# Detectability of nonparametric signals: higher criticism versus likelihood ratio

**Marc Ditzhaus**[*],[†]

*Institute of Statistic, Ulm University, Germany, and*
*Mathematical Institute, Heinrich-Heine University Düsseldorf, Germany*
*e-mail:* marc.ditzhaus@hhu.de


**and**


**Arnold Janssen**[*],[†]

*Mathematical Institute, Heinrich-Heine University Düsseldorf, Germany*
*e-mail:* janssena@uni-duesseldorf.de

**Abstract:** We study the signal detection problem in high dimensional noise data (possibly) containing rare and weak signals. Log-likelihood ratio (LLR) tests depend on unknown parameters, but they are needed to judge the quality of detection tests since they determine the detection regions. The popular Tukey's higher criticism (HC) test was shown to achieve the same completely detectable region as the LLR test does for different (mainly) parametric models. We present a novel technique to prove this result for very general signal models, including even nonparametric $p$-value models. Moreover, we address the following questions which are still pending since the initial paper of Donoho and Jin: *What happens on the border of the completely detectable region, the so-called detection boundary? Does HC keep its optimality there?* In particular, we give a complete answer for the heteroscedastic normal mixture model. As a byproduct, we give some new insights about the LLR test's behaviour on the detection boundary by discussing, among others, Pitmans's asymptotic efficiency as an application of Le Cam's theory.

## 1. Introduction

Signal detection in huge data sets becomes more and more important in current research. The number of relevant information is often a quite small part of the data set and hidden there. In genomics, for example, the assumption is often used that the major part of the genes in patients affected by some common

---

diseases like cancer behaves like white noise and a minor part is differentially expressed but only slightly ([8, 15, 21]). Consequently, the number of signals as well as the signal strength is small. This circumstance makes it difficult to decide whether there are any signals. Other application fields are disease surveillance ([30, 34]), local anomaly detection ([35]), cosmology and astronomy ([7, 27]). In the last decade *Tukey's higher criticism* (HC) test ([37, 38, 39]) modified by Donoho and Jin [12] became quite popular for these kind of problems. The reason for HC's popularity is that the area of complete detection coincide for the HC test and the *log-likelihood ratio* (LLR) test under different specific model assumption ([2, 3, 5, 6, 12, 26]). The LLR test, which achieves the highest power among all tests, cannot be applied since it requires the knowledge of the unknown signal strength and proportion. But it serves as an important benchmark and, in particular, determines which kind of signal alternatives are completely detectable at all. That the HC test can completely separate every completely detectable alternative was also shown within sparse linear regression models and binary regression models ([1, 20, 33]). To overcome the problem of an unknown noise distribution Delaigle et al. [9] used a bootstrap version of HC. Moreover, Jager and Wellner [23] suggested a whole family of different tests sharing HC's complete detectability behaviour for the heterogeneous normal mixture model. Recently, Ditzhaus [11] verified that the same is true beyond this specific model. A lot of related literature about HC's possibilities, even beyond signal detection, can be found in the survey paper of Donoho and Jin [13]. For instance, Hall et al. [18] applied HC for classification.

There are (only) a few results concerning the asymptotic power behaviour of the LLR test on the detection boundary, which separates the area of complete detection and the area of no possible detection, see Cai et al. [5] and Ingster [19] for the heteroscedastic and heterogeneous normal mixture models. Since Donoho and Jin [12] the following questions is pending: *How does HC perform on the detection boundary? Does it keep its optimality?* Donoho and Jin [12] specially pointed out: "*Just at the critical point where $r = \rho_*$ (1 + o(1)), our result says nothing; this would be an interesting (but very challenging) area for future work.*"

Our paper's purpose is twofold. First, we want to fill the theoretical gap concerning the tests' power behaviour on the detection boundary and give an answer to the question mentioned before. We quantify the asymptotic power of the LLR test by giving the LLR statistic's limit distribution. On the detection boundary the LLR test has nontrivial asymptotic power, whereas the HC test does not. Consequently, HC is not overall powerful. However, our message is not to scrap the idea of HC. Its power behaviour is still optimal beyond the detection boundary for a long list of models. The second purpose of our paper is to add a $p$-value model with signals coming from a nonparametric alternative to this list of models.

The paper is organized as follows. In Section 1.1 we introduce the general model and the detection testing problem. For the readers' convenience we add to our paper the illustrative Section 1.2. There, all main results are presented by discussing our prime example. The asymptotic results about the benchmark

LLR tests appear in Section 2. The following Section 3 is devoted to the HC statistic and introduce an "HC complete detection" as well as a "trivial HC power" Theorem. Whereas the previous two sections develop the general machinery, Section 4 contains the applications. We discuss a generalizations of the illustrative results from Section 1.2 as well as the heteroscedastic normal mixture model. Although the latter was already studied in great detail we can give some new insights for it. Further examples can be found in Ditzhaus [10, 11]. All proofs are relegated to Appendix B.

### 1.1. The model

Let $\{k_n : n \in \mathbb{N}\} \subset \mathbb{N}$, where $k_n \to \infty$ represents the number of observations. Throughout this paper, if not stated otherwise all limits are meant as $n \to \infty$. Let the following three mutually independent triangular arrays consisting of rowwise independent random variables are given, where values in different spaces are allowed:

- $(Z_{n,i})_{i \leq k_n}$ representing the noisy background, where the distribution $P_{n,i}$ of $Z_{n,i}$ is assumed to be known. In the applications we often assume that $P_{n,i} = P_0$ depends neither on $i$ nor on $n$, and $P_0$ may stand for a distribution of $p$-values under the null.
- $(X_{n,i})_{i \leq k_n}$ representing the signals, where the signal distribution $\mu_{n,i}$ of $X_{n,i}$ is typically unknown.
- $(B_{n,i})_{i \leq k_n}$ representing the appearance of a signal, where $B_{n,i}$ is Bernoulli distributed with typically unknown success probability $0 \leq \varepsilon_{n,i} \leq 1$.

Instead of these random variables we observe

$$Y_{n,i} = \begin{cases} X_{n,i} & \text{if } B_{n,i} = 1 \\ Z_{n,i} & \text{if } B_{n,i} = 0 \end{cases}$$

for all $1 \leq i \leq k_n$. The vector $(Y_{n,1}, \ldots, Y_{n,k_n})$ represents the noise data containing a random amount $\sum_{i=1}^{k_n} B_{n,i}$ of signals. It is easy to check that the distribution $Q_{n,i}$ of $Y_{n,i}$ is given by

$$Q_{n,i} = (1 - \varepsilon_{n,i})P_{n,i} + \varepsilon_{n,i}\mu_{n,i} = P_{n,i} + \varepsilon_{n,i}(\mu_{n,i} - P_{n,i}). \tag{1.1}$$

The additional index $i$, for instance, $\mu_{n,i}$ instead of $\mu_n$, allows to treat two-sample or more general kinds of signal alternatives. We are interested whether there are any signals in the noise data, i.e., whether $B_{n,i} = 1$ for at least one $i = 1, \ldots, k_n$. To be more specific, we study the testing problem

$$\mathcal{H}_{0,n}: \ \varepsilon_{n,i} = 0 \text{ for all i} \qquad \text{versus} \qquad \mathcal{H}_{1,n}: \ \varepsilon_{n,i} > 0 \text{ for at least one } i, \tag{1.2}$$

where we observe pure noise $(Y_{n,1}, \ldots, Y_{n,k_n}) = (Z_{n,1}, \ldots, Z_{n,k_n})$ under the null. We are especially interested in the case of rare signals in the sense that

$$\max_{1 \leq i \leq k_n} \varepsilon_{n,i} \to 0. \tag{1.3}$$

In this setting, we distinguish between the *sparse* ($\sum_{i=1}^{k_n} \varepsilon_{n,i}^2 \to 0$), the *classical* ($\lim_{n\to\infty} \sum_{i=1}^{k_n} \varepsilon_{n,i}^2 \in (0,\infty)$) and the *dense* signal case ($\sum_{i=1}^{k_n} \varepsilon_{n,i}^2 \to \infty$). In the *rowwise identical setting*, where all quantities, $\varepsilon_{n,i} = \varepsilon_n$ etc., are independent of $i$ the parametrization $\varepsilon_n = n^{-\beta}$ for $\beta \in (0,1)$ is standard. Then $\beta < 1/2$ and $\beta > 1/2$ correspond to the dense and sparse case, respectively. We denote $\varepsilon_n = n^{-1/2}$, or in other words $\beta = 1/2$, as the classical case since it is the usual rate of convergence when discussing contiguous alternatives. In the classical case nontrivial power results can be obtained by choosing a signal distribution $\mu_n = \mu \neq P_0 = P_n$, whereas in the sparse case, where less signals are present, only asymptotically singular $\mu_n$ and $P_n$ lead to nontrivial power results. At the same time, asymptotically merging $\mu_n$ and $P_n$ lead to nontrivial results in the dense case, where, relatively, a lot of signals occur. While our applications focus on the most interesting sparse case, the technical machinery applies for all three cases. A huge class of examples for the dense case is examined by Ditzhaus [11].

Another typical assumption in the signal detection literature is

$$\mu_{n,i} \ll P_{n,i} \text{ for all } 1 \leq i \leq k_n, \tag{1.4}$$

which we also suppose throughout this paper. In Section 2.4 we discuss what happens if the assumption of absolute continuity is violated. Following the ideas of Cai and Wu [6] we explain that every model can be reduced to a model such that (1.4) is fulfilled.

**Convention and Notation:** Observe that

$$\frac{\mathrm{d}Q_{n,i}}{\mathrm{d}P_{n,i}} = 1 + \varepsilon_{n,i}\Big(\frac{\mathrm{d}\mu_{n,i}}{\mathrm{d}P_{n,i}} - 1\Big).$$

The distributions $P_{n,i}, \mu_{n,i}, Q_{n,i}$ and the densities $\frac{\mathrm{d}Q_{n,i}}{\mathrm{d}P_{n,i}} \circ pr_i$ shall lie on the same product space, where the projections $pr_i$ on the $i$th coordinate are suppressed throughout the paper to improve the readability. Moreover, we introduce the product measures

$$Q_{(n)} = \bigotimes_{i=1}^{k_n} Q_{n,i} \text{ and } P_{(n)} = \bigotimes_{i=1}^{k_n} P_{n,i}.$$

### 1.2. Illustration of the results and the main contents

In this illustrative section we give an overview of our results by studying a special nonparametric $p$-values model. For simplicity we set $k_n = n$ and restrict to the rowwise identical case, i.e., $\mu_{n,i} = \mu_n$ etc. Testing results are often presented in terms of $p$-values since they allow a comparison of different data types on the same platform. In our context, a quantile transformation like $p_{n,i} = P_n((Y_{n,i}, \infty))$ or $p_{n,i} = P_n((-\infty, Y_{n,i}])$ may be used to get $p$-values. As long as the noise distribution $P_n$ is continuous the $p$-values $p_{n,1}, \ldots, p_{n,n}$ follow under the null a uniform distribution $P_0$, say, on the unit interval $(0,1)$. To benefit from this universal platform without too many or too specific model

assumptions, we consider in this illustrative section from the beginning that $p$-values are present and, in particular, $P_n = P_0$.

Typically, small $p$-values indicates that the alternative is true, or in our case that signals are present. Respecting this we suggest signal distributions $\mu_n$ with a shrinking support $[0, \kappa_n]$, where

$$\kappa_n = n^{-r} \text{ and } \varepsilon_n = n^{-\beta} \tag{1.5}$$

for some $r > 0$. Clearly, $\mu_n$ and $P_0$ are asymptotically singular. Hence, this setting is an example for the sparse case and we restrict our considerations to $\beta \in (1/2, 1)$. In order to obtain such $\mu_n$ the interval $(0, \kappa_n)$ is blown up to $(0, 1)$ and a nonparametric shape function $h$ is used. Let $h : (0, 1) \to (0, \infty)$ be a Lebesgue probability density, i.e., we have $\int h \, dP_0 = 1$, with $\int h^2 \, dP_0 \in (0, \infty)$ and define the signal distribution by its rescaled Lebesgue density

$$\frac{d\mu_n}{dP_0}(x) = \frac{1}{\kappa_n} h\Big(\frac{x}{\kappa_n}\Big) \mathbf{1}\{x \le \kappa_n\}, \ x \in (0, 1). \tag{1.6}$$

Since it could be too restrictive in practice to consider only measures with a shrinking support, in Section 4.1 we add a "small" perturbation to the densities. To sum up, we have a nonparametric testing problem which can be expressed heuristically as

$$\mathcal{H}_{0,n} : \varepsilon_n = 0 \text{ versus } \mathcal{H}_{1,n} : \varepsilon_n > 0, h \in L^2(P_0) \text{ with } h \ge 0, \int h \, dP_0 = 1.$$

The alternative $\mathcal{H}_{1,n}$ is composite since, for example in this specific setting, the signal proportion $\varepsilon_n$ and the signal shape function $h$ are unknown. When we talk about the LLR test below then the LLR test corresponding to the true but unknown $\varepsilon_{n,\text{true}}$ and $h_{\text{true}}$ is meant. This test is optimal for testing $\mathcal{H}_{0,n}$ against the simple alternative $\widetilde{\mathcal{H}}_{1,n} : \{\varepsilon_n = \varepsilon_{n,\text{true}}, h = h_{\text{true}}\}$. In contrast to that, the HC test is designed for the composite alternative while being asymptotically as good as the specific LLR test based on the unknown $\varepsilon_{n,\text{true}}$ and $h_{\text{true}}$. The heuristic phrase "being asymptotically as good as" is explained below in more detail.

The following list of the seven problems I–VII and their solutions regarding our prime example gives the reader a first impression and overview of the results which can be obtained be the general machinery developed in Sections 3 and 2.

   I. *Determination of the detection boundary*: Since the paper of Donoho and Jin [12] the term *detection boundary* is of great interest for the detection problem. This boundary splits the $r$-$\beta$ parametrisation plane into the *completely detectable* and the *undetectable* area. For each pair $(r, \beta)$ from the *completely detectable* area the LLR test, the optimal test, can completely separate the null and the alternative asymptotically. This means that there is a sequence $(\varphi_n)_{n \in \mathbb{N}}$ of LLR tests with nominal levels $E_{P_{(n)}}(\varphi_n) = \alpha_n$ such that $\alpha_n \to 0$ and the power $E_{Q_{(n)}}(\varphi_n)$ under the alternative tends to 1. For each $(r, \beta)$ from the undetectable area the null $\mathcal{H}_{0,n}$ and the
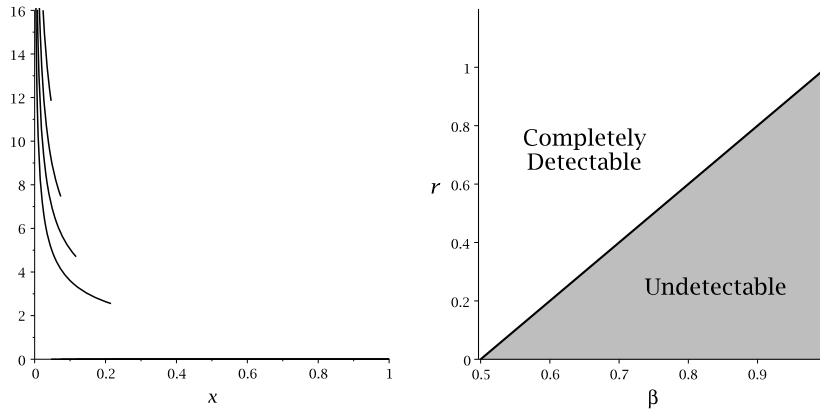
FIG 1. *Left: Plot of $x \mapsto \mathrm{d}\mu_n/\mathrm{d}P_0(x)$ for $h(x) = (1-a)x^{-a}$, $a = 9/20$, $r = 2/3$ and $n \in \{10, 25, 50, 100\}$, see* (1.6). *Right: The nonparametric detection boundary is plotted. Above the boundary is the completely detectable area and underneath is the undetectable area. The limits of the LLR statistic are Gaussian on the solid line under the null as well as under the alternative, and they are real-valued but non-Gaussian on the end of the line (solid circle). The limit under the alternative is equal to $\infty$ with a positive probability.*

alternative $\mathcal{H}_{1,n}$ are asymptotically indistinguishable, i.e. the sum of error probabilities tends to 1 for each possible sequence of tests. Hence, no test yields asymptotically better results than a constant test $\varphi \equiv \alpha \in (0,1)$. For the illustrative model we have a *nonparametric* detection boundary which is independent of the shape function $h$ and given by

$$\rho(\beta) = 2\beta - 1 \text{ for } \beta \in \left(\frac{1}{2}, 1\right].\tag{1.7}$$

The area where $r > \rho(\beta)$ ($r < \rho(\beta)$, resp.) corresponds to the completely detectable area (undetectable area, respectively), see Figure 1.

II. *Gaussian limits on the detection boundary?* For some parametric models the limit distribution of the log-likelihood ratio test statistic $T_n$, see below, was determined, e.g. for the heteroscedastic and heterogeneous normal mixture model, see Cai et al. [5] and Ingster [19]. For our model with $1/2 < \beta < 1$ and $r = \rho(\beta)$ we have

$$T_n = \log \frac{\mathrm{d}Q_{(n)}}{\mathrm{d}P_{(n)}} \xrightarrow{\mathrm{d}} \begin{cases} \xi_1 \sim N(-\frac{\sigma^2(h)}{2}, \sigma^2(h)) & \text{under } \mathcal{H}_{0,n}, \\ \xi_2 \sim N(\frac{\sigma^2(h)}{2}, \sigma^2(h)) & \text{under } \mathcal{H}_{1,n}, \end{cases}$$

where $\sigma^2(h) = \int_0^1 h^2 \, \mathrm{d}P_0$. Observe that the limits only depend on the second moment of $h$ and not on its specific structure.

III. *What happens if we choose the wrong $h$ or $\beta$ for the LLR statistic on the boundary?* Let $(h_1, \beta_1)$ and $(h_2, \beta_2)$ represent two specific models of the illustrative example on the detection boundary, i.e. $\beta_i \in (1/2, 1)$ and $r_i = \rho(\beta_i)$ for $i = 1, 2$. Using Le Cam's LAN theory we can determine

the asymptotic power of the LLR test $\varphi_{n,\beta_2,h_2,\alpha}$ of the model $(h_2,\beta_2)$ of nominal level $\alpha \in (0,1)$ if $(h_1,\beta_1)$ is the true, underlying model:

$$E_{\mathcal{H}_{1,n}(h_1,\beta_1)}(\varphi_{n,\beta_2,h_2,\alpha}) \to \Phi\Big(u_\alpha + \sqrt{\sigma^2(h_1)\mathrm{ARE}}\Big),$$

$$\text{where } \mathrm{ARE} = \frac{(\int_0^1 h_1 h_2 \, dP_0)^2}{\sigma^2(h_1)\sigma^2(h_2)}\mathbf{1}\{\beta_1 = \beta_2\}$$

is Pitman's asymptotic relative efficiency, see Hájek et al. [17], $\Phi$ denotes the distribution function of a standard normal distribution and $u_\alpha$ is the corresponding $\alpha$-quantile, i.e. $\Phi(u_\alpha) = \alpha$. This formula quantifies the loss of power by choosing the wrong $\beta$ or $h$. In particular, the LLR test $\varphi_{n,\beta_2,h_2,\alpha}$ cannot separate the null and the alternative asymptotically, i.e ARE= 0, if the supports of $h_1$ and $h_2$ are disjunct, or if $\beta_1$ and $\beta_2$ are unequal.

IV. *Beyond Gaussian limits on the detection boundary.* Non-Gaussian limits of $T_n$ may occur ([5, 19]). Here, these limits can be observed if the second moment assumption on $h$ is violated, i.e., we have $\int_0^1 h^2 \, dP_0 = \infty$. In this case the limits are infinitely divisible distributed with nontrivial Lévy measure. These Lévy measures depend heavily on the special structure of $h$, details can be found in Theorem 4.5.

V. *Extension of the detection boundary:* We discuss also the case $\beta = 1$, whereas a lot of former research was focused (only) on $\beta < 1$. The case $\beta \geq 1$ was of minor interest reason since the probability that at least one signal is present equals $1 - (1 - \varepsilon_n)^n$, which tends to $1 - e^{-1}$ and 0 if $\beta = 1$ and $\beta > 1$, respectively. In particular, the pair $(\beta, r)$ with $\beta > 1$ and $r > 0$ always belongs to the undetectable area. Hence, $\beta > 1$ do not need to be studied further. But $\beta = 1$ should be taken into account since a new class of limits can be observed. To be more specific, for $\beta = 1$ and $r > 1$ we have

$$T_n \xrightarrow{\mathrm{d}} \begin{cases} \xi_1 \equiv -1 & \text{under } \mathcal{H}_{0,n}, \\ \xi_2 \sim e^{-1}\epsilon_{-1} + (1 - e^{-1})\epsilon_\infty & \text{under } \mathcal{H}_{1,n}, \end{cases}$$

where $\epsilon_a$ denotes the Dirac measure centered in $a \in [-\infty, \infty]$, i.e. $\epsilon_a(A) = \mathbf{1}\{x \in A\}$. As far as we know such nontrivial limits, where $\xi_2$ equals $\infty$ with a positive probability, were not observed for the detection issue until now.

VI. *Optimality of HC.* As already known for different mainly parametric models, we can show also for the illustrative nonparametric $p$-values model that the completely detectable regions of the LLR and the HC test coincide. By this we give a further reason why HC is a good candidate for the signal detection problem.

VII. *No power of HC on the boundary.* We show that on the detection boundary, i.e. $\beta \in (1/2, 1)$ and $r = \rho(\beta)$, the HC test cannot distinguish between the null and the alternative alternative, whereas the LLR test has nontrivial power, compare to II.

Among others, we apply our results to the model (1.6) in a more general form, e.g. $h_{n,i}$, $\kappa_{n,i}$ and $\varepsilon_{n,i}$ may depend on $i$ and $n$. We want to point out that these kind of alternatives were already studied in the context of goodness-of-fit testing by Khmaladze [28]. He used the name *spike chimeric alternatives*. Finally, we want to mention that our general model and the upcoming results also include

VIII. discrete models as the Poisson model of Arias-Castro and Wang [2] (Note that only the results concerning LLR tests apply for discrete models).

IX. the *sparse* $(\sum_{i=1}^{k_n} \varepsilon_{n,i}^2 \to 0)$, the *classical* $(\lim_{n\to\infty} \sum_{i=1}^{k_n} \varepsilon_{n,i}^2 \in (0,\infty))$ and the *dense* case $(\sum_{i=1}^{k_n} \varepsilon_{n,i}^2 \to \infty)$.

## 2. Asymptotic power behaviour of LLR tests

In this section we discuss the asymptotic power behaviour of LLR tests. These tests depend on the unknown signals and, hence, they are not applicable. But they serve as an import benchmark and all new suggested tests should be compare with the optimal LLR tests.

It is well known that at least for a subsequence $T_n$ converges in distribution to a random variable with values on the extended real line $[-\infty, \infty]$ under the null as well as under the alternative, see Lemma 60.6 of Strasser [36]. That is why we can assume without loss of generality that

$$T_n = \sum_{i=1}^{k_n} \log \frac{dQ_{n,i}}{dP_{n,i}}(Y_{n,i}) \xrightarrow{d} \begin{cases} \xi_1 \text{ under } P_{(n)} \text{ (null)}, \\ \xi_2 \text{ under } Q_{(n)} \text{ (alternative)}, \end{cases} \tag{2.1}$$

where $\xi_1$ and $\xi_2$ are random variables on $[-\infty, \infty]$. Regarding the phase diagram on the right side in Figure 1 we are interested in the following three different regions/cases:

(i) (Completely detectable) The LLR test $\varphi_n = \mathbf{1}\{T_n > c_n\}$ with appropriate critical values $c_n \in \mathbb{R}$ can completely separate the null and the alternative asymptotically, i.e. the sum of error probabilities $E_{\mathcal{H}_{0,n}}(\varphi_n) + E_{\mathcal{H}_{1,n}}(1-\varphi_n)$ tends to 0. We will see that this corresponds to $\xi_1 \equiv -\infty$ and $\xi_2 \equiv \infty$.

(ii) (Undetectable) No test sequence $(\psi_n)_{n\in\mathbb{N}}$ can distinguish between the null and the alternative asymptotically, i.e we always have $E_{\mathcal{H}_{0,n}}(\varphi_n) + E_{\mathcal{H}_{1,n}}(1 - \varphi_n) \to 1$. This case corresponds to $\xi_1 \equiv 0 \equiv \xi_2$.

(iii) (Detectable) The LLR test $\varphi_n = \mathbf{1}\{T_n > c_n\}$ with appropriate critical values $c_n \in \mathbb{R}$ can separate the null and the alternative asymptotically but not completely, i.e. $E_{\mathcal{H}_{0,n}}(\varphi_n) + E_{\mathcal{H}_{1,n}}(1 - \varphi_n) \to c \in (0, 1)$.

In the following we denote the completely detectable and the undetectable case as the trivial cases since the limits of $T_n$ are degenerated. We start by discussing these and we present a useful tool to verify these trivial cases/limits of $T_n$. After that we will see that the same tools can be used to determine the nontrivial limits in the detectable case. In the last two subsections we consider the asymptotic relative efficiency, compare to (III) from Section 1.2, and explain what to do when the condition (1.4) is violated.

### 2.1. Trivial limits

In the proofs we work with different distances for probability measure, among others the Hellinger distance and the variational distance. Using theses distances we can classify the different detection regions. We refer the reader to the Appendix B, for further details. Here, we only present our new tool. Let us introduce for all $x > 0$ the following two sums

$$I_{n,1,x} = \sum_{i=1}^{k_n} \varepsilon_{n,i} \mu_{n,i} \Big( \varepsilon_{n,i} \frac{d\mu_{n,i}}{dP_{n,i}} > x \Big) \tag{2.2}$$

$$\text{and } I_{n,2,x} = \sum_{i=1}^{k_n} \varepsilon_{n,i}^2 E_{P_{n,i}} \Big( \Big( \frac{d\mu_{n,i}}{dP_{n,i}} \Big)^2 \mathbf{1}\Big\{ \varepsilon_{n,i} \frac{d\mu_{n,i}}{dP_{n,i}} \le x \Big\} - 1 \Big). \tag{2.3}$$

**Theorem 2.1.** *Let $\tau > 0$ be fixed.*

(a) *The completely detectable case is present if and only if $I_{n,1,\tau}$ or $I_{n,2,\tau}$ tends to $\infty$.*

(b) *We are in the undetectable case if and only if $I_{n,1,\tau}$ as well as $I_{n,2,\tau}$ tends to $0$.*

### 2.2. Nontrivial limits

It turns out that only a special class of distributions $\nu_1$ and $\nu_2$, say, of $\xi_1$ and $\xi_2$ may occur. The results fit in the more general framework of statistical experiments: all nontrivial weak accumulation points with respect to the weak topology of statistical experiments are infinitely divisible statistical experiments in the sense of Le Cam [31], see Le Cam and Yang [32] and [24]. In the following we explain what this means in our situation. Classical infinitely divisible distributions on $(\mathbb{R}, \mathcal{B})$ play a key role for our setting. That is why we want to recall that the characteristic function $\varphi$ of an infinitely divisible distribution on $(\mathbb{R}, \mathcal{B})$ is given by the Lévy-Khintchine formula

$$\varphi(t) = \exp\Big[ i\gamma t - \frac{\sigma^2 t^2}{2} + \int_{\mathbb{R}\setminus\{0\}} \Big( \exp(itx) - 1 - \frac{itx}{1+x^2} \Big) \, d\eta(x) \Big], \, t \in \mathbb{R},$$

where $\gamma \in \mathbb{R}$, $\sigma^2 \in [0, \infty)$ and $\eta$ is a Lévy measure, i.e. $\eta$ is a measure on $\mathbb{R} \setminus \{0\}$ with $\int \min(x^2, 1) \, d\eta < \infty$. The triple $(\gamma, \sigma^2, \eta)$ is called the Lévy-Khintchine triple and is unique. See Gnedenko and Kolmogorov [16] for more details about infinitely divisible distributions. The following theorem gives us a characterisation of all possible limits of $T_n$.

**Theorem 2.2.** (a) *Either $\xi_1$ is real-valued or $\xi_1 \equiv -\infty$ with probability one. In case of the latter $\xi_2 \equiv \infty$ with probability one.*

(b) *Suppose $\xi_1$ is real-valued. Then $a = P(\xi_2 \in \mathbb{R}) > 0$ and we can rewrite $\nu_2 = a\rho + (1-a)\epsilon_\infty$, where $\rho(A) = a^{-1}\nu_2(A \cap \mathbb{R})$ for all $A \in \mathcal{B}([-\infty, \infty])$. Moreover, $\nu_1$ and $\rho = a^{-1}\nu_{2|\mathbb{R}}$ are infinitely divisible distributions on*

$(\mathbb{R}, \mathcal{B})$. *Let* $(\gamma_1, \sigma_1^2, \eta_1)$ *and* $(\gamma_2, \sigma_2^2, \eta_2)$ *be the Lévy-Khintchine triplets of* $\nu_1$ *and* $\rho = a^{-1}\nu_{2|\mathbb{R}}$. *Then we have:*

(i) *The Lévy measures* $\eta_1$ *and* $\eta_2$ *are concentrated on* $(0, \infty)$, *i.e.* $\eta_j(-\infty, 0) = 0$. *and* $\int_{(0,\infty)} e^x \, d\eta_1(x) < \infty$. *Moreover,* $\frac{d\eta_2}{d\eta_1}(x) = e^x$ *for all* $x > 0$.

(ii) *The variances of the Gaussian parts of* $\xi_1$ *and* $\xi_2$ *coincide, i.e.* $\sigma_1^2 = \sigma_2^2$.

(iii) *The drift parameters* $\gamma_1$ *and* $\gamma_2$ *fulfill the formulas:*

$$\log(a) = \gamma_1 + \frac{\sigma_1^2}{2} - \int_{(0,\infty)} \left(1 - e^x + \frac{x}{1+x^2}\right) \, d\eta_1(x), \qquad (2.4)$$

$$\gamma_2 \quad = \gamma_1 + \sigma_1^2 + \int_{(0,\infty)} (e^x - 1)\frac{x}{1+x^2} \, d\eta_1(x). \qquad (2.5)$$

*Remark* 2.3. If $\xi_1$ is real-valued then by Le Cam's first Lemma the null (product) measure $P_{(n)}$ is contiguous with respect to the alternative (product) measure $Q_{(n)}$, i.e. $Q_{(n)}(A_n) \to 0$ implies $P_{(n)}(A_n) \to 0$. If additionally $\xi_2$ is real-valued then $P_{(n)}$ and $Q_{(n)}$ are mutually contiguous, i.e. $Q_{(n)}(A_n) \to 0$ if and only if $P_{(n)}(A_n) \to 0$. Observe that under mutually contiguity a random variable is asymptotically constant under the null $P_{(n)}$ if and only if this is the case under the alternative $Q_{(n)}$.

According to Theorem 2.2(b) the Lévy-Khintchine triplets of $\nu$ and $\rho = a^{-1}\nu_{2|\mathbb{R}}$ are closely related to each other. This was already observed in the context of statistical experiments by Janssen et al. [24].

Now, we know the class of all possible limits and, hence, the questions arises naturally how to determine the distribution of $\xi_1$ and $\xi_2$ for a given setting. To answer this question we first observe that by Theorem 2.2(i) the Lévy measures $\eta_1$ and $\eta_2$ are uniquely determined by their difference $M = \eta_2 - \eta_1$. Combining this, Theorem 2.2(ii) and Theorem 2.2(iii) yields that $M$, $\sigma_1^2$ and $a = \nu_2(\mathbb{R})$ serve to understand the distribution of $\xi_1$ and $\xi_2$ completely. We will see that these three are determined by the limits of the sums given by (2.2) and (2.3). To give a first impression why this is the case we explain briefly the impact of $I_{n,1,x}$. Since the summands of $T_n$ fulfill the so-called condition of infinite smallness, i.e. a finite number of summands has no influence of the sum's convergence behaviour, well-known limit theorems to infinitely divisible distributed random variable can be applied, see, for instance, Gnedenko and Kolmogorov [16]. In the case of real-valued $\xi_1$ we obtain from these theorems

$$\sum_{i=1}^{k_n} P_{n,i}\left(\varepsilon_{n,i}\frac{d\mu_{n,i}}{dP_{n,i}} > e^x - 1 + \varepsilon_{n,i}\right) \to \eta_1(x, \infty) \qquad (2.6)$$

for all x from a dense subset of $(0, \infty)$. If additionally $\xi_2$ is real valued then the same holds for $\eta_2$ when we replace $P_{n,i}$ by $Q_{n,i}$. Combining these and (1.3) shows that $I_{n,1,e^x-1}$ tends to $M(x, \infty) = (\eta_2 - \eta_1)(x, \infty)$ for all x coming from a dense

subset of $(0, \infty)$ if both, $\xi_1$ and $\xi_2$, are real-valued. In the case of $a = \nu_2(\mathbb{R}) = P(\xi_2 \in \mathbb{R}) < 1$ a similar convergence can be observed, namely $I_{n,1,e^x-1}$ tends to $(\eta_2 - \eta_1)(x, \infty) + M(\infty)$, where the mass $M(\infty)$ in the point $\infty$ characterizes $a$ uniquely.

**Theorem 2.4.** *Let $I_{n,1,x}$ and $I_{n,2,x}$, $x > 0$, be defined as in* (2.2) *and* (2.3). *$\xi_1$ is real-valued if and only if the following* (a) *and* (b) *hold:*

(a) *There is a dense subset $\mathcal{D}$ of $(0, \infty)$ and a measure $M$ on $((0, \infty], \mathcal{B}(0, \infty])$ such that for all $x \in \mathcal{D}$*

$$\lim_{n \to \infty} I_{n,1,e^x-1} = M(x, \infty].$$

(b) *For some $\sigma^2 \in [0, \infty)$ we have*

$$\lim_{x \searrow 0} \begin{array}{c} \limsup \\ \liminf \\ n \to \infty \end{array} I_{n,2,x} = \sigma^2,$$

*i.e. this equation holds for $\limsup_{n \to \infty}$ and $\liminf_{n \to \infty}$ simultaneously.*

*If* (a) *and* (b) *hold then using the notation from Theorem* 2.2(b) *we obtain $\nu_2(\mathbb{R}) = \exp(-M(\{\infty\}))$, $\sigma^2 = \sigma_1^2 = \sigma_2^2$ and $\eta_2 - \eta_1 = M_{|(0,\infty)}$.*

*Remark* 2.5. (i) From Theorem 2.2(i) we get for all $x > 0$

$$\frac{\mathrm{d}\eta_1}{\mathrm{d}M}(x) = \frac{1}{\exp(x) - 1} \text{ and } \frac{\mathrm{d}\eta_2}{\mathrm{d}M}(x) = \frac{\exp(x)}{\exp(x) - 1}. \tag{2.7}$$

(ii) Consider the rowwise identical case with a noise distribution independent on $n$, i.e. $P_{n,i} = P_0$, $\mu_{n,i} = \mu_n$ and $\varepsilon_{n,i} = \varepsilon_n$. Thus, $Y_{n,1}, \ldots, Y_{n,k_n}$ are identical $P_0$-distributed under the null. By using techniques of extreme value theory it is sometimes possible to show that

$$\max_{1 \le i \le k_n} \left\{ \varepsilon_n \frac{\mathrm{d}\mu_n}{\mathrm{d}P_0}(Y_{n,i}) \right\} \overset{\mathrm{d}}{\longrightarrow} \widetilde{Y}$$

for a real-valued random variable $\widetilde{Y}$. Note that $\max_{1 \le i \le k_n} \{P_{n,i}(\varepsilon_{n,i} \frac{\mathrm{d}\mu_{n,i}}{\mathrm{d}P_{n,i}} > \tau)\} \le \tau^{-1} \max_{1 \le i \le k_n} \varepsilon_{n,i} \to 0$. Hence, regarding (2.6) we get the following connection to the Lévy measure $\eta_1$ of $\xi_1$:

$$P(\widetilde{Y} > e^x - 1) = \exp(-\eta_1(x, \infty))$$

for all $x$ coming from a dense subset of $(0, \infty)$. This may be useful to get a first impression how to choose $\mu_n$ and $\varepsilon_n$ to obtain nontrivial limits.

### 2.3. *Asymptotic relative efficiency*

In the case of normal distributed limits we have

$$T_n \overset{\mathrm{d}}{\longrightarrow} \begin{cases} \xi_1 \sim N(-\sigma^2/2, \sigma^2) \text{ under } P_{(n)} \text{ (null)}, \\ \xi_2 \sim N(\sigma^2/2, \sigma^2) \text{ under } Q_{(n)} \text{ (alternative)}, \end{cases} \tag{2.8}$$

for some $\sigma \in [0, \infty)$, where $N(0,0)$ denotes the Dirac measure $\epsilon_0$ centered in 0. In the case of $\sigma = 0$ no test sequence can separate between the null and the alternative asymptotically, see Section 2.1. Observe that both normal distributed limits depend only on one parameter, namely $\sigma^2$. In Appendix A, see Theorem A.1, we give many different equivalent conditions for normal distributed $\xi_1$ and $\xi_2$, even the conditions in Theorem 2.2 can be simplified in this case. Further equivalent conditions and closely related results can be found in Section A3 and A4 of Janssen [25]. In this section we restrict ourselves to these kind of limits, excluding the trivial case $\sigma = 0$, and discuss the LLR test's power behaviour if the "wrong" signal distributions and/or the "wrong" signal probabilities are chosen for the test statistic. To be more specific, we fix the triangular schemes of noise distributions $\{P_{n,i} : 1 \leq i \leq n \in \mathbb{N}\}$ and consider for $j = 1, 2$ a triangular scheme of signal distributions $\boldsymbol{\mu^{(j)}} = \{\mu_{n,i}^{(j)} : 1 \leq i \leq n \in \mathbb{N}\}$ as well as one of signal probabilities $\boldsymbol{\varepsilon^{(j)}} = \{\varepsilon_{n,i}^{(j)} : 1 \leq i \leq n \in \mathbb{N}\}$. Let $\boldsymbol{\theta_1} = (\boldsymbol{\mu^{(1)}}, \boldsymbol{\varepsilon^{(1)}})$ be the true, underlying model and $\boldsymbol{\theta_2} = (\boldsymbol{\mu^{(2)}}, \boldsymbol{\varepsilon^{(2)}})$ be the model pre-chosen by the statistician for the LLR test. Denote by $T_n(\boldsymbol{\theta_j})$ and $\varphi_n(\boldsymbol{\theta_j}) = \mathbf{1}\{T_n(\boldsymbol{\theta_j}) > c_{n,j}\}$ the LLR statistic and the LLR test for the model $\boldsymbol{\theta_j}$, $j = 1, 2$. Using Pitman's asymptotic relative efficiency, see Hájek et al. [17], we quantify the loss in terms of the asymptotic power if $\varphi_n(\boldsymbol{\theta_2})$ instead of the optimal $\varphi_n(\boldsymbol{\theta_1})$ is used.

**Theorem 2.6** (LLR power under Gaussian limits)**.** *Suppose that $T_n(\boldsymbol{\theta_j})$, $j \in \{1,2\}$, converges to Gaussian limits, compare to (2.8), with $\sigma_j > 0$. Moreover, assume that for $j, r \in \{1, 2\}$ the limit*

$$\gamma(\boldsymbol{\theta_j}, \boldsymbol{\theta_r}) = \lim_{n \to \infty} \sum_{i=1}^{k_n} \varepsilon_{n,i}^{(j)} \varepsilon_{n,i}^{(r)} Cov_{P_{n,i}} \Big( \frac{\mathrm{d}\mu_{n,i}^{(j)}}{\mathrm{d}P_{n,i}}, \frac{\mathrm{d}\mu_{n,i}^{(r)}}{\mathrm{d}P_{n,i}} \Big) \qquad (2.9)$$

*exists in $\mathbb{R}$. Suppose that $\gamma(\boldsymbol{\theta_j}, \boldsymbol{\theta_j}) = \sigma_j^2$. Let the critical values $c_{n,j}$ be chosen such that both tests $\varphi_n(\boldsymbol{\theta_1})$ and $\varphi_n(\boldsymbol{\theta_2})$ are asymptotically exact of a pre-chosen size $\alpha \in (0,1)$, i.e. $E_{\mathcal{H}_{0,n}}(\varphi_n(\boldsymbol{\theta_1})) \to \alpha$. Then the asymptotic power of the pre-chosen LLR test $\varphi_n(\boldsymbol{\theta_2})$ under the alternative $\mathcal{H}_{1,n}(\boldsymbol{\theta_1})$ of the true, underlying model $\boldsymbol{\theta_1}$ is given by*

$$E_{\mathcal{H}_{1,n}(\boldsymbol{\theta_1})}(\varphi_{n,\boldsymbol{\theta_2}}) \to \Phi\Big( \frac{\gamma(\boldsymbol{\theta_1}, \boldsymbol{\theta_2})}{\sqrt{\gamma(\boldsymbol{\theta_2}, \boldsymbol{\theta_2})}} + u_\alpha \Big)$$

$$= \Phi\Big( \mathrm{sign}(\gamma(\boldsymbol{\theta_1}, \boldsymbol{\theta_2})) \sqrt{\gamma(\boldsymbol{\theta_1}, \boldsymbol{\theta_1}) \mathrm{ARE}} + u_\alpha \Big),$$

$$where \ \mathrm{ARE} = \frac{\gamma(\boldsymbol{\theta_1}, \boldsymbol{\theta_2})^2}{\gamma(\boldsymbol{\theta_1}, \boldsymbol{\theta_1}) \gamma(\boldsymbol{\theta_2}, \boldsymbol{\theta_2})} \in [0, 1]$$

*is Pitman's asymptotic relative efficiency, see Hájek et al. [17].*

*Remark* 2.7. The assumption $\gamma(\boldsymbol{\theta_j}, \boldsymbol{\theta_j}) = \sigma_j^2$ is connected to the classical Lindeberg-condition. It is often but not always fulfilled if (2.8) holds. For example, it is violated in the case $\beta = 3/4$ and $r = \rho(\beta)$ for the heterogeneous normal mixture model, which is discussed in Section 4.2. The good news are that by a

truncation argument we find for every model $\boldsymbol{\theta} = (\boldsymbol{\mu}, \boldsymbol{\varepsilon})$, for which (2.8) holds, another $\widetilde{\boldsymbol{\theta}} = (\widetilde{\boldsymbol{\mu}}, \widetilde{\boldsymbol{\varepsilon}})$ such that the limit $\gamma(\widetilde{\boldsymbol{\theta}}, \widetilde{\boldsymbol{\theta}})$ from (2.9) exists and equals $\sigma^2$ from (2.8), and, moreover, the test's asymptotic behaviour is not effected by replacing $\boldsymbol{\theta}$ by $\widetilde{\boldsymbol{\theta}}$. The details are carried out in Appendix A, see Lemma A.3.

Note that Theorem 2.6 gives the sharp upper bound of the asymptotic power for all tests of asymptotic size $\alpha \in (0, 1)$ if (2.8) holds for the underlying model. The asymptotic relative efficiency ARE is a good tool to quantify the loss of power if the wrong LLR test is used. If ARE $= 1$ there is no loss of power by using $\varphi_n(\boldsymbol{\theta_2})$ and if ARE $= 0$ the test $\varphi_n(\boldsymbol{\theta_2})$ cannot distinguish between the null and the alternative asymptotically. Consider for a moment the rowwise identical case, i.e. $P_{n,i} = P_{n,1}$, $\mu_{n,i}^{(1)} = \mu_{n,1}^{(1)}$ etc. If ARE $\in (0, 1)$ then, heuristically, $(1 - \text{ARE}) \cdot 100\%$ of the observations are wasted. To be more specific, it can be shown that $\varphi_n(\boldsymbol{\theta_2})$ based on all $k_n$ observations $(Y_{n,1}, \ldots, Y_{n,k_n})$ achieves the same power as the optimal test does when only $m = [(1 - ARE)k_n]$ observations $(Y_{n,1}, \ldots, Y_{n,m})$ are used, where $[x]$ is the integer part of $x \in \mathbb{R}$.

### 2.4. Violation of (1.4)

Here, we discuss how to handle a violation of (1.4). This issue was already discussed by Cai and Wu [6], see their Section III.C, in terms of the Hellinger distance to determine the detection boundary. Their idea can be used for our purpose to determine, more generally, the limits of $T_n$, even on the boundary. Instead of the original model it is sufficient to analyse a "closely related" model for which (1.4) is fulfilled.

By Lebesgues' decomposition, see Lemma 1.1 of Strasser [36], there exist a constant $\lambda_{n,i} \in [0, 1]$, a $P_{n,i}$-null set $N_{n,i}$ as well as probability measures $\widetilde{\mu}_{n,i}$ and $\nu_{n,i}$ such that $\widetilde{\mu}_{n,i} \ll P_{n,i}$, $\nu_{n,i}(N_{n,i}) = 1$ and $\mu_{n,i} = (1 - \lambda_{n,i})\widetilde{\mu}_{n,i} + \lambda_{n,i}\nu_{n,i}$. Now, let $\widetilde{Q}_{n,i}$, $\widetilde{Q}_{(n)}$ and $\widetilde{T}_n$ defined as $Q_{n,i}$, $Q_{(n)}$ and $T_n$ replacing $\mu_{n,i}$ and $\varepsilon_{n,i}$ by $\widetilde{\mu}_{n,i}$ and $\widetilde{\varepsilon}_{n,i} = (1 - \lambda_{n,i})\varepsilon_{n,i}$, respectively. Clearly, for this new model (1.4) is fulfilled and our results can be applied to determine the limits of $\widetilde{T}_n$. When knowing these we can immediately give the ones of $T_n$:

**Corollary 2.8.** *Suppose that* (2.1) *is fulfilled for* $\widetilde{T}_n$, $\widetilde{\xi}_1$ *and* $\widetilde{\xi}_2$. *Moreover, assume that* $\sum_{i=1}^{k_n} \varepsilon_{n,i}\lambda_{n,i} \to c \in [0, \infty]$. *Then* (2.1) *holds for* $T_n$, $\xi_1 = \widetilde{\xi}_1 - c$ *and* $\xi_2 = \widetilde{\xi}_2 + X$, *where* $X$ *is independent of* $\widetilde{\xi}_2$ *with* $P(X = -c) = e^{-c}$ *and* $P(X = \infty) = 1 - e^{-c}$. *In particular,* $\xi_1 \equiv -\infty$ *and* $\xi_2 \equiv \infty$ *if* $c = \infty$, *or if* $\widetilde{\xi}_1 \equiv -\infty$ *and* $\widetilde{\xi}_2 \equiv \infty$.

We can state the results of Corollary 2.8 also in terms of distributions. Denote by $\widetilde{\nu}_j$ the distribution of $\widetilde{\xi}_j$. Then $\nu_1 = \widetilde{\nu}_1 * \epsilon_{-c}$ and $\nu_2 = e^{-c}\widetilde{\nu}_2 * \epsilon_{-c} + (1 - e^{-c})\epsilon_\infty$.

## 3. Power of the higher criticism test

In the previous section we discussed the LLR test which can be used to detect simple alternatives from the null. An adaptive and applicable test for alterna-

tives of the whole completely detectable area is Tukey's HC test modified by Donoho and Jin [12]. There are different versions of it. To relax the notation, we decided to use the one dealing with continuously distributed $p$-values having a quantile transformation in mind, see also the explanations at the beginning of Section 1.2. The optimality of HC in a discrete model, namely the Poisson means model, was shown by Arias-Castro and Wang [2]. Our results about the LLR statistic in Section 2 are also valid for discrete models but in this section we only regard continuous ones. The extension to discrete models is a possible project for the future.

The HC statistic for outcomes $p_{n,i} \in [0,1]$ is defined by

$$\mathrm{HC}_n = \sup_{t \in (0,1)} \left| \sqrt{k_n} \, \frac{\mathbb{F}_n(t) - t}{\sqrt{t(1-t)}} \right|,$$

where $\mathbb{F}_n$ is the empirical distribution function of the observation vector $(p_{n,i})_{i \leq k_n}$. For every $t \in (0,1)$ we compare the empirical distribution function and the null/noise distribution function $t \mapsto F(t) = t$. This difference is normalized in the spirit of the central limit theorem. For a fixed $t$ the resulting fraction is asymptotically standard normal distributed. The interval $(0,1)$, over which the supremum is taken, can be replaced by $(0, \alpha_0)$, $(k_n^{-1}, \alpha_0)$ or $(k_n^{-1}, 1-k_n^{-1})$ for some tuning parameter $\alpha_0 \in (0,1)$, see Donoho and Jin [12]. The test statistic can also be defined without taking the absolute value of the fraction. All these versions of the HC statistic would lead here to the same power results. To improve the readability of this section we give the results only for the HC version introduced above. By Jaeschke [22], see also Eicker [14], the limit distribution of $\mathrm{HC}_n$ is known under the null. We have

$$P_{(n)}(a_n \mathrm{HC}_n - b_n \leq x) \to \Lambda(x)^2 = \exp(-2\exp(-x)), \; x \in \mathbb{R}, \qquad (3.1)$$

where $\Lambda$ is the distribution function of a standard Gumbel distribution and the following normalisation constants are used

$$a_n = \sqrt{2\log\log(k_n)} \text{ and } b_n = 2\log\log(k_n) + \frac{1}{2}\log\log\log(k_n) - \frac{1}{2}\log(\pi).$$

Hence, the test $\varphi_{n,HC,\alpha} = \mathbf{1}\{\mathrm{HC}_n > c_n(\alpha)\}$ with

$$c_n(\alpha) = \frac{-\log(-\log(\alpha)/2) + b_n}{a_n} = \sqrt{2\log\log(k_n)}(1 + o(1))$$

is an asymptotically exact level $\alpha \in (0,1)$ test, i.e. $E_{\mathcal{H}_{0,n}}(\varphi_{n,HC,\alpha}) \to \alpha$. But we cannot recommend to use these critical values based on the limiting distribution since the convergence rate is really slow, see Khmaladze and Shinjikashvili [29]. Since the noise distribution is known, standard Monte-Carlo simulations can be used to estimate the $\alpha$-quantile of $\mathrm{HC}_n$ for finite sample size. Alternatively, you can find finite recursion formulas for the exact finite distribution in the paper of Khmaladze and Shinjikashvili [29].

In the following we present our tool for HC.

**Theorem 3.1** (Completely detectable by HC)**.** *Define for all $v \in (0, 1/2)$*

$$H_n(v) = \frac{|\sum_{i=1}^{k_n} \varepsilon_{n,i}(\mu_{n,i}(0,v] - v)| + |\sum_{i=1}^{k_n} \varepsilon_{n,i}(\mu_{n,i}(1-v,1) - v)|}{\sqrt{k_n v}}. \quad (3.2)$$

*Let $(v_n)_{n \in \mathbb{N}}$ be a sequence in the interval $(0, 1/2)$ such that $a_n^{-1} H_n(v_n) \to \infty$ and $\liminf_{n \to \infty} k_n v_n > 0$. Then $a_n HC_n - b_n \to \infty$ in $Q_{(n)}$-probability.*

Basically, we compare the tails near to 0 and 1 of the signal and the noise distribution. This verification method for HC's optimality is an extension of the ones used by Cai et al. [5] and Donoho and Jin [12]. Under the assumptions of Theorem 3.1 the sum of HC's error probabilities tends to 0 for appropriate critical values. In other words, HC can completely separate the null and the alternative.

The same $H_n(v)$ can be used to show that HC has no power under the alternative, i.e. the sum of error probabilities tends to 1 independently how the critical values are chosen.

**Theorem 3.2** (Undetectable by HC)**.** *Suppose that $P_{n,i} = P_n$, $\varepsilon_{n,i} = \varepsilon_n$ and $\mu_{n,i} = \mu_n$ do not depend on $i$. Define $H_n(v)$ as in Theorem 3.1. Moreover, assume that $P_{(n)}$ and $Q_{(n)}$ are mutually contiguous, compare to Remark 2.3. If*

$$a_n \sup\{H_n(v) : v \in [r_n, s_n] \cup [t_n, u_n]\} \to 0, \; where \quad (3.3)$$

$$\frac{\log(r_n)}{\log(k_n)} \to -1, \; \frac{\log(u_n)}{\log(k_n)} \to 0, \; and \; \frac{\log(s_n)}{\log(k_n)}, \frac{\log(t_n)}{\log(k_n)} \to \kappa \in (0,1) \quad (3.4)$$

*for some sequences $r_n, s_n, t_n, u_n \in (0,1)$ then*

$$Q_{(n)}(a_n HC_n - b_n \le x) \to \Lambda(x)^2 = \exp(-2\exp(-x)), \; x \in \mathbb{R}. \quad (3.5)$$

*Remark* 3.3. Suppose that $a_n^2 \sum_{i=1}^{k_n} \varepsilon_{n,i}^2 \to 0$, which is usually fulfilled for sparse signals. From Hölder's inequality $(a_n/\sqrt{k_n}) \sum_{i=1}^{k_n} \varepsilon_{n,i} \to 0$ follows. Hence, it is easy to see that the statements of Theorems 3.1 and 3.2 remain true if $H_n(v)$ is replaced by

$$\widetilde{H}_n(v) = \frac{1}{\sqrt{k_n v_n}} \Big( \sum_{i=1}^{k_n} \varepsilon_{n,i}(\mu_{n,i}(0,v] + \mu_{n,i}(1-v,1)) \Big), \; v \in \Big(0, \frac{1}{2}\Big).$$

## 4. Application to practical detection models

### *4.1. Nonparametric alternatives for p-values*

Here, we discuss a generalisation of the $p$-values model (1.6). In particular, we suppose $P_{n,i} = P_0 = \lambda_{|(0,1)}$. In contrast to Section 1.2, we now consider that the shape function $h_{n,i}$, the shrinking parameter $\kappa_{n,i} > 0$ and the signal probability $\varepsilon_{n,i}$ may depend on $i$. The assumption that the signal distribution

has a shrinking support can be too restrictive for practice. But the approach allows an extension of the model in the way that we add a perturbation $r_{n,i}$. Throughout this section we consider signal distributions $\mu_{n,i}$ given by

$$\frac{\mathrm{d}\mu_{n,i}}{\mathrm{d}P_0}(u) = \frac{1}{\kappa_{n,i}} h_{n,i}\Big(\frac{u}{\kappa_{n,i}}\Big) + r_{n,i}(u) \geq 0 \text{ with } \int_0^1 r_{n,i} \,\mathrm{d}P_0 = 0, \qquad (4.1)$$

where $h_{n,i}$ is close to some $h \in L^1(P_0)$ and the perturbation $r_{n,i}$ is "small" in the sense that

$$\sum_{i=1}^{k_n} \varepsilon_{n,i}^2 \int_0^1 r_{n,i}^2 \,\mathrm{d}P_0 \to 0. \qquad (4.2)$$

Instead of (1.5) we suppose that

$$\max_{1 \leq i \leq k_n} (\varepsilon_{n,i} + \kappa_{n,i}) \to 0.$$

Since we already presented the results concerning this model for the rowwise identical case $\mu_{n,i} = \mu_n$ and $\varepsilon_{n,i} = \varepsilon_n$ in Section 1.2, the theorems are stated only in their general versions here.

**Theorem 4.1.** *Suppose that*

$$\sum_{i=1}^{k_n} \frac{\varepsilon_{n,i}^2}{\kappa_{n,i}} \to K \in [0,\infty] \ and \ \max_{1 \leq i \leq k_n} \int_0^1 (h_{n,i} - h)^2 \,\mathrm{d}P_0 \to 0 \qquad (4.3)$$

*for some* $h, h_{n,i} \in L^2(P_0)$. *Without loss of generality we can suppose that*

$$\frac{\varepsilon_{n,1}}{\kappa_{n,1}} \leq \frac{\varepsilon_{n,2}}{\kappa_{n,2}} \leq \cdots \leq \frac{\varepsilon_{n,k_n}}{\kappa_{n,k_n}}.$$

(a) *(Undetectable case) If* $K = 0$ *then the undetectable case is present.*
(b) *(Completely detectable case) If* $K = \infty$,

$$\sum_{i=r_n}^{k_n} \varepsilon_{n,i} \to \infty \ and \ \sum_{i=1}^{r_n} \frac{\varepsilon_{n,i}^2}{\kappa_{n,i}} \to \infty \qquad (4.4)$$

*for some* $r_n \in \{1, \ldots, k_n\}$ *then we are in the completely detectable case.*
(c) *If* $\sup_{n\in\mathbb{N}} \sum_{i=1}^{k_n} \varepsilon_{n,i} < \infty$ *or* $K < \infty$ *then every accumulation point* $\xi_1$ *(in the sense of convergence in distribution) of* $T_n$, *compare to* (2.1), *is real-valued under the null. In particular, if* $K \in [0,\infty)$ *and*

$$\max_{1 \leq i \leq k_n} \frac{\varepsilon_{n,i}}{\kappa_{n,i}} = \frac{\varepsilon_{n,k_n}}{\kappa_{n,k_n}} \to 0 \qquad (4.5)$$

*then the limits of* $T_n$ *are Gaussian and* (2.8) *holds for* $\sigma^2 = \sigma^2(h) = K \int_0^1 h^2 \,\mathrm{d}P_0$.

(d) *In the spirit of Section 2.3, let $\boldsymbol{\theta_j} = \{(h_{n,i}^{(j)}, \kappa_{n,i}^{(j)}, \varepsilon_{n,i}^{(j)})_{i \leq k_n} : n \in \mathbb{N}\}$ denote a model for $j = 1,2$ such that (4.3) and (4.5) hold for some $K^{(j)} \in (0,\infty)$ and $h^{(j)} \in L^2(P_0)$. Then all assumptions of Theorem 2.6 are satisfied with*

$$\gamma(\boldsymbol{\theta_1}, \boldsymbol{\theta_2}) = \lim_{n\to\infty} \sum_{i=1}^{k_n} \frac{\varepsilon_{n,i}^{(1)}\varepsilon_{n,i}^{(2)}}{\kappa_{n,i}^{(1)}\kappa_{n,i}^{(2)}} \int_0^{\min\{\kappa_{n,i}^{(1)}, \kappa_{n,i}^{(2)}\}} h_{n,i}^{(1)}(x/\kappa_{n,i}^{(1)}) h_{n,i}^{(2)}(x/\kappa_{n,i}^{(2)}) \, \mathrm{d}x$$

*if this limit exists.*

The detection boundary introduced in (1.7) follows immediately from Theorem 4.1(a) and (b). The asymptotic behaviour on this boundary, discussed in II, can be deduced from Theorem 4.1(c). As stated in V, the case $\beta = 1$ is of special interest. If $\beta = 1$ and $r < 1$ then the pair $(\beta, r) = (1, r)$ belongs to the undetectable area by Theorem 4.1(a). But, if in addition to $\beta = 1$ we have either $r = 1$ or $r > 1$ then we obtain non-Gaussian limits $\xi_1$ and $\xi_2$, note that (4.5) is not fulfilled anymore. Details about the actual limits' distributions are presented in the subsequent Theorem 4.3 and Remark 4.4. Using Theorem 4.1(d) we can calculate the asymptotic relative efficiency ARE if the LLR test $\varphi_n(\boldsymbol{\theta_2})$ is used although $\boldsymbol{\theta_1}$ is the underlying model, see III and the following Remark 4.2. In addition to the rowwise identical scenario, the general formulation of Theorem 4.1 allows also a discussion, for instance, of a two-sample alternative with mainly $\varepsilon_{n,i} = 0$ and only sparse positive $\varepsilon_{n,i} > 0$.

*Remark* 4.2. Suppose the conditions of Theorem 4.1(d) are fulfilled.

(i) (No power under different shrinking) Assume that $\kappa_{n,i}^{(1)}(\kappa_{n,i}^{(2)})^{-1}$ converges uniformly for $i \in \{1, \ldots, k_n\}$ to 0 or to $\infty$. From Cauchy Schwartz's inequality we get $\gamma(\boldsymbol{\theta_1}, \boldsymbol{\theta_2}) = 0$ and, hence, ARE $= 0$.

(ii) If $\varepsilon_{n,i}^{(1)} = \varepsilon_{n,i}^{(2)}$ and $\kappa_{n,i}^{(1)} = \kappa_{n,i}^{(2)}$ in Theorem 4.1(d) then $\gamma(\boldsymbol{\theta_1}, \boldsymbol{\theta_2})$ can be expressed in terms of $K^{(1)} = K^{(2)}$, $h^{(1)}$ and $h^{(2)}$. In particular, we obtain

$$\text{ARE} = \frac{<h^{(1)}, h^{(2)}>^2}{<h^{(1)}, h^{(1)}><h^{(2)}, h^{(2)}>}, \text{ where } <f, g> = \int_0^1 fg \, \mathrm{d}P_0.$$

If $\varepsilon_{n,i} = \varepsilon_n$ and $\kappa_{n,i} = \kappa_n$ does not depend on $i = 1, \ldots, k_n$ then (4.4) is fulfilled for $r_n = [k_n/2]$ if and only if $K = \infty$ and $k_n\varepsilon_n \to \infty$. Combining this and Theorem 4.1 yields the detection boundary presented in I from Section 1.2 and the Gaussian limits introduced in II on this boundary if $\beta < 1$. Next, we give the generalisation of the result stated in IV from Section 1.2 concerning the case $\beta = 1$.

**Theorem 4.3** (Extreme case $\beta = 1$). *Let $\kappa_{n,i} = k_n^{-r}$, $r > 0$, and $\varepsilon_{n,i} = k_n^{-1}$. Let $\mathcal{D}$ be a dense subset of $(0, \infty)$ and $M$ be a measure on $(0, \infty]$ with $M(\{\infty\}) = 0$ such that $M(x, \infty) < \infty$ for all $x \in \mathcal{D}$ and*

$$\max_{1 \leq i \leq n} \left| \int_0^1 h_{n,i} \mathbf{1}\{h_{n,i} > e^x - 1\} \, \mathrm{d}P_0 - M(x, \infty) \right| \to 0. \tag{4.6}$$

*Then (2.1) holds for $\xi_1$ and $\xi_2$ given as follows:*

(a) *(Undetectable case) If $r < 1$ then $\xi_1 \equiv \xi_2 \equiv 0$.*

(b) *If $r = 1$ then $\xi_j$, $j \in \{1,2\}$, is infinitely divisible with Lévy-Khintchine triplet $(\gamma_j, 0, \eta_j)$, where $\gamma_j$ and $\eta_j$ are given by (2.4), (2.5) and (2.7).*

(c) *If $r > 1$ then $\xi_1 \equiv -1$ and $\xi_2 \sim e^{-1}\epsilon_{-1} + (1 - e^{-1})\epsilon_{\infty}$.*

*Remark* 4.4. Let $h \in L^1(P_0)$. Suppose that $h_{n,i} = h_n$, $\int_0^1 |h_n - h| \, \mathrm{d}P_0 \to 0$ and $P_0(u \in (0,1) : h(u) = x) = 0$ for all $x > 0$. Note that the latter is always fulfilled for strictly monotone $h$. Then (4.6) holds for $M$ given by $M(x, \infty) = \int_0^1 h \, \mathbf{1}\{h > e^x - 1\} \, \mathrm{d}P_0$. Consequently, if $r = 1$ then $\eta_1 = \mathcal{L}(\log(h + 1)|P_0)$, or in other words $\eta_1$ equals the distribution of $\log(h(U) + 1)$ for a uniformly distributed $U$ on $(0,1)$.

Note that we need for the statements in Theorem 4.3 and Remark 4.4 only $h \in L^1(P_0)$, and not $h \in L^2(P_0)$ as in Theorem 4.1. It is also possible to determine the detection boundary if $h \notin L^2(P_0)$. In this case we get nontrivial Lévy measures on the whole detection boundary depending heavily on the shape of $h$ comparable to the situation in Theorem 4.3(b). In the following we discuss an example for $h \in L^1(P_0) \setminus L^2(P_0)$.

**Theorem 4.5.** *Let $h_{n,i}(x) = h(x) = (1 - \alpha)x^{-\alpha}$ for all $x \in (0,1)$ and some $\alpha \in [1/2, 1)$. Moreover, let $k_n = n$, $\varepsilon_{n,i} = n^{-\beta}$, $\beta \in (1/2, 1)$, and $\kappa_{n,i} = n^{-r}$, $r > 0$. Then the detection boundary is given by*

$$\rho^{\#}(\beta, \alpha) \ = \ \min\left(0, \frac{\beta - \alpha}{1 - \alpha}\right). \tag{4.7}$$

*In detail, $r < \rho^{\#}(\beta, \alpha)$ (resp. $r > \rho^{\#}(\beta, \alpha)$) leads to the undetectable case (resp. completely detectable case). If $r = \rho^{\#}(\beta, \alpha)$ then $T_n$ converges to infinitely divisible $\xi_j$, $j \in \{1,2\}$, with Lévy-Khintchine triplet $(\gamma_j, 0, \eta_j)$ under $\mathcal{H}_{0,n}$ and $\mathcal{H}_{1,n}$, respectively. $\gamma_j$ and $\eta_j$ are uniquely determined by (2.4), (2.5) and*

$$\frac{\mathrm{d}\eta_j}{\mathrm{d}\lambda}(x) = \frac{(1 - \alpha)^{\frac{1}{\alpha}}}{\alpha} e^x (e^x - 1)^{-\frac{1}{\alpha} - 1}, \ x > 0,$$

The limit in Theorem 4.5 for $r = \rho^{\#}(\beta, \alpha)$ does not coincide with the one for $\beta = 1$ from Theorem 4.3(b) with $h_{n,i}(x) = (1 - \alpha)x^{-\alpha}$. Moreover, note that the case $\alpha < 1/2$ is included in Theorem 4.1, see also Figure 1 as well as I and II in the introduction.

Let us now consider the HC test. Since the given model is one for $p$-values the observations do not need to be transformed. Hence, the HC test is based on $p_{n,i} = Y_{n,i}$.

**Theorem 4.6** (Higher criticism). *Consider the model*

(i) *from Section 1.2, where $h \in L^{2+\delta}(P_0)$ for some $\delta \in (0,1)$, or*

(ii) *from Theorem 4.5.*

*Then the areas of complete detection of the HC and the LLR test coincide. HC cannot distinguish between the null and the alternative asymptotically if $r \leq 1$ and $r = \rho(\beta)$ or $r = \rho^{\#}(\beta, \alpha)$, respectively, i.e. on the detection boundary.*

Moreover, under the model assumptions of Theorem *4.3* with $h_{n,i} = h_n$ *HC cannot distinguish between the null and the alternative asymptotically if* $\beta = r = 1$.

## *4.2. Heteroscedastic normal mixtures*

The heteroscedastic normal mixture model was already studied essentially in the literature (e.g., [5, 12, 19]). Nevertheless, we can give, as a further application of our results, some new insights about it concerning the extension of the detection boundary and the asymptotic power of the HC test on the boundary. But we first introduce the model. Let $k_n = n$, $P_{n,i} = P_0 = N(0,1)$ and $\mu_{n,i} = \mu_n = N(\vartheta_n, \sigma_0^2)$, $\sigma_0 > 0$, where the parametrisation $\varepsilon_{n,i} = \varepsilon_n = n^{-\beta}$ and $\vartheta_n = \sqrt{2r \log n}$ with $\beta \in (1/2, 1)$ and $r > 0$ is used. The detection boundary given by

$$
\rho(\beta, \sigma_0) = \begin{cases} 
\left(2 - \sigma_0^2\right)\left(\beta - \frac{1}{2}\right) & \text{if } \frac{1}{2} < \beta \leq 1 - \frac{\sigma_0^2}{4},\ \sigma_0 < \sqrt{2}, & \text{(I)} \\
\left(1 - \sigma_0\sqrt{1-\beta}\right)^2 & \text{if } 1 - \frac{\sigma_0^2}{4} < \beta < 1,\ \sigma_0 < \sqrt{2}, & \text{(II)} \\
0 & \text{if } \frac{1}{2} < \beta \leq 1 - \frac{1}{\sigma_0^2},\ \sigma_0 \geq \sqrt{2}, & \text{(III)} \\
\left(1 - \sigma_0\sqrt{1-\beta}\right)^2 & \text{if } 1 - \frac{1}{\sigma_0^2} < \beta < 1,\ \sigma_0 \geq \sqrt{2}, & \text{(IV)}
\end{cases}
\tag{4.8}
$$

and the limits of $T_n$ on it were already determined by Cai et al. [5] and Ingster [19]. The detection boundary is plotted for different $\sigma_0$ in Figure 2. Moreover, it was shown that the completely detectable areas of the LLR and HC tests coincide, see Cai et al. [5], Donoho and Jin [12]. All these results can be proven by using our methods, see Ditzhaus [10]. Note that the HC test is applied to the vector $(p_{n,i})_{i \leq k_n}$ of $p$-values, which we get by transforming each observations $Y_{n,i}$ to $p_{n,i} = 1 - \Phi(Y_{n,i})$.

**Proposition 4.7** (see Theorems 5 and 6 of [5])**.**  *(a) If* $r < \rho(\beta, \sigma_0)$ *then we are in the undetectable case, i.e. no test can distinguish between the null* $\mathcal{H}_{0,n}$ *and the alternative* $\mathcal{H}_{n,1}$ *asymptotically.*

(b) *If* $r > \rho(\beta, \sigma_0)$ *then the LLR as well as the HC test can completely separate the null and the alternative asymptotically.*

(c) *Suppose that* $r = \rho(\beta, \sigma_0)$. *Moreover, add a logarithmic term in the parametrisation of* $\varepsilon_n$ *as follows:*

$$
\varepsilon_n = n^{-\beta}\left(\log(n)\right)^{E(\beta,\sigma_0)} \quad \text{with } E(\beta, \sigma_0) = \begin{cases} 0 & \text{on (I).} \\ \frac{1}{2} - \frac{\sqrt{1-\beta}}{2\sigma_0} & \text{else.} \end{cases}
\tag{4.9}
$$

*In the following we discuss the different parts (I), (II) and (IV) of the detection boundary.*

(i) *(Gaussian limits) Consider part (I). Then* $(2.8)$ *holds for*

$$
\sigma^2 = \left(\sigma_0\sqrt{2 - \sigma_0^2}\right)^{-1}\left(1 - \frac{1}{2}\mathbf{1}\left\{\beta = 1 - \frac{\sigma_0^2}{4}\right\}\right).
$$

(ii) *Consider the parts (II) and (IV). Then* (2.1) *holds for infinitely divisible $\xi_1$ and $\xi_2$ with Lévy-Khintchine triplets $(\gamma_1, 0, \eta_1)$ and $(\gamma_2, 0, \eta_2)$, respectively, where $\eta_1$, $\eta_2$ are given by*

$$\frac{\mathrm{d}\eta_1}{\mathrm{d}\lambda}(x) = \frac{1}{c_1}\left(e^x - 1\right)^{c_2-3} e^x \; and \; \frac{\mathrm{d}\eta_2}{\mathrm{d}\lambda}(x) = e^x \frac{\mathrm{d}\eta_1}{\mathrm{d}\lambda}(x), \; x > 0,$$

*with $c_1 = 2\sqrt{\pi}\sigma_0^{c_3}c_4$, $c_2 = c_4^{-1}(\sigma_0 - 2\sqrt{1-\beta})$, $c_3 = c_4^{-1}\sigma_0 - \sqrt{1-\beta}$ and $c_4 = \sigma_0 - \sqrt{1-\beta}$, and $\gamma_1$ and $\gamma_2$ fulfill* (2.4) *and* (2.5) *with $\sigma^2 = 0$.*

*Remark* 4.8. By carefully reading the proof of Cai et al. [5], see in particular the top of page 658, there must be an additional factor $1/2$ in the exponent of the logarithmic term in their definition of $\varepsilon_n$ as in our (4.9).

Applying our Theorem 3.2 we can show, as already postulated, that HC has no asymptotic power on the boundary.

**Theorem 4.9** (HC on the boundary). *Let $r = \rho(\beta, \sigma_0) > 0$, $\beta \in (1/2, 1)$. Moreover, reparametrize $\varepsilon_n$ on the quadratic part of the boundary as we did in* (4.9). *Then the HC test has no (asymptotic) power, whereas the LLR does so.*

In (4.8) the detection boundary is (only) defined for $\beta < 1$. As we already did in the previous section, we can extend this boundary for $\beta = 1$ by a infinite vertical line starting in $(r, \beta) = (1, 1)$, see Figure 2. Again, we observe on this line unusual limits of $T_n$.

**Theorem 4.10** (Detection boundary extension). *Let $\beta = 1$. In this case we use the original/non-reparametrized definition of $\varepsilon_n$, i.e., $\varepsilon_n = n^{-1}$.*

(i) *If $r < 1$ then the pair $(\beta, r) = (1, r)$ belongs to the undetectable region.*
(ii) *If $\beta = 1$ and $r = 1$ then $\xi_1 \equiv -1/2$ and $\xi_2 \sim e^{-1/2}\epsilon_{-1/2} + (1 - e^{-1/2})\epsilon_\infty$.*
(iii) *If $\beta = 1$ and $r > 1$ then $\xi_1 \equiv -1$ and $\xi_2 \sim e^{-1}\epsilon_{-1} + (1 - e^{-1})\epsilon_\infty$.*

The results concerning ARE can also be applied for the heteroscedastic models. Fix the variance parameter $\sigma_0 > 0$. Let $\boldsymbol{\theta_1} = (\beta_1, r_1)$ and $\boldsymbol{\theta_2} = (\beta_2, r_2)$ represent two models from the linear part (I) of the detection boundary leading to Gaussian limits of $T_n$. Suppose that the models are different, i.e. $\beta_1 \neq \beta_2$. By applying Theorem 2.6 and simple calculations, which are omitted to the reader, ARE $= 0$ can be shown. That means that the LLR test $\varphi_n(\boldsymbol{\theta_2})$ can not distinguish between the null and the alternative asymptotically when $\boldsymbol{\theta_1}$ is the true, underlying model. As already mentioned $\gamma(\boldsymbol{\theta_j}, \boldsymbol{\theta_j}) = \sigma_j^2$ does not hold if $\beta_j = 1 - \sigma_0^2/4$. In this case make use of the truncation Lemma A.3.

Cai et al. [5] already considered the dense case $\beta < 1/2$. In this case $\sigma_0^2 \neq 1$ always leads to the completely detectable case independently of how the signal strength $\vartheta_n$ is chosen. Thus, only the heterogeneous case $\sigma_0^2 = 1$ is of real interest. In this case the parametrisation $\vartheta_n = n^r$ is used for $r > 0$. The corresponding detection boundary is given by $\rho(\beta) = 1/2 - \beta$ and is plotted in Figure 2. The HC test achieves the same region of complete detection, see Cai et al. [5]. Our results concerning the tests' power behaviour on the detection
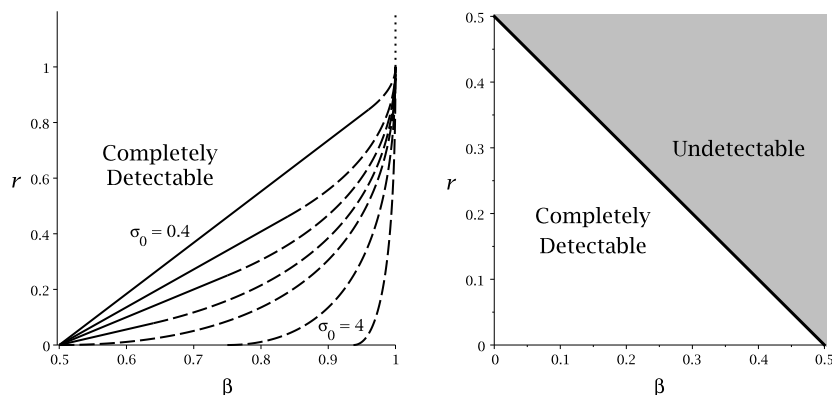
Fig 2. *Detection boundaries for the heteroscedastic normal mixture model. Left: (Sparse case for $\sigma_0 \in \{0.4, 0.8, 1, 1.2, \sqrt{2}, 2, 4\}$) Above the boundary is the completely detectable area and underneath is the undetectable area for both tests (LLR and HC). The limits $\xi_1$ and $\xi_2$ are Gaussian on the linear part (solid) and non-Gaussian on the quadratic part (dashed). In both cases the HC test has no asymptotic power. On the vertical dotted line $P(\xi_2 \in \mathbb{R}) \in (0, 1)$. Right: (dense case for $\sigma_0^2 = 1$) Above the boundary is the undetectable area and underneath is the completely detectable area for both tests. On the boundary the limits $\xi_1$ and $\xi_2$ are Gaussian and the HC test has no power.*

boundary can also be applied. In short, on the detection boundary (2.8) holds for some $\sigma > 0$ and the HC test has no asymptotic power there. This is even possible to a general class of one-parametric exponential families including the dense heterogeneous normal mixtures. Further details concerning the dense case can be found in Ditzhaus [10, 11].

## Appendix A: Gaussian limits

Gaussian limits $\xi_1$ and $\xi_2$, compare to (2.8), are of special interest, for example regarding Theorem 2.6. Recall that the degenerate case is included as $\sigma = 0$. In the following we give several equivalent conditions for Gaussian limits.

**Theorem A.1** (Gaussian limits)**.** *The conditions* (a)-(i) *are equivalent:*

(a) *$\xi_1$ and $\xi_2$ are Gaussian or $\xi_1 = \xi_2 \equiv 0$ with probability one.*
(b) *$\xi_1 \sim N(-\frac{\sigma^2}{2}, \sigma^2)$ for some $\sigma^2 \in [0, \infty)$.*
(c) *$\xi_2 \sim N(\frac{\sigma^2}{2}, \sigma^2)$ for some $\sigma^2 \in [0, \infty)$.*
(d) *$\xi_2$ is real-valued and $\xi_1 \sim N(a, \sigma^2)$ for some $a \in \mathbb{R}$, $\sigma^2 \in [0, \infty)$.*
(e) *$\xi_2 \sim N(a, \sigma^2)$ for some $a \in \mathbb{R}$, $\sigma^2 \in [0, \infty)$.*
(f) *$Z_n$ given by* (A.1) *converges in distribution under $P_{(n)}$ to some normal distributed $Z \sim N(0, \sigma^2)$ for some $\sigma^2 \in [0, \infty)$:*

$$Z_n = \sum_{i=1}^{k_n} \varepsilon_{n,i} \Big( \frac{\mathrm{d}\mu_{n,i}}{\mathrm{d}P_{n,i}} - 1 \Big) \xrightarrow{\mathrm{d}} Z. \tag{A.1}$$

(g) $\xi_2$ *is real-valued and* $\max_{1 \leq i \leq k_n} \frac{\mathrm{d}Q_{n,i}}{\mathrm{d}P_{n,i}} \to 1$ *in* $P_{(n)}$-*probability.*

(h) $\xi_2$ *is real-valued and* $\max_{1 \leq i \leq k_n} \varepsilon_{n,i} \frac{\mathrm{d}\mu_{n,i}}{\mathrm{d}P_{n,i}} \to 0$ *in* $P_{(n)}$-*probability.*

(i) *For some* $\tau \in (0, \infty)$ *and all* $x > 0$ *we have* $I_{n,1,x} \to 0$ *and* $I_{n,2,\tau} \to \sigma^2 \in [0, \infty)$.

*If one of the conditions* (b)–(f) *or* (i) *is fulfilled for some* $\sigma^2 \in [0, \infty)$ *then the others do so for the same* $\sigma^2$.

*Remark* A.2. Theorem A.1(i) holds for some $\tau > 0$ if and only if it does for all.

To apply Theorem 2.6 $\gamma(\boldsymbol{\theta}, \boldsymbol{\theta}) = \sigma^2$ is needed, where $\sigma^2$ comes from the previous section and $\boldsymbol{\theta}$ denotes the underlying model, compare to the notation in Section 2.3. As already mentioned there are examples, for which this equation fails although $\xi_1$ and $\xi_2$ are normal distributed. But by truncation we can always ensure the equality without changing the asymptotic results.

**Lemma A.3** (Truncation). *Let the assumptions of Theorem A.1 and one of its equivalent conditions* (a)-(i) *be fulfilled. In order to use a truncation argument define*

$$\widetilde{\varepsilon}_{n,i} = \varepsilon_{n,i}\mu_{n,i}\Big(\varepsilon_{n,i}\frac{\mathrm{d}\mu_{n,i}}{\mathrm{d}P_{n,i}} \leq \tau\Big) \text{ for some } \tau > 0$$

*and let* $\widetilde{\mu}_{n,i}$ *be given as follows: if* $\widetilde{\varepsilon}_{n,i} = 0$ *then* $\frac{\mathrm{d}\widetilde{\mu}_{n,i}}{\mathrm{d}P_{n,i}} = 1$, *and otherwise*

$$\frac{\mathrm{d}\widetilde{\mu}_{n,i}}{\mathrm{d}P_{n,i}} = \frac{\mathrm{d}\mu_{n,i}}{\mathrm{d}P_{n,i}}\mathbf{1}\Big\{\varepsilon_{n,i}\frac{\mathrm{d}\mu_{n,i}}{\mathrm{d}P_{n,i}} \leq \tau\Big\}\Big[\mu_{n,i}\Big(\varepsilon_{n,i}\frac{\mathrm{d}\mu_{n,i}}{\mathrm{d}P_{n,i}} \leq \tau\Big)\Big]^{-1}.$$

*All our asymptotic results in this paper remain the same if we replace* $\mu_{n,i}$ *and* $\varepsilon_{n,i}$ *by* $\widetilde{\mu}_{n,i}$ *and* $\widetilde{\varepsilon}_{n,i}$.

## Appendix B: Proofs

In the following we give all the proofs. These are not given in the order of their appearance since we apply, for example, Theorem 2.4 to verify Theorem 2.2. Before giving the proofs we introduce some useful properties of binary experiments and generalise limit theorems of Gnedenko and Kolmogorov [16] to infinitely divisible distributions.

### B.1. Binary experiments and distances for probability measures

Binary experiments classify different types of signal detectability. This gives us a first rough insight in the different detection regions for our signal detection problem. This standard approach is recalled for a sequence of binary experiments $\{\widetilde{P}_{(n)}, \widetilde{Q}_{(n)}\}$, $n \in \mathbb{N} \cup \{0\}$, where the underlying measurable spaces $(\Omega_n, \mathcal{A}_n)$

may change with $n$. Recall the equivalence of the weak convergences in (B.1) and (B.2) on $[-\infty, \infty]$:

$$\mathcal{L}\left(\log \frac{\mathrm{d}\widetilde{Q}_{(n)}}{\mathrm{d}\widetilde{P}_{(n)}}\Big|\widetilde{P}_{(n)}\right) \xrightarrow{\mathrm{w}} \mathcal{L}\left(\log \frac{\mathrm{d}\widetilde{Q}_{(0)}}{\mathrm{d}\widetilde{P}_{(0)}}\Big|\widetilde{P}_{(0)}\right) = \nu_1 \text{ (say)}, \qquad \text{(B.1)}$$

$$\mathcal{L}\left(\log \frac{\mathrm{d}\widetilde{Q}_{(n)}}{\mathrm{d}\widetilde{P}_{(n)}}\Big|\widetilde{Q}_{(n)}\right) \xrightarrow{\mathrm{w}} \mathcal{L}\left(\log \frac{\mathrm{d}\widetilde{Q}_{(0)}}{\mathrm{d}\widetilde{P}_{(0)}}\Big|\widetilde{Q}_{(0)}\right) = \nu_2 \text{ (say)}. \qquad \text{(B.2)}$$

Following Le Cam we say that $\{\widetilde{P}_{(n)}, \widetilde{Q}_{(n)}\}$ converges weakly to $\{\nu_1, \nu_2\}$ ($\{\widetilde{P}_{(0)}, \widetilde{Q}_{(0)}\}$, respectively) if and only if (B.1) or (B.2) is fulfilled. Note that every sequence of binary experiments has at least one accumulation point in the sense of weak convergence, see Lemma 60.6 of Strasser [36]. In general $\nu_1$ is a measure on $\mathbb{R} \cup \{-\infty\}$ and $\nu_2$ is one on $\mathbb{R} \cup \{\infty\}$ connected by

$$\frac{\mathrm{d}\nu_{2|\mathbb{R}}}{\mathrm{d}\nu_{1|\mathbb{R}}}(x) = e^x \text{ and } \nu_2(\{-\infty\}) = 1 - \int e^x \,\mathrm{d}\nu_1(x). \qquad \text{(B.3)}$$

Using the terminology of weak convergence of binary experiments we can express the different types of (asymptotic) detectability as follows:

- *completely detectable:* $\{P_{(n)}, Q_{(n)}\}$ converges weakly to the so called full informative experiment $\{\nu_1, \nu_2\} = \{\epsilon_{-\infty}, \epsilon_{\infty}\}$.
- *undetectable:* $\{P_{(n)}, Q_{(n)}\}$ converges weakly to the so called uninformative experiment $\{\nu_1, \nu_2\} = \{\epsilon_0, \epsilon_0\}$.
- *detectable:* None (weak) accumulation point of $\{P_{(n)}, Q_{(n)}\}$ is the uninformative experiment $\{\nu_1, \nu_2\} = \{\epsilon_0, \epsilon_0\}$.

The variational distance of probability measures $\widetilde{P}$ and $\widetilde{Q}$ on a common measure space $(\widetilde{\Omega}, \widetilde{\mathcal{A}})$ is given by

$$||\widetilde{P} - \widetilde{Q}|| = \sup\{E_{\widetilde{P}}(\varphi) - E_{\widetilde{Q}}(\varphi) : \text{ measurable } \varphi : \widetilde{\Omega} \to [0, 1]\}, \qquad \text{(B.4)}$$

see Lemma 2.3 of Strasser [36]. It is easy to show that weak convergence of $\{\widetilde{P}_{(n)}, \widetilde{Q}_{(n)}\}$ to $\{\widetilde{P}_{(0)}, \widetilde{Q}_{(0)}\}$ implies convergence of the variational distance $||\widetilde{P}_{(n)} - \widetilde{Q}_{(n)}|| \to ||\widetilde{P}_{(0)} - \widetilde{Q}_{(0)}||$. Our three cases can be reformulated to:

- *completely detectable:* $||P_{(n)} - Q_{(n)}||$ tends to 1.
- *undetectable:* $||P_{(n)} - Q_{(n)}||$ tends to 0.
- *detectable:* We have $\liminf_{n \to \infty} ||P_{(n)} - Q_{(n)}|| > 0$.

For product measures the Hellinger distance $d$ is useful:

$$d^2(\widetilde{P}, \widetilde{Q}) = \frac{1}{2} \int \left(\left(\frac{\mathrm{d}\widetilde{P}}{\mathrm{d}\nu}\right)^{\frac{1}{2}} - \left(\frac{\mathrm{d}\widetilde{Q}}{\mathrm{d}\nu}\right)^{\frac{1}{2}}\right)^2 \mathrm{d}\nu = 1 - \int \left(\frac{\mathrm{d}\widetilde{P}}{\mathrm{d}\nu}\frac{\mathrm{d}\widetilde{Q}}{\mathrm{d}\nu}\right)^{\frac{1}{2}} \mathrm{d}\nu, \qquad \text{(B.5)}$$

where $\widetilde{P}, \widetilde{Q} \ll \nu$. Since $d^2(\widetilde{P}, \widetilde{Q}) \le ||\widetilde{P} - \widetilde{Q}|| \le \sqrt{2}\, d(\widetilde{P}, \widetilde{Q})$, see Lemma 2.15 of Strasser [36], we obtain from (1.1) and (1.3) that

$$\max_{i=1,\ldots,k_n} d^2(P_{n,i}, Q_{n,i}) \le \max_{1 \le i \le k_n} ||P_{n,i} - Q_{n,i}|| \le \max_{1 \le i \le k_n} \varepsilon_{n,i} \to 0. \qquad \text{(B.6)}$$

Consequently, $d^2(P_{(n)}, Q_{(n)}) = 1 - \prod_{i=1}^{k_n}(1 - d^2(P_{n,i}, Q_{n,i}))$ tends to $b \in [0,1]$ if and only if $-\log(1-b)$ is the limit of

$$D_n = \sum_{i=1}^{k_n} d^2(P_{n,i}, Q_{n,i}). \tag{B.7}$$

To sum up, we get the following characterisation of the trivial detection regions.

**Lemma B.1.** *(a) We are in the undetectable case if and only if $D_n \to 0$.*
*(b) We are in completely detectable case if and only if $D_n \to \infty$.*

Note that from the connection between the variational distance and the Hellinger distance we obtain

$$\frac{1}{2}\sum_{i=1}^{k_n} \varepsilon_{n,i}^2 \,||P_{n,i} - \mu_{n,i}||^2 \le D_n \le \sum_{i=1}^{k_n} \varepsilon_{n,i}\,||P_{n,i} - \mu_{n,i}||. \tag{B.8}$$

### B.2. Limit theorems

For the readers' convenience let us recall well known convergence results of Gnedenko and Kolmogorov [16] which we use rapidly. Let $(Y_{n,i})_{1 \le i \le k_n}$ be a triangular array of row-wise independent, infinitesimal, real-valued random variables on some probability space $(\Omega, \mathcal{A}, P)$. In our case we have

$$\sum_{i=1}^{k_n} P(Y_{n,i} \le x) = 0 \tag{B.9}$$

for all fixed $x < 0$ if $n \ge N_x$ is sufficiently large. Combining this with (9) of Chap. 3.18, Theorem 4.25.4 and the subsequent remark of Gnedenko and Kolmogorov [16] yields:

**Theorem B.2.** *We have distributional convergence*

$$\sum_{i=1}^{k_n} Y_{k_n,i} \overset{\mathrm{d}}{\longrightarrow} Y$$

*to some real-valued $Y$ on $(\Omega, \mathcal{A}, P)$ if and only if the following conditions* (i)-(iii) *hold.*

*(i) There is a Lévy measure $\eta$ on $\mathbb{R} \setminus \{0\}$ such that $\eta(-\infty, 0) = 0$ and*

$$\sum_{i=1}^{k_n} P\left(Y_{k_n,i} > x\right) \to \eta(x, \infty) \in \mathbb{R} \text{ as } n \to \infty$$

*for all $x \in C_+(\eta)$, i.e. for all continuity points of $t \mapsto \eta(t, \infty)$, $t > 0$.*

(ii) *There exists some constant $\sigma^2 \in [0, \infty)$ such that*

$$\sigma^2 = \lim_{\varepsilon \searrow 0} \limsup_{n \to \infty} \liminf_{n \to \infty} \sum_{i=1}^{k_n} \int_{\{|Y_{k_n,i}| < \varepsilon\}} Y_{k_n,i}^2 \, \mathrm{d}P - \sum_{i=1}^{k_n} \left( \int_{\{|Y_{k_n,i}| < \varepsilon\}} Y_{k_n,i} \, \mathrm{d}P \right)^2.$$

(iii) *There is some constant $\gamma \in \mathbb{R}$ and $\tau_0 \in C_+(\eta)$ such that*

$$\lim_{n \to \infty} \sum_{i=1}^{k_n} \int Y_{k_n,i} \mathbf{1}\{|Y_{k_n,i}| < \tau_0\} \, \mathrm{d}P$$

$$= \gamma + \int_{(-\tau_0, \tau_0) \setminus \{0\}} \frac{x^3}{1 + x^2} \, \mathrm{d}\eta(x) - \int_{\mathbb{R} \setminus [-\tau_0, \tau_0]} \frac{x}{1 + x^2} \, \mathrm{d}\eta(x).$$

*Under* (i)-(iii) $Y$ *is infinitely divisible with Lévy-Khintchine triplet* $(\gamma, \sigma^2, \eta)$.

As stated in Theorem 2.4, we have to deal also with positive weights in $\infty$ for the limits since $\nu_2 = \rho + (1 - a)\epsilon_{-\infty}$, where $a < 1$ may occur.

**Theorem B.3.** *Suppose that the conditions* (ii) *and* (iii) *of Theorem B.2 hold for some $\tau_0 \in C_+(M_0)$. Assume that the following* (a) *and* (b) *hold.*

(a) *There is a dense subset $\mathcal{D}$ of $(0, \infty)$ and a measure $M_0$ on $(0, \infty]$ with*

$$\sum_{i=1}^{k_n} P\left(Y_{k_n,i} > x\right) \to M_0(x, \infty] \in \mathbb{R} \text{ for all } x \in \mathcal{D}.$$

(b) *There exists some $\tau_1 > 0$ such that*

$$\limsup_{n \to \infty} \sum_{i=1}^{k_n} \int_{\{|Y_{k_n,i}| < \tau_1\}} Y_{k_n,i}^2 \, \mathrm{d}P < \infty.$$

*Then,*

$$\mathcal{L}\left(\sum_{i=1}^{k_n} Y_{n,i}\right) \xrightarrow{\mathrm{w}} e^{-M_0(\{\infty\})}\nu + (1 - e^{-M_0(\{\infty\})})\epsilon_\infty,$$

*where $\nu$ is a infinitely divisible measure on $\mathbb{R}$ with Lévy-Khintchine triplet $(\gamma, \sigma^2, \eta)$ and Lévy measure $\eta = M_{0|(0,\infty)}$.*

*Proof.* Put $\eta = M_{0|(0,\infty)}$. Let the sequence $(M_n)_{n \in \mathbb{N}}$ consists of measures on $(0, \infty]$ given by $M_n(x, \infty] = \sum_{i=1}^{k_n} P(Y_{n,i} > x)$, $x > 0$. Clearly, $M_{n|(0,\infty)} \xrightarrow{\mathrm{w}} \eta$ and $\limsup_{n \to \infty} \int_{(0,\tau_1)} t^2 \, \mathrm{d}M_n(t) < \infty$. Thus, we obtain $\int \min(t^2, 1) \, \mathrm{d}\eta(t) < \infty$, which proves that $\eta$ is a Lévy measure. Define $Z_{n,u} = \sum_{i=1}^{k_n} Y_{n,i}\mathbf{1}\{Y_{n,i} \leq u\}$ for all $u \in \mathcal{D}$, $u > \tau_0$. By Theorem B.2 $Z_{n,u}$ converges in distribution to $X_u$, where $X_u$ is infinitely divisible with Lévy-Khintchine triplet $(\gamma_u, \sigma^2, \eta_u)$, Lévy measure $\eta_u = \eta_{(0,u]}$ and shift term

$$\gamma_u = \gamma - \int_{(u,\infty)} \frac{x}{1 + x^2} \, \mathrm{d}\eta(x).$$

Since $\eta$ is Lévy measure it is easy to verify $\gamma_u \to \gamma$ as $\mathcal{D} \ni u \to \infty$. By this and Theorem 3.19.2 of Gnedenko and Kolmogorov [16] $X_u$ converges in distribution to $X$ as $\mathcal{D} \ni u \to \infty$, where $X \sim \nu$. Now, let $(u_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{D}$ which tends to $\infty$ slowly enough such that $\sum_{i=1}^{k_n} P\left(Y_{k_n,i} > u_n\right) \to M_0(\{\infty\})$. Standard arguments, see Theorem 3.2 of Billingsley [4], imply that $Z_{n,u_n}$ converges in distribution to $X$ since for all $\delta > 0$

$$\limsup_{n\to\infty} P\left(\left|Z_{n,u} - Z_{n,u_n}\right| \geq \delta\right) \leq M_0(u, \infty) \to 0 \text{ as } \mathcal{D} \ni u \to \infty.$$

The basic idea to determine the limit distribution of $\sum_{i=1}^{k_n} Y_{n,i}$ is to condition on $C_n = \{\max_{1 \leq i \leq k_n} Y_{n,i} \leq u_n\}$. Note that for all $t \in \mathbb{R}$

$$P\left(\sum_{i=1}^{k_n} Y_{n,i} \leq t\right) = P(Z_{n,u_n} \leq t | C_n)P(C_n) + P\left(\sum_{i=1}^{k_n} Y_{n,i} \leq t, \max_{1 \leq i \leq k_n} Y_{n,i} > u_n\right),$$

where the latter summand tends to 0. Moreover, observe that

$$1 - P(C_n) = \prod_{i=1}^{k_n}\left(1 - P(Y_{n,i} > u_n)\right) \to e^{-M_0(\{\infty\})}.$$

It is remains to show that $Z_{n,u_n}$ tends to $X$ conditioned on $C_n$. Conditioned on $C_n$ we have $Z_{n,u_n} = \sum_{i=1}^{k_n} Y_{n,i}\mathbf{1}\{Y_{n,i} \leq u_n\}$ and $(Y_{n,i}\mathbf{1}\{Y_{n,i} \leq u_n\})_{i \leq k_n}$ is a rowwise independent and infinitesimal triangular array. Hence, we can apply Theorem B.2 to $Z_{n,u_n}$ conditioned on $C_n$. Finally, by basic calculations Theorem B.2(i)-(iii) are fulfilled for the same $\eta$, $\sigma^2$ and $\gamma$ given by the Lévy-Khintchine triplet of the limit X of $Z_{n,u_n}$, e.g. we have for all $x \in \mathcal{D}$

$$\sum_{i=1}^{k_n} P\left(Y_{n,i}\mathbf{1}\{Y_{n,i} \leq u_n\} > x | C_n\right) = \sum_{i=1}^{k_n} \frac{P(Y_{n,i} > x) - P(Y_{n,i} > u_n)}{P(Y_{n,i} \leq u_n)} \to \eta(x, \infty)$$

since $\min_{1 \leq i \leq k_n} P(Y_{n,i} \leq u_n) \geq 1 - \max_{1 \leq i \leq k_n} P(Y_{n,i} \geq 1) \to 1$. $\qquad\square$

### B.3. Proofs of Section 2 and Appendix A

#### B.3.1. Proof of Theorem 2.1

The statement of Theorem 2.1 follows immediately from the following lemma.

**Lemma B.4.** *Let $I_{n,1,x}$ and $I_{n,2,x}$, $x > 0$, be defined as in in (2.2) and (2.3), respectively. Let $D_n$ be defined as in (B.7). Then for all $\tau > 0$ there exists a constant $C_\tau > 0$ such that*

$$D_n \leq \left(\frac{1}{2} + \max_{1 \leq i \leq k_n} \varepsilon_{n,i}\right) I_{n,1,\tau} + I_{n,2,\tau}, \tag{B.10}$$

$$D_n \geq C_\tau \max\left\{I_{n,1,\tau}, I_{n,2,\tau} - \frac{2}{\tau} I_{n,1,\tau} \max_{1 \leq i \leq k_n} \varepsilon_{n,i}\right\}. \tag{B.11}$$

*Remark* B.5. The idea and the proof of the upper bound of $D_n$ in (B.10) is based on the argumentation of Cai et al. [5] on pp. 21f.

*Proof of Lemma B.4.* To shorten the notation, we define

$$A_{n,i,x} = \left\{ \varepsilon_{n,i} \frac{\mathrm{d}\mu_{n,i}}{\mathrm{d}P_{n,i}} > x \right\} \text{ for all } x > 0. \tag{B.12}$$

We can deduce from (B.5) that

$$D_n \leq \sum_{i=1}^{k_n} \mathbb{E}_{P_{n,i}} \left( 1 - \sqrt{1 - \varepsilon_{n,i} + \varepsilon_{n,i} \frac{\mathrm{d}\mu_{n,i}}{\mathrm{d}P_{n,i}} \mathbf{1}(A_{n,i,\tau}^c)} \right). \tag{B.13}$$

Note that $1 - \sqrt{1+t} \leq -t/2 + t^2$ for all $t \geq -1$. Applying this (pointwisely) to the integrand in (B.13) with $t = \varepsilon_{n,i}(\frac{\mathrm{d}\mu_{n,i}}{\mathrm{d}P_{n,i}} \mathbf{1}(A_{n,i,\tau}^c) - 1)$ yields (B.10).

We split the proof of (B.11) into two steps. First, define for all $x > 0$

$$\widetilde{I}_{n,2,x} = \sum_{i=1}^{k_n} \int_{A_{n,i,x}^c} \varepsilon_{n,i}^2 \left( \frac{\mathrm{d}\mu_{n,i}}{\mathrm{d}P_{n,i}} - 1 \right)^2 \mathrm{d}P_{n,i}. \tag{B.14}$$

For $\varepsilon_n^{\max} = \max_{1 \leq i \leq k_n} \varepsilon_{n,i}$ we can deduce from $\varepsilon_n^{\max} \geq (\varepsilon_n^{\max})^2$ and

$$\sum_{i=1}^{k_n} P_{n,i}(Y_{n,i} > x) \leq \sum_{i=1}^{k_n} P_{n,i}(A_{n,i,e^x-1}) \leq \frac{1}{e^x - 1} I_{n,1,e^x-1} \tag{B.15}$$

$$\text{that } -\frac{2\varepsilon_n^{\max}}{x} I_{n,1,x} \leq \widetilde{I}_{n,2,x} - I_{n,2,x} \leq 2\varepsilon_n^{\max} I_{n,1,x} \tag{B.16}$$

for all $x > 0$. Since $\mathrm{d}Q_{n,i}/\mathrm{d}P_{n,i}$ is bounded from above by $1 + \tau$ on $A_{n,i,\tau}^c$ we obtain

$$2D_n \geq \sum_{i=1}^{k_n} \int \left( 1 - \frac{\mathrm{d}Q_{n,i}}{\mathrm{d}P_{n,i}} \right)^2 \left( 1 + \left( \frac{\mathrm{d}Q_{n,i}}{\mathrm{d}P_{n,i}} \right)^{1/2} \right)^{-2} \mathbf{1}(A_{n,i,\tau}^c) \, \mathrm{d}P_{n,i}$$

$$\geq \frac{\widetilde{I}_{n,2,\tau}}{(1 + \sqrt{1+\tau})^2}.$$

Combining this and (B.16) gives us the first bound in (B.14) for appropriate $C_\tau$. Second, set $C = 1/(\sqrt{\tau/2 + 1} + 1) < 1/2$. Note that on $A_{n,i,\tau}$

$$\left( \frac{\mathrm{d}Q_{n,i}}{\mathrm{d}P_{n,i}} \right)^{1/2} - 1 = \left( \frac{\mathrm{d}Q_{n,i}}{\mathrm{d}P_{n,i}} - 1 \right) \left( \left( \frac{\mathrm{d}Q_{n,i}}{\mathrm{d}P_{n,i}} \right)^{1/2} + 1 \right)^{-1} \leq C \left( \frac{\mathrm{d}Q_{n,i}}{\mathrm{d}P_{n,i}} - 1 \right).$$

Consequently,

$$2 D_n \geq \sum_{i=1}^{k_n} E_{P_{n,i}} \left( \left( \frac{\mathrm{d}Q_{n,i}}{\mathrm{d}P_{n,i}} - 1 - 2 \left( \left( \frac{\mathrm{d}Q_{n,i}}{\mathrm{d}P_{n,i}} \right)^{1/2} - 1 \right) \right) \mathbf{1}(A_{n,i,\tau}) \right)$$

$$\geq (1 - 2C) \left( 1 - \frac{\max_{1 \leq i \leq k_n} \varepsilon_{n,i}}{\tau} \right) \sum_{i=1}^{k_n} \varepsilon_{n,i} \mu_{n,i} (A_{n,i,\tau}).$$

$\square$

*B.3.2. Proof of Theorem 2.2(b)*

The statements follows from Remark (8.6) and Lemma (8.7) of Janssen et al. [24] as we explain in the following. Let $C^2_{lok}(\mathbb{R})$ be set of all bounded functions $f : \mathbb{R} \to \mathbb{R}$ that are twice differentiable with continuous derivatives in some neighbourhood of 0. Denote by $f^{(k)}(0)$ the $k$th derivative of $f$ at 0. The Lévy-Khintchine triplet of a infinitely divisible measure $\nu$ is equal to $(\gamma, \sigma^2, \eta)$ if and only if the *generating functional* $A : C^2_{lok}(\mathbb{R}) \to \mathbb{R}$ admits the Lévy-Khintchine representation

$$A(f) = f^{(1)}(0)\gamma + \sigma^2 f^{(2)}(0) + \int_{\mathbb{R}\setminus\{0\}} \left( f(x) - f(0) - \frac{f^{(1)}(0)x}{1+x^2} \right) \mathrm{d}\eta(x)$$

for all $f \in C^2_{lok}(\mathbb{R})$. For the actual definition of $A$ and more details about it we refer the reader to Janssen et al. [24], in particular to (8.1)-(8.4).

**Lemma B.6.** *Let $\{\widetilde{\nu}_1, \widetilde{\nu}_2\}$ be some binary experiment in its standard form, compare to* (B.1) *and* (B.2), *such that $\widetilde{\nu}_1(\mathbb{R}) = \widetilde{\nu}_2(\mathbb{R}) = 1$ and $\widetilde{\nu}_1$ is infinitely divisible with Lévy-Khintchine triplet $(\gamma, \sigma^2, \eta)$. Then $\widetilde{\nu}_2$ is also infinitely divisible with Lévy-Khintchine triplet $(\gamma_2, \sigma_2^2, \eta_2)$, where $\sigma_1^2 = \sigma_2^2$, $\eta_2 \ll \eta_1$ with Radon-Nikodym derivative $x \mapsto \mathrm{d}\eta_2/\mathrm{d}\eta_1(x) = e^x$ and*

$$\gamma_1 + \frac{\sigma_1^2}{2} - \int \left( 1 - e^x + \frac{x}{x^2+1} \right) \mathrm{d}\eta_1(x) = 0, \tag{B.17}$$

$$\gamma_2 = \gamma_1 + \sigma_1^2 + \int (e^x - 1)\frac{x}{1+x^2} \, \mathrm{d}\eta_1(x). \tag{B.18}$$

*Remark* B.7. Since $\int x^2 \mathbf{1}(|x| \leq 1) \, \mathrm{d}\eta_1(x), \int e^x \mathbf{1}(|x| \geq 1) \, \mathrm{d}\eta_1(x) < \infty$, see Lemma (8.7)(a) of Janssen et al. [24], the integrals in (B.17) and (B.18) are finite.

*Proof of Lemma B.6.* Let $A$ be the generating functional of $\widetilde{\nu}_1$. Combining $\int \exp \mathrm{d}\widetilde{\nu}_1 = \widetilde{\nu}_2(\mathbb{R}) = 1$ and Lemma (8.7)(b) and (c) from Janssen et al. [24] we deduce that $A(\exp) = 0$ and $C^2_{lok}(\mathbb{R}) \ni f \mapsto A(\exp f)$ is the generating functional of $\widetilde{\nu}_2$ and, in particular, $\widetilde{\nu}_2$ is infinitely divisible. Using the Lévy-Khintchine representation of $A$ immediately yields that $A(\exp)$ is equal to the left side of (B.17), which proves (B.17). From $f(0)A(\exp) = 0$ we get for all $f \in C^2_{lok}(\mathbb{R})$

$$A(f \exp) = f^{(1)}(0)\left(\gamma_1 + \sigma_1^2 + \int (e^x - 1)\frac{x}{1+x^2} \, \mathrm{d}\eta_1(x)\right)$$
$$+ f^{(2)}(0)\frac{\sigma_1^2}{2} + \int \left( f(x) - f(0) - \frac{f^{(1)}(0)x}{1+x^2} \right) e^x \, \mathrm{d}\eta_1(x).$$

Consequently, the statements about $(\gamma_2, \sigma_2^2, \eta_2)$ follow. $\square$

Now, we prove Theorem 2.2(b). Since $\frac{\mathrm{d}Q_{n,i}}{\mathrm{d}P_{n,i}} \geq 1 - \max_{1 \leq i \leq k_n} \varepsilon_{n,i} \to 1$ (B.9) is fulfilled and by Theorem B.2 $\eta_1$ is concentrated on $(0, \infty)$. Now, consider

$\{\widetilde{\nu}_1, \widetilde{\nu}_2\} = \{\nu_1 * \epsilon_{-\log(a)}, (a^{-1}\nu_{2|\mathbb{R}}) * \epsilon_{-\log(a)}\}$. This binary experiment is in its standard from since

$$\frac{\mathrm{d}\widetilde{\nu}_2}{\mathrm{d}\widetilde{\nu}_1}(x) = a^{-1}\frac{\mathrm{d}\nu_2}{\mathrm{d}\nu_1}(x + \log(a)) = \exp(x),\ x \in \mathbb{R}.$$

Clearly, $\widetilde{\nu}_1$ is infinitely divisible with Lévy characteristic $(\gamma_1 - \log(a), \sigma_1^2, \eta_1)$ and $\widetilde{\nu}_1(\mathbb{R}) = \widetilde{\nu}_2(\mathbb{R}) = 1$. Applying Lemma B.6 proves that $\widetilde{\nu}_2$ is infinitely divisible and so is $\rho = a^{-1}\nu_{2|\mathbb{R}}$. Moreover, is easy to check that we obtain all statements about the Lévy-Khintchine triplets.

### B.3.3. Proof of Theorem 2.4

We carried out two different proofs for Theorem 2.4. The first one relies on infinitely divisible statistical experiments and accompanying Poisson experiments, and arguments from Chap. 4, 5, 9, 10 of Janssen et al. [24] are used. The second one is based on traditional limit theorems for real-valued random variables. Since, probably, the second one is easier to follow for the readers who are not experts in the field of statistical experiments we decided to present only the second proof.

At the end of the proof we will verify the following lemma.

**Lemma B.8.** *Suppose that* (a) *and* (b) *hold. Then the sums in Theorem B.2* (ii) *and* (iii) *and in Theorem B.3*(a) *and* (b) *for* $Y_{n,i}$ *defined by*

$$Y_{n,i} = \log\frac{\mathrm{d}Q_{n,i}}{\mathrm{d}P_{n,i}} \tag{B.19}$$

*are upper bounded for every* $x > 0$ *and all sufficiently small* $\tau_0, \tau_1 \in \mathcal{D}$, *respectively, under* $P_{(n)}$ *as well as under* $Q_{(n)}$. *In particular, Theorem B.2*(ii) *is fulfilled for* $\sigma^2$ *under* $P_{(n)}$.

Let us first assume that (a) and (b) are fulfilled. Define $Y_{n,i}$ as in (B.19). Regarding Lemma B.8 and using typical sub-subsequence arguments we can assume without loss of generality that Theorem B.2(i) and (ii) as well as Theorem B.3(a) and (b) hold for a measure $M_1$ (resp. $M_2$), $\sigma_1 \geq 0$ ($\sigma_2 \geq 0$, resp.) and $\gamma_1 \in \mathbb{R}$ ($\gamma_2 \in \mathbb{R}$, resp.) under $P_{(n)}$ ($Q_{(n)}$, resp.). In particular, by Lemma B.8 $\sigma_1^2 = \sigma^2$. Note that $\eta_j = M_{j|(0,\infty)}$ is a Lévy measure. From (B.15) we obtain $M_1(\{\infty\}) = 0$ and so $\xi_1$, the limit of $T_n$ under $P_{(n)}$, is real-valued. Moreover, since $\max_{1 \leq i \leq k_n} \varepsilon_{n,i} \to 0$ and $\varepsilon_{n,i}\mu_{n,i}(A_{n,i,e^x-1+\varepsilon_{n,i}}) = Q_{n,i}(Y_{n,i} > x) - (1 - \varepsilon_{n,i})P_{n,i}(Y_{n,i} > x)$ we can deduce that $M_{|(0,\infty)} = \eta_2 - \eta_1$ and $M_2(\{\infty\}) = M(\{\infty\})$. Finally, the proof for the first assertion is completed by Theorem 2.2(b).

Now, let $\xi_1$ be not equal to $-\infty$ with probability one. By Theorem 2.1(a) we have $\sup_{n \in \mathbb{N}} I_{n,1,\tau} + I_{n,2,\tau} < \infty$ for all $\tau > 0$. Hence, for each subsequence there is a subsequence such that (a) for some measure $M$ and (b) for some $\sigma^2$ are fulfilled. From Theorem 2.2(b) and the first assertion proved above we obtain: $\xi_1$ is real-valued, and $M$ and $\sigma^2$ are uniquely determined by the distribution of

$\xi_1$ and so do not depend on the special choice of the subsequence, which proves the second assertion (and Theorem 2.2(a)).

*Proof of Lemma B.8.* First, observe that by (B.15) the sum in Theorem B.3(a) is upper bounded under $P_{(n)}$ as well as under $Q_{(n)}$ for all $\tau > 0$. By (1.3)

$$B_{n,i,\tau} = \{|Y_{n,i}| \le \tau\} = A^c_{n,i,t_{n,i}(\tau)} \tag{B.20}$$

if $n \ge N_\tau$ is sufficiently large, where $t_{n,i}(\tau) = e^\tau - 1 + \varepsilon_{n,i} \in [e^\tau - 1, e^\tau]$. Define $\widetilde{I}_{n,2,x}$ as in (B.14). By Taylor's formula there exists some random variable $R_{n,i,\tau}$ with $R_{n,i,\tau} = 0$ on $B^c_{n,i,\tau}$ such that we have on $B_{n,i,\tau}$

$$Y_{n,i} = \varepsilon_{n,i}\Big(\frac{\mathrm{d}\mu_{n,i}}{\mathrm{d}P_{n,i}} - 1\Big) - \varepsilon^2_{n,i}\Big(\frac{\mathrm{d}\mu_{n,i}}{\mathrm{d}P_{n,i}} - 1\Big)^2\Big(\frac{1}{2} + R_{n,i,\tau}\Big) \tag{B.21}$$

and $\max_{1\le i\le k_n} |R_{n,i,\tau}| \le C_\tau$ for some constant $C_\tau \in (0,\infty)$ with $C_\tau \to 0$ as $\tau \searrow 0$. Combining this and (B.15) yields

$$\Big|\sum_{i=1}^{k_n} \int_{B_{n,i,\tau}} Y_{n,i}\, \mathrm{d}P_{n,i}\Big| \le \Big(1 + \frac{1}{e^\tau - 1}\Big)I_{n,1,e^\tau-1} + \Big(\frac{1}{2} + C_\tau\Big)\widetilde{I}_{n,2,e^\tau},$$

where by (B.16) the upper bound is bounded itself for all sufficiently small $\tau > 0$. Since $Q_{n,i} = (1 - \varepsilon_{n,i})P_{n,i} + \varepsilon_{n,i}\mu_{n,i}$ and $\frac{\mathrm{d}\mu_{n,i}}{\mathrm{d}P_{n,i}} \le e^\tau$ on $B_{n,i,\tau}$ we obtain similarly the following upper bound of $|\sum_{i=1}^{k_n} \int_{B_{n,i,\tau}} Y_{n,i}\, \mathrm{d}Q_{n,i}|$:

$$\Big|\sum_{i=1}^{k_n} \int_{B_{n,i,\tau}} Y_{n,i}\, \mathrm{d}P_{n,i}\Big| + I_{n,1,e^\tau-1} + \Big(1 + \Big(\frac{1}{2} + C_\tau\Big)e^\tau\Big)\widetilde{I}_{n,2,e^\tau},$$

which itself is bounded for all small $\tau > 0$, see also (B.16). In the last step we discuss the sum in Theorem B.2(ii). On $B_{n,i,\tau}$ we obtain the following inequalities from (B.21) for all sufficiently small $\tau > 0$ such that $C_\tau \le \frac{1}{2}$:

$$\varepsilon_{n,i}\Big|\frac{\mathrm{d}\mu_{n,i}}{\mathrm{d}P_{n,i}} - 1\Big|(2 - e^\tau - 2\varepsilon_{n,i}) \le |Y_{n,i}| \le \varepsilon_{n,i}\Big|\frac{\mathrm{d}\mu_{n,i}}{\mathrm{d}P_{n,i}} - 1\Big|(e^\tau + 2\varepsilon_{n,i}).$$

From this, (B.16) and $\frac{\mathrm{d}Q_{n,i}}{\mathrm{d}P_{n,i}} \le e^\tau + \max_{1\le i\le k_n} \varepsilon_{n,i}$ on $B_{n,i,\tau}$ we conclude

$$\lim_{\tau\searrow 0} \frac{\limsup}{\liminf}_{n\to\infty} \sum_{i=1}^{k_n} \int_{B_{n,i,\tau}} Y^2_{n,i}\, \mathrm{d}Q_{n,i} \le \lim_{\tau\searrow 0} \frac{\limsup}{\liminf}_{n\to\infty} \sum_{i=1}^{k_n} \int_{B_{n,i,\tau}} Y^2_{n,i}\, \mathrm{d}P_{n,i} = \sigma^2.$$

Since $(a+b)^2 \le 4a^2 + 4b^2$ we have for all sufficiently small $\tau > 0$ that

$$\frac{1}{4}\sum_{i=1}^{k_n}\Big(\int_{B_{n,i,\tau}} Y_{n,i}\, \mathrm{d}P_{n,i}\Big)^2$$

$$\le \sum_{i=1}^{k_n}\Big(\varepsilon_{n,i}\int_{B^c_{n,i,\tau}} 1 - \frac{\mathrm{d}\mu_{n,i}}{\mathrm{d}P_{n,i}}\, \mathrm{d}P_{n,i}\Big)^2 + \Big(\int_{B_{n,i,\tau}} \varepsilon^2_{n,i}\Big(\frac{\mathrm{d}\mu_{n,i}}{\mathrm{d}P_{n,i}} - 1\Big)^2 \mathrm{d}P_{n,i}\Big)^2$$

$$\le \Big(\max_{1\le i\le k_n} \varepsilon_{n,i}\Big)(1 + \tau)I_{n,1,e^\tau-1} + \widetilde{I}_{n,2,e^\tau}\Big(e^\tau - 1 + \max_{1\le i\le k_n} \varepsilon_{n,i}\Big)^2.$$

Hence, by (B.16) $\lim_{\tau \searrow 0} \limsup_{n \to \infty} (\int_{B_{n,i,\tau}} Y_{n,i} \, \mathrm{d}P_{n,i})^2 = 0$ and, consequently, Theorem B.2(ii) is fulfilled for $\sigma^2$ under $P_{(n)}$. $\qquad\qquad\qquad\square$

### B.3.4. Proof of Theorem 2.2(a)

We verified Theorem 2.2(a) while proving the second assertion of Theorem 2.4.

### B.3.5. Proof of Theorem A.1

The equivalence of (a)-(e) follows from (B.3) and is standard for binary experiments, see Strasser [36]. The equivalence of (g) and (h) follows from (1.1) and (1.3). Define $A_{n,i,x}$ as in (B.12).

Equivalence of (b) and (i): By Theorem 2.4 $I_{n,1,x} \to 0$ for all $x > 0$ holds also under (b). Hence, we can suppose that this convergence is fulfilled subsequently. Fix $\tau > 0$. Then

$$0 \le E_{P_{n,i}} \left( \varepsilon_{n,i}^2 \Big( \frac{\mathrm{d}\mu_{n,i}}{\mathrm{d}P_{n,i}} \Big)^2 \mathbf{1} \Big\{ \frac{\mathrm{d}\mu_{n,i}}{\mathrm{d}P_{n,i}} \in (x, \tau] \Big\} \right) \le \tau \, \varepsilon_{n,i} \mu_{n,i}(A_{n,i,x})$$

holds for all $x \in (0, \tau]$ and so $I_{n,2,x} - I_{n,2,\tau} \to 0$ does. Consequently, (i) holds if and only if Theorem 2.4(a) and (b) do so for the same $\sigma^2 \in [0, \infty)$ and $M \equiv 0$. Hence, the equivalence of (b) and (i) follows from Theorem 2.4.

Equivalence of (f) and (i): Define $Y_{n,i}$ as in (B.19) and set $\widetilde{Y}_{n,i} = f(Y_{n,i})$ for $f(x) = \exp(x) - 1$, $x \in \mathbb{R}$. Note that $f(0) = 0$ and $f'(0) = f''(0) = 1$. From this, a Taylor expansion, compare to (B.21), and Theorem B.2 we obtain that $\sum_{i=1}^{k_n} Y_{n,i}$ converges in distribution to $X$ with Lévy-Khintchine triplet $(0, \sigma^2, 0)$ if and only if $\sum_{i=1}^{k_n} \widetilde{Y}_{n,i}$ does so to $\widetilde{X}$ with Lévy-Khintchine triplet $(-\sigma^2/2, \sigma^2, 0)$.

Equivalence of (d) and (h): Throughout this proof step we can assume that $\xi_2$ is real-valued and so is $\xi_1$, see Theorem 2.2(a). By the first Lemma of Le Cam $P_{(n)}$ and $Q_{(n)}$ are mutually contiguous, see also Remark 2.3. Hence, (h) is true if and only if for all $x > 0$

$$0 \gets Q_{(n)} \Big( \max_{1 \le i \le k_n} \varepsilon_{n,i} \frac{\mathrm{d}\mu_{n,i}}{\mathrm{d}P_{n,i}} > x \Big) = 1 - \prod_{i=1}^{k_n} \Big( 1 - Q_{n,i}(A_{n,i,x}) \Big).$$

Combining this and (B.15) yields that (h) is fulfilled if and only if $I_{n,1,x} \to 0$ for all $x > 0$. Finally, note that $\xi_1$ is normal distributed if and only if it has trivial Lévy measure $\eta_1 \equiv 0$, which by Theorem 2.4 is true if and only if $I_{n,1,x} \to 0$ for all $x > 0$.

### B.3.6. Proof of Lemma A.3

Let $\widetilde{Q}_{n,i}$ and $\widetilde{Q}_{(n)}$ be defined as $Q_{n,i}$ and $Q_{(n)}$ replacing $\mu_{n,i}$ and $\varepsilon_{n,i}$ by $\widetilde{\mu}_{n,i}$ and $\widetilde{\varepsilon}_{n,i}$. For the statement in Lemma A.3 it is sufficient to show that $\{Q_{(n)}, \widetilde{Q}_{n,i}\}$ tend weakly to the uninformative experiment $\{\epsilon_0, \epsilon_0\}$. The main task for this purpose is to verify $\sum_{i=1}^{k_n} ||Q_{n,i}(\boldsymbol{\theta}) - Q_{n,i}(\widetilde{\boldsymbol{\theta}})|| \to 0$, which is left to the reader.

### B.3.7. Proof of Theorem 2.6

Denote by $Z_n(\boldsymbol{\theta_1})$ and $Z_n(\boldsymbol{\theta_2})$ the statistic introduced in (A.1) for the model $\boldsymbol{\theta_1}$ and $\boldsymbol{\theta_2}$, respectively. Since these statistics are linear, the multivariate central limit theorem implies distributional convergence $(Z_n(\boldsymbol{\theta_1}), Z_n(\boldsymbol{\theta_2})) \xrightarrow{\mathrm{d}} \widetilde{Z} \sim N((0,0), (\gamma(\boldsymbol{\theta_i}, \boldsymbol{\theta_j}))_{1 \le i,j \le 2})$ under $P_{(n)}$. In the next step we verify for $j = 1, 2$

$$T_n(\boldsymbol{\theta_j}) = Z_n(\boldsymbol{\theta_j}) - \frac{\gamma(\boldsymbol{\theta_j}, \boldsymbol{\theta_j})}{2} + R_{n,j}, \tag{B.22}$$

where $R_{n,j}$ converges in $P_{(n)}$-probability to 0. Let $j \in \{1, 2\}$ be fixed. Define $Y_{n,i}^{(j)} = \varepsilon_{n,i}^{(j)}(\mathrm{d}\mu_{n,i}^{(j)}/\mathrm{d}P_{n,i}^{(j)} - 1)$. Note that by Taylor's Theorem $\log(1 + x) = x - \frac{x^2}{2} + (2/3)x^3(1 + y_x)^{-3}$ for $|y_x| \le x$. Since $\max_{1 \le i \le k_n} Y_{n,i}^{(j)} \to 0$ in $P_{(n)}$-probability, see Theorem A.1, it remains to shown that $\sum_{i=1}^{k_n} (Y_{n,i}^{(j)})^2 \to \gamma(\boldsymbol{\theta_j}, \boldsymbol{\theta_j})$ in $P_{(n)}$-probability. It is well know that this follows immediately if the Lindeberg condition is fulfilled for the triangular array $(Y_{n,i}^{(j)})_{i \le k_n}$ under $P_{(n)}$. Observe that combining Theorem A.1(f) and the assumption $\gamma(\boldsymbol{\theta_j}, \boldsymbol{\theta_j}) = \sigma_j^2$ yields the desired Lindeberg condition and, finally, (B.22).

From (B.22) and the asymptotic normality of the vector $(Z_n(\boldsymbol{\theta_1}), Z_n(\boldsymbol{\theta_2}))$ we obtain $(T_n(\boldsymbol{\theta_1}), T_n(\boldsymbol{\theta_2})) \xrightarrow{\mathrm{d}} \widehat{Z} \sim N((-\gamma(\boldsymbol{\theta_j}, \boldsymbol{\theta_j})/2)_{j=1,2}, (\gamma(\boldsymbol{\theta_i}, \boldsymbol{\theta_j}))_{1 \le i,j \le 2})$. Consequently, by the third lemma of Le Cam we get under $Q_{(n)}(\boldsymbol{\theta_1})$

$$(T_n(\boldsymbol{\theta_1}), T_n(\boldsymbol{\theta_2})) \xrightarrow{\mathrm{d}} \widehat{Z} \sim N\left(\left(\frac{-\gamma(\boldsymbol{\theta_j}, \boldsymbol{\theta_j})}{2} + \gamma(\boldsymbol{\theta_1}, \boldsymbol{\theta_j})\right)_{j=1,2}, (\gamma(\boldsymbol{\theta_i}, \boldsymbol{\theta_j}))_{1 \le i,j \le 2}\right).$$

Finally, the desired statement can be concluded.

### B.3.8. Proof of Corollary 2.8

Define

$$\varepsilon_{n,i}^* = \frac{\varepsilon_{n,i}(1 - \lambda_{n,i})}{1 - \varepsilon_{n,i}\lambda_{n,i}}.$$

Now, let $Q_{n,i}^*$, $Q_{(n)}^*$ and $T_n^*$ defined as $\widetilde{Q}_{n,i}$, $\widetilde{Q}_{(n)}$ and $\widetilde{T}_n$ replacing $\widetilde{\varepsilon}_{n,i}$ by $\varepsilon_{n,i}^*$. Since $\widetilde{\varepsilon}_{n,i} = \varepsilon_{n,i}^*(1 + a_{n,i})$ with $\max_{1 \le i \le k_n} |a_{n,i}| \to 0$ it can easily be seen by Theorems 2.2 and 2.4 that (2.1) also holds for $T_n^*$ with the same limits $\widetilde{\xi}_1$ and $\widetilde{\xi}_2$. Note that

$$\frac{\mathrm{d}Q_{n,i}}{\mathrm{d}P_{n,i}} = (1 - \varepsilon_{n,i}\lambda_{n,i})\frac{\mathrm{d}Q_{n,i}^*}{\mathrm{d}P_{n,i}} + \infty \, \mathbf{1}_{N_{n,i}}\mathbf{1}\{\lambda_{n,i} > 0\}.$$

Since $\bigotimes_{i=1}^{k_n} N_{n,i}$ is a $P_{(n)}$-null set we obtain that $P_{(n)}$-almost surely

$$\log\left(\frac{\mathrm{d}Q_{(n)}}{\mathrm{d}P_{(n)}}\right) = \log\left(\frac{\mathrm{d}Q_{(n)}^*}{\mathrm{d}P_{(n)}}\right) + \sum_{i=1}^{k_n} \log(1 - \lambda_{n,i}\varepsilon_{n,i}).$$

Combining this and $\sum_{i=1}^{k_n} \log(1 - \lambda_{n,i}\varepsilon_{n,i}) \to -c$ yields that $T_n \xrightarrow{d} \widetilde{\xi}_1 - c = \xi_1$ under $P_{(n)}$. By Section B.1 we obtain that $T_n$ converges in distribution to some $\xi_2$ under $Q_{(n)}$ and by (B.3) we get the desired representation $\nu_2 = e^{-c}\widetilde{\nu}_2 * \epsilon_{-c} + (1 - e^{-c})\epsilon_\infty$ of $\xi_2$'s distribution.

### B.4. Proofs of Section 3

To shorten the notation we define

$$Z_n(t) = \sqrt{n}\,\frac{\mathbb{F}_n(t) - t}{\sqrt{t(1-t)}},\ t \in (0,1).$$

Then,

$$\mathrm{HC}_n = \sup_{t \in (0,1)} |Z_n(t)|.$$

#### B.4.1. Proof of Theorem 3.1

First, note that

$$a_n HC_n - b_n = \sqrt{2}\,\log\log(k_n)\left(\frac{\mathrm{HC}_n}{\sqrt{\log\log(k_n)}} - \sqrt{2} + o(1)\right).$$

That is why it sufficient to show that for some $\gamma > 0$

$$Q_{(n)}\Big(\frac{|Z_n(v_n)|}{\sqrt{\log\log k_n}} \leq \sqrt{2} + \gamma\Big) \to 0 \tag{B.23}$$

$$\text{or } Q_{(n)}\Big(\frac{|Z_n(1 - v_n)|}{\sqrt{\log\log k_n}} \leq \sqrt{2} + \gamma\Big) \to 0. \tag{B.24}$$

To verify this we apply Chebyshev's inequality. Note that for every real-valued random variable $Z$ on some probability space $(\Omega, \mathcal{A}, P)$ with finite expectation we have

$$P\left(|Z| \leq \frac{|E(Z)|}{2}\right) = P\left(|Z - E(Z)| \geq \frac{|E(Z)|}{2}\right) \leq 4\frac{\mathrm{Var}_P(Z)}{E_P(Z)^2}. \tag{B.25}$$

Consequently, we need to determine first the expectation and variance for $Z_n(v)$ for $v \in \{v_n, 1 - v_n\}$:

$$E_{Q_{(n)}}(Z_n(v)) = \sqrt{k_n}\,\frac{k_n^{-1}\sum_{i=1}^{k_n} Q_{n,i}(0,v] - v}{\sqrt{v(1-v)}} = \frac{\sum_{i=1}^{k_n} \varepsilon_{n,i}\,(\mu_{n,i}\,(0,v] - v)}{\sqrt{k_n\,v(1-v)}},$$

$$\mathrm{Var}_{Q_{(n)}}(Z_n(v)) = \frac{1}{k_n}\,\frac{\sum_{i=1}^{k_n} Q_{n,i}(0,v]\,(1 - Q_{n,i}(0,v])}{v(1-v)}$$

$$\leq \min\Big\{\frac{\sum_{i=1}^{k_n} Q_{n,i}(0,v]}{k_n\,v(1-v)}\,,\ \frac{\sum_{i=1}^{k_n}(1 - Q_{n,i}(0,v])}{k_n\,v(1-v)}\Big\}$$

$$= \min\Big\{\frac{1}{1-v} + \frac{E_{Q_{(n)}}\,[Z_n\,(v)]}{\sqrt{k_n\,v(1-v)}}\,,\ \frac{1}{v} - \frac{E_{Q_{(n)}}\,[Z_n\,(v)]}{\sqrt{k_n\,v(1-v)}}\Big\}.$$

By assumption we have

$$\frac{|\sum_{i=1}^{k_n} \varepsilon_{n,i}(\mu_{n,i}(0, v_n] - v_n)|}{\sqrt{k_n v_n \log \log(k_n)}} \to \infty \tag{B.26}$$

$$\text{or} \quad \frac{|\sum_{i=1}^{k_n} \varepsilon_{n,i}(\mu_{n,i}(1 - v_n, 1) - v_n)|}{\sqrt{k_n v_n \log \log k_n}} \to \infty. \tag{B.27}$$

Suppose that (B.26) holds. Then

$$\left| \frac{E_{Q_{(n)}}(Z_n(v_n))}{\sqrt{\log \log(k_n)}} \right| \to \infty \text{ and } \frac{\text{Var}_{Q_{(n)}}(Z_n(v_n))}{E_{Q_{(n)}}(Z_n(v_n))^2} \to 0.$$

Combining this and (B.25) yields that (B.23) is fulfilled for all $\gamma > 0$. Analogously, if (B.27) is true then (B.24) holds for all $\gamma > 0$.

### B.4.2. Proof of Theorem 3.2

Let $G_n$ be the distribution function of $Q_{n,1}$, i.e. $G_n(v) = Q_{n,1}([0, v])$, $v \in (0, 1)$. Let $U_1, U_2, \ldots$ be a sequence of independent, uniformly on $(0, 1)$ distributed random variables on the same probability space $(\Omega, \mathcal{A}, P)$. Note $(U_1, \ldots, U_{k_n}) \sim P_{(n)}$ and $(G_n^{-1}(U_1), \ldots, G_n^{-1}(U_{k_n})) \sim Q_{(n)}$, where $G_n^{-1}$ denotes the left continuous quantile function of $Q_{n,1}$. Moreover, denote the interval $(r_n, s_n) \cup (t_n, u_n)$ by $J_{n,1}$ and $[1 - u_n, 1 - t_n] \cup [1 - s_n, 1 - r_n]$ by $J_{n,2}$. By (3.3) it is easy to see that we can replace $r_n$ by any $r_n' \geq r_n$ such that $\log(r_n') = (-1 + o(1)) \log(n)$. In particular, we can assume without loss of generality that $k_n r_n \geq 1$ and, analogously, $u_n < 1/2$. From Corollaries 2 and 3 as well as (1) and (2) of Theorem of Jaeschke [22], which also hold for the statistics $W_n, \widehat{V}_n, \widehat{W}_n$ introduced at the beginning of subsection 2 therein, we can deduce that

$$a_n \sup_{v \in (0,1) \setminus (J_{n,1} \cup J_{n,2})} \left\{ \left| \frac{\sum_{i=1}^{k_n} (\mathbf{1}\{U_i \leq v\} - v)}{\sqrt{k_n v(1 - v)}} \right| \right\} - b_n \xrightarrow{\text{P}} -\infty \tag{B.28}$$

$$\text{and } a_n \sup_{v \in (0,1)} \left\{ \left| \frac{\sum_{i=1}^{k_n} (\mathbf{1}\{U_i \leq v\} - v)}{\sqrt{k_n v(1 - v)}} \right| \right\} - b_n \xrightarrow{\text{d}} Y, \tag{B.29}$$

where the distribution function of $Y$ equals $\Lambda^2$, see (3.1). By (B.28), the mutually contiguity of $P_{(n)}$ and $Q_{(n)}$ and the equivalence "$G_n(v) \geq u \Leftrightarrow v \geq G_n^{-1}(u)$" it is sufficient for (3.5) to verify

$$a_n \sup_{v \in J_{n,1} \cup J_{n,2}} \left\{ \frac{\sum_{i=1}^{k_n} (\mathbf{1}\{U_i \leq G_n(v)\} - v)}{\sqrt{k_n v(1 - v)}} \right\} - b_n \xrightarrow{\text{d}} Y. \tag{B.30}$$

For this purpose we define

$$\Delta_{n,1}(v) = \frac{\sum_{i=1}^{k_n} \left( \mathbf{1}\{U_i \le G_n(v)\} - G_n(v) \right)}{\sqrt{n\, G_n(v)(1 - G_n(v))}},$$

$$\Delta_{n,2}(v) = \sqrt{\frac{G_n(v)}{v}}, \ \Delta_{n,3}(v) = \sqrt{\frac{1 - G_n(v)}{(1 - v)}}, \ \Delta_{n,4}(v) = \sqrt{k_n}\, \frac{G_n(v) - v}{\sqrt{v(1 - v)}}.$$

Clearly,

$$\frac{\sum_{i=1}^{k_n} \left( \mathbf{1}\{U_i \le G_n(v)\} - v \right)}{\sqrt{k_n\, v(1 - v)}} = \Delta_{n,1}(v)\Delta_{n,2}(v)\Delta_{n,3}(v) + \Delta_{n,4}(v).$$

Hence, the proof of (B.30) falls naturally into the following steps:

$$\sup_{v \in J_{n,1} \cup J_{n,2}} |\Delta_{n,j}(v) - 1| \to 0 \text{ for } j \in \{2, 3\}, \tag{B.31}$$

$$a_n \sup_{v \in J_{n,1} \cup J_{n,2}} |\Delta_{n,4}(v)| \to 0, \tag{B.32}$$

$$a_n \sup_{v \in J_{n,1} \cup J_{n,2}} \{|\Delta_{n,1}(v)|\} - b_n \xrightarrow{\text{d}} Y. \tag{B.33}$$

First, observe that $(1 - \varepsilon_n)v \le G_n(v) \le v + \varepsilon_n(1 - v)$ for all $v \in (0, 1)$. Hence, we have for all $v_1 \in (0, 1/2]$ and $v_2 \in [1/2, 1)$ that

$$\frac{1 - G_n(v_1)}{1 - v_1}, \ \frac{G_n(v_2)}{v_2} \in (1 - \varepsilon_n, 1 + \varepsilon_n).$$

Moreover, we have for all $v_1 \in J_{n,1}$ and all $v_2 \in J_{n,2}$ that

$$\left| \frac{G_n(v_1)}{v_1} - 1 \right| = \frac{\varepsilon_n |\mu_n(0, v_1] - v_1|}{v_1} \le \frac{H_n(v_1)}{\sqrt{k_n r_n}} \le a_n H_n(v_1), \tag{B.34}$$

$$\left| \frac{1 - G_n(v_2)}{1 - v_2} - 1 \right| = \frac{\varepsilon_n |\mu_n(v_2, 1) - (1 - v_2)|}{1 - v_2} \le a_n H_n(1 - v_2). \tag{B.35}$$

Consequently, (B.31) follows. Similarly to the above, we obtain

$$|\Delta_{n,4}(v_1)| \le \frac{H_n(v_1)}{\sqrt{1 - u_n}} \le \frac{1}{\sqrt{2}} H_n(v_1) \text{ and } |\Delta_{n,4}(v_2)| \le \frac{1}{\sqrt{2}} H_n(1 - v_2).$$

for all $v_1 \in J_{n,1}$ and $v_2 \in J_{n,2}$. From this we obtain (B.32). Clearly,

$$\sup_{v \in J_{n,1} \cup J_{n,2}} |\Delta_{n,1}(v)| = \sup_{v \in \widetilde{J}_{n,1} \cup \widetilde{J}_{n,2}} \left| \frac{\sum_{i=1}^{k_n} (\mathbf{1}\{U_i \le v\} - v)}{\sqrt{k_n v(1 - v)}} \right|,$$

where $\widetilde{J}_{n,1} = [\widetilde{r}_n, \widetilde{s}_n] \cup [\widetilde{t}_n, \widetilde{u}_n]$ by $\widehat{J}_{n,2} = (1 - \widehat{u}_n, 1 - \widehat{t}_n) \cup (1 - \widehat{s}_n, 1 - \widehat{r}_n)$ with $\widetilde{r}_n = G_n(r_n)$, $\widetilde{s}_n = G_n(s_n)$, $\widetilde{t}_n = G_n(t_n)$, $\widetilde{u}_n = G_n(u_n)$, $\widehat{r}_n = 1 - G_n(1 - r_n)$, $\widehat{s}_n = 1 - G_n(1 - s_n)$, $\widehat{t}_n = 1 - G_n(1 - t_n)$ and $\widehat{u}_n = 1 - G_n(1 - u_n)$. From (B.34), (B.35) and (3.3) we deduce that $(\widetilde{r}_n, \widetilde{s}_n, \widetilde{t}_n, \widetilde{u}_n)$ and $(\widehat{r}_n, \widehat{s}_n, \widehat{t}_n, \widehat{u}_n)$ fulfil (3.4). Finally, (B.33) follows from (B.28) and (B.29) (with the new parameters).

## B.5. Proofs of Section 4.1

Before we prove the theorems stated in Section 4.1 we want to point out the following: We can always assume that there is no perturbation, i.e. $r_{n,i} = 0$ for all $i, n$, see Lemma B.9. Note that we will assume this in all upcoming proofs concerning Section 4.1 without recalling it every time.

**Lemma B.9** (Perturbation). *Let us consider the situation in Section 4.1. Let $\mu_{n,i}^*$, $Q_{n,i}^*$ and $Q_{(n)}^*$ be defined as $\mu_{n,i}$, $Q_{n,i}$ and $Q_{(n)}$ setting $r_{n,i} = 0$ for all $i, n$. Then* (4.2) *is a sufficient that $\{Q_{(n)}, Q_{(n)}^*\}$ converges weakly to the uninformative experiment $\{\epsilon_0, \epsilon_0\}$. In other words, if* (4.2) *is fulfilled then the perturbation by $(r_{n,i})_{i \leq k_n}$ does not affect the asymptotic results.*

*Proof.* It is sufficient to show that $\sum_{i=1}^{k_n} d^2(Q_{n,i}, Q_{n,i}(\widetilde{\boldsymbol{\theta}})) \to 0$ under (4.2). This convergence follows immediately from the first representation of the Hellinger distance in (B.5) and the third binomial formula:

$$
\begin{aligned}
&2d^2(Q_{n,i}, Q_{n,i}(\widetilde{\boldsymbol{\theta}})) \\
&= \int_0^1 \frac{(\varepsilon_{n,i} r_{n,i})^2}{(\sqrt{(1 - \varepsilon_{n,i}) + \varepsilon_{n,i} h_{n,i}} + \sqrt{(1 - \varepsilon_{n,i}) + \varepsilon_{n,i} h_{n,i} + \varepsilon_{n,i} r_{n,i}})^2} \, \mathrm{d}\lambda \\
&\leq \frac{\varepsilon_{n,i}^2}{1 - \varepsilon_{n,i}} \int_0^1 r_{n,i}^2 \, \mathrm{d}\lambda. \qquad \qquad \qquad \qquad \qquad \qquad \qquad \qquad \square
\end{aligned}
$$

### B.5.1. Proof of Theorem 4.1

First, observe that

$$
I_{n,1,x} = \sum_{i=1}^{k_n} \varepsilon_{n,i} \int_0^1 h_{n,i} \, \mathbf{1}\left\{ \frac{\varepsilon_{n,i}}{\kappa_{n,i}} h_{n,i} > x \right\} \mathrm{d}\lambda, \tag{B.36}
$$

$$
I_{n,2,x} = \sum_{i=1}^{k_n} \left( \frac{\varepsilon_{n,i}^2}{\kappa_{n,i}} \int_0^1 h_{n,i}^2 \, \mathbf{1}\left\{ \frac{\varepsilon_{n,i}}{\kappa_{n,i}} h_{n,i} \leq x \right\} \mathrm{d}\lambda \right) - \sum_{i=1}^{k_n} \varepsilon_{n,i}^2. \tag{B.37}
$$

Moreover, note that

$$
I_{n,1,x} \leq \frac{1}{x} \sum_{i=1}^{k_n} \frac{\varepsilon_{n,i}^2}{\kappa_{n,i}} \int h_{n,i}^2 \, \mathbf{1}\left\{ \frac{\varepsilon_{n,i}}{\kappa_{n,i}} h_{n,i} > x \right\} \mathrm{d}\lambda, \tag{B.38}
$$

$$
I_{n,1,x} + I_{n,2,x} \leq \max\{1, x^{-1}\} \left( \max_{1 \leq i \leq k_n} \int_0^1 h_{n,i}^2 \, \mathrm{d}\lambda \right) \sum_{i=1}^{k_n} \frac{\varepsilon_{n,i}^2}{\kappa_{n,i}}
$$

By these and Theorem 2.1 $K = 0$ corresponds to the undetectable case and no accumulation point of $\{P_{(n)}, Q_{(n)}\}$ is full informative if $K \in (0, \infty)$. By Lemma B.1(b) and (B.8) the latter is also valid if $\limsup_{n \to \infty} \sum_{i=1}^{k_n} \varepsilon_{n,i} < \infty$. Consequently, (a) and the first statement in (c) are verified. Now, let us

suppose that $K \in (0, \infty)$ and (4.5) holds. Clearly, $\sum_{i=1}^{k_n} \varepsilon_{n,i}^2 \to 0$. By (B.37) and (B.38) $I_{n,1,x} \to 0$ and $I_{n,2,x} \to K \int h^2 \, dP_0 = \sigma^2$ for all $x > 0$. Hence, applying Theorem 2.4 completes the proof of (c).

Now, let the assumptions of (b) hold. Without loss of generality we can assume that $\sum_{i=1}^{k_n} \varepsilon_{n,i}^2 \to C_1 < \infty$ and $\varepsilon_{n,r_n}/\kappa_{n,r_n} \to C \in [0, \infty]$ since otherwise we use standard sub-subsequence arguments and make use of (B.8). If $C \geq 1$ then for all sufficiently large $n \in \mathbb{N}$

$$I_{n,1,x} \geq \sum_{i=r_n}^{k_n} \varepsilon_{n,i} \int_0^1 h_{n,i} \mathbf{1}\Big\{\frac{\varepsilon_{n,r_n}}{\kappa_{n,r_n}} h_{n,i} > x\Big\} \, d\lambda$$

$$\geq \Big(\sum_{i=r_n}^{k_n} \varepsilon_{n,i}\Big) \min_{1 \leq i \leq k_n} \int_0^1 h_{n,i} \mathbf{1}\Big\{\frac{1}{2} h_{n,i} > x\Big\} \, d\lambda$$

and so by (4.3) $I_{n,1,x} \to \infty$ for all sufficiently small $x > 0$. If $C < 1$ then

$$I_{n,2,x} \geq \Big(\sum_{i=1}^{r_n} \frac{\varepsilon_{n,i}^2}{\kappa_{n,i}}\Big) \min_{1 \leq i \leq k_n} \int_0^1 h_{n,i}^2 \mathbf{1}\Big\{2h_{n,i} \leq x\Big\} \, d\lambda - C_1$$

and so by (4.3) $I_{n,2,x} \to \infty$ for all sufficiently large $x > 0$. Hence, applying Theorem 2.1 verifies (b). Finally, note that $K < \infty$ implies $\sum_{i=1}^{k_n} \varepsilon_{n,i}^2 \to 0$. Keeping this in mind the proof of (d) is trivial (and omitted to the reader).

### B.5.2. Proof of Theorem 4.3

By (B.37)

$$-\frac{1}{k_n} \leq I_{n,2,x} \leq \frac{x}{k_n} \sum_{i=1}^{k_n} \int_0^1 h_{n,i} \mathbf{1}\{k_n^{r-1} h_{n,i} \leq x\} \, d\lambda \leq x$$

and so $\lim_{x \searrow 0} \limsup_{n \to \infty} \liminf I_{n,2,x} = 0$.

Combining (B.36) and (4.6) yields for all $x \in \mathcal{D}$ that $I_{n,1,e^x-1}$ equals

$$\frac{1}{k_n} \sum_{i=1}^{k_n} \int_0^1 h_{n,i} \mathbf{1}\Big\{k_n^{r-1} h_{n,i} > e^x - 1\Big\} \, d\lambda \to \mathbf{1}\{r > 1\} + M(x, \infty)\mathbf{1}\{r = 1\}.$$

Consequently, applying Theorem 2.4 and Theorem 2.1 completes the proof.

### B.5.3. Proof of Theorem 4.5

It is easy to verify that by (B.36) and (B.37)

$$I_{n,1,x} = \min\Big\{n^{1-\beta}, n^{\frac{1}{\alpha}(\alpha-\beta+r(1-\alpha))}\Big(\frac{x}{1-\alpha}\Big)^{1-\frac{1}{\alpha}}\Big\}, \quad I_{n,2,x} \leq n^{1-2\beta+r}.$$

Note that $1 - 2\beta + r < 0$ if $r < \rho^*(\beta, \alpha)$, or if $r = \rho^*(\beta, \alpha)$ and $\alpha > 1/2$. Moreover, in the case of $\alpha = 1/2$, $r = \rho^*(\beta, \alpha) = 2\beta - 1$ we have

$$I_{n,2,x} = \frac{1}{2} \log(2xn^{1-\beta}) - n^{1-2\beta} \to 0.$$

Combining these, Theorem 2.4, Theorem 2.1 and (2.7) completes the proof.

### B.5.4. Proof of Theorem 4.6

To shorten the notation, set $\mu_n = \mu_{n,1}$, $\kappa_n = \kappa_{n,i}$ and $\varepsilon_n = \varepsilon_{n,i}$. Since the support of $\mu_n$ is $(0, \kappa_n)$ with $\kappa_n \to 0$ and, clearly, $a_n k_n \varepsilon_n^2 = a_n k_n^{1-2\beta} \to 0$ we deduce from Remark 3.3 that we can replace $H_n(v)$ in Theorems 3.1 and 3.2 by

$$\widehat{H}_n(v) = k_n^{\frac{1}{2}-\beta} v^{-\frac{1}{2}} \mu_n(0, v) = k_n^{\frac{1}{2}-\beta} v^{-\frac{1}{2}} \int_0^{\min\{vk_n^r, 1\}} h \, d\lambda.$$

We give the proof for the model (i) and the one from Theorem 4.3 in the case of $r = 1$. The model (ii) is much simpler and left to the reader.

First, consider $\beta = r = 1$. Let $r_n = k_n^{-1} a_n^3$, $s_n = t_n$ and $u_n = (\log k_n)^{-1}$. Clearly, (3.4) holds. Moreover,

$$a_n \sup\{\widehat{H}_n(v) : v \in [r_n, u_n]\} \leq a_n k_n^{\frac{1}{2}-\beta} r_n^{-\frac{1}{2}} \to 0.$$

Hence, by Theorem 3.2 the HC test has no power asymptotically.

Now, consider the model from Section 1.2 with $h \in L^{2+\delta}(P_0)$ for some $\delta \in (0, 1)$. In particular, we have $k_n = n$. First, let $r > \rho(\beta) = 1 - 2\beta$ and $\beta < 1$. Set $v_n = n^{-\min\{1, r\}}$. Clearly, $n^r v_n \geq 1$ and

$$a_n^{-1} \widehat{H}_n(v_n) = a_n^{-1} n^{1/2-\beta+\min\{1,r\}/2} \to \infty.$$

By this, Theorems 3.1 and 4.1 the areas of complete detection $(r > \rho(\beta))$ coincide for the HC and the LLR test. It remains to discuss $r = \rho(\beta) = 2\beta - 1$ and $\beta < 1$. Set $r_n = n^{-1}$, $s_n = n^{-r} a_n^{-4(1+2/\delta)}$, $t_n = n^{-r} a_n^4$ and $u_n = (\log n)^{-1}$. Clearly, (3.4) holds. By Hölder's inequality there is some $c_0 > 0$ such that

$$\mu_n(0, v] \leq \left(\int_0^1 h^{2+\delta} \, d\lambda\right)^{1/(2+\delta)} \left(\int_0^{vn^r} d\lambda\right)^{1-1/(2+\delta)} \leq c_0 \, (vn^r)^{1-1/(2+\delta)}$$

for all $v \in (0, 1)$. Hence, we obtain

$$a_n \sqrt{n} \varepsilon_n \sup_{v \in [r_n, s_n]} \left\{\frac{\mu_n(0, v]}{\sqrt{v}}\right\} \leq a_n \, n^{1/2-\beta} c_0 s_n^{1/2-1/(2+\delta)} n^{r-r/(2+\delta)} \leq c_0 \, a_n^{-1} \to 0.$$

Moreover,

$$a_n \sqrt{n} \varepsilon_n \sup_{v \in [t_n, u_n]} \left\{\frac{\mu_n(0, v]}{\sqrt{v}}\right\} \leq a_n n^{1/2-\beta} t_n^{-1/2} = a_n^{-1} \to 0.$$

Finally, by Theorem 3.2 the HC test has no power asymptotically.

### B.6. Proofs for Section 4.2

*B.6.1. Proof of Theorem 4.9*

First, remind that we apply the HC statistic to $p_{n,i} = 1 - \Phi(Y_{n,i})$. Hence, without loss of generality we can write $\mu_n = N(\vartheta_n, \sigma_0^2)^{1-\Phi}$. Note that

$$\mu_n(0, v] = 1 - \Phi\Big(-\frac{\Phi^{-1}(v) + \vartheta_n}{\sigma_0}\Big), \ v \in (0, 1). \tag{B.39}$$

Moreover, we have for all $v \in (0, 1/2)$

$$\mu_n(1 - v, 1] = 1 - \Phi\Big(\frac{-\Phi^{-1}(v) + \vartheta_n}{\sigma_0}\Big) \leq \mu_n(0, v]. \tag{B.40}$$

Observe that by Remark 2.3 and Proposition 4.7 $P_{(n)}$ and $Q_{(n)}$ are mutually contiguous. Clearly, this is not affected by the transformation to $p$-values. Consequently, by (B.40), Theorem 3.2 and Remark 3.3 it is sufficient to show that

$$a_n \sqrt{n}\varepsilon_n \sup_{v \in (n^{-1+\lambda_n}, 1/2]} \frac{\mu_n(0, v]}{\sqrt{v}} \to 0 \text{ with } \lambda_n = \frac{(\log\log(n))^2}{\log(n)},$$

i.e. $r_n = n^{-1+\lambda_n}$, $s_n = t_n$ and $u_n = 1/2$. Let $\delta > 0$ be sufficiently small that $2\delta < 1 - r$ and $2\delta \leq \beta - 1/2 - r/2$, where $2\beta - 1 - r$ is positive. Then

$$a_n \sqrt{n}\varepsilon_n \sup_{v \in (n^{-r-2\delta}, 1/2]} \Big\{\frac{\mu_n(0, v]}{\sqrt{v}}\Big\} \leq a_n(\log(n))^{E(\beta, \sigma_0)} n^{1/2 - \beta + r/2 + \delta} \to 0.$$

Consequently, by Theorem 3.2 it remains to show that

$$a_n n^{1/2 - \beta}(\log(n))^{E(\beta, \sigma_0)} \sup_{\kappa \in [r + 2\delta, 1 - \lambda_n]} n^{\kappa/2}\mu_n(0, n^{-\kappa}] \to 0.$$

For this purpose, a fine analysis of the tail behaviour of $\Phi$ is required.

**Lemma B.10.** *We have*

$$\frac{x}{\sqrt{2\pi}(1 + x^2)} \exp\Big(-\frac{1}{2}x^2\Big) \leq 1 - \Phi(x) \leq \frac{1}{\sqrt{2\pi}x} \exp\Big(-\frac{1}{2}x^2\Big) \tag{B.41}$$

*for all $x > 0$. Moreover, there is some $U > 0$ such that for all $u \in (0, U)$*

$$-\Phi^{-1}(u) = \Phi^{-1}(1 - u) \geq \sqrt{2\log(u^{-1})}\Big(1 - \frac{7 + \log\log(u^{-1})}{4\log(u^{-1})}\Big). \tag{B.42}$$

*Proof.* From integration by parts we obtain for all $x > 0$

$$1 - \Phi(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} \frac{1}{t} \, te^{-t^2/2} \, \mathrm{d}t = \frac{1}{x\sqrt{2\pi}}e^{-x^2/2} - \int_x^\infty \frac{1}{t^2\sqrt{2\pi}}e^{-t^2/2} \, \mathrm{d}t.$$

Hence, the upper bound in (B.41) follows. Since the integral on the right-hand side is smaller than $x^{-2}(1 - \Phi(x))$ also the lower bound follows. Clearly, $\Phi^{-1}$ is increasing and $\Phi^{-1}(1-u) \to \infty$ as $u \searrow 0$. Let $U > 0$ such that $\Phi^{-1}(1-U) > 1$. By applying (B.41) for $x = \Phi^{-1}(1-u)$ with $u \in (0, U)$

$$\Phi^{-1}(1-u) \le \sqrt{-2\log(u\sqrt{2\pi}\Phi^{-1}(1-u))} \le \sqrt{-2\log(u)}. \qquad \text{(B.43)}$$

Obviously, by (B.41) we have $(1/6x)\exp(-x^2/2) \le 1 - \Phi(x)$ for all $x > 1$. By setting again $x = \Phi^{-1}(1-u)$ for $u \in (0, U)$ we obtain from this, (B.43) and $\sqrt{1-y} \ge 1 - y/2 - y^2$ for all $y \in (0, 1)$ that

$$\Phi^{-1}(1-u) \ge \sqrt{2\log(u^{-1})}\sqrt{1 - \frac{\log(6) + \log(2)/2 + \log\log(u^{-1})/2}{\log(u^{-1})}}$$

$$\ge \sqrt{2\log(u^{-1})}\Big(1 - \frac{3 + \log\log(u^{-1})/2}{2\log(u^{-1})} - \Big(\frac{3 + \log\log(u^{-1})/2}{\log(u^{-1})}\Big)^2\Big).$$

Finally, by choosing $U > 0$ sufficiently small we get (B.42). $\qquad \square$

From now on, let $n \in \mathbb{N}$ be sufficiently large such that $n^{-1+\lambda_n} < U$ and so (B.42) holds for all $u = n^{-\kappa}$, $\kappa \le 1 - \lambda_n$. We obtain for all $\kappa \in [r + 2\delta, 1 - \lambda_n]$

$$-\Phi^{-1}(n^{-\kappa}) - \vartheta_n \ge \sqrt{2\log(n)}\Big(\sqrt{\kappa} - \sqrt{r} - \frac{\log(\kappa) + \log\log(n) + 7}{4\sqrt{\kappa}\log(n)}\Big) =: w_n(\kappa).$$

Hence, by (B.39) and (B.41) there is $c > 0$ such that for all $\kappa \in [r + 2\delta, 1 - \lambda_n]$

$$n^{\frac{1}{2}\kappa}\mu_n(0, n^{-\kappa}] \le n^{\frac{1}{2}\kappa}\Big(1 - \Phi\Big(\frac{w_n(\kappa)}{\sigma_0}\Big)\Big) \le n^{\frac{1}{2}\kappa}\frac{\sigma_0}{w_n(\kappa)}\exp\Big(-\frac{1}{2\sigma_0^2}w_n(\kappa)^2\Big)$$

$$\le cn^{E_1(\kappa)}(\log(n))^{E_2(\kappa)} \text{ with } E_2(\kappa) = -\frac{1}{2} + \frac{1}{2}\frac{\sqrt{\kappa} - \sqrt{r}}{\sigma_0^2\sqrt{\kappa}}$$

$$\text{and } E_1(\kappa) = \frac{1}{2}\kappa + \sigma_0^{-2}(2\sqrt{\kappa r} - \kappa - r).$$

Since we are interested in the supremum of all $\kappa \in [r + 2\delta, 1 - \lambda_n]$ we need to find the (uniquely) point $\kappa_n^* \in [r + 2\delta, 1 - \lambda_n]$ attaining the maximum of $[r + 2\delta, 1 - \lambda_n] \ni \kappa \to E_1(\kappa)$. For this purpose we need to discuss two cases.

First, let $\sigma_0 < \sqrt{2}$ and $r < (2 - \sigma_0^2)^2/4$ (or equivalently $\beta < 1 - \sigma_0^2/4$). Then $E(\beta, \sigma_0) = 0$, $\varepsilon_n = n^{-\beta}$ and $r = (2 - \sigma_0^2)(\beta - 1/2)$. Without loss of generality we assume that $r + 2\delta < 4r(2 - \sigma_0^2)^{-2} < (1 - \delta)^2$ and $\delta(2 - \sigma_0^2)/(4\sigma_0^2) < 1/8$. Then it is easy to verify that $\kappa_n^* = \kappa^* = 4r/(2 - \sigma_0^2)^2$ and $E_1(\kappa_n^*) = r/(2 - \sigma_0^2)$. Since $E_2$ is increasing we have for all sufficiently large $n \in \mathbb{N}$ that

$$a_n\sqrt{n}\varepsilon_n \sup_{\kappa \in [r+2\delta, 1-\lambda_n]} n^{\kappa/2}\mu_n(0, n^{-\kappa}] = a_n \sup_{\kappa \in [r+2\delta, \kappa^*(1-\delta)^{-2}]} n^{\kappa/2+1/2-\beta}\mu_n(0, n^{-\kappa}]$$

$$\le a_n c\, n^{E_1(\kappa^*)+1/2-\beta}(\log(n))^{E_2(\kappa^*(1-\delta)^{-2})}$$

$$\le a_n c\,(\log(n))^{-1/8} \to 0.$$

Second, let $(\beta, \sigma_0) \in (1-1/\sigma_0^2, 1) \times (\sqrt{2}, \infty)$ or $(\beta, \sigma_0) \in [1-\sigma_0^2/4, 1) \times (0, \sqrt{2})$. Clearly, $E_1$ and $E_2$ are increasing in $[r+2\delta, 1]$. Hence, $\kappa_n^* = 1 - \lambda_n$. Since $r = (1-\sigma_0\sqrt{1-\beta})^2$, $1/2 - 1/\sigma_0^2 + 2\sqrt{r}/\sigma_0^2 - r/\sigma_0^2 = \beta - 1/2$ and $\sqrt{1-\lambda_n} \leq 1 - \lambda_n/2$ we obtain that

$$
\begin{aligned}
E_1(1-\lambda_n) &= \beta - \frac{1}{2} + \lambda_n\left(\frac{1}{\sigma_0^2} - \frac{1}{2}\right) + \frac{2}{\sigma_0^2}\sqrt{r}(\sqrt{1-\lambda_n} - 1) \\
&\leq \beta - \frac{1}{2} - K(\beta, \sigma_0^2)\lambda_n, \text{ where} \\
K(\beta, \sigma_0^2) &= \frac{1}{2} - \frac{1}{\sigma_0}\sqrt{1-\beta}\begin{cases} = 0 & \text{if } \beta = 1 - \frac{1}{4}\sigma_0^2, \ \sigma_0 < \sqrt{2}. \\ > 0 & \text{else.} \end{cases}
\end{aligned}
$$

Moreover, $E_2(1) = -\frac{1}{4} < 0$ if $\beta = 1 - \sigma_0^2/4$, $\sigma_0^2 < \sqrt{2}$. Consequently,

$$
\begin{aligned}
a_n\sqrt{n}\varepsilon_n &\sup_{\kappa \in [r+2\delta, 1-\lambda_n]} n^{\kappa/2}\mu_n(0, n^{-\kappa}] \\
&\leq a_n c\, n^{E_1(1-\lambda_n)+1/2-\beta}(\log(n))^{E_2(1)+E(\beta,\sigma_0^2)} \\
&\leq a_n c(\log(n))^{E_2(1)+E(\beta,\sigma_0^2)-K(\beta,\sigma_0^2)\log\log(n)} \to 0.
\end{aligned}
$$

### B.6.2. Proof of Theorem 4.10

By careful calculations we obtain

$$
\frac{1}{n}\frac{\mathrm{d}\mu_n}{\mathrm{d}P_0}(x+\vartheta_n) = \frac{1}{\sigma_0}\exp\left(\frac{\sigma_0^2-1}{2\sigma_0^2}x^2 + x\sqrt{2r\log n} + (r-1)\log n\right).
$$

Define $C_{n,\tau} = \{x \in \mathbb{R} : n^{-1}\frac{\mathrm{d}\mu_n}{\mathrm{d}P_0}(x+\vartheta_n) > \tau\}$, $\tau > 0$. It is easy to see that $\mathbf{1}\{x \in C_{n,\tau}\} \to \mathbf{1}\{r = 1, x > 0\} + \mathbf{1}\{r > 1\}$ for $x \neq 0$. From this and Lebesgue's dominated convergence theorem we deduce that

$$
I_{n,1,\tau} = \int \mathbf{1}\{x \in C_{n,\tau}\}\mathrm{d}N(0,1)(x) \to \mathbf{1}\{r > 1\} - \frac{1}{2}\mathbf{1}\{r = 1\}.
$$

Moreover,

$$
I_{n,2,\tau} \leq \tau \int \frac{\mathrm{d}\mu_n}{\mathrm{d}P_0}\mathbf{1}\left\{\frac{1}{n}\frac{\mathrm{d}\mu_n}{\mathrm{d}P_0} \leq \tau\right\}\mathrm{d}P_0 \leq \tau.
$$

Finally, combining Theorem 2.4 and Theorem 2.1 yields the statement.

### References

[1] ARIAS-CASTRO, E. AND CANDÈS, E. J. AND PLAN, Y. (2015). Global testing under sparse alternatives: ANOVA, multiple comparisons and the higher criticism. *Ann. Statist.* **39**, no.5, 2533–2556. MR2906877

[2] ARIAS-CASTRO, E. AND WANG, M. (2015). The sparse Poisson means model. *Electron. J. Stat.* **9**, no. 2, 2170–2201. MR3406276

[3] ARIAS-CASTRO, E. AND WANG, M. (2017). Distribution-free tests for sparse heterogeneous mixtures. *TEST* **26**, no. 1, 71–94. MR3613606

[4] BILLINGSLEY, P. (1999). *Convergence of Probability Measures*, 2nd ed. Wiley, New York. MR1700749

[5] CAI, T., JENG, J. AND JIN, J. (2011). Optimal detection of heterogeneous and heteroscedastic mixtures. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **73**, no. 5, 629–662. MR2867452

[6] CAI, T. AND WU, Y. (2014). Optimal Detection of Sparse Mixtures Against a Given Null Distribution. *IEEE Trans. Inform. Theory* **60**, no. 4, 2217-2232. MR3181520

[7] CAYON, L., JIN, J. AND TREASTER, A. (2004). Higher Criticism statisitc: Detecting and identifying non-Gaussianity in the WMAP first year data. *Mon. Not. Roy. Astron. Soc.* **362**, 826–832.

[8] DAI, H., CHARNIGO, R., SRIVASTAVA, T., TALEBIZADEH, Z. AND QING, S. (2012). Integrating P-values for genetic and genomic data analysis. *J. Biom. Biostat.*, 3–7.

[9] Delaigle, A., Hall, J. & Jin, J. (2011). Robustness and accuracy of methods for high dimensional data analysis based on Student's t statistic. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **73**, 283–301. MR2815777

[10] DITZHAUS, M. (2017). *The power of tests for signal detection under high-dimensional data.* PhD-thesis, Heinrich-Heine-University Duesseldorf. https://docserv.uni-duesseldorf.de/servlets/DocumentServlet?id=42808

[11] DITZHAUS, M. (2018). *Signal detection via Phi-divergences for general mixtures.* arXiv:1803.06519v1.

[12] DONOHO, D. AND JIN, J. (2004). Higher criticism for detecting sparse heterogeneous mixtures. *Ann. Statist.* **32**, no. 3, 962–994. MR2065195

[13] DONOHO, D. AND JIN, J. (2015). Higher Criticism for Large-Scale Inference, Especially for Rare and Weak Effects. *Statist. Sci.* **30** , no. 1, 1–25. MR3317751

[14] EICKER, F. (1979). The asymptotic distribution of the suprema of the standardized empirical processes. *Ann. Stat.* **7**, 116–138. MR0515688

[15] GOLDSTEIN, D.B. (2009). Common genetic variation and human traits. *New England J. Med.* **360**, 1696–1698.

[16] GNEDENKO, B.V. AND KOLMOGOROV, A.N. (1954). *Limit distribution for sums of independent random variables,* Addison–Wesley, Reading, MA. Translated and annotated by K. L. Chung. MR0062975

[17] HÁJEK, J., ŠIDÁK, Z. AND SEN, P. K. (1999). *Theory of rank tests.* Probability and Mathematical Statistics, second edition. Academic Press, Inc., San Diego, CA. MR1680991

[18] Hall, P., Pittelkow, Y. & Ghosh, M. (2008). Theoretical measures of relative performance of classifiers for high dimensional data with small sample sizes. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **70**, 158–173. MR2412636

[19] INGSTER, Y. (1997). Some problems of hypothesis testing leading to infinitely divisible distributions. *Math. Methods Statist.* **6**, no. 1, 47–69.

MR1456646

[20] INGSTER, Y. I. AND TSYBAKOV, A. B. AND VERZELEN, N. (2010). Detection boundary in sparse regression. *Electron. J. Stat.* **4**, 1476–1526. MR2747131

[21] IYENGAR, S. K. AND ELSTON, R.C. (2007). The genetic basis of complex traits: Rare vvariant or "common gene, common disease"? *Methods Mol. Biol.* **376**, 71–84.

[22] JAESCHKE, D. (1979). The asymptotic distribution of the suprema of the standardized empirical distribution function on subintervals. *Ann. Stat.* **7**, no. 1, 108–115. MR0515687

[23] JAGER, L. AND WELLNER, J. (2007). Goodness-of-fit tests via phi-divergences. *Ann. Stat.* **35**, no. 5, 2008–2053. MR2363962

[24] JANSSEN, A., MILBRODT, H. AND STRASSER, H. (1985). *Infinitely divisible statistical experiments.* Lecture notes in Statistic **27**, Springer-Verlag, Berlin. MR0788883

[25] JANSSEN, A (1990). Statistical experiments with non-regular densities. In: Janssen, A. and Mason, D. M., Non-Standard Rank Tests. Lecture Notes Stat. **65**, 183–240. MR1080968

[26] JIN, J. (2004). Detecting a target in very noisy data from multiple looks. *A festschrift for Herman Rubin*, 255–286, IMS Lecture Notes Monogr. Ser., 45, Inst. Math. Statist., Beachwood, OH. MR2126903

[27] JIN, J., STARK, J.-L., DONOHO, D., AGHANIM, N. AND FORNI, O. (2005). Cosmological non-Gaussian signature detection: Comparing performance of different statistical tests. *J. Appl. Signal Processing* **15**, 2470–2485. MR2210857

[28] KHMALADZE, E.V. (1998). Goodness of fit tests for Chimeric alternatives.*Statist. Neerlandica* **52**, no. 1, 90–111. MR1615550

[29] Khmaladze, E. and Shinjikashvili, E. (2001). Calculation of noncrossing probabilities for Poisson processes and its corollaries. *Adv. in Appl. Probab.* **33**, 702–716. MR1860097

[30] KULLDORFF, M., HEFFERNAN, R., HARTMAN, J., ASSUNCAO, R. AND MOSTASHARI, F. (2005). A space-time permutation scan statistic for disease outbreak detection. *PLoS Med* **2**, no. 3, e59.

[31] LE CAM, L. (1986). *Asymptotic methods in statistical decision theory.* Springer Series in Statistics. Springer-Verlag, New York. MR0856411

[32] LE CAM, L. AND YANG, G. L. (2000). *Asymptotics in statistics.* Second edition. Springer Series in Statistics. Springer Verlag, New York. MR1784901

[33] MUKHERJEE, R., PILLAI, N. S. AND LIN, X (2015). Hypothesis testing for high-dimensional sparse binary regression. *Ann. Statist.* **43**, no. 1, 352–381. MR3311863

[34] NEILL, D. AND LINGWALL, J. (2007). A nonparametric scan statistic for multivariate disease surveillance. *Advances in Disease Surveillance* **4**, 106–116.

[35] SALIGRAMA, V. and ZHAO, M. (2012). Local anomaly detection. *JMLR W& CP* **22**, 969–983.

[36] STRASSER, H. (1985). *Mathematical Theory of Statistics,* De Gruyter, Berlin/New York. MR0812467

[37] TUKEY, J. W. (1976). *T13 N: The higher Criticism.* Coures Notes. Stat 411. Princetion Univ.

[38] TUKEY, J. W. (1989). *Higher Criticism for individual significances in serveral tables or parts of tables.* Internal working paper, Princeton Univ.

[39] TUKEY, J. W. (1994). *The Collected Works of John W. Tukey: Multiple Comparisons, Volume VIII.* Chapman and Hall, London. MR1263027