] Worms, J. and Worms, R.: New estimators of the extreme value index under random right censoring, for heavy-tailed distributions. *Extremes*, vol. 17 pp. 337 - 358 (2014).