On the Asymptotics of the LRT in Mixture Models

Dissertation

zur Erlangung des Doktorgrades
des Fachbereichs Mathematik
der Universität Hamburg

vorgelegt von

Astrid Ruck
aus Husum

Hamburg
2005

Als Dissertation angenommen vom Fachbereich
Mathematik der Universität Hamburg
auf Grund der Gutachten von Prof. Dr. H. Daduna
und Prof. Dr. W. Seidel

Hamburg, den 19. April 2005

Prof. Dr. A. Kreuzer
Dekan des Fachbereichs Mathematik

Contents

1. Introduction

Finite mixture models are often used for a mathematical description of heterogeneous populations which are composed by a finite number of subpopulations with respect to several random phenomena or features. Within the scope of statistics, finite mixture models are especially used for discriminant analysis, clustering or image analysis. Their practical applications are in a variety of areas such as economics, medicine or biology. Likewise, theoretical statistical studies are combined with concrete applications of mixture models.

For example, let us assume that we have a population of animals consisting of two components, either male or female. Additionally, we wish to construct a classification rule with respect to these component types based on the feature of length which is e.g. normally distributed on each component type. A random sample provides a two population mixture of normal distributions. The a priori probabilities or mixing proportions for each class may be known or unknown. Likewise, the underlying parameters of the distribution, here given by mean and variance, could either be known or unknown. Clearly, this example can easily be extended to more complicated applications. These may consider more than one feature and may be based on p different populations. Furthermore, there could be several distributions depending on the features and the number of components may be unknown.

A short survey of various statistical problems of our kind and the corresponding models can be found in the first chapter of Lindsay (1995). His book also addresses elaborately the theoretical aspects of mixtures. A comprehensive illustration of the main issues in finite mixture models is given by McLachlan and Peel (2000), whereas Böhning (2000) is concerned, in particular, with likelihood methods.

In our work, we investigate the problem of testing for the numbers of components in a mixture. More precisely, we consider the likelihood ratio test (LRT) for testing Q populations against p populations.

Let $\mathcal{F} = \{f_{\boldsymbol{\gamma}} : \boldsymbol{\gamma} \in \Gamma\}$, $\Gamma \subset I\!R^k$, be a family of probability densities with respect to some σ-finite measure ν on a measurable space $(\mathfrak{X}, \mathcal{X})$, $\mathfrak{X} \subset I\!R^m$. The model of all p population mixtures of $\mathcal{F}$ is given by

$$\mathcal{G}_p = \left\{ g_{\boldsymbol{\pi},(\boldsymbol{\gamma}^1,\ldots,\boldsymbol{\gamma}^p)} = \sum_{l=1}^{p} \pi_l f_{\boldsymbol{\gamma}^l} : \boldsymbol{\pi} \in \mathit{\Pi}_p, \boldsymbol{\gamma}^1, \ldots, \boldsymbol{\gamma}^p \in \Gamma \right\}$$

with $\mathit{\Pi}_p = \left\{ \boldsymbol{\pi} = (\pi_1, \ldots, \pi_p)' \in [0,1]^p : \sum\limits_{l=1}^{p} \pi_l = 1 \right\}$.

The specific attributes of the lth component population are described by the component parameter $\boldsymbol{\gamma}^l$. The quantities $\pi_1, \ldots, \pi_p$ are typically referred to as mixing proportions or weights and the functions $f_{\boldsymbol{\gamma}^1}, \ldots, f_{\boldsymbol{\gamma}^p}$ are called the component densities of the mixture $\sum\limits_{l=1}^{p} \pi_l f_{\boldsymbol{\gamma}^l}$ which is obviously also a density. Accordingly, finite mixture distribution functions are defined by their weighted component distribution functions. We define the true order of $g \in \mathcal{G}_p$ as the smallest integer t such that g has a representation $\sum_{i=1}^{t} \pi_i f_{\boldsymbol{\gamma^i}}$ and all the $\boldsymbol{\gamma^i}$, $i = 1, \ldots, t$ are different and all the associated π_i are not null, i.e $\boldsymbol{\pi} \in \Pi_t \cap (0,1)^t$. (Additional assumptions will ensure that different component parameters $\boldsymbol{\gamma^i}$, $i = 1, \ldots, t$ are equivalent to different component densities $f_{\boldsymbol{\gamma^i}}$, $i = 1, \ldots, t$ later on.)

Notice that we use bold letters to represent a column vector, e.g. $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_k)'$, where the superscript $'$ denotes the transposition of a vector or a matrix. Every single component of a vector is supplied with a subscript index, while a sequence of vectors is supplied with superscript indices, e.g. $\boldsymbol{\gamma^1}, \ldots, \boldsymbol{\gamma^p}$. In addition, we use the statistical terminology of Witting (1985) unless an alternative reference is given.

As mentioned above, we study the problem of testing for the number of components in a mixture. There are diverse statistical techniques such as Bayesian approaches, e.g. the Bayesian information criterion (BIC), for estimating the number of components. The LRT provides a general method to construct a test for the number of components in a mixture. We are interested in the recent field of research in asymptotic theory of the LRT, its applicability to exponential families and ways to calculate of the asymptotic distribution of the log-LRT statistic.

Assume we wish to test the null hypothesis that the true density is given by a mixture of Q components versus the alternative that the true density is given by a mixture of p components for some $Q < p$; i.e.

$$\boldsymbol{H_0 : g \in \mathcal{G}_Q} \quad \text{against} \quad \boldsymbol{H_1 : g \in \mathcal{G}_p \backslash \mathcal{G}_Q} \ .$$

Let the test be based on an independent identically distributed (i.i.d.) sample $\boldsymbol{X_1}, \boldsymbol{X_2}, \ldots, \boldsymbol{X_n}$ of size n, with true density of $\boldsymbol{X_i}$, $i = 1, \ldots, n$, given by

$$g^0 = \sum_{l=1}^{p} \pi_l^0 f_{\boldsymbol{\gamma^{0,l}}} .$$

Notice that true parameters, such as mixing weights and component parameters as well as the corresponding true mixture are supplied with a superscript 0.

Given a realization $\boldsymbol{x_1}, \ldots, \boldsymbol{x_n}$ of $\boldsymbol{X_1}, \ldots, \boldsymbol{X_n}$, we define the likelihood function of $\boldsymbol{\pi} \in \Pi_p$ and $\boldsymbol{\gamma^1}, \ldots, \boldsymbol{\gamma^p} \in \Gamma$ by

$$L_{(\boldsymbol{x_1}, \ldots, \boldsymbol{x_n})}\Big(\boldsymbol{\pi}, (\boldsymbol{\gamma^1}, \ldots, \boldsymbol{\gamma^p})\Big) = \prod_{i=1}^{n} g_{\boldsymbol{\pi}, (\boldsymbol{\gamma^1}, \ldots, \boldsymbol{\gamma^p})}(\boldsymbol{x_i}).$$

The LRT statistic has the form

$$\Lambda(\boldsymbol{x_1}, \ldots, \boldsymbol{x_n}) = \frac{\sup\limits_{g \in \mathcal{G}_p} \prod\limits_{i=1}^{n} g(\boldsymbol{x_i})}{\sup\limits_{g \in \mathcal{G}_Q} \prod\limits_{i=1}^{n} g(\boldsymbol{x_i})},$$

where the numerator is equivalent to

$$\sup \Big\{ L_{(\boldsymbol{x_1}, \ldots, \boldsymbol{x_n})}\Big(\boldsymbol{\pi}, (\boldsymbol{\gamma^1}, \ldots, \boldsymbol{\gamma^p})\Big) : \boldsymbol{\pi} \in \Pi_p, \boldsymbol{\gamma^1}, \ldots, \boldsymbol{\gamma^p} \in \Gamma \Big\}$$

and the denominator is equivalent to

$$\sup\Big\{L_{(\boldsymbol{x_1},\ldots,\boldsymbol{x_n})}\Big(\boldsymbol{\pi},(\boldsymbol{\gamma^1},\ldots,\boldsymbol{\gamma^Q})\Big):\boldsymbol{\pi}\in\mathit{\Pi}_Q,\boldsymbol{\gamma^1},\ldots,\boldsymbol{\gamma^Q}\in\Gamma\Big\}.$$

If the test statistic $\Lambda(\boldsymbol{x_1},\ldots,\boldsymbol{x_n})$ is sufficiently large, the null hypothesis H_0 will be rejected. Following the literature we use the log-LRT statistic

$$\log\Big(\Lambda(\boldsymbol{x_1},\ldots,\boldsymbol{x_n})\Big) = \sup_{g\in\mathcal{G}_p}\sum_{i=1}^{n}\log\Big(g(\boldsymbol{x_i})\Big)-\sup_{g\in\mathcal{G}_Q}\sum_{i=1}^{n}\log\Big(g(\boldsymbol{x_i})\Big). \tag{1.1}$$

In order to decide whether to reject H_0 or not, we need the distribution of the log-LRT statistic under the null hypothesis. Unfortunately, in general there is no explicit expression known and the usual chi-squared asymptotic of Wilks (1938) does not hold for $\frac{1}{2}\log(\Lambda(\boldsymbol{x_1},\ldots,\boldsymbol{x_n}))$.

It is well known that the classical distribution theory of the LRT is not applicable on mixture models and "even in the simple case of testing homogeneity against a mixture with two components, a lot of problems remain to be solved and a complete answer has not yet been given", see Garel (2005). In the classical likelihood theory, it is assumed that the parameters describing the null hypothesis are uniquely defined and belong to the interior of the parameter space. In mixture models the parameters belonging to the null hypothesis are on the boundary of the parameter space rather than within it and even if the mixtures are identifiable, the corresponding parameters (under the null hypothesis) may be non-identifiable, see Gosh and Sen (1985). Chernoff (1954) shows how to deal with the problem when the null hypothesis is in a certain sense on the boundary. Thus the real problem of testing $g=g^0$ against $g\neq g^0$ results from the fact that g^0 has several representations. More precisely, the maximum likelihood estimator may not converge to a unique accumulation point (under the null hypothesis). In general, the lack of identifiability leads to a degenerate Fisher information matrix and as a consequence the usual concept to derive an asymptotic fails.

Gosh and Sen (1985) were the first to develop an expression of the asymptotic LRT statistic for testing one population against two populations. They solve the problem of non-identifiability of the usual parameterization with the aid of a separation condition applied to the parameters. Additionally, they assume the parameter space to be compact. This assumption conforms to a result of Hartigan (1985) who proves that for simple normal mixtures with unbounded mean, the log-LRT statistic diverges to infinity in probability when n tends to infinity.

Recent research has removed this separation condition and there is now a great variety of approaches to this problem. Frequently, mixtures with respect to a concrete distribution are investigated under a reparameterization of the mixture model. For instance, Chernoff and Lander (1995) study the asymptotic distribution of the LRT for homogeneity against a mixture of two binomial distributions. In their work several versions of binomial mixture problems are studied under an appropriate parameterization. Lemdani and Pons (1997) investigate mixtures of binomial distributions in LR tests for linkage between genes and markers using ad hoc reparameterizations. Leroux (1992) proposes a maximum penalized likelihood method by estimating a mixture distribution and proves that this method produces a consistent estimator in the sense of weak convergence. Furthermore, he illustrates his results by estimating a Poisson mixture. Keribin (2000) proves the almost sure consistency of the maximum penalized likelihood estimator for an appropriate penalization sequence using the reparameterization developed

by Dacunha-Castelle and Gassiat (1999). She shows that the corresponding theory holds for Poisson mixtures as well as for specific Gaussian mixtures with unknown mean as described below. Garel (2001) considers the problem of testing homogeneity against a mixture of two components in several cases of univariate Gaussian mixtures. Additionally, Liu and Shao (2004) have recently published asymptotic results of the LRT in a two component univariate normal mean mixture model.
An outline of further theoretical results based on special mixture models is given in McLachlan and Peel (2000), section 6.5.1. Garel (2005) also presents a general survey on the theory and methodology of mixture distributions.

Dacunha-Castelle and Gassiat (1997, 1999) propose a general theory for deriving the limiting distribution of the LRT statistic for testing Q populations against p populations. They focus on a reparameterization of the mixture model such that one of the "new" parameters is identifiable at the previously non-identifiable point. Another general approach to the asymptotic LRT statistic for testing for the number of components has been developed in a recent work by Liu and Shao (2003). In their applications to finite mixture models they use the parameterization of Dacunha-Castelle and Gassiat (1999) as well as a special new parameterization.

The very latest work by Garel (2005) presents an asymptotic theory of the LRT, however neither using a separation condition nor a reparameterization. His results require assumptions based on the second order derivatives of the log-density and hold for testing homogeneity against a two population mixture based on a one-dimensional parameter space.

Our work is based on the general asymptotic theory developed by Dacunha-Castelle and Gassiat (1999) [DCG99] for k-dimensional parameter spaces. As previously mentioned they use a local reparameterization (called "locally conic") of mixtures to deal with the identifiability problem. Roughly speaking, under the assumption that the true density

$$g^0 \quad = \quad \sum_{l=1}^{q} \pi_l^0 f_{\gamma^{0,l}} \quad \text{with } \pi_1^0, \ldots, \pi_q^0 \in (0,1),\ q \leq Q,$$

belongs to the null hypothesis, they reparameterize $\mathcal{G}_p$ by two parameters θ and $\boldsymbol{\beta}$ such that θ describes a "normalized distance" to $(\boldsymbol{\pi^0}, \boldsymbol{\gamma^{0,1}}, \ldots, \boldsymbol{\gamma^{0,q}})$ and the parameter vector $\boldsymbol{\beta}$ describes a corresponding "direction." In principle, the reparameterization is identifiable in the parameter θ at $\theta = 0$ even if it is non-identifiable in $\boldsymbol{\beta}$.
As a main result they obtain that the log-LRT statistic for testing Q against p populations converges in distribution to a difference of two functionals of Gaussian processes with index sets directly related to the parameter sets corresponding to the model of Q and p populations, respectively. More precisely, [DCG99] claim in Theorem 3.2. that the log-LRT statistic

$$T_n(p) \quad = \quad \sup_{g \in \mathcal{G}_p} \sum_{i=1}^{n} \log\Big(g(\boldsymbol{x_i})\Big) - \sum_{i=1}^{n} \log\Big(g^0(\boldsymbol{x_i})\Big) \tag{1.2}$$

(for testing $g = g^0 \in \mathcal{G}_Q$ against $g \in \mathcal{G}_p \backslash \{g^0\}$) converges in distribution to

$$\frac{1}{2} \sup_{d \in \mathcal{D}} \big(\xi_d\big)^2 \cdot \mathbb{1}_{\xi_d \geq 0},$$

where $\big(\xi_d\big)_{d \in \mathcal{D}}$ is a centered Gaussian process on a class of functions $\mathcal{D}$ which is given by the set of directional scores for $\theta = 0$ of $\mathcal{G}_p$.

Since $\log\big(\Lambda(\boldsymbol{x_1},\ldots,\boldsymbol{x_n})\big) = T_n(p) - T_n(Q)$ according to (1.1) and (1.2), it follows that

$$\log\Big(\Lambda(\boldsymbol{x_1},\ldots,\boldsymbol{x_n})\Big) \xrightarrow{\mathcal{L}} \frac{1}{2}\sup_{d\in\mathcal{D}}\big(\xi_d\big)^2\cdot\mathbb{1}_{\xi_d\geq 0} - \frac{1}{2}\sup_{d\in\mathcal{D}_0}\big(\xi_d\big)^2\cdot\mathbb{1}_{\xi_d\geq 0}, \tag{1.3}$$

where $\mathcal{D}_0 \subset \mathcal{D}$ is the set of directional scores for $\theta = 0$ of $\mathcal{G}_Q$. ("$\xrightarrow{\mathcal{L}}$" denotes convergence in distribution.)

Finally, an orthogonal dissection yields

$$\frac{1}{2}\sup_{d\in\mathcal{D}}\big(\xi_d\big)^2\cdot\mathbb{1}_{\xi_d\geq 0} - \frac{1}{2}\sup_{d\in\mathcal{D}_0}\big(\xi_d\big)^2\cdot\mathbb{1}_{\xi_d\geq 0} \;=\; \frac{1}{2}\sup_{u\in\mathcal{U}}\big(\xi_u\big)^2\cdot\mathbb{1}_{\xi_u\geq 0},$$

where $\mathcal{U}$ is a subset of the orthogonal space of $\mathcal{D}_0$ (see [DCG99], Theorem 3.6).

For the special case of testing one population against a two population mixture (with one-dimensional component parameter space Γ) [DCG99] give an equivalent representation of formula (1.3), namely

$$\log\Big(\Lambda(\boldsymbol{x_1},\ldots,\boldsymbol{x_n})\Big) \xrightarrow{\mathcal{L}} \frac{1}{2}\sup_{\gamma\in\Gamma\setminus\{\gamma^{0,1}\}}\big(\xi_{h(\gamma)}\big)^2\cdot\mathbb{1}_{\xi_{h(\gamma)}\geq 0} \tag{1.4}$$

(see [DaCa99], Corollary 3.7).

Principally, our work is concerned with the following questions:

- Which mixture models fulfill the sufficient conditions from [DCG99] for the existence of an asymptotic distribution?

When endeavouring this question, we found that the sufficient conditions for the existence of an asymptotic distribution according to [DCG99] are not satisfied for all mixtures of exponential families. For instance, the sufficient conditions from [DCG99] for testing Q populations against p populations are not satisfied for univariate Gaussian mixtures if both parameters mean and variance are unknown. However, in the case of testing $g = g^0$ against $g = (1-\pi)f_{\boldsymbol{\gamma}^0} + \pi f_{\boldsymbol{\gamma}}$, $\pi \in (0,1)$, $\boldsymbol{\gamma} \in \Gamma\setminus\{\boldsymbol{\gamma^0}\}$ and true density $g^0 = f_{\boldsymbol{\gamma}^0}$, i.e. for so-called simple mixtures or contamination models, the asymptotic theory by Dacunha-Castelle and Gassiat (1997) [DCG97] does hold for the aforementioned Gaussian families (see [DCG97], Remark 4.2). The reason for this is a simplified reparameterization of the contamination model (see [DCG97], p. 296) compared with the reparameterization of $\mathcal{G}_2$ as given in [DCG99], although the simple mixture model is a subset of $\mathcal{G}_2$.

Chapter 3 investigates in detail the applicability of the theory from [DCG99] to exponential families. We derive sufficient conditions for the existence of a limiting distribution which depends mainly on the statistic and the parameter function of the corresponding exponential families. Considering a minimal parameterization of these families, our conditions are based on the analytical properties of the statistic and the ν-affine independence of its components as well as on the form of the Jacobian matrix of the parameter function and the affine independence of its corresponding components. Additionally, we introduce Proposition 3.11 as an elementary tool to check one of the required conditions from [DCG99]. The advantage of Proposition 3.11 is, that it allows us to bound rather complicated suprema required in [DCG99] by utilizing certain suprema of scalar products of the parameter function and the statistic. As a consequence we need not to analyze the geometric structure of the densities $\{f_\gamma : \gamma \in \Gamma\}$ for the establishment of the bound of the aforementioned suprema, which makes it more widely applicable. We apply Proposition 3.11 to all examples of exponential families considered in our work, see section 3.2. In our main Theorem 3.13 we collect all sufficient conditions on the statistic and the parameter function, which ensure the existence of an asymptotic distribution according to [DCG99]. It will become apparent that the structure of the underlying parameter space Γ is of considerable importance. To be more precise, the parameter space may depend on the true component parameters. It seems to exist a connection between this dependency and the canonical parameter space Z_*, since it occurs in all of our examples with $Z_* \neq I\!R^k$, see section 3.2.

Finally, we prove with the help of Theorem 3.13 the applicability of the method, presented in [DCG99], for various examples (see section 3.2). Keribin (2000) has already shown that it works for mixtures of Poisson distributions, for multivariate Gaussian mixtures with unknown mean and known covariance and with known mean and unknown covariance matrix, respectively, where in both cases the covariance matrix is assumed to be a multiple of the identity matrix. Thanks to Theorem 3.13 we are able to generalize her results on the Gaussian mixtures to arbitrary covariance matrices in the first case and to covariance matrices of the form $\Sigma = diag((\sigma_1)^2, \ldots, (\sigma_k)^2)$ in the second case.

- Is it possible to generalize the sufficient conditions for the existence of an asymptotic distribution of the LRT statistic introduced in [DCG99]?

[DCG99] assume two conditions, (P0) and (P1), to ensure the existence of the limiting distribution of the LRT statistic. On p. 1187 they state "assumption (P0) is probably not optimal. It should be possible to prove the result using only derivatives up to order 3." In our work we modify property (P1), only, since the densities of an exponential family have at least as many derivatives as the parameter function of the exponential family, which is infinitely often differentiable in all our examples. Our first modification is a generalization of (P1) for testing $Q = q$ populations against $q + 1$ populations with parameter space $\Gamma \subset I\!R^k$. Our second modification of (P1) applies to the test $g \in \mathcal{G}_1$ against $g \in \mathcal{G}_{2,M} \backslash \mathcal{G}_1$, where $\mathcal{G}_{2,M}$ is a two population mixture model with restricted parameter space $\tilde{\Gamma} \subset \Gamma \times \Gamma$, $\Gamma \subset I\!R^2$. In section 2.2 we show that the asymptotic theory from [DCG99] with respect to our modifications still holds. Finally, in the case of testing homogeneity against a two population mixture, we generalize Corollary 3.7 from [DCG99] from a one-dimensional parameter space to a k-dimensional space (see section 2.2.4). In this specific case our first modification imply (P1). The resulting limiting process has an index set $\Gamma \backslash \{\boldsymbol{\gamma}^{\mathbf{0,1}}\}$ corresponding to (1.4) instead of an index set of functions as given in (1.3), which, obviously, simplifies the calculation of the quantiles of the limiting distribution.

- Which mixture models fulfill the modified sufficient conditions for the existence of an asymptotic distribution?

In chapter 3 we investigate the applicability of the theory from [DCG99] with respect to our modifications for exponential families. Again we derive sufficient conditions for the existence of an asymptotic distribution which depend on the statistic and the parameter function. Theorem 3.13 addresses various combinations of properties of the statistic and the parameter function, which imply the sufficient conditions suggested in [DCG99] as well as our modifications. Best to our knowledge, we are the first to give sufficient conditions for the existence of a limiting distribution for testing homogeneity against a mixture of bivariate Gaussian distributions with known mean and arbitrary unknown covariance matrix, a situation where condition (P1) from [DCG99] is not satisfied (see Example 3.19). In Example 3.18 we show that our second modification holds for testing homogeneity against a mixture of two populations of univariate normal distributions with unknown mean and unknown variance, while condition (P1) from [DCG99] is not fulfilled. Our results corroborate Garel's conjecture that an asymptotic distribution of the LRT statistic exists for this test, see Garel (2001). The verification of his conjecture with the help of the methods of [DCG99] can be carried out, only, if one restricts the parameter space to the two population mixture model.

- How to calculate the asymptotic distribution of the log-LRT statistic?

In order to answer this question, we calculate in chapter 4 the asymptotic distribution of the log-LRT statistic for testing homogeneity against a two population mixture of univariate normal distributions. The corresponding limiting process can be simulated, since it is a centered Gaussian process, with the aid of its finite-dimensional marginal distributions. In this context we introduce a modified spectral factorization on the resulting covariance matrix, whose outcome is a product of surprisingly low dimensional matrices. This low dimensionality gives rise to the assumption that the limiting distribution is close to a low dimensional Gaussian one. As a consequence we observe considerably shorter computation times compared with that of the exact quantiles. A power study, based on both kinds of quantiles, demonstrates that they produce nearly the same power, even for a small sample size of $n = 1000$, although the asymptotic quantiles are larger than the exact ones.

2. Asymptotic Theory of the Likelihood Ratio Test (LRT) in Finite Mixtures

In the introduction we already mentioned that we are interested in testing for the number of components in finite mixtures which lead to tests of hypothesis. More precisely, we investigate the LRT statistic for testing

$$\boldsymbol{H_0 : g \in \mathcal{G}_Q} \ \text{ against } \ \boldsymbol{H_1 : g \in \mathcal{G}_p \backslash \mathcal{G}_Q} \quad , \ Q < p,$$

for a given true density $g^0 \in \mathcal{G}_Q$ belonging to the null hypothesis. Thereby, we use the definitions of finite mixtures as given in the introduction. Our work is based on the modified and revised article [DCG99] for its part refers to the earlier article [DCG97] in some passages. Any modification we make to some assumptions of [DCG99] will be marked correspondingly. The main purpose of this chapter is to verify that the concept from [DCG99] remains unchanged under slight generalizations of some of the sufficient conditions for deriving an asymptotic distribution.

Let us start with a first modification of some assumptions from [DCG99] with respect to the parameter space and underlying true parameters. Its meaning will be given below.

(A1) $\Gamma \subset \mathbb{R}^k$ is compact, $\boldsymbol{\gamma^{0,1}}, \ldots, \boldsymbol{\gamma^{0,q}} \in \Gamma$ are distinct accumulation points of Γ, $\boldsymbol{\pi^0} \in \Pi_q \cap (0,1)^q$ and g^0 has a representation $\sum_{l=1}^{q} \pi_l^0 f_{\boldsymbol{\gamma}^{0,l}}$.

Assumption (A1) implies that the true order of g^0 is equal to q. The requirements on the true component parameters to be accumulation points of Γ and on the true mixing weights to be strictly positive are not required by [DCG99]. We will show in Example 2.23 that Lemma 5.1 in [DCG99] may not hold without assuming that the true mixing weights are strictly positive. This assumption is an essential modification since Lemma 5.1 plays a decisive role in deriving a limiting distribution. Although this restriction on the true mixing weights will not change the central ideas of the asymptotic theory of [DCG99], it can very well lead to another testing problem, namely to the test of "Q populations against p populations" for $q \leq Q < p$. In other words, if $g^0 \in \mathcal{G}_q$ is an arbitrary true density with less than q mixture components then the method in [DCG99] for deriving an asymptotic distribution does not work for testing "q populations against p populations". Thus all mixing weights have to be strictly positive. As a consequence, we distinguish between the number Q of mixing components under the null hypothesis and the true order $q \leq Q$ of the true mixture g^0.

The compactness assumption on Γ is essential, since Hartigan (1985) proves that the log-LRT statistic may diverge to infinity if the component parameter space Γ is unbounded.

As previously mentioned, the mixture model $\mathcal{G}_p$ of all p population mixtures of ν-densities of $\mathcal{F} = \{f_\gamma : \gamma \in \Gamma\}$ is not identifiable for the parameters $\boldsymbol{\pi} = (\pi_1, \dots, \pi_p)'$ and $\boldsymbol{\gamma}^{\mathbf{1}}, \dots, \boldsymbol{\gamma}^{\boldsymbol{p}}$. [DCG99] develop their theory under the assumption that $\mathcal{G}_p$ is p-weakly identifiable in the following sense:

$$\text{(ID)} \qquad \sum_{l=1}^{p} \pi_l^a f_{\gamma^{a,l}} \;=\; \sum_{l=1}^{p} \pi_l^b f_{\gamma^{b,l}} \quad \nu\text{−a.e.} \quad \Leftrightarrow \quad \sum_{l=1}^{p} \pi_l^a \mathbb{1}_{\gamma^{a,l}} \;=\; \sum_{l=1}^{p} \pi_l^b \mathbb{1}_{\gamma^{b,l}}.$$

In other words $\mathcal{G}_p$ is p-weakly identifiable if the parameter is the discrete mixing probability distribution on Γ.

Notice that the p-weakly identifiability does not lead to unique representations of mixtures. For example, $g \in \mathcal{G}_p$ with representation $g_{\boldsymbol{\pi},(\gamma^1,\dots,\gamma^p)}$ also has the representation $g_{\boldsymbol{\pi}_\sigma,(\gamma^{\sigma(1)},\dots,\gamma^{\sigma(p)})}$ for any permutation σ of the index set $\{1, \dots, p\}$ and $g = f_\gamma \in \mathcal{G}_p$ is equal to $\sum_{l=1}^{p} \pi_l f_\gamma$ for any $\boldsymbol{\pi} \in \mathcal{T}\!\!\mathcal{T}_p$.

The main meaning of (ID) is that no density $f_\gamma \in \mathcal{F}$ can be ν-a.e. represented as a positive weighted sum of densities of $\mathcal{F}$ with component parameters distinct from $\boldsymbol{\gamma}$. More precisely, one has for any $l \in \{2, \dots, p\}$

$$\sum_{i=1}^{l} \pi_i f_{\gamma^i} \notin \mathcal{F} \quad \nu\text{−a.e.}$$

for any $\boldsymbol{\pi} \in \mathcal{T}\!\!\mathcal{T}_l \cap (0,1)^l$ and for any $f_{\gamma^1}, \dots, f_{\gamma^l} \in \mathcal{F}$ with $\boldsymbol{\gamma}^{\mathbf{1}}, \dots, \boldsymbol{\gamma}^{\boldsymbol{l}} \in \Gamma$ being distinct.

Remark 2.1 *$\mathcal{G}_p$ is p-weakly identifiable iff for any set of distinct points $\boldsymbol{\gamma}^{\mathbf{1}}, \dots, \boldsymbol{\gamma}^{\boldsymbol{l}} \in \Gamma$, $l \le p$, the corresponding densities $f_{\gamma^1}, \dots, f_{\gamma^l} \in \mathcal{F}$ are ν-a.e. linearly independent over the field of real numbers.*

Remark 2.1 is an implication of a theorem by Yakowitz and Spragins (1968) [YS68], p. 210, which says that "a necessary and sufficient condition that the class of *all* finite mixtures of the family $\tilde{\mathcal{F}}$ (of cumulative distribution functions) be identifiable is that $\tilde{\mathcal{F}}$ is a linearly independent set over the field of real numbers." Since the equation on the left hand side of (ID) holds ν-a.e. we can also use the theorem from [YS68] for our family $\mathcal{F}$ of densities. [YS68] also prove identifiability for some specific distribution families which generate identifiable mixtures. For instance, they show that finite arbitrary m-dimensional Gaussian mixtures are identifiable. Consequently, p population Gaussian mixtures are p-weakly identifiable.

Under (ID) (and (A1)) the true density g^0 still has several representations. [DCG99], p. 1180, reparameterize $\mathcal{G}_p$ with the aid of two parameters. One of the "new" parameters, essentially a positive real distance to $(\boldsymbol{\pi}^{\mathbf{0}}, \boldsymbol{\gamma}^{\mathbf{0},\mathbf{1}}, \dots, \boldsymbol{\gamma}^{\mathbf{0},\boldsymbol{q}})$, is chosen in such a way that g^0 is identifiable at that point characterizing the true density even if the other "new" parameter is not identifiable at that point. Thus the first parameter is the only one that is identifiable under the null hypothesis.

One fundamental assumption by [DCG99] for their theory is that there exists a "locally conic parameterization" of $\mathcal{G}_p$ through two parameters $\theta \in [0, M]$ and $\boldsymbol{\beta} \in \mathcal{B}$ with

$$\mathcal{G}_p \;=\; \left\{ g_{(\theta,\boldsymbol{\beta})} : \ (\theta, \boldsymbol{\beta}) \in \mathcal{T} \right\}, \quad \mathcal{T} \subset [0, M] \times \mathcal{B}, \tag{2.1}$$

where $M > 0$ and $\mathcal{B} \subset I\!R^{p(k+1)}$.

According to [DCG99], p. 1181, we give the following

Definition 2.2 (locally conic model)
Let $M > 0$, $\mathcal{B} \subset I\!R^{p(k+1)}$ be a Borel measurable and bounded set and $\mathcal{T} \subset [0, M] \times \mathcal{B}$ be endowed with the product topology of $I\!R$ and $\mathcal{B}$. Let $\overline{\mathcal{T}}$ be the (compact) closure of $\mathcal{T}$.

$\mathcal{G}_p$ given in (2.1) is called a locally conic model if

(LC1) $g_{(\theta,\boldsymbol{\beta})} = g^0 \quad \Leftrightarrow \quad \theta = 0$ *and*

(LC2) *For all $\boldsymbol{\beta} \in \mathcal{B}$ either $\theta_{\boldsymbol{\beta}} = \sup\Big\{t : \; [0, t] \times \{\boldsymbol{\beta}\} \subset \overline{\mathcal{T}}\Big\} > 0$ or there exists some $c > 0$ such that $[0, c] \times \{\boldsymbol{\beta}\} \cap \overline{\mathcal{T}} = \emptyset$.*

(LC1) implies that $g_{(\theta,\boldsymbol{\beta})}$ is identifiable for θ at $\theta = 0$ while it is not identifiable for $\boldsymbol{\beta}$ at $\theta = 0$ since $g_{(0,\boldsymbol{\beta})} = g^0$ for all $\boldsymbol{\beta} \in \mathcal{B}$. Let $(\theta_n, \boldsymbol{\beta_n})_{n \in I\!N}$ be any sequence in $\overline{\mathcal{T}}$ with $\lim_{n\to\infty} \theta_n = 0$ and let $\tilde{\boldsymbol{\beta}} = \boldsymbol{\beta}_{\boldsymbol{n}^0}$ for some $n^0 \in I\!N$. If $\theta_{\tilde{\boldsymbol{\beta}}} > 0$ then (LC2) says that the corresponding sequence of mixtures $\big(g_{(\theta_n,\tilde{\boldsymbol{\beta}})} : (\theta_n, \tilde{\boldsymbol{\beta}}) \in \overline{\mathcal{T}}\big)_{n \in I\!N}$ is defined in a right neighbourhood of $\theta = 0$. Hence, (under further suitable conditions) asymptotic expansions of $g_{(\theta,\boldsymbol{\beta})}$ with respect to θ are possible in $[0, \theta_{\boldsymbol{\beta}}) \neq \emptyset$ while using the right derivatives at $\theta = 0$.

Notice that we require measurability in the sense of Borel and boundedness of $\mathcal{B} \subset I\!R^{p(k+1)}$ for $\mathcal{G}_p$ being a locally conic model instead of compactness as [DCG99] do. The reason for this will become clear by the definition of $\mathcal{B}$ given in Proposition 2.10. In view of section 2.2 it is sufficient to ensure that $\overline{\mathcal{B}}$ is Borel measurable and compact.

2.1. Locally Conic Parameterization of the p Populations Mixture Model

In this section we introduce a reparameterization of $\mathcal{G}_p$ according to [DCG99], p. 1183-1184. We will define the new parameters in an explicit way while [DCG99] give an implicit definition of them. Though the two representations are equivalent, we think that an explicit presentation is easier to handle when calculating the asymptotic distribution. Although [DCG99], p. 1184, claim that their parameterization is locally conic by construction, we will show this with the aid of our modification (A1) and another modification (A2), namely that the parameter space Γ is convex in a specific way.

The reparameterization is directly connected to the index set of a Gaussian process, which forms the fundamental expression of the asymptotic distribution of a corresponding LRT statistic. In the introduction, p. 4, we mentioned that this index set is given by the directional score functions for $\theta = 0$ under an alternative parameterization of $\mathcal{G}_p$. In subsection 2.2.1 we will define this kind of functional class and specify its properties.

Firstly, we introduce the following notations for partial derivatives:

$\partial_{\gamma_{i_1}\cdots\gamma_{i_h}} f_{\boldsymbol{\gamma}}$ denotes the hth partial derivative of $f_{\boldsymbol{\gamma}}$ with respect to $\gamma_{i_1} \ldots \gamma_{i_h}$ and

$\partial_{\gamma_{i_1}\cdots\gamma_{i_h}} f_{\boldsymbol{\gamma}}|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,l}}$ denotes the hth partial derivative of $f_{\boldsymbol{\gamma}}$ with respect to $\gamma_{i_1} \ldots \gamma_{i_h}$ at the point $\boldsymbol{\gamma}^{\boldsymbol{0},\boldsymbol{l}}$.

The first central point in [DCG99] is to represent each density of $\mathcal{G}_p$ as a perturbation of the true density

$$g^0 = \sum_{l=1}^{q} \pi_l^0 f_{\gamma^{0,l}}$$

according to (A1). The reparameterization results from perturbing the true mixing weights as well as the true mixing component parameters and adding a perturbation as $(p-q)$-mixture. This leads to

$$g_{(\tilde{\theta},\boldsymbol{\beta})} = \sum_{i=1}^{p-q} \lambda_i \tilde{\theta} f_{\gamma^i} + \sum_{l=1}^{q} \left(\pi_l^0 + \rho_l \tilde{\theta}\right) f_{\gamma^{0,l}+\delta^l \tilde{\theta}}$$

for suitable $\tilde{\theta} \geq 0$ and $\boldsymbol{\beta} = (\lambda_1, \dots, \lambda_{p-q}, \boldsymbol{\gamma^1}, \dots, \boldsymbol{\gamma^{p-q}}, \boldsymbol{\delta^1}, \dots, \boldsymbol{\delta^q}, \rho_1, \dots, \rho_q)' \in I\!R^{p(k+1)}$ with

$$\begin{array}{rl} \lambda_i \geq 0,\ \boldsymbol{\gamma^i} \in \Gamma \subset I\!R^k & \text{for } i = 1, \dots, p-q, \\ \boldsymbol{\delta^l} \in I\!R^k,\ \rho_l \in I\!R & \text{for } l = 1, \dots, q \end{array} \tag{2.2}$$

and

$$\sum_{i=1}^{p-q} \lambda_i = -\sum_{l=1}^{q} \rho_l \geq 0. \tag{2.3}$$

Clearly, one has to ensure that the mixing weights $\pi_l^0 + \rho_l \tilde{\theta}$ are nonnegative and that $\boldsymbol{\gamma^{0,l}} + \boldsymbol{\delta^l} \tilde{\theta} \in \Gamma$ for all $l = 1, \dots, q$. Formula (2.3) assures that the sum of mixing weights is equal to 1 in the above representation of $g_{(\tilde{\theta},\boldsymbol{\beta})}$. By construction the parameter $\tilde{\theta}$ can be interpreted as a distance to the characteristic parameters $(\boldsymbol{\pi^0}, \boldsymbol{\gamma^{0,1}}, \dots, \boldsymbol{\gamma^{0,q}})$ of the true density and $\boldsymbol{\beta}$ represents a corresponding direction.

The second central point in [DCG99] is the assumption that

$$\mathcal{D} = \left\{ \frac{\partial_\theta g_{(\theta,\boldsymbol{\beta})}|_{\theta=0}}{g^0} : \boldsymbol{\beta} \in \mathcal{B} \right\} \tag{2.4}$$

is a so-called "Donsker class" of functions (see Definition 2.12) and the corresponding closure $\overline{\mathcal{D}}$ is a subset of the unit sphere in the Hilbert space $L_2(g^0\nu)$. This normalization is equivalent to the directional Fisher information being uniformly equal to 1. For this reason [DCG99] introduce a normalizing factor by $N(\boldsymbol{\beta}) = \|(\partial_{\tilde{\theta}} g_{(\tilde{\theta},\boldsymbol{\beta})}|_{\tilde{\theta}=0})/g^0\|_{L_2(g^0\nu)}$. Simple computations result in

$$N(\boldsymbol{\beta}) = \left\| \sum_{l=1}^{q} \pi_l^0 \sum_{i=1}^{k} \delta_i^l \frac{\partial_{\gamma_i} f_{\boldsymbol{\gamma}}|_{\boldsymbol{\gamma}=\boldsymbol{\gamma^{0,l}}}}{g^0} + \sum_{i=1}^{p-q} \lambda_i \frac{f_{\boldsymbol{\gamma^i}}}{g^0} + \sum_{l=1}^{q} \rho_l \frac{f_{\boldsymbol{\gamma^{0,l}}}}{g^0} \right\|_{L_2(g^0\nu)} \tag{2.5}$$

since it holds $\partial_{\tilde{\theta}} f_{\boldsymbol{\gamma^{0,l}}+\boldsymbol{\delta^l}\tilde{\theta}}|_{\tilde{\theta}=0} = \sum_{i=1}^{k} \delta_i^l \, \partial_{\gamma_i} f_{\boldsymbol{\gamma}}|_{\boldsymbol{\gamma}=\boldsymbol{\gamma^{0,l}}}$.

In order to avoid the cases $N(\boldsymbol{\beta}) = 0$ and $N(\boldsymbol{\beta}) = \infty$ (temporary) restrictions on the parameter space of $\boldsymbol{\beta}$ and on the component densities as well as on their partial derivatives are introduced:

- **Restrictions on the component densities and their derivatives:**

(P1 t) For any set of distinct points $\boldsymbol{\gamma^1}, \ldots, \boldsymbol{\gamma^{p-q}} \in \Gamma \backslash \{\boldsymbol{\gamma^{0,1}}, \ldots, \boldsymbol{\gamma^{0,q}}\} \subset I\!R^k$ the functions

$$\frac{f_{\boldsymbol{\gamma^i}}}{g^0}, \ \frac{f_{\boldsymbol{\gamma^{0,l}}}}{g^0}, \frac{\partial_{\gamma_1} f_{\boldsymbol{\gamma}}\big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma^{0,l}}}}{g^0}, \ldots, \frac{\partial_{\gamma_k} f_{\boldsymbol{\gamma}}\big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma^{0,l}}}}{g^0}, l = 1, \ldots, q, \ i = 1, \ldots, p-q$$

are linearly independent in $L_2(g^0\nu)$.

- **Restriction on the parameter space of $\boldsymbol{\beta}$:**
Let $(\lambda_1, \ldots, \lambda_{p-q}, \boldsymbol{\delta^1}, \ldots, \boldsymbol{\delta^q}, \rho_1, \ldots, \rho_q)'$ be defined on the unit sphere of $I\!R^{p+qk}$, i.e.

$$\sum_{i=1}^{p-q} (\lambda_i)^2 + \sum_{l=1}^{q} (\rho_l)^2 + \sum_{l=1}^{q} \|\boldsymbol{\delta^l}\|^2 \quad = \quad 1. \tag{2.6}$$

We define the (temporary) parameter space for

$$\boldsymbol{\beta} \ = \ (\lambda_1, \ldots, \lambda_{p-q}, \boldsymbol{\gamma^1}, \ldots, \boldsymbol{\gamma^{p-q}}, \boldsymbol{\delta^1}, \ldots, \boldsymbol{\delta^q}, \rho_1, \ldots, \rho_q)'$$

by

$$\tilde{\mathcal{B}} \ = \ \Big\{\boldsymbol{\beta} \in I\!R^{p(k+1)} : \ \boldsymbol{\beta} \text{ fulfills (2.2), (2.3), (2.6) and } \boldsymbol{\gamma^1}, \ldots, \boldsymbol{\gamma^{p-q}} \in \Gamma \backslash \{\boldsymbol{\gamma^{0,1}}, \ldots, \boldsymbol{\gamma^{0,q}}\}\Big\}. \tag{2.7}$$

If (A1) is satisfied and $\boldsymbol{\gamma^1}, \ldots, \boldsymbol{\gamma^{p-q}} \in \Gamma \backslash \{\boldsymbol{\gamma^{0,1}}, \ldots, \boldsymbol{\gamma^{0,q}}\}$ are distinct points then the p-weakly identifiability assumption (ID) is a necessary and sufficient condition for $\frac{f_{\boldsymbol{\gamma^i}}}{g^0}, \frac{f_{\boldsymbol{\gamma^{0,l}}}}{g^0}$, $l = 1, \ldots, q$, $i = 1, \ldots, p-q$ being linearly independent in $L_2(g^0\nu)$.

It follows directly from (P1 t) that $N(\boldsymbol{\beta})$ is finite for all $\boldsymbol{\beta} \in \tilde{\mathcal{B}}$ since linear combinations of functions in $L_2(g^0\nu)$ also belong to $L_2(g^0\nu)$. Furthermore, the linear independence in $L_2(g^0\nu)$ given by (P1 t) implies $N(\boldsymbol{\beta}) = 0 \Leftrightarrow (\lambda_1, \ldots, \lambda_{p-q}, \boldsymbol{\delta^1}, \ldots, \boldsymbol{\delta^q}, \rho_1, \ldots, \rho_q)' = \mathbf{0}$ since $\pi_l > 0$, $l = 1, \ldots, q$, according to (A1). By construction of $\tilde{\mathcal{B}}$ (i.e. thanks to (2.6)) we obtain

Lemma 2.3 *Let assumption* (A1) *and property* (P1 t) *be satisfied. Then one has*

$$0 < N(\boldsymbol{\beta}) < \infty \qquad \textit{for all } \boldsymbol{\beta} \in \tilde{\mathcal{B}}.$$

In Proposition 2.10 we will give a final definition of $\mathcal{B}$, $\mathcal{B} \subset \tilde{\mathcal{B}}$, ensuring that $\mathcal{G}_p$ in (2.1) is a locally conic model.

In view of subsection 2.2.2 it is of particular importance how fast $N(\boldsymbol{\beta_n})$ converges to 0 for a corresponding sequence $(\boldsymbol{\beta_n})_{n \in I\!N}$. A "suitable" convergence speed of $N(\boldsymbol{\beta_n})$ with limiting value 0 ensures that the remaining terms of an expansion of the likelihood function will converge uniformly in probability to 0. For this purpose, the temporary property (P1 t) will later be restricted appropriately.

Remark 2.4 *Let* (A1) *and property* (P1 t) *be satisfied. Let* $(\boldsymbol{\beta_n})_{n \in I\!N}$ *be a sequence in* $\tilde{\mathcal{B}}$ *and*

$$\begin{aligned} J(l) \quad &= \quad \Big\{i \in \{1, \ldots, p-q\} : f_{\boldsymbol{\gamma^{i,n}}} \overset{n \to \infty}{\longrightarrow} f_{\boldsymbol{\gamma^{0,l}}}\Big\}, \ l = 1, \ldots, q, \\ J \quad &= \quad \Big\{l \in \{1, \ldots, q\} : \ J(l) \neq \emptyset\Big\}. \end{aligned}$$

Then we have

$$\lim_{n\to\infty} N(\boldsymbol{\beta_n}) = 0 \quad\Leftrightarrow\quad \lim_{n\to\infty} \boldsymbol{\delta^{l,n}} = \mathbf{0} \text{ for all } l\in\{1,\dots,q\}, \tag{2.8}$$

$$\lim_{n\to\infty} \lambda_{i,n} = 0 \text{ for all } i\in\{1,\dots,p-q\}\backslash\big(J(1)\cup\ldots\cup J(q)\big), \tag{2.9}$$

$$\lim_{n\to\infty} \rho_{l,n} = 0 \text{ for all } l\in\{1,\dots,q\}\backslash J \textit{ and} \tag{2.10}$$

$$\lim_{n\to\infty}\Big(\sum_{u\in J(l)} \lambda_{u,n}\Big) = -\lim_{n\to\infty}\rho_{l,n} \text{ for all } l\in J. \tag{2.11}$$

Proof of Remark 2.4:
(P1 t) and (A1), i.e. $\pi_l > 0$ for $l = 1,\dots,q$, lead directly to (2.8). Correspondingly, (P1 t) leads to (2.10) and (2.11). According to (2.2) we have $\lambda_{u,n}\geq 0$ for all $u\in\{1,\dots,p-q\}\backslash\big(J(1)\cup\ldots\cup J(q)\big)$ which leads to (2.9) due to (P1 t). □

Under the conditions (A1) and (P1 t) we define a normalized reparameterization of $\mathcal{G}_p$ by

$$g_{(\theta,\boldsymbol{\beta})} \;=\; \sum_{i=1}^{p-q}\lambda_i\frac{\theta}{N(\boldsymbol{\beta})}f_{\boldsymbol{\gamma}^i} + \sum_{l=1}^{q}\Big(\pi_l^0+\rho_l\frac{\theta}{N(\boldsymbol{\beta})}\Big)f_{\boldsymbol{\gamma}^{0,l}+\boldsymbol{\delta}^l\frac{\theta}{N(\boldsymbol{\beta})}} \tag{2.12}$$

on a suitable parameter space $\mathcal{T}\subset[0,M]\times\mathcal{B}$ such that (2.1), (LC1) and (LC2) hold. (Simple computations show that (2.12) really leads to $\big\|\big(\partial_\theta g_{(\theta,\boldsymbol{\beta})}|_{\theta=0}\big)/g^0\big\|_{L_2(g^0\nu)} = 1$.)

To describe $g\in\mathcal{G}_p$ (which is not identifiable) by (2.12), we need an assignment to associate the "old" parameters of g with those of g^0, i.e. $\boldsymbol{\gamma^{0,1}},\dots,\boldsymbol{\gamma^{0,q}}$ and $\pi_0^1,\dots,\pi_0^q$. According to [DCG99] we define for any permutation σ of the index set $\{1,\dots,p\}$ the parameters both θ_σ and

$$\boldsymbol{\beta_\sigma} \;=\; (\lambda_{1,\sigma},\dots,\lambda_{p-q,\sigma},\boldsymbol{\gamma^{1,\sigma}},\dots,\boldsymbol{\gamma^{p-q,\sigma}},\boldsymbol{\delta^{1,\sigma}}\dots,\boldsymbol{\delta^{q,\sigma}},\rho_{1,\sigma},\dots,\rho_{q,\sigma})'$$

in such a way that $g_{(\theta_\sigma,\boldsymbol{\beta_\sigma})} = g$ holds. This assignment is not effected by the order of the succession of the "old" parameters .

Lemma 2.5 *Let assumption* (A1) *and property* (P1 t) *be satisfied. Let $\mathcal{G}_p$ be the mixture model of all p population mixtures of ν-densities $\mathcal{F} = \{f_{\boldsymbol{\gamma}} : \boldsymbol{\gamma}\in\Gamma\}$, $f_{\boldsymbol{\gamma}}$ continuous on Γ. Let $g\in\mathcal{G}_p\backslash\{g^0\}$ have a representation $g = g_{\boldsymbol{\pi},(\gamma^1,\dots,\gamma^p)}$. For any permutation σ of the index set $\{1,\dots,p\}$ we define*

$$R(\boldsymbol{\pi_\sigma},\boldsymbol{\gamma^{\sigma(1)}},\dots,\boldsymbol{\gamma^{\sigma(p)}}) \;=\; \Bigg(\sum_{i=1}^{p-q}(\pi_{\sigma(i)})^2+\sum_{l=1}^{q}(\pi_{\sigma(p-q+l)}-\pi_l^0)^2+\sum_{l=1}^{q}\|\boldsymbol{\gamma^{\sigma(p-q+l)}}-\boldsymbol{\gamma^{0,l}}\|^2\Bigg)^{\frac{1}{2}}. \tag{2.13}$$

Then $R(\boldsymbol{\pi_\sigma},\boldsymbol{\gamma^{\sigma(1)}},\dots,\boldsymbol{\gamma^{\sigma(p)}}) > 0$ holds. Further, we define $\boldsymbol{\beta_\sigma}$ by

$$\forall l = 1,\dots,p-q: \quad \boldsymbol{\gamma^{l,\sigma}} \;=\; \boldsymbol{\gamma^{\sigma(l)}}, \tag{2.14}$$

$$\forall l = 1,\dots,p-q: \quad \lambda_{l,\sigma} \;=\; \Big(R(\boldsymbol{\pi_\sigma},\boldsymbol{\gamma^{\sigma(1)}},\dots,\boldsymbol{\gamma^{\sigma(p)}})\Big)^{-1}\pi_{\sigma(l)}, \tag{2.15}$$

$$\forall l = 1,\dots,q: \quad \boldsymbol{\delta^{l,\sigma}} \;=\; \Big(R(\boldsymbol{\pi_\sigma},\boldsymbol{\gamma^{\sigma(1)}},\dots,\boldsymbol{\gamma^{\sigma(p)}})\Big)^{-1}\Big(\boldsymbol{\gamma^{\sigma(p-q+l)}}-\boldsymbol{\gamma^{0,l}}\Big), \tag{2.16}$$

$$\forall l = 1,\dots,q: \quad \rho_{l,\sigma} \;=\; \Big(R(\boldsymbol{\pi_\sigma},\boldsymbol{\gamma^{\sigma(1)}},\dots,\boldsymbol{\gamma^{\sigma(p)}})\Big)^{-1}\Big(\pi_{\sigma(p-q+l)}-\pi_l^0\Big) \tag{2.17}$$

and

$$\theta_\sigma = R(\boldsymbol{\pi_\sigma}, \boldsymbol{\gamma}^{\boldsymbol{\sigma(1)}}, \ldots, \boldsymbol{\gamma}^{\boldsymbol{\sigma(p)}}) N(\boldsymbol{\beta_\sigma}). \tag{2.18}$$

Then the following statements hold: $\boldsymbol{\beta_\sigma} \in \tilde{\mathcal{B}}$,

$$\theta_\sigma = \left\| \sum_{l=1}^{q} \pi_l^0 \sum_{i=1}^{k} \left(\gamma_i^{\sigma(p-q+l)} - \gamma_i^{0,l}\right) \frac{\partial_{\gamma_i} f_{\boldsymbol{\gamma}}|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,l}}}{g^0} + \sum_{l=1}^{p-q} \pi_{\sigma(i)} \frac{f_{\boldsymbol{\gamma}^{\sigma(i)}}}{g^0} + \sum_{l=1}^{q} \left(\pi_{\sigma(p-q+l)} - \pi_l^0\right) \frac{f_{\boldsymbol{\gamma}^{0,l}}}{g^0} \right\|_{L_2(g^0\nu)}, \tag{2.19}$$

$$\frac{\theta_\sigma}{N(\boldsymbol{\beta_\sigma})} \leq p + 2q \sup_{\boldsymbol{\gamma}\in\Gamma} \|\boldsymbol{\gamma}\|^2 \tag{2.20}$$

and $g_{(\theta_\sigma, \boldsymbol{\beta_\sigma})} = g$.

If one compares our explicit representation of $(\theta_\sigma, \boldsymbol{\beta_\sigma})$ for $g \neq g^0$ with the implicit one from [DCG99] (see here (2.22)-(2.24) as well as [DCG99], p. 1184) then it is obvious that both representations are equivalent. The reparameterization in [DCG99] is implicit because in defining the components of $\boldsymbol{\beta_\sigma}$ they use $N(\boldsymbol{\beta_\sigma})$ already. The advantage of our explicit description is that the components of $\boldsymbol{\beta_\sigma}$ do not depend on the normalizing factor $N(\boldsymbol{\beta_\sigma})$. In other words, the components of $\boldsymbol{\beta_\sigma}$ depend only on the component parameters and on the mixing weights but not on the corresponding component density itself.

Notice that we reparameterize $\mathcal{G}_p$ only for $g \neq g^0$ while [DCG99] do not mention this restriction. However, in the special case of $g = g^0$ the components of $\boldsymbol{\beta}$ (especially (2.15)-(2.17)) are not well-defined if (LC1) holds. In the next section we will additionally assume that $f_{\boldsymbol{\gamma}}$ possesses partial derivative up to order 5.

Proof of Lemma 2.5:
$R(\boldsymbol{\pi_\sigma}, \boldsymbol{\gamma}^{\boldsymbol{\sigma(1)}}, \ldots, \boldsymbol{\gamma}^{\boldsymbol{\sigma(p)}}) > 0$:
Assume that $R(\boldsymbol{\pi_\sigma}, \boldsymbol{\gamma}^{\boldsymbol{\sigma(1)}}, \ldots, \boldsymbol{\gamma}^{\boldsymbol{\sigma(p)}}) = 0$ then it follows $\pi_{\sigma(i)} = 0$ for $i = 1, \ldots, p-q$ and $\pi_{\sigma(p-q+l)} = \pi_l^0$, $\boldsymbol{\gamma}^{\boldsymbol{\sigma(p-q+l)}} = \boldsymbol{\gamma}^{\boldsymbol{0,l}}$ for $l = 1, \ldots, q$. This is in contradiction to the assumption $g \neq g^0$.

$\boldsymbol{\beta_\sigma} \in \tilde{\mathcal{B}}$:
(2.2) is a consequence of $R(\boldsymbol{\pi_\sigma}, \boldsymbol{\gamma}^{\boldsymbol{\sigma(1)}}, \ldots, \boldsymbol{\gamma}^{\boldsymbol{\sigma(p)}}) > 0$ and the definitions (2.15)-(2.17). From (2.15), (2.17) and $\boldsymbol{\pi_\sigma} \in \Pi_p$ it follows (2.3). Inserting (2.15), (2.16) and (2.17) in the left hand side of (2.6) results in

$$R(\boldsymbol{\pi_\sigma}, \boldsymbol{\gamma}^{\boldsymbol{\sigma(1)}}, \ldots, \boldsymbol{\gamma}^{\boldsymbol{\sigma(p)}})^{-2} \left(\sum_{i=1}^{p-q} (\pi_{\sigma(i)})^2 + \sum_{l=1}^{q} (\pi_{\sigma(p-q+l)} - \pi_l^0)^2 + \sum_{l=1}^{q} \|\boldsymbol{\gamma}^{\boldsymbol{\sigma(p-q+l)}} - \boldsymbol{\gamma}^{\boldsymbol{0,l}}\|^2 \right) \overset{(2.13)}{=} 1.$$

Thus (2.6) is satisfied.

Inserting the definitions of the components of $\boldsymbol{\beta_\sigma}$ (2.14)-(2.17) in the right hand side of the definition (2.5) of $N(\boldsymbol{\beta_\sigma})$, we obtain that equation (2.5) is equivalent to

$$N(\boldsymbol{\beta_\sigma}) = R(\boldsymbol{\pi_\sigma}, \boldsymbol{\gamma}^{\boldsymbol{\sigma(1)}}, \ldots, \boldsymbol{\gamma}^{\boldsymbol{\sigma(p)}})^{-1} \left\| \sum_{l=1}^{q} \pi_l^0 \sum_{i=1}^{k} \left(\gamma_i^{\sigma(p-q+l)} - \gamma_i^{0,l}\right) \frac{\partial_{\gamma_i} f_{\boldsymbol{\gamma}}|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,l}}}{g^0} + \sum_{i=1}^{p-q} \pi_{\sigma(l)} \frac{f_{\boldsymbol{\gamma}^{\sigma(i)}}}{g^0} + \sum_{l=1}^{q} \left(\pi_{\sigma(p-q+l)} - \pi_l^0\right) \frac{f_{\boldsymbol{\gamma}^{0,l}}}{g^0} \right\|_{L_2(g^0\nu)}. \tag{2.21}$$

Using the fact that $R(\boldsymbol{\pi_\sigma}, \boldsymbol{\gamma^{\sigma(1)}}, \ldots, \boldsymbol{\gamma^{\sigma(p)}})^{-1} = N(\boldsymbol{\beta_\sigma})/\theta_\sigma$ we obtain that (2.21) is equivalent to (2.19).

Formula (2.20) is a direct consequence of $\theta_\sigma/N(\boldsymbol{\beta_\sigma}) = R(\boldsymbol{\pi_\sigma}, \boldsymbol{\gamma^{\sigma(1)}}, \ldots, \boldsymbol{\gamma^{\sigma(p)}})$ and (2.13).

Inserting $(\theta_\sigma, \boldsymbol{\beta_\sigma})$ in formula (2.12) results in $g_{(\theta_\sigma, \boldsymbol{\beta_\sigma})} = g$. □

The following example shows that the above parameterization is not locally conic yet.

Example 2.6 *Let* (A1), (ID) *and* (P1 t) *be satisfied and let the true mixing density be* $g^0 = \frac{1}{4} f_{\gamma^{0,1}} + \frac{3}{4} f_{\gamma^{0,2}}$ *for* $(\gamma^{0,1}, \gamma^{0,2}) = (0, \frac{1}{2})$. *Let* $\boldsymbol{\pi} = (0, \frac{3}{4}, \frac{1}{4})'$ *and* $(\gamma^1, \gamma^2, \gamma^3) = (1, \frac{1}{2}, 0)$ *then* $g_{\boldsymbol{\pi}, (\gamma^1, \gamma^2, \gamma^3)} = 0 f_{\gamma^1} + \frac{3}{4} f_{\gamma^2} + \frac{1}{4} f_{\gamma^3} = g^0$. *Let the permutation* σ *of the index set* $\{1, 2, 3\}$ *be the identity.*

Using the aspects of Lemma 2.5 *we obtain*

$$R(\boldsymbol{\pi_\sigma}, \gamma^{1,\sigma}, \gamma^{2,\sigma}, \gamma^{3,\sigma}) \overset{(2.13)}{=} 1 \qquad \textit{and} \qquad \theta_\sigma \overset{(2.18)}{=} N(\boldsymbol{\beta_\sigma})$$

with $\boldsymbol{\beta_\sigma} = (\lambda_{1,\sigma}, \gamma^{1,\sigma}, \delta^{1,\sigma}, \delta^{2,\sigma}, \rho_{1,\sigma}, \rho_{2,\sigma})' = \left(0, 1, \frac{1}{2}, -\frac{1}{2}, \frac{1}{2}, -\frac{1}{2}\right)'$ *due to* (2.14)-(2.17). *Clearly,* $\boldsymbol{\beta_\sigma} \in \tilde{\mathcal{B}}$ *(see* (2.7)*). Remark* 2.4 *leads to* $N(\boldsymbol{\beta_\sigma}) > 0$. *Hence, we have* $\theta_\sigma > 0$ *although* $g_{(\theta_\sigma, \boldsymbol{\beta_\sigma})} = g^0$. *Thus condition* (LC1) *of a locally conic parameterization is not satisfied.*

The problem in Example 2.6 is that we choose a disadvantageous permutation. More precisely, we associate the component parameter γ^2 with $\gamma^{0,1}$ and γ^3 with $\gamma^{0,2}$ which have big distances, instead of associating γ^2 with $\gamma^{0,2}$ and γ^3 with $\gamma^{0,1}$ which have a small distances. There are several possibilities of choosing a promising permutation. A natural approach is given in [DCG99]. They obtain a permutation σ by associating step by step the nearest points $\boldsymbol{\gamma^l}$, $l = 1, \ldots, p$, of a given mixture g with the true parameters $\boldsymbol{\gamma^{0,1}}, \ldots, \boldsymbol{\gamma^{0,q}}$.

Algorithm 2.7 Choosing a permutation $\boldsymbol{\sigma}$ of $\{1, \ldots, p\}$:
For any mixture $g \in \mathcal{G}_p$ *we consider only those representations with truly involved component densities, i.e. which have a strictly positive mixing weight and appear only once in the weighted sum of component densities. Without loss of generality let* $\{1, \ldots, a\}$, $a \le p$, *be the corresponding index set with respect to the strictly positive mixing weights of* g. *Moreover, let* $\{i_1, \ldots, i_a\}$ *be a permutation of the index set* $\{1, \ldots, a\}$ *as well as* $\{l_1, \ldots, l_q\}$ *be a permutation of* $\{1, \ldots, q\}$, *such that*

$$\begin{aligned}
\|\boldsymbol{\gamma^{i_1}} - \boldsymbol{\gamma^{0,l_1}}\| &= \min_{\substack{l \in \{1,\ldots,q\} \\ i \in \{1,\ldots,a\}}} \|\boldsymbol{\gamma^{i}} - \boldsymbol{\gamma^{0,l}}\|, \\
\|\boldsymbol{\gamma^{i_2}} - \boldsymbol{\gamma^{0,l_2}}\| &= \min_{\substack{l \in \{1,\ldots,q\} \setminus \{l_1\} \\ i \in \{1,\ldots,a\} \setminus \{i_1\}}} \|\boldsymbol{\gamma^{i}} - \boldsymbol{\gamma^{0,l}}\|, \\
&\vdots \\
\|\boldsymbol{\gamma^{i_b}} - \boldsymbol{\gamma^{0,l_b}}\| &= \min_{\substack{l \in \{1,\ldots,q\} \setminus \{l_1,\ldots,l_{b-1}\} \\ i \in \{1,\ldots,a\} \setminus \{i_1,\ldots,i_{b-1}\}}} \|\boldsymbol{\gamma^{i}} - \boldsymbol{\gamma^{0,l}}\| \quad , \; b = \min\{a, q\}
\end{aligned}$$

and

$$\sum_{j=1}^{a} \|\boldsymbol{\gamma^{i_j}} - \boldsymbol{\gamma^{0,l_j}}\| \le \sum_{j=1}^{a} \|\boldsymbol{\gamma^{\tilde{\sigma}(j)}} - \boldsymbol{\gamma^{0,l_j}}\| \quad \textit{for any permutation } \tilde{\sigma} \textit{ of } \{1, \ldots, a\}$$

hold. Consequently, one has $\|\boldsymbol{\gamma^{i_1}} - \boldsymbol{\gamma^{0,l_1}}\| \le \|\boldsymbol{\gamma^{i_2}} - \boldsymbol{\gamma^{0,l_2}}\| \le \ldots \le \|\boldsymbol{\gamma^{i_b}} - \boldsymbol{\gamma^{0,l_b}}\|$.

We define the permutation σ of the index set $\{1,\ldots,p\}$ by

$$\sigma(p-q+l_j) \;=\; i_j \quad \textit{for } j=1,\ldots,b$$

and complete the permutation σ in some ordered way. We call $\sigma(g)$ a permutation with respect to g.

Notice that $\sigma(g)$ is not necessarily unique, but it will be seen in the proof of Proposition 2.10 that it leads to the identifiability of $g_{(\theta_{\sigma(g)},\boldsymbol{\beta}_{\boldsymbol{\sigma}(\boldsymbol{g})})}$ in $\theta_{\sigma(g)}=0$ which is equivalent to (LC1). In the same manner the parameter $\theta_{\sigma(g)}$ can be interpreted as distance to the characteristic parameter $(\boldsymbol{\pi}^{\mathbf{0}},\boldsymbol{\gamma}^{\mathbf{0,1}},\ldots,\boldsymbol{\gamma}^{\mathbf{0},\boldsymbol{q}})$ with $\theta_{\sigma(g)}=0$ iff $g=g^0$. But $\theta_{\sigma(g)}$ is not a metric distance since Algorithm 2.7 considers the distances of the component parameters only and not of the component densities of g.

In view of the definition of the "new" parameter space $\mathcal{T}$ (see Proposition 2.10) we introduce the following assumption which is not required in [DCG99]:

$$\text{(A2)} \qquad \forall\boldsymbol{\gamma}\in\Gamma:\ \mathcal{S}_{q,\boldsymbol{\gamma}}=\left\{t_{q+1}\boldsymbol{\gamma}+\sum_{l=1}^{q}t_l\boldsymbol{\gamma}^{\mathbf{0},\boldsymbol{l}}:\sum_{l=1}^{q+1}t_l=1 \text{ and } t_l\geq 0 \text{ for } l=1,\ldots,q+1\right\}\subset\Gamma.$$

Thus $\mathcal{S}_{q,\boldsymbol{\gamma}}$ is the convex hull of $\{\boldsymbol{\gamma},\boldsymbol{\gamma}^{\mathbf{0,1}},\ldots,\boldsymbol{\gamma}^{\mathbf{0},\boldsymbol{q}}\}$. The special case of $t_{q+1}=0$ leads to

Remark 2.8 *Let* (A2) *be satisfied. Then we obtain*

$$\mathcal{S}_q \;=\; \left\{\sum_{l=1}^{q}t_l\boldsymbol{\gamma}^{\mathbf{0},\boldsymbol{l}}:\sum_{l=1}^{q}t_l=1 \textit{ and } t_l\geq 0 \textit{ for } l=1,\ldots,q\right\}\subset\Gamma,$$

i.e. the convex hull of $\{\boldsymbol{\gamma}^{\mathbf{0,1}},\ldots,\boldsymbol{\gamma}^{\mathbf{0},\boldsymbol{q}}\}$ is also a subset of Γ under (A2).

The conditions of the following lemma are a sufficient condition for (LC2) of Definition 2.2 to hold.

Lemma 2.9 *Let* (A1), (A2) *and* (P1 t) *be satisfied. Let $\mathcal{G}_p$ be the mixture model of all p population mixtures of ν-densities $\mathcal{F}=\{f_{\boldsymbol{\gamma}}:\boldsymbol{\gamma}\in\Gamma\}$, $f_{\boldsymbol{\gamma}}$ continuous on Γ. Furthermore, let $g_{(\theta_{\sigma(g)},\boldsymbol{\beta}_{\boldsymbol{\sigma}(\boldsymbol{g})})}=g\in\mathcal{G}_p\backslash\{g^0\}$ be defined as in Lemma 2.5 with permutation $\sigma(g)$ of $\{1,\ldots,p\}$ given by Algorithm 2.7. Then $g_{(\theta,\boldsymbol{\beta}_{\boldsymbol{\sigma}(\boldsymbol{g})})}\in\mathcal{G}_p$ for all $\theta\in[0,\theta_{\sigma(g)}]$.*

Proof of Lemma 2.9: According to formula (2.12) one has for $(\theta,\boldsymbol{\beta}_{\boldsymbol{\sigma}})\in[0,\theta_\sigma]\times\tilde{\mathcal{B}}$

$$g_{(\theta,\boldsymbol{\beta}_{\boldsymbol{\sigma}(\boldsymbol{g})})} \;=\; \sum_{i=1}^{p-q}\lambda_{i,\sigma(g)}\frac{\theta}{N(\boldsymbol{\beta}_{\boldsymbol{\sigma}(\boldsymbol{g})})}f_{\boldsymbol{\gamma}^{i,\sigma(g)}}+\sum_{l=1}^{q}\left(\pi_l^0+\rho_{l,\sigma(g)}\frac{\theta}{N(\boldsymbol{\beta}_{\boldsymbol{\sigma}(\boldsymbol{g})})}\right)f_{\boldsymbol{\gamma}^{\mathbf{0},\boldsymbol{l}}+\boldsymbol{\delta}^{l,\sigma(g)}\frac{\theta}{N(\boldsymbol{\beta}_{\boldsymbol{\sigma}(\boldsymbol{g})})}}.$$

It is sufficient to verify the following statements for all $\theta\in[0,\theta_{\sigma(g)}]$:

i) $\left(\lambda_{1,\sigma(g)}\frac{\theta}{N(\boldsymbol{\beta}_{\boldsymbol{\sigma}(\boldsymbol{g})})},\ldots,\lambda_{p-q,\sigma(g)}\frac{\theta}{N(\boldsymbol{\beta}_{\boldsymbol{\sigma}(\boldsymbol{g})})},\pi_1^0+\rho_{1,\sigma(g)}\frac{\theta}{N(\boldsymbol{\beta}_{\boldsymbol{\sigma}(\boldsymbol{g})})},\ldots,\pi_q^0+\rho_{q,\sigma(g)}\frac{\theta}{N(\boldsymbol{\beta}_{\boldsymbol{\sigma}(\boldsymbol{g})})}\right)'\in\mathcal{\Pi}_p:$

$$\begin{aligned}
&\text{For} \quad \rho_{l,\sigma(g)} \in [0,1] \quad \text{one has} \quad \pi_l^0 + \rho_{l,\sigma(g)} \frac{\theta}{N(\boldsymbol{\beta}_{\sigma(g)})} \in \left[\pi_l^0, \pi_l^0 + \rho_{l,\sigma(g)} \frac{\theta_{\sigma(g)}}{N(\boldsymbol{\beta}_{\sigma(g)})}\right] \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \overset{(2.17),(2.18)}{=} \left[\pi_l^0, \pi_{\sigma(p-q+l)}\right] \subset (0,1]. \\
&\text{For} \quad \rho_{l,\sigma(g)} \in [-1,0) \quad \text{one has} \quad \pi_l^0 + \rho_{l,\sigma(g)} \frac{\theta}{N(\boldsymbol{\beta}_{\sigma(g)})} \in \left[\pi_l^0 + \rho_{l,\sigma(g)} \frac{\theta_{\sigma(g)}}{N(\boldsymbol{\beta}_{\sigma(g)})}, \pi_l^0\right) \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \overset{(2.17),(2.18)}{=} \left[\pi_{\sigma(p-q+l)}, \pi_l^0\right) \subset [0,1).
\end{aligned}$$

It follows from condition (2.2) that $\lambda_{i,\sigma(g)} \frac{\theta}{N(\boldsymbol{\beta}_{\sigma(g)})} \geq 0$ for all $i = 1, \ldots, p-q$. Moreover, $\sum_{i=1}^{p-q} \lambda_{i,\sigma(g)} \frac{\theta}{N(\boldsymbol{\beta}_{\sigma(g)})} + \sum_{l=1}^{p} \left(\pi_l^0 + \rho_{l,\sigma(g)} \frac{\theta}{N(\boldsymbol{\beta}_{\sigma(g)})}\right) \overset{(2.3)}{=} \sum_{l=1}^{p} \pi_l^0 = 1.$

ii) $\boldsymbol{\gamma}^{0,l} + \boldsymbol{\delta}^{l,\sigma(g)} \frac{\theta}{N(\boldsymbol{\beta}_{\sigma(g)})} \in \Gamma$ for $l = 1, \ldots, q$:

There exists some $t_l \in [0,1]$ such that $\theta = t_l \theta_{\sigma(g)}$. From (A2) follows

$$\boldsymbol{\gamma}^{0,l} + \boldsymbol{\delta}^{l,\sigma(g)} \frac{t_l \theta_{\sigma(g)}}{N(\boldsymbol{\beta}_{\sigma(g)})} = (1-t_l)\boldsymbol{\gamma}^{0,l} + t_l\left(\boldsymbol{\gamma}^{0,l} + \boldsymbol{\delta}^{l,\sigma(g)} \frac{\theta_{\sigma(g)}}{N(\boldsymbol{\beta}_{\sigma(g)})}\right) \in \Gamma. \quad \square$$

Proposition 2.10 *Let* (A1), (A2), (ID) *and* (P1 t) *be satisfied. Let $\mathcal{G}_p$ be the mixture model of all p population mixtures of ν-densities $\mathcal{F} = \{f_{\boldsymbol{\gamma}} : \boldsymbol{\gamma} \in \Gamma\}$, $f_{\boldsymbol{\gamma}}$ continuous on Γ. For $g \in \mathcal{G}_p$ let the permutation $\sigma(g)$ be given by Algorithm* 2.7. *We define*

$$\begin{aligned}
\mathcal{B} &= \left\{\boldsymbol{\beta}_{\sigma(g)} : g \in \mathcal{G}_p \backslash \{g^0\}\right\} \quad \textit{and} \\
\mathcal{T} &= \left\{(\theta, \boldsymbol{\beta}_{\sigma(g)}) : \theta \leq \theta_{\sigma(g)}, g \in \mathcal{G}_p \backslash \{g^0\}\right\}.
\end{aligned}$$

Then there exists some $M > 0$ such that $\mathcal{T} \subset [0,M] \times \mathcal{B}$ and $\left\{g_{(\theta,\boldsymbol{\beta})} : (\theta, \boldsymbol{\beta}) \in \mathcal{T}\right\}$ is a locally conic model of $\mathcal{G}_p$.

Notice that $\mathcal{B}$ may not be compact which leads us to the boundedness assumption of $\mathcal{B}$ in Definition 2.2. In the definition of $\mathcal{B}$ and $\mathcal{T}$, respectively, we consider only permutations $\sigma(g)$ for $g \in \mathcal{G}_p \backslash \{g^0\}$ instead of $g \in \mathcal{G}_p$ as [DCG99] do, p. 1184. As mentioned below Lemma 2.5, this restriction ensures that all $\boldsymbol{\beta} \in \mathcal{B}$ are well defined. Furthermore, it follows from the proof of Lemma 2.5 that $\mathcal{B} \subset \tilde{\mathcal{B}}$.

In the proof Proposition 2.10 will be seen that for any $\boldsymbol{\beta} \in \mathcal{B}$ one has $\theta_{\boldsymbol{\beta}} > 0$, i.e. the first condition of (LC2) holds for any $\boldsymbol{\beta} \in \mathcal{B}$ and we do need the second condition of (LC2). The reason for this is assumption (A2) which is not required in [DCG99]. As afore mentioned, p. 4, [DCG99] prove that the limiting distribution of the LRT statistic is essentially a function of the supremum of a Gaussian process on a class of functions. Additionally, [DCG99] say that these functions are indexed by the subset of $\mathcal{B}$ which is "the intersection of all directions approaching 0 in $\mathcal{T}$" (see [DCG99], p. 1184-1185). Thus they use directions fulfilling the first condition of (LC2) for defining the asymptotic distribution only. Our minor modification (A2) of Γ has considerable advantages when calculating the asymptotic distribution. Thanks to (A2) the Γ in the parameter set $\mathcal{B}$ is explicit given while it may be implicitly given in [DCG99].

Proof of Proposition 2.10:
Let $M = \sup\limits_{g \in \mathcal{G}_p \backslash \{g^0\}} \theta_{\sigma(g)}$. According to the representation (2.19) of $\theta_{\sigma(g)}$ we obtain with the aid

of the Minkowski inequality

$$
\begin{aligned}
M &\leq q \max_{l=1,\dots,q} \left\{ \left(\max_{\gamma\in\Gamma} \|\gamma - \gamma^{0,l}\| \right) \left\| \frac{\partial_{\gamma_i} f_\gamma|_{\gamma=\gamma^{0,l}}}{g^0} \right\|_{L_2(g^0\nu)} \right\} + (p-q) \max_{\gamma\in\Gamma} \left\| \frac{f_\gamma}{g^0} \right\|_{L_2(g^0\nu)} \\
&\quad + q \max_{l=1,\dots,q} \left\| \frac{f_{\gamma^{0,l}}}{g^0} \right\|_{L_2(g^0\nu)},
\end{aligned}
$$

since f_γ is continuous on Γ. From the compactness assumption (A1) for Γ and from property (P1 t), i.e. $\frac{\partial_{\gamma_i} f_\gamma|_{\gamma=\gamma^{0,l}}}{g^0}, \frac{f_\gamma}{g^0}, \frac{f_{\gamma^{0,l}}}{g^0} \in L_2(g^0\nu)$, it follows that $M < \infty$.
By construction we have

$$
\begin{aligned}
\mathcal{B} = \Big\{ \boldsymbol{\beta} \in I\!R^{p(k+1)} : \; & \boldsymbol{\beta} \text{ fulfills (2.2), (2.3), (2.6) and } \|\boldsymbol{\delta}^l\| \frac{\theta}{N(\boldsymbol{\beta})} \leq \|\boldsymbol{\gamma}^i - \boldsymbol{\gamma}^{0,l}\|, \\
& \text{for all } i = 1,\dots,p-q \text{ and for all } l = 1,\dots,q, \\
& \boldsymbol{\gamma}^1,\dots,\boldsymbol{\gamma}^{p-q} \in \Gamma \backslash \{\boldsymbol{\gamma}^{0,1},\dots,\boldsymbol{\gamma}^{0,q}\} \text{ being distinct} \Big\},
\end{aligned}
$$

with $\boldsymbol{\beta} = (\lambda_1,\dots,\lambda_{p-q},\boldsymbol{\gamma}^1,\dots,\boldsymbol{\gamma}^{p-q},\boldsymbol{\delta}^1,\dots,\boldsymbol{\delta}^q,\rho_1,\dots,\rho_q)'$. Each condition (2.2), (2.3), (2.6) describes a plain and unit sphere, respectively, which are of course measurable sets. For any $i = 1,\dots,p-q$ and any $l = 1,\dots,q$ each function $\|\boldsymbol{\gamma}^i - \boldsymbol{\gamma}^{0,l}\|$ is a composition of a projection of $\boldsymbol{\beta}$ on $\boldsymbol{\gamma}^i$, a difference to fixed $\boldsymbol{\gamma}^{0,l}$, a sum of squares and a square root, which are continuous functions. Hence $\{\boldsymbol{\beta} : \|\boldsymbol{\gamma}^i - \boldsymbol{\gamma}^{0,l}\|\}$ is a measurable set. According to Lemma 2.3 we have $N(\boldsymbol{\beta}_{\boldsymbol{\sigma}}) > 0$. $N(\boldsymbol{\beta}_{\boldsymbol{\sigma}})$ is a composition of linear combinations of fixed functions and $f_{\boldsymbol{\gamma}^i}$, $i = 1,\dots,p-q$, respectively, with coefficients being components of $(\lambda_1,\dots,\lambda_{p-q},\boldsymbol{\delta}^1,\dots,\boldsymbol{\delta}^q,\rho_1,\dots,\rho_q)$, which are projections of $\boldsymbol{\beta}$, and finally applying the $L_2(g^0\nu)$-norm. Thereby, is $\boldsymbol{\gamma}^i$ is also a projection of $\boldsymbol{\beta}$ and $f_{\boldsymbol{\gamma}^i}$ is continuous on Γ due to assumption. Thus $N(\boldsymbol{\beta})$ is a continuous function in $\boldsymbol{\beta}$. Furthermore, θ can also be described as function of $\boldsymbol{\beta}$ being a composition of continuous functions in $\boldsymbol{\beta}$, which we used before to describe $N(\boldsymbol{\beta})$. Thus we obtain that $\{\boldsymbol{\beta} : \|\boldsymbol{\delta}^l\| \frac{\theta}{N(\boldsymbol{\beta})} \leq \|\boldsymbol{\gamma}^i - \boldsymbol{\gamma}^{0,l}\|\}$ is a measurable set for $i = 1,\dots,p-q$ and any $l = 1,\dots,q$. As the intersection of all these measurable sets is non-empty, we obtain that $\mathcal{B}$ is a measurable set.

(LC1) Obviously, we have $g_{(0,\boldsymbol{\beta})} \overset{(2.12)}{=} g^0$ for all $\boldsymbol{\beta} \in \mathcal{B}$.

Suppose that there exists some $\boldsymbol{\beta} \in \mathcal{B}$ and $\theta > 0$ such that $g_{(\theta,\boldsymbol{\beta})} = g^0$. Then there also exists some $g \in \mathcal{G}_p$ such that $\boldsymbol{\beta} = \boldsymbol{\beta}_{\boldsymbol{\sigma}(\boldsymbol{g})}$ and $\theta \leq \theta_{\sigma(g)}$ for a corresponding permutation $\sigma(g)$ given by Algorithm 2.7. Let $t \in (0,1]$ such that $\theta = t\theta_{\sigma(g)}$ holds. According to Lemma 2.5 we obtain the following components of $\boldsymbol{\beta}_{\boldsymbol{\sigma}(\boldsymbol{g})}$:

$$
\forall l = 1,\dots,p-q: \quad \lambda_{l,\sigma(g)} = \frac{N(\boldsymbol{\beta}_{\boldsymbol{\sigma}(\boldsymbol{g})})}{\theta_{\sigma(g)}} \pi_{\sigma(g)(l)}, \tag{2.22}
$$

$$
\forall l = 1,\dots,q: \quad \boldsymbol{\delta}^{l,\boldsymbol{\sigma}(\boldsymbol{g})} = \frac{N(\boldsymbol{\beta}_{\boldsymbol{\sigma}(\boldsymbol{g})})}{\theta_{\sigma(g)}} \left(\boldsymbol{\gamma}^{\boldsymbol{\sigma}(\boldsymbol{g})(\boldsymbol{p}-\boldsymbol{q}+\boldsymbol{l})} - \boldsymbol{\gamma}^{0,l} \right) \text{ and} \tag{2.23}
$$

$$
\forall l = 1,\dots,q: \quad \rho_{l,\sigma(g)} = \frac{N(\boldsymbol{\beta}_{\boldsymbol{\sigma}(\boldsymbol{g})})}{\theta_{\sigma(g)}} \left(\pi_{\sigma(g)(p-q+l)} - \pi_l^0 \right). \tag{2.24}
$$

If $g_{(\theta_{\sigma(g)},\boldsymbol{\beta}_{\boldsymbol{\sigma}(\boldsymbol{g})})} = g^0$ then it follows by Algorithm 2.7 that $\boldsymbol{\gamma}^{\boldsymbol{\sigma}(\boldsymbol{g})(\boldsymbol{p}-\boldsymbol{q}+\boldsymbol{l})} = \boldsymbol{\gamma}^{0,l}$ for all $l = 1,\dots,q$. Since only truly involved component densities are used by the algorithm, we have $\pi_{\sigma(g)(p-q+l)} = \pi_l^0$ for all $l = 1,\dots,q$ and $\pi_{\sigma(g)(l)} = 0$ for all $l = 1,\dots,p-q$.

Hence, it follows $\boldsymbol{\delta}^{\boldsymbol{l},\boldsymbol{\sigma}(\boldsymbol{g})} = \mathbf{0}$ and $\rho_{l,\sigma(g)} = 0$ for all $l = 1,\ldots,q$ as well as $\lambda_{l,\sigma(g)} = 0$ for all $l = 1,\ldots,p-q$. Thus property (2.6) of $\mathcal{B}$ does not hold, which is in contradiction to $\boldsymbol{\beta} = \boldsymbol{\beta}_{\boldsymbol{\sigma}(\boldsymbol{g})} \in \mathcal{B}$.

Let $g_{(\theta_{\sigma(g)},\boldsymbol{\beta}_{\boldsymbol{\sigma}(\boldsymbol{g})})} \in \mathcal{G}_p\backslash\{g^0\}$. According to formula (2.12) $g_{(\theta,\boldsymbol{\beta}_{\boldsymbol{\sigma}(\boldsymbol{g})})}$ has a representation

$$\sum_{i=1}^{p-q}\lambda_{i,\sigma(g)}\frac{\theta}{N(\boldsymbol{\beta}_{\boldsymbol{\sigma}(\boldsymbol{g})})}f_{\boldsymbol{\gamma}^{\boldsymbol{i},\boldsymbol{\sigma}(\boldsymbol{g})}} + \sum_{l=1}^{q}\left(\pi_l^0 + \rho_{l,\sigma(g)}\frac{\theta}{N(\boldsymbol{\beta}_{\boldsymbol{\sigma}(\boldsymbol{g})})}\right)f_{\boldsymbol{\gamma}^{\boldsymbol{0},\boldsymbol{l}}+\boldsymbol{\delta}^{\boldsymbol{l},\boldsymbol{\sigma}(\boldsymbol{g})}\frac{\theta}{N(\boldsymbol{\beta}_{\boldsymbol{\sigma}(\boldsymbol{g})})}} = \sum_{l=1}^{q}\pi_l^0 f_{\boldsymbol{\gamma}^{\boldsymbol{0},\boldsymbol{l}}}. \tag{2.25}$$

As $\boldsymbol{\gamma}^{\boldsymbol{i},\boldsymbol{\sigma}(\boldsymbol{g})} \in \Gamma\backslash\{\boldsymbol{\gamma}^{\mathbf{0},\mathbf{1}},\ldots,\boldsymbol{\gamma}^{\mathbf{0},\boldsymbol{q}}\}$, $i = 1,\ldots,p-q$, (see (2.7)) we obtain $\lambda_{i,\sigma(g)} = 0$ for all $i = 1,\ldots,p-q$ due to assumption (ID). Furthermore, (ID) leads to

$$\left\{\boldsymbol{\gamma}^{\boldsymbol{0},\boldsymbol{l}} + \boldsymbol{\delta}^{\boldsymbol{l},\boldsymbol{\sigma}(\boldsymbol{g})}\frac{\theta}{N(\boldsymbol{\beta}_{\boldsymbol{\sigma}(\boldsymbol{g})})} : l = 1,\ldots,q\right\} = \left\{\boldsymbol{\gamma}^{\boldsymbol{0},\boldsymbol{l}} : l = 1,\ldots,q\right\}. \tag{2.26}$$

1. Case: $\boldsymbol{\delta}^{\boldsymbol{l},\boldsymbol{\sigma}(\boldsymbol{g})} = \mathbf{0}$ for all $l = 1,\ldots,q$.
   Then it follows from (2.25) that $\rho_{l,\sigma(g)} = 0$ for all $l = 1,\ldots,q$, which contradicts the assumption $\boldsymbol{\beta}_{\boldsymbol{\sigma}(\boldsymbol{g})} \in \mathcal{B}$ since (2.6) is not satisfied.
2. Case: There exist some $i,j \in \{1,\ldots,q\}$, $i \neq j$, such that $\boldsymbol{\delta}^{\boldsymbol{i},\boldsymbol{\sigma}(\boldsymbol{g})} \neq \mathbf{0}$ and

   $$\boldsymbol{\gamma}^{\boldsymbol{0},\boldsymbol{j}} = \boldsymbol{\gamma}^{\boldsymbol{0},\boldsymbol{i}} + \boldsymbol{\delta}^{\boldsymbol{i},\boldsymbol{\sigma}(\boldsymbol{g})}\frac{\theta}{N(\boldsymbol{\beta}_{\boldsymbol{\sigma}(\boldsymbol{g})})}. \tag{2.27}$$

   As $g_{(\theta_{\sigma(g)},\boldsymbol{\beta}_{\boldsymbol{\sigma}(\boldsymbol{g})})} \in \mathcal{G}_p\backslash\{g^0\}$ we have according to (2.16) and (2.18)

   $$\boldsymbol{\gamma}^{\boldsymbol{\sigma}(\boldsymbol{g})(\boldsymbol{p}-\boldsymbol{q}+\boldsymbol{i})} - \boldsymbol{\gamma}^{\boldsymbol{0},\boldsymbol{i}} = \boldsymbol{\delta}^{\boldsymbol{i},\boldsymbol{\sigma}(\boldsymbol{g})}\frac{\theta_{\sigma(g)}(1+t-t)}{N(\boldsymbol{\beta}_{\boldsymbol{\sigma}(\boldsymbol{g})})}. \tag{2.28}$$

   Using $\theta = t\theta_{\sigma(g)}$ and (2.27) lead to

   $$\begin{aligned} \boldsymbol{\gamma}^{\boldsymbol{\sigma}(\boldsymbol{g})(\boldsymbol{p}-\boldsymbol{q}+\boldsymbol{i})} - \boldsymbol{\gamma}^{\boldsymbol{0},\boldsymbol{i}} &= \boldsymbol{\delta}^{\boldsymbol{i},\boldsymbol{\sigma}(\boldsymbol{g})}\frac{\theta_{\sigma(g)}(1-t)}{N(\boldsymbol{\beta}_{\boldsymbol{\sigma}(\boldsymbol{g})})} + \boldsymbol{\gamma}^{\boldsymbol{0},\boldsymbol{j}} - \boldsymbol{\gamma}^{\boldsymbol{0},\boldsymbol{i}} \\ \Leftrightarrow \quad \boldsymbol{\gamma}^{\boldsymbol{\sigma}(\boldsymbol{g})(\boldsymbol{p}-\boldsymbol{q}+\boldsymbol{i})} - \boldsymbol{\gamma}^{\boldsymbol{0},\boldsymbol{j}} &= \boldsymbol{\delta}^{\boldsymbol{i},\boldsymbol{\sigma}(\boldsymbol{g})}\frac{\theta_{\sigma(g)}(1-t)}{N(\boldsymbol{\beta}_{\boldsymbol{\sigma}(\boldsymbol{g})})}. \end{aligned} \tag{2.29}$$

   Additionally, one has

   $$\begin{aligned} \|\boldsymbol{\gamma}^{\boldsymbol{\sigma}(\boldsymbol{g})(\boldsymbol{p}-\boldsymbol{q}+\boldsymbol{i})} - \boldsymbol{\gamma}^{\boldsymbol{0},\boldsymbol{i}}\| &\overset{(2.28)}{=} \frac{\theta_{\sigma(g)}}{N(\boldsymbol{\beta}_{\boldsymbol{\sigma}(\boldsymbol{g})})}\|\boldsymbol{\delta}^{\boldsymbol{i},\boldsymbol{\sigma}(\boldsymbol{g})}\| \\ &\overset{t\in(0,1]}{>} \frac{\theta_{\sigma(g)}(1-t)}{N(\boldsymbol{\beta}_{\boldsymbol{\sigma}(\boldsymbol{g})})}\|\boldsymbol{\delta}^{\boldsymbol{i},\boldsymbol{\sigma}(\boldsymbol{g})}\| \\ &\overset{(2.29)}{=} \|\boldsymbol{\gamma}^{\boldsymbol{\sigma}(\boldsymbol{g})(\boldsymbol{p}-\boldsymbol{q}+\boldsymbol{i})} - \boldsymbol{\gamma}^{\boldsymbol{0},\boldsymbol{j}}\| \end{aligned}$$

   which contradicts the property of $\sigma(g)$ given by Algorithm 2.7, i.e. it is in contradiction to $\|\boldsymbol{\gamma}^{\boldsymbol{\sigma}(\boldsymbol{g})(\boldsymbol{p}-\boldsymbol{q}+\boldsymbol{i})} - \boldsymbol{\gamma}^{\boldsymbol{0},\boldsymbol{i}}\| \leq \|\boldsymbol{\gamma}^{\boldsymbol{\sigma}(\boldsymbol{g})(\boldsymbol{p}-\boldsymbol{q}+\boldsymbol{i})} - \boldsymbol{\gamma}^{\boldsymbol{0},\boldsymbol{j}}\|$.

Hence, there does not exist some pair $\boldsymbol{\beta} \in \mathcal{B}$ and $\theta > 0$ such that $g_{(\theta,\boldsymbol{\beta})} = g^0$.

(LC2) follows directly from the definition of $\mathcal{T}$ and Lemma 2.9. □

2.2. Asymptotic Results

The asymptotic theory from [DCG99] is based on the properties (P0) and (P1) given below which are related to a family $\mathcal{F} = \{f_\gamma : \gamma \in \Gamma\}$ of probability densities (with respect to some σ-finite measure ν). Roughly speaking, property (P0) a) implies that the parameter θ can be consistently estimated and properties (P0) b) and c) say that the densities, which are involved in the likelihood function, possess partial derivatives up to the fifth order in a right neighbourhood of $\theta = 0$. [DCG99] use properties (P0) and (P1) to maximize an appropriate Taylor expansion of the log-likelihood function to find an asymptotic distribution of the LRT statistic. The limiting distribution is essentially given by a function of the supremum of a Gaussian process indexed by a functional class $\mathcal{D}$.

We give two different modifications (M1) and (M2) of (P1) only, though [DCG99], p. 1187, mention that "assumption (P0) is probably not optimal. It should be possible to prove the result using only derivatives up to order 3." But with regard to applying their theory to exponential families we are confronted with the problem that (P1) is not satisfied in "simple" cases such as mixtures of univariate Gaussian families when both parameter, mean and variance, are unknown. In contrast, it will be seen that in many cases it is possible to overcome difficulties with respect to property (P0) via restrictions on the parameter space.

Actually, the first modification consists of (P1 t) and (M1) while the second modification consists of (P1 t) and (M2). Both lead to an extension on the applicability of the theory from [DCG99] since we permit certain linearly dependencies between the partial derivatives of the first and second order, while the first order partial derivatives are assumed to be linearly independent in $L_2(g^0\nu)$ due to (P1 t). The first modification by (P1 t) and (M1) leads under (P0) to a limiting distribution of the LRT statistic for testing $g \in \mathcal{G}_q$ against $g \in \mathcal{G}_{q+1}\backslash\mathcal{G}_q$ with $Q = q \geq 1$, where $q = 1$ means testing one population against two populations. This modification is based on a k-dimensional parameter space. It assumes linear independence in $L_2(g^0\nu)$ of all true component densities, its corresponding first partial derivatives and a subset of second order partial derivatives to which, definitely, the diagonal entries of the appropriate Hesse matrices belong. In Example 3.19 we show that this modification holds for testing homogeneity against a two population bivariate Gaussian mixture with known mean and arbitrary unknown covariance matrix on an appropriate parameter space. For this kind of test problem property (P1) from [DCG99] is not satisfied. Notice that (P1) is a specific case of (P1 t) plus (M1) for testing one population against two populations. The second modification by (P1 t) and (M2) holds for testing one against two populations provided there exists a suitable 2-dimensional parameter space depending on the distances between the true component parameter vector and the component vectors. In Example 3.18 we show that it holds for testing homogeneity against two populations of univariate normal distributions with both unknown mean and variance (while (P1) from [DCG99] is not satisfied). Our results corroborate Garel's conjecture that an asymptotic distribution of the LRT statistic exists for this test, see Garel (2001). The verification of his conjecture with the help of the methods from [DCG99] can be carried out, only, if one restricts the parameter space to the two population mixture model.
The expressions for (M1) and (M2), respectively, have their origin in maintaining the concept of [DCG99] and are of a technical nature.

In Lemma 2.16 we identify the (compact) closure of the functional classes $\mathcal{D}$ with the aid of the corresponding modifications. In Proposition 2.18 we claim that $\overline{\mathcal{D}}$ under (P1 t) as well as (M1) and (M2), respectively, are so-called "Donsker classes". Moreover, we show in the proof of Lemma 2.20 that our modifications of (P1) leave the main idea of proof from [DCG99] unaffected and that they are sufficient conditions for the existence of the corresponding asymptotic distribution. Property (P1 t) assures that we could reprove the fundamental ideas of the theory from [DCG99]. For instance, the basic Lemma (see Lemma A.2 here, as well as [DCG99], Lemma 5.3) for deriving a limiting distribution of the LRT statistic with the method from [DCG99] would not hold. Finally, we show in Example 2.23 that Lemma 5.1 in [DCG99] does not hold if some true mixing weight is 0 which is possible according to the assumptions in [DCG99]. For this reason we introduced assumption (A1).

In view of later use we repeat assumptions (A1), (A2) and (ID) in order to show all assumptions, properties and modifications collectively.

(A1) $\Gamma \subset I\!R^k$ is compact, $\boldsymbol{\gamma}^{\mathbf{0,1}}, \ldots, \boldsymbol{\gamma}^{\mathbf{0,q}} \in \Gamma$ are distinct accumulation points of Γ, $\boldsymbol{\pi}^{\mathbf{0}} \in \mathit{\Pi}_q \cap (0,1)^q$ and g^0 has a representation $\sum\limits_{l=1}^{q} \pi_l^0 f_{\boldsymbol{\gamma}^{0,l}}$.

(A2) $\forall \boldsymbol{\gamma} \in \Gamma: \ \mathcal{S}_{q,\boldsymbol{\gamma}} = \left\{ t_{q+1}\boldsymbol{\gamma} + \sum\limits_{l=1}^{q} t_l \boldsymbol{\gamma}^{\mathbf{0,l}} : \sum\limits_{l=1}^{q+1} t_l = 1 \text{ and } t_l \geq 0 \text{ for } l = 1, \ldots, q+1 \right\} \subset \Gamma.$

(ID) $\sum\limits_{l=1}^{p} \pi_l^a f_{\boldsymbol{\gamma}^{a,l}} \quad = \quad \sum\limits_{l=1}^{p} \pi_l^b f_{\boldsymbol{\gamma}^{b,l}} \quad \nu\text{-a.e.} \quad \Leftrightarrow \quad \sum\limits_{l=1}^{p} \pi_l^a \mathbb{1}_{\boldsymbol{\gamma}^{a,l}} \quad = \quad \sum\limits_{l=1}^{p} \pi_l^b \mathbb{1}_{\boldsymbol{\gamma}^{b,l}}.$

Let us specify the properties given in [DCG99], p. 1185, as well as our modifications.

(P0) a) There exists a function h in $L_1(g^0\nu)$ such that for all $\boldsymbol{\gamma} \in \Gamma$, $\left|\log\left(f_{\boldsymbol{\gamma}}\right)\right| \leq h$ ν-a.e. holds.

b) $f_{\boldsymbol{\gamma}}$ possesses partial derivatives up to order 5 on Γ which are the right and left partial derivatives, respectively, on the boundary of Γ. For all $j \leq 5$ and all $i_1, \ldots, i_j \in \{1, \ldots, k\}$ one has

$$\frac{\partial_{\gamma_{i_1}} \ldots \partial_{\gamma_{i_j}} f_{\boldsymbol{\gamma}}\big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,l}}}{g^0} \in L_3(g^0\nu) \qquad , \ l = 1, \ldots, q.$$

c) For all $j \leq 5$ and all $i_1, \ldots, i_j \in \{1, \ldots, k\}$ there exists a function m_j and some $\epsilon > 0$ such that for all $l = 1, \ldots, q$

$$\sup_{\boldsymbol{\gamma} \in \overline{U}_\epsilon(\boldsymbol{\gamma}^{0,l})} \left| \frac{\partial_{\gamma_{i_1}} \ldots \partial_{\gamma_{i_j}} f_{\boldsymbol{\gamma}}}{g^0} \right| \leq m_{i_1,\ldots,i_j}$$

and $E_{g^0\nu}\left(m^3_{i_1,\ldots,i_j}\right) < \infty$ hold.
($\overline{U}_\epsilon(\boldsymbol{\gamma}^{\mathbf{0,l}}) = \{\boldsymbol{\gamma} \in \Gamma : \|\boldsymbol{\gamma} - \boldsymbol{\gamma}^{\mathbf{0,l}}\| \leq \epsilon\}$ is not empty since $\boldsymbol{\gamma}^{\mathbf{0,l}}$ is an accumulation point of Γ according to (A1).)

(P1) For any integer p_1 and p_2 such that $p_1 + p_2 \leq p - q$, for any set of distinct points $\boldsymbol{\gamma}^{\boldsymbol{l}} \in \Gamma \backslash \{\boldsymbol{\gamma}^{\mathbf{0,1}}, \ldots, \boldsymbol{\gamma}^{\mathbf{0,q}}\}$ with $l \in \{1, \ldots, p_1\}$ and any permutation σ of $\{1, \ldots, q\}$ the functions

$$\frac{f_{\boldsymbol{\gamma}^{i'}}}{g^0}, \ \frac{f_{\boldsymbol{\gamma}^{0,l}}}{g^0}, \ \frac{\partial_{\gamma_j} f_{\boldsymbol{\gamma}}\big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,l}}}{g^0}, \ \frac{\partial_{\gamma_i\gamma_j} f_{\boldsymbol{\gamma}}\big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,l'}}}{g^0},$$
$$i' \in \{1, \ldots, p_1\}, \ l' \in \{\sigma(1), \ldots, \sigma(p_2)\}, \ l = 1, \ldots, q, \ i,j = 1, \ldots, k \text{ with } i \leq j$$

are linearly independent in $L_2(g^0\nu)$. (For $p_1 = 0$ we define $\{1, \ldots, p_1\} = \emptyset$ and correspondingly we define $\{\sigma(1), \ldots, \sigma(p_2)\} = \emptyset$ for $p_2 = 0$.)

For testing $g \in \mathcal{G}_q$ against $g \in \mathcal{G}_{q+1} \backslash \mathcal{G}_q$ with $Q = q \geq 1$, we have $(p_1, p_2) \in \{(0,0),(0,1),(1,0)\}$ according to (P1). For the case of $p_2 = 0$ property (P1) is equivalent to (P1 t) and we modify only the case $(p_1, p_2) = (0,1)$ by (M1) and (M2), respectively.

(P1 t) For any set of distinct points $\boldsymbol{\gamma}^{\mathbf{1}}, \ldots, \boldsymbol{\gamma}^{\boldsymbol{p-q}} \in \Gamma \backslash \{\boldsymbol{\gamma}^{\mathbf{0,1}}, \ldots, \boldsymbol{\gamma}^{\mathbf{0,q}}\} \subset I\!R^k$ the functions

$$\frac{f_{\boldsymbol{\gamma}^i}}{g^0}, \ \frac{f_{\boldsymbol{\gamma}^{0,l}}}{g^0}, \frac{\partial_{\gamma_1} f_{\boldsymbol{\gamma}}\big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,l}}}{g^0}, \ldots, \frac{\partial_{\gamma_k} f_{\boldsymbol{\gamma}}\big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,l}}}{g^0} \ , l = 1, \ldots, q, \ i = 1, \ldots, p-q$$

are linearly independent in $L_2(g^0\nu)$.

(M1) For any $h \in \{1, \ldots, q\}$ let $U_h \subset \Big\{(i,j) \in \{1, \ldots, k\}^2 : i < j\Big\}$ be chosen in such a way that

$$\mathcal{H}_h = \left\{ \frac{f_{\boldsymbol{\gamma}^{0,l}}}{g^0}, \ \frac{\partial_{\gamma_i} f_{\boldsymbol{\gamma}}\big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,l}}}{g^0}, \ \frac{\partial_{\gamma_i\gamma_i} f_{\boldsymbol{\gamma}}\big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,h}}}{g^0}, \ \frac{\partial_{\gamma_{i'}\gamma_{j'}} f_{\boldsymbol{\gamma}}\big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,h}}}{g^0} : \right.$$
$$\left. l = 1, \ldots, q, \ i = 1, \ldots, k, \ (i', j') \in U_h \right\}$$

is a set of linearly independent functions in $L_2(g^0\nu)$ of maximal cardinality. For $h = 1, \ldots, q$ let $U_h^c = \Big\{(i,j) \in \{1, \ldots, k\}^2 : i < j, \ (i,j) \notin U_h\Big\}$ then

$$\mathcal{H}_h^c = \left\{ \frac{\partial_{\gamma_i\gamma_j} f_{\boldsymbol{\gamma}}\big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,h}}}{g^0} : \ (i,j) \in U_h^c \right\}$$

is a set of linearly independent functions in $L_2(g^0\nu)$ and for any $(u,v) \in U_h^c$ there exists exactly one $i \in \{1, \ldots, k\}$ such that

$$\partial_{\gamma_u\gamma_v} f_{\boldsymbol{\gamma}}\big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,h}} = c_i^h \ \partial_{\gamma_i\gamma_i} f_{\boldsymbol{\gamma}}\big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,h}} \quad \text{for some } c_i^h \neq 0 \tag{2.30}$$

with $(\min\{u,i\}, \max\{u,i\})$, $(\min\{v,i\}, \max\{v,i\}) \in U_h$.

Additionally, we define for any $h \in \{1, \ldots, q\}$ and $(u,v) \in U_h^c$ the index sets $H_h(u,v) = \Big\{i \in \{1, \ldots, k\} : \text{(2.30) is fulfilled}\Big\}$, $H_h = \bigcup_{(u,v) \in U_h^c} H_h(u,v)$ and $H_h^c = \{1, \ldots, k\} \backslash H_h$. (Then for any $(u,v) \in U_h^c$ the cardinality of $H_h(u,v)$ is equal to 1.)

Thereby, the condition $(\min\{u,i\}, \max\{u,i\})$, $(\min\{v,i\}, \max\{v,i\}) \in U_h$ is of technical nature and is used in the proof of Lemma 2.22 which is a fundamental working tool for statements about the convergence behaviour of the LRT statistic under our modification (P1 t) and (M1). Obviously, (P1 t) plus (M1) are equal to (P1) if $U_h^c = \emptyset$ for any $h \in \{1, \ldots, q\}$ holds.

As mentioned above (P1 t) and (M1) hold for testing homogeneity against a two population bivariate Gaussian mixture with known mean and arbitrary unknown covariance matrix on an appropriate parameter space, see Example 3.19. For this kind of test problem property (P1) from [DCG99] is not satisfied since for the corresponding densities $f_{(\sigma_1,\sigma_2,\varrho)}$ we have, for instance, $\partial_{\sigma_1}\partial_{\sigma_2} f_{(\sigma_1,\sigma_2,\varrho)}\big|_{(\sigma_1,\sigma_2,\varrho)'=(\sigma_1^0,\sigma_2^0,0)'} = \frac{1}{\sigma_1^0\sigma_2^0}\partial_\varrho\partial_\varrho f_{(\sigma_1,\sigma_2,\varrho)}\big|_{(\sigma_1,\sigma_2,\varrho)'=(\sigma_1^0,\sigma_2^0,0)'}$ which contradicts (P1).

For the specific case of testing $g \in \mathcal{G}_1 = \{f_{\boldsymbol{\gamma}} : \boldsymbol{\gamma} \in \Gamma\}$, $\Gamma \subset I\!R^2$, against $g \in \mathcal{G}_{2,M} \backslash \mathcal{G}_1$ with $\mathcal{G}_{2,M} = \left\{\pi f_{\boldsymbol{\gamma^1}} + (1-\pi) f_{\boldsymbol{\gamma^2}} : (\boldsymbol{\gamma^1}, \boldsymbol{\gamma^2})' \in \tilde{\Gamma},\, \pi \in [0,1]\right\}$ and specific $\tilde{\Gamma} \subset \Gamma \times \Gamma$, we give the following modified property

(M2) Let $\Gamma \subset I\!R^2$ such that the restricted two component parameter set

$$\tilde{\Gamma} = \left\{(\boldsymbol{\gamma^1}, \boldsymbol{\gamma^2})' \in \Gamma \times \Gamma : \|\boldsymbol{\gamma^1} - \boldsymbol{\gamma^{0,1}}\| \leq \|\boldsymbol{\gamma^2} - \boldsymbol{\gamma^{0,1}}\|,\ |\gamma_1^2 - \gamma_1^{0,1}| \leq |\gamma_2^2 - \gamma_2^{0,1}|\right\},$$

and the functions (parameterized on Γ)

$$\frac{f_{\boldsymbol{\gamma^{0,1}}}}{g^0}, \frac{\partial_{\gamma_1} f_{\boldsymbol{\gamma}}\big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma^{0,1}}}}{g^0}, \frac{\partial_{\gamma_2} f_{\boldsymbol{\gamma}}\big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma^{0,1}}}}{g^0}, \frac{\partial_{\gamma_1\gamma_2} f_{\boldsymbol{\gamma}}\big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma^{0,1}}}}{g^0} \text{ and } \frac{\partial_{\gamma_2\gamma_2} f_{\boldsymbol{\gamma}}\big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma^{0,1}}}}{g^0}$$

are linearly independent in $L_2(g^0\nu)$. Furthermore, there exists a constant $c \neq 0$ such that $c\, \frac{\partial_{\gamma_2} f_{\boldsymbol{\gamma}}\big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma^{0,1}}}}{g^0} = \frac{\partial_{\gamma_1\gamma_1} f_{\boldsymbol{\gamma}}\big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma^{0,1}}}}{g^0}$ holds.

As aforementioned (P1 t) and (M2) hold for testing homogeneity against two populations of univariate normal distributions with both parameters unknown, mean and variance, see Example 3.18. In this example $\boldsymbol{\gamma^2} = (\mu^2, \sigma^2)'$ describes the component parameter with the largest distance to the true parameter vector $(\mu^{0,1}, \sigma^{0,1})' = (0,1)'$ and $\tilde{\Gamma}$ depends on the distances $|\mu^2 - \mu^{0,1}| \leq |\sigma^2 - \sigma^{0,1}|$. For instance, if g^0 is the standardized normal density and $0 < \varepsilon = \|(\mu^1, \sigma^1)' - (0,1)'\|$ then Figure 2.1 shows the parameter space of $(\mu^2, \sigma^2)'$.

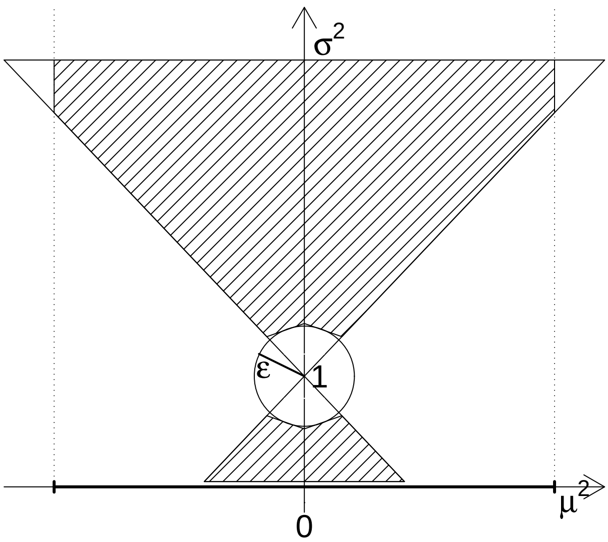

Figure 2.1.: $\{(\mu^2, \sigma^2)' \in \Gamma : \varepsilon \leq \|(\mu^2, \sigma^2)' - (0,1)'\|,\ |\mu^2 - \mu^{0,1}| \leq |\sigma^2 - \sigma^{0,1}|\}$

Notice that $\partial_\mu \partial_\mu f_{(\mu,\sigma)} = \frac{1}{\sigma} \partial_\sigma f_{(\mu,\sigma)}$ is in contradiction to (P1).

Similar to Proposition 2.10 we have

Lemma 2.11 *Let* $\Gamma \subset I\!R^2$ *and* $\mathcal{G}_{2,M} = \left\{\pi f_{\boldsymbol{\gamma^1}} + (1-\pi) f_{\boldsymbol{\gamma^2}} : (\boldsymbol{\gamma^1}, \boldsymbol{\gamma^2})' \in \tilde{\Gamma},\, \pi \in [0,1]\right\}$, $\tilde{\Gamma} \subset \Gamma \times \Gamma$, *be a mixture model of* 2 *population mixtures of* ν*-densities* $\mathcal{F} = \{f_{\boldsymbol{\gamma}} : \boldsymbol{\gamma} \in \Gamma\}$, $f_{\boldsymbol{\gamma}}$ *continuous on* Γ, *based on the restricted parameter space*

$$\tilde{\Gamma} = \left\{(\boldsymbol{\gamma^1}, \boldsymbol{\gamma^2})' \in \Gamma \times \Gamma : \|\boldsymbol{\gamma^1} - \boldsymbol{\gamma^{0,1}}\| \leq \|\boldsymbol{\gamma^2} - \boldsymbol{\gamma^{0,1}}\|,\ |\gamma_1^2 - \gamma_1^{0,1}| \leq |\gamma_2^2 - \gamma_2^{0,1}|\right\}.$$

Let (A1), (A2), (ID) *and* (P1 t) *be satisfied. For*

$$\tilde{\mathcal{B}}_{2,M} = \Big\{\boldsymbol{\beta} = (\lambda, \boldsymbol{\gamma}, \boldsymbol{\delta}, \rho)' \in I\!R^6 : \boldsymbol{\beta} \text{ fulfills } (2.2),\ (2.3),\ (2.6),\ \boldsymbol{\gamma} \in \Gamma \backslash \{\boldsymbol{\gamma}^{\mathbf{0,1}}\} \text{ with } (\boldsymbol{\gamma}^{\mathbf{0,1}} + \boldsymbol{\delta}\tfrac{\theta}{N(\boldsymbol{\beta})}, \boldsymbol{\gamma})' \in \tilde{\Gamma}\Big\}$$

let $g_{(\theta_{\sigma(g)},\boldsymbol{\beta}_{\boldsymbol{\sigma(g)}})} = g \in \mathcal{G}_{2,M}\backslash\{g^0\}$ *be defined as in Lemma* 2.5 *with permutation* $\sigma(g)$ *of* $\{1,2\}$ *given by Algorithm* 2.7. *We define*

$$\begin{aligned} \mathcal{B} &= \Big\{\boldsymbol{\beta}_{\boldsymbol{\sigma(g)}} : g \in \mathcal{G}_{2,M}\backslash\{g^0\}\Big\} \quad \textit{and} \\ \mathcal{T} &= \Big\{(\theta, \boldsymbol{\beta}_{\boldsymbol{\sigma(g)}}) : \theta \le \theta_{\sigma(g)}, g \in \mathcal{G}_{2,M}\backslash\{g^0\}\Big\}. \end{aligned}$$

Then there exists some $M > 0$ *such that* $\mathcal{T} \subset [0, M] \times \mathcal{B}$ *and* $\Big\{g_{(\theta,\boldsymbol{\beta})} : (\theta, \boldsymbol{\beta}) \in \mathcal{T}\Big\}$ *is a locally conic model of* $\mathcal{G}_{2,M}$.

Due to the parameterization in Lemma 2.5 (which uses Algorithm 2.7) $\boldsymbol{\beta} = (\lambda, \boldsymbol{\gamma}, \boldsymbol{\delta}, \rho) \in \tilde{\mathcal{B}}_{2,M}$ is equivalent to $\|(\boldsymbol{\gamma}^{\mathbf{0,1}} + \boldsymbol{\delta}\frac{\theta}{N(\boldsymbol{\beta})}) - \boldsymbol{\gamma}^{\mathbf{0,1}}\| \le \|\boldsymbol{\gamma} - \boldsymbol{\gamma}^{\mathbf{0,1}}\|$. Property $(\boldsymbol{\gamma}^{\mathbf{0,1}} + \boldsymbol{\delta}\frac{\theta}{N(\boldsymbol{\beta})}, \boldsymbol{\gamma})' \in \tilde{\Gamma}$ of $\tilde{\mathcal{B}}_{2,M}$ leads to $|\gamma_1 - \gamma_1^{0,1}| \le |\gamma_2 - \gamma_2^{0,1}|$, where $\{\boldsymbol{\beta} : |\gamma_1 - \gamma_1^{0,1}| \le |\gamma_2 - \gamma_2^{0,1}|\}$ is a measurable set. The proof of Lemma 2.11 follows the same lines as the proof of Proposition 2.10 and uses Lemma 2.9 which clearly holds for $\tilde{\mathcal{B}}_{2,M} \subset \tilde{\mathcal{B}}$.

We base our following work on the assumptions (A1), (A2) and (ID) as well as on the aforementioned definitions and propositions.

2.2.1. Donsker Classes

[DCG99] claim that under a locally conic parameterization and the regularity properties (P0) and (P1) it is possible to make a Taylor expansion of the log-likelihood function near the identifiable point $\theta = 0$. Afterwards they maximize the corresponding expansion, firstly with respect to the distance parameter θ and then with respect to the non-identifiable parameter $\boldsymbol{\beta}$. The resulting first term of the expansion maximized in θ is a sum of (square-integrable) score functions and can be interpreted as empirical process. It is necessary that the remaining terms of the expansion converge uniformly to 0 (in probability) to obtain an asymptotic result. Correspondingly, it is also necessary that the aforementioned empirical process has a uniform convergence behaviour. The latter leads to so-called "Donsker classes" (see Definition 2.12). Roughly speaking, a Donsker class is a set of square-integrable functions for which a corresponding empirical process satisfies a uniform version of the central limit theorem and converges in distribution to a centered Gaussian process (see e.g. van der Vaart and Wellner (1996), p. 81).

We begin this subsection with an outline of Donsker classes. After that we identify the (compact) closure of the corresponding set $\mathcal{D}$ of score functions resulting from our modifications (M1) and (M2), respectively. Finally, we verify that these classes of functions are Donsker classes (see Proposition 2.18).

According to Liu and Shao (2003), p. 819, we give the following

Definition 2.12 (*P*-Donsker class)
Let $(\Omega, \mathcal{A}, P)$ *be a probability space and let* $\mathcal{D} \subset \mathcal{L}_2(P)$ *be a family of* $\mathcal{A}$*-measurable, real-valued*

square-integrable functions for P. Moreover, let the $\mathcal{D}$-indexed empirical process $(\mathbb{G}_f^n)_{f\in\mathcal{D}}$ be given by

$$\mathbb{G}_f^n(\boldsymbol{X}) \quad = \quad \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big(f(\boldsymbol{X_i}) - Pf\Big) \text{ with } Pf = \int f\,dP,$$

where $\boldsymbol{X_1}, \boldsymbol{X_2}, \ldots$ are i.i.d. with law P and $\boldsymbol{X} = (\boldsymbol{X_1}, \boldsymbol{X_2}, \ldots)'$. Let $\big(\xi_f\big)_{f\in\mathcal{D}}$ be a centered Gaussian process on $(\Omega, \mathcal{A}, P)$ with covariance function

$$Cov_P\big(\xi_{f_1}, \xi_{f_2}\big) = P(f_1 - Pf_1)(f_2 - Pf_2) = Pf_1f_2 - Pf_1Pf_2$$

for $f_1, f_2 \in \mathcal{D}$.

$\mathcal{D}$ is called a P-Donsker class, if there exists a version $(\tilde{\mathbb{G}}_f^n)_{f\in\mathcal{D}}$ of $(\mathbb{G}_f^n)_{f\in\mathcal{D}}$ such that in probability,

$$\sup_{f\in\mathcal{D}} \big|\, \tilde{\mathbb{G}}_f^n - \xi_f \big| \quad \longrightarrow \quad 0$$

as n tends to infinity, and the Gaussian process $(\xi_f^n)_{f\in\mathcal{D}}$ has continuous paths with respect to the pseudometric $\rho_P(\cdot,\cdot)$ on $\mathcal{D}$ given by $\rho_P(f_1, f_2) = \Big(P(f_1 - f_2)^2 - (Pf_1 - Pf_2)^2\Big)^{\frac{1}{2}}$ for $f_1, f_2 \in \mathcal{D}$.

(In contrast to a metric a pseudometric $\rho_P(\cdot,\cdot)$ may vanish for different elements of a null set, i.e. $\rho_P(f_1, f_2) = 0$ for some $f_1 \neq f_2$.)

For further details see also Dudley (1999), p. 91-94.

Whether or not a given class $\mathcal{D}$ of square-integrable functions is a Donsker class depends on its size. One way to measure the size of $\mathcal{D}$ is given by Dudley (1999), p. 234,

Definition 2.13 (bracketing numbers, metric entropy with bracketing)
Let $(\Omega, \mathcal{A})$ be a measurable space and let $\mathcal{L}^0(\Omega, \mathcal{A})$ denote the set of all real-valued $\mathcal{A}$-measurable functions on Ω. For $f, g \in \mathcal{L}^0(\Omega, \mathcal{A})$ the set $[f, g] = \{h \in \mathcal{L}^0(\Omega, \mathcal{A}) : f \leq h \leq g\}$ will be called a bracket. Given a probability space $(\Omega, \mathcal{A}, P)$, $1 \leq q \leq \infty$, $\mathcal{D} \subset \mathcal{L}_q(P)$ (the family of $\mathcal{A}$-measurable, real-valued q-integrable functions for P) with usual seminorm $\|\cdot\|_{\mathcal{L}_q}$ and $\varepsilon > 0$, the bracketing number $N_{[\,]}^{(q)}(\varepsilon, \mathcal{D}, P)$ denotes the smallest integer m such that for some $f_1, \ldots, f_m$ and $g_1, \ldots, g_m$ in $\mathcal{L}_q(P)$ with $\|g_i - f_i\|_{\mathcal{L}_q} \leq \varepsilon$ for $i = 1, \ldots, m$,

$$\mathcal{D} \quad \subset \bigcup_{i=1}^{m} [f_i, g_i].$$

In this context $\log N_{[\,]}^{(q)}(\varepsilon, \mathcal{D}, P)$ will be called a metric entropy with bracketing.

In other words, the bracketing number is the minimum number of ε-brackets needed to cover $\mathcal{D}$.

In the definition of bracketing numbers, the upper and lower bounds f_i and g_i of the brackets are not required to belong to $\mathcal{D}$, but are assumed to have finite norms. According to Dudley (1999) (Theorem 7.2.1, p. 239) one has

Theorem 2.14 (M. Ossiander)
Let $(\Omega, \mathcal{A}, P)$ be a probability space and let $\mathcal{D} \subset \mathcal{L}_2(P)$ be such that

$$\int_0^1 \Big(\log N_{[\,]}^{(2)}(\varepsilon, \mathcal{D}, P) \Big)^{\frac{1}{2}} d\varepsilon \quad < \quad \infty.$$

Then $\mathcal{D}$ is a P-Donsker class.

One operation that preserves the Donsker property is for instance

Theorem 2.15 (van der Vaart and Wellner (1996), Theorem 2.10.1)
If $\mathcal{D}$ is Donsker and $\mathcal{K} \subset \mathcal{D}$, then $\mathcal{K}$ is Donsker.

In our work, the general form of the set of score functions is $\mathcal{D} = \left\{ \frac{\partial_\theta g_{(\theta,\beta)}|_{\theta=0}}{g^0} : \beta \in \mathcal{B} \right\}$. For testing Q populations against p populations under a true density g^0 of true order $q \leq Q$ the corresponding functional class has the form of

$$\mathcal{D} \;=\; \left\{ \frac{1}{N(\beta)} \left(\sum_{l=1}^{q} \pi_l^0 \sum_{i=1}^{k} \delta_i^l \frac{\partial_{\gamma_i} f_\gamma|_{\gamma=\gamma^{0,l}}}{g^0} + \sum_{i=1}^{p-q} \lambda_i \frac{f_{\gamma^i}}{g^0} + \sum_{l=1}^{q} \rho_l \frac{f_{\gamma^{0,l}}}{g^0} \right) : \; \beta \in \mathcal{B} \right\} \quad (2.31)$$

according to the representation (2.12) of $g_{(\theta,\beta)} \in \mathcal{G}_p$. The elements of $\mathcal{D}$ are given by functions $d_\beta : \mathfrak{X} \to I\!R$, $\mathfrak{X} \subset I\!R^m$.

[DCG99] prove that $\overline{\mathcal{D}}$ is a Donsker class under the assumptions (P0) and (P1) (see p. 1195-1196). (Notice, however, that in their corresponding Proposition 3.1 they do not assume (P1) and write $\mathcal{D}$ instead of $\overline{\mathcal{D}}$.) Their concept is firstly the identification of $\overline{\mathcal{D}}$ and secondly, the application of the Theorem of Ossiander. They find the closure of $\mathcal{D}$ by letting $p - q - p_1$ components $\boldsymbol{\gamma}^{i,n}$ of $\boldsymbol{\beta_n}$ converge to some of the true components $\boldsymbol{\gamma}^{0,l}$ for $n \to \infty$ (see p. 1195). Without giving any reasons they assume that $p_1 \leq p - q - 1$ components do not converge to some true component. This means that at least one component parameter $\boldsymbol{\gamma}^{i,n}$ of $\boldsymbol{\beta_n}$ converges to some true parameter for $n \to \infty$. If all true mixing weights are strictly positive (as assumed by (A1)) and (P1 t) is assumed to be satisfied, this may lead to $N(\boldsymbol{\beta_n}) = 0$, see Remark 2.4. But if some of the true mixing weights is 0, it may happen that no component parameter $\boldsymbol{\gamma}^{i,n}$ of $\boldsymbol{\beta_n}$ converges to some $\boldsymbol{\gamma}^{0,l}$ for $n \to \infty$. One way of dealing with this problem is to exclude such cases in Algorithm 2.7. (It will be seen in Example 2.23 that our assumption of strictly positive true mixing weights is necessary to obtain an asymptotic distribution with the method from [DCG99].)

We follow the concept from [DCG99] and identify the closures of the corresponding sets of score functions by

Lemma 2.16 *Let* (A1), (A2) *and* (ID) *be satisfied. Let $\overline{\mathcal{D}}$ denote the (compact) closure of $\mathcal{D}$.*

a) *Let $Q = q$, $p = q + 1$ and let* (P0), (P1 t) *and* (M1) *be satisfied. Then $\overline{\mathcal{D}}$ is given by* *functions of the form*

$$\begin{aligned}
&\sum_{l=1}^{p_1} \mu_l \frac{f_{\gamma^l}}{g^0} + \sum_{l=1}^{q} \tilde{\rho}_l \frac{f_{\gamma^{0,l}}}{g^0} + \sum_{l=1}^{q} \sum_{i=1}^{k} \lambda_{l,i} \frac{\partial_{\gamma_i} f_{\gamma}|_{\gamma=\gamma^{0,l}}}{g^0} \\
+ & \sum_{l=1}^{p_2} \frac{1}{\sqrt{2}} \left(\sum_{(i,j)\in U_h} \alpha_i^h \alpha_j^h \frac{\partial_{\gamma_i\gamma_j} f_{\gamma}|_{\gamma=\gamma^{0,h}}}{g^0} + \frac{1}{2} \sum_{i \in H_h^c} (\alpha_i^h)^2 \frac{\partial_{\gamma_i\gamma_i} f_{\gamma}|_{\gamma=\gamma^{0,h}}}{g^0} \right. \\
& \left. + \sum_{(u,v)\in U_h^c} \sum_{i\in H_h(u,v)} \left(c_i^h \alpha_u^h \alpha_v^h + \frac{1}{2} (\alpha_i^h)^2 \right) \frac{\partial_{\gamma_i\gamma_i} f_{\gamma}|_{\gamma=\gamma^{0,h}}}{g^0} \right) \qquad (2.32)
\end{aligned}$$

for some $h \in \{1, \ldots, q\}$ *and* $c_i^h \neq 0$, $p_1, p_2 \in \{0, 1\}$ *such that* $p_1 + p_2 \leq 1$, $\mu_l \geq 0$ *and* $\sum_{l=1}^{p_1} \mu_l + \sum_{l=1}^{q} \tilde{\rho}_l = 0$ *(and cardinality of* $H_h(u, v)$ *equal to one).*
Furthermore, $\overline{\mathcal{D}}$ *is a subset of the unit sphere in the Hilbert space* $L_2(g^0\nu)$.

b) *Let* $Q = q = 1$, $p = 2$ *and let* (P0), (P1 t) *and* (M2) *be satisfied. Let* $\mathcal{G}_1 = \{f_{\boldsymbol{\gamma}} : \boldsymbol{\gamma} \in \Gamma\}$, $\Gamma \subset I\!R^2$, *and* $\mathcal{G}_{2,M} = \left\{\pi f_{\boldsymbol{\gamma}^1} + (1 - \pi) f_{\boldsymbol{\gamma}^2} : (\boldsymbol{\gamma}^1, \boldsymbol{\gamma}^2)' \in \tilde{\Gamma}, \pi \in [0, 1]\right\}$, $\tilde{\Gamma} \subset \Gamma \times \Gamma$, *and* $\mathcal{B} = \{\boldsymbol{\beta}_{\boldsymbol{\sigma}(g)} : g \in \mathcal{G}_{2,M} \backslash \{f_{\boldsymbol{\gamma}^{0,1}}\}\}$ *from Lemma* 2.11. *Then* $\overline{\mathcal{D}}$ *is given by functions of the form*

$$\begin{aligned}
&\sum_{l=1}^{p_1} \mu_l \frac{f_{\gamma^l}}{g^0} + \tilde{\rho}_1 \frac{f_{\gamma^{0,1}}}{g^0} + \sum_{i=1}^{2} \lambda_{1,i} \frac{\partial_{\gamma_i} f_{\gamma}|_{\gamma=\gamma^{0,1}}}{g^0} \\
+ & \sum_{l=1}^{p_2} \frac{1}{\sqrt{2}} \left(\alpha_1 \alpha_2 \frac{\partial_{\gamma_1\gamma_2} f_{\gamma}|_{\gamma=\gamma^{0,1}}}{g^0} + \frac{1}{2} (\alpha_2)^2 \frac{\partial_{\gamma_2\gamma_2} f_{\gamma}|_{\gamma=\gamma^{0,1}}}{g^0} \right) \qquad (2.33)
\end{aligned}$$

with $p_1, p_2 \in \{0, 1\}$ *such that* $p_1 + p_2 \leq 1$, $\mu_l \geq 0$ *and* $\sum_{l=1}^{p_1} \mu_l + \tilde{\rho}_1 = 0$.
Furthermore, $\overline{\mathcal{D}}$ *is a subset of the unit sphere in the Hilbert space* $L_2(g^0\nu)$.

Please note, that in (2.32) as well as in (2.33) the corresponding normalizing factor is included in the coefficients of the functions. Before we give the proof of Lemma 2.16 we introduce

Lemma 2.17 *Let the functions* $h_1, \ldots, h_r$ *be linearly independent in* $L_2(g^0\nu)$ *and let* $\boldsymbol{z} \in I\!R^r$. *Then there exist some* $c, c' > 0$ *such that*

$$c' \,\|\boldsymbol{z}\| \quad \geq \quad \left\| \sum_{i=1}^{r} z_i h_i \right\|_{L_2(g^0\nu)} \quad \geq \quad c\,\|\boldsymbol{z}\| \quad \geq \quad 0, \qquad (2.34)$$

where $\|\cdot\|$ *denotes the Euclidean norm. For* $\boldsymbol{z} \neq \boldsymbol{0}$ *the expressions in* (2.34) *are strictly positive.*

Lemma 2.17 is often used as a basic tool for evaluating a $L_2(g^0\nu)$-norm with the aid of an Euclidean norm.

Proof of Lemma 2.17:

$A = \left(\langle h_i, h_j \rangle_{L_2(g^0\nu)}\right)_{i,j=1,\ldots,r}$ and $\boldsymbol{z} \in I\!R^r$ lead to $\left\| \sum_{i=1}^{r} z_i h_i \right\|_{L_2(g^0\nu)} = \left(\boldsymbol{z}' A \boldsymbol{z}\right)^{\frac{1}{2}} \geq 0$ since the scalar product in the Hilbert space $L_2(g^0\nu)$ is positive definite. Consequently, the (symmetric) matrix A is also positive definite and all its corresponding eigenvalues reordered by $\eta_{min} =$

$\eta_1 \ldots \leq \eta_r = \eta_{max}$ are strictly positive. The assertion follows from the theorem by Rayleigh-Reiz (see Horn and Johnson (1985), p. 176) which says that $\eta_1 \boldsymbol{z'z} \leq \boldsymbol{z'}A\boldsymbol{z} \leq \eta_r \boldsymbol{z'z}$ holds for all $\boldsymbol{z} \in \mathbb{R}^r$. □

As usual we introduce the Landau order symbols for asymptotic behaviour of functions by

$$\begin{aligned} O(f) &= \Big\{ g : \mathbb{N} \to \mathbb{R}_+ : \ \exists c > 0 \ \exists n_0 \in \mathbb{N} \ \forall n \geq n_0 \text{ one has } g(n) \leq cf(n) \Big\}, \\ o(f) &= \Big\{ g : \mathbb{N} \to \mathbb{R}_+ : \ \lim_{n\to\infty} \frac{g(n)}{f(n)} = 0 \Big\}. \end{aligned}$$

Proof of Lemma 2.16:

a) Firstly, we search the set of all possible accumulation points of $\mathcal{D}$ in the Hilbert space $L_2(g^0\nu)$. According to (2.31)(which was derived by (A1), (A2), (ID), (P1 t)) the functions in $\mathcal{D}$ have the form

$$d_{\boldsymbol{\beta}} = \frac{1}{N(\boldsymbol{\beta})} \left(\sum_{l=1}^{q} \pi_l^0 \sum_{i=1}^{k} \delta_i^l \frac{\partial_{\gamma_i} f_{\boldsymbol{\gamma}}|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,l}}}{g^0} + \lambda_1 \frac{f_{\boldsymbol{\gamma}^1}}{g^0} + \sum_{l=1}^{q} \rho_l \frac{f_{\boldsymbol{\gamma}^{0,l}}}{g^0} \right) \text{ for } \boldsymbol{\beta} \in \mathcal{B}. \quad (2.35)$$

Then $d_{\boldsymbol{\beta}}$ reads as $d_{\boldsymbol{\beta}}^{num} \|d_{\boldsymbol{\beta}}^{num}\|^{-1}_{L_2(g^0\nu)}$ with $N(\boldsymbol{\beta}) = \|d_{\boldsymbol{\beta}}^{num}\|_{L_2(g^0\nu)}$. $d_{\boldsymbol{\beta}}^{num}$ is continuous in $\boldsymbol{\beta} = (\lambda_1, \boldsymbol{\gamma^1}, \boldsymbol{\delta^1}, \ldots, \boldsymbol{\delta^q}, \rho_1, \ldots, \rho_q)'$ on $\mathcal{B}$ since $f_{\boldsymbol{\gamma}}$ is continuous in $\boldsymbol{\gamma}$, because it possesses partial derivatives up to order 5 on Γ according to (P0)b), and linear combinations of continuous functions (due to (P0)) are also continuous. As the $L_2(g^0\nu)$-norm is also a continuous function, we have $d : \mathcal{B} \to \mathcal{D}$ with $\boldsymbol{\beta} \mapsto d_{\boldsymbol{\beta}}$ is a continuous function. We investigate the sequences $(d_{\boldsymbol{\beta_n}})_{n\in\mathbb{N}}$ with $\lim_{n\to\infty} \boldsymbol{\beta_n} = \tilde{\boldsymbol{\beta}} \in \overline{\mathcal{B}}\backslash\mathcal{B}$ for $\lim_{n\to\infty} N(\boldsymbol{\beta_n}) > 0$ and $\lim_{n\to\infty} N(\boldsymbol{\beta_n}) = 0$, respectively. In the first case we obtain $\lim_{n\to\infty} d_{\boldsymbol{\beta_n}} = d_{\tilde{\boldsymbol{\beta}}}$ which has a form according to (2.35). In the second case we use Remark 2.4 to describe the possible accumulation points of $\mathcal{D}$ in $L_2(g^0\nu)$ with the aid of the corresponding accumulation points $\tilde{\boldsymbol{\beta}} = (\tilde{\lambda}_1, \tilde{\boldsymbol{\gamma}}^{\boldsymbol{1}}, \tilde{\boldsymbol{\delta}}^{\boldsymbol{1}}, \ldots, \tilde{\boldsymbol{\delta}}^{\boldsymbol{q}}, \tilde{\rho}_1, \ldots, \tilde{\rho}_q)' \in \overline{\mathcal{B}}\backslash\mathcal{B}$:

i) Suppose that $\boldsymbol{\gamma^{1,n}} \to \tilde{\boldsymbol{\gamma}}^{\boldsymbol{1}} \in \Gamma\backslash\{\boldsymbol{\gamma^{0,1}}, \ldots, \boldsymbol{\gamma^{0,q}}\}$ for $n \to \infty$.
It follows from Remark 2.4 that $\tilde{\lambda}_1 = 0$, $\tilde{\rho}_l = 0$ and $\tilde{\boldsymbol{\delta}}^{\boldsymbol{l}} = \boldsymbol{0}$ for all $l = 1, \ldots q$. However, this contradicts property (2.6) of $\mathcal{B}$.

ii) Suppose that there exists some $h \in \{1, \ldots, q\}$ such that $\boldsymbol{\gamma^{1,n}} \to \boldsymbol{\gamma^{0,h}}$ for $n \to \infty$. Without loss of generality let the permutation σ of the index set $\{1, \ldots, q+1\}$ be the identity. Then there exists some $n_0 \in \mathbb{N}$ such that for any $n \geq n_0$ we have $\|\boldsymbol{\gamma^{1,n}} - \boldsymbol{\gamma^{0,h}}\| \geq \|\boldsymbol{\gamma^{l+1,n}} - \boldsymbol{\gamma^{0,l}}\|$, $l = 1, \ldots, q$, due to Algorithm 2.7. Due to the assumption $\boldsymbol{\gamma^{1,n}} \to \boldsymbol{\gamma^{0,h}}$ it follows $\|\boldsymbol{\gamma^{1,n}} - \boldsymbol{\gamma^{0,h}}\| \to 0$ for $n \to \infty$. Additionally, it follows $\boldsymbol{\delta^{l,n}} \to \boldsymbol{0}$ for $l = 1, \ldots, q$ and $n \to \infty$, see (2.16). As a consequence we have $\tilde{\boldsymbol{\delta}}^{\boldsymbol{l}} = \boldsymbol{0}$ for $l = 1, \ldots, q$. Moreover, it follows from (2.10) that $\tilde{\rho}_l = 0$ for $l \in \{1, \ldots q\}\backslash\{h\}$ and from (2.11) as well as from (2.6) that $\tilde{\lambda}_1 = -\tilde{\rho}_h = \frac{1}{\sqrt{2}}$. This implies that

$$\tilde{\boldsymbol{\beta}}^{\boldsymbol{h}} = \Big(\frac{1}{\sqrt{2}}, \boldsymbol{\gamma^{0,h}}, \boldsymbol{0}, \ldots, \boldsymbol{0}, \underbrace{0, \ldots, 0}_{h-1}, -\frac{1}{\sqrt{2}}, \underbrace{0, \ldots, 0}_{q-h} \Big)' \in \overline{\mathcal{B}}\backslash\mathcal{B}.$$

As we would like to describe the possible accumulation points on $\mathcal{D}$, we are interested in the form of $\lim_{n\to\infty} d_{\boldsymbol{\beta^h_n}}$ for some $h \in \{1, \ldots, q\}$ using the fact that $d : \mathcal{B} \to \mathcal{D}$ is

continuous (as mentioned above) and $\lim_{n\to\infty} \boldsymbol{\beta_n^h} = \tilde{\boldsymbol{\beta}}^{\boldsymbol{h}}$. Similar to [DCG99], p. 1195, we make a Taylor expansion of the numerator $d^{num}_{\boldsymbol{\beta_n^h}}$ of $d_{\boldsymbol{\beta_n^h}}$ up to the second order at the point $\tilde{\boldsymbol{\beta}}^{\boldsymbol{h}}$. Since the denominator of $d_{\boldsymbol{\beta_n^h}}$ is given by $\|d^{num}_{\boldsymbol{\beta_n^h}}\|_{L_2(g^0\nu)}$ we make the same expansion with the argument of the $L_2(g^0\nu)$-norm. Afterward we verify that the resulting expression of $d_{\boldsymbol{\beta_n^h}}$ has bounded coefficients with respect to the linearly independent functions given by (P1 t) and (M1).
For $h = 1,\dots,q$ we obtain

$$\begin{aligned}
d^{num}_{\boldsymbol{\beta_n^h}} &= \Big(\lambda_{1,n} + \rho_{h,n}\Big)\frac{f_{\boldsymbol{\gamma^{0,h}}}}{g^0} + \sum_{\substack{l=1\\ l\neq h}}^{q} \rho_{l,n}\frac{f_{\boldsymbol{\gamma^{0,l}}}}{g^0} \\
&+ \sum_{i=1}^{k}\Big(\lambda_{1,n}\Big(\gamma_i^{1,n} - \gamma_i^{0,h}\Big) + \pi_h^0\delta_i^{h,n}\Big)\frac{\partial_{\gamma_i} f_{\boldsymbol{\gamma}}|_{\boldsymbol{\gamma}=\boldsymbol{\gamma^{0,h}}}}{g^0} + \sum_{\substack{l=1\\ l\neq h}}^{q}\sum_{i=1}^{k}\pi_l^0\delta_i^{l,n}\frac{\partial_{\gamma_i} f_{\boldsymbol{\gamma}}|_{\boldsymbol{\gamma}=\boldsymbol{\gamma^{0,l}}}}{g^0} \\
&+ \frac{1}{2\sqrt{2}}\sum_{i=1}^{k}\sum_{j=1}^{k}\Big(\gamma_i^{1,n} - \gamma_i^{0,h}\Big)\Big(\gamma_j^{1,n} - \gamma_j^{0,h}\Big)\frac{\partial_{\gamma_i\gamma_j} f_{\boldsymbol{\gamma}}|_{\boldsymbol{\gamma}=\boldsymbol{\gamma^{0,h}}}}{g^0} + o(1). \qquad (2.36)
\end{aligned}$$

Let $|U_h|$ denote the cardinality of U_h (given by (M1)). Then $d_{\boldsymbol{\beta_n^h}}$ has $r_h = q + qk + k + |U_h|$ terms with respect to the linearly $L_2(g^0\nu)$-independent functions given by $\mathcal{H}_h = \{\frac{f_1}{g^0},\dots,\frac{f_{r_h}}{g^0}\}$. According to (2.36) the corresponding sequences of coefficients $z_1^{h,n},\dots,z_{r_h}^{h,n}$ of $\frac{f_1}{g^0},\dots,\frac{f_{r_h}}{g^0}$ are given by

$$\begin{array}{rl}
\lambda_{1,n} + \rho_{h,n}, & \rho_{l,n}, l = 1,\dots,q,\ l\neq h, \\
\lambda_{1,n}(\gamma_i^{1,n} - \gamma_i^{0,h}) + \pi_h^0\delta_i^{h,n}, & \pi_l^0\delta_i^{l,n}, l = 1,\dots,q,\ l\neq h,\ i = 1,\dots,k, \\
\frac{1}{2\sqrt{2}}(\gamma_i^{1,n} - \gamma_i^{0,h})^2,\ i\in H_h^c, & \frac{1}{\sqrt{2}}(\gamma_i^{1,n} - \gamma_i^{0,h})(\gamma_j^{1,n} - \gamma_j^{0,h}), (i,j)\in U_h \text{ for } i<j
\end{array}$$

and

$$\frac{1}{\sqrt{2}}\Big(\frac{1}{2}(\gamma_i^{1,n} - \gamma_i^{0,h})^2 + c_i^h(\gamma_u^{1,n} - \gamma_u^{0,h})(\gamma_v^{1,n} - \gamma_v^{0,h})\Big),\ i\in H_h = \textstyle\bigcup_{(u,v)\in U_h^c} H_h(u,v) \text{ and } c_i^h \neq 0.$$

Applying the same Taylor expansion on the argument of the denominator $\|d^{num}_{\boldsymbol{\beta_n^h}}\|_{L_2(g^0\nu)}$ leads to sequences of coefficients $z_j^{h,n}/\|\sum_{i=1}^{r_h} z_i^{h,n}\frac{f_i}{g^0} + o(1)\|_{L_2(g^0\nu)}$, $j\in\{1,\dots,r_h\}$, for $d^{num}_{\boldsymbol{\beta_n^h}}/\|d^{num}_{\boldsymbol{\beta_n^h}}\|_{L_2(g^0\nu)}$. For $\boldsymbol{z^{h,n}} = (z_1^{h,n},\dots,z_{r_h}^{h,n})'$ we obtain

$$\begin{aligned}
\lim_{n\to\infty}|z_j^{h,n}|\Big\|\sum_{i=1}^{r_h} z_i^{h,n}\frac{f_i}{g^0} + o(1)\Big\|^{-1}_{L_2(g^0\nu)} &= \lim_{n\to\infty}|z_j^{h,n}|\Big\|\Big(\sum_{i=1}^{r_h} z_i^{h,n}\frac{f_i}{g^0}\Big)(1+o(1))\Big\|^{-1}_{L_2(g^0\nu)} \\
&\overset{\text{Lem.2.17}}{\leq} \lim_{n\to\infty}\frac{|z_j^{h,n}|}{c\,\|\boldsymbol{z^{h,n}}\|(1+o(1))} \\
&\leq \lim_{n\to\infty}\frac{|z_j^{h,n}|}{c\sqrt{(z_j^{h,n})^2}(1+o(1))} \\
&= \frac{1}{c}
\end{aligned}$$

for some $c > 0$. The form of (2.35) and (2.36), respectively, and the continuity of the coefficients lead to accumulation points μ_l, $\tilde{\rho}_l$, $\lambda_{l,i}$ and $\frac{1}{\sqrt{2}}\alpha_i^h\alpha_j^h$ of the sequences of coefficients such that (2.32) holds.
Furthermore, we have that any $d \in \overline{\mathcal{D}}$ is square-integrable since all functions in (2.32) are square-integrable due to (P1 t) and (M1) and their corresponding coefficients are bounded. $\overline{\mathcal{D}}$ is a subset of the unit sphere in the Hilbert space $L_2(g^0\nu)$ since any $d \in \overline{\mathcal{D}}$ has a form of $d^{num}\|d^{num}\|^{-1}_{L_2(g^0\nu)}$ which implies $\|d\|_{L_2(g^0\nu)} = 1$.

b) According to (2.31) (which was derived by (A1), (A2), (ID), (P1 t)) and $q = 1$, $p = 2$ the functions in $\mathcal{D}$ have the form

$$d_{\boldsymbol{\beta}} = \frac{1}{N(\boldsymbol{\beta})}\left(\sum_{i=1}^{2}\delta_i^1\frac{\partial_{\gamma_i}f_{\boldsymbol{\gamma}}|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,1}}}{g^0} + \lambda_1\frac{f_{\boldsymbol{\gamma}^1}}{g^0} + \rho_1\frac{f_{\boldsymbol{\gamma}^{0,1}}}{g^0}\right) \text{ for } \boldsymbol{\beta} \in \mathcal{B}. \tag{2.37}$$

Let $d^{num}_{\boldsymbol{\beta}}$ be the numerator of $d_{\boldsymbol{\beta}}$. The $d^{num}_{\boldsymbol{\beta}}$ is continuous in $\boldsymbol{\beta} = (\lambda_1, \boldsymbol{\gamma}^1, \boldsymbol{\delta}^1, \rho_1)'$ on $\mathcal{B}$ since $f_{\boldsymbol{\gamma}}$ is continuous in $\boldsymbol{\gamma}$, because it possesses partial derivatives up to order 5 on Γ which are the right and the left partial derivatives, respectively, on the boundary of Γ according to (P0)b), and linear combinations of continuous functions (due to (P0)) are also continuous. As the $L_2(g^0\nu)$-norm is also a continuous function, we have $d : \mathcal{B} \to \mathcal{D}$ with $\boldsymbol{\beta} \mapsto d_{\boldsymbol{\beta}}$ is a continuous function. As in the proof of a) we investigate the sequences $(d_{\boldsymbol{\beta_n}})_{n\in\mathbb{N}}$ with $\lim_{n\to\infty}\boldsymbol{\beta_n} = \tilde{\boldsymbol{\beta}} \in \overline{\mathcal{B}}\backslash\mathcal{B}$ for $\lim_{n\to\infty} N(\boldsymbol{\beta_n}) > 0$ and $\lim_{n\to\infty} N(\boldsymbol{\beta_n}) = 0$, respectively. Since the first case leads to a form as in (2.37) we study the second case. Remark 2.4 leads to a unique limit point in $\overline{\mathcal{B}}\backslash\mathcal{B}$, namely $\tilde{\boldsymbol{\beta}} = (\frac{1}{\sqrt{2}}, \boldsymbol{\gamma}^{0,1}, \mathbf{0}, -\frac{1}{\sqrt{2}})'$. For all sequences $(\boldsymbol{\gamma}^{1,n})_{n\in\mathbb{N}}$ tending to $\boldsymbol{\gamma}^{0,1}$ we have $|\gamma_1^{1,n} - \gamma_1^{0,1}| \leq |\gamma_2^{1,n} - \gamma_2^{0,1}|$ according to the construction of $\mathcal{B}$ based on $\tilde{\Gamma}$. Such kind of sequences are are also considered in the proof of a) and thus we follow its method. As above we make a Taylor expansion of the numerator $d^{num}_{\boldsymbol{\beta_n}}$ as well as of the argument of the denominator $\|d^{num}_{\boldsymbol{\beta_n}}\|_{L_2(g^0\nu)}$ of $d_{\boldsymbol{\beta_n}}$ up to the second order at the point $\tilde{\boldsymbol{\beta}}$. While using property (M2) we obtain

$$\begin{aligned}
d^{num}_{\boldsymbol{\beta_n}} = {} & \left(\lambda_{1,n} + \rho_{1,n}\right)\frac{f_{\boldsymbol{\gamma}^{0,1}}}{g^0} + \left(\lambda_{1,n}\left(\gamma_1^{1,n} - \gamma_1^{0,1}\right) + \delta_1^{1,n}\right)\frac{\partial_{\gamma_1}f_{\boldsymbol{\gamma}}|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,1}}}{g^0} \\
& + \left(\lambda_{1,n}\left(\gamma_2^{1,n} - \gamma_2^{0,1}\right) + \delta_2^{1,n} + \frac{c}{2\sqrt{2}}\left(\gamma_1^{1,n} - \gamma_1^{0,1}\right)^2\right)\frac{\partial_{\gamma_2}f_{\boldsymbol{\gamma}}|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,1}}}{g^0} \\
& + \frac{1}{\sqrt{2}}\left(\gamma_1^{1,n} - \gamma_1^{0,1}\right)\left(\gamma_2^{1,n} - \gamma_2^{0,1}\right)\frac{\partial_{\gamma_1\gamma_2}f_{\boldsymbol{\gamma}}|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,1}}}{g^0} \\
& + \frac{1}{2\sqrt{2}}\left(\gamma_2^{1,n} - \gamma_2^{0,1}\right)^2\frac{\partial_{\gamma_2\gamma_2}f_{\boldsymbol{\gamma}}|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,1}}}{g^0} + o(1)
\end{aligned} \tag{2.38}$$

for some $c \neq 0$, where $\lambda_{1,n} + \rho_{1,n} = 0$ according to (2.2). The remainder of the proof is in the same manner as with a). □

Notice that we obtain the statements of Lemma 2.16 by Taylor expansions in a point $\tilde{\boldsymbol{\beta}}$ while [DCG99], p. 1195, expand in the true component parameters $\boldsymbol{\gamma}^{0,1}, \ldots, \boldsymbol{\gamma}^{0,q}$.

Proposition 2.18 *Let* (A1), (A2) *and* (ID) *be satisfied.*

a) *Suppose that* $Q = q$, $p = q + 1$, (P0), (P1 t) *and* (M1) *hold. Then* $\overline{\mathcal{D}}$ *is a Donsker class.*

b) *Suppose that* $q = 1, p = 2$, (P0), (P1 t) *and* (M2) *hold. Let* $\mathcal{G}_{2,M}$ *and* $\mathcal{B}$ *from Lemma* 2.16 b). *Then* $\overline{\mathcal{D}}$ *is a Donsker class.*

The verification of Proposition 2.18 follows the concept by Keribin (2000), p. 60.

Proof of Proposition 2.18:

a) Let $\overline{\mathcal{B}}$ be the index set of $\overline{\mathcal{D}}$, $\boldsymbol{\beta^*} = (\beta_1^*, \ldots, \beta_K^*)' \in \mathcal{B}$ and $K = (q+1)(k+1) = dim(\overline{\mathcal{B}})$. Firstly, we show that for any $\varepsilon > 0$ there exists some $\tau > 0$ such that

$$\left\| \inf_{\beta \in B^*} d_{\boldsymbol{\beta}} - \sup_{\beta \in B^*} d_{\boldsymbol{\beta}} \right\|^2_{L_2(g^0\nu)} < \varepsilon \qquad \text{with } B^* = \mathop{\times}_{i=1}^{K} \left[\beta_i^* - \tfrac{\tau}{2}, \beta_i^* + \tfrac{\tau}{2}\right] \cap \overline{\mathcal{B}}. \tag{2.39}$$

Let $(\boldsymbol{\beta_n})_{n\in I\!N}$ be a sequence with $\lim_{n\to\infty} \boldsymbol{\beta_n} = \boldsymbol{\beta^*}$.

1. Suppose that $\lim_{n\to\infty} N(\boldsymbol{\beta_n}) > 0$ holds.
   As shown in the proof of Lemma 2.16 a) $d : \mathcal{B} \to \mathcal{D}$ with $\boldsymbol{\beta} \mapsto d_{\boldsymbol{\beta}}$ is a continuous function. If $\boldsymbol{\beta^*} \in \overline{\mathcal{B}} \backslash \mathcal{B}$ we also have $\lim_{n\to\infty} d_{\boldsymbol{\beta_n}} = d_{\boldsymbol{\beta^*}}$ (see the proof of Lemma 2.16 a)). Additionally, all functions in $\mathcal{D}$ are $g^0\nu$- square-integrable as well as $d_{\boldsymbol{\beta^*}}$ due to (P1 t). Thus for any $\varepsilon > 0$ there exists some $\tau > 0$ such that (2.39) holds.
2. Suppose that $\lim_{n\to\infty} N(\boldsymbol{\beta_n}) = 0$ holds.
   We write

$$\left\| \inf_{\beta \in B^*} d_{\boldsymbol{\beta}} - \sup_{\beta \in B^*} d_{\boldsymbol{\beta}} \right\|^2_{L_2(g^0\nu)} = \left\| \inf_{\beta \in B^*} \frac{d^{num}_{\boldsymbol{\beta}}}{\|d^{num}_{\boldsymbol{\beta}}\|_{L_2(g^0\nu)}} - \sup_{\beta \in B^*} \frac{d^{num}_{\boldsymbol{\beta}}}{\|d^{num}_{\boldsymbol{\beta}}\|_{L_2(g^0\nu)}} \right\|^2_{L_2(g^0\nu)}. \tag{2.40}$$

According to the proof of Lemma 2.16 a) there exists some $h \in \{1, \ldots, q\}$ such that $\lim_{n\to\infty} \boldsymbol{\beta_n} = \boldsymbol{\beta^*} = \boldsymbol{\tilde{\beta}^h} = (\frac{1}{\sqrt{2}}, \boldsymbol{\gamma^{0,h}}, \mathbf{0}, \ldots, \mathbf{0}, \underbrace{0, \ldots, 0}_{h-1}, -\frac{1}{\sqrt{2}}, \underbrace{0, \ldots, 0}_{q-h})' \in \overline{\mathcal{B}} \backslash \mathcal{B}$. In the same manner as with the proof of Lemma 2.16 a) we make a Taylor expansion of $d^{num}_{\boldsymbol{\beta}}$ up to the second order at the point $\boldsymbol{\tilde{\beta}^h}$ and obtain

$$d^{num}_{\boldsymbol{\beta}} = \left(\sum_{i=1}^{r_h} z_i^h(\boldsymbol{\beta}) \frac{f_i}{g^0} \right) (1 + o(1))$$

where $\mathcal{H}_h = \{\frac{f_1}{g^0}, \ldots, \frac{f_{r_h}}{g^0}\}$ is a fixed set of $L_2(g^0\nu)$-independent functions and $z_1^h(\boldsymbol{\beta})$, $\ldots$, $z_{r_h}^h(\boldsymbol{\beta})$ are their corresponding coefficients, see p. 29. Due to Lemma 2.17 and $\mathcal{H}_h$ is fixed, the right hand side of (2.40) is less or equal to

$$\left\| \inf_{\beta \in B^*} \frac{\left(\sum\limits_{i=1}^{r_h} z_i^h(\boldsymbol{\beta}) \frac{f_i}{g^0}\right)(1 + o(1))}{c' \, \|\boldsymbol{z^h(\beta)}\| (1 + o(1))} - \sup_{\beta \in B^*} \frac{\left(\sum\limits_{i=1}^{r_h} z_i^h(\boldsymbol{\beta}) \frac{f_i}{g^0}\right)(1 + o(1))}{c \, \|\boldsymbol{z^h(\beta)}\| (1 + o(1))} \right\|^2_{L_2(g^0\nu)}$$

for some $c, c' > 0$. For any $i = 1, \ldots, r_h$ the coefficients $z_i^h(\boldsymbol{\beta})$ are continuous in $\boldsymbol{\beta}$ according to their form (see p. 29) and $\lim_{n\to\infty} \frac{z_i^h(\boldsymbol{\beta})}{\|\boldsymbol{z^h(\beta)}\|} \le 1$. Thus for any $\varepsilon > 0$ there exists some $\tau > 0$ such that (2.39) holds.

Without loss of generality let $\varepsilon > \tau > 0$ such that (2.39) holds. Then we obtain a partition of $\overline{\mathcal{D}}$ of cardinality of $O(\frac{1}{\varepsilon^K})$. This partition of $\overline{\mathcal{D}}$ implies that the bracketing number $N^{(2)}_{[\,]}(\varepsilon, \overline{\mathcal{D}}, g^0\nu) = O(\frac{1}{\varepsilon^K})$. Since $\int\limits_0^1 \Big\{ \log\Big(\varepsilon^{-K}\Big)\Big\}^{\frac{1}{2}} d\varepsilon \leq K^{\frac{1}{2}}\Big\{ -\int\limits_0^{1/e} \log(\varepsilon)\, d\varepsilon + (1-\frac{1}{e})\Big\} < \infty$ the assertion follows from Theorem 2.14.

b) The verification of $\overline{\mathcal{D}}$ being a Donsker class is along the same line. Notice that we assume $B^* \subset \overline{\mathcal{B}}$ in (2.39). As shown in the proof of Lemma 2.16 b) $d : \mathcal{B} \to \mathcal{D}$ with $\boldsymbol{\beta} \mapsto d_{\boldsymbol{\beta}}$ is a continuous function. The corresponding coefficients of the $L_2(g^0\nu)$ linearly independent functions in (2.38) are also continuous on $\mathcal{B}$. Thus we follow the proof in the same manner as in a). $\square$

At the end of this subsection we would like to make a short remark on the following statement by Liu and Shao (2003) [LS03] , p. 825, “Note that the most widely used mixture models are those from the exponential families... For these models the Donsker class condition can be directly verified using Lemma 3.2... we assume that the P-Donsker class condition is satisfied and focus on deriving the index set.” Thereby, Lemma 3.2 in [LS03] is helpful when their generalized score function is Lipschitz with respect to a corresponding (component and mixing) parameter and the parameter space is compact. Additionally, [LS03], p. 826, claim “The locally conic parameterization approach of Dacunha-Castelle and Gassiat (1997, 1999) is very useful in identifying the index set.” According to [DCG99] one has to verify that the closure of $\mathcal{D} = \{d_{\boldsymbol{\beta}} : \boldsymbol{\beta} \in \mathcal{B}\}$ (see (2.35)) is a Donsker class, where the space $\mathcal{B} = \Big\{\boldsymbol{\beta}_{\boldsymbol{\sigma}(\boldsymbol{g})} : g \in \mathcal{G}_p\backslash\{g^0\}\Big\}$ may not be compact according to the remarks after our Proposition 2.10. Thus Lemma 3.2 from [LS03] may not be applicable to the set of score functions given in [DCG99] directly. In chapter 3 we will establish that the sufficient conditions for the existence of an asymptotic distribution according to [DCG99] are not satisfied for all mixtures of exponential families.

2.2.2. Asymptotic Results for Testing $g = g^0$ Against $g \in \mathcal{G}_p\backslash\{g^0\}$

We base this subsection (as usual) on the assumptions (A1), (A2), (ID) as well as on the aforementioned definitions and propositions. Especially, we use the fact that $\mathcal{G}_p$ and $\mathcal{G}_{2,M}$, respectively, have a locally conic model. Furthermore, we assume that $\mathcal{D} = \{d_{\boldsymbol{\beta}} : \boldsymbol{\beta} \in \mathcal{B}\}$ where $\mathcal{B}$ is given by Proposition 2.10 and Lemma 2.11, respectively.

For any $g \in \mathcal{G}_p$ let the LRT statistic for testing $\boldsymbol{H_0 : g = g^0}$ against $\boldsymbol{H_1 : g \in \mathcal{G}_p\backslash\{g^0\}}$ be given by

$$T_n(p) \quad = \sup_{g_{(\theta,\boldsymbol{\beta})}\in\mathcal{G}_p} l_n(\theta,\boldsymbol{\beta}) - l_n(0) \tag{2.41}$$

where $l_n(\theta,\boldsymbol{\beta}) = l_n(g_{(\theta,\boldsymbol{\beta})}) = \sum\limits_{i=1}^{n} \log\big(g_{(\theta,\boldsymbol{\beta})}(\boldsymbol{x_i})\big)$ denotes the log-likelihood function of θ and $\boldsymbol{\beta}$ and $l_n(0) = \sum\limits_{i=1}^{n} \log\big(g^0(\boldsymbol{x_i})\big)$ for some realization $\boldsymbol{x_1},\ldots,\boldsymbol{x_n}$ of an i.i.d. sample $\boldsymbol{X_1},\ldots,\boldsymbol{X_n}$.

[DCG99] derive the asymptotic distribution of $T_n(p)$ in six steps. Firstly, they define a maximizer $\hat{\theta}_{\boldsymbol{\beta}}(n)$ of the log-likelihood function $l_n(\theta,\boldsymbol{\beta})$ for fixed value of $\boldsymbol{\beta} \in \mathcal{B}$. Secondly, they use property (P0) and the assumption that the reparameterization is locally conic in order to conclude that the supremum with respect to $\boldsymbol{\beta} \in \mathcal{B}$ of $\hat{\theta}_{\boldsymbol{\beta}}(n)$ converges in probability to 0 as n tends to infinity (see [DCG99], Proposition 3.3). Thirdly, they expand the log-likelihood function

$$\begin{aligned} l_n(\theta, \boldsymbol{\beta}) - l_n(0) &= \sum_{i=1}^{n} \frac{g_{(\theta,\boldsymbol{\beta})}(\boldsymbol{x_i}) - g^0(\boldsymbol{x_i})}{g^0(\boldsymbol{x_i})} - \frac{1}{2}\sum_{i=1}^{n}\left(\frac{g_{(\theta,\boldsymbol{\beta})}(\boldsymbol{x_i}) - g^0(\boldsymbol{x_i})}{g^0(\boldsymbol{x_i})}\right)^2 \\ &\quad + \frac{1}{3}\sum_{i=1}^{n} U_i \left(\frac{g_{(\theta,\boldsymbol{\beta})}(\boldsymbol{x_i}) - g^0(\boldsymbol{x_i})}{g^0(\boldsymbol{x_i})}\right)^3 \end{aligned} \tag{2.42}$$

for θ tending to 0, where $|U_i| \leq 1$ (see [DCG99], p. 1196). Their idea is to make a Taylor expansion of $g_{(\theta,\boldsymbol{\beta})}$ with respect to θ at the point $\theta = 0$ and afterwards to maximize the resulting log-likelihood function over θ and then over $\boldsymbol{\beta}$. Unfortunately, the aforementioned Taylor expansion does not lead to uniformly bounded coefficients with respect to the linearly independent functions given by (P1) or given in our modifications by (P1 t) as well as by (M1) and (M2), respectively. [DCG99] solve the problem by defining, fourthly, a (random) partition of the parameter space $\mathcal{B} = A_n \cup B_n$ in such a way that the above-mentioned coefficients (with respect to the linearly independent functions) are uniformly bounded on A_n while they are not on the other set B_n. More precisely, [DCG99] define the partitions of $\mathcal{B}$ by

$$\begin{aligned} A_n &= \left\{\boldsymbol{\beta} \in \mathcal{B} : \frac{\sup_{l\in\{1,\dots,q\}} \|\boldsymbol{\delta^l}\|}{N(\boldsymbol{\beta})^2} \leq \frac{1}{\eta_n^\alpha}\right\} \quad \text{and} \\ B_n &= \left\{\boldsymbol{\beta} \in \mathcal{B} : \exists\, l \in \{1,\dots,q\} \text{ with } \frac{\|\boldsymbol{\delta^l}\|}{N(\boldsymbol{\beta})^2} \geq \frac{1}{\eta_n^\alpha}\right\} \end{aligned}$$

for some $\alpha < \frac{3}{4}$ and $\eta_n = \sup_{\boldsymbol{\beta}\in\mathcal{B}} \hat{\theta}_{\boldsymbol{\beta}}(n)$ given in step two (see [DCG99], p. 1186). To obtain a general overview of the behaviour of the likelihood function on region A_n and B_n respectively, we refer to Kerbin (2000), p. 58-59.

On B_n the normalizing factor $N(\boldsymbol{\beta})$ and all parameters $\boldsymbol{\delta^l}$, $l = 1, \dots, q$, tend uniformly to $\mathbf{0}$, such that there exists some l such that $\frac{\|\boldsymbol{\delta^l}\|}{N(\boldsymbol{\beta})^2}$ is not uniformly bounded. The latter ratio is found in the Taylor expansion of the likelihood function and leads to an expansion up to order 5 on B_n, while it is sufficient to expand up to order 2 on A_n. (For this reason partial derivatives up to order 5 are demanded by property (P0).) This partition leads to

$$T_n(p) = \sup\left\{\sup_{\boldsymbol{\beta}\in A_n} l_n(\hat{\theta}_{\boldsymbol{\beta}}(n), \boldsymbol{\beta}) - l_n(0),\ \sup_{\boldsymbol{\beta}\in B_n} l_n(\hat{\theta}_{\boldsymbol{\beta}}(n), \boldsymbol{\beta}) - l_n(0)\right\}. \tag{2.43}$$

Fifthly, and that is the most difficult point, they check that the remaining terms with respect to the Taylor expansion of the likelihood function are $o(1)$ uniform in probability over $\boldsymbol{\beta}$ in A_n and B_n, respectively.
In this work, we use the terms "uniformly $O_P(\cdot)$" and "uniformly $o_P(\cdot)$" in the following sense

Definition 2.19 (Prakasa Rao (1987), Definition 1.15.2a)
Let $\{r_n : n \in \mathbb{N}\}$ be a sequence of positive real numbers. Let $\{Y_n(\phi) : \phi \in \Phi, n \in \mathbb{N}\}$ be a sequence of random vectors with $Y_n(\phi)$ defined on a probability space $(\Omega_n, \mathcal{A}_n, P_n)$ and taking values in the extended $\mathbb{R}^k$. Then $Y_n(\phi) = O_P(r_n)$ if for every $\delta > 0$, there exists a constant $C > 0$ and $N(\delta, C)$ such that

$$P_n\Big(\sup_{\phi\in\Phi} |Y_n(\phi)|/r_n \leq C\Big) \ \geq\ 1 - \delta \quad \textit{for all } n \geq N(\delta, C)$$

and $Y_n = o_P(r_n)$ *if for every* $\delta > 0$ *and* $\epsilon > 0$, *there exists a constant* $N(\delta, \epsilon)$ *such that*

$$P_n\Big(\sup_{\phi\in\Phi} |Y_n(\phi)/r_n| \le \epsilon\Big) \quad \ge \quad 1-\delta \quad \textit{for all } n \ge N(\delta,\epsilon).$$

Our modifications lead to the same results as given in the fifth step, namely

Lemma 2.20 *Let* (A1), (A2) *and* (ID) *be satisfied. Suppose that at least one of the following assumptions holds:*

a) (P0) *and* (P1),

b) (P0), (P1 t) *and* (M1) *for* $Q = q$, $p = q + 1$,

c) (P0), (P1 t) *and* (M2) *for* $q = 1$, $p = 2$ *and* $\mathcal{G}_{2,M}$, $\mathcal{B}$ *from Lemma* 2.16 b).

Then the following statements hold:

1) $\sup_{\beta\in A_n} l_n(\hat\theta_{\beta}(n), \beta) - l_n(0)$ *converges in distribution to the variable* $\frac{1}{2}\sup_{d\in\mathcal{D}} \left(\xi_d\right)^2 \cdot \mathbb{1}_{\xi_d\ge 0}$,

2) $\sup_{\beta\in B_n} l_n(\hat\theta_{\beta}(n), \beta) - l_n(0)$ *is bounded from above by* $\sup_{\beta\in A_n} l_n(\hat\theta_{\beta}(n), \beta) - l_n(0) + o_P(1)$.

(Thereby $(\xi_d)_{d\in\mathcal{D}}$ *is a centered Gaussian process with covariance given by the usual scalar product in* $L_2(g^0\nu)$, *see Definition* 2.12.*)*
(Under assumption a) *statement* 1) *and* 2) *are equal to* [DCG99], *Lemma* 3.4 *and Lemma* 3.5, *respectively.)*

The proof of Lemma 2.20 is given in the appendix since it follows mainly the same lines as the proof of Lemma 3.4 and Lemma 3.5 in [DCG99], respectively. It is marked correspondingly, whenever the underlying modified assumptions are used. Lemma 2.22 given below is a fundamental working tool for statements about the convergence behaviour of the remaining terms of the Taylor expansions on A_n and B_n, respectively. In Example 2.23 we show that Lemma 2.22 does not hold if some true mixing weight is equal to 0. For this reason we have assumed in contrast to [DCG99] by (A1) that all true mixing weights are strictly positive.

The sixth step is conform to [DCG99]

Theorem 2.21 *Let* (A1), (A2) *and* (ID) *be satisfied. Suppose that at least one of the assumptions* a), b) *or* c) *from Lemma* 2.20 *holds. Then* $T_n(p)$ *converges in distribution to the variable*

$$\frac{1}{2}\sup_{d\in\mathcal{D}} \left(\xi_d\right)^2 \cdot \mathbb{1}_{\xi_d\ge 0}.$$

(Under assumption a) *the statement is equal to* [DCG99], *Theorem* 3.2.*)*

Proof of Theorem 2.21:
Due to (2.43) and Lemma 2.20 2) we have

$$T_n(p) \quad = \quad \sup\left\{ \sup_{\beta\in A_n} l_n(\hat\theta_{\beta}(n), \beta) - l_n(0),\ \sup_{\beta\in A_n} l_n(\hat\theta_{\beta}(n), \beta) - l_n(0) + o_P(1)\right\}.$$

Using Lemma 2.20 1) leads to the assertion. □

Lemma 2.22 *Let* (A1), (A2) *and* (ID) *be satisfied. Suppose that at least one of the assumptions* a), b) *or* c) *from Lemma* 2.20 *holds. Then there exists a constant $a \geq 0$ such that for any $\boldsymbol{\beta} \in \mathcal{B}$*

$$\frac{\sup_{l=1,\dots,q} \|\boldsymbol{\delta^l}\|^2}{N(\boldsymbol{\beta})} \leq a. \tag{2.44}$$

(Under assumption a) *the statement is equal to* [DCG99], *Lemma* 5.1.*)*

Proof of Lemma 2.22:

b) Suppose that $\|\boldsymbol{\delta^h}\|^2/N(\boldsymbol{\beta})$ is unbounded for some $h \in \{1,\dots,q\}$. Then there exists a sequence $\boldsymbol{\beta_n} = (\lambda_{1,n}, \boldsymbol{\gamma^{1,n}}, \boldsymbol{\delta^{1,n}}, \dots, \boldsymbol{\delta^{q,n}}, \rho_{1,n}, \dots, \rho_{q,n})' \in \mathcal{B}$ such that

$$\lim_{n\to\infty} \frac{\|\boldsymbol{\delta^{h,n}}\|^2}{N(\boldsymbol{\beta_n})} = \infty.$$

According to (2.6) we have $\boldsymbol{\delta^{h,n}} \in [-1,1]^k$, $n \in I\!N$, which leads to $\lim_{n\to\infty} N(\boldsymbol{\beta_n}) = 0$. Corresponding to the proof of Lemma 2.16 a) we write $N(\boldsymbol{\beta_n}) = \|d^{num}_{\boldsymbol{\beta^h_n}}\|_{L_2(g^0\nu)}$ with $d^{num}_{\boldsymbol{\beta^h_n}}$ given by (2.36) and $\lim_{n\to\infty} \boldsymbol{\beta^h_n} = (\frac{1}{\sqrt{2}}, \boldsymbol{\gamma^{0,h}}, \boldsymbol{0}, \dots, \boldsymbol{0}, 0, \dots, 0, -\frac{1}{\sqrt{2}}, 0, \dots, 0)' \in \overline{\mathcal{B}} \backslash \mathcal{B}$. Let $\mathcal{H}_h = \{\frac{f_1}{g^0}, \dots, \frac{f_{r_h}}{g^0}\}$ and let the corresponding coefficients $z_i^{h,n}$ for $i \in \{1,\dots,r_h\}$ also be given as in the proof of Lemma 2.16 a). Then it follows from Lemma 2.17

$$\lim_{n\to\infty} \frac{\|\boldsymbol{\delta^{h,n}}\|^2}{N(\boldsymbol{\beta^h_n})} \leq \lim_{n\to\infty} \frac{1}{c} \frac{\|\boldsymbol{\delta^{h,n}}\|^2}{\|\boldsymbol{z^{h,n}}\|(1+o(1))}$$

for some $c > 0$. Using only coefficients of functions in $\mathcal{H}_h$ depending on $\boldsymbol{\gamma^{1,n}} - \boldsymbol{\gamma^{0,h}}$ we obtain

$$\lim_{n\to\infty} \frac{\|\boldsymbol{\delta^{h,n}}\|^2}{N(\boldsymbol{\beta^h_n})} \leq \lim_{n\to\infty} \frac{1}{c} \frac{\|\boldsymbol{\delta^{h,n}}\|^2}{\Big(\|\lambda_{1,n}(\boldsymbol{\gamma^{1,n}} - \boldsymbol{\gamma^{0,h}}) + \pi_h^0 \boldsymbol{\delta^{h,n}}\|^2 + p_n\Big)^{\frac{1}{2}}(1+o(1))} \tag{2.45}$$

where

$$p_n = \frac{1}{2} \sum_{(i,j)\in U_h} (a_i^{h,n} a_j^{h,n})^2 + \frac{1}{8} \sum_{i \in H_h^c} (a_i^{h,n})^4 + \frac{1}{2} \sum_{\substack{(u,v)\in U_h^c \\ i \in H_h(u,v)}} \Big(\frac{1}{2}(a_i^{h,n})^2 + c_i^h a_u^{h,n} a_v^{h,n}\Big)^2 \tag{2.46}$$

and $\boldsymbol{a^{h,n}} = \boldsymbol{\gamma^{1,n}} - \boldsymbol{\gamma^{0,h}}$ converges to $\boldsymbol{0}$ for $n \to \infty$, $c_i^h \neq 0$ for $i \in H_h(u,v)$.

We investigate the right hand side of (2.45) with the aid of the convergence speed of $\boldsymbol{\delta^{h,n}}$ and $\lambda_{1,n}\, \boldsymbol{a^{h,n}}$:

1. Case: $\|\boldsymbol{\delta^{h,n}}\| = o\Big(\|\lambda_{1,n}\, \boldsymbol{a^{h,n}}\|\Big)$ and 2. Case: $\|\lambda_{1,n}\, \boldsymbol{a^{h,n}}\| = o\Big(\|\boldsymbol{\delta^{h,n}}\|\Big)$
   It is obvious that

$$\|\lambda_{1,n}\, \boldsymbol{a^{h,n}} + \pi_h^0 \boldsymbol{\delta^{h,n}}\|^2 \geq \Big(\|\lambda_{1,n}\, \boldsymbol{a^{h,n}}\| - \pi_h^0 \|\boldsymbol{\delta^{h,n}}\|\Big)^2.$$

Since the two sequences $\|\lambda_{1,n}\, \boldsymbol{a^{h,n}}\|$ and $\|\boldsymbol{\delta^{h,n}}\|$ converge to 0, there exists some $n_0 \in I\!N$ such that

$$\|\lambda_{1,n}\, \boldsymbol{a^{h,n}} + \pi_h^0 \boldsymbol{\delta^{h,n}}\|^2 \quad \geq \quad \Big(\|\lambda_{1,n}\, \boldsymbol{a^{h,n}}\| - \pi_h^0 \|\boldsymbol{\delta^{h,n}}\|\Big)^4 \quad \text{for all } n \geq n_0.$$

Thus it follows directly that the right hand side of (2.45) is upper bounded by

$$\lim_{n\to\infty} \frac{1}{c} \frac{\|\boldsymbol{\delta^{h,n}}\|^2}{\Big(\|\lambda_{1,n}\, \boldsymbol{a^{h,n}}\| - \|\pi_h^0 \boldsymbol{\delta^{h,n}}\|\Big)^2 (1+o(1))} \quad < \quad \infty.$$

3. Case: $\lambda_{1,n}\, \boldsymbol{a^{h,n}} = -\pi_h^0\, \boldsymbol{\delta^{h,n}}\, (1+o(1))$
Let $c_{max} = \max\Big\{1, \max\{|c_i^h| : i \in H_h\}\Big\}$. First of all we prove that the right hand side of (2.45) is upper bounded by

$$\lim_{n\to\infty} \frac{1}{c} \frac{\|\boldsymbol{\delta^{h,n}}\|^2}{\Big(\| - \pi_h^0 \boldsymbol{\delta^{h,n}} o(1)\|^2 + \frac{1}{8\, c_{max}} \sum_{i=1}^{k} (a_i^{h,n})^4\Big)^{\frac{1}{2}} (1+o(1))}. \tag{2.47}$$

Afterwards we verify that (2.47) is bounded.

According to (M1) we have $\{1,\ldots,k\} = H_h^c \cup H_h$ with $H_h = \bigcup_{(u,v)\in U_h^c} H_h(u,v)$. Simple computations lead to

$$\begin{aligned} p_n &\overset{(2.46)}{=} \frac{1}{2} \sum_{(i,j)\in U_h} (a_i^{h,n} a_j^{h,n})^2 + \frac{1}{8} \sum_{i=1}^{k} (a_i^{h,n})^4 + \frac{1}{2} \sum_{\substack{(u,v)\in U_h^c \\ i\in H_h(u,v)}} (c_i^h a_i^{h,n} a_u^{h,n} a_i^{h,n} a_v^{h,n}) \\ &\qquad + \frac{1}{2} \sum_{\substack{(u,v)\in U_h^c \\ i\in H_h(u,v)}} (c_i^h a_u^{h,n} a_v^{h,n})^2 \\ &\geq \frac{1}{8\, c_{max}} \sum_{i=1}^{k} (a_i^{h,n})^4 + \frac{1}{2} \sum_{(i,j)\in U_h} (a_i^{h,n} a_j^{h,n})^2 + \frac{1}{2\, c_{max}} \sum_{\substack{(u,v)\in U_h^c \\ i\in H_h(u,v)}} (c_i^h a_i^{h,n} a_u^{h,n} a_i^{h,n} a_v^{h,n}) \end{aligned} \tag{2.48}$$

where $c_i^h \neq 0$ for $i \in H_h(u,v)$.
For any $(u,v) \in U_h^c$ $(u < v)$ with the corresponding (unique) $i \in H_h(u,v)$ we write $(m_{u,i}, M_{u,i}) = (min\{u,i\}, max\{u,i\})$ as well as $(m_{v,i}, M_{v,i}) = (min\{v,i\}, max\{v,i\})$. You may recall that both pair of indices belong to U_h as assumed in (M1), i.e.

$$M = \Big\{(m_{u,i}, M_{u,i}), (m_{v,i}, M_{v,i}) : \text{with } (u,v) \in U_h^c \text{ and } i \in H_h(u,v)\Big\} \subset U_h.$$

As $M \subset U_h$ we have

$$\begin{aligned} \frac{1}{2} \sum_{(i,j)\in U_h} (a_i^{h,n} a_j^{h,n})^2 &\geq \frac{1}{2} \sum_{(m_{i,w}, M_{i,w})\in M} (a_i^{h,n} a_w^{h,n})^2 \\ &= \frac{1}{2} \sum_{\substack{(m_{i,u}, M_{i,u}), \\ (m_{i,v}, M_{i,v})\in M \\ u<v}} \Big((a_i^{h,n} a_u^{h,n})^2 + (a_i^{h,n} a_v^{h,n})^2\Big). \end{aligned} \tag{2.49}$$

By definition of c_{max} one has $0 < |\frac{c_i^h}{c_{max}}| \leq 1$ for any $i \in H_h$ which leads to

$$\frac{1}{2} \sum_{(i,j) \in U_h} (a_i^{h,n} a_j^{h,n})^2 + \frac{1}{2\, c_{max}} \sum_{\substack{(u,v) \in U_h^c \\ i \in H_h(u,v)}} (c_i^h a_i^{h,n} a_u^{h,n} a_i^{h,n} a_v^{h,n})$$

$$\overset{(2.49)}{\geq} \sum_{\substack{(m_{i,u}, M_{i,u}), \\ (m_{i,v}, M_{i,v}) \in M \\ u<v}} \left(\left(\frac{1}{\sqrt{2}} a_i^{h,n} a_u^{h,n} \right)^2 + \frac{c_i^h}{2\, c_{max}} a_i^{h,n} a_u^{h,n} a_i^{h,n} a_v^{h,n} + \left(\frac{1}{\sqrt{2}} \frac{c_i^h}{2\, c_{max}} a_i^{h,n} a_v^{h,n} \right)^2 \right). \tag{2.50}$$

Obviously, the right hand side of (2.50) is positive and due to (2.48) we have $p_n \geq \frac{1}{8\, c_{max}} \sum_{i=1}^{k} (a_i^{h,n})^4 \geq 0$. Thus the right hand side of (2.45) is upper bounded by (2.47). According to the above assumption we have $\boldsymbol{\delta^{h,n}} = -\lambda_{1,n} \boldsymbol{a^{h,n}} / \left(\pi_h^0 \,(1 + o(1)) \right)$. Let $a_{max}^{h,n} = \max\{|a_i^{h,n}| : i = 1, \ldots, k\}$. Then (2.47) is upper bounded by

$$\lim_{n \to \infty} \frac{\lambda_{1,n}}{c} \frac{\left(k\, (a_{max}^{h,n})^2 \right) / \left(\pi_h^0 \,(1 + o(1)) \right)}{\left(\frac{1}{8\, c_{max}} \sum_{i=1}^{k} (a_i^{h,n})^4 \right)^{\frac{1}{2}} (1 + o(1))}$$

$$\leq \lim_{n \to \infty} \frac{\lambda_{1,n}}{c} \frac{k / \left(\pi_h^0 \,(1 + o(1)) \right)}{\left(\frac{1}{8\, c_{max}} \left(1 + \sum_{\substack{i=1 \\ i \neq max}}^{k} \left(a_i^{h,n} / a_{max}^{h,n} \right)^4 \right) \right)^{\frac{1}{2}} (1 + o(1))}$$

which is obviously finite for $\pi_h^0 > 0$ (due to (A1)).

a) see [DCG99], p. 1194-1195. Similar to the proof of b), especially formula (2.45), we obtain

$$\lim_{n \to \infty} \frac{\|\boldsymbol{\delta^{h,n}}\|}{N(\boldsymbol{\beta_n})}$$

$$\leq \frac{\|\boldsymbol{\delta^{l,n}}\|^2}{c\,(1 + o(1))} \left(\Big\| \sum_{j \in J_l} \lambda_j (\boldsymbol{\gamma^{j,n}} - \boldsymbol{\gamma^{0,l}}) + \pi_l^0 \boldsymbol{\delta^{l,n}} \Big\|^2 + \frac{1}{4} \Big(\sum_{j \in J_l} \lambda_j \|(\boldsymbol{\gamma^{j,n}} - \boldsymbol{\gamma^{0,l}})\|^2 \Big)^2 \right)^{-\frac{1}{2}},$$

where J_l denotes the set of indices i such that $\boldsymbol{\gamma^{i,n}}$ tends to $\boldsymbol{\gamma^{0,l}}$. Notice that [DCG99], p. 1195, do not mention the factor π_l^0 of $\boldsymbol{\delta^{l,n}}$. However, the very existence of the factor π_l^0 demands that all the mixing weights be strictly positive. Otherwise (2.44) does not hold, which is shown in Example 2.23.

c) The proof is almost in the same manner as the proof given in b). We suppose that there exists a sequence $\boldsymbol{\beta_n} = (\lambda_{1,n}, \boldsymbol{\gamma^{1,n}}, \boldsymbol{\delta^{1,n}}, \rho_{1,n})' \in \mathcal{B}$ such that $\|\boldsymbol{\delta^{1,n}}\|^2 / N(\boldsymbol{\beta_n})$ is unbounded and therefore implies that $\lim_{n \to \infty} \boldsymbol{\beta_n} = (\frac{1}{\sqrt{2}}, \boldsymbol{\gamma^{0,1}}, \mathbf{0}, -\frac{1}{\sqrt{2}})'$. Corresponding to the proof of Lemma 2.16 b) we write $N(\boldsymbol{\beta_n}) = \|d_{\boldsymbol{\beta_n}}^{num}\|_{L_2(g^0\nu)}$ with $d_{\boldsymbol{\beta_n}}^{num}$ given by (2.38).

Then Lemma 2.17 leads to

$$\lim_{n\to\infty} \frac{\|\boldsymbol{\delta^{1,n}}\|^2}{N(\boldsymbol{\beta_n})}$$

$$\overset{(2.38)}{\leq} \lim_{n\to\infty} \frac{\|\boldsymbol{\delta^{1,n}}\|^2}{\eta\,(1+o(1))} \left(\left(\lambda_{1,n}a_{1,n} + \delta_1^{1,n}\right)^2 + \left(\lambda_{1,n}a_{2,n} + \delta_2^{1,n} + \tfrac{c}{2\sqrt{2}}(a_{1,n})^2\right)^2 + p_n \right)^{-\frac{1}{2}} \tag{2.51}$$

for some $\eta > 0,\ c \neq 0$ and $\boldsymbol{a_n} = (a_{1,n}, a_{2,n})' = \boldsymbol{\gamma^{1,n}} - \boldsymbol{\gamma^{0,1}}$, $p_n = \frac{1}{2}(a_{2,n})^2\left((a_{1,n})^2 + \frac{1}{4}(a_{2,n})^2\right)$.

1. Case: $\|\boldsymbol{\delta^{h,n}}\| = o\left(\left\| \left(\lambda_{1,n}a_{1,n},\ \lambda_{1,n}a_{2,n} + \frac{c}{2\sqrt{2}}(a_{1,n})^2\right)' \right\|\right)$ and
2. Case: $\left\| \left(\lambda_{1,n}a_{1,n},\ \lambda_{1,n}a_{2,n} + \frac{c}{2\sqrt{2}}(a_{1,n})^2\right)' \right\| = o\left(\|\boldsymbol{\delta^{h,n}}\|\right)$ can be verified corresponding to the first and second case in b), respectively.
3. Case: $\left(\lambda_{1,n}a_{1,n},\ \lambda_{1,n}a_{2,n} + \frac{c}{2\sqrt{2}}(a_{1,n})^2\right)' = -\boldsymbol{\delta^{1,n}}\,(1+o(1))$
   Then the right hand side of (2.51) is equal to

$$\lim_{n\to\infty} \frac{1}{\eta\,(1+o(1))} \frac{\left[(\lambda_{1,n}a_{1,n})^2 + \left(\lambda_{1,n}a_{2,n} + \frac{c}{2\sqrt{2}}(a_{1,n})^2\right)^2\right](1+o(1))^{-2}}{\sqrt{(-\delta_1^{1,n}o(1))^2 + (-\delta_2^{1,n}o(1))^2 + \frac{1}{2}(a_{2,n})^2\left((a_{1,n})^2 + \frac{1}{4}(a_{2,n})^2\right)}}$$

   Due to the parameterization in Lemma 2.5 (which uses Algorithm 2.7) we have $\|(\boldsymbol{\gamma^{0,1}} + \boldsymbol{\delta}\frac{\theta}{N(\boldsymbol{\beta})}) - \boldsymbol{\gamma^{0,1}}\| \leq \|\boldsymbol{\gamma^1} - \boldsymbol{\gamma^{0,1}}\| = \|\boldsymbol{a_n}\|$ and according to the form of $\tilde{\Gamma}$ we have $|a_{1,n}| = |\gamma_1 - \gamma_1^{0,1}| \leq |\gamma_2 - \gamma_2^{0,1}| = |a_{2,n}|$. As $a_{1,n}$ and $a_{2,n}$ converge to 0 we have $(a_{1,n})^2 \leq a_{2,n}$ for $n \to \infty$ which implies that the right hand side of (2.51) is less or equal to

$$\lim_{n\to\infty} \frac{1}{\eta\,(1+o(1))} \left(\frac{\left[(\lambda_{1,n}a_{2,n})^2 + \left(\lambda_{1,n}|a_{2,n}| + \frac{c}{2\sqrt{2}}|a_{2,n}|\right)^2\right]^2 (1+o(1))^{-4}}{\frac{1}{2}(a_{2,n})^2\left((a_{1,n})^2 + \frac{1}{4}(a_{2,n})^2\right)} \right)^{\frac{1}{2}}$$

$$\leq \lim_{n\to\infty} \frac{1}{\eta\,(1+o(1))} \left(\frac{\left(\lambda_{1,n} + (\lambda_{1,n} + \frac{c}{2\sqrt{2}})^2\right)^2 (a_{2,n})^4}{\frac{1}{8}(a_{2,n})^4} \right)^{\frac{1}{2}}.$$

   It follows $\lim_{n\to\infty} \frac{\|\boldsymbol{\delta^{1,n}}\|^2}{N(\boldsymbol{\beta_n})} < \infty$ since $\lim_{n\to\infty} \lambda_{1,n} = \frac{1}{\sqrt{2}}$. □

The following example shows the necessity of assumption (A1), i.e. the necessity of strictly positive true mixing weights, for Lemma 2.22. Although this restriction does not change the central ideas of [DCG99] it leads us to the test of "Q populations against p populations" instead of "q populations against p populations" which we illustrate by

Example 2.23 *Let* (A1),(A2), (ID), (P0) *and* (P1) *be satisfied. Suppose that we wish to test* $Q = 2$ *populations against* $p = 4$ *populations. Let the true density be given by* $g^0 = 0f_{\gamma^{0,1}} + 1f_{\gamma^{0,2}}$ *for* $\gamma^{0,1} = 0, \gamma^{0,2} = 1$ *and* $\boldsymbol{\pi^0} = (0,1)'$ *though the true order of* g *is equal to* 1. *According to* (2.5) *the normalizing factor has the form*

$$N(\boldsymbol{\beta}) = \left\| \delta^2 \frac{\partial_\gamma f_\gamma|_{\gamma=\gamma^{0,2}}}{g^0} + \lambda_1 \frac{f_{\gamma^1}}{g^0} + \lambda_2 \frac{f_{\gamma^2}}{g^0} + \rho_1 \frac{f_{\gamma^{0,1}}}{g^0} + \rho_2 \frac{f_{\gamma^{0,2}}}{g^0} \right\|_{L_2(g^0\nu)} = \|d_{\boldsymbol{\beta}}^{num}\|_{L_2(g^0\nu)}.$$

Let $\boldsymbol{\beta_n} = (\lambda_{1,n}, \lambda_{2,n}, \gamma^{1,n}, \gamma^{2,n}, \delta^{1,n}, \delta^{2,n}, \rho_{1,n}, \rho_{2,n})' \in \mathcal{B}$ *be given by*

$$\lambda_{1,n} = n^{-\frac{3}{2}} = -\rho_{1,n}, \quad \lambda_{2,n} = \frac{1}{\sqrt{2}}\left(1 - \frac{2}{n^{\frac{3}{2}}} - \frac{1}{n^2} - \frac{1}{n^8}\right)^{\frac{1}{2}} = -\rho_{2,n}, \quad \delta^{1,n} = \frac{1}{n}, \quad \delta^{2,n} = \frac{1}{n^4}$$

and

$$\gamma^{1,n} = \gamma^{0,1} + \frac{1}{n}, \quad \gamma^{2,n} = \gamma^{0,2} + \frac{1}{n^4}.$$

(Simple computations show that (2.6) *is satisfied.)*
We obtain $\lim_{n\to\infty} \boldsymbol{\beta_n} = (0, \frac{1}{\sqrt{2}}, \gamma^{0,1}, \gamma^{0,2}, 0, 0, 0, -\frac{1}{\sqrt{2}})' = \tilde{\boldsymbol{\beta}} \in \overline{\mathcal{B}} \backslash \mathcal{B}$. *A Taylor expansion up to the second order leads to*

$$N(\boldsymbol{\beta_n}) \;=\; \left\| d^{num}_{\tilde{\boldsymbol{\beta}}_n} + (\partial_{\boldsymbol{\beta}} d^{num}_{\boldsymbol{\beta}}|_{\boldsymbol{\beta}=\tilde{\boldsymbol{\beta}}})(\boldsymbol{\beta_n} - \tilde{\boldsymbol{\beta}}) + \frac{1}{2}(\boldsymbol{\beta_n} - \tilde{\boldsymbol{\beta}})'(\partial^2_{\boldsymbol{\beta}} d^{num}_{\boldsymbol{\beta}}|_{\boldsymbol{\beta}=\tilde{\boldsymbol{\beta}}})(\boldsymbol{\beta_n} - \tilde{\boldsymbol{\beta}}) + o(1) \right\|_{L_2(g^0\nu)}$$

with

$$\begin{aligned}
d^{num}_{\tilde{\boldsymbol{\beta}}_n} &= \frac{1}{\sqrt{2}}\frac{f_{\gamma^{0,2}}}{g^0} - \frac{1}{\sqrt{2}}\frac{f_{\gamma^{0,2}}}{g^0} \quad = \quad 0, \\
(\partial_{\boldsymbol{\beta}} d^{num}_{\boldsymbol{\beta}}|_{\boldsymbol{\beta}=\tilde{\boldsymbol{\beta}}})(\boldsymbol{\beta_n} - \tilde{\boldsymbol{\beta}}) &= \frac{f_{\gamma^{0,1}}}{g^0}(\lambda_{1,n} + \rho_{1,n}) + \frac{f_{\gamma^{0,2}}}{g^0}\Big(\lambda_{2,n} + \rho_{2,n} + \frac{1}{\sqrt{2}} - \frac{1}{\sqrt{2}}\Big) \\
&\quad + \Big(\frac{1}{\sqrt{2}}(\gamma^{2,n} - \gamma^{0,2}) + \delta^{2,n}\Big)\frac{\partial_{\gamma^{2,n}} f_{\gamma^{2,n}}|_{\gamma^{2,n}=\gamma^{0,2}}}{g^0}, \\
\frac{1}{2}(\boldsymbol{\beta_n} - \tilde{\boldsymbol{\beta}})'(\partial^2_{\boldsymbol{\beta}} d^{num}_{\boldsymbol{\beta}}|_{\boldsymbol{\beta}=\tilde{\boldsymbol{\beta}}})(\boldsymbol{\beta_n} - \tilde{\boldsymbol{\beta}}) &= \frac{1}{2\sqrt{2}}(\gamma^{2,n} - \gamma^{0,2})^2\frac{\partial^2_{\gamma^{2,n}} f_{\gamma^{2,n}}|_{\gamma^{2,n}=\gamma^{0,2}}}{g^0} \\
&\quad + \lambda_{1,n}(\gamma^{1,n} - \gamma^{0,1})\frac{\partial_{\gamma^{1,n}} f_{\gamma^{1,n}}|_{\gamma^{1,n}=\gamma^{0,1}}}{g^0} \\
&\quad + (\lambda_{2,n} - \frac{1}{\sqrt{2}})(\gamma^{2,n} - \gamma^{0,2})\frac{\partial_{\gamma^{2,n}} f_{\gamma^{2,n}}|_{\gamma^{2,n}=\gamma^{0,2}}}{g^0}.
\end{aligned}$$

Hence,

$$\begin{aligned}
&N(\boldsymbol{\beta_n}) \\
= \;& \Bigg\| \lambda_{1,n}(\gamma^{1,n} - \gamma^{0,1})\frac{\partial_{\gamma^{1,n}} f_{\gamma^{1,n}}|_{\gamma^{1,n}=\gamma^{0,1}}}{g^0} + \Big(\delta^{2,n} + \lambda_{2,n}(\gamma^{2,n} - \gamma^{0,2})\Big)\frac{\partial_{\gamma^{2,n}} f_{\gamma^{2,n}}|_{\gamma^{2,n}=\gamma^{0,2}}}{g^0} \\
&\qquad + \frac{1}{2\sqrt{2}}(\gamma^{2,n} - \gamma^{0,2})^2\frac{\partial^2_{\gamma^{2,n}} f_{\gamma^{2,n}}|_{\gamma^{2,n}=\gamma^{0,2}}}{g^0} \Bigg\|_{L_2(g^0\nu)} \\
\le \;& c'\Big\{\Big(\lambda_{1,n}(\gamma^{1,n} - \gamma^{0,1})\Big)^2 + \Big(\delta^{2,n} + \lambda_{2,n}(\gamma^{2,n} - \gamma^{0,2})\Big)^2 + \frac{1}{8}(\gamma^{2,n} - \gamma^{0,2})^4\Big\}^{\frac{1}{2}} \qquad (2.52)
\end{aligned}$$

for some $c' > 0$, *where the last relation is a consequence of Lemma* 2.17. *Finally, it results from the form of* $\boldsymbol{\beta_n}$

$$\lim_{n\to\infty} \frac{\sup_{l=1,2} |\delta^{l,n}|^2}{N(\boldsymbol{\beta_n})} \quad = \quad \lim_{n\to\infty} \frac{|\delta^{1,n}|^2}{N(\boldsymbol{\beta_n})}$$

and

$$\begin{aligned}
& \lim_{n\to\infty} \frac{\sup_{l=1,2}|\delta^{l,n}|^2}{N(\boldsymbol{\beta_n})} \\
\overset{(2.52)}{\geq}\; & \lim_{n\to\infty} \frac{1}{c'}\,\frac{1}{n^2}\left\{\left(n^{-\frac{3}{2}}\frac{1}{n}\right)^2 + \left(\frac{1}{n^4}+\frac{1}{\sqrt{2}}\left[1-\frac{2}{n^{\frac{3}{2}}}-\frac{1}{n^2}-\frac{1}{n^8}\right]^{\frac{1}{2}}\frac{1}{n^4}\right)^2 - \frac{1}{8}\left(\frac{1}{n^4}\right)^4\right\}^{-\frac{1}{2}} \\
=\; & \infty.
\end{aligned}$$

2.2.3. Asymptotic Results for Testing Q Populations Against p Populations

In the last subsection we studied the convergence behaviour of the statistic

$$T_n(p) \quad = \quad \sup_{g_{(\theta,\boldsymbol{\beta})}\in\mathcal{G}_p} l_n(\theta,\boldsymbol{\beta}) - l_n(0)$$

for the testing problem of $g = g^0$ against $g \in \mathcal{G}_p\backslash\{g^0\}$. This statistic and its limiting distribution form the basic expression for the general test

$$\boldsymbol{H_0 : g \in \mathcal{G}_Q} \quad \text{against} \quad \boldsymbol{H_1 : g \in \mathcal{G}_p\backslash\mathcal{G}_Q}$$

for a given true density $g^0 \in \mathcal{G}_Q$ of true order q belonging to the null hypothesis. According to definition (1.1) of the corresponding log-LRT statistic and the monotonicity properties of the logarithm we obtain

$$\begin{aligned}
\log\Big(\Lambda_n(\boldsymbol{x_1},\ldots,\boldsymbol{x_n})\Big) &= \sup_{g_{(\theta,\boldsymbol{\beta})}\in\mathcal{G}_p} l_n(\theta,\boldsymbol{\beta}) - l_n(0) \quad - \sup_{g_{(\theta,\boldsymbol{\beta})}\in\mathcal{G}_Q} l_n(\theta,\boldsymbol{\beta}) + l_n(0) \\
&= T_n(p) - T_n(Q)
\end{aligned}$$

for a suitable parameterization of $\mathcal{G}_Q$. Of course, we need a locally conic parameterization of $\mathcal{G}_Q$ as well as the Donsker class property of the corresponding set of score functions $\mathcal{D}_0$ to ensure the existence of an asymptotic distribution of $T_n(Q)$. The sufficient conditions for a convergence in distribution of $T_n(Q)$ (under our modifications) are given by

Lemma 2.24 *Let* (A1), (A2) *and* (ID) *be satisfied. Then the following statements hold:*

1) *Let each* $g_{(\theta_{\sigma(g)},\boldsymbol{\beta_{\sigma(g)}})} \in \mathcal{G}_Q\backslash\{g^0\}$ *be defined as in Lemma* 2.5 *with permutation* $\sigma(g)$ *of* $\{1,\ldots,Q\}$ *given by Algorithm* 2.7. *We define*

$$\begin{aligned}
\mathcal{B}_0 &= \Big\{\boldsymbol{\beta_{\sigma(g)}} : g \in \mathcal{G}_Q\backslash\{g^0\}\Big\} \quad \textit{and} \\
\mathcal{T}_0 &= \Big\{(\theta,\boldsymbol{\beta_{\sigma(g)}}) : \theta \leq \theta_{\sigma(g)}, g \in \mathcal{G}_Q\backslash\{g^0\}\Big\}.
\end{aligned}$$

Then there exists some $M_0 > 0$ *such that* $\mathcal{T}_0 \subset [0, M_0] \times \mathcal{B}_0$ *and* $\Big\{g_{(\theta,\boldsymbol{\beta})} : (\theta,\boldsymbol{\beta}) \in \mathcal{T}_0\Big\}$ *is a locally conic model of* $\mathcal{G}_Q$.

2) *Suppose that at least one of the following conditions holds:*

   a) (P0) *and* (P1),

   b) (P0), (P1 t) *and* (M1) *for* $Q = q$, $p = q + 1$,

c) (P0), (P1 t) *and* (M2) *for* $q = 1$ *and* $p = 2$.

Then the closure $\overline{\mathcal{D}}_0$ *of* $\mathcal{D}_0 = \left\{ \frac{\partial_\theta g(\theta,\beta)|_{\theta=0}}{g^0} : \beta \in \mathcal{B}_0 \right\}$ *is a Donsker class.*

Notice that under assumption c) of Lemma 2.24 $\mathcal{B}_0$ is defined in the usual way, i.e. without any further restrictions as used for its corresponding $\mathcal{B}$ (see Lemma 2.16).

Proof of Lemma 2.24:

1) follows directly from Proposition 2.10. Clearly we have $\mathcal{B}_0 \subset \mathcal{B}$ and thus $\mathcal{D}_0 \subset \mathcal{D}$.

2) is a consequence of Theorem 2.15 which claims that subsets of Donsker classes are also Donsker classes. □

The representation of $\mathcal{D}_0$ is exactly of the same manner as given by (2.31), namely

$$\mathcal{D}_0 = \left\{ \frac{1}{N(\beta)} \left(\sum_{l=1}^{q} \pi_l^0 \sum_{i=1}^{k} \delta_i^l \frac{\partial_{\gamma_i} f_\gamma|_{\gamma=\gamma^{0,l}}}{g^0} + \sum_{i=1}^{Q-q} \lambda_i \frac{f_{\gamma^i}}{g^0} + \sum_{l=1}^{q} \rho_l \frac{f_{\gamma^{0,l}}}{g^0} \right) : \beta \in \mathcal{B}_0 \right\}, \quad (2.53)$$

where the middle sum vanishes if $Q = q$ and the last sum vanishes if $q = 1$ according to (2.3) (since $\rho_1 = 0$.)

Thus it follows from Lemma 2.24 (and previous statements on $T_n(p)$ analogously applied to $T_n(Q)$) that Theorem 2.21 also holds for $T_n(Q)$. Consequently, we obtain

Theorem 2.25 *Let* (A1), (A2) *and* (ID) *be satisfied. Suppose that at least one of the following assumptions holds:*

a) (P0) *and* (P1),

b) (P0), (P1 t) *and* (M1) *for* $Q = q$, $p = q + 1$,

c) (P0), (P1 t) *and* (M2) *for* $q = 1$, $p = 2$ *and* $\mathcal{G}_{2,M}$, $\mathcal{B}$ *from Lemma* 2.16 b).

Let $\mathcal{B}_0$ *be given as in Lemma* 2.24. *Then the log-LRT statistic for testing Q populations against p populations*

$$\log\Big(\Lambda_n(\boldsymbol{x_1}, \ldots, \boldsymbol{x_n})\Big) = \log\Big(T_n(p)\Big) - \log\Big(T_n(Q)\Big)$$

converges in distribution to the following variable

$$\frac{1}{2} \sup_{d \in \mathcal{D}} \left(\xi_d\right)^2 \cdot \mathbb{1}_{\xi_d \geq 0} - \frac{1}{2} \sup_{d \in \mathcal{D}_0} \left(\xi_d\right)^2 \cdot \mathbb{1}_{\xi_d \geq 0}, \quad (2.54)$$

where $\left(\xi_d\right)_{\mathcal{D}}$ *and* $\left(\xi_d\right)_{\mathcal{D}_0}$ *are centered Gaussian processes with covariance given by the usual scalar product in* $L_2(g^0\nu)$.

It is possible to describe the asymptotic distribution by *one* Gaussian process instead of using the difference of *two* Gaussian processes (2.54). This can be done by using the fact that $\mathcal{D}_0$ is a subset of $\mathcal{D}$ (see [DCG99], Theorem 3.6). In the next subsection we give the corresponding limiting expression for testing one population against two populations under our first modification by (P1 t) and (M1).

2.2.4. Testing One Population Against Two Populations

In this subsection we derive a limiting process with an index set given by the component parameter space $\Gamma\backslash\{\boldsymbol{\gamma}^{\mathbf{0}}\}$ instead of an index set given by a Donsker class of functions. In order to apply the asymptotic theory in an actual testing problem this of course simplifies the calculation of the quantiles of the limiting distribution.

We construct the corresponding Gaussian process according to Theorem 3.6 and Corollary 3.7 from [DCG99]. Since [DCG99] consider only a one-dimensional component parameter space Γ in Corollary 3.7, we generalize their result on component parameter spaces with $dim(\Gamma) = k$ using our first modification (P1 t) plus (M1). As mentioned above property (P1) is a special case of this modification for testing one population against two populations. A generalization of Corollary 3.7 from [DCG99] with the aid of our second modification (P1 t) and (M2) is not possible because of the structure of the parameter space belonging to the two population mixtures.

Let the true density be given by $g^0 = f_{\boldsymbol{\gamma}^0} \in \mathcal{G}_1$. Since $Q = q = 1$ and $p = 2$ we obtain from (2.31) and (2.53) that

$$\mathcal{D} = \left\{ \frac{1}{N(\boldsymbol{\beta})} \left(\sum_{i=1}^{k} \delta_i \frac{\partial_{\gamma_i} f_{\boldsymbol{\gamma}}|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^0}}{f_{\boldsymbol{\gamma}^0}} + \lambda \frac{f_{\boldsymbol{\gamma}} - f_{\boldsymbol{\gamma}^0}}{f_{\boldsymbol{\gamma}^0}} \right) : \ \boldsymbol{\beta} = (\lambda, \boldsymbol{\gamma}, \boldsymbol{\delta}, -\lambda)' \in \mathcal{B} \right\}, \tag{2.55}$$

$$\mathcal{D}_0 = \left\{ \frac{1}{N(\boldsymbol{\beta})} \left(\sum_{i=1}^{k} \delta_i \frac{\partial_{\gamma_i} f_{\boldsymbol{\gamma}}|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^0}}{f_{\boldsymbol{\gamma}^0}} \right) : \ \boldsymbol{\beta} = \boldsymbol{\delta} \in \mathcal{B}_0 \right\} \quad \text{with } \mathcal{B}_0 \overset{(2.6)}{=} \left\{ \boldsymbol{\delta} : \|\boldsymbol{\delta}\|^2 = 1 \right\}. \tag{2.56}$$

Thereby $\mathcal{D}$ can also be indexed by $\tilde{\boldsymbol{\beta}} = (\boldsymbol{\gamma}, \boldsymbol{\delta})' \in \tilde{\mathcal{B}} = \Gamma\backslash\{\boldsymbol{\gamma}^{\mathbf{0}}\} \times \{\boldsymbol{\delta} : \|\boldsymbol{\delta}\|^2 \leq 1\}$ because of Proposition 2.10 and (2.6).

For $i = 1, \ldots, k$ and $\boldsymbol{\gamma} \in \Gamma\backslash\{\boldsymbol{\gamma}^{\mathbf{0}}\} \subset I\!R^k$ let

$$e_i = \frac{\partial_{\gamma_i} f_{\boldsymbol{\gamma}}\big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^0}}{f_{\boldsymbol{\gamma}^0}} \left\| \frac{\partial_{\gamma_i} f_{\boldsymbol{\gamma}}\big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^0}}{f_{\boldsymbol{\gamma}^0}} \right\|^{-1}_{L_2(f_{\boldsymbol{\gamma}^0}\nu)}$$

and

$$v_{\boldsymbol{\gamma}} = \frac{f_{\boldsymbol{\gamma}} - f_{\boldsymbol{\gamma}^0}}{f_{\boldsymbol{\gamma}^0}} \left\| \frac{f_{\boldsymbol{\gamma}} - f_{\boldsymbol{\gamma}^0}}{f_{\boldsymbol{\gamma}^0}} \right\|^{-1}_{L_2(f_{\boldsymbol{\gamma}^0}\nu)}.$$

Then $e_1, \ldots, e_k$ and $v_{\boldsymbol{\gamma}}$ are linearly independent in $L_2(f_{\boldsymbol{\gamma}^0}\nu)$ according to (P1 t). Further let $\{o_1, \ldots, o_k, h_{\boldsymbol{\gamma}}\}$ be the corresponding set of orthonormal functions in $L_2(f_{\boldsymbol{\gamma}^0}\nu)$. Then $\mathcal{D}_0$ can be described by the set $\{o_1, \ldots, o_k\}$ of orthonormal functions in $L_2(f_{\boldsymbol{\gamma}^0}\nu)$

$$\mathcal{D}_0 = \left\{ \sum_{i=1}^{k} \tau_i o_i : \sum_{i=1}^{k} \tau_i^2 = 1 \right\}$$

since $o_1, \ldots, o_k \in \mathcal{D}_0$ and $\sum\limits_{i=1}^{k} \tau_i^2 = 1$ lead to $\left(\sum\limits_{i=1}^{k} \tau_i o_i\right) \left\| \sum\limits_{i=1}^{k} \tau_i o_i \right\|^{-1} = \sum\limits_{i=1}^{k} \tau_i o_i \in \mathcal{D}_0$. Correspondingly, $\mathcal{D}$ is the set of functions of the form

$$\mathcal{D} = \left\{ \sum_{i=1}^{k} \tau_i o_i + \tau h_{\boldsymbol{\gamma}} : \sum_{i=1}^{k} \tau_i^2 + \tau^2 = 1, \ \tau \geq 0 \right\}. \tag{2.57}$$

The set $\{o_1, \ldots, o_k\}$ can be constructed by the orthogonalization procedure of Schmidt. This procedure and an additional normalization lead to

$$h_\gamma \quad = \quad \frac{\frac{f_\gamma - f_{\gamma^0}}{f_{\gamma^0}} - \sum_{i=1}^{k} o_i \left\langle o_i, \frac{f_\gamma - f_{\gamma^0}}{f_{\gamma^0}} \right\rangle_{L_2(f_{\gamma^0}\nu)}}{\left\| \frac{f_\gamma - f_{\gamma^0}}{f_{\gamma^0}} - \sum_{i=1}^{k} o_i \left\langle o_i, \frac{f_\gamma - f_{\gamma^0}}{f_{\gamma^0}} \right\rangle_{L_2(f_{\gamma^0}\nu)} \right\|_{L_2(f_{\gamma^0}\nu)}}. \tag{2.58}$$

Corollary 2.26 *Let* (A1), (A2), (ID), (P0), (P1 t) *and* (M1) *be satisfied for* $q = 1$ *and* $p = 2$. *Further let* h_γ *be given by formula* (2.58). *Then the log-LRT statistic for testing one population against two populations*

$$\log\Big(\Lambda_n(\boldsymbol{x_1}, \ldots, \boldsymbol{x_n})\Big) \quad = \quad \log\Big(T_n(2)\Big) - \log\Big(T_n(1)\Big)$$

converges in distribution to the following variable

$$\frac{1}{2} \sup_{\gamma \in \Gamma \setminus \{\gamma^0\}} \left(\xi_{h_\gamma}\right)^2 \cdot \mathbb{1}_{\xi_{h_\gamma} \geq 0}.$$

Proof of Corollary 2.26.:
Let $\mathcal{H} = \Big\{h_\gamma : \boldsymbol{\gamma} \in \Gamma \setminus \{\boldsymbol{\gamma}^{\mathbf{0}}\}\Big\}$. The representation (2.57) of $\mathcal{D}$ leads to

$$\sup_{d \in \mathcal{D}} \left(\xi_d\right)^2 \cdot \mathbb{1}_{\xi_d \geq 0} \quad = \quad \sup_{\substack{h_\gamma \in \mathcal{H} \\ d_0 \in \mathcal{D}_0}} \; \sup_{\substack{\tau \geq 0 \\ \kappa^2 + \tau^2 = 1}} \left(\xi_{\kappa d_0 + \tau h_\gamma}\right)^2 \cdot \mathbb{1}_{\xi_{\kappa d_0 + \tau h_\gamma} \geq 0}.$$

Since $(\xi_d)_{d \in \mathcal{D}}$ is a linear process (see e.g. Dudley (1999), p. 92) it follows that

$$\xi_{\kappa d_0 + \tau h_\gamma} \quad = \quad \kappa \xi_{d_0} + \tau \xi_{h_\gamma}.$$

Let $\angle(a, b)$ denote the angle between a and b. For fixed ξ_{d_0} and ξ_{h_γ} we obtain

$$\begin{aligned} & \sup_{\substack{\tau \geq 0 \\ \kappa^2 + \tau^2 = 1}} \left(\kappa \xi_{d_0} + \tau \xi_{h_\gamma}\right)^2 \cdot \mathbb{1}_{\kappa \xi_{d_0} + \tau h_\gamma \geq 0} \\ = \; & \sup_{\substack{\tau \geq 0 \\ \kappa^2 + \tau^2 = 1}} \left(\|(\kappa, \tau)'\| \, \|(\xi_{d_0}, \xi_{h_\gamma})'\| \, \cos\left\{ \angle\Big((\kappa, \tau)', (\xi_{d_0}, \xi_{h_\gamma})'\Big) \right\} \right)^2 \cdot \mathbb{1}_{\xi_{\kappa d_0 + \tau h_\gamma} \geq 0}. \end{aligned} \tag{2.59}$$

If $\xi_{h_\gamma} \geq 0$ then the supremum of the right hand side of (2.59) is in $(\kappa, \tau)' = \|(\xi_{d_0}, \xi_{h_\gamma})'\|^{-1} (\xi_{d_0}, \xi_{h_\gamma})'$ and it follows

$$\sup_{\substack{\tau \geq 0 \\ \kappa^2 + \tau^2 = 1}} \left(\kappa \xi_{d_0} + \tau \xi_{h_\gamma}\right)^2 \cdot \mathbb{1}_{\kappa \xi_{d_0} + \tau h_\gamma \geq 0} \quad = \quad (\xi_{d_0})^2 + (\xi_{h_\gamma})^2 \mathbb{1}_{\xi_{h_\gamma} \geq 0}. \tag{2.60}$$

If $\xi_{h_\gamma} < 0$ then the supremum of the right hand side of (2.59) will be reached for $\tau = 0$ and $\kappa = 1$ or $\kappa = -1$ whichever ξ_{d_0} is positive or negative. Thus (2.60) also holds in this case. Consequently, we have

$$\sup_{d \in \mathcal{D}} \left(\xi_d\right)^2 \cdot \mathbb{1}_{\xi_d \geq 0} \quad = \quad \sup_{d_0 \in \mathcal{D}_0} (\xi_{d_0})^2 + \sup_{h_\gamma \in \mathcal{H}} (\xi_{h_\gamma})^2 \mathbb{1}_{\xi_{h_\gamma} \geq 0}.$$

Since $\mathcal{D}_0$ is a symmetrical set it follows that

$$\sup_{d_0 \in \mathcal{D}_0} (\xi_{d_0})^2 \mathbb{1}_{\xi_{d_0} \geq 0} \quad = \quad \sup_{d_0 \in \mathcal{D}_0} (\xi_{d_0})^2$$

and Theorem 2.25 leads to an asymptotic distribution given by the following variable

$$\frac{1}{2} \sup_{d \in \mathcal{D}} (\xi_d)^2 \mathbb{1}_{\xi_d \geq 0} - \frac{1}{2} \sup_{d_0 \in \mathcal{D}_0} (\xi_{d_0})^2 \mathbb{1}_{\xi_{d_0} \geq 0} \quad = \quad \frac{1}{2} \sup_{\boldsymbol{\gamma} \in \Gamma \setminus \{\boldsymbol{\gamma}^{\mathbf{0}}\}} (\xi_{h_{\boldsymbol{\gamma}}})^2 \mathbb{1}_{\xi_{h_{\boldsymbol{\gamma}}} \geq 0}. \tag{2.61}$$

Thereby (2.61) is a consequence of $\boldsymbol{\gamma} \mapsto h_{\boldsymbol{\gamma}}$ being a continuous function which is also bijective according to (P1 t). $\square$

3. Application to Mixtures of Exponential Families

In the variety of published literature about mixture models and their applications to testing and estimating problems, examples of mixtures of specific distribution families are often given (see introduction). Thus mixture models are of theoretical interest as well as interesting for applications. In this chapter we deal with the question to which mixture families the theory from [DCG99] is applicable. As mentioned above, Keribin (2000) applies her results to Poisson distributions as well as to multivariate Gaussian families with unknown mean and known covariance matrix being a multiple of the identity matrix. Likewise, Ciuperca (1998) follows the theory from [DCG97] and shows that it is not applicable to simple mixtures of translation families of exponential densities. Due to the observation that all aforementioned promising distribution families belong to exponential families and that the application, given in Ciuperca (1998), does not belong to them, we decided to study mixtures of exponential families.

Our main purpose for this chapter is to develop sufficient conditions to apply the theory from [DCG99] which depend on the statistic and the parameter function of corresponding densities in exponential representation. These conditions are easy to handle if the component parameter space is of dimension one or two, but their complexity increases with the dimensionality of the parameter space. If the parameter function has a diagonal Jacobian matrix then a verification of our derived sufficient conditions is quite simple. The latter special case leads to an elementary proof that the theory from [DCG99] is applicable to m-dimensional Gaussian mixtures with unknown mean and known arbitrary covariance matrix (see Example 3.15).

3.1. Basic Definitions and Properties of Exponential Families

In this section we provide a basis of definitions and properties of exponential families which will be subsequently used. This general survey is given according to Witting (1985). Further we also refer to Barndorff-Nielsen (1978) even if we do not cite him explicitly.

Definition 3.1 (exponential family)(Witting (1985), Definition 1.150)
Let $(\mathfrak{X}, \mathcal{X})$ be a measurable space with $\mathfrak{X} \subset \mathbb{R}^m$. A family $\mathcal{P} = \{P_{\boldsymbol{\gamma}} : \boldsymbol{\gamma} \in \Gamma\}$, $\Gamma \subset \mathbb{R}^\kappa$, of probability measures on $(\mathfrak{X}, \mathcal{X})$ is said to be an exponential family provided there exists a σ-finite measure ν on $(\mathfrak{X}, \mathcal{X})$ such that $\mathcal{P}$ is dominated by ν and there exists an integer $k \in \mathbb{N}$, real-valued functions $A, \zeta_1, \ldots, \zeta_k : \Gamma \to \mathbb{R}$ as well as real-valued $\mathcal{X}$-measurable functions $r, T_1, \ldots, T_k : \mathfrak{X} \to \mathbb{R}$ and for every $P_{\boldsymbol{\gamma}} \in \mathcal{P}$ its corresponding ν-density has an exponential representation

$$f_{\boldsymbol{\gamma}}(\boldsymbol{x}) \quad = \quad A(\boldsymbol{\gamma}) \exp\left\{\left\langle \boldsymbol{\zeta}(\boldsymbol{\gamma}), \boldsymbol{T}(\boldsymbol{x})\right\rangle\right\} r(\boldsymbol{x}) \qquad \nu - a.s., \tag{3.1}$$

where $\boldsymbol{\zeta} = (\zeta_1, \ldots, \zeta_k)'$, $\boldsymbol{T} = (T_1, \ldots, T_k)'$, $\langle \boldsymbol{\zeta}(\boldsymbol{\gamma}), \boldsymbol{T}(\boldsymbol{x}) \rangle = \sum_{i=1}^{k} \zeta_i(\boldsymbol{\gamma}) T_i(\boldsymbol{x})$ *and normalizing factor* $A(\boldsymbol{\gamma}) = \Big(\int \exp\left\{ \langle \boldsymbol{\zeta}(\boldsymbol{\gamma}), \boldsymbol{T}(\boldsymbol{x}) \rangle \right\} r(\boldsymbol{x})\, \nu(d\boldsymbol{x}) \Big)^{-1}$. $\mathcal{P}$ *is also called k-parametric exponential family in* $\boldsymbol{\zeta}$ *and* $\boldsymbol{T}$.

As $f_{\boldsymbol{\gamma}}$ is a non-negative $\mathcal{X}$-measurable function we assume without loss of generality that $r(\boldsymbol{x}) \geq 0$ for all $\boldsymbol{x} \in \mathfrak{X}$ and $A(\boldsymbol{\gamma}) > 0$ for all $\boldsymbol{\gamma} \in \Gamma$. According to Witting (1985), p. 143, we define

$$\mu(B) \quad = \quad \int_B r(\boldsymbol{x})\, \nu(d\boldsymbol{x}), \quad B \in \mathcal{X}, \tag{3.2}$$

as an equivalent σ-finite measure to the probability measures $P_{\boldsymbol{\gamma}}$, $\boldsymbol{\gamma} \in \Gamma$, i.e. $P_{\boldsymbol{\gamma}}(N) = 0$ for all $\boldsymbol{\gamma} \in \Gamma \Leftrightarrow \mu(N) = 0$. Then every probability measure $P_{\boldsymbol{\gamma}} \in \mathcal{P}$ has also a corresponding μ-density

$$p_{\boldsymbol{\gamma}}(\boldsymbol{x}) \quad = \quad \mathcal{C}(\boldsymbol{\zeta}(\boldsymbol{\gamma})) \exp\Big\{ \langle \boldsymbol{\zeta}(\boldsymbol{\gamma}), \boldsymbol{T}(\boldsymbol{x}) \rangle \Big\} \qquad \mu - a.s., \tag{3.3}$$

where $\mathcal{C}(\boldsymbol{\zeta}(\boldsymbol{\gamma})) = A(\boldsymbol{\gamma}) > 0$ is an appropriate normalizing factor. Consequently, all statistical properties of $\mathcal{P}$ result from the character of $\exp\{\langle \boldsymbol{\zeta}(\boldsymbol{\gamma}), \boldsymbol{T}(\boldsymbol{x}) \rangle\}$.

Notice that neither the parameter function $\boldsymbol{\zeta}$ nor the statistic $\boldsymbol{T}$ is uniquely defined. Thus $\mathcal{P}$ has several representations. In the next sections we often make use of

Definition 3.2 (canonical parameterization)
Let $\mathcal{P}$ *be a k-parametric exponential family in* $\boldsymbol{\zeta}$ *and* $\boldsymbol{T}$ *such that the mapping* $\boldsymbol{\zeta} : \Gamma \to \boldsymbol{\zeta}(\Gamma) = Z \subset I\!R^k$ *is one-to-one. The family* $\{P_{\boldsymbol{\zeta}(\boldsymbol{\gamma})} : \boldsymbol{\gamma} \in \Gamma\}$ *is a parameterization of* $\mathcal{P}$ *and is equal to* $\{P_{\boldsymbol{\zeta}(\boldsymbol{\gamma})} : \boldsymbol{\zeta}(\boldsymbol{\gamma}) \in Z\}$ *since the function* $\boldsymbol{\zeta}(\boldsymbol{\gamma})$ *is injective. According to Witting, p. 149, we also write* $\{P_{\boldsymbol{\zeta}} : \boldsymbol{\zeta} \in Z\}$. *This parameterization of* $\mathcal{P}$ *is called canonical. Thus* $\boldsymbol{\zeta}, \boldsymbol{\zeta}(\boldsymbol{\gamma})$ *denote points in* Z *and if we mean the function* $\boldsymbol{\zeta} : \Gamma \to \boldsymbol{\zeta}(\Gamma)$ *it will be marked correspondingly. The set*

$$Z_* \quad = \quad \Big\{ \boldsymbol{\zeta} \in I\!R^k : 0 < \int \exp\Big\{ \langle \boldsymbol{\zeta}, \boldsymbol{T}(\boldsymbol{x}) \rangle \Big\} \mu(d\boldsymbol{x}) < \infty \Big\} \tag{3.4}$$

is called canonical parameter space. Clearly, one has $Z \subset Z_* \subset I\!R^k$.
The family $\mathcal{P} = \{P_{\boldsymbol{\zeta}} : \boldsymbol{\zeta} \in Z\}$ *in canonical parameterization has* μ*-densities of the form*

$$\tilde{p}_{\boldsymbol{\zeta}}(\boldsymbol{x}) \quad = \quad \mathcal{C}(\boldsymbol{\zeta}) \exp\Big\{ \langle \boldsymbol{\zeta}, \boldsymbol{T}(\boldsymbol{x}) \rangle \Big\} \qquad \mu - a.s., \tag{3.5}$$

which is equivalent to

$$\tilde{p}_{\boldsymbol{\zeta}}(\boldsymbol{x}) \quad = \quad \exp\Big\{ \langle \boldsymbol{\zeta}, \boldsymbol{T}(\boldsymbol{x}) \rangle - K(\boldsymbol{\zeta}) \Big\} \quad \mu - a.s, \tag{3.6}$$

where the normalizing factor is given by

$$\mathcal{C}(\boldsymbol{\zeta}) \quad = \quad \Big(\int \exp\{\langle \boldsymbol{\zeta}, \boldsymbol{T}(\boldsymbol{x}) \rangle\}\, \mu(d\boldsymbol{x}) \Big)^{-1} \tag{3.7}$$

and

$$K(\boldsymbol{\zeta}) \quad = \quad \log\Big(\int \exp\{\langle \boldsymbol{\zeta}, \boldsymbol{T}(\boldsymbol{x}) \rangle\}\, \mu(d\boldsymbol{x}) \Big) \quad = \quad -\log(\mathcal{C}(\boldsymbol{\zeta}))$$

is called log-Laplace transform or cumulate generating function. Under a canonical representation of $\mathcal{P}$ *the measurable function* $\boldsymbol{T}$ *is called generating statistic.*

According to Barndorff-Nielsen (1978), p. 112, we have

Definition 3.3 (minimal parameterization, order)
Let $(\mathfrak{X}, \mathcal{X})$ be a measurable space with $\mathfrak{X} \subset I\!R^m$ and let $\mathcal{P} = \{P_{\boldsymbol{\gamma}} : \boldsymbol{\gamma} \in \Gamma\}$, $\Gamma \subset I\!R^\kappa$, be a family of probability measures on $(\mathfrak{X}, \mathcal{X})$. Let $k_{min}(\nu)$ be the smallest integer such that the probability measures in $\mathcal{P}$ with respect to ν are representable as in (3.1) for $A, \zeta_1, \ldots, \zeta_{k_{min}(\nu)} : \Gamma \to I\!R$ as well as for $r, T_1, \ldots, T_{k_{min}(\nu)} : \mathfrak{X} \to I\!R$. $k_{min(\nu)}$ is called the order of $\mathcal{P}$ and is denoted by $ord(\mathcal{P})$. Any representation (3.1) with $ord(\mathcal{P}) = k_{min(\nu)}$ is said to minimal .

Factually, $k_{min}(\nu)$ is an integer independent of ν and $dim(\Gamma)$. In Proposition 1.153 given in Witting (1985) necessary and sufficient conditions for minimal representations of (3.1) are given. They are based on the following

Definition 3.4 (affinely independent, $\boldsymbol{\nu}$-affinely independent)
For $\Gamma \subset I\!R^\kappa$ and $\mathfrak{X} \subset I\!R^m$ let $\zeta_i : \Gamma \to I\!R$ and $T_i : \mathfrak{X} \to I\!R$, $i = 1, \ldots, k$.

a) *The functions $\zeta_1, \ldots, \zeta_k$ are called affinely independent if the functions $1, \zeta_1, \ldots, \zeta_k$ are linearly independent, i.e. for $\boldsymbol{a} \in I\!R^k$ and $a_0 \in I\!R$ the following statement holds:*

$$\big\langle \boldsymbol{\zeta}(\boldsymbol{\gamma}), \boldsymbol{a} \big\rangle = a_0 \quad \text{for all } \boldsymbol{\gamma} \in \Gamma \quad \Rightarrow \quad \boldsymbol{a} = \boldsymbol{0},\ a_0 = 0.$$

b) *$T_1, \ldots, T_k$ are called ν-affinely independent if for any ν-null set the functions $T_1, \ldots, T_k$ are affinely independent on the corresponding complementary set, i.e. for $\boldsymbol{b} \in I\!R^k$ and $b_0 \in I\!R$ the following statement holds:*

$$\big\langle \boldsymbol{b}, \boldsymbol{T}(\boldsymbol{x}) \big\rangle = b_0 \quad \nu\text{-}a.s. \quad \Rightarrow \quad \boldsymbol{b} = \boldsymbol{0},\ b_0 = 0.$$

Proposition 3.5 (Witting (1985), Proposition 1.153)
An exponential family $\mathcal{P}$ in $\boldsymbol{\zeta}$ and T is minimal k-parametric iff its ν-densities have a representation as in (3.1) and both of the following conditions are satisfied:

a) *The functions $\zeta_1, \ldots, \zeta_k$ are affinely independent.*

b) *The functions $T_1, \ldots, T_k$ are ν-affinely independent.*

As mentioned above ν and μ are equivalent measures and thus $T_1, \ldots, T_k$ are ν-affinely independent iff they are μ-affinely independent.

Notice that even if $\mathcal{P}$ is minimal, its corresponding parameter function $\boldsymbol{\zeta}$ and statistic $\boldsymbol{T}$ are only uniquely defined with respect to non-degenerate affine transformations and non-degenerate ν-affine transformations, respectively (see Witting (1985), Corollary 1.154).

A basic tool for the next sections is given by

Proposition 3.6 (Witting (1985), Proposition 1.164)
Let $\mathcal{P}$ be a k-parametric exponential family in $\boldsymbol{\zeta}$ and $\boldsymbol{T}$ with μ-densities in canonical representation (3.5) or (3.6) and canonical parameter space Z_*.

Then the generating statistic $\boldsymbol{T}$ *possesses mixed moments of arbitrary order with respect to* $\tilde{p}_{\boldsymbol{\zeta}}\mu$, $\boldsymbol{\zeta} \in Z_*$. *For all* $\boldsymbol{\zeta} \in int(Z_*)$ *the functions* $\boldsymbol{\zeta} \mapsto E_{\boldsymbol{\zeta}} T_1^{l_1} \dots T_k^{l_k}$, $\boldsymbol{\zeta} \mapsto K(\boldsymbol{\zeta})$ *and* $\boldsymbol{\zeta} \mapsto \mathcal{C}(\boldsymbol{\zeta})$ *are continuous and infinitely often differentiable. In particular, for all* $\boldsymbol{\zeta} \in int(Z_*)$ *one has*

$$E_{\boldsymbol{\zeta}} \boldsymbol{T} = \partial_{\boldsymbol{\zeta}} K(\boldsymbol{\zeta}) = -\partial_{\boldsymbol{\zeta}} \log \Big(\mathcal{C}(\boldsymbol{\zeta})\Big), \tag{3.8}$$

$$Cov_{\boldsymbol{\zeta}} \boldsymbol{T} = \partial^2_{\boldsymbol{\zeta}} K(\boldsymbol{\zeta}) = -\partial^2_{\boldsymbol{\zeta}} \log \Big(\mathcal{C}(\boldsymbol{\zeta})\Big), \tag{3.9}$$

$$E_{\boldsymbol{\zeta}} T_1^{l_1} \dots T_k^{l_k} = \mathcal{C}(\boldsymbol{\zeta}) \partial^{l_1}_{\zeta_1} \dots \partial^{l_k}_{\zeta_k} \int \exp\Big\{\langle \boldsymbol{\zeta}, \boldsymbol{T}(\boldsymbol{x})\rangle\Big\} \mu(d\boldsymbol{x}) \quad \forall (l_1, \dots, l_k) \in I\!N_0^k, \tag{3.10}$$

where $\partial_{\boldsymbol{\zeta}} K(\boldsymbol{\zeta})$ *and* $\partial^2_{\boldsymbol{\zeta}} K(\boldsymbol{\zeta})$ *denote the gradient and the Hessian matrix of* $K(\boldsymbol{\zeta})$, *respectively. Further* $\partial^{l_i}_{\zeta_i} h_{\boldsymbol{\zeta}}$, $i \in \{1, \dots, k\}$, *denotes the* l_i*th partial derivative of* $h_{\boldsymbol{\zeta}}$ *with respect to* ζ_i.

In our following work we assume that $\mathcal{P} = \{P_{\boldsymbol{\gamma}} : \boldsymbol{\gamma} \in \Gamma\}$, $\Gamma \subset I\!R^k$, is a minimal k-parametric exponential family in $\boldsymbol{\zeta}$ and $\boldsymbol{T}$ and that $\mathcal{F} = \{f_{\boldsymbol{\gamma}} : \boldsymbol{\gamma} \in \Gamma\}$ is the corresponding family of ν-densities in exponential representation (3.1) with respect to the probability measures of $\mathcal{P}$. Furthermore, we assume that $\boldsymbol{\zeta} : \Gamma \to Z = \boldsymbol{\zeta}(\Gamma)(\subset I\!R^k)$ is a bijective function which assures the existence of a canonical parameterization of $\mathcal{P}$. According to the above definitions and properties we make use of

$$p_{\boldsymbol{\gamma}}(\boldsymbol{x}) \overset{(3.3)}{=} \mathcal{C}(\boldsymbol{\zeta}(\boldsymbol{\gamma})) \exp\Big\{\langle \boldsymbol{\zeta}(\boldsymbol{\gamma}), \boldsymbol{T}(\boldsymbol{x})\rangle\Big\} \overset{(3.5)}{=} \tilde{p}_{\boldsymbol{\zeta}(\boldsymbol{\gamma})}(\boldsymbol{x}) \qquad \mu - a.s., \tag{3.11}$$

$$f_{\boldsymbol{\gamma}}(\boldsymbol{x}) \overset{(3.1)}{=} r(\boldsymbol{x}) p_{\boldsymbol{\gamma}}(\boldsymbol{x}) = r(\boldsymbol{x}) \tilde{p}_{\boldsymbol{\zeta}(\boldsymbol{\gamma})}(\boldsymbol{x}) \qquad \nu - a.s., \tag{3.12}$$

where $\mathcal{C}(\boldsymbol{\zeta}(\boldsymbol{\gamma})) = A(\boldsymbol{\gamma})$ and the measure μ defined in (3.2). Correspondingly, we define the true μ-density (in canonical parameterization) by

$$w^0 = \sum_{l=1}^{q} \pi_l^0 p_{\boldsymbol{\gamma}^{0,l}} \quad \mu - a.s., \qquad \tilde{w}^0 = \sum_{l=1}^{q} \pi_l^0 \tilde{p}_{\boldsymbol{\zeta}^{0,l}} \quad \mu - a.s. \tag{3.13}$$

for $\boldsymbol{\zeta}^{\mathbf{0},\boldsymbol{l}} = \boldsymbol{\zeta}(\boldsymbol{\gamma}^{\mathbf{0},\boldsymbol{l}})$, $l = 1, \dots, q$. Clearly, we obtain

$$g^0(\boldsymbol{x}) = \sum_{l=1}^{q} \pi_l^0 f_{\boldsymbol{\gamma}^{0,l}}(\boldsymbol{x}) = r(\boldsymbol{x}) w^0(\boldsymbol{x}) = r(\boldsymbol{x}) \tilde{w}^0(\boldsymbol{x}) \quad \nu - a.s. \tag{3.14}$$

From now on we assume that the latter equations ν-a.s. and μ-a.s., respectively, hold.

3.2. Applications to Mixtures in Exponential Representation

In this section we present our main Theorem 3.13 on the applicability of the asymptotic theory from [DCG99] for testing "Q populations against p populations" to exponential families. More precisely, in Theorem 3.13 we give sufficient conditions for the ν-densities of $\mathcal{F} = \{f_{\boldsymbol{\gamma}} : \boldsymbol{\gamma} \in \Gamma\}$ to hold with respect to an exponential family in $\boldsymbol{\zeta}$ and $\boldsymbol{T}$ such that the properties (P0), (P1), (P1 t), (M1) and (M2) hold (see p. 21-23). For this we assume that (A1), (A2) as well as (ID) are satisfied. We base our statements on the last section, especially on the equations (3.11)-(3.14) and on the fact that $\boldsymbol{T}$ possesses mixed moments of arbitrary order with respect to $\tilde{p}_{\boldsymbol{\zeta}}\mu$, $\boldsymbol{\zeta} \in Z_*$, due to Proposition 3.6. It will be seen that the structure of the component parameter space Γ is of essential importance.

First of all, we derive a general expression for arbitrary partial derivatives $\partial_{\gamma_{i_1}} \ldots \partial_{\gamma_{i_n}} f_{\boldsymbol{\gamma}}$ (see Lemma 3.9) since all properties (P0), (P1), (P1 t), (M1) and (M2) require certain properties for the partial derivatives of $f_{\boldsymbol{\gamma}}$ with respect to $\boldsymbol{\gamma}$ to hold. This expression results from the fact that $f_{\boldsymbol{\gamma}}(\boldsymbol{x}) = r(\boldsymbol{x})p_{\boldsymbol{\gamma}}(\boldsymbol{x}) = r(\boldsymbol{x})\tilde{p}_{\boldsymbol{\zeta}(\boldsymbol{\gamma})}(\boldsymbol{x})$ and that $r(\boldsymbol{x})$ does not depend on $\boldsymbol{\gamma}$. Beforehand we make use of the analytic properties of $\mathcal{C}(\boldsymbol{\zeta})$ given by Proposition 3.6 to obtain a general expression for $\partial_{\zeta_{i_1}} \ldots \partial_{\zeta_{i_n}} \tilde{p}_{\boldsymbol{\zeta}}$ (see Lemma 3.7).

The following (technical) lemma gives a general (complicated) expression of the partial derivatives of $\tilde{p}_{\boldsymbol{\zeta}}$ with respect to $\boldsymbol{\zeta}$, but its basic statement will be summarized in Remark 3.8. Roughly speaking, Lemma 3.7 says that any partial derivative (with respect to $\boldsymbol{\zeta}$) of $\tilde{p}_{\boldsymbol{\zeta}}(\boldsymbol{x})$ is a product of $\tilde{p}_{\boldsymbol{\zeta}}(\boldsymbol{x})$ itself and some polynomial in $\boldsymbol{T}(\boldsymbol{x})$.

Lemma 3.7 *Let $\mathcal{P}$ be a k-parametric exponential family in $\boldsymbol{\zeta}$ and $\boldsymbol{T}$ with μ-densities in canonical representation* (3.5) *or* (3.6) *and canonical parameter space Z_*. For $n \in I\!N$ let $i_1, \ldots, i_n \in \{1, \ldots, k\}$ and $A^n = \{1, \ldots, n\}$. Then we have*

$$\partial_{\zeta_{i_1}} \ldots \partial_{\zeta_{i_n}} \tilde{p}_{\boldsymbol{\zeta}}(\boldsymbol{x}) \;=\; \sum_{\substack{B_1, \ldots, B_n \subset A^n \\ \text{disjoint}}} \left\{ a_{(B_1, \ldots, B_n)} \Big(\prod_{l=1}^{n} E_{\boldsymbol{\zeta}} \Big(\prod_{j \in B_l} T_{i_j} \Big) \Big) \prod_{j \in A^n \setminus \left(\bigcup_{t=1}^{n} B_t \right)} T_{i_j}(\boldsymbol{x}) \right\},$$

where $a_{(B_1, \ldots, B_n)} \in I\!R$ are suitable factors and $E_{\boldsymbol{\zeta}}\Big(\prod_{j \in \emptyset} T_{i_j} \Big) = \prod_{j \in \emptyset} T_{i_j}(\boldsymbol{x}) = 1$.
(Notice that B_i, $i = 1, \ldots, n$, may take the form $B_i = \emptyset$ and that we sum all combinations of disjoint subsets of A^n. Furthermore, there may exist some $r, s \in A^n$, $r \neq s$, such that $i_r = i_s$.)

The proof of Lemma 3.7 via induction is given in the appendix.

As any factor $a_{(B_1, \ldots, B_n)} \prod_{l=1}^{n} E_{\boldsymbol{\zeta}}(\prod_{j \in B_l} T_{i_j})$ of Lemma 3.7 is real-valued it follows

Remark 3.8 *Let the assumptions of Lemma 3.7 be satisfied. For $n \in I\!N$ and $i_1, \ldots, i_n \in \{1, \ldots, k\}$*

$$\partial_{\zeta_{i_1}} \ldots \partial_{\zeta_{i_n}} \tilde{p}_{\boldsymbol{\zeta}}(\boldsymbol{x}) \;=\; P^{\boldsymbol{\zeta}}_{\zeta_{i_1} \ldots \zeta_{i_n}}(\boldsymbol{T}(\boldsymbol{x}))\, \tilde{p}_{\boldsymbol{\zeta}}(\boldsymbol{x}),$$

where $P^{\boldsymbol{\zeta}}_{\zeta_{i_1} \ldots \zeta_{i_n}}(\boldsymbol{T}(\boldsymbol{x}))$ is a polynomial in the components of $\boldsymbol{T}(\boldsymbol{x})$, i.e. $T_1(\boldsymbol{x}), \ldots, T_k(\boldsymbol{x})$. In particular, one has

$$\partial_{\zeta_{i_1}} \tilde{p}_{\boldsymbol{\zeta}}(\boldsymbol{x}) \overset{\text{(A.20)}}{=} \Big(T_{i_1}(\boldsymbol{x}) - E_{\boldsymbol{\zeta}} T_{i_1} \Big) \tilde{p}_{\boldsymbol{\zeta}}(\boldsymbol{x}), \tag{3.15}$$

$$\partial_{\zeta_{i_1}} \partial_{\zeta_{i_2}} \tilde{p}_{\boldsymbol{\zeta}}(\boldsymbol{x}) \overset{\text{(3.15),(A.23)}}{=} \Big\{ \Big(T_{i_1}(\boldsymbol{x}) - E_{\boldsymbol{\zeta}} T_{i_1} \Big) \Big(T_{i_2}(\boldsymbol{x}) - E_{\boldsymbol{\zeta}} T_{i_2} \Big) - Cov_{\boldsymbol{\zeta}} T_{i_1} T_{i_2} \Big\} \tilde{p}_{\boldsymbol{\zeta}}(\boldsymbol{x}). \tag{3.16}$$

Using the fact that $f_{\boldsymbol{\gamma}}(\boldsymbol{x}) \overset{(3.12)}{=} r(\boldsymbol{x})\tilde{p}_{\boldsymbol{\zeta}(\boldsymbol{\gamma})}(\boldsymbol{x})$, where $r(\boldsymbol{x})$ is independent of $\boldsymbol{\gamma}$, we obtain the partial derivatives of $f_{\boldsymbol{\gamma}}$ with respect to $\boldsymbol{\gamma}$ by

Lemma 3.9 *Let $\mathcal{P} = \{P_{\boldsymbol{\gamma}} : \boldsymbol{\gamma} \in \Gamma\}$, $\Gamma \subset I\!R^k$, be a k-parametric exponential family in $\boldsymbol{\zeta}$ and $\boldsymbol{T}$ and let $\mathcal{F} = \{f_{\boldsymbol{\gamma}} : \boldsymbol{\gamma} \in \Gamma\}$ is the corresponding family of ν-densities in exponential representation* (3.1) *with respect to the probability measures of $\mathcal{P}$. Further let $\boldsymbol{\zeta} : \Gamma \to Z$ be a bijective function which ensures the existence of a canonical parameterization $\{P_{\boldsymbol{\zeta}} : \boldsymbol{\zeta} \in Z\}$ of $\mathcal{P}$ with μ-densities* *of the form* (3.5) *or* (3.6). *If* $\boldsymbol{\zeta}:\Gamma\to Z$ *possesses partial derivatives up to order* n, $n\in\mathbb{N}$, *then we have*

$$\begin{aligned}&\partial_{\gamma_{i_1}}\ldots\partial_{\gamma_{i_n}} f_{\boldsymbol{\gamma}}(\boldsymbol{x})\\ =\ & r(\boldsymbol{x})\sum_{m=1}^{n}\sum_{j_1=1}^{k}\cdots\sum_{j_m=1}^{k}\left\{\left(\partial_{\zeta_{j_1}(\boldsymbol{\gamma})}\ldots\partial_{\zeta_{j_m}(\boldsymbol{\gamma})}\tilde{p}_{\boldsymbol{\zeta}(\boldsymbol{\gamma})}(\boldsymbol{x})\right)\sum_{\substack{A_1^n,\ldots,A_m^n\neq\emptyset\\ \text{disjoint}\\ A_1^n\cup\ldots\cup A_m^n=A^n}}\left(b_{(A_1^n,\ldots,A_m^n)}\prod_{i=1}^{m}\partial_{A_i^n}\zeta_{j_i}(\boldsymbol{\gamma})\right)\right\},\end{aligned}$$

where $i_1,\ldots,i_n\in\{1,\ldots,k\}$, $A^n=\{1,\ldots,n\}$, $\partial_{\{1,\ldots,l\}}\zeta_{j_i}(\boldsymbol{\gamma})=\partial_{\gamma_{i_1}}\ldots\partial_{\gamma_{i_l}}\zeta_{j_i}(\boldsymbol{\gamma})$ *for* $l\in\{1,\ldots,n\}$ *and suitable factors* $b_{(A_1^n,\ldots,A_m^n)}\in\mathbb{R}$.
(Notice that we sum all combinations of disjoint non-empty subsets of A^n*. Furthermore, there may exist some* $r,s\in A^n$, $r\neq s$, *such that* $i_r=i_s$*.)*

The proof of Lemma 3.9 via induction is given in the appendix.

Lemma 3.9 says, if the parameter function $\boldsymbol{\zeta}:\Gamma\to Z$ possesses derivatives up to order n then $\partial_{\gamma_{i_1}}\ldots\partial_{\gamma_{i_n}} f_{\boldsymbol{\gamma}}(\boldsymbol{x})$ is equal to a product of $f_{\boldsymbol{\gamma}}(\boldsymbol{x})=r(\boldsymbol{x})\tilde{p}_{\boldsymbol{\zeta}(\boldsymbol{\gamma})}(\boldsymbol{x})$ and some polynomial in the components of the statistic $\boldsymbol{T}(\boldsymbol{x})$ due to Lemma 3.7. Notice that the real-valued factors $b_{(A_1^n,\ldots,A_m^n)}\prod_{i=1}^{m}\partial_{A_i^n}\zeta_{j_i}(\boldsymbol{\gamma})$ do not depend on $\boldsymbol{x}$.

For any polynomial $P(\boldsymbol{T}(\boldsymbol{x}))$ in the components of $\boldsymbol{T}(\boldsymbol{x})$ and $m\in\mathbb{N}$ it holds

$$\begin{aligned}\int\left|\frac{P(\boldsymbol{T}(\boldsymbol{x}))f_{\boldsymbol{\gamma}^{0,l}}(\boldsymbol{x})}{g^0(\boldsymbol{x})}\right|^m g^0(\boldsymbol{x})\,\nu(d\boldsymbol{x}) &\leq \int\left|\frac{P(\boldsymbol{T}(\boldsymbol{x}))f_{\boldsymbol{\gamma}^{0,l}}(\boldsymbol{x})}{\pi_l^0 f_{\boldsymbol{\gamma}^{0,l}}(\boldsymbol{x})}\right|^m g^0(\boldsymbol{x})\,\nu(d\boldsymbol{x})\\ &\overset{(3.13),(3.14)}{=} \sum_{j=1}^{q}\int\left|\frac{P(\boldsymbol{T}(\boldsymbol{x}))}{\pi_l^0}\right|^m \tilde{p}_{\boldsymbol{\zeta}^{0,j}}(\boldsymbol{x})\,\mu(d\boldsymbol{x}) < \infty\end{aligned}$$

since $\boldsymbol{T}$ possesses mixed moments of arbitrary order with respect to $\tilde{p}_{\boldsymbol{\gamma}^{0,j}}\mu$, $j=1,\ldots,q$, due to Proposition 3.6 and $\pi_l^0>0$ according to assumption (A1). As a consequence we obtain

$$\frac{\partial_{\gamma_{i_1}}\ldots\partial_{\gamma_{i_n}} f_{\boldsymbol{\gamma}}|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,l}}}{g^0}\in L_m(g^0\nu)\qquad\text{for all } n,m\in\mathbb{N}\tag{3.17}$$

if the assumptions of Lemma 3.9 are satisfied and $\boldsymbol{\zeta}:\Gamma\to Z$ possesses partial derivatives up to order n. For $m=3$ and $n=1,\ldots,5$ formula (3.17) is equivalent to property (P0)b). If $m=2$ then (3.17) leads to $\frac{f_{\boldsymbol{\gamma}^{0,l}}}{g^0}$, $\frac{\partial_{\gamma_i} f_{\boldsymbol{\gamma}}|_{\boldsymbol{\gamma}^{0,l}}}{g^0}$, $\frac{\partial_{\gamma_i\gamma_j} f_{\boldsymbol{\gamma}}|_{\boldsymbol{\gamma}^{0,l}}}{g^0}\in L_2(g^0\nu)$ as assumed by (P1), (P1 t), (M1) and (M2). In properties (P1) and (P1 t) it is additionally assumed that $\frac{f_{\boldsymbol{\gamma}}}{g^0}\in L_2(g^0\nu)$ for any $\boldsymbol{\gamma}\in\Gamma\backslash\{\boldsymbol{\gamma}^{0,1},\ldots,\boldsymbol{\gamma}^{0,q}\}$ and according to property (P0)c) we also need a sufficient condition for $\frac{f_{\boldsymbol{\gamma}}}{g^0}$ being in $L_3(g^0\nu)$ for any $\boldsymbol{\gamma}\in\overline{U}_\varepsilon(\boldsymbol{\gamma}^{0,l})=\{\boldsymbol{\gamma}\in\Gamma:\ \|\boldsymbol{\gamma}-\boldsymbol{\gamma}^{0,l}\|\leq\varepsilon\}$, $l=1,\ldots,q$ and some $\varepsilon>0$.

Lemma 3.10 *Let the assumptions of Lemma* 3.9 *as well as* (A1) *be satisfied. (As aforementioned* $\boldsymbol{\zeta},\boldsymbol{\zeta}(\boldsymbol{\gamma})$ *denote points in* Z *and if we mean the function* $\boldsymbol{\zeta}:\Gamma\to\boldsymbol{\zeta}(\Gamma)$ *it will be marked correspondingly.) For* $m\in\mathbb{N}$ *let*

$$\tilde{Z}_m\ =\ \left\{\boldsymbol{\zeta}\in Z_*: m\boldsymbol{\zeta}-(m-1)\boldsymbol{\zeta}(\boldsymbol{\gamma}^{0,l})\in Z_*\ \text{for some } l\in\{1,\ldots,q\}\right\}.$$

Then the following statements hold :

a) $\boldsymbol{\zeta}(\boldsymbol{\gamma}) \in \tilde{Z}_m \;\Rightarrow\; \frac{f_{\boldsymbol{\gamma}}}{g^0} \in L_m(g^0\nu)$,

b) *for* $q=1$: $\boldsymbol{\zeta}(\boldsymbol{\gamma}) \in \tilde{Z}_m \;\Leftrightarrow\; \frac{f_{\boldsymbol{\gamma}}}{g^0} \in L_m(g^0\nu)$.

Proof of Lemma 3.10:

a) Clearly, one obtains $\boldsymbol{\zeta^{0,l}} = \boldsymbol{\zeta}(\boldsymbol{\gamma^{0,l}}) \in \tilde{Z}_m$ for any $l \in \{1,\ldots,q\}$. Thus the parameter $\boldsymbol{\zeta} = \boldsymbol{\zeta}(\boldsymbol{\gamma}) \in \tilde{Z}_m$ leads to

$$0 \quad < \quad \mathcal{C}(\boldsymbol{\zeta}),\ \mathcal{C}(\boldsymbol{\zeta^{0,l}}),\ \mathcal{C}(m\boldsymbol{\zeta} - (m-1)\boldsymbol{\zeta^{0,l}}) \quad < \quad \infty \tag{3.18}$$

for some $l \in \{1,\ldots,q\}$. (Notice that $m\boldsymbol{\zeta} - (m-1)\boldsymbol{\zeta^{0,l}}$ describes a point in the canonical parameter space Z_* which leads to $0 < \mathcal{C}(m\boldsymbol{\zeta} - (m-1)\boldsymbol{\zeta^{0,l}}) < \infty$ according to Definition 3.2.) Consequently,

$$\begin{aligned}
0 \quad &< \quad \int \left|\frac{f_{\boldsymbol{\gamma}}(\boldsymbol{x})}{g^0(\boldsymbol{x})}\right|^m g^0(\boldsymbol{x})\,\nu(d\boldsymbol{x}) && (3.19)\\
&\overset{(3.12)}{\leq} \int \frac{|\tilde{p}_{\boldsymbol{\zeta}}(\boldsymbol{x})|^m}{|\pi_l^0 \tilde{p}_{\boldsymbol{\zeta^{0,l}}}(\boldsymbol{x})|^{m-1}}\,\mu(d\boldsymbol{x}) && (3.20)\\
&\overset{(3.5)}{=} \int \frac{\big(\mathcal{C}(\boldsymbol{\zeta})\big)^m}{(\pi_l^0\mathcal{C}(\boldsymbol{\zeta^{0,l}}))^{m-1}}\ \exp\Big\{m\langle\boldsymbol{\zeta},\boldsymbol{T}(\boldsymbol{x})\rangle - (m-1)\big\langle\boldsymbol{\zeta^{0,l}},\boldsymbol{T}(\boldsymbol{x})\big\rangle\Big\}\,\mu(d\boldsymbol{x}) &&\\
&\overset{(3.5)}{=} \frac{\big(\mathcal{C}(\boldsymbol{\zeta})\big)^m}{(\pi_l^0\mathcal{C}(\boldsymbol{\zeta^{0,l}}))^{m-1}\mathcal{C}(m\boldsymbol{\zeta} - (m-1)\boldsymbol{\zeta^{0,l}})} \int \tilde{p}_{m\boldsymbol{\zeta}-(m-1)\boldsymbol{\zeta^{0,l}}}(\boldsymbol{x})\,\mu(d\boldsymbol{x}) && (3.21)
\end{aligned}$$

which is finite according to (3.18) and to (A1) (i.e. $\pi_l^0 > 0$).

b) Obviously, equality in formula (3.20) holds iff $q = 1$. Suppose that $\boldsymbol{\zeta}$, $\boldsymbol{\zeta^{0,1}} \in Z_*$ and $m\boldsymbol{\zeta} - (m-1)\boldsymbol{\zeta^{0,1}} \notin Z_*$. Then the expression in (3.21) is either equal to 0 which is in contradiction to (3.19) or the expression in (3.21) is infinite which leads to $\frac{f_{\boldsymbol{\gamma}}}{g^0} \notin L_m(g^0\nu)$. Consequently, it follows assertion b) with the aid of a). □

For $m \in I\!N$ and any $\boldsymbol{\gamma}$ such that $\boldsymbol{\zeta}(\boldsymbol{\gamma}) \in \tilde{Z}_m$ holds and under the assumptions of Lemma 3.9 and (A1), we obtain according to (3.17)

$$\frac{\partial_{\gamma_{i_1}} \ldots \partial_{\gamma_{i_n}} f_{\boldsymbol{\gamma}}}{g^0} \in L_m(g^0\nu) \qquad \text{for all } n \in I\!N. \tag{3.22}$$

In our context, it may be a difficult task to verify property (P0)c); i.e. to ensure for $m = 3$, $n = 1,\ldots,5$, that the supremum of (3.22) on $\overline{U}_\varepsilon(\boldsymbol{\gamma^{0,l}})$, $l = 1,\ldots,q$, is a function which belongs to $L_3(g^0\nu)$. In Proposition 3.11 we give a sufficient condition for applying (P0)c) which is based on the trivial facts that the exponential function is strictly monotonic increasing, g^0 has fixed parameter components $\boldsymbol{\gamma^{0,1}},\ldots,\boldsymbol{\gamma^{0,q}}$ and that $\boldsymbol{T}$ possesses mixed moments of arbitrary order with respect to $\tilde{w}^0\mu$ which is equivalent to $g^0\nu$. In the proof of Proposition 3.11 we construct a corresponding supremum function $m_{j_1,\ldots,j_n}$, $n = 1,\ldots,5$, essentially by a product of $(g^0)^{-1}$ and majorant densities $\tilde{p}_{\boldsymbol{u^i}}(\boldsymbol{x_i})$ defined on appropriate regions G_i, $i = 1,\ldots,m$. The advantage is that we investigate the supremum of a scalar product $\big\langle\boldsymbol{\zeta}(\boldsymbol{\gamma}),\boldsymbol{T}(\boldsymbol{x})\big\rangle$ on $\overline{U}_\varepsilon(\boldsymbol{\gamma^{0,l}})$ instead of the supremum of a ratio of two densities via (3.22). The sufficient condition $\boldsymbol{u^i} \in \tilde{Z}_3$

in Proposition 3.11 ensures that $\tilde{p}_{\boldsymbol{u^i}}(\boldsymbol{x_i}) \in L_3(g^0\nu)$. If $Z_* = \mathbb{R}^k$ then the latter condition certainly holds. But if $Z_* \neq \mathbb{R}^k$ as, for instance, for Gaussian families with unknown covariance matrix, families of negative binomial distributions and Gamma distributions, one has to choose a suitable component parameter space Γ (see section 3.2). (Notice that we do not assume that $\boldsymbol{\zeta}(\overline{U}_\varepsilon(\boldsymbol{\gamma^{0,l}})) \subset \tilde{Z}_3$, $l = 1, \ldots, q$.)

In section 3.2 we use Proposition 3.11 as an elementary (simple) tool for verifying condition (P0)c) in all examples of exponential families we considered in our work. Thanks to its generality we do not need to analyze the geometric structure of the densities of $\mathcal{F} = \{f_{\boldsymbol{\gamma}} : \boldsymbol{\gamma} \in \Gamma\}$ to obtain an appropriate supremum function $m_{j_1,\ldots,j_n}$, $n = 1, \ldots, 5$. It will be seen in Example 3.15 that although we generalize the example of Gaussian mixtures given by Keribin (2000), p. 59-60, our method is much easier to handle since $Z_* = \mathbb{R}^k$.

Proposition 3.11 *Let the assumptions of Lemma* 3.9 *and assumption* (A1) *be satisfied. Suppose that* $\boldsymbol{\zeta} : \Gamma \to Z = \boldsymbol{\zeta}(\Gamma)$ *possesses partial derivatives up to order* 5. *Furthermore, suppose that there exists some* $\varepsilon > 0$, *a disjoint decomposition* $G_1 \cup \ldots \cup G_m = \mathfrak{X}$ *and some parameters* $\boldsymbol{u^i} \in \tilde{Z}_3$, $i = 1, \ldots, m$, *(see Lemma* 3.10*) such that for any* $l = 1, \ldots, q$ *and any* $\boldsymbol{x} \in G_i$, $i = 1, \ldots, m$,

$$\sup_{\boldsymbol{\gamma} \in \overline{U}_\varepsilon(\boldsymbol{\gamma^{0,l}})} \big\langle \boldsymbol{\zeta}(\boldsymbol{\gamma}), \boldsymbol{T}(\boldsymbol{x}) \big\rangle \;\leq\; \big\langle \boldsymbol{u^i}, \boldsymbol{T}(\boldsymbol{x}) \big\rangle \qquad g^0\nu - a.s.$$

Then for all $n \leq 5$ *and all* $j_1, \ldots, j_n \in \{1, \ldots, k\}$ *there exists a function* $m_{j_1,\ldots,j_n}$ *such that for any* $l = 1, \ldots, q$

$$\sup_{\boldsymbol{\gamma} \in \overline{U}_\varepsilon(\boldsymbol{\gamma^{0,l}})} \left|\frac{f_{\boldsymbol{\gamma}}}{g^0}\right| \left|P^{\boldsymbol{\zeta}(\boldsymbol{\gamma})}_{\zeta_{j_1}\ldots\zeta_{j_n}}\right| \;\leq\; m_{j_1,\ldots,j_n} \tag{3.23}$$

and $E_{g^0\nu}(m^3_{j_1,\ldots,j_n}) < \infty$ *hold.*

Proof of Proposition 3.11:
Obviously, for $n \in \{1, \ldots, 5\}$ one has

$$\sup_{\boldsymbol{\gamma} \in \overline{U}_\varepsilon(\boldsymbol{\gamma^{0,l}})} \left|\frac{f_{\boldsymbol{\gamma}}}{g^0}\right| \left|P^{\boldsymbol{\zeta}(\boldsymbol{\gamma})}_{\zeta_{j_1}\ldots\zeta_{j_n}}\right| \;\leq\; \left(\sup_{\boldsymbol{\gamma} \in \overline{U}_\varepsilon(\boldsymbol{\gamma^{0,l}})} \left|\frac{f_{\boldsymbol{\gamma}}}{g^0}\right|\right)\left(\sup_{\boldsymbol{\gamma} \in \overline{U}_\varepsilon(\boldsymbol{\gamma^{0,l}})} \left|P^{\boldsymbol{\zeta}(\boldsymbol{\gamma})}_{\zeta_{j_1}\ldots\zeta_{j_n}}\right|\right). \tag{3.24}$$

Firstly, we evaluate $\displaystyle\sup_{\boldsymbol{\gamma} \in \overline{U}_\varepsilon(\boldsymbol{\gamma^{0,l}})} \left|P^{\boldsymbol{\zeta}(\boldsymbol{\gamma})}_{\zeta_{j_1}\ldots\zeta_{j_n}}\right|$:

For any $\boldsymbol{x} \in \mathfrak{X}$ one has according to Lemma 3.7 and Remark 3.8

$$P^{\boldsymbol{\zeta}(\boldsymbol{\gamma})}_{\zeta_{j_1}\ldots\zeta_{j_n}}(\boldsymbol{T}(\boldsymbol{x})) \;=\; \sum_{\substack{B_1,\ldots,B_n \subset A^n \\ \text{disjoint}}} \left\{ a_{(B_1,\ldots,B_n)} \left(\prod_{l=1}^n E_{\boldsymbol{\zeta}(\boldsymbol{\gamma})}\Big(\prod_{s \in B_l} T_{j_s} \Big) \right) \prod_{\substack{s \in A^n \setminus B^n \\ B^n = B_1 \cup \ldots \cup B_n}} T_{j_s}(\boldsymbol{x}) \right\},$$

where $A^n = \{1, \ldots, n\}$ and $a_{(B_1,\ldots,B_n)} \in \mathbb{R}$. Since $\boldsymbol{\zeta} \mapsto E_{\boldsymbol{\zeta}} T_1^{l_1} \ldots T_k^{l_k}$ is continuous due to Proposition 3.6 and $\boldsymbol{\zeta} : \Gamma \to Z$ is also continuous, there exists some $C^{0,l} > 0$ such that

$$\sup_{\boldsymbol{\gamma} \in \overline{U}_\varepsilon(\boldsymbol{\gamma^{0,l}})} \max_{\substack{B_1,\ldots,B_n \subset A^n \\ \text{disjoint}}} \left| a_{(B_1,\ldots,B_n)} \left(\prod_{l=1}^n E_{\boldsymbol{\zeta}(\boldsymbol{\gamma})}\Big(\prod_{s \in B_l} T_{j_s} \Big) \right) \right| \;\leq\; C^{0,l} \;<\; \infty.$$

Thus we obtain

$$\sup_{\gamma\in\overline{U}_\varepsilon(\boldsymbol{\gamma}^{0,l})}\left|P^{\boldsymbol{\zeta}(\gamma)}_{\zeta_{j_1}\ldots\zeta_{j_n}}(\boldsymbol{T}(\boldsymbol{x}))\right| \;\le\; C^{0,l}\sum_{\substack{B_1,\ldots,B_n\subset A^n\\ \text{disjoint}}}\left|\prod_{\substack{s\in A^n\setminus B^n\\ B^n=B_1\cup\ldots\cup B_n}} T_{j_s}(\boldsymbol{x})\right| \quad,\ \boldsymbol{x}\in\mathfrak{X}. \tag{3.25}$$

Secondly, for any $\boldsymbol{x}\in\mathfrak{X}$ one has

$$\sup_{\gamma\in\overline{U}_\varepsilon(\boldsymbol{\gamma}^{0,l})}\frac{f_{\boldsymbol{\gamma}}(\boldsymbol{x})}{g^0(\boldsymbol{x})}\ \stackrel{(3.1)}{\le}\ \frac{1}{g^0(\boldsymbol{x})}\left(\sup_{\gamma\in\overline{U}_\varepsilon(\boldsymbol{\gamma}^{0,l})}\mathcal{C}(\boldsymbol{\zeta}(\gamma))\right)\left(\sup_{\gamma\in\overline{U}_\varepsilon(\boldsymbol{\gamma}^{0,l})}\exp\left\{\left\langle\boldsymbol{\zeta}(\gamma),\boldsymbol{T}(\boldsymbol{x})\right\rangle\right\}\right)$$

because $g^0=\sum_{l=1}^{q}\pi_l^0 f_{\boldsymbol{\gamma}^{0,l}}$ has fixed component parameters. Since $\boldsymbol{\zeta}:\Gamma\to Z$ as well as $\mathcal{C}:Z_*\to I\!R$ are continuous according to assumption and Proposition 3.6, respectively, it follows

$$U^{0,l}\;=\;\sup_{\gamma\in\overline{U}_\varepsilon(\boldsymbol{\gamma}^{0,l})}\mathcal{C}(\boldsymbol{\zeta}(\gamma))\;<\;\infty.$$

Using formula (3.25) and the above assumptions we define for all $n\le 5$ and all $j_1,\ldots,j_n\in\{1,\ldots,k\}$ and for any $\boldsymbol{x}\in G_i$, $i=1,\ldots,m$

$$m_{j_1,\ldots,j_n}(\boldsymbol{x})\;=\;C\,U\,(g^0(\boldsymbol{x}))^{-1}\exp\left\{\left\langle\boldsymbol{u^i},\boldsymbol{T}(\boldsymbol{x})\right\rangle\right\}\sum_{\substack{B_1,\ldots,B_n\subset A^n\\ \text{disjoint}}}\left|\prod_{\substack{s\in A^n\setminus B^n\\ B^n=B_1\cup\ldots\cup B_n}} T_{j_s}(\boldsymbol{x})\right|.$$

where $A^n=\{1,\ldots,n\}$, $C=\max_{l=1,\ldots,q}C^{0,l}$ and $U=\max_{l=1,\ldots,q}U^{0,l}$.

By construction the relation (3.23) is satisfied and

$$E_{g^0\nu}(m^3_{j_1,\ldots,j_n})\ \stackrel{(3.5)}{=}\ (C\,U)^3\sum_{i=1}^{m}\int_{G_i}\frac{(r(\boldsymbol{x})\tilde{p}_{\boldsymbol{u^i}}(\boldsymbol{x}))^3}{(\mathcal{C}(\boldsymbol{u^i}))^3(g^0(\boldsymbol{x}))^2}\left(\sum_{\substack{B_1,\ldots,B_n\subset A^n\\ \text{disjoint}}}\left|\prod_{\substack{s\in A^n\setminus B^n\\ B^n=B_1\cup\ldots\cup B_n}} T_{j_s}(\boldsymbol{x})\right|\right)^3\nu(d\boldsymbol{x}).$$

Since $\boldsymbol{u^i}\in\tilde{Z}_3$, $i=1,\ldots,m$, (due to the assumption) it follows that $0<\mathcal{C}(\boldsymbol{u^i})<\infty$ and there exists some $l_i\in\{1,\ldots,q\}$ such that $3\boldsymbol{u^i}-2\boldsymbol{\zeta^{0,l_i}}\in Z_*$. Consequently, we have for $i=1,\ldots,m$

$$\begin{aligned}\frac{(r(\boldsymbol{x})\tilde{p}_{\boldsymbol{u^i}}(\boldsymbol{x}))^3}{(\mathcal{C}(\boldsymbol{u^i}))^3(g^0(\boldsymbol{x}))^2}\ &\stackrel{(3.5)}{=}\ \frac{(r(\boldsymbol{x}))^3\exp\left\{\left\langle 3\boldsymbol{u^i},\boldsymbol{T}(\boldsymbol{x})\right\rangle\right\}}{(r(\boldsymbol{x})w^0(\boldsymbol{x}))^2}\\ &\le\ \frac{r(\boldsymbol{x})}{(\pi^0_{l_i})^2}\frac{\exp\left\{\left\langle 3\boldsymbol{u^i}-2\boldsymbol{\zeta^{0,l_i}},\boldsymbol{T}(\boldsymbol{x})\right\rangle\right\}}{(\mathcal{C}(\boldsymbol{\zeta^{0,l_i}}))^2}\\ &=\ \frac{r(\boldsymbol{x})}{(\pi^0_{l_i}\mathcal{C}(\boldsymbol{\zeta^{0,l_i}}))^2\,\mathcal{C}(3\boldsymbol{u^i}-2\boldsymbol{\zeta^{0,l_i}})}\tilde{p}_{3\boldsymbol{u^i}-2\boldsymbol{\zeta}^{0,l_i}}(\boldsymbol{x})\end{aligned}$$

and we obtain $E_{g^0\nu}(m^3_{j_1,\ldots,j_n})$ is less or equal to

$$(C\,U)^3\sum_{i=1}^{m}\left\{\frac{1}{(\pi^0_{l_i}\mathcal{C}(\boldsymbol{\zeta^{0,l_i}}))^2\,\mathcal{C}(3\boldsymbol{u^i}-2\boldsymbol{\zeta^{0,l_i}})}\int_{G_i}\tilde{p}_{3\boldsymbol{u^i}-2\boldsymbol{\zeta}^{0,l_i}}(\boldsymbol{x})\left(\sum_{\substack{B_1,\ldots,B_n\subset A^n\\ \text{disjoint}}}\left|\prod_{\substack{s\in A^n\setminus B^n\\ B^n=B_1\cup\ldots\cup B_n}} T_{j_s}(\boldsymbol{x})\right|\right)^3\mu(d\boldsymbol{x})\right\}.$$

Finally, it follows that $E_{g^0\nu}(m^3_{j_1,\ldots,j_n}) < \infty$ for all $n \leq 5$ and all $j_1, \ldots, j_n \in \{1, \ldots, k\}$ since the generating statistic $\boldsymbol{T}$ possesses mixed moments of arbitrary order with respect to $\tilde{p}_{3\boldsymbol{u}^i - \boldsymbol{\zeta}^{0,l_i}}\,\mu$ because of Proposition 3.6. $\square$

We give two further sufficient conditions for applying (P0)c) by the following lemma. Thereby, the second condition can be interpreted as a special case of Proposition 3.11. The first condition holds for any family $\mathcal{F} = \{f_\gamma : \gamma \in \Gamma\}$ with $f_\gamma(\boldsymbol{x})/f_{\gamma^{0,l}}(\boldsymbol{x})$ being bounded on $\overline{U}_\epsilon(\boldsymbol{\gamma}^{0,l}) \times \mathfrak{X}$ and $\boldsymbol{\zeta}(\boldsymbol{\gamma}) - \boldsymbol{\zeta}(\boldsymbol{\gamma}^{0,l}) \in Z_*$ for any $\boldsymbol{\gamma} \in \overline{U}_\epsilon(\boldsymbol{\gamma}^{0,l})$, $l = 1, \ldots, q$. Please note, that $\boldsymbol{0} \in Z_*$ does not necessary hold (e.g. see Example 3.17). The latter condition ensures that (3.26) is well-defined:

$$\begin{aligned} \frac{f_\gamma(\boldsymbol{x})}{f_{\gamma^{0,l}}(\boldsymbol{x})} &= \frac{\mathcal{C}(\boldsymbol{\zeta}(\boldsymbol{\gamma}))}{\mathcal{C}(\boldsymbol{\zeta}(\boldsymbol{\gamma}^{0,l}))} \exp\left\{ \left\langle \boldsymbol{\zeta}(\boldsymbol{\gamma}) - \boldsymbol{\zeta}(\boldsymbol{\gamma}^{0,l}), \boldsymbol{T}(\boldsymbol{x}) \right\rangle \right\} \\ &= \frac{\mathcal{C}(\boldsymbol{\zeta}(\boldsymbol{\gamma}))}{\mathcal{C}(\boldsymbol{\zeta}(\boldsymbol{\gamma}^{0,l}))\mathcal{C}(\boldsymbol{\zeta}(\boldsymbol{\gamma}) - \boldsymbol{\zeta}(\boldsymbol{\gamma}^{0,l}))} \tilde{p}_{\boldsymbol{\zeta}(\boldsymbol{\gamma}) - \boldsymbol{\zeta}(\boldsymbol{\gamma}^{0,l})}(\boldsymbol{x}). \end{aligned} \tag{3.26}$$

If for all $\boldsymbol{x} \in \mathfrak{X}$ the function $r(\boldsymbol{x}) = c$ (> 0) then for all $\boldsymbol{x} \in \mathfrak{X}$ we have $f_{\boldsymbol{\zeta}^{-1}(\boldsymbol{\zeta}(\boldsymbol{\gamma}) - \boldsymbol{\zeta}(\boldsymbol{\gamma}^{0,l}))}(\boldsymbol{x}) = c\,\tilde{p}_{\boldsymbol{\zeta}(\boldsymbol{\gamma}) - \boldsymbol{\zeta}(\boldsymbol{\gamma}^{0,l})}(\boldsymbol{x})$. Please note, if $r(\boldsymbol{x})$ is not a constant function then $\tilde{p}_{\boldsymbol{\zeta}(\boldsymbol{\gamma}) - \boldsymbol{\zeta}(\boldsymbol{\gamma}^{0,l})}(\boldsymbol{x})$ may be unbounded on $\mathfrak{X}$.

Lemma 3.12 *Let the assumptions of Lemma* 3.9 *and assumption* (A1) *be satisfied. Suppose that $\boldsymbol{\zeta} : \Gamma \to Z = \boldsymbol{\zeta}(\Gamma)$ possesses partial derivatives up to order* 5*. Further suppose that there exists some $\varepsilon > 0$ such that one of the following two conditions is fulfilled for any $l = 1, \ldots, q$:*

1) a) $S^{0,l} = \left\{ \boldsymbol{\zeta}(\boldsymbol{\gamma}) - \boldsymbol{\zeta}(\boldsymbol{\gamma}^{0,l}) : \boldsymbol{\gamma} \in \overline{U}_\varepsilon(\boldsymbol{\gamma}^{0,l}) \right\} \subset Z_*$ *and*
   b) *there exists some constant $M^{0,l} < \infty$ such that $\sup_{\boldsymbol{\zeta} \in S^{0,l}} \tilde{p}_{\boldsymbol{\zeta}}(\boldsymbol{x}) \leq M^{0,l}$ $g^0\nu$-a.s.*

2) *Suppose that the assumptions of Proposition* 3.11 *hold and there exists some $i \in \{1, \ldots, m\}$ such that $K = G_i \subset \mathfrak{X}$ is compact with*

$$\int_K \frac{1}{(g^0(\boldsymbol{x}))^2} \Bigg(\sum_{\substack{B_1, \ldots, B_n \subset A^n \\ \text{disjoint}}} \Bigg| \prod_{\substack{s \in A^n \setminus B^n \\ B^n = B_1 \cup \ldots \cup B_n}} T_{j_s}(\boldsymbol{x}) \Bigg| \Bigg)^3 \nu(d\boldsymbol{x}) \quad < \quad \infty$$

for all $n \leq 5$ and all $j_1, \ldots, j_n \in \{1, \ldots, k\}$, where $A^n = \{1, \ldots, n\}$. Further suppose that there exists some constant $M \in I\!R$ such that

$$\sup_{\boldsymbol{\gamma} \in \overline{U}_\varepsilon(\boldsymbol{\gamma}^{0,l})} \left\langle \boldsymbol{\zeta}(\boldsymbol{\gamma}), \boldsymbol{T}(\boldsymbol{x}) \right\rangle \quad \leq \quad M \quad \text{for } \boldsymbol{x} \in K.$$

Then for all $n \leq 5$ and all $j_1, \ldots, j_n \in \{1, \ldots, k\}$ there exists a function $m_{j_1,\ldots,j_n}$ such that for all $l = 1, \ldots, q$ (3.23) *as well as $E_{g^0\nu}(m^3_{j_1,\ldots,j_n}) < \infty$ hold.*

Proof of Lemma 3.12: According to the proof of Proposition 3.11 we evaluate the factors of (3.24).

1) For any $\boldsymbol{x} \in \mathfrak{X}$ one has

$$\sup_{\boldsymbol{\gamma} \in \overline{U}_\varepsilon(\boldsymbol{\gamma}^{0,l})} \frac{f_\gamma(\boldsymbol{x})}{g^0(\boldsymbol{x})} \overset{(3.26)}{\leq} \sup_{\boldsymbol{\gamma} \in \overline{U}_\varepsilon(\boldsymbol{\gamma}^{0,l})} \frac{1}{\pi^0_l} \frac{\mathcal{C}(\boldsymbol{\zeta}(\boldsymbol{\gamma}))}{\mathcal{C}(\boldsymbol{\zeta}(\boldsymbol{\gamma}^{0,l}))\,\mathcal{C}(\boldsymbol{\zeta}(\boldsymbol{\gamma}) - \boldsymbol{\zeta}(\boldsymbol{\gamma}^{0,l}))} \tilde{p}_{\boldsymbol{\zeta}(\boldsymbol{\gamma}) - \boldsymbol{\zeta}(\boldsymbol{\gamma}^{0,l})}(\boldsymbol{x}).$$

The ratio of the latter expression on the right hand side is bounded on $\overline{U}_\varepsilon(\boldsymbol{\gamma}^{0,l})$ by some $U^{0,l} < \infty$, since $\mathcal{C} : Z_* \to I\!R$ is continuous according to Proposition 3.6 and $\boldsymbol{\zeta} : \Gamma \to Z$ is also continuous. Hence, we obtain due to 1)a) and b)

$$\sup_{\boldsymbol{\gamma}\in\overline{U}_\varepsilon(\boldsymbol{\gamma}^{0,l})} \frac{f_{\boldsymbol{\gamma}}(\boldsymbol{x})}{g^0(\boldsymbol{x})} \leq \frac{1}{\pi_l^0} U^{0,l} M^{0,l} < \infty \quad g^0\nu - a.s.$$

Using expression (3.25) for $\sup_{\boldsymbol{\gamma}\in\overline{U}_\varepsilon(\boldsymbol{\gamma}^{0,l})} \left| P^{\boldsymbol{\zeta}(\boldsymbol{\gamma})}_{\zeta_{j_1}\ldots\zeta_{j_n}}(\boldsymbol{T}(\boldsymbol{x}))\right|$, we define for all $n \leq 5$ and all $j_1, \ldots, j_n \in \{1, \ldots, k\}$ the function

$$m_{j_1,\ldots,j_n}(\boldsymbol{x}) = \begin{cases} C\,M \sum\limits_{\substack{B_1,\ldots,B_n\subset A^n\\ \text{disjoint}}} \left| \prod\limits_{\substack{s\in A^n\setminus B^n\\ B^n=B_1\cup\ldots\cup B_n}} T_{j_s}(\boldsymbol{x})\right| & , \ \boldsymbol{x} \in \mathfrak{X}\setminus N \\ \infty & , \ \boldsymbol{x} \in N \end{cases}$$

where $A^n = \{1, \ldots, n\}$, $N = \left\{\boldsymbol{x} \in \mathfrak{X} : \sup_{\boldsymbol{\zeta}\in S^{0,l}} \tilde{p}_{\boldsymbol{\zeta}}(\boldsymbol{x}) = \infty\right\}$, $C = \max_{l=1,\ldots,q} C^{0,l}$ and $M = \max_{l=1,\ldots,q} \frac{1}{\pi_l^0} U^{0,l} M^{0,l}$.

By construction relation (3.23) is satisfied and $E_{g^0\nu}(m^3_{j_1,\ldots,j_n}) = E_{\tilde{w}^0\mu}(m^3_{j_1,\ldots,j_n}) < \infty$ holds for all $n \leq 5$ and all $j_1, \ldots, j_n \in \{1, \ldots, k\}$ since the generating statistic $\boldsymbol{T}$ possesses mixed moments of arbitrary order with respect to $\tilde{w}^0\mu$ thanks to Proposition 3.6 and N is a $g^0\nu$-null set according to assumption 1)b).

2) is in the same manner as the proof of Proposition 3.11. □

Now we introduce our main theorem. An illustration of its conditions is given below. Afterwards, we present the proof of Theorem 3.13.

Theorem 3.13 *Let $\mathcal{P} = \{P_{\boldsymbol{\gamma}} : \boldsymbol{\gamma} \in \Gamma\}$, $\Gamma \subset I\!R^k$, be a minimal k-parametric exponential family in $\boldsymbol{\zeta}$ and $\boldsymbol{T}$ and let $\mathcal{F} = \{f_{\boldsymbol{\gamma}} : \boldsymbol{\gamma} \in \Gamma\}$ is the corresponding family of ν-densities in exponential representation* (3.1) *with respect to the probability measures of $\mathcal{P}$. Further let $\boldsymbol{\zeta} : \Gamma \to Z$ be a bijective function which ensures the existence of a canonical parameterization $\{P_{\boldsymbol{\zeta}} : \boldsymbol{\zeta} \in Z\}$ of $\mathcal{P}$ with μ-densities of the form* (3.5) *or* (3.6)*. Furthermore, we assume that* (A1), (A2) *as well as* (ID) *are satisfied. Let us consider the following conditions:*

1) *For $r : \mathfrak{X} \to [0, \infty)$ one has $|\log(r)| \in L_1(g^0\nu)$.*

2) *$\boldsymbol{\zeta} : \Gamma \to Z = \boldsymbol{\zeta}(\Gamma)$ possesses continuous partial derivatives up to order* 5.

3) *For $l = 1, \ldots, q$ the Jacobian matrix $\partial_{\boldsymbol{\gamma}}\boldsymbol{\zeta}(\boldsymbol{\gamma})|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,l}}$ is regular.*

4) *$\boldsymbol{\zeta}(\boldsymbol{\gamma}) \in \tilde{Z}_2$ for $\boldsymbol{\gamma} \in \Gamma\setminus\{\boldsymbol{\gamma}^{0,1}, \ldots, \boldsymbol{\gamma}^{0,q}\}$ (see Lemma* 3.10*).*

5) *For $l = 1, \ldots, q$ and for all $n \leq 5$ and all $j_1, \ldots, j_n \in \{1, \ldots, k\}$ there exists a function $m_{j_1,\ldots,j_n}$ and some $\varepsilon > 0$ such that*

$$\sup_{\boldsymbol{\gamma}\in\overline{U}_\varepsilon(\boldsymbol{\gamma}^{0,l})} \left|\frac{f_{\boldsymbol{\gamma}}}{g^0}\right| \left| P^{\boldsymbol{\zeta}(\boldsymbol{\gamma})}_{\zeta_{j_1}\ldots\zeta_{j_n}} \right| \leq m_{j_1,\ldots,j_n}$$

and $E_{g^0\nu}(m^3_{j_1,\ldots,j_n}) < \infty$ hold.

6) *The functions* $T_1, \ldots, T_k, T_iT_j$, $1 \le i \le j \le k$, *are* ν*-affinely independent.*

7) *For* $l = 1, \ldots, q$ *and all* $1 \le i \le j \le k$

$$c^{0,l}_{i,j} \sum_{u=1}^{k} \sum_{v \le u}^{k} \Big\{ \tau_{u,v} \Big[\Big(\partial_{\gamma_i} \zeta_v(\boldsymbol{\gamma}) \big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,l}} \Big) \Big(\partial_{\gamma_j} \zeta_u(\boldsymbol{\gamma}) \big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,l}} \Big) + \Big(\partial_{\gamma_j} \zeta_v(\boldsymbol{\gamma}) \big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,l}} \Big) \Big(\partial_{\gamma_i} \zeta_u(\boldsymbol{\gamma}) \big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,l}} \Big) \Big] \Big\} = 0 \quad \Rightarrow \quad c^{0,l}_{i,j} = 0 \tag{3.27}$$

where $\tau_{u,v} = \begin{cases} 1 & , \ \text{for } u \neq v \\ \frac{1}{2} & , \ \text{for } u = v \end{cases}$ *for* $u, v \in \{1, \ldots, k\}$.

8) *Let* $U \subset \Big\{(u,v) \in \{1, \ldots, k\}^2 : u < v\Big\}$ *and* $U^c = \Big\{(u,v) \in \{1, \ldots, k\}^2 : u < v, (u,v) \notin U\Big\}$ *such that*

$$\mathcal{H} = \Big\{ T_1, \ldots, T_k, \ T_1T_1, \ T_2T_2, \ldots, T_kT_k, \ T_uT_v, \ (u,v) \in U \Big\}$$

is a set of ν*-affinely independent functions of maximal cardinality. Then for any* $(u,v) \in U^c$ *there exists a unique* $\kappa \in \{1, \ldots, k\}$ *and* $\alpha_\kappa \neq 0$ *such that* $T_uT_v = \alpha_\kappa T_\kappa T_\kappa$ *with* $(\min\{u,\kappa\}, \max\{u,k\})$, $(\min\{v,\kappa\}, \max\{v,k\}) \in U$ *holds.*

9) *Let* $Q = q$ *and* $p = q + 1$ *and let* U *and* U^c *from* 8)*. For* $(u,v) \in U^c$ *let*

$$H(u,v) = \Big\{ \kappa \in \{1, \ldots, k\} : T_uT_v = \alpha_\kappa T_\kappa T_\kappa \text{ for some } \alpha_\kappa \neq 0 \Big\},$$

$H = \bigcup_{(u,v) \in U^c} H(u,v)$, $H^c = \{1, \ldots, k\} \backslash H$ *and* $I = \{(1,1), \ldots, (k,k)\}$*. Then* $|H(u,v)| = 1$ *for* $(u,v) \in U^c$ *and for any* $(i,j) \in U \cup I$ *and any* $h \in \{1, \ldots, q\}$ *one has*

$$\begin{aligned} c^h_{i,j} \Big\{ & \sum_{(u,v) \in U} \Big[\Big(\partial_{\gamma_i} \zeta_v(\boldsymbol{\gamma}) \big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,h}} \Big) \Big(\partial_{\gamma_j} \zeta_u(\boldsymbol{\gamma}) \big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,h}} \Big) \\ & \qquad + \Big(\partial_{\gamma_j} \zeta_v(\boldsymbol{\gamma}) \big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,h}} \Big) \Big(\partial_{\gamma_i} \zeta_u(\boldsymbol{\gamma}) \big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,h}} \Big) \Big] \\ & + \sum_{u \in H^c} \Big[\Big(\partial_{\gamma_i} \zeta_u(\boldsymbol{\gamma}) \big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,h}} \Big) \Big(\partial_{\gamma_j} \zeta_u(\boldsymbol{\gamma}) \big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,h}} \Big) \Big] \\ & + \sum_{\substack{(u,v) \in U^c \\ \kappa \in H(u,v)}} \Big[\alpha_\kappa \Big(\partial_{\gamma_i} \zeta_v(\boldsymbol{\gamma}) \big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,h}} \Big) \Big(\partial_{\gamma_j} \zeta_u(\boldsymbol{\gamma}) \big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,h}} \Big) \\ & \qquad + \alpha_\kappa \Big(\partial_{\gamma_j} \zeta_v(\boldsymbol{\gamma}) \big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,h}} \Big) \Big(\partial_{\gamma_i} \zeta_u(\boldsymbol{\gamma}) \big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,h}} \Big) \\ & \qquad + \Big(\partial_{\gamma_j} \zeta_\kappa(\boldsymbol{\gamma}) \big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,h}} \Big) \Big(\partial_{\gamma_i} \zeta_\kappa(\boldsymbol{\gamma}) \big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,h}} \Big) \Big] \Big\} = 0 \quad \Rightarrow \quad c^h_{i,j} = 0. \end{aligned} \tag{3.28}$$

10) *Let* $\mathcal{G}_1 = \Big\{ f_{\boldsymbol{\gamma}} : \boldsymbol{\gamma} \in \Gamma \Big\}$ *for* $\Gamma \subset I\!R^2$ *and* $\mathcal{G}_{2,M} = \Big\{ \pi_1 f_{\boldsymbol{\gamma}^1} + \pi_2 f_{\boldsymbol{\gamma}^2} : (\boldsymbol{\gamma}^1, \boldsymbol{\gamma}^2) \in \tilde{\Gamma}, \ \boldsymbol{\pi} \in \Pi_2 \Big\}$, *where*

$$\tilde{\Gamma} = \Big\{ (\boldsymbol{\gamma}^1, \boldsymbol{\gamma}^2) \in \Gamma \times \Gamma : \|\boldsymbol{\gamma}^1 - \boldsymbol{\gamma}^{0,1}\| \le \|\boldsymbol{\gamma}^2 - \boldsymbol{\gamma}^{0,1}\|, \ |\gamma_1^2 - \gamma_1^{0,1}| \le |\gamma_2^2 - \gamma_2^{0,1}| \Big\}.$$

Furthermore, $\partial_{\gamma_2} \zeta_2(\boldsymbol{\gamma})|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,1}} \neq 0$, *the functions* T_1, T_2, T_1T_2, T_2T_2 *are* ν*-affinely independent and there exists a constant* $\alpha \neq 0$ *such that* $T_2 = \alpha T_1 T_1$ *holds.*

Then the following statements hold:

$$\begin{array}{rcll} 1),\,2),\,5) & \Rightarrow & \text{(P0)} & , \\ 2),\,3),\,4) & \Rightarrow & \text{(P1 t)} & , \\ 2),\,3),\,4),\,6),\,7) & \Rightarrow & \text{(P1)} & , \end{array} \qquad \begin{array}{rcl} 2),\,3),\,4),\,8),\,9) & \Rightarrow & \text{(M1)}, \\ 2),\,3),\,4),\,10) & \Rightarrow & \text{(M2)}. \end{array}$$

Before we prove Theorem 3.13 we give an illustration of its conditions 1)-10):

1) Property (P0)a) and condition 1) are equivalent since $|\log(f_{\boldsymbol{\gamma}})| = |\log(\tilde{p}_{\boldsymbol{\zeta}(\boldsymbol{\gamma})})| + |\log(r)|$ and $|\log(f_{\boldsymbol{\gamma}})| \in L_1(g^0\nu)$ for any $\boldsymbol{\gamma} \in \Gamma$ (what will be seen in the proof of Theorem 3.13).

2) As mentioned above, all properties (P0), (P1), (P1 t), (M1) and (M2) require the existence of partial derivatives of $f_{\boldsymbol{\gamma}}$ with respect to $\boldsymbol{\gamma}$. Since (P0) requires partial derivatives up to the fifth order, condition 2) is introduced which is a sufficient condition to apply Lemma 3.9.

3) Property (P1 t) assumes among other things the linear independence of $\frac{\partial_{\gamma_i} f_{\boldsymbol{\gamma}}|_{\boldsymbol{\gamma}^{0,l}}}{g^0}$, $i = 1, \ldots, k$, $l = 1, \ldots, q$, in $L_2(g^0\nu)$. We will prove that (P1 t) holds iff $\partial_{\boldsymbol{\zeta}}\boldsymbol{\zeta}(\boldsymbol{\gamma})|_{\boldsymbol{\gamma}^{0,l}}$, $l = 1, \ldots, q$, is regular. The argumentation is based on the fact that $T_1, \ldots, T_k$ are ν-affinely independent due to Theorem 3.5 (since $\mathcal{P}$ is minimal k-parametric). This concept is also of interest to ensure (P1), (M1) and (M2) since (P1 t) is a special case of these properties.

4) According to Lemma 3.10 condition 4) leads to $\frac{f_{\boldsymbol{\gamma}}}{g^0} \in L_2(g^0\nu)$ as assumed by (P1) and (P1 t).

5) Evaluating $\partial_{\gamma_{i_1}} \ldots \partial_{\gamma_{i_n}} f_{\boldsymbol{\gamma}}$ (for all $n \le 5$ and all $i_1, \ldots, i_n \in \{1, \ldots, k\}$) according to Lemma 3.9 leads to the sufficient condition 5) for applying property (P0)c). As previously mentioned Proposition 3.11 gives a sufficient condition for condition 5) to hold.

6) Roughly speaking, property (P1) assumes that $\frac{f_{\boldsymbol{\gamma}}}{g^0}$, $\frac{f_{\boldsymbol{\gamma}^{0,l}}}{g^0}$, $\frac{\partial_{\gamma_i} f_{\boldsymbol{\gamma}}|_{\boldsymbol{\gamma}^{0,l}}}{g^0}$, $\frac{\partial_{\gamma_i\gamma_j} f_{\boldsymbol{\gamma}}|_{\boldsymbol{\gamma}^{0,l}}}{g^0}$ are linearly independent in $L_2(g^0\nu)$. We give sufficient conditions for (P1) to hold by condition 6) on $\boldsymbol{T}$ and condition 7) on the parameter function $\boldsymbol{\zeta}$, respectively, since these functions describe the properties of the corresponding exponential family as well as the properties of the corresponding ν-densities and their derivatives. Thereby, condition 6) as well as condition 7) result from evaluating the linear combination of the above-mentioned functions.

7) Condition 7) is very technical. The higher the dimension k of the parameter space Γ, the more difficult it will be to verify condition 7). Simple computations show that a sufficient condition for $c_{i,i}^{0,l} = 0$, $i = 1, \ldots, k$, to hold in (3.27) is given by

$$\left(\sum_{j=1}^{k} \partial_{\gamma_i} \zeta_j(\boldsymbol{\gamma}) \big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,l}} \right)^2 \neq 0. \tag{3.29}$$

This means that the sum of elements in each column of the Jacobian matrix $\partial_{\boldsymbol{\gamma}}\boldsymbol{\zeta}(\boldsymbol{\gamma})|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,l}}$ is not equal to 0. Notice that property 7) only uses the first order partial derivatives of $\boldsymbol{\zeta} : \Gamma \to \boldsymbol{\zeta}(\Gamma)$.

8) Corresponding to item 6) and 7), sufficient conditions for (M1) to hold are given by condition 8) and 9). In condition 8) the set U is fixed since we consider exponential families in *fixed* $\boldsymbol{\zeta} : \Gamma \to I\!R^k$ and *fixed* $\boldsymbol{T} : \mathfrak{X} \to I\!R^k$. This means that U does not depend on any $h \in \{1, \ldots, q\}$ while the sets U_h given in (M1) do. But in condition 9) we make requirements on the derivatives of $\boldsymbol{\zeta} : \Gamma \to \boldsymbol{\zeta}(\Gamma)$ at the points $\boldsymbol{\gamma}^{\mathbf{0},\boldsymbol{h}}$, $h = 1, \ldots, q$.

9) The technical condition 9) is equivalent to condition 7) if $\dim(\Gamma) = k \leq 2$. Simple computations show that a sufficient condition for $c^h_{i,i} = 0$, $i = 1, \dots, k$, to hold in (3.28) is given by

$$\left(\sum_{j=1}^{k} \partial_{\gamma_i}\zeta_j(\boldsymbol{\gamma})\big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,h}}\right)^2 - \sum_{\substack{(u,v)\in U^c \\ \kappa\in H(u,v)}} \Big((2-2\alpha_\kappa)(\partial_{\gamma_i}\zeta_u(\boldsymbol{\gamma})\big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,h}})(\partial_{\gamma_i}\zeta_v(\boldsymbol{\gamma})\big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,h}})\Big) \neq 0. \quad (3.30)$$

If $\alpha_\kappa = 1$ for all $\kappa \in H = \bigcup_{(u,v)\in U^c} H(u,v)$ then (3.30) is equivalent to (3.29).

Some useful sufficient (and necessary) conditions for condition 7) and condition 9) to hold are given below in Lemma 3.14.

10) Finally, the technical condition 10) leads (together with 2), 3) and 4)) to (M2).

Notice that condition 1) may imply a restriction of $\mathfrak{X}$ while condition 4) and 5) hold for appropriate parameter space Γ.

Lemma 3.14 *Let the assumptions of Lemma* 3.9 *as well as* (A1) *be satisfied. Let conditions* 3), 7) *as well as* 9) *of Theorem* 3.13 *be satisfied. Then the following statements hold:*

1) a) *If $k = 1$ then conditions* 3) *and* 7) *are equivalent.*
   b) *If $k \in \{1, 2\}$ then condition* 7) *and condition* 9) *are equivalent since $U^c = \emptyset$.*

2) *Suppose that for $l = 1, \dots, q$ the Jacobian matrix of $\boldsymbol{\zeta} : \Gamma \to \boldsymbol{\zeta}(\Gamma)$ with respect to $\boldsymbol{\gamma}$ at the point $\boldsymbol{\gamma}^{0,l}$ is diagonal. Then*
   a) *conditions* 3) *and* 7) *are equivalent,*
   b) *conditions* 3) *and* 9) *are equivalent.*

Proof of Lemma 3.14:

1) a) For $k = 1$ we have $dim(\Gamma) = dim(Z) = 1$ and for $l \in \{1, \dots, q\}$ it is obvious that

$$\left\{c^{0,l}_{1,1}\left(\partial_{\gamma_1}\zeta_1(\boldsymbol{\gamma})\big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,l}}\right)^2 = 0 \ \Rightarrow\ c^{0,l}_{1,1} = 0\right\} \ \Leftrightarrow\ \det\left(\partial_{\gamma_1}\zeta_1(\boldsymbol{\gamma})\big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,l}}\right) \neq 0$$
$$\Leftrightarrow\ \partial_{\gamma_1}\zeta_1(\boldsymbol{\gamma})\big|_{\boldsymbol{\gamma}} \text{ is regular.}$$

   Additionally, we have $\mathcal{H} = \{T_1, T_1T_1\}$ and $U = U^c = \emptyset$ according to condition 8) of Theorem 3.13.

   b) For $k = 1$ condition 7) and 9) are equivalent since $U = U^c = \emptyset$ (see 1)a)) and $H^c = \{1\}$.
   For $k = 2$ we use the abbreviation $\partial_i\zeta_u = \partial_{\gamma_i}\zeta_u(\boldsymbol{\gamma})|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,l}}$, $i, u = 1, 2$, $l = 1, \dots, q$. Condition 7) is equivalent to

$$\forall l \in \{1, \dots, q\}: \quad \begin{aligned} c^{0,l}_{1,1}\left(\partial_1\zeta_1 + \partial_1\zeta_2\right)^2 = 0 \ &\Rightarrow\ c^{0,l}_{1,1} = 0, \\ c^{0,l}_{1,2}\left(\partial_1\zeta_1 + \partial_1\zeta_2\right)\left(\partial_2\zeta_1 + \partial_2\zeta_2\right) = 0 \ &\Rightarrow\ c^{0,l}_{1,2} = 0, \\ c^{0,l}_{2,2}\left(\partial_2\zeta_1 + \partial_2\zeta_2\right)^2 = 0 \ &\Rightarrow\ c^{0,l}_{2,2} = 0. \end{aligned}$$

Assume that there exist some $\alpha_1, \alpha_2 \neq 0$ such that $T_1T_2 = \alpha_1 T_1 T_1$ or $T_1T_2 = \alpha_2 T_2 T_2$ hold. This contradicts the assumption that T_1 and T_2 are linearly independent in $L_2(g^0\nu)$. Thus we have $\mathcal{H} = \{T_1, T_2, T_1T_1, T_2T_2, T_1T_2\}$, $U = \{(1,2)\}$ and $U^c = \emptyset$ according to condition 8) of Theorem 3.13. As $H^c = \{1,2\}$ condition 9) is equivalent to

$$\begin{aligned}
\forall h \in \{1,\dots,q\}: \quad & c_{1,1}^h\Big\{\Big[2\Big(\partial_1\zeta_1\Big)\Big(\partial_1\zeta_2\Big)\Big] + \Big[\Big(\partial_1\zeta_1\Big)^2 + \Big(\partial_1\zeta_2\Big)^2\Big]\Big\} = 0 \;\Rightarrow\; c_{1,1}^{0,h} = 0,\\
& c_{1,2}^h\Big\{\Big[\Big(\partial_1\zeta_2\Big)\Big(\partial_2\zeta_1\Big) + \Big(\partial_2\zeta_2\Big)\Big(\partial_1\zeta_1\Big)\Big] + \\
& \qquad \Big[\Big(\partial_1\zeta_1\Big)\Big(\partial_2\zeta_1\Big) + \Big(\partial_1\zeta_2\Big)\Big(\partial_2\zeta_2\Big)\Big]\Big\} = 0 \;\Rightarrow\; c_{1,2}^{0,h} = 0,\\
& c_{2,2}^h\Big\{\Big[2\Big(\partial_2\zeta_1\Big)\Big(\partial_2\zeta_2\Big)\Big] + \Big[\Big(\partial_2\zeta_1\Big)^2 + \Big(\partial_2\zeta_2\Big)^2\Big]\Big\} = 0 \;\Rightarrow\; c_{2,2}^{0,h} = 0.
\end{aligned}$$

Obviously, condition 7) and 9) are equivalent for $k = 2$.

2) a) For $l \in \{1,\dots,q\}$ and $1 \le i \le j \le k$ formula (3.27) is equivalent to

$$\Big\{c_{i,j}^{0,l}\big(\partial_{\gamma_i}\zeta_i(\boldsymbol{\gamma})|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,l}}\big)\big(\partial_{\gamma_j}\zeta_j(\boldsymbol{\gamma})|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,l}}\big) = 0\Big\} \;\Rightarrow\; c_{i,j}^{0,l} = 0$$

which is equivalent to

$$\det\Big(\partial_{\boldsymbol{\gamma}}\boldsymbol{\zeta}(\boldsymbol{\gamma})\big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,l}}\Big) = \prod_{j=1}^{k}\big(\partial_{\gamma_j}\zeta_j(\boldsymbol{\gamma})|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,l}}\big) \neq 0. \tag{3.31}$$

b) Let U be given as in condition 8) of Theorem 3.13 and let $I = \{(1,1),\dots,(k,k)\}$. For $h \in \{1,\dots,q\}$ and $(i,j) \in U \cup I$ formula (3.28) is equivalent to

$$\begin{aligned}
c_{i,j}^h\ \alpha_\kappa^h\big(\partial_{\gamma_i}\zeta_i(\boldsymbol{\gamma})|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,h}}\big)\big(\partial_{\gamma_j}\zeta_j(\boldsymbol{\gamma})|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,h}}\big) = 0 \qquad & \text{for } (u,v) = (i,j) \in U^c \text{ and}\\
& H_{(u,v)} = \{\kappa\},\ \alpha_k \neq 0,\\
c_{i,j}^h\ \big(\partial_{\gamma_i}\zeta_i(\boldsymbol{\gamma})|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,h}}\big)\big(\partial_{\gamma_j}\zeta_j(\boldsymbol{\gamma})|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,h}}\big) = 0 \qquad & \text{for } (u,v) = (i,j) \in U \cup I
\end{aligned}$$

$\Rightarrow c_{i,j}^h = 0$
which is obviously equivalent to (3.31). □

Although Lemma 3.14 does not shed much light on the meaning of the technical conditions 7) and 9), it allows us to show that the asymptotic theory from [DCG99] is applicable to m-dimensional Gaussian mixtures with unknown mean and arbitrary known covariance matrix, see Example 3.15.

Proof of Theorem 3.13:

(P0) a) For any $\boldsymbol{\gamma} \in \Gamma$, $\Gamma \subset \mathbb{R}^k$, and any $\boldsymbol{x} \in \mathfrak{X}$ one has

$$\big|\log\big(f_{\boldsymbol{\gamma}}(\boldsymbol{x})\big)\big| \overset{(3.1)}{\le} \Big\{\max_{\boldsymbol{\gamma}\in\Gamma}\big|\log\{\mathcal{C}(\boldsymbol{\zeta}(\boldsymbol{\gamma}))\}\big|\Big\} + \Big\{\max_{\boldsymbol{\gamma}\in\Gamma}\sum_{i=1}^{k}\big|\zeta_i(\boldsymbol{\gamma})T_i(\boldsymbol{x})\big|\Big\} + |\log(r(\boldsymbol{x}))|.$$

Obviously, $\kappa_1 = \max_{\gamma\in\Gamma} \left|\log\{C(\boldsymbol{\zeta}(\boldsymbol{\gamma}))\}\right| < \infty$ since Γ is compact due to (A1) and $\boldsymbol{\zeta} \mapsto C(\boldsymbol{\zeta})$ as well as $\boldsymbol{\gamma} \mapsto \boldsymbol{\zeta}(\boldsymbol{\gamma})$ are continuous according to Proposition 3.6 and condition 2), respectively. Since $\kappa_2 = \max\limits_{\gamma\in\Gamma}\max\limits_{1\le i\le k} |\zeta_i(\boldsymbol{\gamma})|$ is also finite, we define a function $h : \mathfrak{X} \to I\!R$ by

$$h(\boldsymbol{x}) \quad = \quad \kappa_1 + \kappa_2 \sum_{i=1}^{k} \left|T_i(\boldsymbol{x})\right| + |\log(r(\boldsymbol{x}))| \quad , \ \boldsymbol{x} \in \mathfrak{X}.$$

From $g^0(\boldsymbol{x}) \overset{(3.14)}{=} r(\boldsymbol{x}) \sum\limits_{l=1}^{q} \pi_l^0 \tilde{p}_{\boldsymbol{\zeta}(\boldsymbol{\gamma}^{0,\iota})}(\boldsymbol{x})$ it follows that

$$\begin{aligned}&\int \left|h(\boldsymbol{x})\right| g^0(\boldsymbol{x})\, \nu(d\boldsymbol{x})\\ =\quad & \kappa_1 + \kappa_2 \sum_{l=1}^{q} \left(\pi_l^0 \sum_{i=1}^{k} \int \left|T_i(\boldsymbol{x})\right| \tilde{p}_{\boldsymbol{\zeta}(\boldsymbol{\gamma}^{0,\iota})}(\boldsymbol{x})\, \mu(d\boldsymbol{x})\right) + \int |\log(r(\boldsymbol{x}))| g^0(\boldsymbol{x})\, \nu(d\boldsymbol{x}).\end{aligned}$$

According to Proposition 3.6 the generating statistic $\boldsymbol{T} = (T_1, \ldots, T_k)'$ possesses mixed moments of arbitrary order with respect to $\tilde{p}_{\boldsymbol{\zeta}(\boldsymbol{\gamma}^{0,\iota})}\mu$, $l = 1, \ldots, q$, and thus we obtain $h \in L_1(g^0\nu)$ due to condition 1).

b) From condition 2), Lemma 3.9 and $g^0(\boldsymbol{x}) \overset{(3.14)}{=} r(\boldsymbol{x})\tilde{w}^0(\boldsymbol{x})$ it follows for $l \in \{1, \ldots, q\}$ and for all $n \le 5$ and all $i_1, \ldots, i_n \in \{1, \ldots, k\}$ that

$$\begin{aligned}&\int \left|\frac{\partial_{\gamma_{i_1}} \ldots \partial_{\gamma_{i_n}} f_{\boldsymbol{\gamma}}(\boldsymbol{x})\big|_{\boldsymbol{\gamma}^{0,\iota}}}{g^0(\boldsymbol{x})}\right|^3 g^0(\boldsymbol{x})\, \nu(d\boldsymbol{x})\\ \le\quad & \kappa^3 \sum_{m=1}^{n} \sum_{j_1=1}^{k} \cdots \sum_{j_m=1}^{k} \int \left|\frac{r(\boldsymbol{x})\, \partial_{\zeta_{j_1}(\boldsymbol{\gamma})} \ldots \partial_{\zeta_{j_m}(\boldsymbol{\gamma})} \tilde{p}_{\boldsymbol{\zeta}(\boldsymbol{\gamma})}(\boldsymbol{x})\big|_{\boldsymbol{\gamma}^{0,\iota}}}{r(\boldsymbol{x})\tilde{w}^0(\boldsymbol{x})}\right|^3 \tilde{w}^0(\boldsymbol{x})\, \mu(d\boldsymbol{x}) \qquad (3.32)\end{aligned}$$

holds, where $A^n = \{1, \ldots, n\}$ and

$$\kappa \quad = \quad \max_{m\in\{1,\ldots,n\}} \sum_{\substack{A_1^n,\ldots,A_m^n \neq \emptyset\\ \text{disjoint}\\ A_1^n \cup\ldots\cup A_m^n = A^n}} \left|b_{(A_1^n,\ldots,A_m^n)} \prod_{i=1}^{m} \partial_{A_i^n} \zeta_{j_i}(\boldsymbol{\gamma})\big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,\iota}}\right| \quad < \quad \infty \qquad (3.33)$$

using the compactness assumption on Γ according to (A1) as well as the continuity of the partial derivatives of $\boldsymbol{\zeta} : \Gamma \to Z$ due to assumption 2). Remark 3.8 and $\pi_l^0 > 0$ (due to (A1)) imply that the left hand side of formula (3.32) is less or equal to

$$\begin{aligned}&\kappa^3 \sum_{m=1}^{n} \sum_{j_1=1}^{k} \cdots \sum_{j_m=1}^{k} \int \left|P_{\zeta_{j_1}\ldots\zeta_{j_m}}^{\boldsymbol{\zeta}(\boldsymbol{\gamma}^{0,\iota})}(\boldsymbol{T}(\boldsymbol{x}))\right|^3 \left|\frac{\tilde{p}_{\boldsymbol{\zeta}(\boldsymbol{\gamma}^{0,\iota})}(\boldsymbol{x})}{\pi_l^0 \tilde{p}_{\boldsymbol{\zeta}(\boldsymbol{\gamma}^{0,\iota})}(\boldsymbol{x})}\right|^3 \tilde{w}^0(\boldsymbol{x})\, \mu(d\boldsymbol{x})\\ =\quad & \kappa^3 \sum_{i=1}^{q} \left(\pi_i^0 \sum_{m=1}^{n} \sum_{j_1=1}^{k} \cdots \sum_{j_m=1}^{k} |\pi_l^0|^{-3} \int \left|P_{\zeta_{j_1}\ldots\zeta_{j_m}}^{\boldsymbol{\zeta}(\boldsymbol{\gamma}^{0,\iota})}(\boldsymbol{T}(\boldsymbol{x}))\right|^3 \tilde{p}_{\boldsymbol{\zeta}(\boldsymbol{\gamma}^{0,i})}(\boldsymbol{x})\, \mu(d\boldsymbol{x})\right).\end{aligned}$$

According to Remark 3.8 $P_{\zeta_{j_1}\ldots\zeta_{j_m}}^{\boldsymbol{\zeta}(\boldsymbol{\gamma}^{0,\iota})}(\boldsymbol{T}(\boldsymbol{x}))$ is a polynomial in the components of $\boldsymbol{T}(\boldsymbol{x})$ and since $\boldsymbol{T} = (T_1, \ldots, T_k)'$ possesses mixed moments with respect to $\tilde{p}_{\boldsymbol{\zeta}(\boldsymbol{\gamma}^{0,i})}\mu$ of arbitrary order thanks to Proposition 3.6 it follows directly (P0)b).

c) According to (A1) all true component parameters $\boldsymbol{\gamma}^{\mathbf{0},\boldsymbol{l}}$, $l = 1,\ldots,q$, are distinct accumulation points of Γ. Hence, there exists some $\epsilon > 0$ such that for all $l = 1,\ldots,q$ one has $\{\boldsymbol{\gamma}^{\mathbf{0},\boldsymbol{l}}\} \neq \bar{U}_\epsilon(\boldsymbol{\gamma}^{\mathbf{0},\boldsymbol{l}}) \subset \Gamma$. Due to assumption 2) we define κ_l according to (3.33) and we obtain $\kappa = \max_{l=1,\ldots,q} \kappa_l < \infty$. Thus it follows from Lemma 3.9 that

$$\begin{aligned}
&\sup_{\boldsymbol{\gamma}\in\overline{U}_\varepsilon(\boldsymbol{\gamma}^{0,l})} \left| \frac{\partial_{\gamma_{i_1}} \ldots \partial_{\gamma_{i_n}} f_{\boldsymbol{\gamma}}(\boldsymbol{x})}{g^0(\boldsymbol{x})} \right| \\
\leq\quad & \kappa \sum_{m=1}^{n} \sum_{j_1=1}^{k} \cdots \sum_{j_m=1}^{k} \sup_{\boldsymbol{\gamma}\in\overline{U}_\varepsilon(\boldsymbol{\gamma}^{0,l})} \left| \frac{r(\boldsymbol{x})\, \partial_{\zeta_{j_1}(\boldsymbol{\gamma})} \ldots \partial_{\zeta_{j_m}(\boldsymbol{\gamma})} \tilde{p}_{\boldsymbol{\zeta}(\boldsymbol{\gamma})}(\boldsymbol{x})}{g^0(\boldsymbol{x})} \right| \\
\overset{\text{Rem.3.8}}{=}\quad & \kappa \sum_{m=1}^{n} \sum_{j_1=1}^{k} \cdots \sum_{j_m=1}^{k} \sup_{\boldsymbol{\gamma}\in\overline{U}_\varepsilon(\boldsymbol{\gamma}^{0,l})} \left| \frac{f_{\boldsymbol{\gamma}}(\boldsymbol{x})}{g^0(\boldsymbol{x})} \right| \left| P^{\boldsymbol{\zeta}(\boldsymbol{\gamma})}_{\zeta_{j_1}\ldots\zeta_{j_n}}(\boldsymbol{T}(\boldsymbol{x})) \right|
\end{aligned}$$

for all $n \leq 5$ and and all $i_1,\ldots,i_n \in \{1,\ldots,k\}$. Finally, assumption 5) leads to property (P0)c).

(P1 t) Let $\boldsymbol{\gamma}^{\boldsymbol{i}} \in \Gamma \backslash \{\boldsymbol{\gamma}^{\mathbf{0},\mathbf{1}},\ldots,\boldsymbol{\gamma}^{\mathbf{0},\boldsymbol{q}}\}$ for $i \in \{1,\ldots,p-q\}$ be a set of distinct points. Assumption 2) and 4) (together with Proposition 3.6, Remark 3.8 and Lemma 3.10) imply that the functions

$$\frac{f_{\boldsymbol{\gamma}^i}}{g^0},\ \frac{f_{\boldsymbol{\gamma}^{0,l}}}{g^0},\ \frac{\partial_{\gamma_1} f_{\boldsymbol{\gamma}}\big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,l}}}{g^0},\ldots,\frac{\partial_{\gamma_k} f_{\boldsymbol{\gamma}}\big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,l}}}{g^0}\ , l = 1,\ldots,q,\ i \in \{1,\ldots,p-q\}$$

belong to $L_2(g^0\nu)$. We will verify that

$$\begin{aligned}
&\sum_{l=1}^{p-q} a_l f_{\boldsymbol{\gamma}^l}(\boldsymbol{x}) + \sum_{l=1}^{q} b_l^0 f_{\boldsymbol{\gamma}^{0,l}}(\boldsymbol{x}) + \sum_{l=1}^{q}\sum_{i=1}^{k} c_i^{0,l} \partial_{\gamma_i} f_{\boldsymbol{\gamma}}(\boldsymbol{x})\big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,l}} = 0 \quad \nu\text{–a.s.} \qquad (3.34)\\
\Leftrightarrow\quad & a_1 = \ldots = a_{p-q} = b_l^0 = c_i^{0,l} = 0 \qquad \text{for all } l = 1,\ldots,q \text{ and all } i = 1,\ldots,k.
\end{aligned}$$

Due to Lemma 3.9 the sum of the first order derivatives on the left hand side of (3.34) is equal to

$$\begin{aligned}
& r(\boldsymbol{x}) \sum_{l=1}^{q}\sum_{i=1}^{k} \left\{ c_i^{0,l} \sum_{j=1}^{k} \left(\partial_{\zeta_j(\boldsymbol{\gamma})} \tilde{p}_{\boldsymbol{\zeta}(\boldsymbol{\gamma})}(\boldsymbol{x})\big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,l}} \right) \left(\partial_{\gamma_i} \zeta_j(\boldsymbol{\gamma})\big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,l}} \right) \right\} \\
\overset{(3.15)}{=}\ & r(\boldsymbol{x}) \sum_{l=1}^{q}\sum_{j=1}^{k} \left\{ \left(T_j(\boldsymbol{x}) - E_{\boldsymbol{\zeta}^{0,l}} T_j \right) \tilde{p}_{\boldsymbol{\zeta}^{0,l}}(\boldsymbol{x}) \sum_{i=1}^{k} c_i^{0,l} \left(\partial_{\gamma_i} \zeta_j(\boldsymbol{\gamma})\big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,l}} \right) \right\}. \qquad (3.35)
\end{aligned}$$

Thus (3.34) is equivalent to

$$\begin{aligned}
& \sum_{l=1}^{p-q} P_{A^l,B_l}(\boldsymbol{T}(\boldsymbol{x})) \exp\Big(\langle \boldsymbol{\zeta}(\boldsymbol{\gamma}^{\boldsymbol{l}}), \boldsymbol{T}(\boldsymbol{x}) \rangle \Big) + \sum_{l=1}^{q} P_{A^{0,l},B_{0,l}}(\boldsymbol{T}(\boldsymbol{x})) \exp\Big(\langle \boldsymbol{\zeta}(\boldsymbol{\gamma}^{\mathbf{0},\boldsymbol{l}}), \boldsymbol{T}(\boldsymbol{x}) \rangle \Big) \\
=\ & 0 \quad \nu - a.s. \qquad (3.36)
\end{aligned}$$

for suitable $A^l = (A_1^l,\ldots,A_k^l)' \in \mathbb{R}^k$, $B_l \in \mathbb{R}$ for $l = 1,\ldots,p-q$ and suitable $A^{0,l} = (A_1^{0,l},\ldots,A_k^{0,l})' \in \mathbb{R}^k$, $B_{0,l} \in \mathbb{R}$ for $l = 1,\ldots,q$, and polynomials in the components of $\boldsymbol{T}(\boldsymbol{x})$ given by

$$P_{A^l,B_l}(\boldsymbol{T}(\boldsymbol{x})) \quad = \quad \sum_{i=1}^{k} A_i^l T_i(\boldsymbol{x}) + B_l.$$

Notice that $\boldsymbol{\zeta}(\boldsymbol{\gamma}^{1}),\dots,\boldsymbol{\zeta}(\boldsymbol{\gamma}^{p-q}),\ \boldsymbol{\zeta}(\boldsymbol{\gamma}^{0,1}),\dots,\boldsymbol{\zeta}(\boldsymbol{\gamma}^{0,q})$ are distinct points since $\boldsymbol{\gamma}^{1},\dots,\boldsymbol{\gamma}^{p-q}$, $\boldsymbol{\gamma}^{0,1},\dots,\boldsymbol{\gamma}^{0,q}$ are assumed to be distinct and $\boldsymbol{\zeta}:\Gamma\to\boldsymbol{\zeta}(\Gamma)$ is bijective. Additionally, $T_1,\dots,T_k$ are ν-affinely independent (see Proposition 3.5) and thus we obtain similar to Lemma A.3 that (3.36) and (3.34) are equivalent to

$$\begin{aligned} \forall l\in\{1,\dots,p-q\}: \quad & P_{A^l,B_l}(\boldsymbol{T}(\boldsymbol{x})) = 0 \quad \nu-a.s. \quad \text{and} \\ \forall l\in\{1,\dots,q\}: \quad & P_{A^{0,l},B_{0,l}}(\boldsymbol{T}(\boldsymbol{x})) = 0 \quad \nu-a.s. \end{aligned}$$

Using again the ν-affine independence of $T_1,\dots,T_k$ (see Definition 3.4) it follows

$$\begin{aligned} \forall l\in\{1,\dots,p-q\}: \quad & A^{l}=\mathbf{0},\ B_l=0 \quad \text{and} \\ \forall l\in\{1,\dots,q\}: \quad & A^{0,l}=\mathbf{0},\ B_{0,l}=0. \end{aligned}$$

Notice that the first two multiple sums on the left hand side of (3.34) do not depend on the components of $\boldsymbol{T}(\boldsymbol{x})$. Hence, we verify separately if the first two multiple sums of (3.34) are null then $a_1=\dots=a_{p-q}=b_1^0=\dots=b^q=0$ and if the right hand side of (3.35) is null then $c_i^{0,l}=0$ for all $l=1,\dots,q$ and all $i=1,\dots,k$.

According to Remark 2.1 (due to (ID)) the functions $f_{\boldsymbol{\gamma}^{0,l}}$, $l=1,\dots,q$ are ν-a.e. linearly independent over the field of real numbers. As $f_{\boldsymbol{\gamma}^{0,l}}(\boldsymbol{x})=r(\boldsymbol{x})\tilde{p}_{\boldsymbol{\zeta}^{0,l}}(\boldsymbol{x})$ holds, formula (3.35) is equal to 0 ν-a.s. iff

$$\forall l=1,\dots,q: \quad \sum_{j=1}^{k}\left\{T_j(\boldsymbol{x})f_{\boldsymbol{\gamma}^{0,l}}(\boldsymbol{x})\sum_{i=1}^{k}c_i^{0,l}\left(\partial_{\gamma_i}\zeta_j(\boldsymbol{\gamma})\big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,l}}\right)\right\} = 0 \quad \nu-a.s. \tag{3.37}$$

hold. Since the partial derivatives of $\boldsymbol{\zeta}(\boldsymbol{\gamma}):\Gamma\to\boldsymbol{\zeta}(\Gamma)$ do not depend on $\boldsymbol{x}$ and the functions $T_1,\dots,T_k$ are ν-affinely independent (see Proposition 3.5), formula (3.37) implies that

$$\forall l=1,\dots,q: \quad \sum_{i=1}^{k}c_i^{0,l}\left(\partial_{\gamma_i}\zeta_j(\boldsymbol{\gamma})\big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,l}}\right)=0 \quad \text{for all } j=1,\dots,k$$

which is equivalent to

$$\forall l=1,\dots,q: \quad \sum_{i=1}^{k}c_i^{0,l}\ \partial_{\gamma_i}\boldsymbol{\zeta}(\boldsymbol{\gamma})\big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,l}}=\mathbf{0}. \tag{3.38}$$

Thereby, for any $l=1,\dots,q$, (3.38) is a linear combination of all column vectors of the Jacobian matrix $\partial_{\boldsymbol{\gamma}}\boldsymbol{\zeta}(\boldsymbol{\gamma})|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,l}}$ which is regular according to assumption 3). Thus we obtain

$$\sum_{l=1}^{q}\sum_{i=1}^{k}c_i^{0,l}\partial_{\gamma_i}f_{\boldsymbol{\gamma}}(\boldsymbol{x})\big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,l}}=0 \quad \nu-\text{a.s.} \quad \Leftrightarrow \quad c_i^{0,l} = 0 \ \text{ for all } l=1,\dots,q \text{ and } i=1,\dots,k.$$

The further use of Remark 2.1 (due to (ID)) implies that $f_{\boldsymbol{\gamma}^{i}}, f_{\boldsymbol{\gamma}^{0,l}}$, $i=1,\dots,p-q$, $l=1,\dots,q$, are ν-a.s. linearly independent over the field of real numbers.

Thus it follows $a_1=\dots=a_{p-q}=b_l^0=c_i^{0,l}=0$ for all $l=1,\dots,q$ and all $i=1,\dots,k$.

(P1) Let $p_1 \le p - q$ and let $\boldsymbol{\gamma}^{\boldsymbol{l}} \in \Gamma \backslash \{\boldsymbol{\gamma}^{\mathbf{0,1}}, \dots, \boldsymbol{\gamma}^{\mathbf{0,q}}\}$ for $l \in \{1, \dots, p_1\}$ be a set of distinct points. Assumption 2) and 4) (together with Proposition 3.6, Remark 3.8 and Lemma 3.10) imply that the functions

$$\frac{f_{\boldsymbol{\gamma}^{i'}}}{g^0}, \; \frac{f_{\boldsymbol{\gamma}^{0,l}}}{g^0}, \; \frac{\partial_{\gamma_j} f_{\boldsymbol{\gamma}}\big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,l}}}{g^0}, \; \frac{\partial_{\gamma_i\gamma_j} f_{\boldsymbol{\gamma}}\big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,l'}}}{g^0},$$
$$i' \in \{1, \dots, p_1\}, \; l' \in \{\sigma(1), \dots, \sigma(p_2)\}, \; l = 1, \dots, q, \; i, j = 1, \dots, k \text{ with } i \le j$$

belong to $L_2(g^0\nu)$. We will show that

$$\begin{aligned} &\sum_{l=1}^{p_1} a_l f_{\boldsymbol{\gamma}^l}(\boldsymbol{x}) + \sum_{l=1}^{q} b_l^0 f_{\boldsymbol{\gamma}^{0,l}}(\boldsymbol{x}) + \sum_{l=1}^{q}\sum_{i=1}^{k} c_i^{0,l} \partial_{\gamma_i} f_{\boldsymbol{\gamma}}(\boldsymbol{x})\big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,l}} + \sum_{l=1}^{q}\sum_{i=1}^{k}\sum_{j=i}^{k} c_{i,j}^{0,l}\, \partial_{\gamma_i\gamma_j} f_{\boldsymbol{\gamma}}(\boldsymbol{x})\big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,l}} \\ &= \quad 0 \qquad \nu - a.s. \qquad (3.39) \\ &\Leftrightarrow \quad a_1 = \ldots = a_{p_1} = b_l^0 = c_i^{0,l} = c_{i,j}^{0,l} = 0 \qquad \text{for all } l = 1, \dots, q \text{ and } 1 \le i \le j \le k. \end{aligned}$$

Due to Lemma 3.9 the sum of first order partial derivatives on the left hand side of (3.39) is given by (3.35) and the sum of the second order partial derivatives on the left hand side of (3.39) is equal to

$$\begin{aligned} \sum_{l=1}^{q}\sum_{i=1}^{k}\sum_{j=i}^{k} \Bigg\{ c_{i,j}^{0,l} \Bigg[&\sum_{u=1}^{k}\sum_{v=1}^{k} \Big(\partial_{\zeta_u(\boldsymbol{\gamma})\zeta_v(\boldsymbol{\gamma})} \tilde{p}_{\boldsymbol{\zeta}(\boldsymbol{\gamma})}(\boldsymbol{x})\big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,l}}\Big) \Big(\partial_{\gamma_i} \zeta_v(\boldsymbol{\gamma})\big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,l}}\Big) \Big(\partial_{\gamma_j} \zeta_u(\boldsymbol{\gamma})\big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,l}}\Big) \\ &+ \sum_{v=1}^{k} \Big(\partial_{\zeta_v(\boldsymbol{\gamma})} \tilde{p}_{\boldsymbol{\zeta}(\boldsymbol{\gamma})}(\boldsymbol{x})\big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,l}}\Big) \Big(\partial_{\gamma_i\gamma_j} \zeta_v(\boldsymbol{\gamma})\big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,l}}\Big)\Bigg]\Bigg\} r(\boldsymbol{x}). \qquad (3.40) \end{aligned}$$

According to Remark 3.8 the partial derivatives of $\tilde{p}_{\boldsymbol{\zeta}(\boldsymbol{\gamma})}$ with respect to $\boldsymbol{\zeta}(\boldsymbol{\gamma})$ are given by the product of a corresponding polynomial in the components of $\boldsymbol{T}(\boldsymbol{x})$ and $\tilde{p}_{\boldsymbol{\zeta}(\boldsymbol{\gamma})}(\boldsymbol{x}) = r(\boldsymbol{x}) f_{\boldsymbol{\gamma}}(\boldsymbol{x})$ itself. Further it follows from Remark 3.8 that polynomials in $T_u(\boldsymbol{x})T_v(\boldsymbol{x})$ appear only when evaluating the corresponding terms of the second order partial derivatives $\partial_{\zeta_u(\boldsymbol{\gamma})\zeta_v(\boldsymbol{\gamma})} \tilde{p}_{\boldsymbol{\zeta}(\boldsymbol{\gamma})}(\boldsymbol{x})\big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,l}}$. Thus (3.39) is equivalent to

$$\begin{aligned} &\sum_{l=1}^{p_1} P_{\mathcal{A}^l, B^l, C_l}(\boldsymbol{T}(\boldsymbol{x})) \exp\Big(\big\langle \boldsymbol{\zeta}(\boldsymbol{\gamma}^{\boldsymbol{l}}), \boldsymbol{T}(\boldsymbol{x})\big\rangle\Big) + \sum_{l=1}^{q} P_{\mathcal{A}^{0,l}, B^{0,l}, C_{0,l}}(\boldsymbol{T}(\boldsymbol{x})) \exp\Big(\big\langle \boldsymbol{\zeta}(\boldsymbol{\gamma}^{\mathbf{0,l}}), \boldsymbol{T}(\boldsymbol{x})\big\rangle\Big) \\ &= \; 0 \quad \nu - a.s. \qquad (3.41) \end{aligned}$$

for suitable $\mathcal{A}^l = (\mathcal{A}_{i,j}^l : 1 \le i \le j \le k)' \in I\!R^{\frac{k(k+1)}{2}}$, $B^l = (B_1^l, \dots, B_k^l)' \in I\!R^k$, $C_l \in I\!R$ for $l = 1, \dots, p_1$ and for suitable $\mathcal{A}^{0,l} = (\mathcal{A}_{i,j}^{0,l} : 1 \le i \le j \le k)' \in I\!R^{\frac{k(k+1)}{2}}$, $B^{0,l} = (B_1^{0,l}, \dots, B_k^{0,l})' \in I\!R^k$, $C_{0,l} \in I\!R$ for $l = 1, \dots, q$, and polynomials in the components of $\boldsymbol{T}(\boldsymbol{x})$ given by

$$P_{\mathcal{A}^l, B^l, C_l}(\boldsymbol{T}(\boldsymbol{x})) \;=\; \sum_{i=1}^{k}\sum_{j=i}^{k} \mathcal{A}_{i,j}^l T_i(\boldsymbol{x}) T_j(\boldsymbol{x}) + \sum_{i=1}^{k} B_i^l T_i(\boldsymbol{x}) + C_l.$$

Notice that $\boldsymbol{\zeta}(\boldsymbol{\gamma}^{\mathbf{1}}), \dots, \boldsymbol{\zeta}(\boldsymbol{\gamma}^{\boldsymbol{p_1}})$, $\boldsymbol{\zeta}(\boldsymbol{\gamma}^{\mathbf{0,1}}), \dots, \boldsymbol{\zeta}(\boldsymbol{\gamma}^{\mathbf{0,q}})$ are distinct points since $\boldsymbol{\gamma}^{\mathbf{1}}, \dots, \boldsymbol{\gamma}^{\boldsymbol{p_1}}$, $\boldsymbol{\gamma}^{\mathbf{0,1}}, \dots, \boldsymbol{\gamma}^{\mathbf{0,q}}$ are assumed to be distinct and $\boldsymbol{\zeta} : \Gamma \to \boldsymbol{\zeta}(\Gamma)$ is bijective. Additionally,

$T_1, \dots, T_k, T_iT_j$, $1 \le i \le j \le k$, are ν-affinely independent according to assumption 6) and thus we obtain due Lemma A.3 that (3.41) is equivalent to

$$\begin{array}{rrcll} \forall l \in \{1, \dots, p_1\}: & P_{\mathcal{A}^l, B^l, C_l}(\boldsymbol{T}(\boldsymbol{x})) & = & 0 \quad \nu - a.s. & \quad \text{and} \\ \forall l \in \{1, \dots, q\}: & P_{\mathcal{A}^{0,l}, B^{0,l}, C_{0,l}}(\boldsymbol{T}(\boldsymbol{x})) & = & 0 \quad \nu - a.s. & \end{array}$$

Using again the ν-affine independence of $T_1, \dots, T_k, T_iT_j$, $1 \le i \le j \le k$, (see Definition 3.4) it follows

$$\begin{array}{rll} \forall l \in \{1, \dots, p_1\}: & \mathcal{A}^l = \mathbf{0},\ B^l = \mathbf{0},\ C_l = 0 & \quad \text{and} \\ \forall l \in \{1, \dots, q\}: & \mathcal{A}^{0,l} = \mathbf{0},\ B^{0,l} = \mathbf{0},\ C_{0,l} = 0. & \end{array}$$

Hence, we verify separately if the first three multiple sums on the left hand side of (3.39) are equal to null then $a_1 = \ldots = a_{p_1} = b_l^0 = c_i^{0,l} = 0$ for all $l = 1, \dots, q$, $i = 1, \dots, k$, and if the last multiple sum on the left hand side of (3.39) is equal to null then $c_{i,j}^{0,l} = 0$ for all $l = 1, \dots, q, 1 \le i \le j \le k$.

As aforesaid polynomials in $T_u(\boldsymbol{x})T_v(\boldsymbol{x})$ appear on the left hand side of (3.39) only when evaluating the corresponding terms in (3.40) of the second order partial derivatives $\partial_{\zeta_u(\gamma)\zeta_v(\gamma)}\tilde{p}_{\boldsymbol{\zeta}(\boldsymbol{\gamma})}(\boldsymbol{x})\big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,l}}$. Thus we show first of all, that for any $l \in \{1, \dots, q\}$ the following statement holds:

$$\sum_{i=1}^{k}\sum_{j=i}^{k}\left\{c_{i,j}^{0,l}\left[\sum_{u=1}^{k}\sum_{v=1}^{k}\Big(T_u(\boldsymbol{x})T_v(\boldsymbol{x})\Big)\Big(\partial_{\gamma_i}\zeta_v(\boldsymbol{\gamma})\big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,l}}\Big)\Big(\partial_{\gamma_j}\zeta_u(\boldsymbol{\gamma})\big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,l}}\Big)\right]f_{\boldsymbol{\gamma}^{0,l}}(\boldsymbol{x})\right\} = 0\ \nu\text{–}a.s.$$

$$\Rightarrow \quad c_{i,j}^{0,l} = 0 \text{ for all } i, j \in \{1, \dots k\} \text{ with } i \le j. \tag{3.42}$$

Since $f_{\boldsymbol{\gamma}^{0,l}}(\boldsymbol{x}) \neq 0$ occurs in all terms we study (3.42) without the factor $f_{\boldsymbol{\gamma}^{0,l}}(\boldsymbol{x})$. For $u, v \in \{1, \dots, k\}$ let

$$\tau_{u,v} = \begin{cases} 1 & , \quad \text{for } u \neq v \\ \frac{1}{2} & , \quad \text{for } u = v. \end{cases}$$

Then the equation on left hand side of the implication (3.42) (without the factor $f_{\boldsymbol{\gamma}^{0,l}}(\boldsymbol{x})$) is ν-a.s. is equivalent to

$$\sum_{i=1}^{k}\sum_{j=i}^{k}\left\{c_{i,j}^{0,l}\sum_{u=1}^{k}\sum_{v\le u}\left\{\Big(\tau_{u,v}T_u(\boldsymbol{x})T_v(\boldsymbol{x})\Big)\left[\Big(\partial_{\gamma_i}\zeta_v(\boldsymbol{\gamma})\big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,l}}\Big)\Big(\partial_{\gamma_j}\zeta_u(\boldsymbol{\gamma})\big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,l}}\Big)\right.\right.\right.$$
$$\left.\left.\left. + \Big(\partial_{\gamma_j}\zeta_v(\boldsymbol{\gamma})\big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,l}}\Big)\Big(\partial_{\gamma_i}\zeta_u(\boldsymbol{\gamma})\big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,l}}\Big)\right]\right\}\right\} = 0\ \nu - a.s.$$

Assumption 6) and 7) lead to $c_{i,j}^{0,l} = 0$ for all $i, j \in \{1, \dots k\}$ with $i \le j$.

The remaining polynomials in the components of $\boldsymbol{T}(\boldsymbol{x})$ of (3.39) are given by $T_j(\boldsymbol{x})$, $j = 1, \dots, k$ and appear in formula (3.35) when evaluating the corresponding terms of the first order partial derivatives $\partial_{\zeta_j(\gamma)}\tilde{p}_{\boldsymbol{\zeta}(\boldsymbol{\gamma})}(\boldsymbol{x})\big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,l}}$. The rest of the proof follows the same lines as the proof of (P1 t) and uses condition 3).

(M1) Assumption 2) and 4) (together with Proposition 3.6, Remark 3.8 and Lemma 3.10) imply that the functions

$$\frac{f_{\boldsymbol{\gamma}^{0,l}}}{g^0},\ \frac{\partial_{\gamma_i} f_{\boldsymbol{\gamma}}\big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,l}}}{g^0},\ \frac{\partial_{\gamma_i\gamma_i} f_{\boldsymbol{\gamma}}\big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,h}}}{g^0},\ \frac{\partial_{\gamma_{i'}\gamma_{j'}} f_{\boldsymbol{\gamma}}\big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,h}}}{g^0}$$
$$l=1,\dots,q,\ i=1,\dots,k,\ (i',j')\in U \text{ for } h\in\{1,\dots,q\}$$

belong to $L_2(g^0\nu)$. We will show that

$$\begin{aligned}\sum_{l=1}^q b_l f_{\boldsymbol{\gamma}^{0,l}}(\boldsymbol{x}) + \sum_{l=1}^q\sum_{i=1}^k c_i^l \partial_{\gamma_i} f_{\boldsymbol{\gamma}}(\boldsymbol{x})\big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,l}} + \sum_{i=1}^k c_{i,i}^h\, \partial_{\gamma_i\gamma_i} f_{\boldsymbol{\gamma}}(\boldsymbol{x})\big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,h}} +&\\ \sum_{(i,j)\in U} c_{i,j}^h\, \partial_{\gamma_i\gamma_j} f_{\boldsymbol{\gamma}}(\boldsymbol{x})\big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,h}} &= 0\ \ \nu-a.s. \quad (3.43)\\ \Leftrightarrow\quad b_l = c_i^l = c_{i,i}^h = c_{i,j}^h = 0 \qquad \text{for all } l=1,\dots,q,\ i=1,\dots,k \text{ and } (i,j)\in U.&\end{aligned}$$

We follow the idea of the proof of (P1) and consider first of all those expressions in which polynomials in $T_u(\boldsymbol{x})T_v(\boldsymbol{x})$ appear. More precisely, corresponding to the remarks after (3.42) we omit the factor $f_{\boldsymbol{\gamma}^{0,h}} = r(\boldsymbol{x})\tilde{p}_{\boldsymbol{\zeta}(\boldsymbol{\gamma}^{0,h})}(\boldsymbol{x})$ and verify that

$$\begin{aligned}\sum_{u=1}^k\sum_{v\le u}\Big\{\Big(\tau_{u,v}T_u(\boldsymbol{x})T_v(\boldsymbol{x})\Big)\sum_{(i,j)\in U\cup I} c_{i,j}^h\Big[\Big(\partial_{\gamma_i}\zeta_v(\boldsymbol{\gamma})\big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,h}}\Big)\Big(\partial_{\gamma_j}\zeta_u(\boldsymbol{\gamma})\big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,h}}\Big) +&\\ \Big(\partial_{\gamma_j}\zeta_v(\boldsymbol{\gamma})\big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,h}}\Big)\Big(\partial_{\gamma_i}\zeta_u(\boldsymbol{\gamma})\big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,h}}\Big)\Big]\Big\} &= 0\ \ \nu-a.s.\\ \Rightarrow\quad c_{i,j}^h = 0 \qquad \text{for all } (i,j)\in U\cup I,\ \text{where } I=\{(1,1),(2,2),\dots,(k,k)\}.& \quad (3.44)\end{aligned}$$

The equation on the left hand side of the implication (3.44) is equivalent to

$$\begin{aligned}\sum_{(i,j)\in U\cup I}\Big\{c_{i,j}^h\Big\{&\sum_{(u,v)\in U}\Big(T_u(\boldsymbol{x})T_v(\boldsymbol{x})\Big)\Big[\Big(\partial_{\gamma_i}\zeta_v(\boldsymbol{\gamma})\big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,h}}\Big)\Big(\partial_{\gamma_j}\zeta_u(\boldsymbol{\gamma})\big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,h}}\Big)\\ &\qquad+\Big(\partial_{\gamma_j}\zeta_v(\boldsymbol{\gamma})\big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,h}}\Big)\Big(\partial_{\gamma_i}\zeta_u(\boldsymbol{\gamma})\big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,h}}\Big)\Big]\\ &+\sum_{u\in H^c}\Big(T_u(\boldsymbol{x})T_u(\boldsymbol{x})\Big)\Big[\Big(\partial_{\gamma_i}\zeta_u(\boldsymbol{\gamma})\big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,h}}\Big)\Big(\partial_{\gamma_j}\zeta_u(\boldsymbol{\gamma})\big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,h}}\Big)\Big]\\ &+\sum_{\substack{(u,v)\in U^c\\ \kappa\in H(u,v)}}\Big(T_\kappa(\boldsymbol{x})T_\kappa(\boldsymbol{x})\Big)\Big[\ \alpha_\kappa\Big(\partial_{\gamma_i}\zeta_v(\boldsymbol{\gamma})\big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,h}}\Big)\Big(\partial_{\gamma_j}\zeta_u(\boldsymbol{\gamma})\big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,h}}\Big)\\ &\qquad+\alpha_\kappa\Big(\partial_{\gamma_j}\zeta_v(\boldsymbol{\gamma})\big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,h}}\Big)\Big(\partial_{\gamma_i}\zeta_u(\boldsymbol{\gamma})\big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,h}}\Big)\\ &\qquad+\ \Big(\partial_{\gamma_j}\zeta_\kappa(\boldsymbol{\gamma})\big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,h}}\Big)\Big(\partial_{\gamma_i}\zeta_\kappa(\boldsymbol{\gamma})\big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,h}}\Big)\Big]\Big\}\Big\}\\ =\quad 0\ \ \nu-a.s.,&\end{aligned}$$

where H^c and $H(u,v)$ are defined in 9). Assumption 8) and 9) lead to $c_{i,j}^{0,h} = 0$ for all $(i,j)\in U\cup I$.

The remaining polynomials in the components of $\boldsymbol{T}(\boldsymbol{x})$ of (3.43) belong to the first order partial derivatives $\partial_{\zeta_j(\boldsymbol{\gamma})}\tilde{p}_{\boldsymbol{\zeta}(\boldsymbol{\gamma})}(\boldsymbol{x})\big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,h}}$. The rest of the proof follows the same lines as the proof of (P1 t) and uses condition 3).

(M2) Assumption 2) and 4) (together with Proposition 3.6, Remark 3.8 and Lemma 3.10) imply that the functions

$$\frac{f_{\boldsymbol{\gamma}^{0,1}}}{g^0}, \frac{\partial_{\gamma_1} f_{\boldsymbol{\gamma}}\big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,1}}}{g^0}, \frac{\partial_{\gamma_2} f_{\boldsymbol{\gamma}}\big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,1}}}{g^0}, \frac{\partial_{\gamma_1\gamma_2} f_{\boldsymbol{\gamma}}\big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,1}}}{g^0} \text{ and } \frac{\partial_{\gamma_2\gamma_2} f_{\boldsymbol{\gamma}}\big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,1}}}{g^0}$$

belong to $L_2(g^0\nu)$. We will show that

$$\begin{aligned} & b f_{\boldsymbol{\gamma}^{0,1}}(\boldsymbol{x}) + \sum_{i=1}^{2} c_i \partial_{\gamma_i} f_{\boldsymbol{\gamma}}(\boldsymbol{x})\big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,1}} + \sum_{i=1}^{2} c_{i,2} \partial_{\gamma_i\gamma_2} f_{\boldsymbol{\gamma}}(\boldsymbol{x})\big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,1}} = 0 \quad \nu - a.s. \\ \Leftrightarrow \quad & b = c_1 = c_2 = c_{1,2} = c_{2,2} = 0. \end{aligned} \tag{3.45}$$

We follow the idea of the proof of (P1) and consider, firstly, those expressions to which the polynomials $T_u(\boldsymbol{x})T_v(\boldsymbol{x})$ belong. Thus we show

$$\begin{aligned} & r(\boldsymbol{x}) \sum_{i=1}^{2} \left\{ c_{i,2} \left[\sum_{u=1}^{2}\sum_{v=1}^{2} \Big(T_u(\boldsymbol{x})T_v(\boldsymbol{x}) \Big) \Big(\partial_{\gamma_i} \zeta_v(\boldsymbol{\gamma})\big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,1}} \Big) \Big(\partial_{\gamma_2} \zeta_u(\boldsymbol{\gamma})\big|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,1}} \Big) \right] \tilde{p}_{\boldsymbol{\zeta}(\boldsymbol{\gamma}^{0,1})}(\boldsymbol{x}) \right\} \\ = \quad & 0 \quad \nu - a.s. \\ & \Rightarrow \quad c_{1,2} = c_{2,2} = 0. \end{aligned}$$

In the following proof we use the abbreviation $\partial_i \zeta_u = \partial_{\gamma_i} \zeta_u(\boldsymbol{\gamma})|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,1}}$ since the true component parameter $\boldsymbol{\gamma}^{\mathbf{0,1}}$ is uniquely defined for $q = 1$.

According to assumption 10) one has $T_2 = \alpha T_1 T_1$ for $\alpha \neq 0$ and we assign $\alpha T_1 T_1$ to the first order partial derivatives in (3.46). Consequently, it is sufficient to show that

$$T_1(\boldsymbol{x})T_2(\boldsymbol{x}) \Big\{ c_{1,2} \Big((\partial_1\zeta_2)(\partial_2\zeta_1) + (\partial_1\zeta_1)(\partial_2\zeta_2) \Big) + 2c_{2,2}(\partial_2\zeta_1)(\partial_2\zeta_2) \Big\} = 0 \quad \nu - a.s., \tag{3.46}$$

$$T_2(\boldsymbol{x})T_2(\boldsymbol{x}) \Big\{ c_{1,2}(\partial_1\zeta_2)(\partial_2\zeta_2) + c_{2,2}(\partial_2\zeta_2)^2 \Big\} = 0 \quad \nu - a.s. \tag{3.47}$$

$$\Rightarrow \quad c_{1,2} = c_{2,2} = 0.$$

Thereby, we omit the factor $f_{\boldsymbol{\gamma}^{0,1}}$. Using that T_1T_2 and T_2T_2 are ν-affinely independent according to 10) and that $\partial_2\zeta_2 \neq 0$ leads to

$$c_{2,2} \overset{(3.47)}{=} -c_{1,2}\frac{\partial_1\zeta_2}{\partial_2\zeta_2}$$

and inserting the latter expression in (3.46) results in

$$c_{1,2}\Big((\partial_1\zeta_1)(\partial_2\zeta_2) - (\partial_1\zeta_2)(\partial_2\zeta_1) \Big) \quad = \quad c_{1,2} \ \det \Big(\partial_{\boldsymbol{\gamma}} \boldsymbol{\zeta}(\boldsymbol{\gamma})|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,1}} \Big) \quad = \quad 0.$$

It follows directly that $c_{1,2} = c_{2,2} = 0$ since the Jacobian matrix $\partial_{\boldsymbol{\gamma}}\boldsymbol{\zeta}(\boldsymbol{\gamma})|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,1}}$ is regular according to assumption 3). (Notice that $c_{2,2}$ can take any value in $\mathbb{R}$ if $\partial_2\zeta_2 = 0$.)

The remaining polynomials in $T_1(\boldsymbol{x})$ and $T_2(\boldsymbol{x})$ appear in formula (3.35) when evaluating the corresponding terms of the first order partial derivatives. Thus the remaining proof is in the same manner as the proof of (P1 t). □

3.3. Application to Specific Exponential Families

In this section we prove with the help of our main Theorem 3.13 the applicability of the theory from [DCG99] to various examples of finite mixtures families. Our main interest is the applicability to Gaussian families. Thanks to our modification (M1) a generalization of the theory from [DCG99] is applicable to bivariate Gaussian families with unknown covariance matrix and known mean, see Example 3.19. Best to our knowledge, we are the first to give sufficient conditions for deriving a limiting distribution for testing homogeneity against a two population mixture of such bivariate Gaussian families. Furthermore, due to our modification (M2), we apply the theory from [DCG99] to the test homogeneity against a two population mixture of univariate Gaussian with unknown mean and unknown variance, see Example 3.18. Our results corroborate Garel's conjecture (see Garel (2001)) that an asymptotic distribution of the LRT statistic exists for this test, but under a suitable restriction of the parameter space of the two population model, only.

It will be seen in a variety of examples that the structure of the component parameter space $\Gamma \subset I\!R^k$ is of considerable importance to ensure the applicability of Theorem 3.13. If the corresponding canonical parameter space $Z_* = I\!R^k$, we define Γ such that (A1) and (A2) are satisfied. In the case of $Z_* \neq I\!R^k$ we have to make additional restrictions on Γ, and the component parameter space may depend on the true parameters $\boldsymbol{\gamma}^{\mathbf{0},\mathbf{1}}, \ldots, \boldsymbol{\gamma}^{\mathbf{0},\boldsymbol{q}}$.

It will be seen that it is not difficult to choose an appropriate one-dimensional parameter space for testing homogeneity against a p population mixture such that (A1), (A2) as well as the conditions 4) and 5) of Theorem 3.13 hold. For such tests the conditions 7) and 9) of Theorem 3.13 are always satisfied according to Lemma 3.14. Furthermore, if the parameter function $\zeta : \Gamma \to \zeta(\Gamma)$ is bijective, five times differentiable with $\partial_\gamma \zeta(\gamma)|_{\gamma=\gamma^{0,1}} \neq 0$, $|\log(r)| \in L_2(g^0\nu)$ and (ID) holds, then the theory from [DCG99] is applicable. The latter conditions are quite easy to verify in the examples given here. As previously mentioned, Garel (2005) makes assumptions based on the second order derivatives of the log-density for testing homogeneity against a two population mixture founded on a one-dimensional parameter space. In addition to these general assumptions he does not use a reparameterization. In the case of Gaussian mixtures his results are consistent with those given in [DCG99] (see Garel (2001), p. 328). Thus the advantage of the theory from [DCG99] lies in the generality of mixtures and the arbitrary dimensionality of the parameter space.

In the introduction we have already given an outline of examples of the log-LRT in finite mixtures investigated in literature. In the following examples we will connect some of them with our results.

3.3.1. Application to Gaussian Families

In this subsection we make use of Theorem 3.13 to verify the applicability of the asymptotic theory from [DCG99] to several mixture models of Gaussian families. Thus we check the corresponding conditions of Theorem 3.13 as well as (A1) and (A2). The p-weakly identifiability condition (ID) always holds since the family of m-dimensional Gaussian distributions generates identifiable finite mixtures according to Yakowitz and Spragins (1968), Proposition 2 (see here, Remark 2.1).

It is well known that the Lebesgue λ-density of a m-dimensional $N(\boldsymbol{\mu}, \Sigma)$-distribution has the form

$$f_{(\boldsymbol{\mu},\Sigma)}(\boldsymbol{x}) \;=\; \frac{1}{\sqrt{(2\pi)^k \det(\Sigma)}} \exp\Big\{ -\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})'(\Sigma)^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\Big\} \quad \lambda\text{-a.s.} \tag{3.48}$$

where $(\mathfrak{X}, \mathcal{X}) = (I\!R^m, I\!B^m)$.

Keribin (2000), Theorem 4.1, says that the theory from [DCG99] is applicable to multivariate Gaussian mixtures with unknown mean and known covariance matrix $\Sigma = diag(\sigma^2, \dots, \sigma^2)$. In the corresponding proof on the applicability of (P0)c) she derives, firstly, a function $\bar{f}$ majorizing $\mathcal{F} = \{f_{\boldsymbol{\mu}} : \boldsymbol{\mu} \in \Gamma\}$, $\Gamma \subset I\!R^k$ compact. Secondly, she derives a minorant function $\underline{g}^0$ for the true density function $g^0 = \sum\limits_{l=1}^{q} \pi_l^0 f_{\boldsymbol{\gamma}^{0,l}}$. Consequently, she obtains $\sup\limits_{\boldsymbol{\mu}\in\Gamma} \frac{f_{\boldsymbol{\mu}}}{g^0} \le \frac{\bar{f}}{\underline{g}^0}$. Finally, she uses the fact, that the partial derivatives of $f_{\boldsymbol{\mu}}$ are equal to a product of $f_{\boldsymbol{\mu}}$ itself and some polynomial in $\boldsymbol{T}(\boldsymbol{x})$, to show validity of (P0)c). To obtain the minorant and the majorant function, Keribin (2000) uses the fact that for known $diag(\sigma^2, \dots, \sigma^2)$ and any $\boldsymbol{\mu} \in \Gamma$ the height lines of the k-dimensional ($k \ge 2$) Gaussian density function $f_{\boldsymbol{\mu}}(\boldsymbol{x})$ are spheres. Unfortunately, $\bar{f}$ and $\underline{g}^0$ do not suffer for arbitrary covariance matrices. Even in the case of $\sigma_1 \ne \sigma_2$ and $\Sigma = diag((\sigma_1)^2, \dots, (\sigma_k)^2)$, $k \ge 2$, one has to find a new way for the construction of a minorant and a majorant function such that the above mentioned relation holds, since the height lines of the corresponding density function $f_{\boldsymbol{\mu}}(\boldsymbol{x})$ are equal to ellipsoids. Thanks to our Proposition 3.11 we have a quite simple method to verify (P0)c) for multivariate Gaussian mixtures with unknown mean and arbitrary known covariance matrix by using the fact that the canonical parameter space $Z_* = I\!R^k$.

Furthermore, we use Proposition 3.11 as elementary simple tool for the verification of (P0)c) on all examples of exponential families considered in our work. Due to its universality we only need to investigate the supremum of the scalar product of $\langle \boldsymbol{\zeta}(\boldsymbol{\gamma}), \boldsymbol{T}(\boldsymbol{x})\rangle$ in a neighbourhood $\overline{U}_\varepsilon(\boldsymbol{\gamma}^{\mathbf{0},\boldsymbol{l}}) = \{\boldsymbol{\gamma} \in \Gamma : \|\boldsymbol{\gamma} - \boldsymbol{\gamma}^{\mathbf{0},\boldsymbol{l}}\| \le \varepsilon\}$ for $l = 1, \dots, q$, instead of analyzing the specific geometric structure of the densities of $\mathcal{F}$.

Example 3.15 *Let $\mathcal{P} = \{N(\boldsymbol{\mu}, \Sigma^0) : \boldsymbol{\mu} \in \Gamma\}$, $\Gamma \subset I\!R^k$, be a minimal k-parametric family of k-dimensional normal distributions with* **known covariance matrix** Σ^0 *and* **unknown mean** $\boldsymbol{\mu}$. *According to* (3.48) *the corresponding Lebesgue λ-densities in exponential representation* (3.1) *are given by*

$$f_{\boldsymbol{\mu}}(\boldsymbol{x}) \;=\; A(\boldsymbol{\mu}) \exp\Big\{ \big\langle \boldsymbol{\mu}, (\Sigma^0)^{-1}\boldsymbol{x}\big\rangle \Big\} r(\boldsymbol{x}) \quad \lambda\text{-}a.s.$$

with $\boldsymbol{\zeta}(\boldsymbol{\mu}) = \boldsymbol{\mu}$, $\boldsymbol{T}(\boldsymbol{x}) = (\Sigma^0)^{-1}\boldsymbol{x}$, $A(\boldsymbol{\mu}) = \big((2\pi)^k \det(\Sigma^0)\big)^{-\frac{1}{2}} \exp\{-\frac{1}{2}\boldsymbol{\mu}'(\Sigma^0)^{-1}\boldsymbol{\mu}\} = \mathcal{C}(\boldsymbol{\mu})$ and $r(\boldsymbol{x}) = \exp\{-\frac{1}{2}\boldsymbol{x}'(\Sigma^0)^{-1}\boldsymbol{x}\}$ (see e.g. Witting (1985), *p.* 147*). Clearly, one has*

$$Z_* \overset{(3.4),(3.7)}{=} \Big\{\boldsymbol{\mu} \in I\!R^k : 0 < \mathcal{C}(\boldsymbol{\mu})^{-1} < \infty\Big\} \;=\; I\!R^k.$$

Let $\Gamma \subset I\!R^k$ be an arbitrary set such that (A1) *and* (A2) *hold. Obviously, $\boldsymbol{\zeta} : \Gamma \to Z = \boldsymbol{\zeta}(\Gamma)$ is bijective.*

Condition 1) *(of Theorem* 3.13*) holds since $|\log(r(\boldsymbol{x}))| = \frac{1}{2}|\boldsymbol{x}'(\Sigma^0)^{-1}\boldsymbol{x}|$ and $\boldsymbol{T}(\boldsymbol{x})$ possesses moments of arbitrary order with respect to $f_{\boldsymbol{\gamma}^{0,l}}\lambda$.*

2) *holds for the identity function $\boldsymbol{\zeta} : \Gamma \to \Gamma$.*

3) *holds since the Jacobian matrix $\partial_{\boldsymbol{\mu}}\boldsymbol{\zeta}(\boldsymbol{\mu})$ is equal to the identity matrix I_k.*

4) *holds since $Z_* = I\!R^k$.*

5) *results from the following construction such that the assumptions of Proposition* 3.11 *are satisfied:*
For $(\Sigma^0)^{-1} = \left(s^0_{i,j}\right)_{i,j=1,\ldots,k}$ *one has* $\langle \boldsymbol{\mu}, (\Sigma^0)^{-1}\boldsymbol{x}\rangle = \sum_{i=1}^{k} x_i \sum_{j=1}^{k} \mu_j s^0_{i,j}$. *From the compactness assumption on* Γ *according to* (A1) *it follows*

$$\sup_{\boldsymbol{\mu}\in\Gamma} \sum_{j=1}^{k} |\mu_j s^0_{i,j}| \quad = \quad S_i \quad < \quad \infty \quad \text{for } i \in I = \{1,\ldots,k\}.$$

Let the power set of I *be given by* $\mathcal{S}(I) = \{M_1,\ldots,M_{2^k}\}$. *We choose the regions* $G_m \subset \mathfrak{X}$, $m = 1,\ldots,2^k$, *depending on the components of* $\boldsymbol{x} = (x_1,\ldots,x_k)'$ *whether they are positive or not, i.e.*

$$G_m \quad = \quad \Big\{(x_1,\ldots,x_k)' \in \mathfrak{X} : x_i \geq 0 \text{ for } i \in M_m \text{ and } x_i < 0 \text{ for } i \in I\backslash M_m\Big\}.$$

For $m = 1,\ldots,2^k$ *we define*

$$\boldsymbol{u^m} = (u^m_1,\ldots,u^m_k)' = \begin{cases} u^m_i = \quad S_i \quad , & i \in M_m \\ u^m_i = -S_i \quad , & i \in I\backslash M_m \end{cases}$$

which leads to $\sup_{\boldsymbol{\mu}\in\overline{U}_\varepsilon(\boldsymbol{\mu}^{0,l})} \sum_{i=1}^{k} x_i \sum_{j=1}^{k} \mu_j s^0_{i,j} \leq \sum_{i=1}^{k} x_i u^m_i$ *for* $\boldsymbol{x} \in G_m$ *and any* $l = 1,\ldots,q$.
Thus it follows together with $Z_* = I\!R^k$ *that the assumptions of Proposition* 3.11 *are satisfied.*

6) *holds since the functions* $T_i(\boldsymbol{x}) = (\Sigma^0)^{-1}x_i, T_i(\boldsymbol{x})T_j(\boldsymbol{x}) = (\Sigma^0)^{-2}x_ix_j$, $1 \leq i \leq j \leq k$, *are obviously* λ-*affinely independent.*

7) *is according to Lemma* 3.14 2)a) *equivalent to condition* 3).

Consequently, properties (P0) *and* (P1) *hold.*

Hence, the theory given in [DCG99] is applicable to multivariate Gaussian mixtures with known covariance matrix, unknown mean and arbitrary Γ such that (A1) as well as (A2) hold due to $Z_* = I\!R^k$.

Our second example is a generalization of Proposition 4.2 of Keribin (2000). She claims (without proof) that the asymptotic theory in [DCG99] is applicable to Gaussian families with an unknown variance lower bounded by a strictly positive number. Unfortunately, $Z_* = (-\infty, 0)^k \neq I\!R^k$ and we have to ensure that $\frac{f_\gamma}{g^0} \in L_2(g^0\lambda)$ holds for any $\boldsymbol{\gamma} \in \Gamma$ as assumed in (P1). We will show that the asymptotic theory from [DCG99] is applicable to Gaussian families with known mean and unknown covariance matrix $\Sigma = diag((\sigma_1)^2,\ldots,(\sigma_k)^2)$ and specific Γ. For testing one population against p populations we choose the parameter space Γ in a common way as, for instance, given by [DCG97], p. 297, and Garel (2001), p. 338 and p. 341. However, in the case of $q > 1$ we cannot choose the remaining true component parameters $\boldsymbol{\gamma}^{\boldsymbol{0},\boldsymbol{l}}$, $l = 2,\ldots,q$, as arbitrary elements of Γ if we want to ensure (P0)c) with the aid of Proposition 3.11.

Remark 3.16 *The fact that the parameter space* Γ *can depend on the true component parameters leads to difficulties in interpretation. The occurrence of this dependency seems to be inherent to investigations of asymptotic distributions with canonical parameter space* $Z_* \neq I\!R^k$. *We find such dependent formulations in* [DCG99] *as well as in Garel* (2001) *who derives the same dependent parameter spaces* Γ *using a completely different approach. Best to our knowledge there are no suggestions in literature how to avoid these dependencies.*

Example 3.17 *Let* $\mathcal{P} = \{N(\boldsymbol{\mu^0}, \Sigma) : \boldsymbol{\gamma} = (\sigma_1, \ldots, \sigma_k)' \in \Gamma\}$, $\Gamma \subset I\!R^k$, *be a minimal* k-*parametric family of* m-*dimensional normal distributions with* **known mean** $\boldsymbol{\mu^0}$ *and* **unknown diagonal covariance matrix** $\Sigma = diag((\sigma_1)^2, \ldots, (\sigma_k)^2)$. *According to* (3.48) *the corresponding Lebesgue* λ-*densities in exponential representation* (3.1) *are given by*

$$f_{\boldsymbol{\gamma}}(\boldsymbol{x}) \quad = \quad \Big((2\pi)^k \prod_{i=1}^k (\sigma_i)^2\Big)^{-\frac{1}{2}} \exp\Big\{ -\frac{1}{2}\sum_{i=1}^k \Big(\frac{x_i - \mu_i^0}{\sigma_i}\Big)^2\Big\} \quad \lambda-a.s.$$

with $\boldsymbol{\zeta}(\boldsymbol{\gamma}) = \Big(-\frac{1}{2}(\sigma_1)^{-2}, \ldots, -\frac{1}{2}(\sigma_k)^{-2}\Big)'$, $\boldsymbol{T}(\boldsymbol{x}) = \Big((x_1 - \mu_1^0)^2, \ldots, (x_k - \mu_k^0)^2\Big)'$, $A(\boldsymbol{\gamma}) = \Big((2\pi)^k \prod_{i=1}^k (\sigma_i)^2\Big)^{-\frac{1}{2}} = \mathcal{C}(\boldsymbol{\zeta}(\boldsymbol{\gamma}))$ *and* $r(\boldsymbol{x}) = 1$ *(see e.g. Witting* (1985), *p.* 144*). Clearly, one has*

$$Z_* \overset{(3.4)}{=} \Big\{\boldsymbol{\zeta} \in I\!R^k : 0 < \int \exp\Big(\langle\boldsymbol{\zeta}, \boldsymbol{T}(\boldsymbol{x})\rangle\Big)\, d\boldsymbol{x} < \infty\Big\} \quad = \quad (-\infty, 0)^k.$$

For $\varepsilon_1, \ldots, \varepsilon_k > 0$ *and* $\boldsymbol{\gamma^{0,1}} = \Big(\sigma_1^{0,1}, \ldots, \sigma_k^{0,1}\Big)' \in \mathop{\times}_{i=1}^{k} (\sqrt{\varepsilon_i}, \infty)$ *we define the parameter space by*

$$\Gamma \quad = \quad \Big[\sqrt{\varepsilon_1}, \Big(2(\sigma_1^{0,1})^2 - \varepsilon_1\Big)^{\frac{1}{2}}\Big] \times \ldots \times \Big[\sqrt{\varepsilon_k}, \Big(2(\sigma_k^{0,1})^2 - \varepsilon_k\Big)^{\frac{1}{2}}\Big].$$

Let the remaining distinct true component parameters be given by

$$\boldsymbol{\gamma^{0,2}}, \ldots, \boldsymbol{\gamma^{0,q}} \quad \in \quad int(\Gamma) \cap \mathop{\times}_{i=1}^{k} \Big(\sqrt{\varepsilon_i}, \sqrt{\tfrac{3}{2}}\,\sigma_i^{0,1}\Big). \tag{3.49}$$

$\boldsymbol{\zeta} : \Gamma \to Z = \boldsymbol{\zeta}(\Gamma)$ *is bijective since* $\Gamma \subset (0, \infty)^k$. *Obviously,* (A1) *as well as* (A2) *are satisfied and it will be shown below that this construction leads to condition* 5) *of Theorem* 3.13.
Condition 1) *(of Theorem* 3.13*) clearly holds since* $|\log(r(\boldsymbol{x}))| = 0$.

2) *holds obviously for the function* $\boldsymbol{\zeta} : \Gamma \to \boldsymbol{\zeta}(\Gamma)$ *since* $\sigma_1, \ldots, \sigma_k \neq 0$.
3) *holds since the Jacobian matrix* $\partial_{\boldsymbol{\gamma}}\boldsymbol{\zeta}(\boldsymbol{\gamma})$ *is equal to the diagonal matrix* $diag\Big(\sigma_1^{-3}, \ldots, \sigma_k^{-3}\Big)$.
4) *By construction of* Γ *it is sufficient to show that there exists some* $l \in \{1, \ldots, q\}$ *such that* $2\boldsymbol{\zeta}(\boldsymbol{\gamma}) - \boldsymbol{\zeta}(\boldsymbol{\gamma^{0,l}}) \in (-\infty, 0)^k$ *for all* $\boldsymbol{\gamma} \in \Gamma\backslash\{\boldsymbol{\gamma^{0,l}}\}$. *Without loss of generality we show that* $2\zeta_1(\boldsymbol{\gamma}) - \zeta_1(\boldsymbol{\gamma^{0,1}}) < 0$ *for all* $\boldsymbol{\gamma} \in \Gamma\backslash\{\boldsymbol{\gamma^{0,1}}\}$ *due to the form of* $\boldsymbol{\zeta} : \Gamma \to Z$.
   Thus it follows 4) *from* $-\frac{2}{2(\sigma_1)^2} + \frac{1}{2(\sigma_1^{0,1})^2} \leq \frac{-2(\sigma_1^{0,1})^2 + (2(\sigma_1^{0,1})^2 - \varepsilon_1)}{2(\sigma_1\sigma_1^{0,1})^2} < 0$ *due to Lemma* 3.10.
5) *results from the following construction by applying Proposition* 3.11*:*
   Let $\sigma_i^M = \sup_{l=1,\ldots,q} \sigma_i^{0,l}$ *for* $i = 1, \ldots, k$, $\boldsymbol{u^1} = \Big(-\frac{1}{2}(\sigma_1^M + \varepsilon)^{-2}, \ldots, -\frac{1}{2}(\sigma_k^M + \varepsilon)^{-2}\Big)'$ *and* $G_1 = I\!R^k$. *Consequently, for any* $l = 1, \ldots, q$ *we have*

$$\sup_{\boldsymbol{\gamma} \in \overline{U}_\varepsilon(\boldsymbol{\gamma^{0,l}})} -\frac{1}{2}\sum_{i=1}^k \Big(\frac{x_i - \mu_i^0}{\sigma_i}\Big)^2 \quad \leq \quad \sum_{i=1}^k u_i^1 (x_i - \mu_i^0)^2 \quad \text{for } \boldsymbol{x} \in G_1.$$

   Additionally, for $i = 1, \ldots, k$ *we obtain* $\boldsymbol{u^1} \in Z_*$ *and* $3u_i^1 - 2\boldsymbol{\zeta}(\sigma_i^{0,1}) = \frac{-3(\sigma_i^{0,1})^2 + 2(\sigma_i^M + \varepsilon)^2}{2((\sigma_i^M + \varepsilon)\sigma_i^{0,1})^2} < 0$ *for* $i = 1, \ldots, k$ *and sufficient small* $\varepsilon > 0$ *via the construction of* Γ *and* (3.49). *Thus* $\boldsymbol{u^1} \in \tilde{Z}_3$ *due to Lemma* 3.10.
6) *holds since the functions* $T_i(\boldsymbol{x}) = (x_i - \mu_i^0)^2$, $T_i(\boldsymbol{x})T_j(\boldsymbol{x}) = (x_i - \mu_i^0)^2(x_j - \mu_i^0)^2$, $1 \leq i \leq j \leq k$, *are obviously* λ-*affinely independent.*
7) *is according to Lemma* 3.14 2)a) *equivalent to condition* 3).

Consequently, properties (P0) *and* (P1) *hold.*

Our third example is an application of our modification (M2). We show that our generalization of the theory given in [DCG99] is applicable for testing homogeneity against a two population mixture of univariate Gaussian population with unknown mean and unknown variance. More precisely, we verify its applicability of testing $g \in \mathcal{G}_1$ against $g \in \mathcal{G}_{2,M} \backslash \mathcal{G}_1$, where the corresponding one population and two population mixture model is given by

$$\mathcal{G}_1 = \Big\{ f_{\gamma} : \gamma = (\mu, \sigma)' \in \Gamma \Big\}, \tag{3.50}$$

$$\mathcal{G}_{2,M} = \Big\{ g_{\boldsymbol{\pi},(\boldsymbol{\gamma^1},\boldsymbol{\gamma^2})} = \pi_1 f_{\boldsymbol{\gamma^1}} + \pi_2 f_{\boldsymbol{\gamma^2}} : (\boldsymbol{\gamma^1}, \boldsymbol{\gamma^2}) \in \tilde{\Gamma},\ \boldsymbol{\pi} \in \mathcal{T}_2 \Big\} \tag{3.51}$$

with $\tilde{\Gamma}$ being specific restriction on $\Gamma \times \Gamma$, see p. 23 Figure 2.1. While the compact parameter space Γ is defined in the same way as in the both latter examples.
As mentioned above the sufficient conditions from [DCG99] for the existence of an limiting distribution are not satisfied for this kind of test without a modification, since $\partial_\mu \partial_\mu f_{(\mu,\sigma)} = \frac{1}{\sigma} \partial_\sigma f_{(\mu,\sigma)}$ contradicts (P1).

Although we consider a restricted two population model, we do not use a separation condition on the parameters as Gosh and Sen (1985) do. Garel (2001), p. 342, gives a conjecture on the asymptotic distribution of the LRT for testing homogeneity against two populations in univariate Gaussian mixtures with unknown mean and unknown variance (without using a separations condition). The verification of his conjecture with the help of the methods found in [DCG99] can be carried out, only, if one restricts the parameter space to the two population mixture model.

Example 3.18 *For $\mu^0 \in I\!R$, $\sigma^0, l, u > 0$, $\varepsilon \in \big(0, (\sigma^0)^2\big)$ and sufficient small $\tilde{\varepsilon} > 0$ let*

$$\Gamma = \Big[\mu^0 - l, \mu^0 + u\Big] \times \Big[\sqrt{\varepsilon}, \sqrt{2(\sigma^0)^2 - \varepsilon}\Big], \tag{3.52}$$

$$\tilde{\Gamma} = \Big\{ (\boldsymbol{\gamma^1}, \boldsymbol{\gamma^2}) \in \Gamma \times \Gamma : \|\boldsymbol{\gamma^1} - \boldsymbol{\gamma^0}\| \le \|\boldsymbol{\gamma^2} - \boldsymbol{\gamma^0}\|,\ |\mu^2 - \mu^0| \le |\sigma^2 - \sigma^0| \Big\}$$

and let the one population and two population mixture model be given by (3.50) *and* (3.51), *respectively. Let $\mathcal{P} = \{N(\mu, \sigma) : \boldsymbol{\gamma} = (\mu, \sigma)' \in \Gamma\}$ be a minimal 2-parametric family of univariate normal distributions with* **unknown standard deviation** *σ and* **unknown mean** *μ. By construction* (A1) *and* (A2) *hold for $q = 1$ and $\boldsymbol{\gamma^0} = (\mu^0, \sigma^0)'$. The corresponding Lebesgue λ-densities of $\mathcal{P}$ in exponential representation* (3.1) *are given by*

$$f_{(\mu,\sigma)}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(-\frac{\mu^2}{2\sigma^2}\Big) \exp\Big(\frac{\mu}{\sigma^2} x - \frac{1}{2\sigma^2} x^2\Big) \quad \lambda - a.s.$$

with $\boldsymbol{\zeta}(\mu, \sigma) = (\mu/\sigma^2, -1/(2\,\sigma^2))'$, $\boldsymbol{T}(x) = (x, x^2)'$, $A(\mu, \sigma) = (2\pi\sigma^2)^{-\frac{1}{2}} \exp(-\mu^2/(2\sigma^2)) = C(\boldsymbol{\zeta}(\mu, \sigma))$ and $r(\boldsymbol{x}) = 1$ (see e.g. Witting (1985), p. 144). $\boldsymbol{\zeta} : \Gamma \to \boldsymbol{\zeta}(\Gamma)$ is bijective since $\sigma > 0$.

As x^2 is growing faster than x for $x \to \pm\infty$ we have

$$Z_* \overset{(3.4)}{=} \Big\{ \boldsymbol{\zeta} \in I\!R^2 : 0 < \int \exp\Big\{ \langle \boldsymbol{\zeta}, \boldsymbol{T}(x) \rangle \Big\}\, dx < \infty \Big\} = I\!R \times (-\infty, 0).$$

Condition 1) *(of Theorem* 3.13*) clearly holds since $|\log(r(\boldsymbol{x}))| = 0$.*
2) *holds obviously for the function $\boldsymbol{\zeta} : \Gamma \to \boldsymbol{\zeta}(\Gamma)$ since $\sigma \neq 0$.*
3) *holds since for any $(\mu, \sigma)' \in \Gamma \subset I\!R \times (0, \infty)$ the Jacobian matrix has the form*

$$\partial_{(\mu,\sigma)} \boldsymbol{\zeta}(\mu, \sigma) = \begin{pmatrix} \sigma^{-2} & -2\mu\sigma^{-3} \\ 0 & \sigma^{-3} \end{pmatrix}.$$

4) *Due to Lemma* 3.10 *it is sufficient to show that for any* $(\mu,\sigma)' \in \Gamma\backslash\{\boldsymbol{\gamma^0}\}$ *one has* $2\boldsymbol{\zeta}(\mu,\sigma) - \boldsymbol{\zeta}(\mu^0,\sigma^0) \in Z_*$.

Thus it follows 4) *from* $2\mu/\sigma^2 - \mu^0/(\sigma^0)^2 \in I\!R$ *and* $-\frac{2}{2\sigma^2} + \frac{1}{2(\sigma^0)^2} \overset{(3.52)}{\leq} \frac{-2(\sigma^0)^2+(2(\sigma^0)^2-\varepsilon)}{2(\sigma\sigma^0)^2} < 0$.

5) *results from the following construction and applying Proposition* 3.11*:*
For $(\sqrt{3/2}-1)\sigma^0 > \delta > 0$ *(as used below) and* $S = \max\{|\mu^0 - l|, |\mu^0 + u|\}$ *it follows that*

$$\sup_{\gamma\in\overline{U}_\delta(\gamma^0)} \left\langle \boldsymbol{\zeta}(\mu,\sigma), \boldsymbol{T}(x)\right\rangle = \sup_{\gamma\in\overline{U}_\delta(\gamma^0)} \frac{2\mu x - x^2}{2\sigma^2} \leq \begin{cases} \dfrac{Sx}{\varepsilon} - \dfrac{x^2}{2(\sigma^0+\delta)^2} &, \quad x \geq 0 \\ \dfrac{-Sx}{\varepsilon} - \dfrac{x^2}{2(\sigma^0+\delta)^2} &, \quad x < 0. \end{cases}$$

Consequently, we define for the regions $G_1 = [0,\infty)$ *and* $G_2 = (-\infty, 0)$ *the corresponding parameters*

$$\boldsymbol{u^1} = \left(\frac{S}{\varepsilon}, -\frac{1}{2(\sigma^0+\delta)^2}\right)' \in Z_* \quad \textit{and} \quad \boldsymbol{u^2} = \left(-\frac{S}{\varepsilon}, -\frac{1}{2(\sigma^0+\delta)^2}\right)' \in Z_*,$$

respectively. Simple computations lead to $3\boldsymbol{u^i} - 2\Big(\frac{\mu^0}{(\sigma^0)^2}, -\frac{1}{2(\sigma^0)^2}\Big)' \in Z_*$ *for* $i = 1,2$ *(since* $(\sqrt{3/2}-1)\sigma^0 > \delta$*). Thus* $\boldsymbol{u^i} \in \tilde{Z}_3$, $i = 1,2$, *due to Lemma* 3.10.

10) *holds since* $\partial_\sigma\zeta_2(\mu,\sigma) = \sigma^{-3} \neq 0$, $T_1(x)$, $T_2(x)$, $T_1(x)T_2(x)$, $T_2(x)T_2(x)$ *given by* x, x^2, x^3 *and* x^4 *are* λ*-affinely independent and* $T_2(x) = x^2 = T_1(x)T_1(x)$.

Consequently, properties (P0), (P1 t) *and* (M2) *hold.*

Our fourth example is an application of our modification (M1) according to Theorem 3.13. Best to our knowledge we are the first to give sufficient conditions for deriving a limiting distribution for testing homogeneity against a mixture of bivariate Gaussian distributions with arbitrary unknown covariance matrix and known mean. As mentioned above the sufficient condition (P1) for the existence of a limiting distribution are not satisfied. For instance, $\partial_{\sigma_1}\partial_{\sigma_2} f_{(\sigma_1,\sigma_2,\varrho)}|_{(\sigma_1,\sigma_2,\varrho)'=(\sigma_1^0,\sigma_2^0,0)'} = \frac{1}{\sigma_1^0\sigma_2^0}\partial_\varrho\partial_\varrho f_{(\sigma_1,\sigma_2,\varrho)}|_{(\sigma_1,\sigma_2,\varrho)'=(\sigma_1^0,\sigma_2^0,0)'}$ which contradicts (P1).

Example 3.19 *Let* $\mathcal{P} = \{N(\boldsymbol{\mu^0},\Sigma) : \boldsymbol{\gamma} = (\sigma_1,\sigma_2,\varrho)' \in \Gamma\}$, $\Gamma \subset (0,\infty)^2 \times (-1,1)$, *be a minimal* 3*-parametric family of bivariate normal distributions with* **unknown covariance matrix** Σ *and* **known mean** $\boldsymbol{\mu^0}$. *Without loss of generality we assume that* $\boldsymbol{\mu^0} = \boldsymbol{0}$ *since normal densities are translation invariant to the conditions of Theorem* 3.13. *Let*

$$\Sigma = \begin{pmatrix} (\sigma_1)^2 & \varrho\sigma_1\sigma_2 \\ \varrho\sigma_1\sigma_2 & (\sigma_2)^2 \end{pmatrix} \quad \textit{and} \quad \Sigma^{-1} = \begin{pmatrix} \frac{1}{(\sigma_1)^2(1-\varrho^2)} & -\frac{\varrho}{\sigma_1\sigma_2(1-\varrho^2)} \\ -\frac{\varrho}{\sigma_1\sigma_2(1-\varrho^2)} & \frac{1}{(\sigma_2)^2(1-\varrho^2)} \end{pmatrix} = \begin{pmatrix} s_{1,1} & s_{1,2} \\ s_{2,1} & s_{2,2} \end{pmatrix}.$$

According to (3.48) *the corresponding Lebesgue* λ*-densities in exponential representation as in* (3.1) *are given by*

$$f_{(\sigma_1,\sigma_2,\varrho)}(x_1,x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\varrho^2}} \exp\left\{-\frac{1}{2(1-\varrho^2)}\left[\left(\frac{x_1}{\sigma_1}\right)^2 - 2\varrho\frac{x_1}{\sigma_1}\frac{x_2}{\sigma_2} + \left(\frac{x_2}{\sigma_2}\right)^2\right]\right\} \ \lambda - a.s.$$

with $\boldsymbol{\zeta}(\sigma_1,\sigma_2,\varrho) = \Big(-s_{1,1}, -s_{2,2}, -s_{1,2}\Big)'$, $\boldsymbol{T}(\boldsymbol{x}) = \Big(\frac{1}{2}(x_1)^2, \frac{1}{2}(x_2)^2, x_1x_2\Big)'$, $r(\boldsymbol{x}) = 1$ *and* $A(\boldsymbol{\gamma}) = \Big(2\pi\sigma_1\sigma_2\sqrt{1-\varrho^2}\Big)^{-1} = \mathcal{C}(\boldsymbol{\zeta}(\sigma_1,\sigma_2,\varrho))$. *We have*

$$\begin{aligned} Z_* &\overset{(3.4),(3.7)}{=} \Big\{\boldsymbol{\zeta}(\sigma_1,\sigma_2,\varrho) \in I\!R^3 : 0 < \mathcal{C}(\boldsymbol{\zeta}(\sigma_1,\sigma_2,\varrho))^{-1} < \infty\Big\} \\ &= \Big\{(\zeta_1,\zeta_2,\zeta_3)' \in (-\infty,0)\times(-\infty,0)\times I\!R : \zeta_1\zeta_2 - (\zeta_3)^2 > 0\Big\}. \end{aligned}$$

In the following we give an algorithm how to define a parameter space

$$\begin{aligned} \Gamma &= \Gamma_{\sigma_1^0} \times \Gamma_{\sigma_2^0} \times \Gamma_\varrho, \\ \Gamma_{\sigma_i^0} &= \Big[\sqrt{\varepsilon_i}, \sqrt{2(\sigma_i^0)^2(1-(\varrho^0)^2)-\varepsilon_i}\,\Big],\ i=1,2, \end{aligned} \tag{3.53}$$

$\varepsilon_i \in \Big(0, 2(\sigma_i^0)^2(1-(\varrho^0)^2)\Big)$, $i=1,2$, *such that conditions* 4), 5) *and* 9) *of Theorem* 3.13 *are satisfied. The idea is to choose some* $\sigma_1^0, \sigma_2^0 > 0$ *and after it to define* Γ_ϱ *successively such that for any* $\varrho^0 \in \Gamma_\varrho$ *the latter mentioned conditions are satisfied:*

1. *Define* $\sigma_1^0, \sigma_2^0 > 0$.
2. *Define an interval* $[r_1, r_2]$ *with* $r_1 \in (-1, 0)$ *and* $r_2 \in (0,1)$ *such that condition* 9) *of Theorem* 3.13 *is satisfied for* $\sigma_1^0, \sigma_2^0 > 0$ *and any* $\varrho^0 \in [r_1, r_2]$. *Lemma* 3.21 1) *ensures the existence of such* r_1, r_2 *and in its proof an algorithm is given how to find* r_1 *and* r_2.
3. *Define a subset* $[A,B] \subset [r_1, r_2]$ *with* $A \in (-1,0)$ *and* $B \in (0,1)$ *such that for* $\sigma_1^0, \sigma_2^0 > 0$ *and* $\varrho^0 \in [A,B]$ *we have* $\boldsymbol{\zeta}(\boldsymbol{\gamma}) \in \tilde{Z}_2$ *for all* $\boldsymbol{\gamma} \in \Gamma_{\sigma_1^0} \times \Gamma_{\sigma_2^0} \times [A,B]$. *Lemma* 3.21 2) *ensures the existence of such* A, B *and in its proof an algorithm is given how to find* A *and* B, *(see* (3.58) and (3.59)*, respectively). As a consequence condition* 4) *of Theorem* 3.13 *is satisfied.*
4. *Define*

$$\Gamma_\varrho = [A,B] \cap \Big[-\tfrac{1}{6}(-1+\sqrt{13}) + \varepsilon_3, \tfrac{1}{6}(-1+\sqrt{13}) - \varepsilon_3\Big]\Big), \tag{3.54}$$

   $\varepsilon_3 \in (0, \frac{1}{6}(-1+\sqrt{13}))$. *Then Lemma* 3.21 3) *ensures the existence of some* $\varepsilon > 0$ *such that* $\boldsymbol{u^1}, \boldsymbol{u^2} \in \tilde{Z}_3$ *(*$\boldsymbol{u^1}, \boldsymbol{u^2}$ *as given below). Thus Proposition* 3.11 *is applicable to our below stated construction to show that condition* 5) *of Theorem* 3.13 *is satisfied.*

If $\varrho^0 = 0$ *then conditions* 4), 5) *and* 9) *of Theorem* 3.13 *are also satisfied for* $\Gamma = \Gamma_{\sigma_1^0} \times \Gamma_{\sigma_2^0} \times [-r, r]$ *with*

$$r = \min\Big\{\Big(\frac{\varepsilon_1}{2(\sigma_1^0)^2-\varepsilon_1}\,\frac{\varepsilon_2}{2(\sigma_2^0)^2-\varepsilon_2} - \varepsilon_4\Big)^{\frac{1}{2}}, 1-\varepsilon_5\Big\}, \tag{3.55}$$

$\varepsilon_4 \in \Big(0, \frac{\varepsilon_1}{2(\sigma_1^0)^2-\varepsilon_1}\,\frac{\varepsilon_2}{2(\sigma_2^0)-\varepsilon_2}\Big)$ *and* $\varepsilon_5 \in (0,1)$ *(what will be seen below).*

Clearly, (A1) *as well as* (A2) *hold and* $\boldsymbol{\zeta}(\boldsymbol{\gamma}) : \Gamma \to \boldsymbol{\zeta}(\Gamma)$ *is bijective.*
Condition 1) *(of Theorem* 3.13*) holds since* $|\log(r(\boldsymbol{x}))| = 0$.
2) *holds for* $\boldsymbol{\zeta} : \Gamma \to \boldsymbol{\zeta}(\Gamma)$ *since* $\sigma_1, \sigma_2 \neq 0$, *and* $\varrho \neq \pm 1$.
3) *holds since for any* $(\sigma_1, \sigma_2, \varrho)' \in \Gamma \subset (0,\infty)^2 \times (-1,1)$ *the Jacobian matrix has the form*

$$\partial_{(\sigma_1,\sigma_2,\varrho)}\boldsymbol{\zeta}(\sigma_1,\sigma_2,\varrho) = \begin{pmatrix} \frac{2}{(\sigma_1)^3(1-\varrho^2)} & 0 & -\frac{2\varrho}{(\sigma_1)^2(1-\varrho^2)^2} \\ 0 & \frac{2}{(\sigma_2)^3(1-\varrho^2)} & -\frac{2\varrho}{(\sigma_2)^2(1-\varrho^2)^2} \\ -\frac{\varrho}{(\sigma_1)^2\sigma_2(1-\varrho^2)} & -\frac{\varrho}{\sigma_1(\sigma_2)^2(1-\varrho^2)} & \frac{1+\varrho^2}{\sigma_1\sigma_2(1-\varrho^2)^2} \end{pmatrix} \tag{3.56}$$

   which is regular since $\det\Big(\partial_{(\sigma_1,\sigma_2,\varrho)}\boldsymbol{\zeta}(\sigma_1,\sigma_2,\varrho)\Big) = 4\Big((\sigma_1)^4(\sigma_2)^4(1-\varrho^2)^3\Big)^{-1} \neq 0.$

4) *If* $\varrho^0 = 0$ *and* $\Gamma = \Gamma_{\sigma_1^0} \times \Gamma_{\sigma_2^0} \times [-r, r]$, r *as in* (3.55), *then condition* 4) *holds due to Lemma* 3.20. *If* $\Gamma = \Gamma_{\sigma_1^0} \times \Gamma_{\sigma_2^0} \times \Gamma_\varrho$, Γ_ϱ *as in* (3.54), *then condition* 4) *holds according to the construction of* Γ_ϱ *and Lemma* 3.21 2).

5) *results from the following construction and applying Proposition* 3.11*:*
Let $\boldsymbol{\gamma^0} = (\sigma_1^0, \sigma_2^0, \varrho^0)' \in \Gamma_{\sigma_1^0} \times \Gamma_{\sigma_2^0} \times \Gamma_\varrho$, Γ_ϱ *as in* (3.54). *(Notice that* $(\sigma_1^0, \sigma_2^0, 0)' \in \Gamma_{\sigma_1^0} \times \Gamma_{\sigma_2^0} \times \Gamma_\varrho$.*) We have*

$$\sup_{\gamma \in \overline{U}_\varepsilon(\boldsymbol{\gamma}^0)} \big\langle \boldsymbol{\zeta}(\mu, \sigma), \boldsymbol{T}(\boldsymbol{x}) \big\rangle \;=\; \sup_{\gamma \in \overline{U}_\varepsilon(\boldsymbol{\gamma}^0)} -\frac{1}{2(1-\varrho^2)} \left[\left(\frac{x_1}{\sigma_1}\right)^2 - 2\varrho \frac{x_1}{\sigma_1}\frac{x_2}{\sigma_2} + \left(\frac{x_2}{\sigma_2}\right)^2 \right].$$

Let $G_1 = \Big\{(x_1, x_2)' \in I\!R^2 : \; x_1 x_2 \ge 0\Big\}$, $G_2 = \Big\{(x_1, x_2)' \in I\!R^2 : \; x_1 x_2 < 0\Big\}$ *and*

$$\begin{aligned} \boldsymbol{u^1} &= \Big(-\frac{1}{(\sigma_1^0+\varepsilon)^2}, -\frac{1}{(\sigma_2^0+\varepsilon)^2}, \frac{\varrho^0+\varepsilon}{(1-(\varrho^0+\varepsilon)^2)(\sigma_1^0-\varepsilon)(\sigma_2^0-\varepsilon)} \Big)', \\ \boldsymbol{u^2} &= \Big(-\frac{1}{(\sigma_1^0+\varepsilon)^2}, -\frac{1}{(\sigma_2^0+\varepsilon)^2}, \frac{\varrho^0-\varepsilon}{(1-(\varrho^0-\varepsilon)^2)(\sigma_1^0+\varepsilon)(\sigma_2^0+\varepsilon)} \Big)' \end{aligned}$$

for some $\varepsilon \in (0, m)$, $m = \min\{\sigma_1^0, \sigma_2^0, 1 - |\varrho^0|, |\varrho^0| \mathbb{1}_{\varrho^0 \neq 0}\}$. *If* $\varrho^0 \ge 0$ *then we have for any* $\boldsymbol{x} \in G_i$, $i = 1, 2$

$$\sup_{\gamma \in \overline{U}_\varepsilon(\boldsymbol{\gamma}^0)} \big\langle \boldsymbol{\zeta}(\mu, \sigma), \boldsymbol{T}(\boldsymbol{x}) \big\rangle \;\le\; \big\langle \boldsymbol{u^i}, \boldsymbol{T}(\boldsymbol{x}) \big\rangle$$

and if $\varrho^0 < 0$ *then we have*

$$\sup_{\gamma \in \overline{U}_\varepsilon(\boldsymbol{\gamma}^0)} \big\langle \boldsymbol{\zeta}(\mu, \sigma), \boldsymbol{T}(\boldsymbol{x}) \big\rangle \;\le\; \begin{cases} \big\langle \boldsymbol{u^2}, \boldsymbol{T}(\boldsymbol{x}) \big\rangle &, \; \boldsymbol{x} \in G_1 \\ \big\langle \boldsymbol{u^1}, \boldsymbol{T}(\boldsymbol{x}) \big\rangle &, \; \boldsymbol{x} \in G_2. \end{cases}$$

In both cases there exists some $\varepsilon \in (0, m)$ *such that* $\mathbf{u^1}, \mathbf{u^2} \in \tilde{Z}_3$ *due to Lemma* 3.21 3). *Thus Proposition* 3.11 *is applicable.*

8) *is fulfilled for* $U = \{(1,3), (2,3)\}$, $U^c = \{(1,2)\}$,

$$\begin{aligned} \mathcal{H} &= \Big\{ T_1(\boldsymbol{x}),\, T_2(\boldsymbol{x}),\, T_3(\boldsymbol{x}),\, T_1(\boldsymbol{x})T_1(\boldsymbol{x}),\, T_2(\boldsymbol{x})T_2(\boldsymbol{x}),\, T_3(\boldsymbol{x})T_3(\boldsymbol{x}),\, T_1(\boldsymbol{x})T_3(\boldsymbol{x}),\, T_2(\boldsymbol{x})T_3(\boldsymbol{x}) \Big\} \\ &= \Big\{ \tfrac{1}{2}x_1^2,\; \tfrac{1}{2}x_2^2,\; x_1x_2,\; \tfrac{1}{4}x_1^4,\; \tfrac{1}{4}x_2^4,\; x_1^2x_2^2,\; \tfrac{1}{2}x_1^3x_2,\; \tfrac{1}{2}x_1x_2^3 \Big\} \end{aligned}$$

since $T_1(\boldsymbol{x})T_2(\boldsymbol{x}) = \frac{1}{4}x_1^2x_2^2 = \frac{1}{4}T_3(\boldsymbol{x})T_3(\boldsymbol{x})$.

9) *If* $\varrho^0 = 0$ *then the Jacobian matrix* $\partial_{(\sigma_1,\sigma_2,\varrho)}\boldsymbol{\zeta}(\sigma_1, \sigma_2, \varrho)|_{\boldsymbol{\gamma}^0 = (\sigma_1^0, \sigma_2^0, 0)}$ *(see* (3.56)*) is equal to a diagonal one. As a consequence condition* 9) *holds due to Lemma* 3.14 2)b). *If* $\varrho^0 \neq 0$, $\varrho^0 \in \Gamma_\varrho$, *then condition* 9) *holds according to the construction of* Γ_ϱ *and Lemma* 3.21 1).

Consequently, properties (P0), (P1) *and* (M1) *are satisfied.*

Lemma 3.20 *Let* $\mathcal{P} = \{N(\boldsymbol{\mu^0}, \Sigma) : \boldsymbol{\gamma} = (\sigma_1, \sigma_2, \varrho)' \in \Gamma\}$ *and be given as in Example* 3.19. *Let* $\boldsymbol{\gamma^0} = (\sigma_1^0, \sigma_2^0, 0)'$ *be the true parameter and* $\Gamma = \Gamma_{\sigma_1^0} \times \Gamma_{\sigma_2^0} \times [-r, r]$ *with* $\Gamma_{\sigma_i^0}$, $i = 1, 2$ *as in* (3.53) *and* r *as in* (3.55). *Then for any* $\boldsymbol{\gamma} = (\sigma_1, \sigma_2, \varrho)' \in \Gamma \backslash \{\boldsymbol{\gamma^0}\}$ *one has* $\boldsymbol{\zeta}(\boldsymbol{\gamma}) \in \tilde{Z}_2$.

Proof of Lemma 3.20:

Due to Lemma 3.10 it is sufficient to show that for any $\boldsymbol{\gamma} = (\sigma_1, \sigma_2, \varrho)' \in \Gamma \backslash \{\boldsymbol{\gamma^0}\}$ one has $2\boldsymbol{\zeta}(\boldsymbol{\gamma}) - \boldsymbol{\zeta}(\boldsymbol{\gamma^0}) \in Z_*$. Thus we show that for any $\boldsymbol{\gamma} = (\sigma_1, \sigma_2, \varrho)' \in \Gamma \backslash \{\boldsymbol{\gamma^0}\}$

$$2\Sigma^{-1} - (\Sigma^0)^{-1} \;=\; \begin{pmatrix} \frac{2(\sigma_1^0)^2 - (\sigma_1)^2(1-\varrho^2)}{(\sigma_1^0)^2(\sigma_1)^2(1-\varrho^2)} & -\frac{2\varrho}{\sigma_1\,\sigma_2(1-\varrho^2)} \\ -\frac{2\varrho}{\sigma_1\,\sigma_2(1-\varrho^2)} & \frac{2(\sigma_2^0)^2 - (\sigma_2)^2(1-\varrho^2)}{(\sigma_2^0)^2(\sigma_2)^2(1-\varrho^2)} \end{pmatrix} \;=\; \begin{pmatrix} z_{1,1} & z_{1,2} \\ z_{2,1} & z_{2,2} \end{pmatrix}$$

is positive definite according to the form of Z_*.

We have $z_{1,1}, z_{2,2} > 0$ for any $(\sigma_1, \sigma_2, \varrho)' \in \Gamma_{\sigma_1^0} \times \Gamma_{\sigma_2^0} \times [-r, r]$ since

$$2(\sigma_i^0)^2 - (\sigma_i)^2(1-\varrho^2) \quad > \quad 2(\sigma_i^0)^2 - \Big(2(\sigma_i^0)^2 - \varepsilon_i\Big)(1-\varrho^2) \quad > \quad 0$$

according to the definition of $\Gamma_{\sigma_i^0}$, $\varepsilon_i \in (0, (\sigma_i^0)^2)$, $i = 1, 2$ and $r \in (0, 1)$. Furthermore, for any $(\sigma_1, \sigma_2, \varrho)' \in \Gamma_{\sigma_1^0} \times \Gamma_{\sigma_2^0} \times [-r, r]$ one has

$$\begin{aligned}
\det\Big(2\Sigma^{-1} - (\Sigma^0)^{-1}\Big) &= \frac{4(\sigma_1^0)^2(\sigma_2^0)^2 - 2(\sigma_1^0)^2(\sigma_2)^2 - 2(\sigma_1)^2(\sigma_2^0)^2 + (\sigma_1)^2(\sigma_2)^2(1-\varrho^2)}{(\sigma_1^0)^2(\sigma_2^0)^2(\sigma_1)^2(\sigma_2)^2(1-\varrho^2)} \\
&= \frac{\Big(2(\sigma_1^0)^2 - (\sigma_1)^2\Big)\Big(2(\sigma_2^0)^2 - (\sigma_2)^2\Big) - \varrho^2(\sigma_1)^2(\sigma_2)^2}{(\sigma_1^0)^2(\sigma_2^0)^2(\sigma_1)^2(\sigma_2)^2(1-\varrho^2)}
\end{aligned}$$

and

$$\det\Big(2\Sigma^{-1} - (\Sigma^0)^{-1}\Big) > 0 \quad \Leftrightarrow \quad \frac{\Big(2(\sigma_1^0)^2 - (\sigma_1)^2\Big)\Big(2(\sigma_2^0)^2 - (\sigma_2)^2\Big)}{(\sigma_1)^2(\sigma_2)^2} > \varrho^2.$$

For any $\boldsymbol{\gamma} = (\sigma_1, \sigma_2, \varrho)' \in \Gamma \backslash \{\boldsymbol{\gamma^0}\}$ we obtain

$$\frac{\Big(2(\sigma_1^0)^2 - (\sigma_1)^2\Big)\Big(2(\sigma_2^0)^2 - (\sigma_2)^2\Big)}{(\sigma_1)^2(\sigma_2)^2} \quad > \quad \frac{\varepsilon_1}{2(\sigma_1^0)^2 - \varepsilon_1}\,\frac{\varepsilon_2}{2(\sigma_2^0)^2 - \varepsilon_2} \quad \overset{(3.55)}{>} \quad r^2 \quad \geq \quad \varrho^2$$

by making use of the form of Γ. Consequently, $2\Sigma^{-1} - (\Sigma^0)^{-1}$ is positive definite. □

Lemma 3.21 *Let $\mathcal{P} = \{N(\boldsymbol{\mu^0}, \Sigma) : \boldsymbol{\gamma} = (\sigma_1, \sigma_2, \varrho)' \in \Gamma\}$ be given as in Example* 3.19.

1) *Let $\sigma_i^0 > 0$, $i = 1, 2$. Then there exist r_1, r_2 with $r_1 \in (-1, 0)$ and $r_2 \in (0, 1)$ such that condition* 9) *of Theorem* 3.13 *is satisfied for any $\varrho^0 \in [r_1, r_2]$.*

2) *Let $\sigma_i^0 > 0$, $i = 1, 2$, and $[r_1, r_2]$ be given as in* 1)*. Then there exist A, B with $A \in (-1, 0)$, $B \in (0, 1)$ and $[A, B] \subset [r_1, r_2]$ such that for any $\boldsymbol{\gamma} = (\sigma_1, \sigma_2, \varrho)' \in \Gamma \backslash \{\boldsymbol{\gamma^0}\}$ one has $\boldsymbol{\zeta}(\boldsymbol{\gamma}) \in \tilde{Z}_2$, where $\boldsymbol{\gamma^0} \in \Gamma = \Gamma_{\sigma_1^0} \times \Gamma_{\sigma_2^0} \times [A, B]$ and*

$$\Gamma_{\sigma_i^0} = \Big[\sqrt{\varepsilon_i}, \sqrt{2(\sigma_i^0)^2(1-(\varrho^0)^2) - \varepsilon_i}\Big], \ i = 1, 2$$

for some ε_i with $(\sigma_i^0)^2(1 - (\varrho^0)^2) > \varepsilon_i > 0$, $i = 1, 2$.

3) *Let $\boldsymbol{\gamma^0} = (\sigma_1^0, \sigma_2^0, \varrho)' \in \Gamma_{\sigma_1^0} \times \Gamma_{\sigma_2^0} \times \Gamma_\varrho$, $\Gamma_{\sigma_i^0}$, $i = 1, 2$, as in* (3.53) *and Γ_ϱ as in* (3.54)*. Then there exists some $\varepsilon \in (0, m)$, $m = \min\{\sigma_1^0, \sigma_2^0, 1 - |\varrho^0|, |\varrho^0|\mathbb{1}_{\varrho^0 \neq 0}\}$, such that*

$$\begin{aligned}
\boldsymbol{u^1} &= \Big(-\frac{1}{(\sigma_1^0+\varepsilon)^2}, -\frac{1}{(\sigma_2^0+\varepsilon)^2}, \frac{\varrho^0+\varepsilon}{(1-\varepsilon^2)(\sigma_1^0-\varepsilon)(\sigma_2^0-\varepsilon)}\Big)' \in \tilde{Z}_3, \\
\boldsymbol{u^2} &= \Big(-\frac{1}{(\sigma_1^0+\varepsilon)^2}, -\frac{1}{(\sigma_2^0+\varepsilon)^2}, \frac{\varrho^0-\varepsilon}{(1-\varepsilon^2)(\sigma_1^0-\varepsilon)(\sigma_2^0-\varepsilon)}\Big)' \in \tilde{Z}_3.
\end{aligned}$$

Proof of Lemma 3.21:

1) For $(i,j) \in U \cup I = \{(1,1),(2,2),(3,3),(1,3),(2,3)\}$ formula (3.28) can be written as

$$c_{i,j}\; p^{i,j}_{(\sigma_1^0,\sigma_2^0)}(\varrho^0) = 0 \quad \Rightarrow \quad c_{i,j} = 0 \quad \textit{for } (i,j) \in U \cup I = \Big\{(1,1),(2,2),(3,3),(1,3),(2,3)\Big\}$$

which certainly holds if $p^{i,j}_{(\sigma_1^0,\sigma_2^0)}(\varrho^0) \neq 0$ for any $(i,j) \in U \cup I$.
In the following we use the abbreviation $\partial_i\zeta_j = \partial_i\zeta_j(\boldsymbol{\gamma})|_{\boldsymbol{\gamma}=(\sigma_1^0,\sigma_2^0,\varrho)'}$ for $i,j = 1,\ldots,3$. Since $\frac{1}{4}T_3(\boldsymbol{x})T_3(\boldsymbol{x}) = \frac{1}{4}(x_1x_2)^2 = T_1(\boldsymbol{x})T_2(\boldsymbol{x})$ we have $U = \{(1,3),(2,3)\}$, $U^c = \{(1,2)\}$, $\alpha_3 = \frac{1}{4}$, $H(1,2) = \{3\}$, $H = \{3\}$ and $H^c = \{1,2\}$. For $\varrho \in (-1,1)$ we obtain (with MATHEMATICA)

$$\begin{aligned}
p^{1,1}_{(\sigma_1^0,\sigma_2^0)}(\varrho) &\overset{(3.30)}{=} \Big(\partial_1\zeta_1 + \partial_1\zeta_2 + \partial_1\zeta_3\Big)^2 - \tfrac{3}{2}(\partial_1\zeta_1)(\partial_1\zeta_2) \\
&\overset{(3.56)}{=} \Big(\frac{2}{(\sigma_1^0)^3(1-\varrho^2)} - \frac{\varrho}{(\sigma_1^0)^2\sigma_2^0(1-\varrho^2)}\Big)^2 \\
&= \frac{(\varrho\sigma_1^0 - 2\sigma_2^0)^2}{(\sigma_1^0)^6\,(\sigma_2^0)^2\,(1-\varrho^2)^2}, \\
p^{2,2}_{(\sigma_1^0,\sigma_2^0)}(\varrho) &\overset{(3.30)}{=} \Big(\partial_2\zeta_1 + \partial_2\zeta_2 + \partial_2\zeta_3\Big)^2 - \tfrac{3}{2}(\partial_2\zeta_1)(\partial_2\zeta_2) \\
&\overset{(3.56)}{=} \frac{(2\sigma_1^0 - \varrho\sigma_2^0)^2}{(\sigma_1^0)^2\,(\sigma_2^0)^6\,(1-\varrho^2)^2},
\end{aligned}$$

$$\begin{aligned}
p^{3,3}_{(\sigma_1^0,\sigma_2^0)}(\varrho) &\overset{(3.30)}{=} \Big(\partial_3\zeta_1 + \partial_3\zeta_2 + \partial_3\zeta_3\Big)^2 - \tfrac{3}{2}(\partial_3\zeta_1)(\partial_3\zeta_2) \\
&\overset{(3.56)}{=} \frac{(\sigma_1^0)^2(\sigma_2^0)^2(1+\varrho^4) - 4\sigma_1^0\sigma_2^0\varrho(1+\varrho^2)\Big((\sigma_1^0)^2 + (\sigma_2^0)^2\Big)}{(\sigma_1^0)^4(\sigma_2^0)^4(1-\varrho^2)^4} + \\
&\quad \frac{4\varrho^2\Big((\sigma_1^0)^4 + (\sigma_1^0)^2(\sigma_2^0)^2 + (\sigma_2^0)^4\Big)}{(\sigma_1^0)^4(\sigma_2^0)^4(1-\varrho^2)^4}, \\
p^{1,3}_{(\sigma_1^0,\sigma_2^0)}(\varrho) &\overset{(3.28)}{=} \sum_{(u,v)\in U}\Big((\partial_1\zeta_u)(\partial_3\zeta_v) + (\partial_3\zeta_u)(\partial_1\zeta_v)\Big) + \sum_{u\in\{1,2\}}(\partial_1\zeta_u)(\partial_3\zeta_u) + \\
&\quad \frac{1}{4}\Big((\partial_1\zeta_1)(\partial_3\zeta_2) + (\partial_3\zeta_1)(\partial_1\zeta_2)\Big) + (\partial_1\zeta_3)(\partial_3\zeta_3) \\
&\overset{(3.56)}{=} \frac{2\sigma_1^0(\sigma_2^0)^2 - \varrho^3(\sigma_1^0)^2\sigma_2^0 - 2\varrho\,\sigma_2^0\,\Big((\sigma_1^0)^2 + 2(\sigma_2^0)^2\Big) + 2\varrho^2\Big((\sigma_1^0)^3 + 2\sigma_1^0(\sigma_2^0)^2\Big)}{(\sigma_1^0)^5\,(\sigma_2^0)^3\,(1-\varrho^2)^3}, \\
p^{2,3}_{(\sigma_1^0,\sigma_2^0)}(\varrho) &\overset{(3.28)}{=} \sum_{(u,v)\in U}\Big((\partial_2\zeta_u)(\partial_3\zeta_v) + (\partial_3\zeta_u)(\partial_2\zeta_v)\Big) + \sum_{u\in\{1,2\}}(\partial_2\zeta_u)(\partial_3\zeta_u) + \\
&\quad \frac{1}{4}\Big((\partial_2\zeta_1)(\partial_3\zeta_2) + (\partial_3\zeta_1)(\partial_2\zeta_2)\Big) + (\partial_2\zeta_3)(\partial_3\zeta_3) \\
&\overset{(3.56)}{=} \frac{2(\sigma_1^0)^2\sigma_2^0 - \varrho^3\sigma_1^0(\sigma_2^0)^2 - 2\varrho\,\sigma_1^0\,\Big(2(\sigma_1^0)^2 + (\sigma_2^0)^2\Big) + 2\varrho^2\Big(2(\sigma_1^0)^2\sigma_2^0 + (\sigma_2^0)^3\Big)}{(\sigma_1^0)^3\,(\sigma_2^0)^5\,(1-\varrho^2)^3}.
\end{aligned}$$

Let $\mathcal{Z} = \{z_1, \ldots, z_t\}$ be the set of all solutions of $p^{i,j}_{\sigma_1^0,\sigma_2^0}(\varrho) = 0$, $(i,j) \in I \cup U$. Then $\mathcal{Z}$ is finite since each nominator of $p^{i,j}_{\sigma_1^0,\sigma_2^0}(\varrho)$ is a polynomial in ϱ. $0 \notin \mathcal{Z}$, since for $\varrho^0 = 0$ the Jacobian matrix $\partial_{(\sigma_1,\sigma_2,\varrho)}\boldsymbol{\zeta}(\sigma_1,\sigma_2,\varrho)|_{\boldsymbol{\gamma}^0=(\sigma_1^0,\sigma_2^0,0)}$ (see (3.56)) is equal to a diagonal one and condition 9) of Theorem 3.13 is satisfied due to Lemma 3.14 2)b). Thus for $\sigma_1^0, \sigma_2^0 > 0$ there exist $r_1 \in (-1,0)$, $r_2 \in (0,1)$ such that for any $\varrho^0 \in [r_1, r_2]$ (3.28) is satisfied. Hence, for $\sigma_1^0, \sigma_2^0 > 0$ for any $\varrho^0 \in [r_1, r_2]$ condition 9) of Theorem 3.13 is satisfied.

2) Due to Lemma 3.10 is sufficient to show the existence of A, B with $A \in (-1,0)$, $B \in (0,1)$ and $[A,B] \subset [r_1,r_2]$ such that $2\boldsymbol{\zeta}(\boldsymbol{\gamma}) - \boldsymbol{\zeta}(\boldsymbol{\gamma}^0) \in Z_*$ for any $\boldsymbol{\gamma} = (\sigma_1,\sigma_2,\varrho)' \in \Gamma_{\sigma_1^0} \times \Gamma_{\sigma_2^0} \times [A,B]$. Firstly, we show that for any $\boldsymbol{\gamma} \in \Gamma_{\sigma_1^0} \times \Gamma_{\sigma_2^0} \times [r_1,r_2]$

$$2\Sigma^{-1} - (\Sigma^0)^{-1} = \begin{pmatrix} \frac{2}{(\sigma_1)^2(1-\varrho^2)} & -\frac{2\varrho}{\sigma_1\sigma_2(1-\varrho^2)} \\ -\frac{2\varrho}{\sigma_1\sigma_2(1-\varrho^2)} & \frac{2}{(\sigma_2)^2(1-\varrho^2)} \end{pmatrix} - \begin{pmatrix} \frac{1}{(\sigma_1^0)^2(1-(\varrho^0)^2)} & -\frac{\varrho^0}{\sigma_1^0\sigma_2^0(1-(\varrho^0)^2)} \\ -\frac{\varrho^0}{\sigma_1^0\sigma_2^0(1-(\varrho^0)^2)} & \frac{1}{(\sigma_2^0)^2(1-(\varrho^0)^2)} \end{pmatrix}$$

is positive definite according to the form of Z_*. For $i = 1, 2$ we have

$$\begin{aligned} & \left(2\Sigma^{-1} - (\Sigma^0)^{-1}\right)_{i,i} = \frac{2}{(\sigma_i)^2(1-\varrho^2)} - \frac{1}{(\sigma_i^0)^2(1-(\varrho^0)^2)} > 0 \\ \Leftrightarrow\ & 2(\sigma_i^0)^2(1-(\varrho^0)^2) > (\sigma_i)^2(1-\varrho^2). \end{aligned}$$

According to the definition of $\Gamma_{\sigma_i^0}$ we have $(\sigma_i)^2 \le 2(\sigma_i^0)^2(1-(\varrho^0)^2) - \varepsilon_i$ for $i = 1,2$ which leads to $\left(2\Sigma^{-1}-(\Sigma^0)^{-1}\right)_{1,1}, \left(2\Sigma^{-1}-(\Sigma^0)^{-1}\right)_{2,2} > 0$ for all $(\sigma_1,\sigma_2,\varrho)' \in \Gamma_{\sigma_1^0}\times\Gamma_{\sigma_2^0}\times[r_1,r_2]$. Thus we show that there exist some A, B with $A \in (-1,0)$, $B \in (0,1)$ and $[A,B] \subset [r_1,r_2]$ such that $\det\left(2\Sigma^{-1} - (\Sigma^0)^{-1}\right) > 0$ for any $(\sigma_1,\sigma_2,\varrho)' \in \Gamma_{\sigma_1^0}\times\Gamma_{\sigma_2^0}\times[A,B]$. We have (with MATHEMATICA)

$$\begin{aligned} & \det\left(2\Sigma^{-1} - (\Sigma^0)^{-1}\right) \\ = & \frac{4\sigma_1^0\sigma_2^0\varrho^0\sigma_1\sigma_2\varrho + 2(\sigma_1^0)^2\left(2(\sigma_2^0)^2(1-(\varrho^0)^2) - (\sigma_2)^2\right) - (\sigma_1)^2\left(2(\sigma_2^0)^2 - (\sigma_2)^2(1-\varrho^2)\right)}{(\sigma_1^0)^2(\sigma_2^0)^2(\sigma_1)^2(\sigma_2)^2(1-(\varrho^0)^2)(1-\varrho^2)}. \end{aligned}$$

Since the denominator of the latter expression is strictly positive and the nominator is equal to $-(\varrho - a(\sigma_1,\sigma_2,\varrho^0))(\varrho - b(\sigma_1,\sigma_2,\varrho^0))$ with

$$\begin{aligned} a(\sigma_1,\sigma_2,\varrho^0) &= \frac{2\sigma_1^0\sigma_2^0\varrho^0\sigma_1\sigma_2 - \sigma_1\sigma_2\sqrt{2(\sigma_1^0)^2-(\sigma_1)^2}\sqrt{2(\sigma_2^0)^2-(\sigma_2)^2}}{(\sigma_1)^2(\sigma_2)^2} \quad \text{and} \\ b(\sigma_1,\sigma_2,\varrho^0) &= \frac{2\sigma_1^0\sigma_2^0\varrho^0\sigma_1\sigma_2 + \sigma_1\sigma_2\sqrt{2(\sigma_1^0)^2-(\sigma_1)^2}\sqrt{2(\sigma_2^0)^2-(\sigma_2)^2}}{(\sigma_1)^2(\sigma_2)^2}, \end{aligned}$$

it follows

$$\begin{aligned} & \forall(\sigma_1,\sigma_2,\varrho)' \text{ with } (\sigma_1,\sigma_2)' \in \Gamma_{\sigma_1^0}\times\Gamma_{\sigma_2^0} \text{ and} \\ & \qquad \varrho \in \Gamma_{\varrho^0} \subset [r_1,r_2] \cap \left((a(\sigma_1,\sigma_2,\varrho^0), b(\sigma_1,\sigma_2,\varrho^0)\right) : \ \det\left(2\Sigma^{-1} - (\Sigma^0)^{-1}\right) > 0. \end{aligned}$$

Furthermore, we have

$$\begin{aligned} & \forall(\sigma_1,\sigma_2)' \in \Gamma_{\sigma_1^0}\times\Gamma_{\sigma_2^0} : \ a(\sigma_1,\sigma_2,\varrho^0) < 0 \text{ and } b(\sigma_1,\sigma_2,\varrho^0) > 0 \qquad (3.57) \\ \Leftrightarrow\ & \forall(\sigma_1,\sigma_2)' \in \Gamma_{\sigma_1^0}\times\Gamma_{\sigma_2^0} : \ 2\sigma_1^0\sigma_2^0|\varrho^0| - \sqrt{2(\sigma_1^0)^2-(\sigma_1)^2}\sqrt{2(\sigma_2^0)^2-(\sigma_2)^2} < 0. \end{aligned}$$

Consequently, (3.57) holds for all $(\sigma_1, \sigma_2, \varrho^0)' \in \Gamma_{\sigma_1^0} \times \Gamma_{\sigma_2^0} \times [-\tilde{r} + \varepsilon_3, \tilde{r} - \varepsilon_3]$, $\tilde{r} = \frac{\sqrt{\varepsilon_1}\sqrt{\varepsilon_2}}{2\sigma_1^0\sigma_2^0}$ and $\varepsilon_3 \in (0, \tilde{r})$. Let

$$A = \max\left\{a(\sigma_1, \sigma_2, \varrho^0): \ (\sigma_1, \sigma_2, \varrho^0)' \in \Gamma_{\sigma_1^0} \times \Gamma_{\sigma_2^0} \times \Big([r_1, r_2] \cap [-\tilde{r} + \varepsilon_3, \tilde{r} - \varepsilon_3]\Big)\right\}, \tag{3.58}$$

$$B = \max\left\{b(\sigma_1, \sigma_2, \varrho^0): \ (\sigma_1, \sigma_2, \varrho^0)' \in \Gamma_{\sigma_1^0} \times \Gamma_{\sigma_2^0} \times \Big([r_1, r_2] \cap [-\tilde{r} + \varepsilon_3, \tilde{r} - \varepsilon_3]\Big)\right\} \tag{3.59}$$

then A and B are well-defined since $a(\sigma_1, \sigma_2, \varrho^0)$ and $b(\sigma_1, \sigma_2, \varrho^0)$ are continuous on $\Gamma_{\sigma_1^0} \times \Gamma_{\sigma_2^0} \times \Big([r_1, r_2] \cap [-\tilde{r} + \varepsilon_3, \tilde{r} - \varepsilon_3]\Big)$. Obviously, $[A, B] \subset [r_1, r_2]$ and $A < 0 < B$. Finally, we obtain that $2\Sigma^{-1} - (\Sigma^0)^{-1}$ is positive definite on $\Gamma = \Gamma_{\sigma_1^0} \times \Gamma_{\sigma_2^0} \times [A, B]$.

3) Due to Lemma 3.10 it is sufficient to show that there exists some $\varepsilon \in (0, m)$, $m = \min\{\sigma_1^0, \sigma_2^0, 1 - |\varrho^0|, |\varrho^0| \mathbb{1}_{\varrho^0 \neq 0}\}$, such that $\boldsymbol{u^1}, \boldsymbol{u^2} \in Z_*$ and $3\boldsymbol{u^i} - 2\boldsymbol{\zeta}(\boldsymbol{\gamma^0}) \in Z_*$.
Obviously, one has $u_1^i, u_2^i < 0$ for $i = 1, 2$. Furthermore, we have

$$u_1^1 u_2^1 - (u_3^1)^2 > 0 \quad \Leftrightarrow \quad \frac{(1 - (\varrho^0 + \varepsilon)^2)^2}{(\varrho^0 + \varepsilon)^2} > \frac{(\sigma_1^0 + \varepsilon)^2(\sigma_2^0 + \varepsilon)^2}{(\sigma_1^0 - \varepsilon)^2(\sigma_2^0 - \varepsilon)^2}$$

for $\varepsilon \in (0, m)$. If $\varrho^0 \geq 0$ then the left hand side of the latter relation is strictly monotonic decreasing in $\varepsilon \in (0, m)$ else it is strictly monotonic increasing. Since the right hand side of the latter relation is strictly monotonic increasing in $\varepsilon \in (0, m)$, both sides are differentiable in ε on $(0, m)$ and

$$\lim_{\varepsilon \to 0} \frac{(1 - (\varrho^0 + \varepsilon)^2)^2}{(\varrho^0 + \varepsilon)^2} \quad > \quad 1 = \lim_{\varepsilon \to 0} \frac{(\sigma_1^0 + \varepsilon)^2(\sigma_2^0 + \varepsilon)^2}{(\sigma_1^0 - \varepsilon)^2(\sigma_2^0 - \varepsilon)^2}$$

due to $\varrho^0 \leq B \leq \frac{\sqrt{\varepsilon_1}\sqrt{\varepsilon_2}}{2\sigma_1^0\sigma_2^0} < \frac{1}{2}$, there exists a sufficient small $\varepsilon \in (0, m)$ such that $u_1^1 u_2^1 - (u_3^1)^2 > 0$. (If $\varrho^0 = 0$ we have $\lim_{\varepsilon \to 0} \frac{(1-(\varrho^0+\varepsilon)^2)^2}{(\varrho^0+\varepsilon)^2} = \infty$.)
Correspondingly, we have

$$u_1^2 u_2^2 - (u_3^2)^2 > 0 \quad \Leftrightarrow \quad \frac{(1 - (\varrho^0 - \varepsilon)^2)^2}{(\varrho^0 - \varepsilon)^2} > \frac{(\sigma_1^0 + \varepsilon)^2(\sigma_2^0 + \varepsilon)^2}{(\sigma_1^0 + \varepsilon)^2(\sigma_2^0 + \varepsilon)^2}$$

for $\varepsilon \in (0, m)$. If $\varrho^0 \geq 0$ then the left hand side of the latter relation is strictly monotonic increasing in $\varepsilon \in (0, m)$ else it is strictly monotonic decreasing. Since the left hand side of the latter relation is also differentiable in ε on $(0, m)$ and we have

$$\lim_{\varepsilon \to 0} \frac{(1 - (\varrho^0 - \varepsilon)^2)^2}{(\varrho^0 - \varepsilon)^2} \quad > \quad 1$$

due to $\varrho^0 \leq B \leq \frac{\sqrt{\varepsilon_1}\sqrt{\varepsilon_2}}{2\sigma_1^0\sigma_2^0} < \frac{1}{2}$, there exists some sufficient small $\varepsilon \in (0, m)$ such that $u_1^2 u_2^2 - (u_3^2)^2 > 0$.
Hence, there exists some sufficient small $\varepsilon \in (0, m)$ such that $\boldsymbol{u^1}, \boldsymbol{u^2} \in Z_*$.
We show that $3\boldsymbol{u^i} - 2\boldsymbol{\zeta}(\boldsymbol{\gamma^0}) \in Z_*$, $i = 1, 2$, is satisfied.
From $\varrho^0 < \frac{1}{2}$ it follows

$$3u_j^i - 2\zeta_j(\boldsymbol{\gamma^0}) \quad = \quad -\frac{3}{(\sigma_j^0 + \varepsilon)^2} + \frac{2}{(\sigma_j^0)^2(1 - (\varrho^0)^2)} \quad < \quad 0, \quad i, j = 1, 2$$

for sufficient small $\varepsilon > 0$. Furthermore, we have

$$\begin{aligned}
&\Big(3u_1^1 - 2\zeta_1(\boldsymbol{\gamma^0})\Big)\Big(3u_2^1 - 2\zeta_2(\boldsymbol{\gamma^0})\Big) - \Big(3u_3^1 - 2\zeta_3(\boldsymbol{\gamma^0})\Big)^2 \\
= \;& \Big(-\frac{3}{(\sigma_1^0+\varepsilon)^2} + \frac{2}{(\sigma_1^0)^2(1-(\varrho^0)^2)}\Big)\Big(-\frac{3}{(\sigma_2^0+\varepsilon)^2} + \frac{2}{(\sigma_2^0)^2(1-(\varrho^0)^2)}\Big) \\
&-\Big(\frac{3(\varrho^0+\varepsilon)}{(1-(\varrho^0+\varepsilon)^2)(\sigma_1^0-\varepsilon)(\sigma_2^0-\varepsilon)} - \frac{2\varrho^0}{\sigma_1^0\sigma_2^0(1-(\varrho^0)^2)}\Big)^2 \\
\neq \;& const
\end{aligned}$$

is differentiable in ε on $(0,m)$, $m = \min\{\sigma_1^0, \sigma_2^0, 1-|\varrho^0|, |\varrho^0|\mathbb{1}_{\varrho^0\neq 0}\}$, and

$$\begin{aligned}
&\lim_{\varepsilon\to 0}\Big(3u_1^1 - 2\zeta_1(\boldsymbol{\gamma^0})\Big)\Big(3u_2^1 - 2\zeta_2(\boldsymbol{\gamma^0})\Big) - \Big(3u_3^1 - 2\zeta_3(\boldsymbol{\gamma^0})\Big)^2 \\
= \;& \frac{9(\varrho^0)^4 - 7(\varrho^0)^2 + 1}{(\sigma_1^0)^2(\sigma_2^0)^2(1-(\varrho^0)^2)^2} \\
= \;& \frac{(\varrho^0 - \frac{-1-\sqrt{13}}{6})(\varrho^0 - \frac{1+\sqrt{13}}{6})(\varrho^0 - \frac{-1+\sqrt{13}}{6})(\varrho^0 - \frac{1+\sqrt{13}}{6})}{(\sigma_1^0)^2(\sigma_2^0)^2(1-(\varrho^0)^2)^2}
\end{aligned}$$

is strictly positive for any $(\sigma_1^0, \sigma_2^0, \varrho^0)' \in \Gamma$ since $\Gamma_\varrho \subset \Big[\frac{1}{6}(1-\sqrt{13}) + \varepsilon_3, \frac{1}{6}(-1+\sqrt{13}) - \varepsilon_3\Big]$, $\varepsilon_3 \in (0, \frac{1}{6}(-1+\sqrt{13})$. Thus there exists a sufficient small $\varepsilon > 0$ such that $3\boldsymbol{u^1} - 2\boldsymbol{\zeta}(\boldsymbol{\gamma^0}) \in Z_*$. Correspondingly, we have

$$\begin{aligned}
3\boldsymbol{u^2} - 2\boldsymbol{\zeta}(\boldsymbol{\gamma^0}) \;=\;& \Big(-\frac{3}{(\sigma_1^0+\varepsilon)^2} + \frac{2}{(\sigma_1^0)^2(1-(\varrho^0)^2)}\Big)\Big(-\frac{3}{(\sigma_2^0+\varepsilon)^2} + \frac{2}{(\sigma_2^0)^2(1-(\varrho^0)^2)}\Big) \\
&-\Big(\frac{3(\varrho^0-\varepsilon)}{(1-(\varrho^0-\varepsilon)^2)(\sigma_1^0+\varepsilon)(\sigma_2^0+\varepsilon)} - \frac{2\varrho^0}{\sigma_1^0\sigma_2^0(1-(\varrho^0)^2)}\Big)^2 \\
\neq \;& const
\end{aligned}$$

is differentiable in ε on $(0,m)$, $m = \min\{\sigma_1^0, \sigma_2^0, 1-|\varrho^0|, |\varrho^0|\mathbb{1}_{\varrho^0\neq 0}\}$, and

$$\begin{aligned}
&\lim_{\varepsilon\to 0}\Big(3u_1^2 - 2\zeta_1(\boldsymbol{\gamma^0})\Big)\Big(3u_2^2 - 2\zeta_2(\boldsymbol{\gamma^0})\Big) - \Big(3u_3^2 - 2\zeta_3(\boldsymbol{\gamma^0})\Big)^2 \\
= \;& \lim_{\varepsilon\to 0}\Big(3u_1^1 - 2\zeta_1(\boldsymbol{\gamma^0})\Big)\Big(3u_2^1 - 2\zeta_2(\boldsymbol{\gamma^0})\Big) - \Big(3u_3^1 - 2\zeta_3(\boldsymbol{\gamma^0})\Big)^2.
\end{aligned}$$

In the same manner as above there exists a sufficient small $\varepsilon > 0$ such that $3\boldsymbol{u^2} - 2\boldsymbol{\zeta}(\boldsymbol{\gamma^0}) \in Z_*$. □

$N(\boldsymbol{\mu}, \Sigma)$	unknown	known	Γ	test	summary
univariate or multivariate	$\boldsymbol{\mu}$	Σ^0	$\Gamma \subset I\!R^k$ aribitrary such that (A1), (A2) hold.	Q against p populations.	Example 3.15: [DCG99] is applicable, generalization of Keribin (2000), Theo.4.1.
univariate or multivariate	$\Sigma = diag(\sigma_1, \ldots, \sigma_k)$	$\boldsymbol{\mu^0}$	$\Gamma = \times_{i=1}^{k} \left[\sqrt{\varepsilon_i}, \sqrt{2(\sigma_i^{0,1})^2 - \varepsilon_i}\right]$, $\boldsymbol{\gamma^{0,1}} = (\sigma_1^{0,1}, \ldots, \sigma_k^{0,1})'$. If $q > 1$ then $\boldsymbol{\gamma^{0,l}} \in int(\Gamma) \cap \times_{i=1}^{k} \left(\sqrt{\varepsilon_i}, \sqrt{\frac{3}{2}}\, \sigma_i^{0,1}\right)$ for $l = 2, \ldots, q$.	Q against p populations.	Example 3.17: [DCG99] is applicable, generalization of Keribin (2000), Prop. 4.2.
univariate	μ, σ	–	$\left[\mu^0 - l, \mu^0 + u\right] \times \left[\sqrt{\varepsilon}, \sqrt{2(\sigma^0)^2 - \varepsilon}\right]$, $\tilde{\Gamma} =$ $\left\{(\boldsymbol{\gamma^1}, \boldsymbol{\gamma^2}) \in \Gamma \times \Gamma : \Vert\boldsymbol{\gamma^1} - \boldsymbol{\gamma^0}\Vert \leq \Vert\boldsymbol{\gamma^2} - \boldsymbol{\gamma^0}\Vert,\ \lvert\mu^2 - \mu^0\rvert \leq \lvert\sigma^2 - \sigma^0\rvert\right\}$.	$g \in \mathcal{G}_1$ against $g \in \mathcal{G}_{2,M} \backslash \mathcal{G}_1$.	Example 3.18: generalization of [DCG99], support of Garel (2001), Conjecture 5.3.
bivariate	$\Sigma = \begin{pmatrix} (\sigma_1)^2 & \varrho\, \sigma_1 \sigma_2 \\ \varrho\, \sigma_1 \sigma_2 & (\sigma_2)^2 \end{pmatrix}$	$\boldsymbol{\mu^0}$	$\Gamma = \Gamma_{\sigma_1^0} \times \Gamma_{\sigma_2^0} \times \Gamma_\varrho$, Γ_ϱ as in (3.54) and for $i = 1, 2$ $\Gamma_{\sigma_i^0} = \left[\sqrt{\varepsilon_i}, \sqrt{2(\sigma_i^0)^2(1 - (\varrho^0)^2) - \varepsilon_i}\right]$. If $\varrho^0 = 0$ then $\Gamma_\varrho = [-r, r]$ with $r =$ $\min\left\{\left(\frac{\varepsilon_1}{2(\sigma_1^0)^2 - \varepsilon_1} \frac{\varepsilon_2}{2(\sigma_2^0)^2 - \varepsilon_2} - \varepsilon_4\right)^{\frac{1}{2}}, 1 - \varepsilon_5\right\}$ is also possible.	One against two populations.	Example 3.19: generalization of [DCG99], best to our knowledge we are the first to give sufficient conditions for deriving an asymptotic.

3.3.2. Application to Discrete Models

In this subsection we apply our main Theorem 3.13 and thus the theory from [DCG99] to several examples of one-parametric discrete exponential families.
Liu and Shao (2003) [LS03], Theorem 3.2, prove that the Donsker class property is automatically fulfilled for discrete models when testing $g = g^0$ against $g \in \mathcal{G}_p \backslash \{g^0\}$ and give a representation of a corresponding Gaussian process. Roughly speaking, they derive the functional index set of a Gaussian process essentially by a set of limits of diverse sequences of a generalized score function with respect to a corresponding (component and mixing) parameter (see [LS03], (3.6)). However, [LS03] do not specify the corresponding parameter space. For testing Q against p populations in general finite mixture models [LS03] say that the index set "can be derived using functions equivalent to the generalized score function. One way to find the equivalent functions is to use Taylor expansions of the likelihood ratios... The local conic parameterization approach of Dacunha-Castelle and Gassiat (1997, 1999) is very useful in identifying" the index set ([LS03], p. 826). Unfortunately, they mention neither that the p-weakly identifiability assumption (ID) used in [DCG99] does not hold for some discrete mixtures nor that the parameter space may depend on the true component parameter. Especially in Example 3.23 we refer to Teicher (1963) who shows that the p-identifiability for mixtures of binomial densities holds only for certain parameters. In Example 3.24 it will be seen that for negative binomial densities the parameter space depends on the true component parameters.

In all our examples we will refer to Teicher (1961) and Teicher (1963), respectively, using his statements about identifiability. According to our notes after Remark 2.1, p. 9, identifiability according to Teicher implies p-weakly identifiability (ID), since Yakowitz and Spragins (1968) make an easy modification of Teicher's definition of the identifiability to include multidimensional cumulative distribution functions.

In our first example we consider Poisson distributions. Keribin (2000), Theorem 4.3, has also shown the applicability of the theory given in [DCG99] for mixtures of Poisson distributions. While she derives a majorant function $\bar{f}$ for $\mathcal{F} = \{f_\gamma : \gamma \in \Gamma\}$ and a minorant function $\underline{g}^0$ for the true density function g^0 to check the applicability of (P0)c) (see here, p. 68), we check it with the aid of our elementary Proposition 3.11 using the fact $Z_* = I\!R$. Consequently, we only need to investigate the supremum of the scalar product of $\langle \zeta(\gamma), T(x) \rangle$ in a neighbourhood $\overline{U}_\varepsilon(\gamma^{0,l}) = \{\gamma \in \Gamma : |\gamma - \gamma^{0,l}| \leq \varepsilon\}$ for $l = 1, \ldots, q$, instead of analyzing the specific geometric structure of the densities of $\mathcal{F}$ as Keribin does.

Notice that condition 6) of Theorem 3.13 is always satisfied for one-parametric exponential families. Furthermore, property 3) and 7) are equivalent according to Lemma 3.14 1)a).

Example 3.22 (Poisson distributions)
Let $\mathcal{P} = \{P(\gamma) : \gamma \in \Gamma\}$, $\Gamma \subset (0, \infty)$, be a one-parametric family of Poisson distributions. Let the σ-finite measure μ on $(\mathfrak{X}, \mathcal{X}) = (I\!R, I\!B)$ be given by $\nu(B) = \sum_{x=0}^{\infty} \mathbb{1}_B(x)$, $B \in \mathcal{X}$. The ν-densities with respect to $\mathcal{P}$ in exponential representation (3.1) are given by

$$f_\gamma(x) \quad = \quad \exp(-\gamma)\frac{\gamma^x}{x!} \quad = \quad \exp(-\gamma)\, \exp\Big\{ \log(\gamma)x \Big\} \frac{1}{x!} \quad \nu - a.s.,$$

where $\zeta(\gamma) = \log(\gamma)$, $T(x) = x$, $A(\gamma) = \exp(-\gamma) = \mathcal{C}(\zeta(\gamma))$ and $r(x) = \frac{1}{x!}$. Clearly, one has

$$Z_* \stackrel{(3.4)}{=} \Big\{ \zeta \in I\!R : 0 < \sum_{x=0}^{\infty} \frac{\exp(\zeta x)}{x!} = \exp(\exp(\zeta)) < \infty \Big\} \quad = \quad I\!R.$$

Let $\Gamma \subset (0, \infty)$ *be an arbitrary set such that* (A1) *and* (A2) *hold. Obviously,* $\zeta : \Gamma \to \zeta(\Gamma)$ *is bijective. It is well known that* $\mathcal{P}$ *is additively closed, i.e. closed with respect to convolution. According to Teicher* (1961)*, p.* 245*, mixtures of additively closed families are identifiable and hence,* (ID) *holds.*
Condition 1) *(of Theorem* 3.13*) follows from*

$$\sum_{l=1}^{q}\left(\pi_l^0 \int |\log(x!)| f_{\gamma^{0,l}}(x)\,\nu(dx)\right) \quad = \quad \sum_{l=1}^{q}\left(\pi_l^0 \exp(-\gamma^{0,l}) \sum_{x=0}^{\infty}\sum_{i=1}^{x} \frac{\log(i)}{i}(\gamma^{0,l})^x\right) \quad < \quad \infty$$

since the power series with center in 0 and coefficient $a_x = \sum_{i=1}^{x} \frac{\log(i)}{i}$ *has a radius of convergence* $r = \limsup_{x\to\infty} |a_x|^{-\frac{1}{x}} = \infty$.

2) *holds obviously for the logarithm function* $\zeta : \Gamma \to \boldsymbol{\zeta}(\Gamma)$.
3) *holds since for any* $\gamma \in \Gamma \subset (0, \infty)$ *the Jacobian matrix is equal to* $\partial_\gamma \log(\gamma) = \gamma^{-1} \neq 0$.
4) *holds since* $Z_* = \mathbb{R}$.
5) *follows from Proposition* 3.11 *for* $G_1 = [0, \infty)$ *and* $u^1 = \sup_{\gamma\in\Gamma} \gamma$ *since for any* $l \in \{1, \ldots, q\}$

$\sup_{\gamma\in\overline{U}_\varepsilon(\gamma^{0,l})} \langle \log(\gamma), x\rangle \leq \langle \log(u^1), x\rangle$ *on* G_1 *and* $Z_* = \mathbb{R}$ *hold.*

6) *follows from dim*$(\Gamma) = 1$.
7) *is equivalent to* 3) *according to Lemma* 3.14 1)a)*.*
Consequently, properties (P0) *and* (P1) *hold.*

Hence, the theory from [DCG99] is applicable to mixtures of Poisson distributions with arbitrary Γ such that (A1) as well as (A2) hold (since $Z_* = \mathbb{R}$).

Our second example is on mixtures of binomial distributions. Notice that the identifiability condition holds only for certain values of parameter. We show that the theory in [DCG99] is applicable for testing q against p populations of $B(m, \gamma)$-mixtures with $m \geq 2p - 1$ and arbitrary Γ such that (A1) as well as (A2) hold since $Z_* = \mathbb{R}$.

Example 3.23 (binomial distributions)
For $m \in \mathbb{N}$ *let* $\mathcal{P} = \{B(m, \gamma) : \gamma \in \Gamma\}$, $\Gamma \subset (0, 1)$*, be a one-parametric family of binomial distributions with respect to the* σ*-finite measure* $\nu(B) = \sum_{x=0}^{m} \mathbb{1}_B(x)$, $B \in \mathcal{X}$, $(\mathfrak{X}, \mathcal{X}) = (\mathbb{R}, \mathbb{B})$.
The ν*-densities with respect to* $\mathcal{P}$ *in exponential representation* (3.1) *are given by*

$$f_\gamma(x) \quad = \quad \binom{m}{x}\gamma^x(1-\gamma)^{m-x} \quad = \quad \binom{m}{x}(1-\gamma)^m \exp\left\{\log\left(\frac{\gamma}{1-\gamma}\right)x\right\} \quad \nu - a.s.,$$

where $\zeta(\gamma) = \log\left(\frac{\gamma}{1-\gamma}\right)$, $T(x) = x$, $A(\gamma) = (1-\gamma)^m = \mathcal{C}(\zeta(\gamma))$ *and* $r(x) = \binom{m}{x}$*. Clearly, one has*

$$Z_* \overset{(3.4)}{=} \left\{\zeta \in \mathbb{R} : 0 < \sum_{x=0}^{m}\binom{m}{x}\exp(\zeta x) < \infty\right\} \quad = \quad \mathbb{R}.$$

Let $\Gamma \subset (0, 1)$ *be an arbitrary set such that* (A1) *and* (A2) *hold. Obviously,* $\zeta : \Gamma \to \zeta(\Gamma)$ *is bijective. According to Teicher* (1963)*, Proposition* 4 (i)*,* $\mathcal{P}$ *is p-identifiable iff* $m \geq 2p - 1$ *(see also Titterington et. al.* (1992)*, Example* 3.1.6*). (We are testing Q populations against p populations.)*
Condition 1) *(of Theorem* 3.13*) follows since* ν *has a finite support.*

2) *is obviously satisfied for* $\zeta : \Gamma \to \zeta(\Gamma)$.
3) *holds since for any* $\gamma \in \Gamma \subset (0,1)$ *the Jacobian matrix is equal to* $\partial_\gamma \log(\frac{\gamma}{1-\gamma}) = (\gamma - \gamma^2)^{-1} \neq 0$.
4) *holds since* $Z_* = I\!R$.
5) *follows from Proposition* 3.11 *for* $G_1 = [0, m+1)$ *and* $u^1 = \sup_{\gamma\in\Gamma} \log\left(\frac{\gamma}{1-\gamma}\right)$ *since for any* $l \in \{1,\dots,q\}$ $\sup_{\gamma \in \overline{U}_\varepsilon(\gamma^{0,l})} \left\langle \log\left(\frac{\gamma}{1-\gamma}\right), x\right\rangle \le \langle u^1, x\rangle$ *on* G_1 *and* $Z_* = I\!R$ *hold.*
6) *follows from* $dim(\Gamma) = 1$.
7) *is equivalent to* 3) *according to Lemma* 3.14 1)a).
Consequently, properties (P0) *and* (P1) *hold.*

Our third example of discrete models (for applying the theory from [DCG99]) is on negative binomial distributions. It shows that the component parameter space Γ depends on the true component parameters. This results from $Z_* = (-\infty, 0) \neq I\!R$ since one has to ensure that $\frac{f_\gamma}{g^0} \in L_2(g^0\nu)$ holds for any $\gamma \in \Gamma$ as assumed in (P1). Though the theory given in [DCG99] is applicable to mixtures of negative binomial distributions, it is applicable to suitable Γ only. As mentioned in the beginning of this subsection, Liu and Shao (2003) do not point out that one has to specify the parameter space somehow when they say that the Donsker class property is automatically fulfilled for discrete models when testing $g = g^0$ against $g \in \mathcal{G}_p \backslash \{g^0\}$, (see Liu and Shao (2003), Theorem 3.2). One possibility of constructing Γ is to choose, firstly, the true components $\boldsymbol{\gamma^{0,l}}$, $l = 1,\dots,q$, and then to conform Γ to the requirements of (A1), (A2) as well as of condition 4) and 5) of Theorem 3.13. We construct Γ conform to the assumptions of Proposition 3.11 to ensure the requirements of condition 5) of Theorem 3.13.

Example 3.24 (negative binomial distributions)
For $m > 0$ *let* $\mathcal{P} = \{\mathrm{NB}(m,\gamma) : \gamma \in \Gamma\}$, $\Gamma \subset (0,1)$, *be a one-parametric exponential family of negative binomial distributions with respect to the* σ*-finite measure* $\nu(B) = \sum_{x=0}^{\infty} 1\!\!1_B(x)$, $B \in \mathcal{X}$, $(\mathfrak{X}, \mathcal{X}) = (I\!R, I\!B)$. *The* ν*-densities with respect to* $\mathcal{P}$ *in exponential representation* (3.1) *are given by*

$$f_\gamma(x) \quad = \quad \gamma^m \exp\Big\{ \log(1-\gamma)x \Big\} \binom{m+x-1}{m-1} \qquad \nu - a.s.,$$

where $\zeta(\gamma) = \log(1-\gamma)$, $T(x) = x$, $A(\gamma) = \gamma^m = \mathcal{C}(\zeta(\gamma))$ *and* $r(x) = \binom{m+x-1}{m-1}$. *Clearly, one has*

$$Z_* \overset{(3.4),(3.7)}{=} \Big\{ \log(1-\gamma) \in I\!R : 0 < \mathcal{C}(\log(1-\gamma))^{-1} = \gamma^{-m} < \infty \Big\} \quad = \quad \zeta(0,1) \quad = \quad (-\infty, 0).$$

For $\delta \in (0,1)$ *let* $\gamma^{0,1},\dots,\gamma^{0,q} \in (0, 1-\delta]$ *be distinct points and let* $\gamma^0_{min} = \min\{\gamma^{0,1},\dots,\gamma^{0,q}\}$. *We define the parameter space by*

$$\Gamma \quad = \quad \Big[1 - (1-\gamma^0_{min})^{\frac{1}{2}}, 1-\delta\Big].$$

Simple computations show that $\Gamma \neq \emptyset$ *since* $\delta \in (0,1)$ *and* $\gamma^0_{min} \in (0, 1-\delta]$. *Obviously,* $\zeta : \Gamma \to \zeta(\Gamma)$ *is bijective on* $\Gamma \subset (0,1)$ *and* (A1) *as well as* (A2) *are satisfied.*

Conditions 1), 2), 3) *(of Theorem* 3.13*) are satisfied due to the form of* $\zeta : \Gamma \to \zeta(\Gamma)$.
4) *Due to Lemma* 3.10 *it is sufficient to show that there exists some* $l \in \{1,\dots,q\}$ *such that* $2\zeta(\gamma) - \zeta(\gamma^{0,l}) \in (-\infty, 0) = Z_*$ *for any* $\gamma \in \Gamma\backslash\{\gamma^{0,l}\}$.

Thus it follows 4) *from* $2\zeta(\gamma) - \zeta(\gamma^0_{min}) = \log\left(\frac{(1-\gamma)^2}{1-\gamma^0_{min}}\right) \in (-\infty, 0) = Z_*$, $\gamma \in \Gamma\backslash\{\gamma^0_{min}\}$.

5) *results from Proposition* 3.11 *for* $G_1 = [0,\infty)$ *and* $u^1 = \log(1-\gamma^0_{min}+\varepsilon)$ *since for any* $l = 1,\dots,q$ *one has*

$$\sup_{\gamma\in\overline{U}_\varepsilon(\gamma^{0,l})} \log(1-\gamma)x \quad\leq\quad u^1 x, \quad x \in G_1 = [0,\infty)$$

and since $\gamma^0_{min} \in (0, 1-\delta]$ *is a fixed parameter with* $(1-\gamma^0_{min})^3 < (1-\gamma^0_{min})^2$ *which implies that there exists some* $\delta > 0$ *such that*

$$\begin{aligned}(1-\gamma^0_{min}+\delta)^3 < (1-\gamma^0_{min})^2 \quad&\Leftrightarrow\quad 3\Big(\log(1-\gamma^0_{min}+\delta)\Big) - 2\Big(\log(1-\gamma^0_{min})\Big) < 0\\ &\Leftrightarrow\quad 3u^1 - 2\zeta(\gamma^0_{min}) \in Z_*.\end{aligned}$$

Thus $u^1 \in \tilde{Z}_3$ *due to Lemma* 3.10.
Consequently, properties (P0) *and* (P1) *hold.*

Please note, $\{\mathrm{NB}(m,\gamma) : (m,\gamma)' \in (0,\infty)\times(0,1)\}$ is not an exponential family (see Brown (1986), Exercise 1.12.1., p. 27). Thus one has to construct Γ somehow to apply Theorem 3.2 by Liu and Shao (2003) on this kind of mixtures.

3.3.3. Application to Gamma Families

In this subsection we make use of Theorem 3.13 to verify the applicability of the asymptotic theory from [DCG99] to gamma distributions. Thus we check the corresponding conditions of Theorem 3.13 as well as (A1) and (A2). The identifiability condition (ID) always holds since all finite mixtures of gamma densities are identifiable according to Teicher (1963), Proposition 2. According to our notes below Remark 2.1, p. 9, identifiability according to Teicher implies p-weakly identifiability (ID), since Yakowitz and Spragins (1968) make an easy modification of Teicher's definition of the identifiability to include multidimensional cumulative distribution functions.

Firstly, we consider mixtures of general gamma distributions. Unfortunately, $Z_* = (0,\infty)\times(-\infty,0) \neq I\!R^2$ and as we have to ensure (among other things) that $\frac{f_{\boldsymbol\gamma}}{g^0} \in L_2(g^0\lambda)$ holds for any $\boldsymbol\gamma \in \Gamma$ as assumed in (P1), we construct a specific component parameter space Γ similar to Example 3.17. If $q > 1$ we cannot choose the remaining true component parameters $\boldsymbol{\gamma^{0,l}}$, $l = 2,\dots,q$ as arbitrary elements of Γ (since $Z_* \neq I\!R^2$) since we want to ensure (P0)c) with the aid of Proposition 3.11.

Secondly, we consider mixtures of one-parametric gamma distributions. As $Z_* = (-\infty,0) \neq I\!R$ we construct a specific component parameter space Γ once again. Thereby, we firstly choose the true parameters and afterwards we construct the parameter space in such a way that the sufficient conditions of Theorem 3.13 hold.

Example 3.25 (gamma distributions)
Let $\mathcal{P} = \{G(\boldsymbol\gamma) : \boldsymbol\gamma = (a,b)' \in \Gamma\}$, $\Gamma \subset (0,\infty)\times(0,\infty)$, *be a minimal* 2*-parametric exponential family of gamma distributions on* $(\mathfrak{X},\mathcal{X}) = (I\!R_+, I\!B_+)$. *Let the gamma function be given by* $g(a) = \int_0^\infty t^{a-1}\exp(-t)\,dt$, $a > 0$. *Then the corresponding Lebesgue* λ*-densities have the form*

$$f_{\boldsymbol\gamma}(x) \quad=\quad (g(a)\,b^a)^{-1}x^{a-1}\exp(-x/b) \quad \lambda - a.s.,$$

and $\boldsymbol{\zeta}(a,b) = (a, -b^{-1})'$, $\boldsymbol{T}(x) = (\log(x), x)'$ *and* $A(a,b) = (g(a)b^a)^{-1} = \mathcal{C}(a, -b^{-1})$ *and* $r(x) = x^{-1}$ *(see e.g. Witting* (1985), *p.* 148*). Obvious is* $\boldsymbol{\zeta} : \Gamma \to \boldsymbol{\zeta}(\Gamma)$ *bijective for any* $\Gamma \subset (0,\infty)^2$. *Clearly,*

$$Z_* \overset{(3.4),(3.7)}{=} \Big\{(a, -b^{-1})' \in I\!R^2 : 0 < \mathcal{C}(a, -b^{-1})^{-1} = g(a)b^a < \infty\Big\} \quad = \quad (0,\infty) \times (-\infty, 0).$$

For $\boldsymbol{\gamma^{0,1}} = (a^{0,1}, b^{0,1})' \in (0,\infty)^2$, $a^{0,1}/2 > \varepsilon_a > 0$, $\varepsilon_b > 0$ *and* $U \geq a^{0,1}$ *we define the parameter space by*

$$\Gamma \quad = \quad \Big[\tfrac{1}{2}a^{0,1} + \varepsilon_a, U\Big] \times \Big[\varepsilon_b, 2b^{0,1} - \varepsilon_b\Big].$$

Let the remaining distinct true parameters

$$\boldsymbol{\gamma^{0,2}}, \ldots, \boldsymbol{\gamma^{0,q}} \quad \in \quad \Gamma \cap \Big[\tfrac{2}{3}a^{0,1} + \varepsilon_a, U\Big] \times \Big[\varepsilon_b, \tfrac{3}{2}b^{0,1} - \varepsilon_b\Big] \tag{3.60}$$

which leads to (A1) *and* (A2) *and it will be shown below that this construction also leads to condition* 4) *and* 5) *of Theorem* 3.13.

Condition 1) *(of Theorem* 3.13*) follows from* $|\log(r(x))| = |\log(x)|$, $x \in I\!R_+$, *and since* $T_1(x) = \log(x)$ *possesses moments of arbitrary order with respect to* $\tilde{p}_{\boldsymbol{\zeta}^{0,\iota}}\lambda$ *due to Proposition* 3.6.

2) *is obvious for* $\boldsymbol{\zeta} : \Gamma \to \zeta(\Gamma)$.

3) *holds since the Jacobian matrix* $\partial_{(a,b)}\boldsymbol{\zeta}(a,b) = diag(1, b^{-2})$, $b > 0$.

4) *Due to Lemma* 3.10 *it is sufficient to show that there exists some* $l \in \{1, \ldots, q\}$ *such that* $2\boldsymbol{\zeta}(\boldsymbol{\gamma}) - \boldsymbol{\zeta}(\boldsymbol{\gamma^{0,l}}) \in (0,\infty) \times (-\infty, 0)$ *for any* $\boldsymbol{\gamma} \in \Gamma \backslash \{\boldsymbol{\gamma^{0,l}}\}$.
Thus it follows 4) *from* $2a - a^{0,1} \geq 2(a^{0,1}/2 + \varepsilon_a) - a^{0,1} > 0$ *and*
$(-2b^{0,1} + b)/(b\, b^{0,1}) \leq (-2b^{0,1} + (2b^{0,1} - \varepsilon_b))/(b\, b^{0,1}) < 0$, $\boldsymbol{\gamma} = (a,b)' \in \Gamma\backslash\{\boldsymbol{\gamma^{0,1}}\}$.

5) *results from the following construction such that Proposition* 3.11 *is applicable:*
Let $a^m = \min_{l=1,\ldots,q} a^{0,l}$, $a^M = \max_{l=1,\ldots,q} a^{0,l}$ *and* $b^M = \max_{l=1,\ldots,q} b^{0,l}$. *For any* $l = 1, \ldots, q$ *we obtain*

$$\sup_{\gamma \in \overline{U}_\varepsilon(\boldsymbol{\gamma^{0,\iota}})} \Big(a\log(x) - \frac{x}{b}\Big) \leq \begin{cases} (a^m - \varepsilon)\log(x) - \dfrac{x}{b^M + \varepsilon} & , \quad x \in G_1 = (0,1) \\[2ex] (a^M + \varepsilon)\log(x) - \dfrac{x}{b^M + \varepsilon} & , \quad x \in G_2 = [1,\infty). \end{cases}$$

Consequently, we define $\boldsymbol{u^1} = \Big(a^m - \varepsilon, -(b^M + \varepsilon)^{-1}\Big)'$ *and* $\boldsymbol{u^2} = \Big(a^M + \varepsilon, -(b^M + \varepsilon)^{-1}\Big)'$. *It follows* $\boldsymbol{u^1}, \boldsymbol{u^2} \in Z_*$ *and* $3\boldsymbol{u^1} - 2\boldsymbol{\zeta}(\boldsymbol{\gamma^{0,1}}) \in Z_*$ *for* $\varepsilon < \varepsilon_a, \varepsilon_b$ *according to*

$$\begin{array}{rcccccc} 3u_1^1 - 2\boldsymbol{\zeta}(\gamma_1^{0,1}) & = & 3(a^m - \varepsilon) - 2a^{0,1} & \overset{(3.60)}{\geq} & 3\Big(\frac{2}{3}a^{0,1} + \varepsilon_a - \varepsilon\Big) - 2a^{0,1} & > & 0, \\ 3u_2^1 - 2\boldsymbol{\zeta}(\gamma_2^{0,1}) & = & \frac{-3b^{0,1} + 2(b^M + \varepsilon)}{(b^M + \varepsilon)b^{0,1}} & \overset{(3.60)}{\leq} & \frac{-3b^{0,1} + 2(3/2\, b^{0,1} - \varepsilon_b + \varepsilon)}{(b^M + \varepsilon)b^{0,1}} & < & 0. \end{array}$$

Furthermore, it follows $3\boldsymbol{u^2} - 2\boldsymbol{\zeta}(\boldsymbol{\gamma^{0,1}}) \in Z_*$ *from* $3u_1^2 - 2\boldsymbol{\zeta}(\gamma_1^{0,1}) = 3(a^M + \varepsilon) - 2a^{0,1} > 0$ *and* $u_2^2 = u_2^1$. *Thus* $\boldsymbol{u^1}, \boldsymbol{u^2} \in \tilde{Z}_3$ *due to Lemma* 3.10.

6) *holds since the functions* $\log(x)$, x, $(\log(x))^2$, $\log(x)x$ *and* x^2 *are obviously* λ*-affinely independent.*

7) *is according to Lemma* 3.14 2)a) *equivalent to condition* 3).

Consequently, properties (P0) *and* (P1) *hold.*

Example 3.26 (exponential distributions)
Let $\mathcal{P} = \{Exp(\gamma) : \gamma \in \Gamma\}$, $\Gamma \subset (0, \infty)$, be a one-parametric family of exponential distributions on $(\mathfrak{X}, \mathcal{X}) = (I\!R_+, I\!B_+)$. It is well known that the Lebesgue λ-densities in exponential representation (3.1) *are given by*

$$f_\gamma(x) = \gamma \exp(-\gamma x) \quad \lambda - a.s.,$$

where $\zeta(\gamma) = -\gamma$, $T(x) = x$ and $A(\gamma) = \gamma = \mathcal{C}(\zeta(\gamma))$ and $r(x) = 1$. Obvious is $\zeta : \Gamma \to \zeta(\Gamma)$ bijective for any $\Gamma \subset (0, \infty)$. Clearly,

$$Z_* \overset{(3.4)}{=} \left\{\gamma \in I\!R : 0 < \int_{I\!R_+} \exp(\gamma x) < \infty\right\} \quad = \quad (-\infty, 0).$$

We construct a parameter space Γ such that condition 4) *and* 5) *of Theorem* 3.13 *is satisfied as well as* (A1) *and* (A2)*:*
Let $\gamma^{0,1}, \ldots, \gamma^{0,q} \in (0, \infty)$ be distinct points and $\gamma^0_{min} = \min\{\gamma^{0,1}, \ldots, \gamma^{0,q}\}$. For $\varepsilon \in (0, \frac{\gamma^0_{min}}{2})$ and $u \geq \max\{\gamma^{0,1}, \ldots, \gamma^{0,q}\}$ we define the parameter space by

$$\Gamma \quad = \quad \left[\frac{\gamma^0_{min}}{2} + \varepsilon, u\right].$$

Simple computations show that $\Gamma \neq \emptyset$ since $\varepsilon \in (0, \frac{\gamma^0_{min}}{2})$ and $u \geq \gamma^0_{min}$. Clearly, (A1) *and* (A2) *hold.*

Conditions 1) *(of Theorem* 3.13*) follows from $|\log(r)| = 0$.*
2),3) *are satisfied by the form of $\zeta : \Gamma \to \zeta(\Gamma)$.*
4) *Due to Lemma* 3.10 *it is sufficient to show that there exists some $l \in \{1, \ldots, q\}$ such that $2\zeta(\gamma) - \zeta(\gamma^{0,l}) \in (-\infty, 0) = Z_*$ for any $\gamma \in \Gamma \backslash \{\gamma^{0,l}\}$.*
Thus it follows 4) *from $2\zeta(\gamma) - \zeta(\gamma^0_{min}) = -2\gamma + \gamma^0_{min} \leq -2\varepsilon \in (-\infty, 0) = Z_*$, $\gamma \in \Gamma \backslash \{\gamma^{min}\}$.*
5) *results from Proposition* 3.11 *for $G_1 = [0, \infty)$ and $u^1 = -\gamma^0_{min} + \varepsilon$ since for any $l = 1, \ldots, q$ we obtain $\sup_{\gamma \in \overline{U}_\delta(\gamma^{0,l})} \langle -\gamma, x\rangle \leq u^1 x$ on G_1 and $3u^1 - 2\zeta(\gamma^0_{min}) = 3(-\gamma^0_{min} + \varepsilon) + 2\gamma^0_{min} < 0$.*
Thus $u^1 \in \tilde{Z}_3$ due to Lemma 3.10.
6) *follows from $dim(\Gamma) = 1$.*
7) *is equivalent to* 3) *according to Lemma* 3.14 1)a).
Consequently, properties (P0) *and* (P1) *hold.*

4. Calculation of the Asymptotic Distribution of the LRT Statistic for Testing Homogeneity Against Two Populations in Normal Mixtures

We calculate the asymptotic distribution of the log-LRT statistic for a special univariate mixture model of normal distributions. For this purpose, we simulate the corresponding Gaussian process by its finite-dimensional marginal distributions. When we generate the appropriate covariance matrix we observe a surprising effect on the size of its eigenvalues. Firstly, it strikes us that only a few eigenvalues have large values and most of them are approximately 0, but we also observe negative eigenvalues of approximately 0. Of course, these are numerical errors. This could lead to the assumption that the asymptotic is similar to a low dimensional normal distribution. As a consequence we introduce a modified spectral factorization on the resulting covariance matrix which leads to considerably short computation times in comparison to the commonly used factorization. Furthermore, it will be seen that there is a negligible difference between the quantiles of both these versions. We also simulate the exact quantiles of the log-LRT statistic. An investigation of the power shows that in a variety of cases the power of log-LRT based on the asymptotic quantiles has nearly the same power as its exact version. We also observe that the computing times for calculating the asymptotic quantiles are surprisingly short in comparison to the times needed for obtaining the exact quantiles.

Let $\mathcal{P} = \{N(\mu, 1) : \mu \in \Gamma\}$, $\Gamma \subset \mathbb{R}$ compact, be a family of univariate normal distributions with known standard deviation $\sigma^0 = 1$ and unknown mean μ. The set of the corresponding Lebesgue λ-densities is $\mathcal{F} = \{f_\mu : \mu \in \Gamma\}$ with $f_\mu(x) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{(x-\mu)^2}{2})$. Then the model of one population and two population mixtures, respectively, is given by

$$\mathcal{G}_1 = \mathcal{F} \quad \text{and} \quad \mathcal{G}_2 = \Big\{ g_{\pi,(\mu_1,\mu_2)} = \pi f_{\mu_1} + (1-\pi) f_{\mu_2} : \ \pi \in [0,1],\ \mu_1,\ \mu_2 \in \Gamma \Big\}.$$

According to Example 3.15 the asymptotic theory from [DCG99] is applicable to the test

$$\boldsymbol{H_0 :\ g \in \mathcal{G}_1} \text{ against } \boldsymbol{H_1 :\ g \in \mathcal{G}_2 \backslash \mathcal{G}_1}.$$

We assume that the true density is given by the standard normal density $g^0 = f_0$ and define $\Gamma = [-a, a]$, $a > 0$. Let $X_1, \dots, X_n$ be an i.i.d. standard normally distributed sample with realization $x_1, \dots, x_n$. According to Corollary 2.26 and (1.1) the log-LRT statistic is given by

$$\log\big(\Lambda_n(x_1,\dots,x_n)\big) \;=\; \sup_{\substack{\mu_1,\mu_2 \in \Gamma \\ \pi \in [0,1]}} \sum_{i=1}^n \log\Big(g_{\pi,(\mu_1,\mu_2)}(x_i)\Big) - \sum_{i=1}^n \log\Big(f_{\hat{\mu}_n}(x_i)\Big), \qquad (4.1)$$

with $\hat{\mu}_n = (-a)\mathbb{1}_{\bar{x}_n<(-a)}(x_1,\dots,x_n) + \bar{x}_n\mathbb{1}_{\bar{x}_n\in\Gamma}(x_1,\dots,x_n) + a\mathbb{1}_{\bar{x}_n>a}(x_1,\dots,x_n)$ and $\bar{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i$, and converges in distribution to the variable

$$\frac{1}{2}\sup_{\mu\in\Gamma\backslash\{0\}} \left(\xi_{h_\mu}\right)^2 \cdot \mathbb{1}_{\xi_{h_\mu}\geq 0}. \tag{4.2}$$

We obtain the expression of the function $h : (\Gamma\backslash\{0\}, \mathfrak{X}) \to I\!R$ by inserting $k = 1$ in (2.58) and simple calculations of

$$\begin{aligned}
o_1(x) &= \frac{\partial_\mu f_\mu|_{\mu=0}}{f_0}\left\|\frac{\partial_\mu f_\mu|_{\mu=0}}{f_0}\right\|^{-1}_{L_2(f_0\lambda)} &&= x, \\
a(\mu) &= \left\langle o_1, \frac{f_\mu - f_0}{f_0}\right\rangle_{L_2(f_0\lambda)} &&= \mu, \\
h_\mu(x) &= \frac{\frac{f_\mu(x)-f_0(x)}{f_0(x)} - o_1(x)a(\mu)}{\left\|\frac{f_\mu(x)-f_0(x)}{f_0(x)} - o_1(x)a(\mu)\right\|_{L_2(f_0\lambda)}} &&= \frac{\exp\left(\mu x - \frac{\mu^2}{2}\right) - 1 - \mu x}{\left(\exp\left(\mu^2\right) - 1 - \mu^2\right)^{\frac{1}{2}}}.
\end{aligned}$$

Similar simple calculations (see appendix) lead to the covariance function

$$Cov\left(\xi_{h_\mu}, \xi_{h_{\tilde{\mu}}}\right) = \frac{\exp\left(\mu\tilde{\mu}\right) - 1 - \mu\tilde{\mu}}{\left(\exp\left(\mu^2\right) - 1 - \mu^2\right)^{\frac{1}{2}}\left(\exp\left(\tilde{\mu}^2\right) - 1 - \tilde{\mu}^2\right)^{\frac{1}{2}}}, \quad \mu,\ \tilde{\mu}\in\Gamma\backslash\{0\}, \tag{4.3}$$

of the centered Gaussian process $\left(\xi_{h_\mu}\right)_{\mu\in\Gamma\backslash\{0\}}$.
Obviously, $Cov\left(\xi_{h_\mu}, \xi_{h_\mu}\right) = Var\left(\xi_{h_\mu}\right) = 1$ for all $\mu\in\Gamma\backslash\{0\}$. Since $\exp(x) - 1 - x$, $x\in I\!R$, has its global minimum in $x = 0$, one has

Remark 4.1 *For any $\mu, \tilde{\mu}\in\Gamma\backslash\{0\}$ it holds $Cov\left(\xi_{h_\mu}, \xi_{h_{\tilde{\mu}}}\right) > 0$.*

Therefore, we know the limiting process, its index set and its marginal distributions. However, the question is, how do we simulate a supremum of a Gaussian process indexed by functions h_μ? The following proposition gives an answer to this problem.

Proposition 4.2 *Let $\Gamma\subset I\!R^k$ be a compact and convex set with accumulation point in $\boldsymbol{\gamma^0}\in\Gamma$ (i.e. (A1) and (A2) are satisfied). Let $f_{\boldsymbol{\gamma^0}}$ be a probability density with respect to some σ-finite measure ν on a measurable space $(\mathfrak{X}, \mathcal{X})$, $\mathfrak{X}\subset I\!R^m$, and let the function*

$$h : \left(\Gamma\backslash\{\boldsymbol{\gamma^0}\}, \mathfrak{X}\right) \to I\!R \quad \textit{with} \quad (\boldsymbol{\gamma}, \boldsymbol{x}) \mapsto h_{\boldsymbol{\gamma}}(\boldsymbol{x})$$

be continuous. Furthermore, let us suppose that $(\xi_{h_{\boldsymbol{\gamma}}})_{\boldsymbol{\gamma}\in\Gamma\backslash\{\boldsymbol{\gamma^0}\}}$ is a centered Gaussian process with continuous sample paths with respect to the pseudometric $\rho(h_{\boldsymbol{\gamma^1}}, h_{\boldsymbol{\gamma^2}}) = \|h_{\boldsymbol{\gamma^1}} - h_{\boldsymbol{\gamma^2}}\|_{L_2(f_{\boldsymbol{\gamma^0}}\nu)}$. Then for any increasing sequence $(\Gamma_N)_{N\in I\!N}$ of finite subsets of $\Gamma\backslash\{\boldsymbol{\gamma^0}\}$ with $\bigcup_{N=1}^{\infty}\Gamma_N$ dense in $\Gamma\backslash\{\boldsymbol{\gamma^0}\}$ one has

$$\frac{1}{2}\sup_{\boldsymbol{\gamma}\in\Gamma_N} (\xi_{h_{\boldsymbol{\gamma}}})^2 \cdot \mathbb{1}_{\xi_{h_{\boldsymbol{\gamma}}}\geq 0} \overset{\text{a.s.}}{\longrightarrow} \frac{1}{2}\sup_{\boldsymbol{\gamma}\in\Gamma\backslash\{\boldsymbol{\gamma^0}\}} (\xi_{h_{\boldsymbol{\gamma}}})^2 \cdot \mathbb{1}_{\xi_{h_{\boldsymbol{\gamma}}}\geq 0} \quad \textit{as } N\to\infty. \tag{4.4}$$

Proof of Proposition 4.2:
$\frac{1}{2} \sup_{\gamma \in \Gamma_N} (\xi_{h_\gamma})^2 \cdot \mathbb{1}_{\xi_{h_\gamma} \geq 0}$ is the supremum of a finite number of random variables and hence is a random variable for each N. From the assumption $(\xi_{h_\gamma})_{\gamma \in \Gamma \setminus \{\boldsymbol{\gamma^0}\}}$ having continuous sample paths it follows that $\left((\xi_{h_\gamma})^2 \cdot \mathbb{1}_{\xi_{h_\gamma} \geq 0}\right)_{\gamma \in \Gamma \setminus \{\boldsymbol{\gamma^0}\}}$ also has a.s. continuous sample paths. Moreover, from $\bigcup_{N=1}^{\infty} \Gamma_N$ being dense in $\Gamma \setminus \{\boldsymbol{\gamma^0}\}$ it follows (4.4). □

Our proof follows the concept of Leadbetter, Lindgren and Rootzén (1983), p. 146, who prove that $\sup_{\gamma \in \Gamma_N} (\xi_\gamma) \xrightarrow{\text{a.s.}} \sup_{\gamma \in \Gamma} (\xi_\gamma)$ for $\Gamma = [0, 1]$.

To apply Proposition 4.2 to the Gaussian process (4.2) it is sufficient to ensure that $h_\mu(x)$ is continuous on $(\Gamma \setminus \{0\}, \mathfrak{X})$ because all other assumptions are satisfied thanks to Corollary 2.26. The continuity of $h_\mu(x)$ follows from

$$\lim_{\mu \to 0} \frac{\exp\left(\mu x - \frac{\mu^2}{2}\right) - 1 - \mu x}{\left(\exp\left(\mu^2\right) - 1 - \mu^2\right)^{\frac{1}{2}}} \quad = \quad \frac{1}{2}\left(x^2 - 1\right)^2,$$

which can be calculated by applying the rule of l'Hospital to $\lim_{\mu \to 0} (h_\mu(x))^2$. (We calculated the latter expression with MATHEMATICA.)

We simulate the Gaussian process $\left(\xi_{h_\mu}\right)_{\mu \in \Gamma \setminus \{0\}}$ by its finite-dimensional marginal distributions. For this purpose, we discretize $\Gamma \setminus \{0\} = [-a, a] \setminus \{0\}$ in finite sets $\Gamma_N = \{\mu_1, \ldots, \mu_N\}$ using the fact that

$$\boldsymbol{\xi}_N = \left(\xi_{h_{\mu_1}}, \ldots, \xi_{h_{\mu_N}}\right)' \sim N(\mathbf{0}, \Sigma_N) \tag{4.5}$$

and according to (4.3) the covariance matrix is given by

$$\Sigma_N \quad = \quad \left(Cov\left(\xi_{h_{\mu_i}}, \xi_{h_{\mu_j}}\right)\right)_{\mu_i, \mu_j \in \Gamma_N} \in I\!R^{N \times N}. \tag{4.6}$$

It is well known that for any matrix $L_N \in I\!R^{N \times N}$ such that $\Sigma_N = L_N L_N'$ and for standard normally distributed $\boldsymbol{X_N} \sim N(\mathbf{0}, I_N)$ one has

$$L_N \boldsymbol{X_N} \sim N(\mathbf{0}, \Sigma_N).$$

Consequently, all we need is an appropriate factorization of $\boldsymbol{\Sigma_N}$ and discretization of Γ_N, respectively:

- **Spectral factorization of $\boldsymbol{\Sigma_N}$:**
  The factorization $\Sigma_N = L_N L_N'$ is not uniquely defined. We decided to use a spectral factorization

  $$\Sigma_N \quad = \quad V_N \, \Delta_N \, V_N' \quad = \quad \left(V_N \, (\Delta_N)^{\frac{1}{2}}\right)\left(V_N \, (\Delta_N)^{\frac{1}{2}}\right)' \tag{4.7}$$

  where $\Delta_N = diag(\lambda_1, \ldots, \lambda_N)$ is the diagonal matrix of the eigenvalues $\lambda_1, \ldots, \lambda_N$ of Σ_N in increasing order and $V_N = \left(\boldsymbol{v_1}, \ldots, \boldsymbol{v_N}\right)$ is an orthogonal matrix of the corresponding eigenvectors. Clearly, there exist exactly N distinct non-negative eigenvalues since Σ_N is positive definite. Consequently, the diagonal matrix $\left(\Delta_N\right)^{\frac{1}{2}} = diag(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_N})$ is well defined. The NAG-routine F02FAF supports this dissection.

- **Choice of Γ_N:**
  With no loss of generality we discretize $\Gamma \backslash \{0\} = [-a, a] \backslash \{0\}$ in finite equidistant sets $\Gamma_N = \{\mu_1, \dots, \mu_N\}$, $N \in I\!N$ even, by

$$\mu_i = \begin{cases} -a + (i-1)\frac{2a}{N} & , \quad i = 1, \dots, \frac{N}{2} \\ -a + i\frac{2a}{N} & , \quad i = \frac{N}{2}, \dots, N. \end{cases}$$

The choice of a spectral factorization of Σ_N is actually common. However, when we generate the covariance matrix we observe a surprising effect on the size of the eigenvalues. Aforesaid, it strikes us that only a few eigenvalues have large values and most of them are approximately 0, but we also observe negative eigenvalues of approximately 0. Obviously, these are numerical errors. For instance, for $N = 1000$ and $\Gamma = [-10.0, 10]$, $[-5.0, 5.0]$, $[-2.5, 2.5]$, $[-1.0, 1.0]$ nearly half of the eigenvalues (EV) were negative whereas the number of eigenvalues larger than 10^{-9} is bounded by 50, see Example 4.3.

The surprisingly small number of eigenvalues larger than 10^{-9} can lead one to assume that the underlying limiting distribution for our testing problem is a low dimensional normal one.

As a consequence we implement two versions of

- **Modified spectral factorization of Σ_N:**

  1. **Using only eigenvalues larger than 0** and setting all $\lambda_i < 0$ to 0.

  2. **Using only eigenvalues larger than 10^{-9}** and setting all $\lambda_i < 10^{-9}$ to 0.

  In each case we define

$$L_N = V_N \left(\tilde{\Delta}_N\right)^{\frac{1}{2}} \in I\!R^{N \times (N-t)}$$

  for $\tilde{\Delta}_N = diag(0, \dots, 0, \lambda_t, \dots, \lambda_N)$ and $t \in \{1, \dots, N\}$ being the first index such that $\lambda_t > 0$.

In both versions we observe that $L_N L'_N \approx \Sigma_N$ (see Example 4.3). In the following section it will be seen that there is also a negligible difference between the corresponding quantiles of both versions. Hence, simulating a Gaussian process with the aid of the low dimensional marginal distributions of the second version leads to nearly the same results as the use of the first version. These observations are consistent with the aforementioned assumption of an asymptotic distribution of (4.2) being a low dimensional Gaussian one.

The most convincing advantage of simulating a Gaussian process while using the second modified spectral factorization is that it costs fewer matrix operations and also fewer random numbers with almost no loss of accuracy. Consequently, we observe considerably shorter computation times in comparison to the first version.

Example 4.3 *For $M \in \mathbb{R}^{N \times N}$ let the* 1*-norm of M be given by*

$$|M| \;=\; \sum_{i=1}^{N}\sum_{j=1}^{N} |M_{i,j}|$$

and the maximum norm of M be given by

$$\|M\|_\infty \;=\; \max_{(i,j)\in\{1,\ldots,N\}^2} |M_{i,j}|.$$

For $N = 1000$ we obtain

	$\boldsymbol{\Gamma = [-10.0, 10.0]}$		$\boldsymbol{\Gamma = [-5.0, 5.0]}$	
	528 EV > 0	**50 EV** $> 10^{-9}$	**516 EV** > 0	**28 EV** $> 10^{-9}$
$\boldsymbol{\lvert L_N L_N{}' - \Sigma_N \rvert}$	$3.33 \cdot 10^{-15}$	$2.37 \cdot 10^{-11}$	$3.77 \cdot 10^{-15}$	$1.15 \cdot 10^{-11}$
$\boldsymbol{\lVert L_N L_N{}' - \Sigma_N \rVert_\infty}$	$3.21 \cdot 10^{-9}$	$6.51 \cdot 10^{-7}$	$8.27 \cdot 10^{-9}$	$3.07 \cdot 10^{-7}$

	$\boldsymbol{\Gamma = [-2.5, 2.5]}$		$\boldsymbol{\Gamma = [-1.0, 1.0]}$	
	507 EV > 0	**18 EV** $> 10^{-9}$	**508 EV** > 0	**17 EV** $> 10^{-9}$
$\boldsymbol{\lvert L_N L'_N - \Sigma_N \rvert}$	$5.55 \cdot 10^{-15}$	$3.94 \cdot 10^{-12}$	$5.72 \cdot 10^{-14}$	$4.35 \cdot 10^{-12}$
$\boldsymbol{\lVert L_N L_N{}' - \Sigma_N \rVert_\infty}$	$6.99 \cdot 10^{-8}$	$1.95 \cdot 10^{-7}$	$7.35 \cdot 10^{-7}$	$8.79 \cdot 10^{-7}$

Unfortunately, generating the covariance matrix $Cov(\xi_{h_\mu}, \xi_{h_{\tilde\mu}})$ (see (4.3)) for μ or $\tilde\mu$ being approximately 0 leads to unpleasant and confusing numerical inaccuracies. For instance, according to computations with FORTRAN 77 (as well as with MATHEMATICA and R) it seems as if the covariance function is oscillating in $\mu \in [-0.01, 0.01]\backslash\{0\}$ for fixed $\tilde\mu = 10^{-5}$ (or vice versa, because of the symmetry of Σ_N), see Figure 4.1. Furthermore, for $\tilde\mu = -0.002$ we obtain covariance values larger than 1 which contradicts $Var(\xi_{h_\mu}) = 1$, $\mu \in \Gamma\backslash\{0\}$, see Figure 4.2.

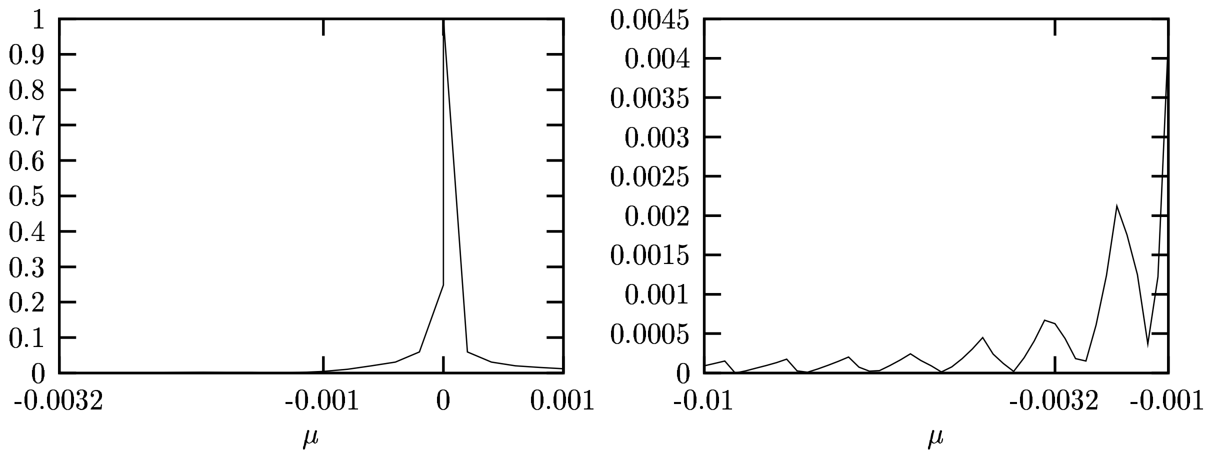

Figure 4.1.: $Cov(\mu, 10^{-5})$

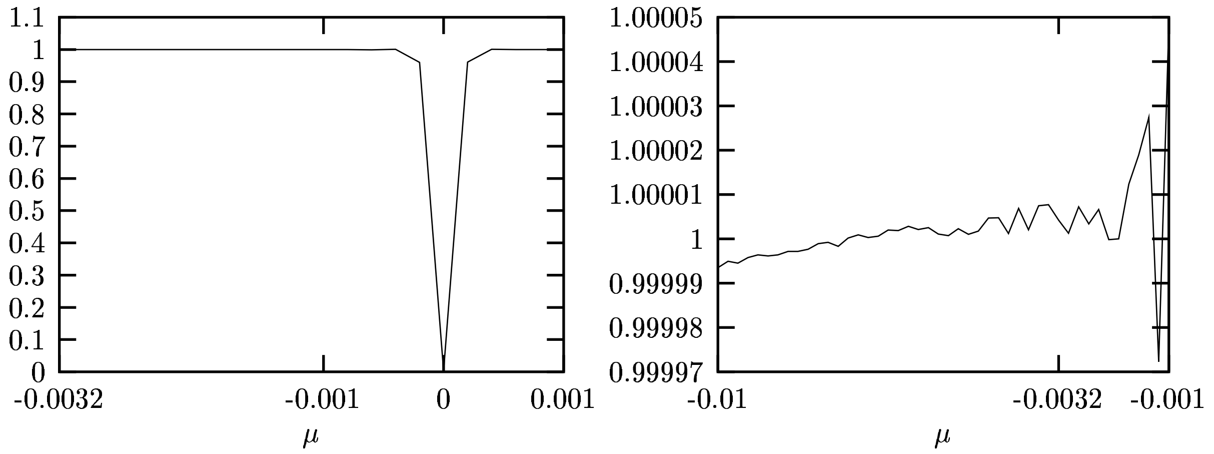

Figure 4.2.: $Cov(\mu, -0.002)$

However, an analysis of the covariance function shows that it is well defined and smooth since

1. $\lim_{\mu \to 0} \dfrac{\exp\left(\mu\tilde{\mu}\right) - 1 - \mu\tilde{\mu}}{\left(\exp\left(\mu^2\right) - 1 - \mu^2\right)^{\frac{1}{2}}} = \dfrac{\tilde{\mu}^2}{\sqrt{2}},$

2. for arbitrary but fixed $\tilde{\mu} \in [-0.01, 0.01] \backslash \{0\}$ the function

$$Cov\left(\xi_{h_\mu}, \xi_{h_{\tilde{\mu}}}\right) : [-0.01, 0.01] \backslash \{0\} \to (0, 1] \tag{4.8}$$

   has an unique (global) maximum in $(\tilde{\mu}, \tilde{\mu})$.

The first item results from the calculation of the square covariance function and the rule of l'Hospital. (We calculated the limit with the help of MATHEMATICA.) The second item is shown in the appendix. Notice that the range $(0, 1]$ given in (4.8) is a consequence of Remark 4.1.

We solve the numerical problem of oscillations and inexact values by using a Taylor expansion of (4.3) for μ or $\tilde{\mu} \in [-0.01, 0.01] \backslash \{0\}$. In all cases we have decided to use finite sums of $n = 100$ terms. This leads to a remainder of $r(|\mu|) \leq \frac{|\mu|^{101}}{101!} \approx 0$ for $\mu \in [-0.01, 0.01] \backslash \{0\}$. We arrange the terms of the resulting Taylor series in such a manner that the whole expression can be calculated in a numerically stable way. Roughly speaking, since the nominator as well as the denominator of (4.3) is approximately 0, we reduce the ratio until we obtain a quotient of interlocking expressions which are near to 1. Furthermore, we distinguish between the case of $\mu, \tilde{\mu} \in [-0.01, 0.01] \backslash \{0\}$ and the case of exactly one of the two parameters belongs to $[-0.01, 0.01] \backslash \{0\}$. A more detailed description of this solution is given in the appendix.

Additionally, an implementation in pseudo FORTRAN 77 code of simulating the variable

$$\frac{1}{2} \sup_{\mu \in \Gamma_N} \left(\xi_{h_\mu}\right)^2 \cdot \mathbb{1}_{\xi_{h_\mu} \geq 0}$$

is given in Program B.1 in the appendix.

Further simulation results of the same testing problem are given in Ruck (2002) who calculates the quantiles of the asymptotic distributions of

$$\frac{1}{2}\sup_{d\in\mathcal{D}}\left(\xi_d\right)^2\cdot\mathbb{1}_{\xi_d\geq 0}-\frac{1}{2}\sup_{d\in\mathcal{D}_0}\left(\xi_d\right)^2\cdot\mathbb{1}_{\xi_d\geq 0}$$

(see Theorem 2.25).

We also refer to Ruck (2001) who uses the above method to calculate the asymptotic distribution of the log-LRT statistic in a contamination mixture model.

For testing homogeneity against a two population mixture with $\Gamma\subset\mathbb{R}^k$, we can also use the method for simulating the corresponding asymptotic distribution thanks to Corollary 2.26 and Proposition 4.2 which are also applicable to k-dimensional parameter spaces here proposed.

4.1. Quantiles of the Asymptotic Distribution

We simulate the quantiles of the asymptotic distributions of $\frac{1}{2}\sup_{\mu\in\Gamma_N}\left(\xi_{h_\mu}\right)^2\cdot\mathbb{1}_{\xi_{h_\mu}\geq 0}$ as described above using 10^7 replications of the empirical test statistic. In each simulation program we start the random generator with the same seed (=0) to obtain repeatable computations. We tabulate the α-quantiles only when eigenvalues (EV) larger than 0 and only eigenvalues larger than 10^{-9} were used, respectively. Moreover, we note the number of the corresponding eigenvalues. We observe for $N=100$ and $N=1000$ that nearly half of the eigenvalues were negative (which are numerical errors) and that the number of eigenvalues larger than 10^{-9} increase with the length of the interval Γ. In our examples this number is bounded by 50 (see $\Gamma=[-10,10]$, $N=1000$). Consequently, the computing times of the quantiles of the second version are noticeably shorter than the computing times of the first version. We also observe that there is a negligible difference between the corresponding quantiles of both versions.

	$\mathbf{\Gamma=[-10.0,10.0]}$		$\mathbf{\Gamma=[-5.0,5.0]}$	
$\mathbf{N=10}$	**10 EV >0**	**10 EV $>10^{-9}$**	**10 EV >0**	**10 EV $>10^{-9}$**
$\alpha=0.9$	2.6524	2.6524	2.4752	2.4752
$\alpha=0.95$	3.2873	3.2873	3.1372	3.1372
$\alpha=0.975$	3.9218	3.9218	3.7953	3.7953
$\alpha=0.99$	4.7664	4.7664	4.6675	4.6675
$\alpha=0.995$	5.4090	5.4090	5.3240	5.3240
$\mathbf{N=100}$	**81 EV >0**	**47 EV $>10^{-9}$**	**67 EV >0**	**25 EV $>10^{-9}$**
$\alpha=0.9$	3.3815	3.3836	2.6964	2.6951
$\alpha=0.95$	4.0914	4.0914	3.3972	3.3949
$\alpha=0.975$	4.7898	4.7918	4.0908	4.0897
$\alpha=0.99$	5.7055	5.7039	4.9994	4.9980
$\alpha=0.995$	6.4008	6.3959	5.6865	5.6857
$\mathbf{N=1000}$	**528 EV >0**	**50 EV $>10^{-9}$**	**516 EV >0**	**28 EV $>10^{-9}$**
$\alpha=0.9$	3.3970	3.3952	2.6967	2.6975
$\alpha=0.95$	4.1089	4.1060	3.3979	3.3999
$\alpha=0.975$	4.8112	4.8059	4.0922	4.0942
$\alpha=0.99$	5.7315	5.7280	5.0040	5.0062
$\alpha=0.995$	6.4207	6.4196	5.6970	5.6968

	$\Gamma = [-2.5, 2.5]$		$\Gamma = [-1.0, 1.0]$	
N = 10	**10 EV > 0**	**10 EV > 10^{-9}**	**10 EV > 0**	**8 EV > 10^{-9}**
$\alpha = 0.9$	1.9716	1.9716	1.3452	1.3465
$\alpha = 0.95$	2.6404	2.6404	1.9730	1.9743
$\alpha = 0.975$	3.3109	3.3109	2.6177	2.6195
$\alpha = 0.99$	4.2018	4.2018	3.4822	3.4887
$\alpha = 0.995$	4.8727	4.8727	4.1444	4.1539
N = 100	**60 EV > 0**	**15 EV > 10^{-9}**	**56 EV > 0**	**10 EV > 10^{-9}**
$\alpha = 0.9$	2.0078	2.0072	1.3484	1.3473
$\alpha = 0.95$	2.6880	2.6862	1.9763	1.9759
$\alpha = 0.975$	3.3660	3.3668	2.6232	2.6214
$\alpha = 0.99$	4.2669	4.2694	3.4942	3.4873
$\alpha = 0.995$	4.9490	4.9476	4.1608	4.1493
N = 1000	**507 EV > 0**	**18 EV > 10^{-9}**	**508 EV > 0**	**17 EV > 10^{-9}**
$\alpha = 0.9$	2.0100	2.0091	1.3490	1.3474
$\alpha = 0.95$	2.6896	2.6860	1.9777	1.9751
$\alpha = 0.975$	3.3695	3.3670	2.6235	2.6223
$\alpha = 0.99$	4.2707	4.2677	3.4913	3.4907
$\alpha = 0.995$	4.9540	4.9492	4.1494	4.1569

An implementation in pseudo FORTRAN 77 code of calculating the quantiles of the asymptotic distribution is given in Program B.2 in the appendix.

4.2. Quantiles of the Log-LRT Statistic

In this section we describe how to simulate the exact quantiles of the log-LRT statistic $\log\big(\Delta_n(x_1,\dots,x_n)\big)$ (4.1). We decide to use an EM-algorithm to calculate

$$\sup_{\substack{\mu_1,\mu_2\in I\!R\\ \pi\in[0,1]}} \sum_{i=1}^{n} \log\Big(g_{\pi,\mu_1,\mu_2}(x_i)\Big) \tag{4.9}$$

because it is known for this iterative algorithm to work for functions with nearly constant regions. While the EM-algorithm conceptually works on the unbounded $I\!R^m$ it can be modified to respect the given boundaries of $\Gamma = [-a, a]$. Our modification works as follows: A starting point is chosen in the interior of Γ. Whenever the unmodified EM-algorithm crosses the bounds of $\Gamma = [-a, a]$ then a local search is performed using the NAG-routine E04JYF starting at the last iteration point inside $[-a, a]$. This is a quasi-Newton algorithm for finding a minimum of a function $F(x_1,\dots,x_n)$, subject to fixed upper and lower bounds on independent variables $x_1,\dots,x_n$, using function values only. For this we make use of the equivalence of minimization and maximization

$$\sup_{\substack{\mu_1,\mu_2\in\Gamma\\ \pi\in[0,1]}} \sum_{i=1}^{n} \log\Big(g_{\pi,\mu_1,\mu_2}(x_i)\Big) \quad = \quad -\min_{\substack{\mu_1,\mu_2\in\Gamma\\ \pi\in[0,1]}} -\sum_{i=1}^{n} \log\Big(g_{\pi,\mu_1,\mu_2}(x_i)\Big).$$

While the results of the EM-algorithm often depend on the chosen starting value M. Heidenreich showed that for the case of normal mixture models there is nearly no dependency on the starting point. So we decide to use $\pi = 0.5$, $\mu_1 = 0.5\bar{x}_n$ and $\mu_2 = 1.5\bar{x}_n$ as starting values, with $\bar{x}_n$ being the average of $x_1,\dots,x_n$, to obtain the maximum relatively fast. Iff the EM-algorithm crosses the bounds of Γ we apply the routine E04JYF with starting values given by the last values of the EM in the interior of Γ.

Details on the inner workings of the EM-algorithm can be found in McLachlan and Krishnan (1997).

We simulate the quantiles of $\log\left(\Delta_n(x_1,\ldots,x_n)\right)$ using 10.000 or 100.000 replications of the empirical test statistic and a stopping criterion based on function values stopped at an accuracy of $n \cdot acc = 10^{-8}$ for $n = 10,\ 100,\ 1000$ and $n \cdot acc = 10^{-5}$ for $n = 10.000$, respectively, or after 100.000 EM-iterations. For each simulation program we start the random generator with the same seed (=0) to obtain repeatable computations. Furthermore, we generate the $N(\mathbf{0}, I_n)$–distributed vector $(x_1,\ldots,x_n)'$ by the NAG-routine G05FDF. An implementation in pseudo FORTRAN 77 code is given in Program B.3 in the appendix.

A table of the α-quantiles for the intervals $\Gamma = [-10.0, 10.0],\ [-5.0, 5.0]$ and $[-2.5, 2.5]$ is given below. In the first two cases the EM never or hardly ever crosses the bounds of Γ, which means that forcing the bounds on the EM-algorithm has no effect on the results. For this reason the quantiles on the intervals $[-10.0, 10.0]$ and $[-5.0, 5.0]$ are the same. There is only a little difference between the quantiles of these intervals and the interval $[-2.5, 2.5]$, too. It seems as if the EM-algorithm converges very slowly for $n = 10.000$ because for fixed $n \cdot acc = 10^{-8}$ the quantiles are increasing with increasing $n = 10, 100, 1000$ whereas the quantiles for $n = 10.000$ calculated with accuracy $n \cdot acc = 10^{-5}$ are much smaller than the quantiles for $n = 1000$.

We also tabulate the corresponding quantiles of the asymptotic distribution given in the previous section as well as the quantiles calculated with the help of Davies' bound (see Davies (1977)). With respect to the latter quantiles we refer to private communication with Bernard Garel. It is clear to see that the α-quantiles calculated by simulating a Gaussian process are nearly the same as those which are calculated with the help of Davies' bound.

	REPL= 100.000		**REPL=10.000**		**Asymp**	**Davies**
	n=10	**n=100**	**n=1.000**	**n=10.000**	**EV $> 10^{-9}$**	
$\boldsymbol{n \cdot acc}$	$\mathbf{10^{-8}}$	$\mathbf{10^{-8}}$	$\mathbf{10^{-8}}$	$\mathbf{10^{-5}}$		
$\mathbf{[-10.0, 10.0]}$						
$\alpha\ =\ 0.9$	0.8785	1.2947	1.4959	1.2768	3.3952	3.4667
$\alpha\ =\ 0.95$	1.4967	1.9762	2.2169	2.0860	4.1060	4.1567
$\alpha\ =\ 0.975$	2.1375	2.6521	2.9161	2.7800	4.8059	4.8474
$\alpha\ =\ 0.99$	2.9777	3.5130	3.9037	3.7566	5.7280	5.7610
$\alpha\ =\ 0.995$	3.6272	4.2338	4.5305	4.3794	6.4196	6.4525
$\mathbf{[-5.0, 5.0]}$						
$\alpha\ =\ 0.9$	0.8785	1.2947	1.4959	1.2768	2.6975	2.7874
$\alpha\ =\ 0.95$	1.4967	1.9762	2.2169	2.0860	3.3999	3.4728
$\alpha\ =\ 0.975$	2.1375	2.6521	2.9161	2.7800	4.0942	4.1599
$\alpha\ =\ 0.99$	2.9777	3.5130	3.9037	3.7566	5.0062	5.0698
$\alpha\ =\ 0.995$	3.6272	4.2338	4.5305	4.3794	5.6968	5.7591
$\mathbf{[-2.5, 2.5]}$						
$\alpha\ =\ 0.9$	0.8750	1.2827	1.4902	1.2767	2.0091	2.1338
$\alpha\ =\ 0.95$	1.4806	1.9377	2.1684	2.0655	2.6860	2.8072
$\alpha\ =\ 0.975$	2.1016	2.5836	2.8310	2.7661	3.3670	3.4853
$\alpha\ =\ 0.99$	2.8892	3.4152	3.7410	3.7376	4.2677	4.3863
$\alpha\ =\ 0.995$	3.4783	4.0511	4.3693	4.3651	4.9492	5.0703

Table 4.1.: Quantiles of the log-LRT statistic and their corresponding asymptotic quantiles

4.3. Power

We simulate the power of level α of the log-LRT for a random sample of size $n = 100$ and 1000 based on the corresponding exact quantiles and the quantiles of the asymptotic distribution as given in Table 4.1 using 10.000 replications. In this paper we restrict ourselves to $\alpha = 0.9$ and refer to Ruck (2002), where power plots are also given for $\alpha = 0.95, 0.975, 0.99$. We observe that the power is symmetrical as well in the mixing weights π and $(1-\pi)$ as in the mixing components f_{μ_1} and f_{μ_2}. Therefore, we only plot the power for the mixing weights $\pi = 0.1,\ 0.25,\ 0.5$ of the fixed component $f_{\mu_1} = f_0$ of the standard normal density and mixing weight $(1-\pi)$ of the mixing component of f_{μ_2}, $\mu_2 \in \Gamma$. For each simulation program we start the random generator with the same seed (=0) to get repeatable computations. Moreover, in any program calculations for $\pi = 0.1,\ 0.25$ and 0.5 are made. As described in section 4.2 we use a combination of an EM-algorithm and a maximization routine of the NAG.

We plot the power of the log-LR tests for the intervals $\Gamma = [-10.0, 10.0]$, $[-5.0, 5.0]$ and $[-2.5, 2.5]$ using an EM-algorithm with an accuracy of $n \cdot acc = 10^{-8}$ and iff it crosses the bounds of Γ, we apply the routine E04JYF. The following figures show the power of the tests using the quantiles of the asymptotic distribution which is generated by a Gaussian process ("Asymp") using only eigenvalues larger than 10^{-9}.

For $\Gamma = [-10.0, 10.0]$ and $\Gamma = [-5.0, 5.0]$ we also plot the power of the tests using the exact quantiles ("LRT") as well as the quantiles of the asymptotic distribution which are generated by the help of Davies' bound ("Davies"). As mentioned above there is no difference between the exact quantiles using the unbounded EM-algorithm and using the modified EM-algorithm. Hence, quantiles calculated using the unbounded EM-algorithm can be interpreted as the exact quantiles.

For $\Gamma = [-2.5, 2.5]$ we also plot the power of the tests using the exact quantiles of the modified EM-algorithm ("LRTE04") and the exact quantiles when the unbounded EM is used ("LRTEM"). The latter power plot is based on the exact quantiles of $\Gamma = [-5.0, 5.0]$ as explained above and for "LRTE04" we use the corresponding quantiles given in Table 4.1.

An implementation in pseudo FORTRAN 77 code of calculating the power is given in Program B.4 in the appendix. The program is initialized by sample size of $n = 1000$, $\Gamma = [-2.5, 2.5]$ and the aforementioned corresponding quantiles.

The power plots reveal the surprising result that the power curves based on the asymptotic quantiles of "Asymp" and "Davies" on $\Gamma = [-10.0, 10.0]$, $[-5.0, 5.0]$ are approximately identical to the curves of the exact quantiles "LRT" minus a positive constant. Therefore, the power of "Asymp" and "Davies" are a lower bound of the power of "LRT". For the interval $\Gamma = [-2.5, 2.5]$ we make the same pleasant observations for "Asymp" and "LRTE04", "LRTEM". The most encouraging fact is that for a sample size of $n = 1000$ this constant is approximately 0. Thus the difference between the exact and the asymptotic quantiles do not make much of a difference to the power. We also observe that the computing times for calculating the asymptotic quantiles are considerably shorter than the computing times of the corresponding combined EM-algorithm even for a small sample size of $n = 1000$. Consequently, the use of our modified spectral factorization using only eigenvalues larger than 10^{-9} is an efficient method of calculating the quantiles of the LRT statistic.

$\mathbf{\Gamma = [-10.0, 10.0]}$, $\boldsymbol{\alpha} = \mathbf{0.9}$ (the curves "Asymp" and "Davies" often overlap)

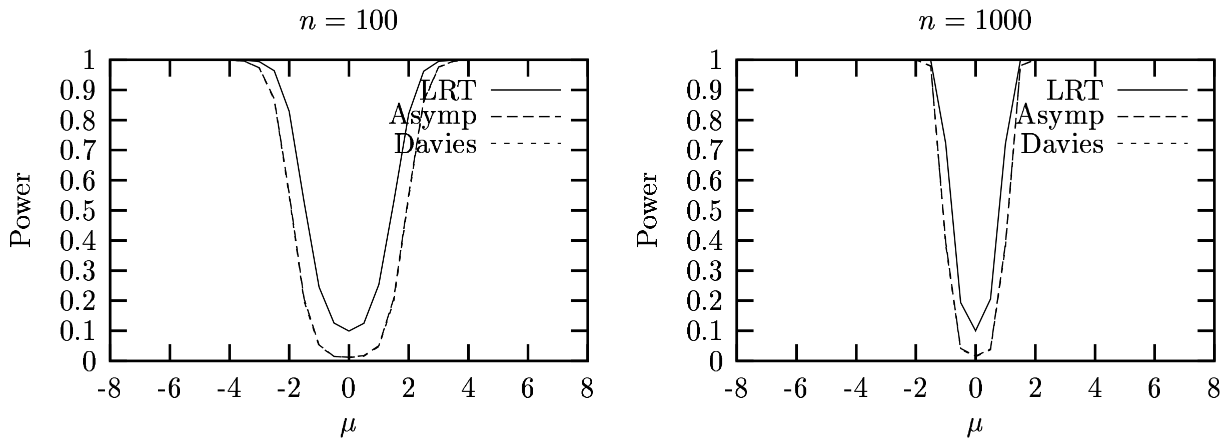

Figure 4.3.: $\pi = 0.1$: $\quad 0.1N(0,1) + (1-0.1)N(\mu,1)$

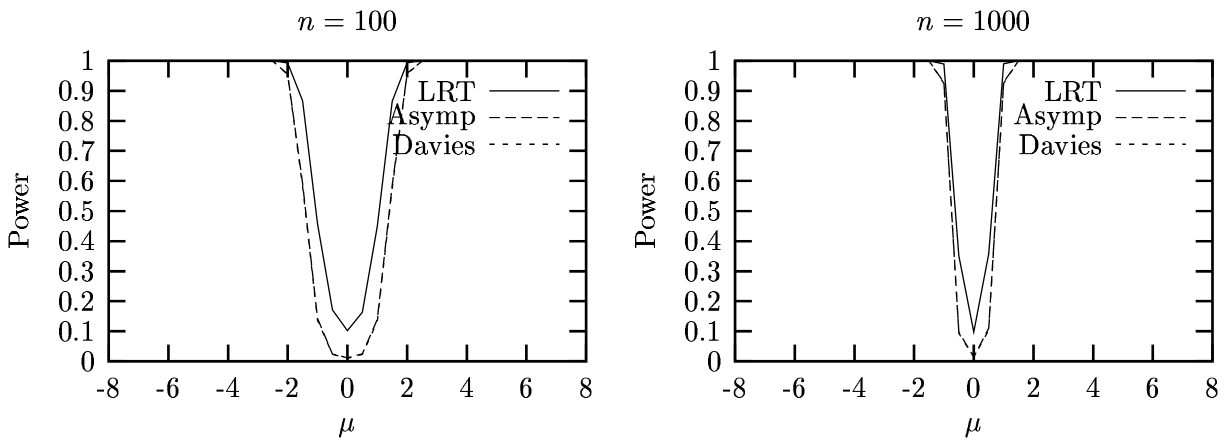

Figure 4.4.: $\pi = 0.25$: $\quad 0.25N(0,1) + (1-0.25)N(\mu,1)$

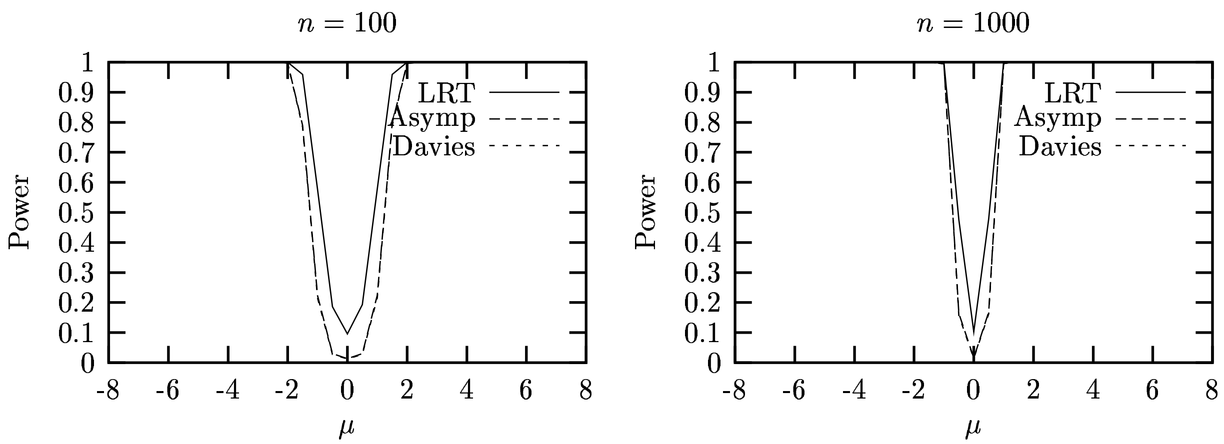

Figure 4.5.: $\pi = 0.5$: $\quad 0.5N(0,1) + (1-0.5)N(\mu,1)$

$\mathbf{\Gamma = [-5.0, 5.0]}$, $\boldsymbol{\alpha = 0.9}$ (the curves "Asymp" and "Davies" often overlap)

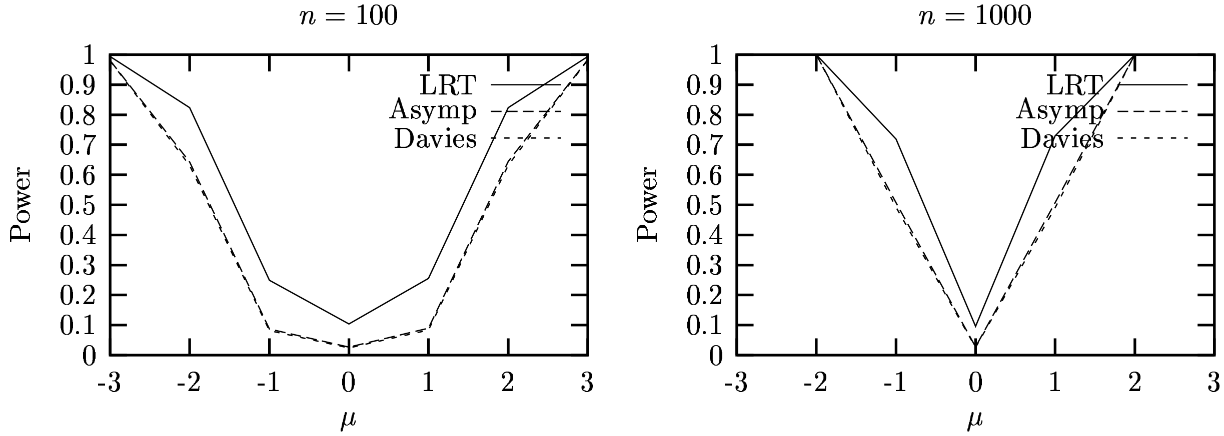

Figure 4.6.: $\pi = 0.1$: $0.1N(0,1) + (1-0.1)N(\mu,1)$

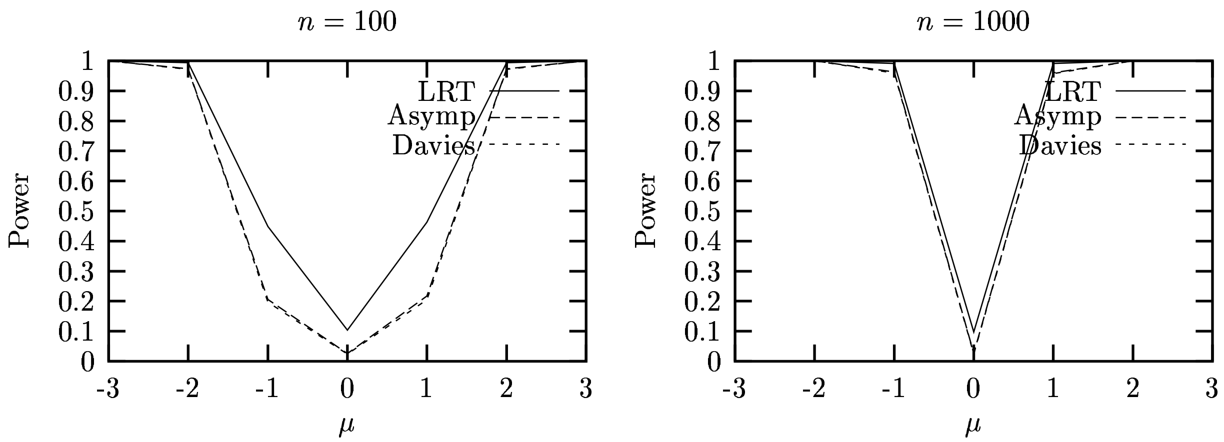

Figure 4.7.: $\pi = 0.25$: $0.25N(0,1) + (1-0.25)N(\mu,1)$

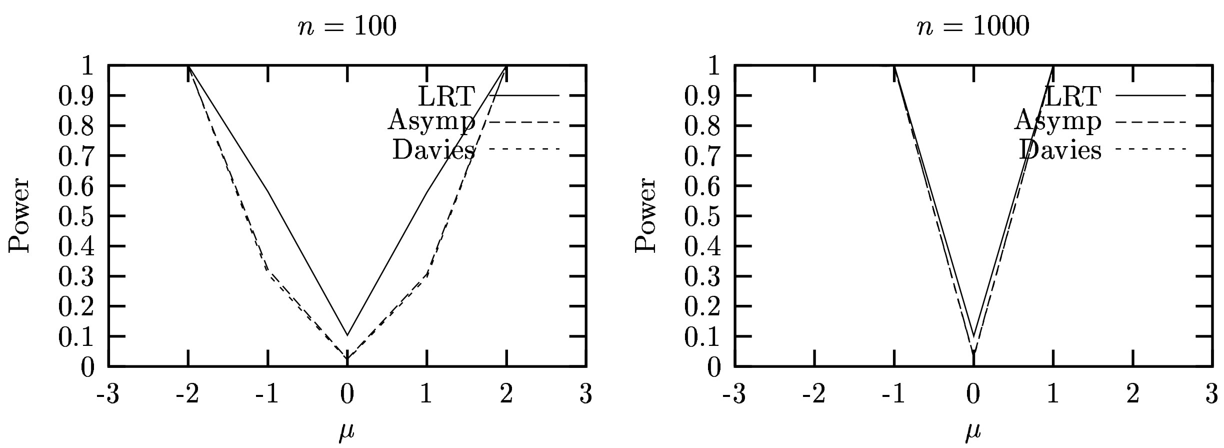

Figure 4.8.: $\pi = 0.5$: $0.5N(0,1) + (1-0.5)N(\mu,1)$

$\mathbf{\Gamma = [-2.5, 2.5]}$, $\boldsymbol{\alpha} = \mathbf{0.9}$ (the curves “LRTE04” and “LRTEM” often overlap)

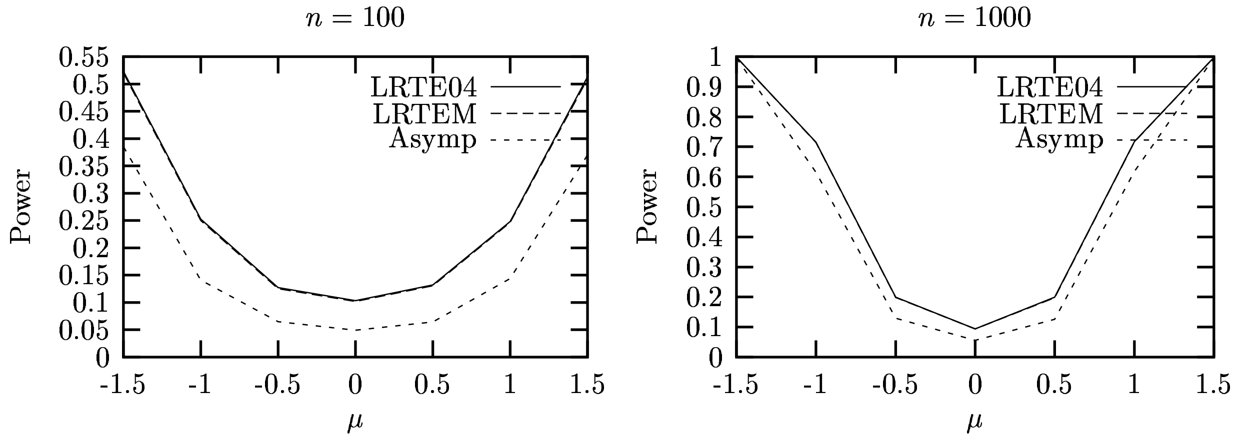

Figure 4.9.: $\pi = 0.1$: $\quad 0.1N(0,1) + (1-0.1)N(\mu,1)$

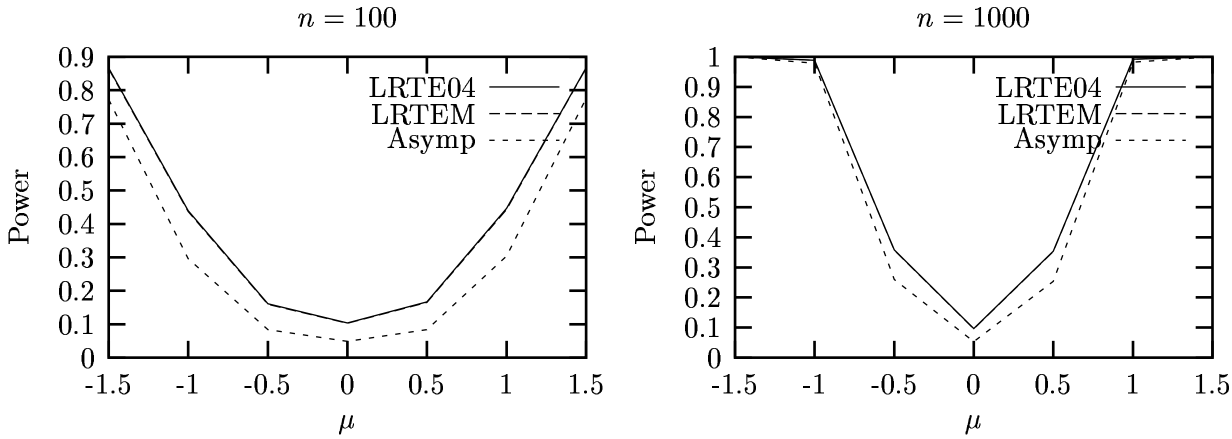

Figure 4.10.: $\pi = 0.25$: $\quad 0.25N(0,1) + (1-0.25)N(\mu,1)$

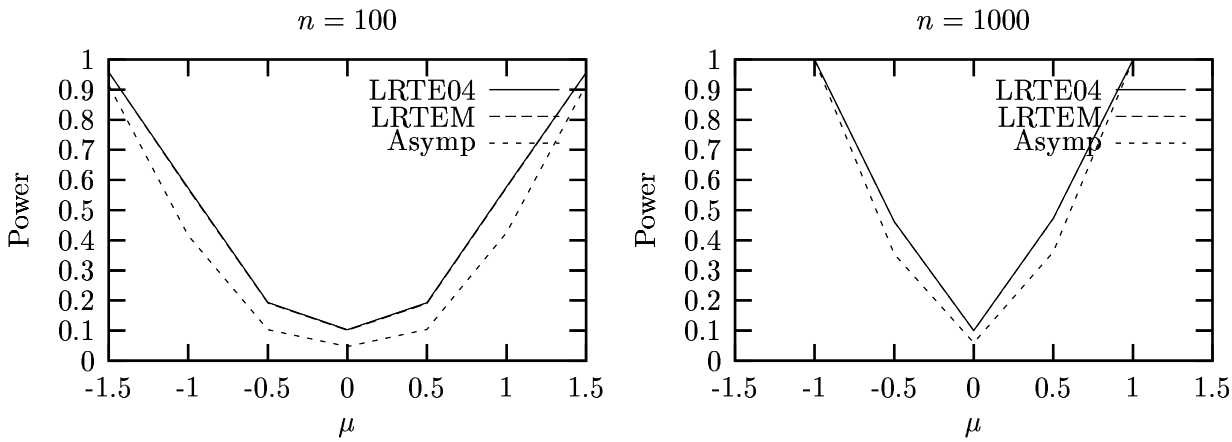

Figure 4.11.: $\pi = 0.5$: $\quad 0.5N(0,1) + (1-0.5)N(\mu,1)$

A. Formal Proofs

A.1. Proof of Lemma 2.20

1) **On $\boldsymbol{A_n}$,** we expand according to [DCG99], p. 1196, $g_{(\theta,\boldsymbol{\beta})}$ up to the order 2 at the point $\theta = 0$. For $\theta > 0$ we have

$$g_{(\theta,\boldsymbol{\beta})}(\boldsymbol{x}) \;=\; g^0(\boldsymbol{x}) + \theta\, \partial_\theta g_{(\theta,\boldsymbol{\beta})}(\boldsymbol{x})|_{\theta=0} + \frac{\theta^2}{2}\, \partial^2_\theta g_{(\theta,\boldsymbol{\beta})}(\boldsymbol{x})|_{\theta=0} + \frac{\theta^3}{3!}\, \partial^3_\theta g_{(\theta,\boldsymbol{\beta})}(\boldsymbol{x})|_{\theta=\theta^*}$$

for some $\theta^* \leq \theta$. Thereby, it results from the representation (2.12) of $g_{(\theta,\boldsymbol{\beta})}$ that

$$\begin{aligned}
\partial_\theta g_{(\theta,\boldsymbol{\beta})}(\boldsymbol{x})|_{\theta=0} &= \frac{1}{N(\boldsymbol{\beta})}\left(\sum_{i=1}^{p-q} \lambda_i f_{\boldsymbol{\gamma}^i}(\boldsymbol{x}) + \sum_{l=1}^{q} \rho_l f_{\boldsymbol{\gamma}^{0,l}}(\boldsymbol{x}) + \sum_{l=1}^{q} \pi^0_l \sum_{i=1}^{k} \delta^l_i \partial_{\gamma_i} f_{\boldsymbol{\gamma}}(\boldsymbol{x})|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,l}}\right)\\
\partial^2_\theta g_{(\theta,\boldsymbol{\beta})}(\boldsymbol{x})|_{\theta=0} &= \frac{2}{N(\boldsymbol{\beta})^2}\sum_{l=1}^{q}\sum_{i_1=1}^{k} \rho_l \delta^l_{i_1}\left(\partial_{\gamma_{i_1}} f_{\boldsymbol{\gamma}}(\boldsymbol{x})|_{\boldsymbol{\gamma}^{0,l}}\right)\\
&\quad + \frac{1}{N(\boldsymbol{\beta})^2}\sum_{l=1}^{q}\sum_{i_1,i_2=1}^{k} \pi^0_l \delta^l_{i_1}\delta^l_{i_2}\partial_{\gamma_{i_1}\gamma_{i_2}} f_{\boldsymbol{\gamma}}(\boldsymbol{x})|_{\boldsymbol{\gamma}^{0,l}},\\
\partial^3_\theta g_{(\theta,\boldsymbol{\beta})}(\boldsymbol{x})|_{\theta=\theta^*} &= \frac{3}{N(\boldsymbol{\beta})^3}\sum_{l=1}^{q}\sum_{i_1,i_2=1}^{k} \rho_l \delta^l_{i_1}\delta^l_{i_2}\left(\partial_{\gamma_{i_1}\gamma_{i_2}} f_{\boldsymbol{\gamma}}(\boldsymbol{x})|_{\boldsymbol{\gamma}^{0,l}+\frac{\theta^*}{N(\boldsymbol{\beta})}\boldsymbol{\delta}^l}\right)\\
&\quad + \frac{1}{N(\boldsymbol{\beta})^3}\sum_{l=1}^{q}\sum_{i_1,i_2,i_3=1}^{k}\left(\pi^0_l + \rho_l\frac{\theta^*}{N(\boldsymbol{\beta})}\right)\delta^l_{i_1}\delta^l_{i_2}\delta^l_{i_3}\left(\partial_{\gamma_{i_1}\gamma_{i_2}\gamma_{i_3}} f_{\boldsymbol{\gamma}}(\boldsymbol{x})|_{\boldsymbol{\gamma}^{0,l}+\frac{\theta^*}{N(\boldsymbol{\beta})}\boldsymbol{\delta}^l}\right)
\end{aligned}$$

hold (see also formula (A.7) or [DCG99], p. 1194). Let m_2 and m_3 be given according to (P0)c). Then we obtain the remainder term

$$\theta\, \partial^3_\theta g_{(\theta,\boldsymbol{\beta})}(\boldsymbol{x})|_{\theta=\theta^*} \;=\; O\left(\left(\sup_{l\in\{1,\ldots,q\}} \frac{\|\boldsymbol{\delta}^l\|^2}{N(\boldsymbol{\beta})^3}\right)\theta\left(m_2(\boldsymbol{x}) + m_3(\boldsymbol{x})\right)g^0(\boldsymbol{x})\right)$$

since $\theta^* \leq \theta$, $\boldsymbol{\delta}^l \in [-1,1]^k$, $\rho_l \in [-1,1]$ and $\frac{\theta^*}{N(\boldsymbol{\beta})}$ is bounded due to (2.20).

According to [DCG99], p. 1196, we define

$$D_n(\boldsymbol{\beta}) \;=\; \sum_{i=1}^{n} \frac{\partial_\theta g_{(\theta,\boldsymbol{\beta})}(\boldsymbol{x_i})|_{\theta=0}}{g^0(\boldsymbol{x_i})}, \tag{A.1}$$

$$F_n(\boldsymbol{\beta}) \;=\; \sum_{i=1}^{n}\sum_{l=1}^{q}\sum_{j=1}^{k} \rho_l \delta^l_j \frac{\partial_{\gamma_j} f_{\boldsymbol{\gamma}}(\boldsymbol{x_i})|_{\boldsymbol{\gamma}^{0,l}}}{g^0(\boldsymbol{x_i})}, \tag{A.2}$$

$$G_n(\boldsymbol{\beta}) \;=\; \sum_{i=1}^{n}\sum_{l=1}^{q}\sum_{j_1,j_2=1}^{k} \pi^0_l \delta^l_{j_1}\delta^l_{j_2} \frac{\partial_{\gamma_{j_1}\gamma_{j_2}} f_{\boldsymbol{\gamma}}(\boldsymbol{x_i})|_{\boldsymbol{\gamma}^{0,l}}}{g^0(\boldsymbol{x_i})}.$$

Then the first expression of the log-likelihood function $l_n(\theta,\boldsymbol{\beta}) - l_n(0)$ given in (2.42) has the form

$$\sum_{i=1}^{n} \frac{g_{(\theta,\boldsymbol{\beta})}(\boldsymbol{x_i}) - g^0(\boldsymbol{x_i})}{g^0(\boldsymbol{x_i})} = \theta D_n(\boldsymbol{\beta}) + \theta^2 \left\{ \frac{1}{N(\boldsymbol{\beta})^2}\Big(F_n(\boldsymbol{\beta}) + G_n(\boldsymbol{\beta})\Big) + O\left(\left(\sup_{l\in\{1,\dots,q\}} \frac{\|\boldsymbol{\delta^l}\|^2}{N(\boldsymbol{\beta})^3}\right)\theta\Big(m_2(\boldsymbol{x}) + m_3(\boldsymbol{x})\Big)g^0(\boldsymbol{x})\right)\right\}.$$

(Notice that in [DCG99], p. 1196, the multiplication of $O(\ldots)$ is a typo.)

According to Proposition 2.18 $\mathcal{D} = \left\{\frac{\partial_\theta g_{(\theta,\boldsymbol{\beta})}|_{\theta=0}}{g^0} : \boldsymbol{\beta} \in \mathcal{B}\right\}$ is a Donsker class. Due to Definition 2.12 $\frac{1}{\sqrt{n}} D_n(\boldsymbol{\beta})$ converges uniformly in distribution. $\sum\limits_{l=1}^{q}\sum\limits_{j=1}^{k} \rho_l \delta_j^l \frac{\partial_{\gamma_j} f_\gamma|_{\gamma^{0,l}}}{g^0}$ and $\sum\limits_{l=1}^{q}\sum\limits_{j_1,j_2=1}^{k} \pi_l^0 \delta_{j_1}^l \delta_{j_2}^l \frac{\partial_{\gamma_{j_1}\gamma_{j_2}} f_\gamma|_{\gamma^{0,l}}}{g^0}$ can be represented as functions of $\overline{\mathcal{D}}$ by using suitable coefficients (see here, Lemma 2.16, and [DCG99] (10), p. 1186). Thus $\frac{1}{\sqrt{n}} F_n(\boldsymbol{\beta})$ and $\frac{1}{\sqrt{n}} G_n(\boldsymbol{\beta})$ converge uniformly in distribution.

All further statements about the asymptotic behaviour of $l_n(\theta,\boldsymbol{\beta}) - l_n(0)$ given in (2.42) also follow the same lines as given in the proof of Lemma 3.4 in [DCG99]. Roughly speaking, they apply the definition of A_n and η_n to show that

$$\frac{\theta^2}{N(\boldsymbol{\beta})^2} F_n(\boldsymbol{\beta}) = o\Big(\theta D_n(\boldsymbol{\beta})\Big) \qquad \text{and} \qquad \frac{\theta^2}{N(\boldsymbol{\beta})^2} G_n(\boldsymbol{\beta}) = o\Big(\theta D_n(\boldsymbol{\beta})\Big)$$

hold, where the $o(\cdot)$ are uniform in probability over $\boldsymbol{\beta}$ in A_n.

Lemma 2.22 leads to the existence of some $a > 0$ such that for any $l = 1,\dots,q$

$$a^{\frac{1}{3}} \geq \left(\frac{\|\boldsymbol{\delta^l}\|^2}{N(\boldsymbol{\beta})}\right)^{\frac{1}{3}}$$

which is equivalent to

$$a^{\frac{1}{3}} \left(\frac{N(\boldsymbol{\beta})^2}{\|\boldsymbol{\delta^l}\|}\right)^{\frac{2}{3}} \geq N(\boldsymbol{\beta}).$$

Hence, there exists some $a > 0$ such that for any $l = 1,\dots,q$

$$\frac{\|\boldsymbol{\delta^l}\|}{N(\boldsymbol{\beta})} \leq a^{\frac{1}{3}} \left(\frac{N(\boldsymbol{\beta})^2}{\|\boldsymbol{\delta^l}\|}\right)^{\frac{2}{3}} \frac{\|\boldsymbol{\delta^l}\|}{N(\boldsymbol{\beta})^2} = a^{\frac{1}{3}} \left(\frac{\|\boldsymbol{\delta^l}\|}{N(\boldsymbol{\beta})^2}\right)^{\frac{1}{3}}.$$

As consequence we obtain, by using the definition of A_n and $\theta < \eta_n$,

$$\frac{\|\boldsymbol{\delta^l}\|^2}{N(\boldsymbol{\beta})^3}\theta \leq \eta_n^{1-\frac{3}{4}\alpha} a^{\frac{1}{3}} \quad , \; l = 1,\dots,q,$$

which converges to 0 since $\alpha < \frac{3}{4}$. (Notice, that [DCG99], p. 1197, do not mention the constant factor $a^{\frac{1}{3}}$ but it makes no difference to the asymptotic behaviour of $O(\ldots)$.)

Then they maximize $\theta D_n(\boldsymbol{\beta})$ with respect to θ which leads to

$$l_n(\hat{\theta}_{\boldsymbol{\beta}}(n),\boldsymbol{\beta}) - l_n(0) \;=\; \frac{1}{2}\frac{D_n(\boldsymbol{\beta})^2}{n}\mathbb{1}_{D_n(\boldsymbol{\beta})\geq 0}(1+o_P(1))$$

uniform over $\boldsymbol{\beta}$ in A_n. Using again Proposition 2.18, i.e. $\mathcal{D}$ is a Donsker class, and that $\mathcal{D} = \cup_n \mathcal{D}_n$ for $\mathcal{D}_n = \left\{\frac{\partial_\theta g_{(\theta,\boldsymbol{\beta})}|_{\theta=0}}{g^0} : \boldsymbol{\beta}\in A_n\right\}$ holds, we obtain according to [DCG99], p. 1197, that

$$\sup_{\boldsymbol{\beta}\in A_n} l_n(\hat{\theta}_{\boldsymbol{\beta}}(n),\boldsymbol{\beta}) - l_n(0) \;=\; \frac{1}{2}\sup_{d\in\mathcal{D}}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n d(\boldsymbol{X_i})\right)^2 \mathbb{1}_{\frac{1}{\sqrt{n}}\sum_{i=1}^n d(\boldsymbol{X_i})\geq 0}(1+o_P(1)) \quad \text{(A.3)}$$

converges in distribution to $\frac{1}{2}\sup_{d\in\mathcal{D}}(\xi_d)^2\mathbb{1}_{\xi_d\geq 0}$.

2) **On $\boldsymbol{B_n}$,** the normalizing factor $N(\boldsymbol{\beta})$ tends to 0 and according to Remark 2.4 all parameter $\boldsymbol{\delta^l}$, $l=1,\dots,q$, also tend to $\mathbf{0}$. A direct consequence of Lemma 2.22 is

Lemma A.1 *Suppose that at least one of the assumptions* a), b) *or* c) *from Lemma* 2.20 *holds. Then there exists a constant* $M>0$ *such that for* $\boldsymbol{\beta}\in B_n$

$$\begin{aligned} N(\boldsymbol{\beta}) &\leq M\eta_n^{2\alpha/3} \quad \textit{and} \\ |\delta_i^l| &\leq M\eta_n^{\alpha/3} \quad \textit{for any } (i,l)\in\{1,\dots,k\}\times\{1,\dots,q\}. \end{aligned}$$

(Under assumption a) *the statement is equal to Lemma* 5.2 *in* [DCG99]*.)*

The proof of Lemma A.1 is a direct consequence of Lemma 2.22 and is based on the relation $N(\boldsymbol{\beta})\leq a^{\frac{1}{3}}/|\phi_{i,l}|^{\frac{2}{3}}$ for $\phi_{i,l}=\delta_i^l/N(\boldsymbol{\beta})^2$ (see the proof of Lemma 5.2 from [DCG99]).

Consequently, the normalizing factor $N(\boldsymbol{\beta})$ and all $|\delta_i^l|$ tend uniformly to 0 on B_n.
We expand according to [DCG99], p. 1198, $g_{(\theta,\boldsymbol{\beta})}$ up to the order 5 by

$$g_{(\theta,\boldsymbol{\beta})}(\boldsymbol{x}) = g^0(\boldsymbol{x}) + \theta\,\partial_\theta g_{(\theta,\boldsymbol{\beta})}(\boldsymbol{x})|_{\theta=0} + \sum_{i=2}^4 \frac{\theta^i}{i!}\partial_\theta^i g_{(\theta,\boldsymbol{\beta})}(\boldsymbol{x})|_{\theta=0} + \frac{\theta^5}{5!}\,\partial_\theta^5 g_{(\theta,\boldsymbol{\beta})}(\boldsymbol{x})|_{\theta=\theta^*} \quad \text{(A.4)}$$

for some $\theta^*\leq\theta$. Using the notations given in the preprint of Dacunha-Castelle and Gassiat (1996) [DCG96], p. 17, as well as the notations in [DCG99], p. 1198, we define

$$v(i,l;i',l') = \left\langle \frac{\partial_{\gamma_i} f_{\boldsymbol{\gamma}}|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,l}}}{g^0}, \frac{\partial_{\gamma_{i'}} f_{\boldsymbol{\gamma}}|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,l'}}}{g^0}\right\rangle_{L_2(g^0\nu)}, \; i,i'\in\{1,\dots,k\} \text{ and } l,l'\in\{1,\dots,q\},$$

$$d_1(\boldsymbol{\beta}) = \sum_{l=1}^q\sum_{i=1}^k \rho_l\delta_i^l \frac{\partial_{\gamma_i} f_{\boldsymbol{\gamma}}|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,l}}}{g^0}, \quad \text{(A.5)}$$

$$\begin{aligned} U(\boldsymbol{\beta}) &= \sum_{l=1}^q\sum_{i=1}^k \rho_l\delta_i^l \left\langle \frac{\partial_{\gamma_i} f_{\boldsymbol{\gamma}}|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^{0,l}}}{g^0}, \frac{\partial_\theta g_{(\theta,\boldsymbol{\beta})}|_{\theta=0}}{g^0}\right\rangle_{L_2(g^0\nu)} = \left\langle d_1(\boldsymbol{\beta}), \frac{\partial_\theta g_{(\theta,\boldsymbol{\beta})}|_{\theta=0}}{g^0}\right\rangle_{L_2(g^0\nu)}, \\ S(\boldsymbol{\beta}) &= \sum_{l,l'=1}^q\sum_{i,i'=1}^k \rho_l\rho_{l'}\delta_i^l\delta_{i'}^{l'} v(i,l;i',l') \;=\; \|d_1(\boldsymbol{\beta})\|^2_{L_2(g^0\nu)} \end{aligned}$$

and we define the polynomial $P_n(\theta,\boldsymbol{\beta})$ of degree 4 in the variable θ by

$$P_n(\theta,\boldsymbol{\beta}) \;=\; \theta D_n(\boldsymbol{\beta}) - \frac{n\theta^2}{2} + \frac{\theta^2}{N(\boldsymbol{\beta})^2}F_n(\boldsymbol{\beta}) - \frac{n\theta^3}{N(\boldsymbol{\beta})^2}U(\boldsymbol{\beta}) - \frac{n\theta^4}{2N(\boldsymbol{\beta})^4}S(\boldsymbol{\beta}) \;=\; \sum_{j=1}^{4} p_j(n,\boldsymbol{\beta})\theta^j$$

where $D_n(\boldsymbol{\beta})$ is given in (A.1) and $F_n(\boldsymbol{\beta})$ is given in (A.2).

From now on we also use [DCG96] since this preprint is sometimes more detailed than [DCG99] and introduces notations which are missed in [DCG99].

In the following we give an outline of the corresponding proof in [DCG99] who verify that for $\theta \le 2\eta_n$ and for $\boldsymbol{\beta} \in B_n$ one has

$$l_n(\theta,\boldsymbol{\beta}) - l_n(0) \;=\; P_n(\theta,\boldsymbol{\beta})(1+o_P(1)) \tag{A.6}$$

uniform over $\boldsymbol{\beta}$ in B_n. We show that the proof also holds under our modifications (P1 t) as well as (M1) and (M2), respectively.

It results from the representation (2.12) of $g_{(\theta,\boldsymbol{\beta})}$ that for $r=2,\dots,5$ the rth derivative of g with respect to θ is given by

$$\begin{aligned}\partial_\theta^r g_{(\theta,\boldsymbol{\beta})}(\boldsymbol{x}) \;=\;& \frac{r}{N(\boldsymbol{\beta})^r}\sum_{l=1}^{q}\sum_{i_1,\dots,i_{r-1}=1}^{k}\rho_l\delta_{i_1}^l\dots\delta_{i_{r-1}}^l\left(\partial_{\gamma_{i_1}\dots\gamma_{i_{r-1}}}f_{\boldsymbol{\gamma}^{0,l}+\boldsymbol{\delta}^l\frac{\theta}{N(\boldsymbol{\beta})}}(\boldsymbol{x})\right)\\ &+\frac{1}{N(\boldsymbol{\beta})^r}\sum_{l=1}^{q}\sum_{i_1,\dots,i_r=1}^{k}\left(\pi_l^0+\rho_l\frac{\theta}{N(\boldsymbol{\beta})}\right)\delta_{i_1}^l\dots\delta_{i_r}^l\left(\partial_{\gamma_{i_1}\dots\gamma_{i_r}}f_{\boldsymbol{\gamma}^{0,l}+\boldsymbol{\delta}^l\frac{\theta}{N(\boldsymbol{\beta})}}(\boldsymbol{x})\right)\end{aligned} \tag{A.7}$$

(see also [DCG99], p. 1194).

Since $|\delta_i^l|$ converges uniformly to 0 on B_n for $l=1,\dots,q$ it follows that

$$\frac{\delta_{i_1}^l\dots\delta_{i_r}^l}{N(\boldsymbol{\beta})^r} \;=\; o\left(\frac{\delta_{i_1}^l\dots\delta_{i_{r-1}}^l}{N(\boldsymbol{\beta})^r}\right)$$

and for $r=2,\dots,5$ we obtain

$$\partial_\theta^r g_{(\theta,\boldsymbol{\beta})}(\boldsymbol{x})|_{\theta=0} \;=\; \left(\frac{r}{N(\boldsymbol{\beta})^r}\sum_{l=1}^{q}\sum_{i_1,\dots,i_{r-1}=1}^{k}\rho_l\delta_{i_1}^l\dots\delta_{i_{r-1}}^l\left(\partial_{\gamma_{i_1}\dots\gamma_{i_{r-1}}}f_{\boldsymbol{\gamma}}(\boldsymbol{x})|_{\boldsymbol{\gamma}^{0,l}}\right)\right)(1+o(1)).$$

According to [DCG96], p. 19-20, we define

$$E_n(\theta,\boldsymbol{\beta}) \;=\; \sum_{i=1}^{n}\sum_{r=2}^{5}\frac{\theta^r}{(r-1)!}\sum_{l=1}^{q}\sum_{i_1,\dots,i_{r-1}=1}^{k}\rho_l\frac{\delta_{i_1}^l\dots\delta_{i_{r-1}}^l}{N(\boldsymbol{\beta})^r}\frac{\partial_{\gamma_{i_1}\dots\gamma_{i_{r-1}}}f_{\boldsymbol{\gamma}}(\boldsymbol{x_i})|_{\boldsymbol{\gamma}^{0,l}}}{g^0(\boldsymbol{x_i})}, \tag{A.8}$$

$$J_n(\theta,\boldsymbol{\beta}) \;=\; \sum_{i=1}^{n}\sup_{l=1,\dots,q}\frac{\|\boldsymbol{\delta}^l\|^5}{N(\boldsymbol{\beta})^5}\theta^5\frac{m_5(\boldsymbol{x_i})}{g^0(\boldsymbol{x_i})}, \tag{A.9}$$

where the function m_5 is given according to (P0)c). We obtain for the first expression of $l_n(\theta,\boldsymbol{\beta})-l_n(0)$ given in (2.42) that

$$\begin{aligned}&\sum_{i=1}^{n}\frac{g_{(\theta,\boldsymbol{\beta})}(\boldsymbol{x_i})-g^0(\boldsymbol{x_i})}{g^0(\boldsymbol{x_i})}\\ \overset{(A.1),(A.4)}{=}\;& \theta D_n(\boldsymbol{\beta}) + \sum_{i=1}^{n}\frac{1}{g^0(\boldsymbol{x_i})}\left(\sum_{j=2}^{4}\frac{\theta^j}{j!}\partial_\theta^j g_{(\theta,\boldsymbol{\beta})}(\boldsymbol{x_i})|_{\theta=0}+\frac{\theta^5}{5!}\partial_\theta^5 g_{(\theta,\boldsymbol{\beta})}(\boldsymbol{x_i})|_{\theta=\theta^*}\right)\\ \overset{(A.8),(A.9)}{=}\;& \theta D_n(\boldsymbol{\beta}) + E_n(\theta,\boldsymbol{\beta})(1+o(1)) + O(J_n(\theta,\boldsymbol{\beta})).\end{aligned} \tag{A.10}$$

The latter equation is a consequence of (P0)c). Inserting (A.10) in (2.42) leads to

$$\begin{aligned} l_n(\theta,\boldsymbol{\beta}) - l_n(0) &= \theta D_n(\boldsymbol{\beta}) + E_n(\theta,\boldsymbol{\beta}) - \frac{1}{2}\Big(\theta D_n(\boldsymbol{\beta}) + E_n(\theta,\boldsymbol{\beta})\Big)^2 + R_n \\ &= P_n(\theta,\boldsymbol{\beta}) + R_n, \end{aligned}$$

where R_n is a sum of terms which are uniformly $o(Q_n(\theta,\boldsymbol{\beta}))$ in $\boldsymbol{\beta}$ for $Q_n(\theta,\boldsymbol{\beta}) = \sup_{1\le j\le 4} |p_j(n,\boldsymbol{\beta})\theta^j|$. An accurate list of the terms belonging to R_n and their asymptotic bounds according to Lemma 2.22 and Lemma A.1 is given in [DCG96], p. 20 et sqq. (Both Lemmas hold under our modifications.)

We follow the argumentation from [DCG99] who prove, firstly, that

$$\sup_{\substack{\theta\le 2\eta_n \\ \boldsymbol{\beta}\in B_n}} P_n(\theta,\boldsymbol{\beta}) \le \left(\sup_{\boldsymbol{\beta}\in A_n} l_n(\hat{\theta}_{\boldsymbol{\beta}}(n),\boldsymbol{\beta}) - l_n(0)\right)(1+o_P(1)) \tag{A.11}$$

holds. To obtain this result they show that

$$\sup_{\substack{\theta\le 2\eta_n \\ \boldsymbol{\beta}\in B_n}} P_n(\theta,\boldsymbol{\beta}) \le \sup_{\boldsymbol{\beta}\in B_n} Z_n(\boldsymbol{\beta}) \tag{A.12}$$

holds for

$$Z_n(\boldsymbol{\beta}) = \frac{1}{2n}\left(\frac{\left(D_n(\boldsymbol{\beta}) - \frac{F_n(\boldsymbol{\beta})U(\boldsymbol{\beta})}{S(\boldsymbol{\beta})}\right)^2}{1-\frac{U(\boldsymbol{\beta})^2}{S(\boldsymbol{\beta})}}\right)\cdot \mathbb{1}_{D_n(\boldsymbol{\beta})-\frac{F_n(\boldsymbol{\beta})U(\boldsymbol{\beta})}{S(\boldsymbol{\beta})}\ge 0} + \frac{1}{2n}\frac{F_n(\boldsymbol{\beta})^2}{S(\boldsymbol{\beta})}. \tag{A.13}$$

For any real λ and any $\mu \ge 0$ such that $\lambda^2+\mu^2=1$ we define

$$d(\lambda,\mu,\boldsymbol{\beta}) = \lambda\frac{d_1(\boldsymbol{\beta})}{\sqrt{S(\boldsymbol{\beta})}} + \mu\left(\frac{\partial_\theta g_{(\theta,\boldsymbol{\beta})}|_{\theta=0}}{g^0} - \frac{d_1(\boldsymbol{\beta})U(\boldsymbol{\beta})}{S(\boldsymbol{\beta})}\right)\left(1-\frac{U(\boldsymbol{\beta})^2}{S(\boldsymbol{\beta})}\right)^{-\frac{1}{2}}$$

which belongs to $\overline{\mathcal{D}}$ using suitable coefficients (under our modifications, too). (Notice that [DCG99], p. 1200, neither mention the square root $\sqrt{S(\boldsymbol{\beta})}$ of the first term nor the square root of the last factor.) Let

$$V_n(\lambda,\mu,\boldsymbol{\beta}) = \frac{\lambda}{\sqrt{S(\boldsymbol{\beta})}} - \mu\frac{U(\boldsymbol{\beta})}{S(\boldsymbol{\beta})}\left(1-\frac{U(\boldsymbol{\beta})^2}{S(\boldsymbol{\beta})}\right)^{-\frac{1}{2}}, \tag{A.14}$$

$$W_n(\mu,\boldsymbol{\beta}) = \mu\left(1-\frac{U(\boldsymbol{\beta})^2}{S(\boldsymbol{\beta})}\right)^{-\frac{1}{2}}. \tag{A.15}$$

This leads to

$$\begin{aligned}
d(\lambda,\mu,\boldsymbol{\beta}) &= \left(d_1(\boldsymbol{\beta})V_n(\lambda,\mu,\boldsymbol{\beta}) + \frac{\partial_\theta g_{(\theta,\boldsymbol{\beta})}|_{\theta=0}}{g^0} W_n(\mu,\boldsymbol{\beta})\right), \\
\left(\sum_{i=1}^n d(\lambda,\mu,\boldsymbol{\beta})(\boldsymbol{X_i})\right)^2 &= V_n(\lambda,\mu,\boldsymbol{\beta})^2 \sum_{i=1}^n\sum_{j=1}^n d_1(\boldsymbol{\beta})(\boldsymbol{X_i}) d_1(\boldsymbol{\beta})(\boldsymbol{X_j}) \\
&\quad + 2\, V_n(\lambda,\mu,\boldsymbol{\beta})\, W_n(\mu,\boldsymbol{\beta}) \sum_{i=1}^n\sum_{j=1}^n d_1(\boldsymbol{\beta})(\boldsymbol{X_i}) \frac{\partial_\theta g_{(\theta,\boldsymbol{\beta})}(\boldsymbol{X_j})|_{\theta=0}}{g^0(\boldsymbol{X_j})} \\
&\quad + W_n(\mu,\boldsymbol{\beta})^2 \sum_{i=1}^n\sum_{j=1}^n \frac{\partial_\theta g_{(\theta,\boldsymbol{\beta})}(\boldsymbol{X_i})|_{\theta=0}}{g^0(\boldsymbol{X_i})} \frac{\partial_\theta g_{(\theta,\boldsymbol{\beta})}(\boldsymbol{X_j})|_{\theta=0}}{g^0(\boldsymbol{X_j})} \\
&\overset{(A.1),(A.2),(A.5)}{=} \Big(F_n(\boldsymbol{\beta})V_n(\lambda,\mu,\boldsymbol{\beta}) + D_n(\boldsymbol{\beta})W_n(\mu,\boldsymbol{\beta})\Big)^2 \\
&\overset{(A.14),(A.15)}{=} \left\{\lambda \frac{F_n(\boldsymbol{\beta})}{\sqrt{S(\boldsymbol{\beta})}} + \mu\left(D_n(\boldsymbol{\beta}) - \frac{F_n(\boldsymbol{\beta})U(\boldsymbol{\beta})}{S(\boldsymbol{\beta})}\right)\left(1 - \frac{U(\boldsymbol{\beta})^2}{S(\boldsymbol{\beta})}\right)^{-\frac{1}{2}}\right\}^2.
\end{aligned}$$

For $\mathcal{T}_n(\boldsymbol{\beta}) = \frac{F_n(\boldsymbol{\beta})}{\sqrt{S(\boldsymbol{\beta})}}$ and $\mathcal{R}_n(\boldsymbol{\beta}) = \left(D_n(\boldsymbol{\beta}) - \frac{F_n(\boldsymbol{\beta})U(\boldsymbol{\beta})}{S(\boldsymbol{\beta})}\right)\left(1 - \frac{U(\boldsymbol{\beta})^2}{S(\boldsymbol{\beta})}\right)^{-\frac{1}{2}}$ one has

$$\begin{aligned}
&\frac{1}{2} \sup_{\substack{\mu\geq 0 \\ \lambda^2+\mu^2=1}} \left(\frac{1}{\sqrt{n}}\sum_{i=1}^n d(\lambda,\mu,\boldsymbol{\beta})(\boldsymbol{X_i})\right)^2 \cdot \mathbb{1}_{(1/\sqrt{n})\sum_{i=1}^n d(\lambda,\mu,\boldsymbol{\beta})(\boldsymbol{X_i})\geq 0} \\
=\ & \frac{1}{2n} \sup_{\substack{\mu\geq 0 \\ \lambda^2+\mu^2=1}} \Big(\lambda\mathcal{T}_n(\boldsymbol{\beta}) + \mu\mathcal{R}_n(\boldsymbol{\beta})\Big)^2 \cdot \mathbb{1}_{(1/\sqrt{n})(\lambda\mathcal{T}_n(\boldsymbol{\beta})+\mu\mathcal{R}_n(\boldsymbol{\beta}))\geq 0} \\
=\ & \frac{1}{2n} \sup_{\substack{\mu\geq 0 \\ \lambda^2+\mu^2=1}} \Big(\|(\lambda,\mu)'\|\, \|(\mathcal{T}_n(\boldsymbol{\beta}),\mathcal{R}_n(\boldsymbol{\beta}))'\| \cos\Big[\angle\Big\{(\lambda,\mu)', (\mathcal{T}_n(\boldsymbol{\beta}),\mathcal{R}_n(\boldsymbol{\beta}))'\Big\}\Big]\Big)^2 \qquad \text{(A.16)} \\
& \cdot \mathbb{1}_{(1/\sqrt{n})(\lambda\mathcal{T}_n(\boldsymbol{\beta})+\mu\mathcal{R}_n(\boldsymbol{\beta}))\geq 0},
\end{aligned}$$

where $\angle(a,b)$ denotes the angle between a and b.

Suppose that there exists some $\boldsymbol{\beta} \in \mathcal{B}$ such that $\mathcal{R}_n(\boldsymbol{\beta}) = 0$. Then one has $D_n(\boldsymbol{\beta}) = \frac{F_n(\boldsymbol{\beta})U(\boldsymbol{\beta})}{S(\boldsymbol{\beta})}$. Due to (A1) and (P1 t) it follows $\lambda_i = 0$, $i = 1,\dots,p-q$ and $\rho_l = 0$, $l = 1,\dots,q$ (see (A.1) and (A.2)). According to (2.6) there exists some $l \in \{1,\dots,q\}$ and $i \in \{1,\dots,k\}$ such that $\delta_i^l \neq 0$ which is in contradiction to $\mathcal{R}_n(\boldsymbol{\beta}) = 0$ thanks to (A1) and (P1 t). Thus we have $\mathcal{R}_n(\boldsymbol{\beta}) \neq 0$ for all $\boldsymbol{\beta} \in \mathcal{B}$.

If $\mathcal{R}_n(\boldsymbol{\beta}) > 0$ then the supremum on the right hand side of (A.16) is in $(\lambda,\mu)' = \|(\mathcal{T}_n(\boldsymbol{\beta}),\mathcal{R}_n(\boldsymbol{\beta}))'\|^{-1}(\mathcal{T}_n(\boldsymbol{\beta}),\mathcal{R}_n(\boldsymbol{\beta}))'$ and it follows

$$\begin{aligned}
&\frac{1}{2n} \sup_{\substack{\mu\geq 0 \\ \lambda^2+\mu^2=1}} \Big(\lambda\mathcal{T}_n(\boldsymbol{\beta}) + \mu\mathcal{R}_n(\boldsymbol{\beta})\Big)^2 \cdot \mathbb{1}_{(1/\sqrt{n})(\lambda\mathcal{T}_n(\boldsymbol{\beta})+\mu\mathcal{R}_n(\boldsymbol{\beta}))\geq 0} \\
=\ & \frac{1}{2n}\Big(\mathcal{T}_n(\boldsymbol{\beta})\Big)^2 + \frac{1}{2n}\Big(\mathcal{R}_n(\boldsymbol{\beta})\Big)^2 \cdot \mathbb{1}_{\mu\mathcal{R}_n(\boldsymbol{\beta})\geq 0}. \qquad \text{(A.17)}
\end{aligned}$$

If $\mathcal{R}_n(\boldsymbol{\beta}) < 0$ then the supremum on the right hand side of (A.16) will be reached for $\mu = 0$ and $\lambda = 1$ or $\lambda = -1$ whichever $\mathcal{T}_n(\boldsymbol{\beta})$ is positive or negative. Thus (A.17) also holds in this case. As the right hand side of (A.17) is equal to the right hand side of (A.13) we obtain

$$Z_n(\boldsymbol{\beta}) = \frac{1}{2} \sup_{\substack{\mu \geq 0 \\ \lambda^2+\mu^2=1}} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n d(\lambda, \mu, \boldsymbol{\beta})(\boldsymbol{X_i}) \right)^2 \cdot \mathbb{1}_{(1/\sqrt{n}) \sum_{i=1}^n d(\lambda,\mu,\boldsymbol{\beta})(\boldsymbol{X_i}) \geq 0}.$$

As mentioned above $d(\lambda, \mu, \boldsymbol{\beta})$ belongs to $\overline{\mathcal{D}}$ and it follows

$$\begin{aligned} \sup_{\substack{\theta \leq 2\eta_n \\ \boldsymbol{\beta} \in B_n}} P_n(\theta, \boldsymbol{\beta}) &\overset{(A.12)}{\leq} \frac{1}{2} \sup_{d \in \mathcal{D}} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n d(\boldsymbol{X_i}) \right)^2 \cdot \mathbb{1}_{(1/\sqrt{n}) \sum_{i=1}^n d(\boldsymbol{X_i}) \geq 0} \\ &\overset{(A.3)}{=} \left(\sup_{\boldsymbol{\beta} \in A_n} l_n(\hat{\theta}_{\boldsymbol{\beta}}(n), \boldsymbol{\beta}) - l_n(0) \right) (1 + o_P(1)). \end{aligned}$$

Thus statement (A.11) holds.

Then [DCG99], p. 1200 et sqq., prove that if at least one term $p_j(n, \boldsymbol{\beta})\theta^j$ of P_n (see p. 103) tends to $+\infty$ it follows that $P_n(\theta, \boldsymbol{\beta})$ tends to $-\infty$, and that all terms have the same order and $R_n = o(P_n(\hat{\theta}, \hat{\boldsymbol{\beta}}))$ at the optimizing value $(\hat{\theta}, \hat{\boldsymbol{\beta}})$. They verify this with the aid of the following technical

Lemma A.2 *Suppose that at least one of the assumptions* a), b) *or* c) *from Lemma* 2.20 *holds. There exists* $\tau > 0$ *such that*

$$\sum_{l,l'=1}^q \sum_{i,i'=1}^k \rho_l \rho_{l'} \phi_{i,l} \phi_{i',l'} n\, v(i, l; i', l') \geq \tau \sum_{l=1}^q \sum_{i=1}^k (\rho_l \phi_{i,l})^2 n\, v(i, l; i, l) \tag{A.18}$$

with $\phi_{i,l} = \delta_i^l / (N(\boldsymbol{\beta}))^2$.

(Under assumption a) the statement is equal to Lemma 5.3 *in* [DCG99].*)*

Proof of Lemma A.2: Lemma A.2 is a simple consequence of (P1 t) which belongs any assumption a), b) and c). According to the definitions given on p. 102 the left hand side of (A.18) is equal to

$$\frac{nS(\boldsymbol{\beta})}{N(\boldsymbol{\beta})^2} = \frac{n \|d_1(\boldsymbol{\beta})\|^2_{L_2(g^0\nu)}}{N(\boldsymbol{\beta})^2} = \frac{n}{N(\boldsymbol{\beta})^2} \left\| \sum_{l=1}^q \sum_{i=1}^k \rho_l \delta_i^l \frac{\partial_{\gamma_i} f_\gamma|_{\gamma=\gamma^{0,l}}}{g^0} \right\|^2_{L_2(g^0\nu)}$$

and right hand side of (A.18) is equal to

$$\tau \frac{n}{N(\boldsymbol{\beta})^2} \sum_{l=1}^q \sum_{i=1}^k (\rho_l \delta_i^l)^2 \left\| \frac{\partial_{\gamma_i} f_\gamma|_{\gamma=\gamma^{0,l}}}{g^0} \right\|^2_{L_2(g^0\nu)}.$$

According to (P1 t) the functions $\frac{\partial_{\gamma_i} f_\gamma|_{\gamma=\gamma^{0,l}}}{g^0}$, $i = 1, \ldots, k$ $l = 1, \ldots, q$, are linearly independent in $L_2(g^0\nu)$. Thus the left hand side of (A.18) is null iff all $(\rho_l \delta_i^l) = 0$ which implies that the right hand side of (A.18) is also null. For sufficient small $\tau > 0$ (A.18) holds. □

Finally, we conclude according to [DCG99] that, in the neighbourhood of $(\hat{\theta}, \hat{\boldsymbol{\beta}})$ formula (A.6) holds. Hence, the assertion follows from (A.11). □

A.2. Proof of Lemma 3.7 via induction on n :

According to (3.5) we have $\tilde{p}_{\boldsymbol{\zeta}}(\boldsymbol{x}) = \mathcal{C}(\boldsymbol{\zeta}) \exp\Big\{\big\langle \boldsymbol{\zeta}, \boldsymbol{T}(\boldsymbol{x})\big\rangle\Big\}$. For $i \in \{1,\dots,k\}$ (3.8) leads to

$$\partial_{\zeta_i}\mathcal{C}(\boldsymbol{\zeta}) \quad = \quad -\Big(\partial_{\zeta_i}K(\boldsymbol{\zeta})\Big)\mathcal{C}(\boldsymbol{\zeta}) \quad = \quad -\Big(E_{\boldsymbol{\zeta}}T_i\Big)\mathcal{C}(\boldsymbol{\zeta}). \tag{A.19}$$

$n=1$: For $i_1 \in \{1,\dots,k\}$, $A^1 = \{1\}$, $a_{(\emptyset)} = 1$ and $a_{(\{1\})} = -1$ we obtain

$$\begin{aligned}
\partial_{\zeta_{i_1}}\tilde{p}_{\boldsymbol{\zeta}}(\boldsymbol{x}) &\overset{(3.8),(A.19)}{=} \tilde{p}_{\boldsymbol{\zeta}}(\boldsymbol{x})\Big(T_{i_1}(\boldsymbol{x}) - E_{\boldsymbol{\zeta}}T_{i_1}\Big) \\
&= \tilde{p}_{\boldsymbol{\zeta}}(\boldsymbol{x})\Big\{\Big(E_{\boldsymbol{\zeta}}\Big(\prod_{j\in\emptyset}T_{i_j}\Big)\Big)\prod_{j\in\{1\}}T_{i_j}(\boldsymbol{x}) - \Big(E_{\boldsymbol{\zeta}}\Big(\prod_{j\in\{1\}}T_{i_j}\Big)\Big)\prod_{j\in\emptyset}T_{i_j}(\boldsymbol{x})\Big\} \\
&= \tilde{p}_{\boldsymbol{\zeta}}(\boldsymbol{x})\sum_{B_1\subset A^1}\Big\{a_{(B_1)}\Big(\prod_{l=1}^{1}E_{\boldsymbol{\zeta}}\Big(\prod_{j\in B_l}T_{i_j}\Big)\Big)\prod_{j\in A^1\setminus B_1}T_{i_j}(\boldsymbol{x})\Big\}.
\end{aligned} \tag{A.20}$$

$n \to n+1$: For $i_1,\dots,i_{n+1} \in \{1,\dots,k\}$ the induction assumption leads to

$$\partial_{\zeta_{i_1}}\dots\partial_{\zeta_{i_{n+1}}}\tilde{p}_{\boldsymbol{\zeta}}(\boldsymbol{x}) = \partial_{\zeta_{i_{n+1}}}\tilde{p}_{\boldsymbol{\zeta}}(\boldsymbol{x})\sum_{\substack{B_1,\dots,B_n\subset A^n\\ \text{disjoint}}}\Big\{a_{(B_1,\dots,B_n)}\Big(\prod_{l=1}^{n}E_{\boldsymbol{\zeta}}\Big(\prod_{j\in B_l}T_{i_j}\Big)\Big)\prod_{j\in A^n\setminus\big(\bigcup_{t=1}^{n}B_t\big)}T_{i_j}(\boldsymbol{x})\Big\}, \tag{A.21}$$

where $A^n = \{1,\dots,n\}$ and $a_{(B_1,\dots,B_n)} \in I\!R$ are suitable factors for disjoint $B_1,\dots,B_n \subset A^n$.

For $l \in \{1,\dots,k\}$ and $l_1,\dots,l_k \in I\!N$ we have

$$\begin{aligned}
&\partial_{\zeta_l}E_{\boldsymbol{\zeta}}T_1^{l_1}\dots T_k^{l_k} \\
&\overset{(3.10)}{=} \Big(\partial_{\zeta_l}\mathcal{C}(\boldsymbol{\zeta})\Big)\partial_{\zeta_1}^{l_1}\dots\partial_{\zeta_k}^{l_k}\int\exp\Big(\big\langle\boldsymbol{\zeta},\boldsymbol{T}(\boldsymbol{x})\big\rangle\Big)\,\nu(d\boldsymbol{x}) \\
&\quad + \mathcal{C}(\boldsymbol{\zeta})\partial_{\zeta_l}\partial_{\zeta_1}^{l_1}\dots\partial_{\zeta_k}^{l_k}\int\exp\Big(\big\langle\boldsymbol{\zeta},\boldsymbol{T}(\boldsymbol{x})\big\rangle\Big)\,\nu(d\boldsymbol{x}) \\
&\overset{(A.19),(3.10)}{=} -\Big(E_{\boldsymbol{\zeta}}T_l\Big)\mathcal{C}(\boldsymbol{\zeta})\partial_{\zeta_1}^{l_1}\dots\partial_{\zeta_k}^{l_k}\int\exp\Big(\big\langle\boldsymbol{\zeta},\boldsymbol{T}(\boldsymbol{x})\big\rangle\Big)\,\nu(d\boldsymbol{x}) + E_{\boldsymbol{\zeta}}T_1^{l_1}\dots T_k^{l_k}T_l \\
&\overset{(3.10)}{=} E_{\boldsymbol{\zeta}}T_1^{l_1}\dots T_k^{l_k}T_l - (E_{\boldsymbol{\zeta}}T_l)\;E_{\boldsymbol{\zeta}}T_1^{l_1}\dots T_k^{l_k}.
\end{aligned} \tag{A.22}$$

As a consequence we obtain for any $i_1,\dots,i_{n+1} \in \{1,\dots,k\}$ and any $B_r \subset \{1,\dots,n\}$

$$\partial_{\zeta_{i_{n+1}}}E_{\boldsymbol{\zeta}}\Big(\prod_{j\in B_r}T_{i_j}\Big) \overset{(A.22)}{=} E_{\boldsymbol{\zeta}}\Big(\prod_{j\in B_r\cup\{n+1\}}T_{i_j}\Big) - \Big(E_{\boldsymbol{\zeta}}T_{i_{n+1}}\Big)E_{\boldsymbol{\zeta}}\Big(\prod_{j\in B_r}T_{i_j}\Big). \tag{A.23}$$

Notice that if $B_r = \emptyset$ then both sides of the latter equation are 0.

Using the product rule and $\partial_{\zeta_{i_{n+1}}}\tilde{p}_{\boldsymbol{\zeta}}(\boldsymbol{x}) \overset{(A.20)}{=} \tilde{p}_{\boldsymbol{\zeta}}(\boldsymbol{x})\big(T_{i_{n+1}}(\boldsymbol{x}) - E_{\boldsymbol{\zeta}}T_{i_{n+1}}\big)$ as well as (A.23) the right hand side of (A.21) results in

$$\tilde{p}_{\boldsymbol{\zeta}}(\boldsymbol{x})\Bigg[\Big(T_{i_{n+1}}(\boldsymbol{x}) - E_{\boldsymbol{\zeta}}T_{i_{n+1}}\Big)\sum_{\substack{B_1,\ldots,B_n\subset A^n\\ \text{disjoint}}}\Big\{a_{(B_1,\ldots,B_n)}\Big(\prod_{l=1}^{n}E_{\boldsymbol{\zeta}}\Big(\prod_{j\in B_l}T_{i_j}\Big)\Big)\prod_{j\in A^n\setminus\big(\bigcup_{t=1}^{n}B_t\big)}T_{i_j}(\boldsymbol{x})\Big\} \tag{A.24}$$

$$+\sum_{\substack{B_1,\ldots,B_n\subset A^n\\ \text{disjoint}}}\Big\{a_{(B_1,\ldots,B_n)}\sum_{l=1}^{n}\Big[E_{\boldsymbol{\zeta}}\Big(\prod_{j\in B_l\cup\{n+1\}}T_{i_j}\Big)\Big(\prod_{\substack{r=1\\ r\neq l}}^{n}E_{\boldsymbol{\zeta}}\Big(\prod_{j\in B_r}T_{i_j}\Big)\Big)\Big]\prod_{j\in A^n\setminus\big(\bigcup_{t=1}^{n}B_t\big)}T_{i_j}(\boldsymbol{x})\Big\} \tag{A.25}$$

$$-\sum_{\substack{B_1,\ldots,B_n\subset A^n\\ \text{disjoint}}}\Big\{a_{(B_1,\ldots,B_n)}\sum_{l=1}^{n}\Big[E_{\boldsymbol{\zeta}}T_{i_{n+1}}E_{\boldsymbol{\zeta}}\Big(\prod_{j\in B_l}T_{i_j}\Big)\Big(\prod_{\substack{r=1\\ r\neq l}}^{n}E_{\boldsymbol{\zeta}}\Big(\prod_{j\in B_r}T_{i_j}\Big)\Big)\Big]\prod_{j\in A^n\setminus\big(\bigcup_{t=1}^{n}B_t\big)}T_{i_j}(\boldsymbol{x})\Big\}\Bigg]. \tag{A.26}$$

For $A^{n+1} = \{1,\ldots,n+1\}$ the factor $[\ldots]$ of $\tilde{p}_{\boldsymbol{\zeta}}(\boldsymbol{x})$ can be written as

$$\sum_{\substack{B_1,\ldots,B_n\subset A^n\\ \text{disjoint}\\ B_{n+1}=\emptyset}}\Big\{a_{(B_1,\ldots,B_n)}\Big(\prod_{l=1}^{n+1}E_{\boldsymbol{\zeta}}\Big(\prod_{j\in B_l}T_{i_j}\Big)\Big)\prod_{j\in A^{n+1}\setminus\big(\bigcup_{t=1}^{n+1}B_t\big)}T_{i_j}(\boldsymbol{x})\Big\} -$$

$$\sum_{\substack{B_1,\ldots,B_n\subset A^n\\ \text{disjoint}\\ B_{n+1}=\{n+1\}}}\Big\{a_{(B_1,\ldots,B_n)}E_{\boldsymbol{\zeta}}\Big(\prod_{j\in B_{n+1}}T_{i_j}\Big)\Big(\prod_{\substack{r=1\\ r\neq n+1}}^{n+1}E_{\boldsymbol{\zeta}}\Big(\prod_{j\in B_r}T_{i_j}\Big)\Big)\prod_{j\in A^{n+1}\setminus\big(\bigcup_{t=1}^{n+1}B_t\big)}T_{i_j}(\boldsymbol{x})\Big\} +$$

$$\sum_{\substack{B_1,\ldots,B_n\subset A^n\\ \text{disjoint}\\ B_{n+1}=\{n+1\}}}\Big\{a_{(B_1,\ldots,B_n)}\sum_{l=1}^{n}\Big[E_{\boldsymbol{\zeta}}\Big(\prod_{j\in B_l\cup B_{n+1}}T_{i_j}\Big)\Big(\prod_{\substack{r=1\\ r\neq\{l,n+1\}}}^{n+1}E_{\boldsymbol{\zeta}}\Big(\prod_{j\in B_r}T_{i_j}\Big)\Big)\Big]\prod_{j\in A^{n+1}\setminus\big(\bigcup_{t=1}^{n+1}B_t\big)}T_{i_j}(\boldsymbol{x})\Big\} -$$

$$\sum_{\substack{B_1,\ldots,B_n\subset A^n\\ \text{disjoint}\\ B_{n+1}=\{n+1\}}}\Big\{a_{(B_1,\ldots,B_n)}\sum_{l=1}^{n}\Big[E_{\boldsymbol{\zeta}}\Big(\prod_{j\in B_{n+1}}T_{i_j}\Big)E_{\boldsymbol{\zeta}}\Big(\prod_{j\in B_l}T_{i_j}\Big)\Big(\prod_{\substack{r=1\\ r\neq\{l,n+1\}}}^{n+1}E_{\boldsymbol{\zeta}}\Big(\prod_{j\in B_r}T_{i_j}\Big)\Big)\Big]\prod_{j\in A^{n+1}\setminus\big(\bigcup_{t=1}^{n+1}B_t\big)}T_{i_j}(\boldsymbol{x})\Big\}$$

where the both first sums are equal to (A.24), the third sum is equal to (A.25) and the fourth sum is equal to (A.26).

Furthermore, one has

$$\Big\{\Big(\prod_{l=1}^{n+1}E_{\zeta}\Big(\prod_{j\in B_l}T_{i_j}\Big)\Big)\ :\ B_1,\ldots,B_n\subset A^n \text{ disjoint},\ B_{n+1}=\emptyset\Big\}$$

$$= \Big\{\Big(\prod_{l=1}^{n+1}E_{\zeta}\Big(\prod_{j\in B_l}T_{i_j}\Big)\Big)\ :\ B_1,\ldots,B_{n+1}\subset A^{n+1} \text{ disjoint},\ n+1\notin B_1\cup\ldots\cup B_{n+1}\Big\} \tag{A.27}$$

and

$$\Big\{E_{\boldsymbol{\zeta}}\Big(\prod_{j\in B_{n+1}} T_{i_j}\Big)\Big(\prod_{\substack{r=1\\ r\neq n+1}}^{n+1} E_{\boldsymbol{\zeta}}\Big(\prod_{j\in B_r} T_{i_j}\Big)\Big),\ E_{\boldsymbol{\zeta}}\Big(\prod_{j\in B_l\cup B_{n+1}} T_{i_j}\Big)\Big(\prod_{\substack{r=1\\ r\neq\{l,n+1\}}}^{n+1} E_{\boldsymbol{\zeta}}\Big(\prod_{j\in B_r} T_{i_j}\Big)\Big),$$

$$E_{\boldsymbol{\zeta}}\Big(\prod_{j\in B_{n+1}} T_{i_j}\Big)E_{\boldsymbol{\zeta}}\Big(\prod_{j\in B_l} T_{i_j}\Big)\Big(\prod_{\substack{r=1\\ r\neq\{l,n+1\}}}^{n+1} E_{\boldsymbol{\zeta}}\Big(\prod_{j\in B_r} T_{i_j}\Big)\Big),\ l=1,\dots,n:$$

$$B_1,\dots,B_n\subset A^n \text{ disjoint},\ B_{n+1}=\{n+1\}\Big\}$$

$$= \Big\{\Big(\prod_{r=1}^{n+1} E_{\boldsymbol{\zeta}}\Big(\prod_{j\in B_r} T_{i_j}\Big)\Big):\ B_1,\dots,B_{n+1}\subset A^{n+1} \text{ disjoint},\ n+1\in B_{n+1}\Big\}. \tag{A.28}$$

The union of (A.27) and (A.28) results in

$$\Big\{\Big(\prod_{r=1}^{n+1} E_{\boldsymbol{\zeta}}\Big(\prod_{j\in B_r} T_{i_j}\Big)\Big): B_1,\dots,B_{n+1}\subset A^{n+1} \text{ disjoint}\Big\}.$$

Hence, there exist suitable coefficients $a_{(B_1,\dots,B_{n+1})}\in I\!R$ such that

$$\begin{aligned}&\partial_{\zeta_{i_1}}\dots\partial_{\zeta_{i_{n+1}}}\tilde p_{\boldsymbol{\zeta}}(\boldsymbol{x})\\ =&\sum_{\substack{B_1,\dots,B_{n+1}\subset A^{n+1}\\ \text{disjoint}}}\Big\{a_{(B_1,\dots,B_{n+1})}\Big(\prod_{l=1}^{n+1} E_{\boldsymbol{\zeta}}\Big(\prod_{j\in B_l} T_{i_j}\Big)\Big)\prod_{j\in A^{n+1}\setminus\big(\bigcup_{t=1}^{n+1} B_t\big)} T_{i_j}(\boldsymbol{x})\Big\}.\end{aligned}$$

□

A.3. Proof of Lemma 3.9 via induction on n :

For any $n\in I\!N$ and $i_1,\dots,i_n\in\{1,\dots,k\}$ we have due to (3.12)

$$\partial_{\gamma_{i_1}}\dots\partial_{\gamma_{i_n}} f_{\boldsymbol{\gamma}}(\boldsymbol{x}) = \partial_{\gamma_{i_1}}\dots\partial_{\gamma_{i_n}} r(\boldsymbol{x})\tilde p_{\boldsymbol{\zeta}(\boldsymbol{\gamma})}(\boldsymbol{x}).$$

Since $r(\boldsymbol{x})$ does not depend on $\boldsymbol{\gamma}$ we consider $\partial_{\gamma_{i_1}}\dots\partial_{\gamma_{i_n}}\tilde p_{\boldsymbol{\zeta}(\boldsymbol{\gamma})}(\boldsymbol{x})$.

$n=1$: For $i_1\in\{1,\dots,k\}$, $A^1=\{1\}$ and $b_{(\{1\})}=1$ we obtain

$$\begin{aligned}\partial_{\gamma_{i_1}}\tilde p_{\boldsymbol{\zeta}(\boldsymbol{\gamma})}(\boldsymbol{x}) &= \Big(\partial_{\zeta_1(\boldsymbol{\gamma})}\tilde p_{\boldsymbol{\zeta}(\boldsymbol{\gamma})}(\boldsymbol{x}),\dots,\partial_{\zeta_k(\boldsymbol{\gamma})}\tilde p_{\boldsymbol{\zeta}(\boldsymbol{\gamma})}(\boldsymbol{x})\Big)\Big(\partial_{\gamma_{i_1}}\zeta_1(\boldsymbol{\gamma}),\dots,\partial_{\gamma_{i_1}}\zeta_k(\boldsymbol{\gamma})\Big)'\\ &= \sum_{m=1}^{1}\sum_{j_m=1}^{k}\Big\{\Big(\partial_{\zeta_{j_m}(\boldsymbol{\gamma})}\tilde p_{\boldsymbol{\zeta}(\boldsymbol{\gamma})}(\boldsymbol{x})\Big)\sum_{A^1_m=\{1\}}\Big(b_{(A^1_m)}\prod_{i=1}^{m}\partial_{A^1_i}\zeta_{j_i}(\boldsymbol{\gamma})\Big)\Big\}.\end{aligned}$$

$n \to n+1$: For $i_1, \ldots, i_{n+1} \in \{1, \ldots, k\}$ the induction assumption implies that

$$\begin{aligned}
&\partial_{\gamma_{i_1}} \ldots \partial_{\gamma_{i_{n+1}}} \tilde{p}_{\boldsymbol{\zeta}(\boldsymbol{\gamma})}(\boldsymbol{x}) \\
= \;& \partial_{\gamma_{i_{n+1}}} \sum_{m=1}^{n} \sum_{j_1=1}^{k} \cdots \sum_{j_m=1}^{k} \Bigg\{ \Big(\partial_{\zeta_{j_1}(\boldsymbol{\gamma})} \ldots \partial_{\zeta_{j_m}(\boldsymbol{\gamma})} \tilde{p}_{\boldsymbol{\zeta}(\boldsymbol{\gamma})}(\boldsymbol{x}) \Big) \sum_{\substack{A_1^n, \ldots, A_m^n \neq \emptyset \\ \text{disjoint} \\ A_1^n \cup \ldots \cup A_m^n = A^n}} \Big(b_{(A_1^n, \ldots, A_m^n)} \prod_{i=1}^{m} \partial_{A_i^n} \zeta_{j_i}(\boldsymbol{\gamma}) \Big) \Bigg\} \\
= \;& \sum_{m=1}^{n} \sum_{j_1=1}^{k} \cdots \sum_{j_m=1}^{k} \Bigg\{ \Big(\partial_{\zeta_{j_1}(\boldsymbol{\gamma})} \ldots \partial_{\zeta_{j_m}(\boldsymbol{\gamma})} \sum_{j_{m+1}=1}^{k} \Big(\partial_{\zeta_{j_{m+1}}(\boldsymbol{\gamma})} \tilde{p}_{\boldsymbol{\zeta}(\boldsymbol{\gamma})}(\boldsymbol{x}) \Big) \Big(\partial_{\gamma_{i_{n+1}}} \zeta_{j_{m+1}}(\boldsymbol{\gamma}) \Big) \Big) \\
& \qquad \sum_{\substack{A_1^n, \ldots, A_m^n \neq \emptyset \\ \text{disjoint} \\ A_1^n \cup \ldots \cup A_m^n = A^n}} \Big(b_{(A_1^n, \ldots, A_m^n)} \prod_{i=1}^{m} \partial_{A_i^n} \zeta_{j_i}(\boldsymbol{\gamma}) \Big) \Bigg\}
\end{aligned} \tag{A.29}$$

$$+ \sum_{m=1}^{n} \sum_{j_1=1}^{k} \cdots \sum_{j_m=1}^{k} \Bigg\{ \Big(\partial_{\zeta_{j_1}(\boldsymbol{\gamma})} \ldots \partial_{\zeta_{j_m}(\boldsymbol{\gamma})} \tilde{p}_{\boldsymbol{\zeta}(\boldsymbol{\gamma})}(\boldsymbol{x}) \Big) \sum_{\substack{A_1^n, \ldots, A_m^n \neq \emptyset \\ \text{disjoint} \\ A_1^n \cup \ldots \cup A_m^n = A^n}} \Big(b_{(A_1^n, \ldots, A_m^n)} \partial_{\gamma_{i_{n+1}}} \prod_{i=1}^{m} \partial_{A_i^n} \zeta_{j_i}(\boldsymbol{\gamma}) \Big) \Bigg\}. \tag{A.30}$$

Calculation of (A.29): For $A^{n+1} = \{1, \ldots, n+1\}$ and $\tilde{b}_{(A_1^{n+1}, \ldots, A_{m+1}^{n+1})} = b_{(A_1^n, \ldots, A_m^n)}$ the sum (A.29) is equal to

$$\begin{aligned}
&\sum_{m=1}^{n} \sum_{j_1=1}^{k} \cdots \sum_{j_{m+1}=1}^{k} \Bigg\{ \Big(\partial_{\zeta_{j_1}(\boldsymbol{\gamma})} \ldots \partial_{\zeta_{j_m}(\boldsymbol{\gamma})} \partial_{\zeta_{j_{m+1}}(\boldsymbol{\gamma})} \tilde{p}_{\boldsymbol{\zeta}(\boldsymbol{\gamma})}(\boldsymbol{x}) \Big) \\
& \qquad \sum_{\substack{A_1^{n+1}, \ldots, A_{m+1}^{n+1} \neq \emptyset \\ \text{disjoint}, A_{m+1}^{n+1} = \{n+1\}, \\ A_1^{n+1} \cup \ldots \cup A_{m+1}^{n+1} = A^{n+1}}} \Big(\tilde{b}_{(A_1^{n+1}, \ldots, A_{m+1}^{n+1})} \prod_{i=1}^{m+1} \partial_{A_i^{n+1}} \zeta_{j_i}(\boldsymbol{\gamma}) \Big) \Bigg\}.
\end{aligned}$$

A shift of index $m+1$ to m implies that the latter expression is equal to

$$\sum_{m=2}^{n+1} \sum_{j_1=1}^{k} \cdots \sum_{j_m=1}^{k} \Bigg\{ \Big(\partial_{\zeta_{j_1}(\boldsymbol{\gamma})} \ldots \partial_{\zeta_{j_m}(\boldsymbol{\gamma})} \tilde{p}_{\boldsymbol{\zeta}(\boldsymbol{\gamma})}(\boldsymbol{x}) \Big) \sum_{\substack{A_1^{n+1}, \ldots, A_m^{n+1} \neq \emptyset \\ \text{disjoint}, A_m^{n+1} = \{n+1\}, \\ A_1^{n+1} \cup \ldots \cup A_m^{n+1} = A^{n+1}}} \Big(\tilde{b}_{(A_1^{n+1}, \ldots, A_m^{n+1})} \prod_{i=1}^{m} \partial_{A_i^{n+1}} \zeta_{j_i}(\boldsymbol{\gamma}) \Big) \Bigg\}. \tag{A.31}$$

For suitable $b^*_{(A_1^{n+1}, \ldots, A_m^{n+1})} \in \mathbb{R}$ (which is replaced as needed by 0) the sum (A.31) is equal to

$$\sum_{m=2}^{n+1} \sum_{j_1=1}^{k} \cdots \sum_{j_m=1}^{k} \Bigg\{ \Big(\partial_{\zeta_{j_1}(\boldsymbol{\gamma})} \ldots \partial_{\zeta_{j_m}(\boldsymbol{\gamma})} \tilde{p}_{\boldsymbol{\zeta}(\boldsymbol{\gamma})}(\boldsymbol{x}) \Big) \sum_{\substack{A_1^{n+1}, \ldots, A_m^{n+1} \neq \emptyset \\ \text{disjoint} \\ A_1^{n+1} \cup \ldots \cup A_m^{n+1} = A^{n+1}}} \Big(b^*_{(A_1^{n+1}, \ldots, A_m^{n+1})} \prod_{i=1}^{m} \partial_{A_i^{n+1}} \zeta_{j_i}(\boldsymbol{\gamma}) \Big) \Bigg\}. \tag{A.32}$$

Calculation of (A.30): We apply the product rule to (A.30). This results in

$$\sum_{m=1}^{n}\sum_{j_1=1}^{k}\cdots\sum_{j_m=1}^{k}\Bigg\{\Big(\partial_{\zeta_{j_1}(\gamma)}\cdots\partial_{\zeta_{j_m}(\gamma)}\tilde{p}_{\boldsymbol{\zeta}(\boldsymbol{\gamma})}(\boldsymbol{x})\Big) \underbrace{\sum_{\substack{A_1^n,\dots,A_m^n\neq\emptyset\\ \text{disjoint}\\ A_1^n\cup\dots\cup A_m^n=A^n}}\Bigg(b_{(A_1^n,\dots,A_m^n)}\sum_{i=1}^{m}\Bigg[\Big(\partial_{A_i^n\cup\{n+1\}}\zeta_{j_i}(\boldsymbol{\gamma})\Big)\prod_{\substack{l=1\\ l\neq i}}^{m}\partial_{A_l^n}\zeta_{j_l}(\boldsymbol{\gamma})\Bigg]\Bigg)}_{(*)}\Bigg\}.$$

Thereby, $(*)$ has a representation

$$\sum_{i=1}^{m}\sum_{\substack{A_j^{n+1}=A_j^n\\ j=1,\dots,m,j\neq i,\\ A_i^{n+1}=A_i^n\cup\{n+1\}}}\Bigg(\tilde{b}_{(A_1^{n+1},\dots,A_m^{n+1})}\prod_{l=1}^{m}\partial_{A_l^{n+1}}\zeta_{j_l}(\boldsymbol{\gamma})\Bigg),$$

where $\tilde{b}_{(A_1^{n+1},\dots,A_m^{n+1})}=b_{(A_1^n,\dots,A_m^n)}$. Thus for suitable $\hat{b}_{(A_1^{n+1},\dots,A_m^{n+1})}\in I\!R$ (which is replaced as needed by 0) (A.30) is equal to

$$\sum_{m=1}^{n}\sum_{j_1=1}^{k}\cdots\sum_{j_m=1}^{k}\Bigg\{\Big(\partial_{\zeta_{j_1}(\gamma)}\cdots\partial_{\zeta_{j_m}(\gamma)}\tilde{p}_{\boldsymbol{\zeta}(\boldsymbol{\gamma})}(\boldsymbol{x})\Big)\sum_{\substack{A_1^{n+1},\dots,A_m^{n+1}\neq\emptyset\\ \text{disjoint}\\ A_1^{n+1}\cup\dots\cup A_m^{n+1}=A^{n+1}}}\Bigg(\hat{b}_{(A_1^{n+1},\dots,A_m^{n+1})}\prod_{l=1}^{m}\partial_{A_l^{n+1}}\zeta_{j_l}(\boldsymbol{\gamma})\Bigg)\Bigg\}. \tag{A.33}$$

Finally, we obtain $\partial_{\gamma_{i_1}}\dots\partial_{\gamma_{i_{n+1}}}p_{\boldsymbol{\zeta}(\boldsymbol{\gamma})}(\boldsymbol{x})$ by summarizing (A.32) and (A.33) which leads to

$$\sum_{m=1}^{n+1}\sum_{j_1=1}^{k}\cdots\sum_{j_m=1}^{k}\Bigg\{\Big(\partial_{\zeta_{j_1}(\gamma)}\cdots\partial_{\zeta_{j_m}(\gamma)}\tilde{p}_{\boldsymbol{\zeta}(\boldsymbol{\gamma})}(\boldsymbol{x})\Big)\sum_{\substack{A_1^{n+1},\dots,A_m^{n+1}\neq\emptyset\\ \text{disjoint}\\ A_1^{n+1}\cup\dots\cup A_m^{n+1}=A^{n+1}}}\Bigg(b_{(A_1^{n+1},\dots,A_m^{n+1})}\prod_{i=1}^{m}\partial_{A_i^{n+1}}\zeta_{j_i}(\boldsymbol{\gamma})\Bigg)\Bigg\}$$

for suitable factors $b_{(A_1^{n+1},\dots,A_m^{n+1})}\in I\!R$. □

A.4. Lemma A.3

Lemma A.3 *Let $k\in I\!N$, $V\subset\{(i,j):1\le j<i\le k\}$ and $U=V\cup\{(i,i):i=1,\dots,k\}$. Let the polynomial $P_{A^l,\boldsymbol{b^l},c_l}:I\!R^k\to I\!R$ be given by*

$$P_{A^l,\boldsymbol{b^l},c_l}(\boldsymbol{x}) \;=\; \sum_{(i,j)\in U}a_{i,j}^l x_i x_j+\sum_{i=1}^{k}b_i^l x_i+c_l,$$

where $A^l=\Big(a_{i,j}^l\,:\,(i,j)\in U\Big)\in I\!R^{|U|}$, $\boldsymbol{b^l}=(b_1^l,\dots,b_k^l)'\in I\!R^k$ and $c_l\in I\!R$. For $n\in N$ let $\boldsymbol{\zeta^1},\dots,\boldsymbol{\zeta^n}$ be distinct points in a dense set $Z\subset I\!R^k$. Then we have

$$\sum_{l=1}^{n}P_{A^l,\boldsymbol{b^l},c_l}(\boldsymbol{x})\exp\Big(\langle\boldsymbol{\zeta^l},\boldsymbol{x}\rangle\Big)\equiv 0 \quad\Leftrightarrow\quad \forall l\in\{1,\dots,n\}:\ P_{A^l,\boldsymbol{b^l},c_l}(\boldsymbol{x})\equiv 0.$$

Proof of Lemma A.3 via induction on n :
$n = 1$: Obviously, $P_{A^1,\boldsymbol{b^1},c_1}(\boldsymbol{x}) \exp\left(\langle \boldsymbol{\zeta^1}, \boldsymbol{x}\rangle\right) \equiv 0 \;\Leftrightarrow\; P_{A^1,\boldsymbol{b^1},c_1}(\boldsymbol{x}) \equiv 0$.

$n \to n+1$:

$$\sum_{l=1}^{n+1} P_{A^l,\boldsymbol{b^l},c_l}(\boldsymbol{x}) \exp\Big(\langle \boldsymbol{\zeta^l}, \boldsymbol{x}\rangle\Big) \equiv 0 \tag{A.34}$$

$$\Leftrightarrow \quad \sum_{l=1}^{n} P_{A^l,\boldsymbol{b^l},c_l}(\boldsymbol{x}) \exp\Big(\langle \boldsymbol{\zeta^l} - \boldsymbol{\zeta^{n+1}}, \boldsymbol{x}\rangle\Big) \equiv -\Big(\sum_{(i,j)\in U} a_{i,j}^{n+1} x_i x_j + \sum_{i=1}^{k} b_i^{n+1} x_i + c_{n+1}\Big).$$

Thereby, we assume with no loss of generality that $\zeta_r^l - \zeta_r^{n+1} \neq 0$ for all $l = 1,\ldots,n$ and all $r = 1,\ldots,k$. (Otherwise multiply both sides of (A.34) with $\exp\left(\langle \boldsymbol{\zeta}, \boldsymbol{x}\rangle\right)$ for a suitable $\boldsymbol{\zeta} \in Z$ using the fact that Z is dense.) Consequently, we have

$$\forall r \in \{1,\ldots,k\}: \qquad \partial_{x_r}^3 \sum_{l=1}^{n} P_{A^l,\boldsymbol{b^l},c_l}(\boldsymbol{x}) \exp\Big(\langle \boldsymbol{\zeta^l} - \boldsymbol{\zeta^{n+1}}, \boldsymbol{x}\rangle\Big) \equiv 0. \tag{A.35}$$

Simple calculations lead to

$$\begin{aligned} &\partial_{x_r}^3 P_{A^l,\boldsymbol{b^l},c_l}(\boldsymbol{x}) \exp\Big(\langle \boldsymbol{\zeta^l} - \boldsymbol{\zeta^{n+1}}, \boldsymbol{x}\rangle\Big) \\ = \;& \Big\{ 3(\zeta_r^l - \zeta_r^{n+1})\Big(\partial_{x_r}^2 P_{A^l,\boldsymbol{b^l},c_l}(\boldsymbol{x})\Big) + 3(\zeta_r^l - \zeta_r^{n+1})^2\Big(\partial_{x_r} P_{A^l,\boldsymbol{b^l},c_l}(\boldsymbol{x})\Big) + \\ & (\zeta_r^l - \zeta_r^{n+1})^3\Big(P_{A^l,\boldsymbol{b^l},c_l}(\boldsymbol{x})\Big)\Big\} \exp\Big(\langle \boldsymbol{\zeta^l} - \boldsymbol{\zeta^{n+1}}, \boldsymbol{x}\rangle\Big). \end{aligned} \tag{A.36}$$

Inserting

$$\begin{aligned} \partial_{x_r} P_{A^l,\boldsymbol{b^l},c_l}(\boldsymbol{x}) &= \sum_{\{j:(r,j)\in V\}} a_{r,j}^l x_j + \sum_{\{i:(i,r)\in V\}} a_{i,r}^l x_i + 2a_{r,r}^l x_r + b_r^l, \\ \partial_{x_r}^2 P_{A^l,\boldsymbol{b^l},c_l}(\boldsymbol{x}) &= 2a_{r,r}^l \end{aligned}$$

in the right hand side of (A.36) results in

$$\partial_{x_r}^3 P_{A^l,\boldsymbol{b^l},c_l}(\boldsymbol{x}) \exp\Big(\langle \boldsymbol{\zeta^l} - \boldsymbol{\zeta^{n+1}}, \boldsymbol{x}\rangle\Big) \;=\; P_{A^{l,r},\boldsymbol{b^{l,r}},c_{l,r}}(\boldsymbol{x}) \exp\Big(\langle \boldsymbol{\zeta^l} - \boldsymbol{\zeta^{n+1}}, \boldsymbol{x}\rangle\Big)$$

where $A^{l,r} = \Big((\zeta_r^l - \zeta_r^{n+1})^3 a_{i,j}^l : (i,j) \in U\Big)$, $\boldsymbol{b^{l,r}} = (b_1^{l,r},\ldots,b_k^{l,r})'$ with

$$b_i^{l,r} = \begin{cases} (\zeta_r^l - \zeta_r^{n+1})^3 b_r^l + 6(\zeta_r^l - \zeta_r^{n+1})^2 a_{r,r}^l &, \; i = r \\ (\zeta_r^l - \zeta_r^{n+1})^3 b_i^l + 3(\zeta_r^l - \zeta_r^{n+1})^2 a_{r,i}^l &, \; i \in \{u : (r,u) \in V\} \\ (\zeta_r^l - \zeta_r^{n+1})^3 b_i^l + 3(\zeta_r^l - \zeta_r^{n+1})^2 a_{i,r}^l &, \; i \in \{u : (u,r) \in V\} \\ (\zeta_r^l - \zeta_r^{n+1})^3 b_i^l &, \; i \in \{1,\ldots,k\}\backslash\{u : (r,u),(u,r) \in U\} \end{cases}$$

and $c_{l,r} = (\zeta_r^l - \zeta_r^{n+1})^3 c_l + 3(\zeta_r^l - \zeta_r^{n+1})^2 b_r^l + 6(\zeta_r^l - \zeta_r^{n+1}) a_{r,r}^l$. Hence, (A.35) is equivalent to

$$\forall r \in \{1,\ldots,k\}: \qquad \sum_{l=1}^{n} P_{A^{l,r},\boldsymbol{b^{l,r}},c_{l,r}}(\boldsymbol{x}) \exp\Big(\langle \boldsymbol{\zeta^l} - \boldsymbol{\zeta^{n+1}}, \boldsymbol{x}\rangle\Big) \equiv 0 \tag{A.37}$$

which is equivalent to

$$\forall r \in \{1, \ldots, k\}\ \forall l \in \{1, \ldots, n\}: \quad P_{A^{l,r}, \boldsymbol{b}^{l,r}, c_{l,r}}(\boldsymbol{x}) \equiv 0$$

due to the induction assumption. It follows (A.37) is equivalent to

$$\begin{aligned}
\Leftrightarrow \quad & \forall r \in \{1, \ldots, k\}\ \forall l \in \{1, \ldots, n\}: & A^{l,r} = \mathbf{0},\ \boldsymbol{b}^{l,r} = 0,\ c_{l,r} = 0 \\
\Leftrightarrow \quad & \forall r \in \{1, \ldots, k\}\ \forall l \in \{1, \ldots, n\}: & A^{l} = \mathbf{0},\ \boldsymbol{b}^{l} = \mathbf{0},\ c_{l} = 0 \\
\Leftrightarrow \quad & \forall r \in \{1, \ldots, k\}\ \forall l \in \{1, \ldots, n\}: & P_{A^{l}, \boldsymbol{b}^{l}, c_{l}}(\boldsymbol{x}) \equiv 0
\end{aligned}$$

using $\zeta_r^l - \zeta_r^{n+1} \neq 0$ for all $l = 1, \ldots, n$ and all $r = 1, \ldots, k$ and $\boldsymbol{\zeta^1} - \boldsymbol{\zeta^{n-1}}, \ldots, \boldsymbol{\zeta^n} - \boldsymbol{\zeta^{n-1}}$ are distinct points. Consequently, (A.34) is equivalent to $P_{A^l, \boldsymbol{b}^l, c_l}(\boldsymbol{x}) \equiv 0$ for all $l = 1, \ldots, n+1$. □

Proofs ommited in Chapter 4

Calculation of (4.3):

For $\mu,\ \tilde{\mu} \in \Gamma \backslash \{0\}$ the covariance function has the form

$$\begin{aligned}
Cov\Big(\xi_{h_\mu}, \xi_{h_{\tilde{\mu}}}\Big) &= \int_{I\!R} h_\mu(x) h_{\tilde{\mu}}(x) f_0(x)\, dx - \int_{I\!R} h_\mu(x) f_0(x)\, dx \int_{I\!R} h_{\tilde{\mu}}(x) f_0(x)\, dx \\
&= \int_{I\!R} h_\mu(x) h_{\tilde{\mu}}(x) f_0(x)\, dx \qquad \text{(A.38)}
\end{aligned}$$

since $\left(\xi_{h_\mu}\right)_{\mu \in \Gamma \backslash \{0\}}$ is a centered Gaussian process. The definition of h_μ leads to

$$\begin{aligned}
Cov\Big(\xi_{h_\mu}, \xi_{h_{\tilde{\mu}}}\Big) = {} & \Big(\exp(\mu^2) - \mu^2 - 1\Big)^{-\frac{1}{2}} \Big(\exp(\tilde{\mu}^2) - \tilde{\mu}^2 - 1\Big)^{-\frac{1}{2}} \\
& \int_{I\!R} \left(\exp\left\{\mu x - \frac{\mu^2}{2}\right\} - 1 - \mu x\right) \left(\exp\left\{\tilde{\mu} x - \frac{\tilde{\mu}^2}{2}\right\} - 1 - \tilde{\mu} x\right) f_0(x)\, dx. \text{(A.39)}
\end{aligned}$$

The integral of (A.39) is given by

$$\begin{aligned}
& \exp(\mu\tilde{\mu}) \int_{I\!R} f_{\mu+\tilde{\mu}}(x)\, dx - \int_{I\!R} (1 + \mu x) f_{\tilde{\mu}}(x)\, dx - \int_{I\!R} (1 + \tilde{\mu} x) f_\mu(x)\, dx \\
& \qquad + \int_{I\!R} \Big(1 + \mu x + \tilde{\mu} x + \mu\tilde{\mu} x^2\Big) f_0(x)\, dx \\
= {} & \exp(\mu\tilde{\mu}) - (1 + \mu\tilde{\mu}) - (1 + \tilde{\mu}\mu) + (1 + \mu\tilde{\mu}). \qquad \text{(A.40)}
\end{aligned}$$

We obtain (4.3) by inserting (A.40) into (A.39). □

$\boldsymbol{Cov\big(\xi_{h_\mu},\xi_{h_{\tilde\mu}}\big) : \big([-0.01,0.01]\backslash\{0\}\big) \to (0,1]}$ **has a unique maximum in $\boldsymbol{(\tilde\mu,\tilde\mu)}$ for any fixed $\boldsymbol{\tilde\mu \in [-0.01,0.01]\backslash\{0\}}$:**

Let $c \in [-1,1]\backslash\{0\}$, then $[-0.01,0.01]\backslash\{0\}$ is a subset of the range of c. For $\mu \in [-1,1]\backslash\{0\}$ we define

$$f(\mu) \quad = \quad Cov\Big(\xi_{h_\mu},\xi_{h_c}\Big) \quad = \quad \frac{\exp\big(\mu c\big) - 1 - \mu c}{\Big(\exp\big(\mu^2\big) - 1 - \mu^2\Big)^{\frac{1}{2}}\Big(\exp\big(c^2\big) - 1 - c^2\Big)^{\frac{1}{2}}}.$$

We show that f has a unique maximum in c : The derivative of f with respect to μ is given by

$$\partial_\mu f(\mu) = \frac{\Big(c\exp\big(\mu c\big) - c\Big)\Big(\exp\big(\mu^2\big) - 1 - \mu^2\Big) - \Big(\exp\big(\mu c\big) - 1 - \mu c\Big)\Big(\mu\exp\big(\mu^2\big) - \mu\Big)}{\Big(\exp\big(\mu^2\big) - 1 - \mu^2\Big)^{\frac{3}{2}}\Big(\exp\big(c^2\big) - 1 - c^2\Big)^{\frac{1}{2}}}. \quad \text{(A.41)}$$

Obviously, one has $\partial_\mu f(\mu)|_{\mu=c} = 0$ for $c \in [-1,1]\backslash\{0\}$. It is clear that the denominator (A.41) is larger than 0 for all $\mu \in [-1,1]\backslash\{0\}$. Hence, it is sufficient to show that the numerator of (A.41) is greater than 0 for all $-1 \le \mu < c$ and is less than 0 for all $1 \ge \mu > c$ with $\mu \ne 0$.

For $\mu \in [-1,1]\backslash\{0,c\}$ let

$$t(\mu) \quad = \quad -\exp\big(\mu^2 + \mu c\big) + \exp\big(\mu^2\big) + \exp\big(\mu c\big) - 1 + \frac{\mu^2 c\Big(\exp\big(\mu^2\big) - \exp\big(\mu c\big)\Big)}{\mu - c}. \quad \text{(A.42)}$$

Then the numerator of (A.41) is given by $(\mu - c)t(\mu)$. Therefore, it is enough to show that

$$\forall \mu \in [-1,1]\backslash\{0,c\} : \qquad t(\mu) < 0. \quad \text{(A.43)}$$

With the help of

$$\begin{aligned}
\exp\big(\mu^2\big) - \exp\big(\mu c\big) \quad &= \quad \sum_{k=0}^{\infty}\frac{\big(\mu^2\big)^k}{k!} - \sum_{k=0}^{\infty}\frac{\big(\mu c\big)^k}{k!} \\
&= \quad (\mu - c)\sum_{k=0}^{\infty}\left(\frac{\mu^k}{k!}\,\frac{\mu^k - c^k}{\mu - c}\right) \\
&= \quad (\mu - c)\sum_{k=0}^{\infty}\frac{\mu^k}{k!}\left(\sum_{i=0}^{k-1}\mu^{k-1-i}c^i\right)
\end{aligned}$$

and the Taylor expansion of the remaining terms of $t(\mu)$ we obtain

$$\begin{aligned}
t(\mu) \quad &= \quad -1 + \sum_{k=0}^{\infty}\frac{1}{k!}\left(-\big(\mu^2 + \mu c\big)^k + \big(\mu^2\big)^k + \big(\mu c\big)^k + \mu^2 c\mu^k\left(\sum_{i=0}^{k-1}\mu^{k-1-i}c^i\right)\right) \\
&= \quad \sum_{k=1}^{\infty}\frac{1}{k!}\left(-\left(\sum_{i=0}^{k}\binom{k}{i}\big(\mu^2\big)^{k-i}\big(\mu c\big)^i\right) + \big(\mu^2\big)^k + \big(\mu c\big)^k + \mu c\left(\sum_{i=0}^{k-1}\mu^{2k-i}c^i\right)\right) \\
&= \quad \underbrace{\sum_{k=1}^{\infty}\frac{1}{k!}\left(-\sum_{i=1}^{k-1}\binom{k}{i}\mu^{2k-i}c^i\right)}_{=\,g(\mu)} + \underbrace{\sum_{k=1}^{\infty}\frac{1}{k!}\left(\sum_{i=0}^{k-1}\mu^{2k-i+1}c^{i+1}\right)}_{=\,l(\mu)}.
\end{aligned}$$

Extracting c^i, $i \in I\!N$, from $g(\mu)$ leads to

$$\begin{aligned} g(\mu) &= c^1 \sum_{k=2}^{\infty} \frac{1}{k!}\left[-\binom{k}{1}\mu^{2k-1}\right] + c^2 \sum_{k=3}^{\infty} \frac{1}{k!}\left[-\binom{k}{2}\mu^{2k-2}\right] + c^3 \sum_{k=4}^{\infty} \frac{1}{k!}\left[-\binom{k}{3}\mu^{2k-3}\right] + \dots \\ &= \sum_{j=1}^{\infty} c^j \left(\sum_{k=j+1}^{\infty} \frac{1}{k!}\left[-\binom{k}{j}\mu^{2k-j}\right]\right) \end{aligned}$$

and extracting c^i, $i \in I\!N$, from $l(\mu)$ results in

$$\begin{aligned} l(\mu) &= c^1 \sum_{k=1}^{\infty} \frac{1}{k!}\mu^{2k+1} + c^2 \sum_{k=2}^{\infty} \frac{1}{k!}\mu^{2k-1+1} + c^3 \sum_{k=3}^{\infty} \frac{1}{k!}\mu^{2k-2+1} + \dots \\ &= \sum_{j=1}^{\infty} c^j \left(\sum_{k=j}^{\infty} \frac{1}{k!}\mu^{2k-(j-1)+1}\right) \\ &= \sum_{j=1}^{\infty} c^j \left(\sum_{k=j+1}^{\infty} \frac{1}{(k-1)!}\mu^{2k-j}\right). \end{aligned}$$

It follows that

$$\begin{aligned} t(\mu) &= \sum_{j=1}^{\infty} c^j \left(\sum_{k=j+1}^{\infty} \mu^{2k-j}\left[\frac{1}{(k-1)!} - \frac{1}{k!}\binom{k}{j}\right]\right) \\ &= \sum_{j=2}^{\infty} c^j \left(\sum_{k=j+1}^{\infty} \mu^{2k-j}\frac{1}{(k-1)!}\left[1 - \binom{k-1}{j-1}\frac{1}{j}\right]\right) \end{aligned}$$

since for $j = 1$ one has $\frac{1}{(k-1)!} - \frac{1}{k!}\binom{k}{1} = 0$. Furthermore, for $k = j+1$ one has $1 - \binom{(j+1)-1}{j-1}\frac{1}{j} = 0$ which leads to

$$t(\mu) = \sum_{j=2}^{\infty} c^j \left(\sum_{k=j+2}^{\infty} \mu^{2k-j}\frac{1}{(k-1)!}\underbrace{\left[1 - \binom{k-1}{j-1}\frac{1}{j}\right]}_{<\,0}\right) \qquad \text{(A.44)}$$

with $1 - \binom{k-1}{j-1}\frac{1}{j} < 0$ since for $j \in \{2, 3, \dots\}$

$$\begin{aligned} && j+1 &> 2 \\ \Leftrightarrow && (j+2-j+1)(j+2-j+2)\dots(j+2-2)(j+2-1) &> j! \\ \Leftrightarrow && (k-j+1)(k-j+2)\dots(k-2)(k-1) &> j!, \quad \forall k \geq j+1 \\ \Leftrightarrow && \binom{k-1}{j-1} &> j, \quad \forall k \geq j+1. \end{aligned}$$

If $\mu c > 0$ then $c^j\mu^{2k-j} > 0$, $j, k \in I\!N$, since the sum of the exponents is even. This leads to $t(\mu) < 0$.

For $\mu, c \in [-1, 1]\backslash\{0\}$ with $\mu c < 0$ we split (A.44) into the sums with even and odd exponents of c :

$$
\begin{aligned}
& t(\mu) \\
= \quad & \sum_{j=1}^{\infty} c^{2j} \left\{ \sum_{k=2j+2}^{\infty} \mu^{2k-2j} \frac{1}{(k-1)!} \left[1 - \binom{k-1}{2j-1} \frac{1}{2j} \right] \right\} \\
& + \sum_{j=1}^{\infty} c^{2j+1} \left\{ \sum_{k=(2j+1)+2}^{\infty} \mu^{2k-(2j+1)} \frac{1}{(k-1)!} \left[1 - \binom{k-1}{(2j+1)-1} \frac{1}{2j+1} \right] \right\} \\
= \quad & \sum_{j=1}^{\infty} c^{2j} \left\{ \sum_{k=2j+2}^{\infty} \mu^{2k-2j} \frac{1}{(k-1)!} \left[1 - \binom{k-1}{2j-1} \frac{1}{2j} \right] \right. \\
& \qquad \left. + c \sum_{k=2j+3}^{\infty} \mu^{2k-2j-1} \frac{1}{(k-1)!} \left[1 - \binom{k-1}{2j} \frac{1}{2j+1} \right] \right\} \\
= \quad & \sum_{j=1}^{\infty} c^{2j} \left\{ \sum_{k=2j+2}^{\infty} \mu^{2k-2j} \frac{1}{(k-1)!} \left[1 - \binom{k-1}{2j-1} \frac{1}{2j} \right] \right. \\
& \qquad \left. + c \sum_{k=2j+2}^{\infty} \mu^{2k-2j+1} \frac{1}{k!} \left[1 - \binom{k}{2j} \frac{1}{2j+1} \right] \right\} \\
= \quad & \sum_{j=1}^{\infty} c^{2j} \left\{ \sum_{k=2j+2}^{\infty} \mu^{2k-2j} \underbrace{\left(\frac{1}{(k-1)!} \left[1 - \binom{k-1}{2j-1} \frac{1}{2j} \right] + c\mu \frac{1}{k!} \left[1 - \binom{k}{2j} \frac{1}{2j+1} \right] \right)}_{=\, z} \right\}
\end{aligned}
$$

Since $c^{2j}, \mu^{2k-2j} > 0$ for $j \in I\!N$ and $k = 2j+2, 2j+3, \ldots$, we show $z < 0$ with the help of

$$
z \quad < \quad \frac{1}{(k-1)!} \left[1 - \binom{k-1}{2j-1} \frac{1}{2j} \right] - \frac{1}{k!} \left[1 - \binom{k}{2j} \frac{1}{2j+1} \right] \quad < \quad 0. \tag{A.45}
$$

Thereby, the first relation follows from $-1 \leq c\mu < 0$ and

$$
\binom{k}{2j} \quad > \quad \binom{k}{1} \quad \Rightarrow \quad \left[1 - \binom{k}{2j} \frac{1}{2j+1} \right] \quad < \quad 0
$$

because of $j \geq 1$ and $k \geq 2j+2$. For $i \in \{2, 3, \ldots\}$, $j \in I\!N$ and $k = 2j+i$ one has

$$
2 \cdots i \quad < \quad 2j\,((2j+2) \cdots (2j+i-2))(2j+i)
$$

since for any $l = 1, \ldots, i-1$ the lth left factor "l" of the left hand side is less than the lth factor on the right hand side. This implies

$$
\begin{aligned}
0 \quad & > \quad (2j+i-1)i! - ((2j+2) \cdots (2j+i))\, 2j \\
& \stackrel{k=2j+i}{=} \quad (k-1)(k-2j)! - \big((2j+2) \cdots k\big)\, 2j \\
& = \quad (k-1)(k-2j)! - \frac{k!\, 2j}{(2j+1)!}
\end{aligned}
$$

which is equivalent to

$$
0 \quad > \quad (k-1) - \frac{k!\, 2j}{(2j+1)!(k-2j)!} \quad = \quad (k-1) - \binom{k}{2j} \frac{2j}{2j+1}
$$

and

$$\begin{aligned} k-\binom{k}{2j} &< 1-\binom{k}{2j}\frac{1}{2j+1} \\ \Leftrightarrow \quad \frac{1}{(k-1)!}\left[1-\binom{k-1}{2j-1}\frac{1}{2j}\right] &< \frac{1}{k!}\left[1-\binom{k}{2j}\frac{1}{2j+1}\right]. \end{aligned}$$

Obviously, we obtain (A.45)

Additionally, we have

$$\lim_{\mu\to 0}\partial_\mu f(\mu) \stackrel{(A.41)}{=} \frac{c^3}{3\sqrt{2}\Big(\exp\left(c^2\right)-1-c^2\Big)^{\frac{1}{2}}},$$

which results from the calculation of $\lim_{\mu\to 0}(\partial_\mu f(\mu))^2$ and (using 12 times) the rule of l'Hospital (or e.g. the use of MATHEMATICA). □

Taylor expansion of (4.3) for $\boldsymbol{\mu}$ or $\boldsymbol{\tilde{\mu}} \in [-0.01, 0.01]\backslash\{0\}$:
We use the Taylor expansion of the exponential functions of (4.3) depending on the values of μ and $\tilde{\mu}$, respectively. Let

$$\frac{f(\mu,\tilde{\mu})}{g(\mu)g(\tilde{\mu})} = \frac{\Big(\exp\left(\mu\tilde{\mu}\right)-1-\mu\tilde{\mu}\Big)^2}{\Big(\exp\left(\mu^2\right)-1-\mu^2\Big)\Big(\exp\left(\tilde{\mu}^2\right)-1-\tilde{\mu}^2\Big)}. \tag{A.46}$$

According to our observations numerical problems only occur on the range $[-0.01, 0.01]\backslash\{0\}$. As a consequence we expand $f(\mu,\tilde{\mu})$ iff at least one of the two parameters μ and $\tilde{\mu}$ belongs to $[-0.01, 0.01]\backslash\{0\}$. Correspondingly, we make a Taylor expansion of $g(\mu)$ iff $\mu \in [-0.01, 0.01]\backslash\{0\}$. This leads to the following three cases:

1. $\mu \in [-0.01, 0.01]\backslash\{0\}$ and $\tilde{\mu} \notin [-0.01, 0.01]$: We expand (A.46) into

$$\begin{aligned} & \frac{\Big(\sum_{k=0}^{n}\frac{1}{k!}\left(\mu\tilde{\mu}\right)^k-1-\mu\tilde{\mu}\Big)^2}{\Big(\sum_{k=0}^{n}\frac{1}{k!}\left(\mu^2\right)^k-1-\mu^2\Big)\Big(\exp\left(\tilde{\mu}^2\right)-1-\tilde{\mu}^2\Big)} \\ = & \frac{\Big(\frac{\mu^2\tilde{\mu}^2}{2!}+\frac{\mu^3\tilde{\mu}^3}{3!}+\ldots+\frac{\mu^n\tilde{\mu}^n}{n!}\Big)^2}{\Big(\frac{\mu^4}{2!}+\frac{\mu^6}{3!}+\ldots+\frac{\mu^{2n}}{n!}\Big)\Big(\exp\left(\tilde{\mu}^2\right)-1-\tilde{\mu}^2\Big)} \\ = & \frac{\frac{\mu^4}{4}\Big(\tilde{\mu}^2+\frac{\mu\tilde{\mu}^3}{3}+\ldots+\frac{\mu^{n-2}\tilde{\mu}^n}{3\cdot 4\cdot\ldots\cdot n}\Big)^2}{\frac{\mu^4}{2!}\Big(1+\frac{\mu^2}{3}+\ldots+\frac{\mu^{2n-4}}{3\cdot 4\cdot\ldots\cdot n}\Big)\Big(\exp\left(\tilde{\mu}^2\right)-1-\tilde{\mu}^2\Big)} \\ = & \frac{1}{2}\frac{\Big(1+\frac{\mu\tilde{\mu}}{3}\Big(1+\frac{\mu\tilde{\mu}}{4}\Big(1+\ldots\Big(1+\frac{\mu\tilde{\mu}}{n-1}\Big(1+\frac{\mu\tilde{\mu}}{n}\Big)\Big)\ldots\Big)\Big)\Big)^2}{\Big(1+\frac{\mu^2}{3}\Big(1+\frac{\mu^2}{4}\Big(1+\ldots\Big(1+\frac{\mu^2}{n-1}\Big(1+\frac{\mu^2}{n}\Big)\Big)\ldots\Big)\Big)\Big)\Big(\exp\left(\tilde{\mu}^2\right)-1-\tilde{\mu}^2\Big)}. \end{aligned}$$

2. $\tilde{\mu} \in [-0.01, 0.01]\backslash\{0\}$ and $\mu \notin [-0.01, 0.01]$: Analogous to item 1.

3. $\tilde{\mu}, \mu \in [-0.01, 0.01]\backslash\{0\}$: Analogous to item 1. we expand (A.46) into

$$\frac{f(\mu, \tilde{\mu})}{g(\mu)g(\tilde{\mu})},$$

with $f : [-0.01, 0.01] \times [-0.01, 0.01] \to I\!R_+$ given by

$$f(\mu, \tilde{\mu}) = \left(1 + \frac{\mu\tilde{\mu}}{3}\left(1 + \frac{\mu\tilde{\mu}}{4}\left(1 + \ldots\left(1 + \frac{\mu\tilde{\mu}}{n-1}\left(1 + \frac{\mu\tilde{\mu}}{n}\right)\right)\ldots\right)\right)\right)^2$$

and $g : [-0.01, 0.01] \to I\!R_+$ given by

$$g(\mu) = \left(1 + \frac{\mu^2}{3}\left(1 + \frac{\mu^2}{4}\left(1 + \ldots\left(1 + \frac{\mu^2}{n-1}\left(1 + \frac{\mu^2}{n}\right)\right)\ldots\right)\right)\right).$$

In all cases we decide to use finite sums of $n = 100$ terms, since the remainder in Taylor's formula is given by $r(|\mu|) \le \frac{|\mu|^{n+1}}{(n+1)!} \overset{n=100}{<} 1.1 \cdot 10^{-362}$, for all $\mu \in [-0.01, 0.01]\backslash\{0\}$. The terms in the brackets have been arranged in such order that the whole expression can be calculated in a numerically stable way. $\square$

B. Simulation Programs

In this section we use a pseudo FORTRAN 77 code while writing comments in italic script and FORTRAN commands as well as NAG-routines in capital letters.

Program B.1 In "SupGauss" the variable $\frac{1}{2} \sup_{\mu \in \Gamma_N} \left(\xi_{h_\mu}\right)^2 \cdot \mathbb{1}_{\xi_{h_\mu} \geq 0}$ is calculated.

PROGRAM <u>SupGauss</u>

Initialisation

$$\Gamma = [-a, a]$$

$$\mu_i = \begin{cases} -a + (i-1)\frac{2a}{N} & , \quad i = 1, \ldots, \frac{N}{2} \\ -a + i\frac{2a}{N} & , \quad i = \frac{N}{2}, \ldots, N \end{cases}$$

Generating the covariance matrix

DO 5 $i = 1, N$

DO 7 $j = i, N$

IF $\mu_i, \mu_j \notin [-0.01, 0.01]$ THEN

$Cov\left(\xi_{h_{\mu_i}}, \xi_{h_{\mu_j}}\right)$ as in (4.3) and (A.46), respectively

ELSE

$Cov\left(\xi_{h_{\mu_i}}, \xi_{h_{\mu_j}}\right)$ by Taylor expansion as on p. 117

ENDIF

$Cov\left(\xi_{h_{\mu_j}}, \xi_{h_{\mu_i}}\right) = Cov\left(\xi_{h_{\mu_i}}, \xi_{h_{\mu_j}}\right)$

7 CONTINUE

5 CONTINUE

Spectral factorization

CALL F02FAF(...)

results in eigenvalues $\lambda_1 \leq \ldots \leq \lambda_N$

and eigenvectors $\boldsymbol{v_1}, \ldots, \boldsymbol{v_N}$

$t = \min \arg\{\lambda_i \geq 10^{-9} : \; i = 1, \ldots, N\}$

Generating $L_N = (\boldsymbol{v_1}, \ldots, \boldsymbol{v_N})\, diag(0, \ldots, 0, \sqrt{\lambda_t}, \ldots, \sqrt{\lambda_N})$

DO 9 $i = 1, N$

DO 11 $j = 1, N - t$

$L_N(i,j) = v_{i,t+j}\, \sqrt{\lambda_{t+j}}$

11 CONTINUE

9 CONTINUE

Generating i.i.d. $X_1, \dots, X_{N-t} \sim N(0,1)$
DO 13 $i = 1, N-t$
$X_i =$ G05DDF(0,1)
13 CONTINUE

Generating $\boldsymbol{\xi_N} = (\xi_{h_{\mu_1}}, \dots, \xi_{h_{\mu_N}})' \sim N(\mathbf{0}, \Sigma_N)$, *see* (4.6) and (4.7)
DO 15 $i = 1, N$
$\xi_{h_{\mu_i}} = 0$
DO 17 $j = 1, N-t$
$\xi_{h_{\mu_i}} = \xi_{h_{\mu_i}} + L_N(i,j) X_j$
17 CONTINUE
15 CONTINUE

Calculating $\frac{1}{2} \sup_{\mu \in \Gamma_N} \left(\xi_{h_\mu}\right)^2 \cdot \mathbb{1}_{\xi_{h_\mu} \geq 0}$
$\sup = \max\{0,\ \xi_{h_{\mu_i}}$ for $i = 1, \dots, N\}$
$\sup = \frac{1}{2}\sup$

STOP
END SupGauss

Program B.2 In "AsympQuantiles" we calculate the quantiles of the asymptotic distribution as described in section 4.

PROGRAM AsympQuantiles
Initialisation of the random generator
seed = 0
G05CBF(seed)

Initialisation
$\text{repl} = 10^7$
DO 5 $r = 1, \text{repl}$
CAll SupGauss, see Program B.1
$Y_r = \sup$
5 CONTINUE

Calculating the asymptotic quantiles
CALL M01CAF($\boldsymbol{Y}$, 1, repl, "Ascending", ifail)
results in $Y_1 \leq \ldots \leq Y_{\text{repl}}$
$\text{quantil}(\alpha) = Y_{\lceil \text{repl } \alpha \rceil}$

STOP
END AsympQuantiles

For any $x \in \mathbb{R}$ we define $\lceil x \rceil = \min\{n \in \mathbb{N} : n \geq x\}$.

Program B.3 In "LRT-Quantiles" we use an EM-algorithm to calculate the quantiles "quantilEM" on the unbounded space $I\!R$, i.e. (4.9) is maximized. Additionally, we apply the NAG-routine E04JYF iff the EM-algorithm crosses the bounds of $[-a, a]$. This leads to the quantiles "quantilE04" (see (4.1)).

PROGRAM LRT-Quantiles

Initialisation of the random generator

seed = 0

G05CBF(seed)

Initialisation of a boolean variable to copy the results of the EM-algorithm if the bounds of $[-a, a]$ *are not exceeded*

test = TRUE

repl = 10.000 (or 100.000)

DO 5 $r = 1, \text{repl}$

Generating $(x_1, \ldots, x_n)' \sim N(\mathbf{0}, I_N)$

CALL G05FDF(0,1,n,$\boldsymbol{x}$)

Initialising starting values of the EM-algorithm

$$
\begin{aligned}
\pi^{EM}(1) &= 0.5 \\
\bar{x} &= \frac{1}{n}\sum_{i=1}^{n} x_i \\
\mu_1^{EM}(1) &= \bar{x} - 0.5 \\
\mu_2^{EM}(1) &= \bar{x} + 0.5
\end{aligned}
$$

EM-Iterations, maximal 100.000 s(teps)

DO 7 $s = 1, 100.000$

CALL EM($\pi^{EM}(s), \mu_1^{EM}(s), \mu_2^{EM}(s), \ldots$)

results in $sup^{EM}(s+1),\ \pi^{EM}(s+1),\ \mu_1^{EM}(s+1),\ \mu_2^{EM}(s+1)$

IF $\mu_1^{EM}(s+1),\ \mu_2^{EM}(s+1) \notin [-a, a]$ THEN

CALL E04JYF($\ldots$), *a quasi-Newton algorithm for finding a minimum of a function to fixed upper and lower bounds, using function values only*

results in supE04(r)

$$\text{supE04(r)} = -\text{supE04(r)} - \sum_{i=1}^{n} \log\left(f_{\hat{\mu}_n}(x_i)\right) \ (\textit{see}\ (4.1))$$

test = FALSE

ENDIF

Stopping criterion for the EM-algorithm

IF $sup^{EM}(s+1) - sup^{EM}(s - \text{stoppdiff}) < n \cdot acc$ THEN

with stoppdiff = 10, $n \cdot acc = 10^{-8}$

GOTO 9

ENDIF

7 CONTINUE

9 CONTINUE

supEM(r) = $sup^{EM}(s)$
IF (Test = TRUE) THEN
supE04(r) = supEM(r)
ENDIF
5 CONTINUE

Calculating the LRT-quantiles
CALL M01CAF(**supE04**, 1, repl, "Ascending", ifail)
results in supE04(1) $\leq \ldots \leq$ supE04(repl)
quantilE04(α) = supE04$\lceil$ repl $\alpha\rceil$

CALL M01CAF(**supEM**, 1, repl, "Ascending", ifail)
results in supEM(1) $\leq \ldots \leq$ supEM(repl)
quantilEM(α) = supEM$\lceil$ repl $\alpha\rceil$

STOP
END LRT-Quantiles

Program B.4 In the following program the power is calculated as described in section 4.3. We consider the example of $\Gamma = [-2.5, 2.5]$, sample size $n = 1000$ and the corresponding 0.9-quantiles as given in Table 4.1, p. 95.

PROGRAM Power
Initialisation of the random generator
seed = 0
G05CBF(seed)

Initialisation of the 0.9-quantiles given by Table 4.1, *p.* 95

$$\begin{aligned} \text{qAsymp} &= 2.0091 \\ \text{qLRTE04} &= 1.4902 \\ \text{qLRTEM} &= 1.4959 \quad (\textit{see } \Gamma = [-5, 5]) \end{aligned}$$

Initialisation of the mixing weights of the testing sample
DO 5 p = 1, 3
IF p = 1 THEN
$\pi = 0.1$
ELSE IF p = 2 THEN
$\pi = 0.25$
ELSE
$\pi = 0.5$
ENDIF

Initialisation of the component parameters of the testing sample
$\mu_1 = 0$
$\mu_2 = -2$
DO 7 m = 1, 7

```
      μ2 = μ2 + 0.5

      Initialisation of the number of rejection of H0
         powAsymp(π, μ2)  =  0
        powLRTE04(π, μ2)  =  0
        powLRTEM(π, μ2)   =  0

      repl = 10.000
      DO 9 r = 1, repl
        Initialisation of the testing sample
        DO 11 i = 1, n
          Generating Z ~ U[0,1]
          Z = G05CAF(Z)
          IF Z ≤ π THEN
            x_i = G05DDF(μ1,1) (i.e. x_i ~ N(μ1,1))
            ELSE
            x_i = G05DDF(μ2,1) (i.e. x_i ~ N(μ2,1))
            ENDIF
11        CONTINUE

        Initialising starting values of the EM-algorithm (as in Program B.3)
        EM-Iterations, maximal 100.000 s(teps) (as in Program B.3)
        Calculating supE04(r) (as in Program B.3)

        IF supE04(r) > qAsymp THEN
           powAsymp(π, μ2) = powAsymp(π, μ2) + 1
        ENDIF
        IF supE04(r) > qLRTE04 THEN
           powLRTE04(π, μ2) = powLRTE04(π, μ2) + 1
        ENDIF
        IF supE04(r) > qLRTEM THEN
           powLRTEM(π, μ2) = powLRTEM(π, μ2) + 1
        ENDIF
9     CONTINUE

        powAsymp(π, μ2)  =  powAsymp(π, μ2)/repl
       powLRTE04(π, μ2)  =  powLRTE04(π, μ2)/repl
       powLRTEM(π, μ2)   =  powLRTEM(π, μ2)/repl

7     CONTINUE
5   CONTINUE


STOP
END Power
```

Bibliography

Barndorff-Nielsen, O. (1978), *Information and Exponential Families*, Wiley & Sons.

Böhning, D. (2000), *Computer-Assisted Analysis of Mixtures and Applications*, Chapman & Hall/CRC.

Brown, L. D. (1986), *Fundamentals of Statistical Exponential Families with Applications in Statistical Decision Theory*, Lecture Notes-Monograph Series, Vol. 9.

Chernoff, H. (1954), *On the Distribution of the Likelihood Ratio*, Ann. Math. Statist., Vol. 25, 573-578.

Chernoff, H. and Lander E. (1995), *Asymptotic Distribution of the Likelihood Ratio Test that a Mixture of Two Binomials is a Single Binomial*, J. Statist. Plann. Inference, 43, 19-40.

Ciuperca, G. (1998), *Maximum Likelihood Test for Mixtures on Translation Parameter*, Preprint, Univ. Paris XI Orsay.

Dacunha-Castelle, D. and Gassiat, É. (November 1996) [DCG96], *Testing the Order of a Model Using Locally Conic Parametrization: Population Mixtures and Stationary ARMA Processes*, Preprint (see [DCG99]).

Dacunha-Castelle, D. and Gassiat, É. (1997) [DCG97], *Testing in Locally Conic Models, and Application to Mixture Models*, ESAIM: Probability and Statistics, Vol. 1, 285-317.

Dacunha-Castelle, D. and Gassiat, É. (1999) [DCG99], *Testing the Order of a Model Using Locally Conic Parametrization: Population Mixtures and Stationary ARMA Processes*, Ann. Statist., Vol. 27, 1178-1209.

Davies, R. B. (1977), *Hypothesis Testing when a Nuisance Parameter is Present only under the Alternative*, Biometrika, Vol. 64, 247-254.

Dudley, R. M. (1999), *Uniform Central Limit Theorems*, Cambridge University Press.

Garel, B. (2001), *Likelihood Ratio Test for Univariate Gaussian Mixture*, J. Statist. Plann. Inference, 96, 325-350.

Garel, B. (2005), *Asymptotic Theory of the Likelihood Ratio Test for the Identification of a Mixture*, J. Statist. Plann. Inference, 131(2), 271-296.

Gosh, J. K. and Sen, P. K. (1985), *On the Asymptotic Performance of the Log Likelihood Ratio Statistic for the Mixture Model and Related Results*, Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer, Vol. II, 789-806, Le Cam, L. M. and Olshen, R. A. eds.

Hartigan, J. A. (1985), *A Failure of Likelihood Asymptotics for Normal Mixtures*, Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer, Vol. II, 807-810, Le Cam, L. M. and Olshen, R. A. eds.

Horn, R. A. and Johnson, C. R. (1985), *Matrix Analysis*, Cambridge University Press.

Leadbetter, M. R., Lindgren, G., Rootzén, H. (1983), *Extremes and Related Properties of Random Sequences and Processes*, Springer-Verlag New York Inc.

Lemdani, M. and Pons, O. (1997), *Likelihood Ratio Tests for Genetic Linkage*, Statistics & Probability Letters, 33, 15-22.

Leroux, B. G. (1992), *Consistent Estimation of a Mixing Distribution*, Ann. Statist., Vol. 20, No. 3, 1350-1360.

Lindsay, B. G. (1995), *Mixture Models: Theory, Geometry and Applications*, NSF-CBMS Regional Conference Series in Probability and Statistics, Vol. 5, IMS, Hayward, CA.

Liu, X. and Shao, Y. (2003) [LS03], *Asymptotics for Likelihood Ratio Tests under Loss of Identifiability*, Ann. Statist., Vol. 31, No. 3, 807-832.

Liu, X. and Shao, Y. (2004), *Asymptotics for the Likelihood Ratio Test in a Two-Component Normal Mixture Model*, J. Statist. Plann. Inference, 123, 61-81.

Keribin, C. (2000), *Consistent Estimation of the Order of Mixture Models*, Sankhyā, Vol. 62, Series A, 49-66.

McLachlan, G. J. and Krishnan, T. (1997), *The EM Algorithm and Extensions*, Wiley & Sons, Inc.

McLachlan, G. and Peel, D. (2000), *Finite Mixture Models*, Wiley & Sons, Inc.

Prakasa Rao, B. L. S., (1987), *Asymptotic Theory of Statistical Inference*, Wiley & Sons, Inc.

Ruck, A. (2001), *Calculating the (Asymptotic) Distribution of the Log-LRT Statistic in a Contamination Mixture Model*, Diskussionsbeiträge zur Statistik und Quantitativen Ökonomik 93 , Universität der Bundeswehr Hamburg.

Ruck, A. (2002), *Calculating the Asymptotic Distribution of the Log-LRT Statistic for Testing One Against Two Populations in Normal Mixtures*, Diskussionsbeiträge zur Statistik und Quantitativen Ökonomik. 98 , Universität der Bundeswehr Hamburg.

Teicher, H., (1961), *Identifiability of Mixtures*, Ann. Math. Statist., Vol. 32, 244-248.

Teicher, H., (1963), *Identifiability of Finite Mixtures*, Ann. Math. Statist., Vol. 34, 1265-1269.

Titterington, D. M., Smith, A. F. M., Makov, U. E. (1992) *Statistical Analysis of Finite Mixture Distributions*, Wiley & Sons, New York.

Van der Vaart, A. W., Wellner, J. A. (1996), *Weak Convergence and Empirical Processes*, Springer-Verlag, New York, Inc.

Wilks, S. S. (1938), *The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses*, Ann. Math. Statist., Vol. 9, 60-62.

Witting, H. (1985), *Mathematische Statistik, I. Parametrische Verfahren bei festem Stichprobenumfang*, Teubner Stuttgart.

Yakowitz, S. J. and Spragins, J. D. (1968) [YS68], *On the Identifiability of Finite Mixtures*, Ann. Math. Statist., Vol. 39, No. 1, 209-214.

Symbols

$A(\boldsymbol{\gamma})$	normalizing factor, p. 45
A_n, B_n	(random) partitions of $\mathcal{B}$, p. 33
$Cov_{\boldsymbol{\zeta}}\boldsymbol{T} = \partial^2_{\boldsymbol{\zeta}} K(\boldsymbol{\zeta})$	p. 48
$E_{\boldsymbol{\zeta}}\boldsymbol{T} = \partial_{\boldsymbol{\zeta}} K(\boldsymbol{\zeta})$	p. 48
$E_{\boldsymbol{\zeta}} T_1^{l_1} \dots T_k^{l_k}$	p. 48
$f_{\boldsymbol{\gamma}^{\boldsymbol{l}}}$	lth component ν-density
$f_{\boldsymbol{\gamma}}(\boldsymbol{x}) = r(\boldsymbol{x}) p_{\boldsymbol{\gamma}}(\boldsymbol{x})$	p. 48
$g_{\boldsymbol{\pi},(\boldsymbol{\gamma}^{\boldsymbol{1}},\dots,\boldsymbol{\gamma}^{\boldsymbol{p}})} = \sum_{l=1}^{p} \pi_l f_{\boldsymbol{\gamma}^{\boldsymbol{l}}}$	p mixture
$g^0 = g_{\boldsymbol{\pi}^{\boldsymbol{0}},(\boldsymbol{\gamma}^{\boldsymbol{0},\boldsymbol{1}},\dots,\boldsymbol{\gamma}^{\boldsymbol{0},\boldsymbol{q}})}$	true q mixture density
$g^0(\boldsymbol{x}) = r(\boldsymbol{x}) w^0(\boldsymbol{x})$	p. 48
$h_{\boldsymbol{\gamma}}$	p. 43
$H_h, H_h^c, H_h(u,v)$	p. 22
$K(\boldsymbol{\zeta}) = -\log(\mathcal{C}(\boldsymbol{\zeta}))$	log-Laplace transform, p. 46
$l_n(g_{(\theta,\boldsymbol{\beta})}) = \sum_{i=1}^{n} \log\left(g_{(\theta,\boldsymbol{\beta})}(\boldsymbol{x_i})\right)$	log-likelihood function, p. 32
$l_n(\theta,\boldsymbol{\beta}) = l_n(g_{(\theta,\boldsymbol{\beta})})$	p. 32
$L_{(\boldsymbol{x_1},\dots,\boldsymbol{x_n})}\Big(\boldsymbol{\pi}, (\boldsymbol{\gamma}^{\boldsymbol{1}},\dots,\boldsymbol{\gamma}^{\boldsymbol{p}})\Big)$	likelihood function, p. 2
$L_2(g^0\nu)$	set of square-integrable functions with respect to $g^0\nu$
$\log N_{[\,]}^{(q)}(\varepsilon,\mathcal{D},P)$	metric entropy, p. 25
$N(\boldsymbol{\beta})$	normalizing factor, p. 11
$N_{[\,]}^{(q)}(\varepsilon,\mathcal{D},P)$	bracketing number, p. 25
$O(f), o(f)$	Landau order symbols, p. 28
$O_P(\cdot), o_P(\cdot)$	p. 33
$ord(\mathcal{P})$	order of $\mathcal{P}$, p. 47
$p_{\boldsymbol{\gamma}}(\boldsymbol{x}) = \mathcal{C}(\boldsymbol{\zeta}(\boldsymbol{\gamma})) \exp\Big\{\big\langle \boldsymbol{\zeta}(\boldsymbol{\gamma}), \boldsymbol{T}(\boldsymbol{x})\big\rangle\Big\}$	μ-density, p. 46
$\tilde{p}_{\boldsymbol{\zeta}}(\boldsymbol{x}) = \mathcal{C}(\boldsymbol{\zeta}) \exp\Big\{\big\langle \boldsymbol{\zeta}, \boldsymbol{T}(\boldsymbol{x})\big\rangle\Big\}$	μ-density in canonical parameterization, p. 46
$Pf = \int f\,dP$	p. 24
$P^{\boldsymbol{\zeta}}_{\zeta_{i_1}\dots\zeta_{i_n}}(\boldsymbol{T}(\boldsymbol{x}))$	polynomial in $\boldsymbol{T}(\boldsymbol{x})$, p. 49
$r : \mathfrak{X} \to I\!R$	p. 45
$\boldsymbol{T} = (T_1,\dots,T_k)' : \mathfrak{X} \to I\!R^k$	(generating) statistic, p. 45
$T_n(p)$	LRT statistic for testing $g = g^0$ against $g \in \mathcal{G}_p \backslash \{g^0\}$
U_h, U_h^c	p. 22
$\overline{U}_\varepsilon(\boldsymbol{\gamma}^{\boldsymbol{0},\boldsymbol{l}}) = \{\boldsymbol{\gamma} : \Vert\boldsymbol{\gamma} - \boldsymbol{\gamma}^{\boldsymbol{0},\boldsymbol{l}}\Vert \le \varepsilon\}$	
$w^0 = \sum_{l=1}^{q} \pi_l^0 p_{\boldsymbol{\gamma}^{\boldsymbol{0},\boldsymbol{l}}}$	p. 48
$\tilde{w}^0 = \sum_{l=1}^{q} \pi_l^0 \tilde{p}_{\boldsymbol{\zeta}^{\boldsymbol{0},\boldsymbol{l}}}$	p. 48
$\boldsymbol{x_i} = (x_{i,1},\dots,x_{i,m})'$	ith observation
$\boldsymbol{X_i} = (X_{i,1},\dots,X_{i,m})'$	ith sample
$Z = \boldsymbol{\zeta}(\Gamma) \subset I\!R^k$	
$\tilde{Z}_m$	p. 50
Z_*	canonical parameter space, p. 46

$\mathcal{B}$	bounded set of "directional" parameters, p. 10, p. 12
$\mathcal{B}_0 \subset \mathcal{B}$	p. 40
$\mathcal{C}(\boldsymbol{\zeta}(\boldsymbol{\gamma})) = A(\boldsymbol{\gamma})$	normalizing factor, p. 46
$\mathcal{C}(\boldsymbol{\zeta}) = \left(\int \exp\{\langle \boldsymbol{\zeta}, \boldsymbol{T}(\boldsymbol{x}) \rangle\} \, \mu(d\boldsymbol{x}) \right)^{-1}$	normalizing factor, p. 46
$\mathcal{D} = \{d_{\boldsymbol{\beta}} : \boldsymbol{\beta} \in \mathcal{B}\}$	set of directional scores of $g_{(\theta,\boldsymbol{\beta})}$, p. 11, 26
$\mathcal{D}_0 \subset \mathcal{D}$	p. 41
$\mathcal{F} = \{f_{\boldsymbol{\gamma}} : \boldsymbol{\gamma} \in \Gamma\}$	family of probability densities
$\mathcal{G}_p$	model of all p population mixtures of $\mathcal{F}$, p. 1
$\mathcal{H}_h, \mathcal{H}_h^c$	p. 22
$\mathcal{P}$	family of probability measures
$\mathcal{T}_0$	p. 40
$(\mathbb{G}_f^n)_{f \in \mathcal{D}}$	empirical process on $\mathcal{D}$, p. 24
$(\mathfrak{X}, \mathcal{X}),\ \mathfrak{X} \subset \mathbb{R}^m$	measurable space
$[f, g]$	bracket, p. 25
$\partial_{\gamma_{i_1} \ldots \gamma_{i_h}} f_{\boldsymbol{\gamma}}$	hth partial derivative of f with respect to $\gamma_{i_1} \ldots \gamma_{i_h}$
$\partial_{\boldsymbol{\gamma}} f_{\boldsymbol{\gamma}} = (\partial_{\gamma_1} f_{\boldsymbol{\gamma}}, \ldots, \partial_{\gamma_k} f_{\boldsymbol{\gamma}})$	gradient of $f_{\boldsymbol{\gamma}}$
$\partial_{\gamma_i}^h f_{\boldsymbol{\gamma}} = \partial_{\gamma_i \ldots \gamma_i} f_{\boldsymbol{\gamma}}$	hth partial derivative of f with respect to γ_i
$\partial_{\boldsymbol{\gamma}}^2 f_{\boldsymbol{\gamma}}$	Hesse matrix of $f_{\boldsymbol{\gamma}}$
$\boldsymbol{\beta} = (\lambda_1, \ldots, \lambda_{p-q}, \boldsymbol{\gamma^1}, \ldots, \boldsymbol{\gamma^{p-q}}, \boldsymbol{\delta^1}, \ldots, \boldsymbol{\delta^q}, \rho_1, \ldots, \rho_q)'$	"directional" parameter, p. 11, 13
$\boldsymbol{\gamma^l} = (\gamma_1^l, \ldots, \gamma_k^l)'$	lth component parameter
$\boldsymbol{\gamma^{0,l}} = (\gamma_1^{0,l}, \ldots, \gamma_k^{0,l})'$	lth true component parameter
$\Lambda(\boldsymbol{x_1}, \ldots, \boldsymbol{x_n})$	LRT statistic, p. 2
$\boldsymbol{\pi} = (\pi_1, \ldots, \pi_p)'$	vector of mixing weights
$\boldsymbol{\pi^0} = (\pi_1^0, \ldots, \pi_q^0)'$	vector of true mixing weights
Π_p	family of all p-dimensional mixing weights
$\mu(B) = \int_B r(\boldsymbol{x}) \, \nu(d\boldsymbol{x}), \quad B \in \mathcal{X}$	p. 46
$\rho_P(\cdot, \cdot)$	pseudometric, p. 24
θ	"distance" parameter (with respect to g^0), p. 13, 14
$\hat{\theta}_{\boldsymbol{\beta}}$	maximizer of $l_n(\theta, \boldsymbol{\beta})$ for fixed value of $\boldsymbol{\beta} \in \mathcal{B}$, p. 33
$(\xi_f)_{f \in \mathcal{D}}$	centered Gaussian process, p. 24
$\boldsymbol{\zeta} = (\zeta_1, \ldots, \zeta_k)' : \Gamma \to \mathbb{R}^k$	parameter function, p. 45

Index

Acknowledgements

I would like to express my sincere gratitude and deep respect for all the people who were involved in my doctoral studies, especially my supervisors, Prof. Hans Daduna and Prof. Wilfried Seidel, for all of the support and guidance they provided to me:

First of all, I would like to thank both of my supervisors for their openness, patience and kind words of support and encouragement towards the completion of my thesis.

Prof. Hans Daduna guided me towards profound insights and inspiration during numerous lengthy and informative discussions. These discussions led to a greater understanding of the more intricate and complicated aspects of my research work.

Prof. Wilfried Seidel generously shared his vast understanding of simulation processes and numerical methodologies with me. This guidance proved to be highly instructional and useful in regards to my ability to perform stable computations for the asymptotic distribution of the LRT statistic.

I also wish to express my heartfelt gratitude to Prof. Bernard Garel for the inspiration and encouragement I gained through our exchange of ideas and results of our independent work on the calculation of quantiles of the limiting distribution.

Additionally, it was an honor and privilege for me to work together with my outstanding colleagues who constantly provided a friendly and motivating work environment.

Finally, I am greatly indebted to my family and friends who have always stood by me through their love and prayers.

Moreover the word of the Lord came unto me, saying, Jeremiah, what seest thou? And I said, I see a rod of an almond tree. Then said the Lord unto me, Thou hast well seen: for I will hasten my word to perform it. Jeremiah 1, 11-12

Abstract

The subject of this thesis is the asymptotic distribution of the Likelihood Ratio Test (LRT) statistic of hypotheses about the number of components in finite mixture models. The "classical" results of a χ^2-asymptotic do not hold for these tests since the parameters of a mixing distribution are not identifiable among other things. Our starting point are the sufficient conditions for the existence of a limiting distribution of the LRT statistic for testing q against p components given by Dacunha-Castelle and Gassiat (1999). An analysis shows that these requirements do not hold, for instance, for "simple" mixtures of univariate normal distributions with heterogeneity of variances. After our modifications of the sufficient conditions given by Dacunha-Castelle and Gassiat (1999) a wider range of applications of their theory is offered.

Additionally, we derive sufficient conditions for the existence of a limiting distribution of the LRT statistic for mixtures of general minimal exponential families. Afterwards we provide a set of various (new) examples for Gaussian-distributions, Gamma-distributions and discrete distributions possessing these properties. Our first modification offers testing for q against $q+1$ components of bivariate normal distributions with unknown (arbitrary) covariance matrix and our second modification offers testing homogeneity against a two population mixture of univariate normal distributions with heterogeneity of variances, if the component parameter space is suitably restricted.

Finally, a method for simulating the quantiles the asymptotic distribution is introduced. Under the null hypothesis the LRT statistic converges in distribution to a supremum of a functional of a suitable Gaussian process, which we approximate with the aid of its final-dimensional marginal distributions. Considering as example a simple test of homogeneity of univariate normal distributions with unknown mean we give numerical results of this general method. In this context we introduce a modified spectral factorization on the resulting covariance matrix, whose outcome is a product of surprisingly low dimensional matrices and leads to considerably short computation times. A comparison of the power of the simulated asymptotic and exact quantiles shows that even for a small sample size the power is nearly the same.

Zusammenfassung

In dieser Arbeit wird die asymptotische Verteilung der Likelihood Ratio Test (LRT) Statistik von Hypothesen über die Anzahl der Komponenten in endlichen Mischungsmodellen untersucht. Die "klassischen" Resultate einer χ^2-Asymptotik gelten für diese Tests nicht, weil u.A. die Parameter einer Mischungsverteilung nicht identifizierbar sind. Unser Ausgangspunkt sind die hinreichenden Bedingungen für die Existenz einer Grenzverteilung der LRT Statistik für den Test von q gegen p Komponenten von Dacunha-Castelle und Gassiat (1999). Eine Analyse zeigt, daß diese Voraussetzungen, die sich auf allgemeine Mischungsmodelle beziehen, beispielsweise nicht für "einfache" Mischungen von univariaten Normalverteilungen mit Varianzheterogenität erfüllt sind. Durch unsere Modifikationen der hinreichenden Bedingungen von Dacunha-Castelle und Gassiat (1999) wird ihrer Theorie ein breiteres Anwendungsspektrum verliehen.

Zusätzlich leiten wir hinreichende Bedingungen für die Existenz einer Grenzverteilung der LRT Statistik für Mischungen allgemeiner minimaler Exponentialfamilien her. Anschließend geben wir eine Reihe (neuer) Beispiele für Gauß-Verteilungen, Gamma-Verteilungen und diskrete Verteilungen an, die diese Eigenschaften aufweisen. Unsere erste Modifikation ermöglicht einen Test von q gegen $q+1$ Komponenten bivariater Normalverteilungen mit unbekannter (beliebiger) Kovarianzmatrix und unsere zweite Modifikation ermöglicht einen Homogenitätstest gegen zwei Populationen univariater Normalverteilungen mit Varianzheterogenität, wenn der Komponentenparameterraum geeignet eingeschränkt wird.

Schließlich wird eine Methode für die Simulation der Quantile der asymptotischen Verteilung vorgestellt. Unter der Nullhypothese konvergiert die LRT Statistik in Verteilung gegen ein Supremum eines Funktionals eines geeigneten Gaußprozesses, den wir mit Hilfe seiner endlichdimensionalen Randverteilungen approximieren. Am Beispiel eines einfachen Homogentitätstests von univariaten Normalverteilungen mit unbekanntem Erwartungswert geben wir numerische Ergebnisse für dieses allgemeine Verfahren an. In diesem Zusammenhang führen wir eine modifizierte Spektralzerlegung der entsprechenden Kovarianzmatrix ein, die ein Produkt aus überraschend niedrigdimensionalen Matrizen ist und zu bemerkenswert kurzen Rechenzeiten führt. Ein Vergleich der Power der simulierten asymptotischen und exakten Quantile zeigt, daß schon für einen relativ kleinen Stichprobenumfang die Power fast identisch ist.