entropy

MDPI

Article

Composite Tests under Corrupted Data

Michel Broniatowski [1,*], Jana Jurečková [2,3], Ashok Kumar Moses [4] and Emilie Miranda [1,5]

1 Laboratoire de Probabilités, Statistique et Modélisation, Sorbonne Université, 75005 Paris, France; emilie.miranda@upmc.fr
2 Institute of Information Theory and Automation, The Czech Academy of Sciences, 18208 Prague, Czech Republic; jurecko@karlin.mff.cuni.cz
3 Faculty of Mathematics and Physics, Charles University, 18207 Prague, Czech Republic
4 Department of ECE, Indian Institute of Technology, Palakkad 560012, India; ashokm@iitpkd.ac.in
5 Safran Aircraft Engines, 77550 Moissy-Cramayel, France
* Correspondence: michel.broniatowski@sorbonne-universite.fr

Received: 11 November 2018; Accepted: 10 January 2019; Published: 14 January 2019

check for updates

Abstract: This paper focuses on test procedures under corrupted data. We assume that the observations Z_i are mismeasured, due to the presence of measurement errors. Thus, instead of Z_i for $i = 1, \ldots, n$, we observe $X_i = Z_i + \sqrt{\delta} V_i$, with an unknown parameter δ and an unobservable random variable V_i. It is assumed that the random variables Z_i are i.i.d., as are the X_i and the V_i. The test procedure aims at deciding between two simple hyptheses pertaining to the density of the variable Z_i, namely f_0 and g_0. In this setting, the density of the V_i is supposed to be known. The procedure which we propose aggregates likelihood ratios for a collection of values of δ. A new definition of least-favorable hypotheses for the aggregate family of tests is presented, and a relation with the Kullback-Leibler divergence between the sets $(f_\delta)_\delta$ and $(g_\delta)_\delta$ is presented. Finite-sample lower bounds for the power of these tests are presented, both through analytical inequalities and through simulation under the least-favorable hypotheses. Since no optimality holds for the aggregation of likelihood ratio tests, a similar procedure is proposed, replacing the individual likelihood ratio by some divergence based test statistics. It is shown and discussed that the resulting aggregated test may perform better than the aggregate likelihood ratio procedure.

Keywords: composite hypotheses; corrupted data; least-favorable hypotheses; Neyman Pearson test; divergence based testing; Chernoff Stein lemma

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 made based on a sample of such items, each with the same distribution F_0 (with density f_0) or G_0 (with density g_0). The measurement device adds a random noise V_δ to each measurement, mutually independent and independent of the item, with a common distribution function H_δ and density h_δ, where δ is an unknown scaling parameter. Therefore the density of the measurement $X := Z + V_\delta$ is either $f_\delta := f_0 * h_\delta$ or $g_\delta := g_0 * h_\delta$, where $*$ denotes the convolution operation. We denote F_δ (respectively G_δ) to be the distribution function with density f_δ (respectively g_δ).

The problem of interest, studied in [1], 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 [2] enables us to estimate the true log-likelihood ratio (true Kullback-Leibler divergence) even when we only dispose of locally perturbed data by additive measurement errors. The distribution function H_0 of the 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 [1], that for small δ the original likelihood ratio test (LRT) is still the most powerful, only on a slightly changed significance level. The test problem leads to a composite of null and alternative classes $\mathbf{H}_0$ or $\mathbf{H}_1$ of distributions of random variables $Z + V_\delta$ with $V_\delta := \sqrt{\delta}V$, where V has distribution H_1. 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 [3]). Moreover, Eguchi and Copas [4] showed that the overall loss of power caused by a misspecified alternative equals 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 [2] and of [5] enable us 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 [1], and is supplemented with numerical illustration.

Statement of the Test Problem

Our 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 [1], 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; indeed, in the tuning of the offset of a measurement device, it is customary to perform a large number of observations on the noise under a 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 a given δ in $[\delta_{\min}, \delta_{\max}]$ with $\delta_{\min} > 0$, denote by 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. A classical example is when $V_\delta = \sqrt{\delta}V$, but at times we have a 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 i.i.d. measurements $X_i := Z_i + V_{\delta,i}$ is performed, where $\delta > 0$ is unknown, and we consider the family of tests of $\mathbf{H}_0(\delta)$:=[X has density f_δ] vs. $\mathbf{H}_1(\delta)$:=[X has density g_δ], with $\delta \in \Delta = [\delta_{\min}, \delta_{\max}]$. Only the X_i are observed. A class of combined tests of $\mathbf{H}_0$ vs. $\mathbf{H}_1$ is proposed, in the spirit of [6–9].

Under every fixed n, we assume that δ is allowed to run over a finite set p_n of 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 [1], 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

$$\lim_{n\to\infty} \frac{\log p_n}{n} = 0. \tag{1}$$

For the sake of notational brevity, we denote by Δ 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, $F_\delta(B)$ (respectively $G_\delta(B)$) designates the probability of B under the distribution F_δ (respectively 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

$$T_n(X_1, ..., X_n) > A_n, \tag{2}$$

such that

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

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 [8], 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.

The present paper handles some issues with respect to this context. In Section 2, we consider some test procedures 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 a lower bound for its power is provided. In Section 3, we develop an asymptotic approach to the Least Favorable Hypotheses (LFH) for these tests. We 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.

We next consider, in Section 3.3, the performance of the test numerically; indeed, under the least-favorable pair of hypotheses we compare the power of the test (as obtained through simulation) with the theoretical lower bound, as obtained in Section 2. We show that the minimal power, as measured under the LFH, is indeed larger than the theoretical lower bound—this result shows that the simulation results overperform on theoretical bounds. These results are developed in a number of examples.

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 substituting those ratios with other kinds of scores in the family of divergence-based concepts, extending the likelihood ratio in a natural way. Such an approach has a long history, stemming from the seminal book by Liese and Vajda [10]. 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., [11]). We explore the properties of those new tests under the pair 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 an explanation for this fact. This is the scope of Section 4.

2. An Extension of the Likelihood Ratio Test

For any δ in Δ, let

$$T_{n,\delta} := \frac{1}{n}\sum_{i=1}^{n} \log \frac{g_\delta}{f_\delta}(X_i), \tag{3}$$

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. $\mathbf{H1}(\delta)$, 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 by

$$T_n > A_n,$$

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 that

$$P_{\mathbf{H0}}(\mathbf{H1}) \leq p_n \max_{\delta\in\Delta} \max_{\delta'\in\Delta} F_\delta\left(T_{n,\delta'} > A_n\right). \tag{4}$$

An upper bound for $P_{\mathbf{H0}}(\mathbf{H1})$ can be obtained, making use of the Chernoff inequality for the right side of (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 us to define the threshold A_n accordingly.

Turning to the power of this test, we define the risk of second kind by

$$\begin{aligned} P_{H_1}(\mathbf{H0}) &:= \sup_{\eta\in\Delta} G_\eta\left(T_n \leq A_n\right) \qquad (5) \\ &= \sup_{\eta\in\Delta} G_\eta\left(\sup_{\delta\in\Delta} T_{n,\delta} \leq A_n\right) \\ &= \sup_{\eta\in\Delta} G_\eta\left(\bigcap_{\delta\in\Delta} T_{n,\delta} \leq A_n\right) \\ &\leq \sup_{\eta\in\Delta} G_\eta\left(T_{n,\eta} \leq A_n\right), \end{aligned}$$

a crude bound which, in turn, can be bounded from above through the Chernoff inequality, which yields a lower bound for the 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:

$$\sup_{\delta\in\Delta} \sup_{\delta'\in\Delta} \int \log \frac{f_{\delta'}}{g_{\delta'}} f_\delta < 0. \tag{6}$$

Remark 1. *Since*

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

making use of the Chernoff-Stein Lemma (see Theorem A1 in the Appendix A), Hypothesis (6) means 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 (7) and (8), which are defined below, 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

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

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

$$\mathcal{N}_{\delta,\delta'} \text{ is a non void open neighborhood of } 0. \tag{7}$$

Define, further,

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

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

$$\mathcal{M}_\eta \text{ is a non void neighborhood of } 0. \tag{8}$$

Let

$$I_\eta(x) := \sup_t tx - \log E_{G_\eta}\left(\exp\left(tW_\eta\right)\right), \tag{9}$$

and

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

We also assume an accessory condition on the support of $Z_{\delta'}$ and W_η, respectively under F_δ and under G_η (see (A2) and (A5) in the proof of Theorem A1). Suppose the regularity assumptions (7) and (8) are fulfilled for all δ, δ' and η. Assume, further, that p_n fulfills (1).

The following result holds:

Proposition 2. *Whenever (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, that

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

and

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

3. Minimax Tests under Noisy Data, Least-Favorable Hypotheses

3.1. 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 of the definition and existence of such a least-favourable pair of hypotheses in an asymptotic context; indeed, for a fixed sample size, the usual definition only leads to an explicit definition in very specific cases. Unlike in [1], 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

$$\min_{\delta\in\Delta} K\left(F_\delta, G_\delta\right) = K\left(F_{\delta_*}, G_{\delta_*}\right). \tag{10}$$

We shall call the pair of distributions $\left(F_{\underline{\delta}},\ G_{\underline{\delta}}\right)$ least-favorable for the sequence of tests $1\left\{T_n > A_n\right\}$ if they satisfy

$$\begin{aligned} F_\delta\left(T_n \le A_n\right) &\ge F_{\underline{\delta}}\left(T_n \le A_n\right) \\ &\ge G_{\underline{\delta}}\left(T_n \le A_n\right) \ge G_\delta\left(T_n \le A_n\right) \end{aligned} \tag{11}$$

for all $\delta \in \Delta$. The condition of unbiasedness of the test is captured by the central inequality in (11).

Because, for finite n, such a pair can be constructed only in few cases, we should take a recourse of (11) to the asymptotics $n \to \infty$. We shall show that any pair of distributions $\left(F_{\delta_*} G_{\delta_*}\right)$ achieving (10) are least-favorable. Indeed, it satisfies the inequality (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 $\delta \in \Delta$,

$$\begin{aligned} \liminf_{n\to\infty} \frac{1}{n} \log F_{\underline{\delta}}\left(T_n \le A_n\right) &\ge \lim_{n\to\infty} \frac{1}{n} \log G_{\underline{\delta}}\left(T_n \le A_n\right) \\ &\ge \limsup_{n\to\infty} \frac{1}{n} \log G_\delta\left(T_n \le A_n\right). \end{aligned} \tag{12}$$

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,

$$\alpha_n < 1 - \sup_{\delta\in\Delta} d_{TV}\left(F_\delta, G_\delta\right). \tag{13}$$

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

Theorem 3. *For any level α_n satisfying (13), the pair $\left(F_{\delta_*}, G_{\delta_*}\right)$ is a least-favorable pair of hypotheses for the family of tests $1\left\{T_n \ge A_n\right\}$, in the sense of (12).*

3.2. Identifying the Least-Favorable Hypotheses

We now concentrate on (10).

For given $\delta \in [\delta_{\min}, \delta_{\max}]$ with $\delta_{\min} > 0$, the distribution of the r.v. V_δ is obtained through some transformation from the known distribution of V. The classical example is $V_\delta = \sqrt{\delta} V$, which is a scaling, and where $\sqrt{\delta}$ is the signal to noise ratio. 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 4. *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 A5 in the Appendix C.

In the Gaussian case, when h is the standard normal density, Proposition 4 holds, since $h_{\delta+\eta} = h_\delta * h_{\eta-\delta}$ with $h_\varepsilon(x) := (1/\sqrt{\varepsilon})\, h\left(x/\sqrt{\varepsilon}\right)$. In order to model symmetric noise, we may consider a symmetrized Gamma density as follows: Set $h_\delta(x) := (1/2)\, \gamma^+(1,\delta)(x) + (1/2)\, \gamma^-(1,\delta)(x)$, where $\gamma^+(1,\delta)$ designates the Gamma density with scale parameter 1 and shape parameter δ, and $\gamma^-(1,\delta)$ the Gamma density on $\mathbb{R}^-$ with same parameter. Hence a r.v. with density h_δ is symmetrically distributed and has variance 2δ. Clearly, $h_{\delta+\eta}(x) = h_\delta * h_\eta(x)$, which shows that Proposition 4 also holds in this case. Note that, except 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 centred Cauchy distribution may help as a description of heavy tailed symmetric noise, and keeps uni-modality through convolution; it satisfies the requirements of Proposition 4 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 below, we also consider symmetric, exponentially distributed densities (Laplace densities) or symmetric Weibull densities with a given shape parameter. The Weibull distribution also fulfills the condition in Proposition 4, being infinitely divisible (see [12]).

3.3. Numerical Performances of the Minimax Test

As frequently observed, numerical results deduced from theoretical bounds are of poor interest due to the sub-optimality of the involved inequalities. They may be sharpened on specific cases. This motivates the need for simulation. We study 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 log-normal 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 to one another in terms of their symmetrized Kullback-Leibler divergence (the 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 $\beta(\delta)$; and

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

In order to compare the performances 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 that $P\left(\left|V_{\underline{\delta}}\right| > \underline{\delta}\right) = \Phi(1) - \Phi(-1) \sim 0.65$, 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 has 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 modeling.

We present some power functions, making use of the theoretical bounds together with the corresponding ones based on simulation runs. As seen, the performances in the theoretical approach is weak. We have focused on simulation, after some comparison with the theoretical bounds.

3.3.1. 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 performances obtained through simulations under the least-favorable hypotheses.

Theoretical Power Bound

While supposedly valid for finite n, the theoretical power bound given by (A8) still assumes some sort of asymptotics, since a good approximation of the bound implies a fine discretization of Δ to compute $I(A_n) = \inf_{\eta \in \Delta_n} I_\eta(A_n)$. Thus, by condition (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 for the power of the composite test, we compute the power of the test $\mathbf{H}_0(\delta_*)$ against $\mathbf{H}_1(\delta_*)$, where δ_* defines the LFH pair $(F_{\delta_*}, G_{\delta_*})$.

Following Proposition 4, the LFH for the test defined by T_n when the noise follows a Gaussian, a Cauchy, or a symmetrized Weibull distribution is achieved for $(F_{\delta_{\max}}, G_{\delta_{\max}})$.

When the noise follows a Laplace distribution, the pair of LFH is the one that satisfies:

$$(F_{\delta_*}, G_{\delta_*}) = \arg \min_{(F_\delta, G_\delta), \delta \in \Delta_n} K(F_\delta, G_\delta). \tag{14}$$

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

Comparison of the Two Power Curves

As expected, Figures 1–3 show that the theoretical lower bound is always below 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 ($n > 250$).

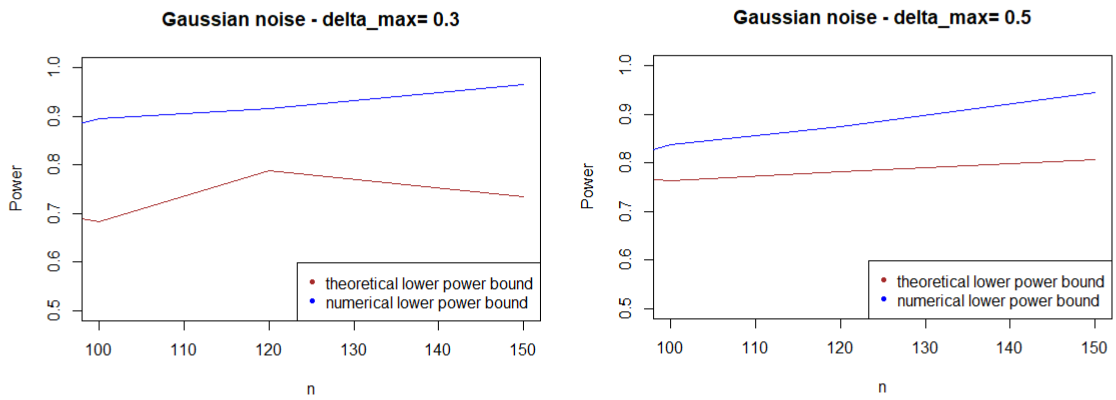

Figure 1. Theoretical and numerical power bound of the test of case A under Gaussian noise (with respect to n), for the first kind risk $\alpha = 0.05$.

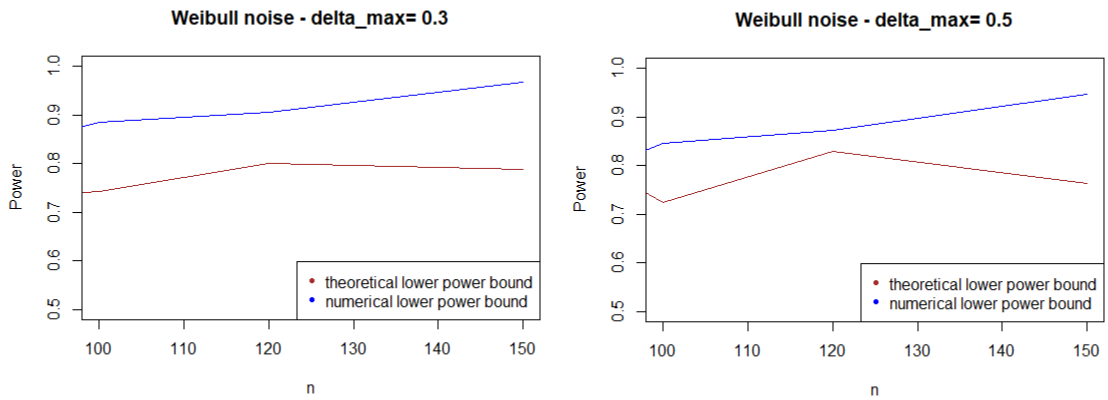

Figure 2. Theoretical and numerical power bound of the test of case A under symmetrized Weibull noise (with respect to n), for the first kind risk $\alpha = 0.05$.

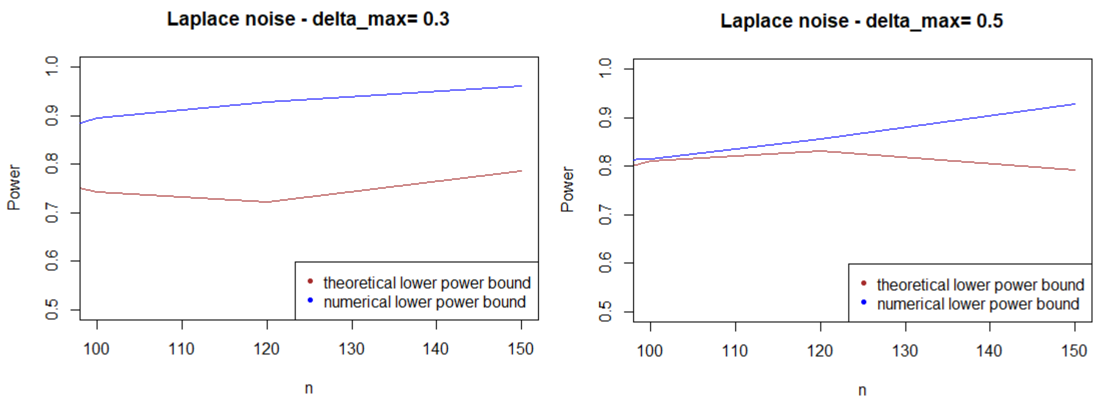

Figure 3. Theoretical and numerical power bound of the test of case A under a symmetrized Laplacian noise (with respect to n), for the first kind risk $\alpha = 0.05$.

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 (5), while the numerical performances are obtained with respect to the least-favorable hypotheses.

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

3.3.2. Case B: The Tail Thickness Problem

The calculation of the moment-generating function, appearing in the formula of $I_\eta(x)$ in (9), is numerically unstable, which renders the computation of the theoretical bound impossible. Thus, in the following sections, the performances of the test will be evaluated numerically, through Monte Carlo replications.

4. Some Alternative Statistics for Testing

4.1. A Family of Composite Tests Based on Divergence Distances

This section provides a similar treatment as above, dealing now with some extensions 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 3.3, we argue through simulation and do not develop the corresponding large deviation approach.

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

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

be 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, [13]).

Associated to this class, consider the function

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

We also consider

$$\begin{aligned} \varphi_1(x) &:= -\log x \\ \varphi_0(x) &:= \frac{1}{x} - 1, \end{aligned}$$

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 illustrates the functions φ_γ, according to γ.

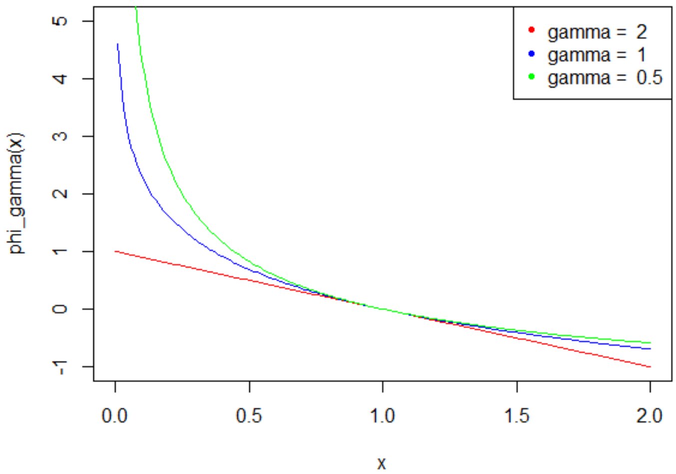

Figure 4. φ_γ for $\gamma = 0.5, 1$, and 2.

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

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

with

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

Define, accordingly, the power of the test, based on T_n^γ under the same hypotheses, by

$$s_\alpha^\gamma := \inf\left\{s : F_{\delta_*}\left(T_n^\gamma > s\right) \leq \alpha\right\}$$

and

$$1 - \beta' := G_{\delta_*}\left(T_n^\gamma > s_\alpha^\gamma\right).$$

First, δ_* defines the pair of hypotheses $(F_{\delta_*}, G_{\delta_*})$, such that the LRT with statistics T^1_{n,δ_*} has maximal power among all tests $H0(\delta_*)$ vs. $H1(\delta_*)$. Furthermore, by Theorem A1, it has minimal power on the logarithmic scale among all tests $H0(\delta)$ vs. $H1(\delta)$.

On the other hand, $(F_{\delta_*}, G_{\delta_*})$ is the LF pair for the test with statistics T^1_n among all pairs (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 considering 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

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

Consider, now, aggregated test statistics T_n^γ. We do not have at hand a similar result, as in Proposition 2. We, thus, consider the behavior of the test $H0(\delta_*)$ vs. $H1(\delta_*)$, although $(F_{\delta_*}, G_{\delta_*})$ 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

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

We will see that it may happen that Δ_n^γ improves over Δ^1_n. 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 hereafter, NP corresponds to the LRT defined between the LFH's $(F_{\delta_*}, G_{\delta_*})$, KL to the test with statistics T_n^1 (hence, as presented Section 2), HELL corresponds to $T_n^{1/2}$, which is associated to the Hellinger power divergence, and G = 2 corresponds to $\gamma = 2$.

4.2. A Practical Choice for Composite Tests Based on Simulation

We consider the same cases A and B, as described in Section 3.3.

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 3.3 as the LFH for the test T_n^1. In both cases A and B, this corresponds to $\delta^* = \delta_{\max}$.

4.2.1. 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 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.

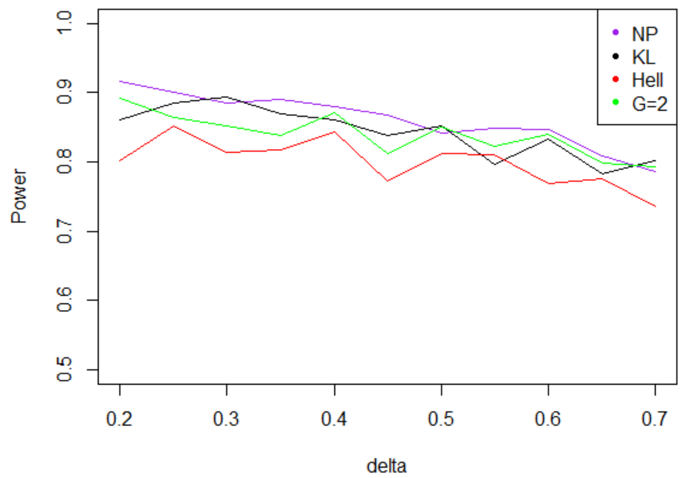

Figure 5. Power of the test of case A under Gaussian noise (with respect to $\delta_{\max}$), for the first kind risk $\alpha = 0.05$ and sample size $n = 100$.

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 alternative hypotheses are extremely well distinguished by the class of tests considered, and this remains true as δ_* increases (cf. Figure 6).

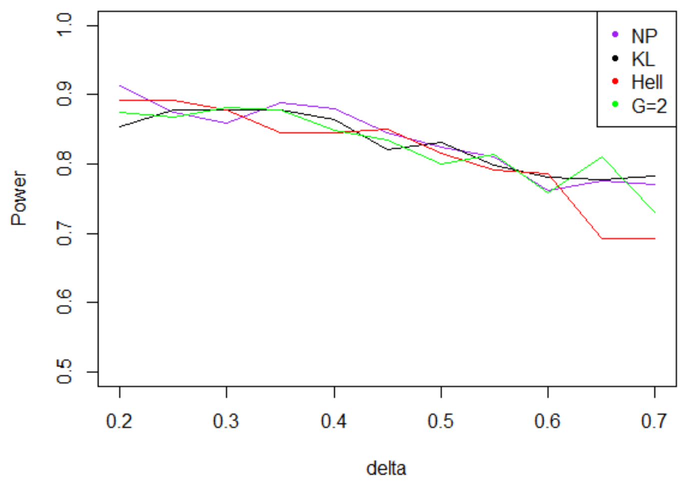

Figure 6. Power of the test of case A under Laplacian noise (with respect to $\delta_{\max}$), for the first kind risk $\alpha = 0.05$ and sample size $n = 100$.

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 7 illustrates that, as $\delta_{\max}$ increases, the power does not decrease much.

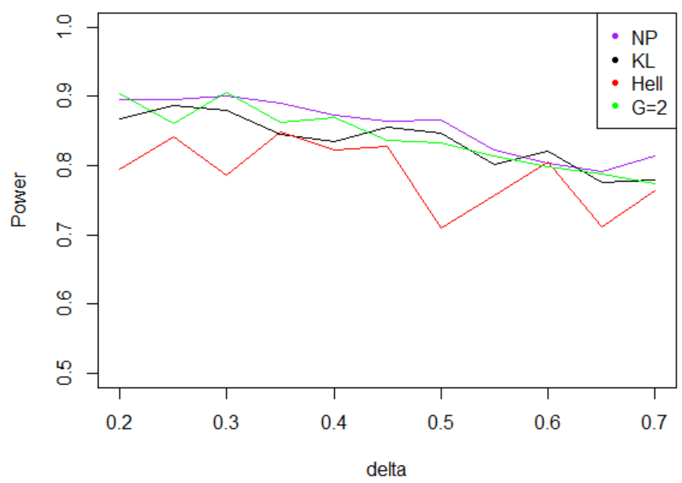

Figure 7. Power of the test of case A under symmetrized Weibull noise (with respect to $\delta_{\max}$), for the first kind risk $\alpha = 0.05$ and sample size $n = 100$.

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 5–8). 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.

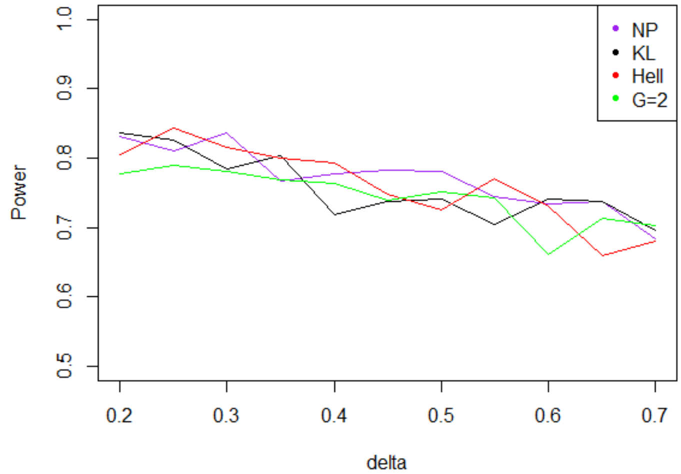

Figure 8. Power of the test of case A under noise following a Cauchy distribution (with respect to $\delta_{\max}$), for the first kind risk $\alpha = 0.05$ and sample size $n = 100$.

4.2.2. 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 4 and statistics T_n^1 provides the best power uniformly upon $\delta_{\max}$. Figure 9 shows a net decrease of the power as $\delta_{\max}$ increases (recall that the power is evaluated under the least favorable alternative $G_{\delta_{\max}}$).

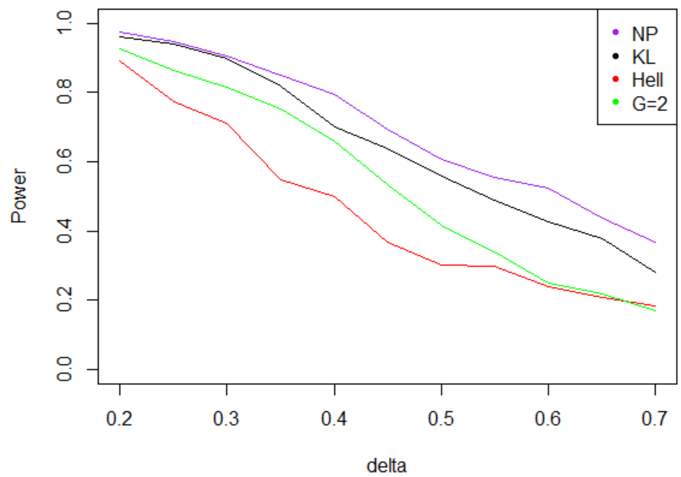

Figure 9. Power of the test of case B under Gaussian noise (with respect to $\delta_{\max}$), for the first kind risk $\alpha = 0.05$ and sample size $n = 100$. The NP curve corresponds to the optimal Neyman Pearson test under $\delta_{\max}$. The KL, Hellinger, and $G = 2$ curves stand respectively 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 $(f_\delta/g_\delta)(X)$ 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 $\phi_\gamma\left((f_\delta/g_\delta)(X)\right)$ under $G_{\delta_{\max}}$ and under $F_{\delta_{\max}}$ are quite distinct for all δ, which in turn implies that the same fact holds for the distributions of T_n^γ for large γ. Indeed, simulations presented in Figure 10 show that the maximal power of the test tends to be achieved when $\gamma = 2$.

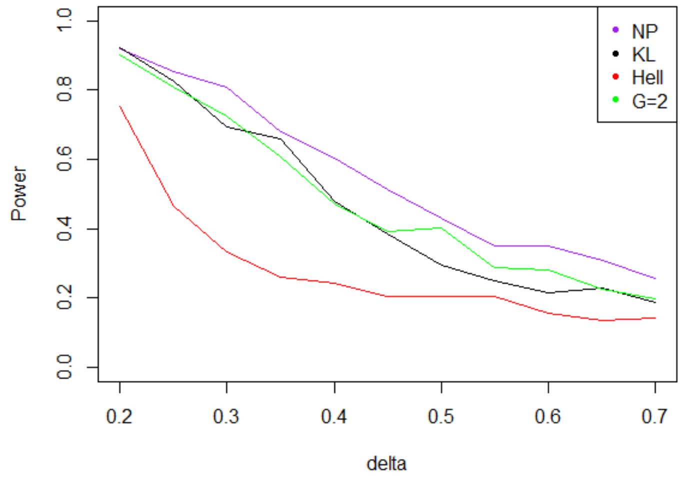

Figure 10. Power of the test of case B under Laplacian noise (with respect to $\delta_{\max}$), for the first kind risk $\alpha = 0.05$ and sample size $n = 100$.

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 11). Indeed, uniformly on δ, and on x, the ratio $(f_\delta / g_\delta)(x)$ is close to 1. Therefore, the distribution of T_n^1 is close to that of $T^1_{n,\delta_{\max}}$, which plays in favor of the KL composite test.

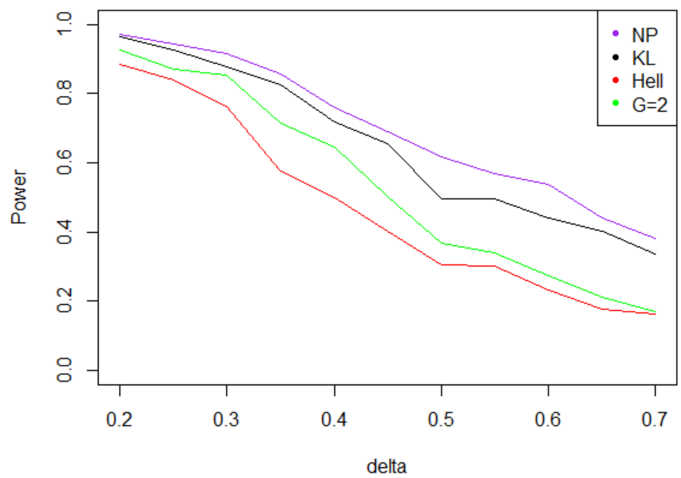

Figure 11. Power of the test of case B under symmetrized Weibull noise (with respect to $\delta_{\max}$), for the first kind risk $\alpha = 0.05$ and sample size $n = 100$.

d. Under a Cauchy distribution, similarly to case A, Figure 12 shows that T_n^γ achieves the maximal power for $\gamma = 1$ and 2, closely followed by $\gamma = 0.5$.

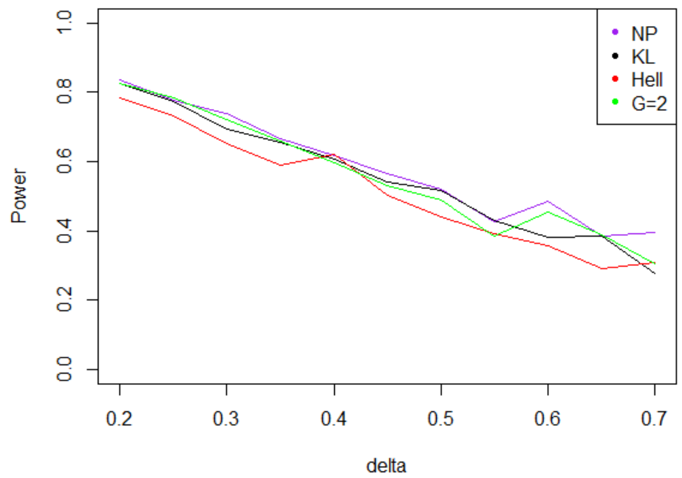

Figure 12. Power of the test of case B under a noise following a Cauchy distribution (with respect to $\delta_{\max}$), for the first kind risk $\alpha = 0.05$ and sample size $n = 100$.

5. Conclusions

We have considered a composite testing problem, where simple hypotheses in either **H0** and **H1** were paired, due to corruption in the data. The test statistics were defined through aggregation of simple likelihood ratio tests. The critical region for this test and a lower bound of its power was 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, and for a number of cases (including corruption by Gaussian, Laplacian, symmetrized Weibull, and Cauchy noise). Whatever this noise, the actual minimal power, as measured through simulation, was quite higher than obtained through analytic developments. Least-favorable hypotheses were defined in an asymptotic sense, and were proved to be the pair of simple hypotheses in **H0** and **H1** which are closest, in terms of the Kullback-Leibler divergence; this holds as a consequence of the Chernoff-Stein Lemma. We, next, considered aggregation of tests where the likelihood ratio was substituted by a divergence-based statistics. This choice extended 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.

Author Contributions: Conceptualization, M.B. and J.J.; Methodology, M.B., J.J., A.K.M. and E.M.; Software, E.M.; Validation, M.B., A.K.M. and E.M.; Formal Analysis, M.B. and J.J.; Writing Original Draft Preparation, M.B., J.J., E.M. and A.K.M.; Supervision, M.B.

Funding: The research of Jana Jurečková was supported by the Grant 18-01137S of the Czech Science Foundation. Michel Broniatowski and M. Ashok Kumar would like to thank the Science and Engineering Research Board of the government of India for the financial support for their collaboration through the VAJRA scheme.

Acknowledgments: The authors are thankful to Jan Kalina for discussion; they also thank two anonymous referees for comments which helped to improve on a former version of this paper.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Proof of Proposition 2

Appendix A.1. The Critical Region of the Test

Define

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

which satisfies

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

Note that, for all δ,

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

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 (6) holds, which means that the classes of distributions (G_δ) and (F_δ) are somehow well separated. This implies that $E_{F_\delta}\left(Z_{\delta'}\right) < 0$, for all δ and δ'.

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

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

Let

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

The function $(\delta, \delta', x) \mapsto J_{\delta,\delta'}(x)$ 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

$$A_n > E_{F_\delta}\left(Z_{\delta'}\right), \tag{A1}$$

for all δ, δ' . Then, by the Chernoff inequality

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

Since A_n should satisfy

$$\exp\left(-n J_{\delta,\delta'}\left(A_n\right)\right) \le \alpha_n,$$

with α_n bounded away from 1, A_n surely satisfies (A1) 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., [14]). Denote

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

We assume that

$$\operatorname*{ess\,sup}_\delta Z_{\delta'} = +\infty, \tag{A2}$$

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

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

It holds that

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

and, as seen previously

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

On the other hand,

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

Let

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

By (A2), the interval $\mathcal{I}$ is not void.

We now define A_n such that (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

$$A_n \geq \max_{\delta,\delta'} E_{F_\delta}\left(Z_{\delta'}\right) = \max_{(\delta,\delta') \in \Delta \times \Delta} K\left(F_\delta, F_{\delta'}\right) - K\left(F_\delta, G_{\delta'}\right), \tag{A3}$$

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 it is the infimum of a finite collection of continuous increasing functions defined on $\mathcal{I}$.

Since

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

given α_n, define

$$A_n := J^{-1}\left(-\frac{1}{n} \log \frac{\alpha_n}{p_n}\right). \tag{A4}$$

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

Appendix A.2. 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 (5),

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

and define

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

It holds that

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

and

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

which is an increasing homeomorphism from $\mathcal{M}_\eta$ 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 $\mathcal{K}_\eta := \left(E_{G_\eta}\left(W_\eta\right), \infty\right)$ 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

$$\operatorname*{ess\,sup}_\eta W_\eta = \infty \tag{A5}$$

for all η.

Assume that A_n satisfies

$$A_n \in \mathcal{K} := \bigcap_{\eta \in \Delta} \mathcal{K}_\eta \tag{A6}$$

namely

$$A_n \geq \sup_{\eta \in \Delta} E_{G_\eta}\left(W_\eta\right) = -\inf_{\eta \in \Delta} K\left(G_\eta, F_\eta\right). \tag{A7}$$

Making use of the 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 it is the infimum of a finite number of continuous increasing functions on the same interval $\mathcal{K}$, which is not void due to (A5).

We have proven that, whenever (A7) holds, a lower bound tor the test of **H0** vs. **H1** is given by

$$\begin{aligned} P_{H_1}(\mathbf{H1}) &\geq 1 - \exp\left(-nI(A_n)\right) \\ &= 1 - \exp\left(-nI\left(J^{-1}\left(-\frac{1}{n}\log\frac{\alpha_n}{p_n}\right)\right)\right). \end{aligned} \tag{A8}$$

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

Appendix A.3. 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 F_δ, $J_{\delta,\delta'}\left(E_\delta\left(Z_{\delta'}\right)\right)$ $= 0$, it follows that $J\left(\sup_{(\delta,\delta')\in\Delta\times\Delta} E_\delta\left(Z_{\delta'}\right)\right) \geq 0$. Since $E_{F_\delta}\left(Z_{\delta'}\right) = K\left(F_\delta, F_{\delta'}\right) - K\left(F_\delta, G_{\delta'}\right) < 0$, whenever 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 (A3) and (A7).

Since $K(G_\delta, F_\delta)$ is positive, it follows from (6) that

$$\sup_{(\delta,\delta')\in\Delta\times\Delta} \int \log \frac{f_{\delta'}}{g_{\delta'}} f_\delta < \sup_{\delta\in\Delta} K(G_\delta, F_\delta), \tag{A9}$$

which implies the following fact:

Let α_n be bounded away from 1. Then (A3) is fulfilled for large n, and therefore there exists A_n such that

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

Furthermore, by (A9), Condition (A7) holds, which yields the lower bound for the power of this test, as stated in (A8).

Appendix B. Proof of Theorem 3

We will repeatedly make use of the following result (Theorem 3 in [15]), which is an extension of the Chernoff-Stein Lemma (see [16]).

Theorem A1. *[Krafft and Plachky] Let x_n, such that*

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

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

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

Remark A2. *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

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

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

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

We shall later verify the validity of the conditions of Theorem A1; namely, that

$$\limsup_{n\to\infty} F_{\delta_*}\left(T_{n,\delta_*} \le A_{n,\delta_*}\right) < 1. \tag{A10}$$

Assuming (A10) we get, by Theorem A1,

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

We shall now prove that

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

Let B_{n,δ_*}, such that

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

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 A1 holds, namely

$$\limsup_{n\to\infty} F_{\delta_*}\left(T_{n,\delta_*}\leq B_{n,\delta_*}\right)<1. \tag{A11}$$

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

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

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

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

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

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

$$\limsup_{n\to\infty} F_{\delta}\left(T_{n,\delta}\leq B_{n,\delta}\right)<1, \tag{A12}$$

it follows, from Theorem A1, that

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

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 (A10)–(A12). We will only verify (A12), as 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 (13). By the law of large numbers, under G_{δ}

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

Therefore, for large n,

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

Since, under F_{δ},

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

this implies that

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

Appendix C. Proof of Proposition 4

We now prove the three lemmas that we used.

Lemma A3. *Let P, Q, and R denote three distributions with respective continuous and bounded densities p, q, and r. Then*

$$K(P*R, Q*R) \le K(P,Q). \tag{A13}$$

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

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

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 $(p*r)$ to be the convolution of p and r,

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

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

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

which is equivalent to

$$K_{\mathcal{P}}(P*R, Q*R) \le K_{\mathcal{P}}(P,Q),$$

where $K_{\mathcal{P}}$ designates the Kullback-Leibler divergence defined on $\mathcal{P}$. Refine the partition and go to the limit (Riemann Integrals), to obtain (A13) □

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 A4. *Let P and Q satisfy the hypotheses of Lemma A3 and let $(R_\delta)_{\delta>0}$ denote a family of p.m.'s on $\mathbb{R}$, and denote accordingly V_δ to be 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.

Proof. Using Lemma A3, it holds that, for positive η,

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

which proves the claim. □

Lemma A5. *Let P, Q, and R be three probability distributions with respective continuous and bounded densities p, q, and r.Assume that*

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

where all involved quantities are assumed to be finite. Then

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

Proof. We proceed as in Lemma A3, using partitions and denoting by $p_1, ..., p_K$ the induced probability of P on $\mathcal{P}$. Then,

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

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

References

1. Broniatowski, M.; Jurečková, J.; Kalina, J. Likelihood ratio testing under measurement errors. *Entropy* **2018**, *20*, 966. [CrossRef]
2. Guo, D. Relative entropy and score function: New information-estimation relationships through arbitrary additive perturbation. In Proceedings of the IEEE International Symposium on Information Theory (ISIT 2009), Seoul, Korea, 28 June–3 July 2009; pp. 814–818.
3. Huber, P.; Strassen, V. Minimax tests and the Neyman-Pearson lemma for capacities. *Ann. Stat.* **1973**, *2*, 251–273. [CrossRef]
4. Eguchi, S.; Copas, J. Interpreting Kullback-Leibler divergence with the Neyman-Pearson lemma. *J. Multivar. Anal.* **2006**, *97*, 2034–2040. [CrossRef]
5. Narayanan, K.R.; Srinivasa, A.R. On the thermodynamic temperature of a general distribution. *arXiv* **2007**, arXiv:0711.1460.
6. Bahadur, R.R. Stochastic comparison of tests. *Ann. Math. Stat.* **1960**, *31*, 276–295. [CrossRef]
7. Bahadur, R.R. *Some Limit Theorems in Statistics*; Society for Industrial and Applied Mathematics: Philadelpha, PA, USA, 1971.
8. Birgé, L. Vitesses maximales de décroissance des erreurs et tests optimaux associés. *Z. Wahrsch. Verw. Gebiete* **1981**, *55*, 261–273. [CrossRef]
9. Tusnády, G. On asymptotically optimal tests. *Ann. Stat.* **1987**, *5*, 385–393. [CrossRef]
10. Liese, F.; Vajda, I. *Convex Statistical Distances*; Teubner: Leipzig, Germany, 1987.
11. Tsallis, C. Possible generalization of BG statistics. *J. Stat. Phys.* **1987**, *52*, 479–485. [CrossRef]
12. Goldie, C. A class of infinitely divisible random variables. *Proc. Camb. Philos. Soc.* **1967**, *63*, 1141–1143. [CrossRef]
13. Basu, A.; Shioya, H.; Park, C. *Statistical Inference: The Minimum Distance Approach*; CRC Press: Boca Raton, FL, USA, 2011.
14. Barndorff-Nielsen, O. *Information and Exponential Families in Statistical Theory*; John Wiley & Sons: New York, NY, USA, 1978.
15. Krafft, O.; Plachky, D. Bounds for the power of likelihood ratio tests and their asymptotic properties. *Ann. Math. Stat.* **1970**, *41*, 1646–1654. [CrossRef]
16. Chernoff, H. Large-sample theory: Parametric case. *Ann. Math. Stat.* **1956**, *27*, 1–22. [CrossRef]