Working Paper Series, N. 4, September 2018

Department of Statistical Sciences
University of Padua
Italy

Likelihood Asymptotics in Nonregular Settings

A Review with Emphasis on the Likelihood Ratio

Brazzale Alessandra R.
Department of Statistical Sciences
University of Padua
Italy

Valentina Mameli
Department of Environmental Sciences, Informatics and Statistics
Ca' Foscari University of Venice
Italy

Abstract: We review the most common situations where one or some of the regularity conditions which underlie likelihood based parametric inference fail. Three main classes of problems will be treated: boundary problems, indeterminate parameters and singular information matrices, and change-point problems. The review is focused on the large- and small-sample properties of the likelihood ratio, though other approaches to hypothesis testing and connections to estimation will be mentioned in passing.

Keywords: boundary point, change-point, finite mixture, first and higher order theory, identifiability, large- and small-sample inference, singular information

UNIVERSITÀ DEGLI STUDI DI PADOVA
DIPARTIMENTO DI SCIENZE STATISTICHE

Contents

Department of Statistical Sciences
Via Cesare Battisti, 241
35121 Padova
Italy

tel: +39 049 8274168
fax: +39 049 8274170
http://www.stat.unipd.it

Corresponding author:
Alessandra R. Brazzale
tel: +39 049 827 4168
alessandra.brazzale@unipd.it
http://www.stat.unipd.it/~brazzale

Likelihood Asymptotics in Nonregular Settings
A Review with Emphasis on the Likelihood Ratio

Brazzale Alessandra R.
Department of Statistical Sciences
University of Padua
Italy

Valentina Mameli
Department of Environmental Sciences, Informatics and Statistics
Ca' Foscari University of Venice
Italy

Abstract: We review the most common situations where one or some of the regularity conditions which underlie likelihood based parametric inference fail. Three main classes of problems will be treated: boundary problems, indeterminate parameters and singular information matrices, and change-point problems. The review is focused on the large- and small-sample properties of the likelihood ratio, though other approaches to hypothesis testing and connections to estimation will be mentioned in passing.

Keywords: boundary point, change-point, finite mixture, first and higher order theory, identifiability, large- and small-sample inference, singular information

1 Introduction

It is commonly believed that under the null hypothesis the three classical tests of likelihood based inference—that is, those based on the Wald, score and likelihood ratio statistics—are asymptotically equivalent and, to the first order of approximation, follow a chi-squared distribution. Less well known is that this statement in order to hold true requires a number of so-called regularity conditions to be valid. These conditions, which are typically of Cramér type (Cramér, 1946, §33.3), require, among others, differentiability of the underlying joint probability or density function up to a suitable order and finiteness of the Fisher information matrix. Models which satisfy these requirements are said to be 'regular' and cover a wide range of applications. However, there are many important cases where one or more conditions break down. A classical example, which is traditionally used to demonstrate the failure of parametric likelihood theory, is Neyman and Scott's (1948) paradox.

Example 1 (Growing number of parameters). Let $(X_1, Y_1), \dots, (X_n, Y_n)$ denote n independent pairs of mutually independent and normally distributed random variables such that for each $i = 1, \dots, n$, X_i and Y_i have mean μ_i and common variance

σ^2. Maximum likelihood yields the estimator

$$\hat{\sigma}_n^2 = \frac{1}{2n}\sum_{i=1}^{n}\{(X_i - \hat{\mu}_i)^2 + (Y_i - \hat{\mu}_i)^2\},$$

with $\hat{\mu}_i = (X_i + Y_i)/2$. Straightforward calculation shows that, for $n \to \infty$, $\hat{\sigma}_n^2$ converges in probability to $\sigma^2/2$ instead of the true value σ^2. The reason is that only a finite number of observations, in fact two, is available for estimating the unknown sample means μ_i. This violates a major requirement which underlies the consistency of the maximum likelihood estimator, namely that the uncertainty of all parameter estimates goes to zero. □

Example 1 is an early formulation of an incidental parameters problem. Other examples of this type are reviewed in Lancaster (2000), who also discusses the relevance of the Neyman-Scott paradox in statistics and economics. Non-regularity may furthermore arise when the parameter space is constrained and the null hypothesis lies on its boundary, or when some of the parameters disappear under the null hypothesis. These situations are not mere mathematical artifacts, but include a large number of models of practical interest, such as mixture distributions and change-point problems, in fields as genetics, reliability, econometrics, and many more. What happens in these situations is probably unknown to the majority, especially among practitioners. The likelihood ratio, for instance, may still follow a χ^2 distribution, but with degrees of freedom which differ from what is generally expected, as is shown by the following simple example.

Example 2 (Translated exponential distribution). Let $X_1, \ldots, X_n$ be an independent and identically distributed sample from an exponential distribution with rate equal to 1. Consider the translation $Y_i = X_i + \theta$, with $\theta > 0$ unknown. Given $Y_{(1)}$, that is, the minimum observed value, the likelihood ratio statistic for testing the hypothesis that $\theta = \theta_0$ is $W(\theta_0) = 2n(Y_{(1)} - \theta_0)$. Straightforward calculation proves that under the null hypothesis $W(\theta_0)$ distributes like a gamma distribution with shape equal to 1 and rate $1/2$, that is, a χ^2_2 distribution. This differs by 1 degree of freedom from the classical χ^2_1 limiting distribution. □

The purpose of this paper is to present the most common situations where one or more regularity conditions fail. A highly cited review of nonregular problems is Smith (1989). Further examples can be found in Barndorff-Nielsen and Cox (1994, §3.8), Davison (2003, §4.6) and Cox (2006, Chapter 7). The majority of existing results contemplate the failure of one condition at a time, but failure of two assumptions simultaneously has also received consideration. Indeed, quite a bit has been produced to deal with nonregular settings, but the results are scattered across the literature. Since it is nearly impossible to cover all aspects of the subject, here, we will focus on the large- and, where possible, small-sample properties of likelihood based parametric test statistics derived under non-standard conditions, that is, when the likelihood function does not fulfil the usual requirements. Special attention will be paid to the likelihood ratio and its limiting distribution. This choice is justified by the widespread use of Wilks' statistic, together with its chi-squared limiting distribution, in almost all areas of research for hypothesis testing, model selection and other related uses. Although beyond the scope of the paper, available solutions for alternative approaches and/or nonparametric and semiparametric models will be mentioned in passing.

First and higher order parametric inference based on the likelihood function of a regular model is reviewed in Section 2 together with the conditions upon which it is based. However, when these are not fulfilled, deriving the finite and/or asymptotic properties of likelihood pivots can be an utmost challenging task. The likelihood ratio, in particular, may converge to a mixture of chi-squared distributions, such as when the true value of the parameter belongs to the boundary of its parameter space, with mixing proportions which are not easy to determine. Or, its asymptotic behaviour may be characterized as the supremum of a squared truncated Gaussian process, which is the common case for finite mixture models. See Sections 3 and 5, respectively. The literature on the properties of first order solutions under nonregular conditions is rich. Contributions which explore the behaviour of higher order approximations in the same settings are however rare. Though the body of small-sample statistics, which include Bartlett's corrected likelihood ratio (Bartlett, 1937) and Barndorff-Nielsen's modified likelihood root (Barndorff-Nielsen, 1986), has increased rapidly during the last four decades, the focus has almost exclusively been on standard problems. See Brazzale et al. (2007) for a rich collection of examples which illustrate the accuracy and use of higher order theory in real life applications. The reason behind this shortage of results no doubt identifies itself with the methodological difficulties which are raised to the n-th power when higher-order solutions are the target.

In the absence of a unifying theory, most of the individual problems have been treated on their own. We will group them into three broad classes of nonregular settings. The first considers the case were the parameter space is bounded and embraces, in particular, testing for a value of the parameter which lies on its boundary; see Section 3. Section 4 is all about models where one part of the parameter vanishes when the remaining one is set to a particular value. The best studied case of indeterminate parameter problem are finite mixture models. Given their widespread use in statistical practice, and their closeness to boundary problems, we will consider them separately in Section 5. Change-point problems are the third broad class of nonregular models, which we will study in Section 6. Most articles investigate the consequences of the failure of one regularity condition at a time. Mixture distributions and change-point problems deserve special attention as they represent situations where two conditions fail simultaneously. A short discussion is provided in Section 7.

2 Likelihood Asymptotics

2.1 First order theory

2.1.1 General notation.

Consider a parametric statistical model with probability density or mass function $f(y;\theta)$, where the parameter θ takes values in a subset $\Theta \subseteq \mathbb{R}^p$, $p \geq 1$, and $y =$

$(y_1, \ldots, y_n)$ is a vector of n observations from $Y = (Y_1, \ldots, Y_n)$. Throughout the paper we will consider these an independent and identically distributed random sample unless stated differently. Let $L(\theta) = L(\theta; y) \propto f(y; \theta)$ and $l(\theta) = \log L(\theta)$ denote the likelihood and the log-likelihood functions, respectively. The maximum likelihood estimate $\hat{\theta}$ of θ is defined as the value of θ which maximizes $L(\theta)$ or equivalently $l(\theta)$. Under mild regularity conditions on the log-likelihood function—which will be discussed later in Section 2.3—$\hat{\theta}$ solves the score equation $u(\theta) = 0$, where $u(\theta) = \partial l(\theta)/\partial\theta$ is the score function. We furthermore define the observed information function $j(\theta) = -\partial^2 l(\theta)/\partial\theta\partial\theta^\top$ and the expected or Fisher information $i(\theta) = E\left[j(\theta; Y)\right]$, where $\theta^\top$ denotes transposition of θ.

2.1.2 Scalar parameter.

The three classical likelihood based statistics for testing $\theta = \theta_0$ are the

Wald statistic, $(\hat{\theta} - \theta_0)^\top j(\hat{\theta})(\hat{\theta} - \theta_0)$;

score statistic, $u(\theta_0)^\top j(\hat{\theta})^{-1} u(\theta_0)$;

likelihood ratio $W(\theta_0) = 2\{l(\hat{\theta}) - l(\theta_0)\}$,

where the observed information $j(\hat{\theta})$ is at times replaced by the Fisher information $i(\theta)$. If the parametric model is regular, the finite sample null distribution of the above three pivots converges to a chi-squared distribution with p degrees of freedom to the order $O(n^{-1})$ as $n \to \infty$. For θ scalar, inference may be based on the corresponding signed versions, that is, on the signed Wald statistic, $(\hat{\theta} - \theta_0) j(\hat{\theta})^{1/2}$, score statistic, $u(\theta_0) j(\theta_0)^{-1/2}$, and likelihood root,

$$r(\theta_0) = \mathrm{sign}(\hat{\theta} - \theta_0)[2\{l(\hat{\theta}) - l(\theta_0)\}]^{1/2},$$

whose asymptotic distribution is standard normal to the order $O(n^{-1/2})$.

2.1.3 Nuisance parameters.

Suppose now that the parameter $\theta = (\psi, \lambda)$ is partitioned into a p_0-dimensional parameter of interest, ψ, and a vector of nuisance parameters λ of dimension $p - p_0$. Large-sample inference for ψ is commonly based on the profile log-likelihood function

$$l_p(\psi) = l(\psi, \hat{\lambda}_\psi),$$

which we obtain by replacing the nuisance parameter λ in $l(\psi, \lambda)$ by the corresponding constraint maximum likelihood estimate $\hat{\lambda}_\psi$, that is, the value of λ which maximizes the log-likelihood $l(\psi, \lambda)$ for fixed ψ. We may then define the profile Wald, score and likelihood ratio statistics for testing $\psi = \psi_0$ as above, but this time in terms of the profile log-likelihood $l_p(\psi)$, with $u_p(\psi) = \partial l_p(\psi)/\partial\psi$ and $j_p(\psi) = \partial l_p(\psi)/\partial\psi\partial\psi^\top$ being the profile score and profile observed information functions, respectively. The asymptotic null distribution of these statistics is chi-squared with p_0 degrees of freedom up to the order $O(n^{-1})$. If ψ is scalar, the corresponding signed versions, $(\hat{\psi} - \psi_0) j_p(\hat{\psi})^{1/2}$, $u_p(\psi_0) j_p(\psi_0)^{-1/2}$, and

$$r_p(\psi_0) = \mathrm{sign}(\hat{\psi} - \psi_0)[2\{l_p(\hat{\psi}) - l_p(\psi_0)\}]^{1/2}, \qquad (1)$$

may be approximated by the standard normal up to the order $O(n^{-1/2})$.

2.2 Higher order theory

The first order results of Section 2.1 are based on limit theorems of probability theory, such as the central limit theorem and the law of large numbers. These theorems provide valid results for a sufficiently large sample size but, due to their at times rather slow convergence rate, may be unreliable if the number of observations is small. In the past four decades, many efforts went into the development of so-called higher order solutions, with the intent to provide nearly exact approximations to the finite-sample distribution of statistics such as the maximum likelihood estimator and the likelihood root.

2.2.1 Density approximations.

The perhaps most influential asymptotic result of higher order likelihood theory is the so-called p^* approximation

$$\begin{aligned} p^*(\hat\theta|a;\theta) &= c(\theta,a)|j(\hat\theta;\hat\theta,a)|^{1/2} \\ &\quad \times \exp\left\{l(\theta;\hat\theta,a)-l(\hat\theta;\hat\theta,a)\right\}, \end{aligned} \tag{2}$$

a remarkably simple formula for the conditional distribution of the maximum likelihood estimator (Barndorff-Nielsen, 1983). To give the density of the maximum likelihood estimator $\hat\theta$ at each point in its sample space, the p^* approximation re-expresses the data $y = (\hat\theta, a)$ as a one-to-one function of the maximum likelihood estimator $\hat\theta$ and of an ancillary statistic a, which needs be known, at least approximately. The technical details are outlined in Barndorff-Nielsen and Cox (1994, §6.2). The p^* formula is exact in exponential families and regression-scale models, and almost exact in many cases of practical interest (Brazzale and Davison, 2008). Typically, the normalizing constant is $c(\theta, a) = (2\pi)^{-p}\{1 + O(n^{-1})\}$, and (2) approximates the conditional density $p(\hat\theta|a;\theta)$ with relative error $O(n^{-3/2})$. Fraser (1988) showed that the formula has wider application than was originally indicated, and that it corresponds more generally to a saddlepoint approximation (see Appendix A.1).

A concurrent density approximation, called tangent exponential model, was obtained by Fraser et al. (1999). This approximation considers the maximum likelihood estimator $\hat\theta^0$ at a fixed value y^0 of y. The tangent exponential model

$$p_{TEM}(s|a;\theta) = c|j(\hat\varphi)|^{-1/2}\times \tag{3}$$

$$\exp\left\{l(\theta;y^0)-l(\hat\theta^0;y^0)+\left[\varphi(\theta)-\varphi(\hat\theta^0)\right]^\top s\right\},$$

is a local exponential family with sufficient statistic $s = \partial l(\hat\theta^0; y)/\partial\theta$ and canonical parameter

$$\varphi(\theta)^\top = l_{;V}(\theta;y^0) = \sum_{i=1}^n dl(\theta;y)/dy_i\big|_{y=y^0} V_i.$$

It shares the same likelihood function of the original model at the fixed point y^0 and, at this point, has the same first derivative with respect to y. The canonical parameter φ is obtained by differentiating the log-likelihood function $l(\theta; y)$ along the columns of a $n \times p$ matrix V whose rows are tangent to the p-dimensional surface in the n-dimensional sample space defined by conditioning on the ancillary statistic a. Note that $l_{;V}$ corresponds to differentiating the original log-likelihood function $l(\theta; y)$ with respect to $\hat{\theta}$ without having to known the transformation from y to $(\hat{\theta}, a)$ explicitly; see Severini (2000, §6.7.2).

In the continuous case the tangent vectors V can be constructed using a vector of pivotal quantities $z = \{z_1(y_1, \theta), \ldots, z_n(y_n, \theta)\}^\top$ where each component $z_i(y_i, \theta)$ has a fixed distribution. The matrix V is defined from z by

$$V = -\left(\frac{\partial z}{\partial y^\top}\right)^{-1} \left(\frac{\partial z}{\partial \theta^\top}\right)\bigg|_{\theta=\hat{\theta}_0, y=y^0}.$$

A common choice is $z_i = F(y_i; \theta)$, with $F(\cdot; \theta)$ the distribution function of Y_i; other choices for z can be found in Fraser et al. (1999).

2.2.2 Tail area approximations.

From the practical point of view what gives the required p-values for testing and/or for the computation of confidence intervals for the parameter of interest is the cumulative distribution function. For scalar θ, direct integration of (2) and (3) yields the two asymptotically equivalent approximations for $F(\hat{\theta}|a; \theta)$, where $\hat{\theta} = \hat{\theta}(y)$ is the observed data point, of the form

$$\Phi^*(r) = \Phi(r) + \phi(r)\left(\frac{1}{r} - \frac{1}{q}\right)$$

and

$$\Phi(r^*) = \Phi\left\{r + \frac{1}{r}\log\left(\frac{q}{r}\right)\right\}.$$

The first type of approximation was originally derived in Lugannani and Rice (1980), while the second goes back to the seminal paper by Barndorff-Nielsen (1983) and involves the famous modified likelihood root r^*. Here,

$$q = j(\hat{\theta})^{-1/2}\left\{l_{;\hat{\theta}}(\hat{\theta}; \hat{\theta}, a) - l_{;\hat{\theta}}(\theta; \hat{\theta}, a)\right\},$$

where $l_{;\hat{\theta}}(\theta; \hat{\theta}, a) = dl(\theta; \hat{\theta}, a)/d\hat{\theta}$ is a so-called sample space derivative, or

$$q = j(\hat{\theta})^{1/2}|l_{\theta;V}(\hat{\theta})|^{-1}\{l_{;V}(\hat{\theta}) - l_{;V}(\theta)\}.$$

For a scalar parameter of interest ψ, in the presence of nuisance parameters λ, the tail area approximations are the same with the likelihood root r replaced by the profile likelihood root $r_p(\psi)$ (1). The extensions of the two basic expressions for the correction term q are

$$q = \frac{|l_{;\hat{\theta}}(\hat{\theta}) - l_{;\hat{\theta}}(\hat{\theta}_\psi) \quad l_{\lambda;\hat{\theta}}(\hat{\theta}_\psi)|}{|l_{\theta;\hat{\theta}}(\hat{\theta})|}\left\{\frac{|j_{\theta\theta}(\hat{\theta})|}{|j_{\lambda\lambda}(\hat{\theta}_\psi)|}\right\}^{1/2}$$

and

$$q = \frac{|l_{;V}(\hat\theta) - l_{;V}(\hat\theta_\psi) \;\; l_{\lambda;V}(\hat\theta_\psi)|}{|l_{\theta;V}(\hat\theta)|} \left\{ \frac{|j_{\theta\theta}(\hat\theta)|}{|j_{\lambda\lambda}(\hat\theta_\psi)|} \right\}^{1/2} .$$

Here $l_{\theta;\hat\theta}(\theta)$ and $l_{\theta;V}(\theta)$ are so-called mixed derivatives which involve derivation with respect to both, the parameter θ and the sample space.

2.3 Regularity conditions

The first step in the derivation of the large- and small-sample approximations and pivots of Sections 2.1 and 2.2 is typically a Taylor series expansion of the log-likelihood function $l(\theta)$, or quantities derived thereof, in $\hat\theta$ around θ. We may illustrate this point by considering the expansion to the order $O_p(n^{-1/2})$ of the likelihood ratio $W(\theta) = 2\{l(\hat\theta) - l(\theta)\}$ for the scalar parameter case.

Example 3 (Asymptotic expansion of likelihood ratio). Let $p = 1$ and $l_m = l_m(\theta) = d^m l(\theta)/d\theta^m$ be the derivative of order $m = 2, 3, \dots$ of $l(\theta)$, the log-likelihood function for θ in a regular parametric model. Remember that $u = u(\theta) = dl(\theta)/d\theta$ represents the score function, while $i = i(\theta)$ is the Fisher information. Taylor series expansion of $l(\hat\theta)$ around θ yields

$$\begin{aligned} l(\hat\theta) - l(\theta) &= (\hat\theta - \theta)u + \frac{1}{2}(\hat\theta-\theta)^2 l_2 \\ &+ \frac{1}{6}(\hat\theta-\theta)^3 l_3 + \frac{1}{24}(\hat\theta-\theta)^4 l_4 + \cdots . \end{aligned} \tag{4}$$

Rewriting (4) using notation (20) and replacing $(\hat\theta - \theta)$ with expansion (22) of Appendix A.2 yields, after suitable rearrangement of the terms

$$\begin{aligned} l(\hat\theta) - l(\theta) &= \frac{1}{2}i^{-1}u^2 \stackrel{\bullet}{+} \frac{1}{6}i^{-2}\left(i^{-1}u\nu_3 + 3H_2\right) l_2 \\ &\stackrel{\bullet}{+} O_p(n^{-1}). \end{aligned} \tag{5}$$

Here $H_2 = l_2 - \nu_2$, with $\nu_m = E[l_m(\theta;Y)]$, for $m = 2, 3$, while a $\bullet$ is used to mark a drop of $n^{-1/2}$. Given that $i^{-1}u^2$ converges asymptotically to the χ^2_1 distribution, the leading term in (5) leads to the well-known result for Wilks' statistic. See Pace and Salvan (1997, §9.4.4) for the details. □

Results like the above require that the model under consideration is 'regular'. This implies first of all that the derivatives of the log-likelihood function can be carried out to whatever order is required, but also that the asymptotic order of expected values of log-likelihood derivatives is proportional to the sample size. There are several ways in which the required regularity conditions may be formulated; see e.g. Cox and Hinkley (1974, p. 281), Barndorff-Nielsen and Cox (1994, §3.8), Azzalini (1996, §3.2.3), Severini (2000, §4.7), Davison (2003, §4.6). Here, we will assume that the following five conditions on the model function $f(y;\theta)$ hold.

Condition 1 All components of θ are identifiable. That is, two model functions, $f(y;\theta^1)$ and $f(y;\theta^2)$, defined by any two different values $\theta^1 \neq \theta^2$ of θ, are distinct almost surely.

Condition 2 The support of $f(y;\theta)$ does not depend on any component of θ.

Condition 3 The parameter space Θ is a compact subspace of $\mathbb{R}^p$, for a fixed value of $p \in \mathbb{N}\backslash\{0\}$, and the true value θ^0 of θ is an interior point of Θ.

Condition 4 The partial derivatives of the log-likelihood function $l(\theta; y)$ with respect to θ up to the order three exist in a neighbourhood of the true parameter value θ^0 almost surely. Furthermore, in such a neighbourhood, n^{-1} times the absolute value of the log-likelihood derivatives of order three are bounded above by a function of Y whose expectation is finite.

Condition 5 The first two Bartlett identities hold, which imply that

$$E[u(\theta;Y)] = 0 \quad \text{and} \quad i(\theta) = \text{Var}(u(\theta;Y)).$$

Conditions 1–5 are used to justify Taylor series expansion, and other similar techniques, applied to the log-likelihood function and quantities derived thereof. They are relevant in many important models of practical interest. For instance, from the perspective of significance testing, Condition 1 fails when under the null hypothesis parameters defined for the whole model become undefined and therefore not estimable; we will say more about this in Section 4.1. Further examples are treated in Sections 4.2 and 5. Failure of Condition 2 is, for instance, addressed in the papers by Hirano and Porter (2003) and Severini (2004). Furthermore, Condition 2 typically does not hold in change-point problems, which will be treated in Section 6. Failure of Condition 3 characterizes the first and most extensively explored nonregular setting known as boundary problem; see Section 3. A prominent example where Condition 4 is not satisfied, is the double exponential, or Laplace, distribution, which is largely implied in quantile regression. For a book-length review of this topic we refer the reader to the monograph by Koenker et al. (2017). Condition 5 is guaranteed if standard results on the interchanging of integration and differentiation hold provided that Condition 2 is satisfied and that the log-likelihood derivatives are continuous functions of θ. A typical situation where this condition fails is when the data under analysis are derived from a probability density which does not belong to the family $f(y;\theta)$ originally chosen to specify the model, a topic of much investigation in robustness theory (Huber and Ronchetti, 2009). A remedy to this situation is provided by Godambe's theory of estimating equations (Godambe, 1991).

The remainder of the paper reviews the most common situations where one or some of Conditions 1–5 fail.

3 Boundary Problems

Boundary problems represent the first and most extensively explored nonregular setting. Besides, small-sample solutions seem to have been addressed only for this nonregular case. A boundary problem arises when the value θ_0 specified by the null hypothesis, or parts of it, fall on the boundary of the parameter space. Informally, the statistical issues in likelihood based inference occur because the maximum likelihood estimate can only fall on the side of θ_0 which belongs to the parameter space Θ. This implies that if the maximum occurs on the boundary, the score function needs not be zero and the related likelihood pivots won't converge to the typical normal or chi-squared distribution. There is a rich variety of examples in literature of boundary problems which include random effects and frailty models and times series analysis. The following example gives a flavour of the statistical issues.

Example 4 (Bivariate normal). Consider an i.i.d. sample $Y_1, \ldots, Y_n$ from the bivariate normal distribution $N_2(\theta, I_2)$, where $\theta = (\theta_1, \theta_2)$, with $\theta_1 \geq 0$ and $\theta_2 \geq 0$, and I_2 is the 2×2 identity matrix. Straightforward calculation shows that the null distribution of the likelihood ratio for testing $H_0 : \theta = (0, 0)$ versus the alternative that at least one equality does not hold converges to a random variable which can informally be written as

$$\frac{1}{4}\chi_0^2 + \frac{1}{2}\chi_1^2 + \frac{1}{4}\chi_2^2. \tag{6}$$

That is, the finite sample distribution of $W(0)$ converges to a mixture of a point mass χ_0^2 in 0 and two chi-squared distributions, χ_1^2 and χ_2^2, with respectively 1 and 2 degrees of freedom (DasGupta, 2008, Example 21.3). □

Distribution (6) is a special case of the so-called chi-bar squared distribution (Kudô, 1963), denoted by $\bar{\chi}^2(\omega, N)$, with cumulative distribution function

$$\Pr(\bar{\chi}^2 \leq c) = \sum_{\nu=0}^{N} \omega_\nu \Pr(\chi_\nu^2 \leq c),$$

which corresponds to a mixture of chi-squared distributions χ_ν^2 with degrees of freedom ν ranging from 0 to N. In some cases, explicit and computationally feasible formulae are available for the $\bar{\chi}^2$ weights $\omega = (\omega_0, \ldots, \omega_N)$. Extensive discussion on their computation and use, with special emphasis on inequality constrained testing, is given in Robertson et al. (1988, Chapters 2 and 3), Wolak (1987), Shapiro (1985, 1988) and Sun (1988).

3.1 General results

The research on boundary problems was initiated by Chernoff (1954). His 1954 seminal paper outlines the asymptotic null distribution of the likelihood ratio for testing whether θ is on one side or the other of a smooth $(p-1)$-dimensional surface. Following geometrical arguments, Chernoff shows that the limiting distribution is a $0.5\chi_0^2 + 0.5\chi_1^2$ mixture, or differently stated a $\bar{\chi}^2(\omega, 1)$ with $\omega = (0.5, 0.5)$, which assumes the value 0 half of the times and follows a chi-squared distribution with one degree of freedom the other half. This generalizes Wilks (1938) result when the parameter space under the null hypothesis is not necessarily a hyperplane. Feder (1968) extends Chernoff's (1954) results to the case where the true parameter value is "near" the boundary of Θ_0 and Θ_1. Here, Θ_0 and Θ_1 identify the two parameter spaces specified by the null and the alternative hypotheses, respectively. More precisely, Feder (1968) assumes that the true parameter value $\theta_n^0 = \theta^0 + o(1)$ is a sequence of points, not necessarily in Θ_0 or Θ_1, such that θ_n^0 approaches their boundary given by the intersection $\bar{\Theta}_0 \cap \bar{\Theta}_1$ of their complementary subsets $\bar{\Theta}_0$ and $\bar{\Theta}_1$. Two related contributions are Moran (1971) and Chant (1974), who investigate the limiting distribution of the maximum likelihood estimator when θ lies on the boundary of a closed parameter space.

The highly cited article by Self and Liang (1987) further extends Chernoff's (1954) ideas and inspired many researchers and fuelled an enormous literature on the subject. Self and Liang (1987) study the asymptotic null distribution of the likelihood ratio statistic for testing the null hypothesis $H_0 : \theta_0 \in \Theta_0$ versus the alternative $H_1 : \theta_0 \in \Theta_1 = \Theta \setminus \Theta_0$ when θ_0 is on the boundary of Θ. The sets Θ and Θ_0 need be regular enough to be approximated at θ_0 by two cones, C_Θ and C_{Θ_0} with vertex at θ_0 (Chernoff, 1954, Definition 2). They show that as long as their Assumptions 1–4 hold—which translate into our Conditions 1–2 and 4–5, with the addition that the likelihood derivatives are taken from the appropriate side—the likelihood ratio asymptotically converges to

$$\sup_{\theta \in C_{\Theta \setminus \{\theta_0\}}} \left\{ -(Z-\theta)^\top i(\theta_0)(Z-\theta) \right\} - \sup_{\theta \in C_{\Theta_0 \setminus \{\theta_0\}}} \left\{ -(Z-\theta)^\top i(\theta_0)(Z-\theta) \right\}. \tag{7}$$

Here, $C_{\Theta \setminus \{\theta_0\}}$ and $C_{\Theta_0 \setminus \{\theta_0\}}$ are the translations of the cones C_Θ and C_{Θ_0}, respectively, such that their vertices are at the origin, and Z is a multivariate Gaussian distribution with zero mean θ and covariance matrix $i(\theta_0)^{-1}$. Self and Liang (1987) present a number of special cases in which the representation given by (7) is used to derived the asymptotic null distribution of the likelihood ratio, which, in most cases, results in a $\bar{\chi}^2$ distribution. The same scenario is considered in Shapiro (1985) who analyses the asymptotic behaviour of the minimum discrepancy function test statistic. Vu and Zhou (1997) derive the large-sample distribution of estimators obtained from estimating functions for models involving covariates. The non-standard asymptotic distribution of the likelihood ratio statistic for the two-way nested variance components model is derived as an example.

The recent results by Kopylev and Sinha (2011) and Sinha et al. (2012) marked a further major step. These authors consider the asymptotic properties of the likelihood ratio when both, parameters of interest and nuisance parameters are present, either of which, or both, may fall on the boundary of the parameter space. The derivation of a closed form expression for the limiting distribution of the likelihood ratio proves to be a very demanding task in these situations. In general, it can be characterized as a chi-bar squared distribution whose weights depend on the number of interest and nuisance parameters present, and on whether these lie on the boundary of or in the interior of the parameter space. For instance, if we wish to test the null hypothesis $H_0 : \theta_1 = \theta_{10}$ against the alternative $H_1 : \theta_1 > \theta_{10}$ under the assumption that the remaining components, $\theta_2, \ldots, \theta_p$, of the parameter θ are interior points, the asymptotic null distribution of the likelihood ratio is a fifty-fifty mixture of a χ^2_0 and a χ^2_1. This result is in agreement with Case 5 of Self and Liang (1987) of which it represents a special case. Yet, limiting distributions other than the $\bar{\chi}^2$ distribution may be found, as shown by Theorem 2.1 of Sinha et al. (2012). Here, the authors consider the situation where the scalar parameter of interest and two nuisance parameters lie on the boundary of the parameter space. They furthermore show that when nuisance parameters are on the boundary, the use of the classical theory may often be anti-conservative. A concise review of the cases considered in Self and Liang (1987), Sinha et al. (2012) and Kopylev and Sinha (2011), along with some interesting examples and an account of the areas of interest coming from the fields of genetics and biology, is given in Kopylev (2012).

3.2 Null variance components

Linear and generalized linear mixed models are a further area of application, where a boundary problem arises as soon as we want to assess the significance of one or more variance components. Stram and Lee (1994) borrow from Self and Liang's (1987) results to test for a non-zero variance component in a linear mixed effects model for longitudinal data. Zhang and Lin (2008) extended their ideas to generalized linear mixed effects models to test if between-subject variation is absent. Both papers assume that the data vector can be partitioned into a large number of independent and identically distributed subvectors. Crainiceanu and Ruppert (2004) relax this assumption for testing the hypothesis of zero scalar variance component in a linear mixed effect model which may include restrictions. More precisely, they consider a model of the form

$$Y = X\beta + Zb + \varepsilon,$$

where Y is a vector of observations of dimension n, X is a $n \times p$ fixed effects design matrix and Z is a $n \times k$ random effects design matrix, b is a k-dimensional vector of random effects that are assumed to follow a multivariate Gaussian distribution with mean 0 and covariance matrix $\sigma_b^2 \Sigma$ of order $k \times k$, ε is assumed to be independent from b and distributed as a normal random vector with zero mean and covariance matrix $\sigma_\varepsilon^2 I_n$, where I_n is the identity matrix. Suppose we are interested in testing

$$H_0 : \beta_{p+1-q} = \beta^0_{p+1-q}, \ldots, \beta_p = \beta^0_p, \quad \text{and} \quad \sigma_b^2 = 0$$

against

$$H_A : \beta_{p+1-q} \neq \beta^0_{p+1-q}, \ldots, \beta_p \neq \beta^0_p, \quad \text{or} \quad \sigma_b^2 > 0.$$

Crainiceanu and Ruppert (2004, Formula 7) show that the finite-sample null distribution of the likelihood ratio statistic distributes as

$$n\left(1 + \frac{\sum_{s=1}^{q} u_s^2}{\sum_{s=1}^{n-p} w_s^2}\right) + \sup_{\lambda \geq 0} f_n(\lambda),$$

where u_s for $s = 1, \ldots, k$ and w_s for $s = 1, \ldots, n-p$ are independent standard normal variables and $\lambda = \sigma_b^2/\sigma_\varepsilon^2$. The function $f_n(\lambda)$ is defined as

$$f_n(\lambda) = n \log\left\{1 + \frac{N_n(\lambda)}{D_n(\lambda)}\right\} - \sum_{s=1}^{k} \log(1 + \lambda\xi_{s,n}),$$

where

$$N_n(\lambda) = \sum_{s=1}^{k} \frac{\lambda\mu_{s,n}}{1 + \lambda\mu_{s,n}} w_s^2,$$

and

$$D_n(\lambda) = \sum_{s=1}^{k} \frac{w_s^2}{1 + \lambda\mu_{s,n}} + \sum_{s=k+1}^{n-p} w_s^2.$$

Here, $\mu_{s,n}$ and $\xi_{s,n}$ are the k eigenvalues of the matrices $\Sigma^{\frac{1}{2}} Z^T P_0 Z \Sigma^{\frac{1}{2}}$ and $\Sigma^{\frac{1}{2}} Z^T Z \Sigma^{\frac{1}{2}}$, respectively. The matrix $P_0 = I_n - X(X^T X)^{-1} X^T$ is the projection matrix onto the orthogonal complement to the subspace spanned by the columns of the design matrix X.

A similar result is derived for the restricted likelihood ratio test (Crainiceanu and Ruppert, 2004, Formula 9). Both results allow one to simulate the finite-sample null distribution of the likelihood ratio statistic once eigenvalues are calculated. This approach has the advantage of being more efficient than the bootstrap, as the speed of the algorithm only depends on the number of random effects k but not on the number of observations n. Theorem 2 of Crainiceanu and Ruppert (2004) furthermore provides the expression of the asymptotic null distribution of the likelihood ratio statistic based on the asymptotic behaviour of the eigenvalues $\mu_{s,n}$ and $\xi_{s,n}$. The resulting distribution, in general, differs from the fifty-fifty mixture of a χ^2_0 and a χ^2_1 which characterizes the independent identical set-up. Applications of Craineceanu and Ruppert's (2004) results include testing for level- or subject- specific effects in a balanced one-way ANOVA, testing for polynomial regression versus a general alternative described by P-splines and testing for a fixed smoothing parameter in a P-spline regression.

3.3 Constrained one-sided tests

The motivating example for Sinha et al. (2012) comes from the context of multistage dose-response models. For a K-stage model, the multistage assumption leads to a simple dose-response function of the form

$$g(d; \beta) = g(\beta_0 + \beta_1 d + \beta_2 d_2 + \cdots + \beta_K d_K),$$

where d is a tested dose and $g(\cdot)$ is a function of interest such as, for instance, the probability of developing a disease. The coefficients $\beta_k \geq 0$, for $k = 1, \dots, K$, are often constrained to be non-negative so that the dose-response function will be non-decreasing. There is no limit on the number of stages K, though in practice K is usually pre-specified to be no larger than the number of non-zero doses tested. Testing whether $\beta_k = 0$ results in a boundary problem and requires the application of a so-called constrained one-sided test. Besides clinical trials, constrained one-sided tests are common in a number of other areas, where the constraints on the parameter space are often natural. Additional situations are testing for over dispersion, clustering of observations and testing for homogeneity in stratified analyses. All these situations amount to having the parameter value under the null hypothesis lying on the boundary of the parameter space.

A first contribution which evaluates the asymptotic properties of constrained one-sided tests is Andrews (2001). The author establishes the limiting distributions of the Wald, score, quasi-likelihood and rescaled quasi-likelihood ratio pivots under the null and the alternative hypothesis. The results are used to test for no conditional heteroscedasticity in a GARCH(1,1) regression model and zero variances in random coefficient models with possibly correlated coefficients. Sen and Silvapulle (2002) review refinements of likelihood based inferential procedures for a number of parametric, semiparametric, and nonparametric models in the non-standard setup where the parameters are subject to inequality constraints. Special emphasis is placed on their applicability, validity, computational flexibility and efficiency. Again, a central role in characterizing the limiting null distribution of the test statistics is played by the chi-bar squared distribution. For a book-length account of constrained statistical inference, we refer the reader to Silvapulle and Sen (2005).

Molenberghs and Verbeke (2007) compare the performance of the Wald, score and likelihood ratio pivots in multivariate one-sided testing. Their suggestion is to consider the likelihood ratio as the default choice for the constrained case as it can be obtained as in the unconstrained case, without additional computation, provided the constraints are properly imposed onto the alternative model. The use of likelihood ratio tests in frailty models is put forward by Claeskens et al. (2008). Their paper relates to previous results by Maller and Zhou (2003). Under minimal conditions on the censoring distribution, Maller and Zhou (2003) find that the likelihood ratio statistic for homogeneity testing asymptotically distributes as a fifty-fifty mixture of a χ^2_0 and a χ^2_1. Claeskens et al. (2008) extend these results to allow for the presence of covariates while also considering the limiting null distribution of the score pivot. The results are illustrated for gamma and positive stable frailty distributions.

3.4 Small-sample results

To our knowledge, the only contribution which explores the higher order properties of likelihood based test statistics in a nonregular setting—besides the paper by Crainiceanu and Ruppert (2004)—is del Castillo and Lopez-Ratera (2006). The authors consider testing for a boundary point in a scalar exponential family. In particular, they consider the family $\mathcal{F}$ of real valued distributions with probability density function

$$f(y;\theta) = e^{\theta y - \kappa(\theta)} f(y), \quad \theta \in \Theta \in \mathbb{R}, \tag{8}$$

where Θ is the set of parameters for which the function $\kappa(\theta) < +\infty$ is finite. $\mathcal{F}$ is said to be the coniugate family of $f(y)$, obtained from its cumulant generating function $\kappa(\theta)$. If Θ is an open convex set, then model (8) represents a regular exponential family. Otherwise, if Θ includes some of its boundary points $\mathcal{F}$ is called a nonregular exponential model. del Castillo and Lopez-Ratera (2006) characterize the asymptotic distribution of the likelihood ratio for testing the null hypothesis $\theta = 0$, where $\Theta = \{c < \theta \leq 0\}$, when the variance of the model, σ_0^2, is finite. The resulting distribution is a fifty-fifty mixture of a χ^2_1 and a χ^2_0, which is similar to the findings by Self and Liang (1987) where one component of the parameter vector lies on the boundary of its parameter space (Case 5). They furthermore show that, when the moments of the model are finite up to the order four and the density of the sample mean is bounded for some n, the signed likelihood root r, its small sample counterpart r^*, and the standardized maximum likelihood estimator $\sqrt{n}\sigma_0\hat{\theta}$, are asymptotically equivalent. In particular, the asymptotic distribution of $\sqrt{n}\hat{\theta}$ is a fifty-fifty mixture of the constant 0 and the negative truncated normal distribution with zero mean and variance given by $1/\sigma_0^2$. The approach is illustrated for testing exponentiality in reliability theory and survival analysis.

4 Indeterminate parameter problems

The 'indeterminate parameter' problem occurs when setting one of the components of $\theta = (\theta_1, \theta_2)$ to a particular value, say $\theta_1 = \theta_{10}$, leads to the disappearance of some or all components of θ_2. The model is no longer identifiable as all model functions $f(y;\theta)$ with $\theta_1 = \theta_{10}$ and θ_2 arbitrary identify the same distribution. Loss of identifiability occurs in areas as diverse as econometrics, reliability theory and survival analysis (Prakasa Rao, 1992). Traditional methods do not work in this situation and the indeterminate parameter problem has been the subject of intensive research. Already in the early '70s Rothenberg (1971) studied the conditions under which a general stochastic model whose probability law is determined by a finite number of parameters is identifiable. More recently, Paulino and Pereira (1994) present a systematic and unified description of the aspects of the theory of identifiability.

When the parameter which represents the true distribution is not unique, the classical likelihood theory of Section 2 is no longer applicable and various difficulties arise in analyzing the asymptotic properties of the likelihood based pivots. Here, we consider the two cases of non-identifiable parameters and singular information matrix.

4.1 Non-identifiable parameters

Liu and Shao (2003) develop a general framework for deriving the asymptotic null distribution of the likelihood ratio statistic for the set of hypotheses $H_0 : \theta \in \Theta_0$ against $H_1 : \theta \in \Theta \setminus \Theta_0$, where $\Theta_0 = \{\theta \in \Theta : P_\theta = P^0\}$ with P_θ being the distribution indexed by θ and P^0 the true distribution. The main idea is to establish a quadratic approximation of the likelihood ratio in a Hellinger neighbourhood of P^0 instead of the Euclidean neighbourhood as it is the common case when the model is identifiable. Under suitable regularity conditions, which assure Hellinger consistency of the maximum likelihood estimator despite loss of identifiability, the asymptotic distribution of the likelihood ratio pivot converges to the supremum of the square of a left-truncated centered Gaussian process with uniformly continuous sample paths (Liu and Shao, 2003, Theorem 2.3). If the likelihood ratio is square integrable, similar results are derived but by using a quadratic approximation of the likelihood ratio based on the Pearson type $\mathcal{L}^2$ distance (Liu and Shao, 2003, Section 3). The results are illustrated using an example on testing the number of components in a finite mixture model.

Motivated by the intent to test a random intercepts model for repeated measurements against an alternative covariance structure allowing for serial correlation, Ritz and Skovgaard (2005) derive the asymptotic distribution of the likelihood ratio statistic, and of the related score test, for a general curved exponential family for which some nuisance parameters vanish under the null hypothesis. The results are then exemplified on the multivariate normal model whose covariance matrix can be written as

$$(\varphi - \varphi_0)\Sigma(\rho) + \gamma_1\Sigma_1 + \cdots + \gamma_k\Sigma_k$$

where $\varphi, \rho, \gamma_1, \ldots, \gamma_1$ are unknown variance parameters and $\Sigma(\rho), \Sigma_1, \ldots, \Sigma_k$ suitable matrices. The null hypothesis $\varphi = \varphi_0$ reduces the model to a random coefficients model, while making the parameter ρ non-identifiable at the same time. The general results are derived without the need to assume compactness of the parameter space, a condition which, as we will see in Section 5, is generally required when some parameter are non-identifiable. The numerical investigation carried out by Ritz and Skovgaard (2005) shows that the limiting distribution of the motivating example lies between a $\bar{\chi}^2(\omega, 1)$, with $\omega = (0.5, 0.5)$ and a χ^2_2 distribution, thus differing somewhat from the commonly encountered situations. The authors furthermore show that their approximation performs well with small or moderate sample sizes, and remains stable over a wide range of parameter values.

Beyond the scope of this paper, though of substantial value because of the insight which the authors provide into the non-identifiable parameters problems, are two papers by Davies and the contributions derived thereof. Davies (1977) investigates the construction of optimal likelihood based tests under loss of identifiability for a two-parameter model when the test statistic is normal. The chi-squared case is considerd in Davies (1987). Asymptotically optimal tests for the same situation are treated in Andrews and Ploberger (1994), who apply their results to tests of one-time structural change with unknown change-point and discuss several other examples. An extension to semi parametric models is considered Song et al. (2009).

Recently, Fortunati et al. (2012) extended a result, which connects parameter identifiability to non-singularity of the information matrix, to the situation in which nuisance parameters are present. This links us to the second case of indeterminate parameter problem we address in this paper.

4.2 Singular information matrix

Strictly connected to the non-identifiable parameters problem is the situation where Fisher's information matrix is singular at the true value θ^0 of the parameter. Singularity of $i(\theta)$ can lead to multiple maxima of the log-likelihood function $l(\theta)$ in a neighbourhood of θ^0 and to inconsistency of the maximum likelihood estimator $\hat{\theta}$. Moreover, the limiting distribution of the likelihood ratio test statistic may not be of chi-squared type.

A first contribution to this topic is the paper by Aitchison and Silvey (1960) who address the problem of singular information matrix when the null hypothesis is specified by suitable constraints on the parameters to allow their estimation. See their Section 6 and previous work by Silvey (1959) on the asymptotic properties of suitably constraint maximum likelihood estimators in the presence of singular information matrices. Aitchison and Silvey (1960) discuss the use and corresponding merits of the two large-sample techniques, which are unconstraint maximum likelihood estimation with its associated Wald test and constrained maximum likelihood estimation with its associated score type test. El-Helbawy and Hassan (1994) also build upon the results by Silvey (1959). In particular, they develop modified formulae for the Wald, score and likelihood ratio pivots which, under standard regularity conditions, follow asymptotically a chi-squared distribution with degrees of freedom specified by the number of constraints. Similarly, Barnabani (2002) proposes to maximize the modified log-likelihood function defined by

$$Q_n(\theta) = \lim_{c \to 0} \left(\frac{1}{n} l(\theta) - \frac{c}{2} ||\theta - \theta^0||^2 \right)$$

when $i(\theta)$ is singular. Here, the constant c governs the quadratic penalization of the normalized log-likelihood function $l(\theta)$, expressed as the squared Euclidean distance of θ from the true value θ^0. The corresponding estimator of θ is consistent and asymptotically normal. We may hence construct a Wald type test statistic which has a chi-squared distribution both under the null and the alternative hypotheses. A most recent contribution is by Jin and Lee (2018) who propose to fit the parameters of models with singular information matrix by adaptive lasso estimators. This papers generalizes Jin and Lee (2017) by allowing the true parameter vector to be on the boundary of the parameter space.

The cornerstone contribution to the development of the theory of singular information matrices is, however, Rotnitzky et al. (2000). These authors address the problem of deriving the asymptotic null distribution of the likelihood ratio statistic for testing the null hypothesis $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$, when θ is a p-dimensional parameter of an identifiable parametric model and the information matrix is singular at θ_0 and has rank $p - 1$. The theory is developed only for independent and identically distributed random variables, though the authors point out that the procedure may straightforwardly be extended to account for independent and non-identically distributed observations. When θ is scalar, the asymptotic properties of the maximum likelihood estimator and of the likelihood ratio statistic depend on the integer m_0, defined as the order of the first partial derivative of the log-likelihood function which does not vanish at $\theta = \theta_0$; see Theorems 1 and 2, respectively, of Rotnitzky et al. (2000). Specifically, if m_0 is odd, the likelihood ratio converges under the null hypothesis to a χ^2_1 random variable with 1 degree of freedom, while for even m_0 it converges to a fifty-fifty $\bar{\chi}^2(\omega, 1)$ mixture. Extensions of these results when the parameter θ is p-dimensional are provided in Theorems 3 and 4 of Rotnitzky et al. (2000).

5 Finite mixture models

Finite mixture models deserve special attention, because of their widespread use in statistical practice but also because of the methodological challenges posed by the derivation of their asymptotic properties. They probably represent the best studied case of an indeterminate parameter problem, though we may equally look at them as a boundary problem. Indeed, testing an hypothesis such as model homogeneity against the alternative that the model be a finite mixture of two or more components will most likely lead to the concurrent failure of two regularity conditions. As we will see in Section 5.1, this occurs because while under the null hypothesis the mixing proportions fall on the boundary of their parameter space, some of the parameters of the corresponding component distributions become indeterminate. Under this set-up, the asymptotic distribution of the likelihood ratio statistic does not follow the commonly believed chi-squared distribution, and its limiting distribution has for long been unknown.

The remainder of the section outlines the many contributions for this class of models with special emphasis on hypothesis testing using the likelihood ratio. A general reference for mixture distributions is the monograph by Lindsay (1995).

5.1 Testing for homogeneity

Consider the two-component mixture model

$$\pi f_1(y;\theta_1) + (1-\pi) f_2(y;\theta_2), \tag{9}$$

where the probability density or mass functions $f_1(y;\theta_1)$ and $f_2(y;\theta_2)$, with $\theta_1 \in \Theta_1 \subseteq \mathbb{R}^{p_1}$ and $\theta_2 \in \Theta_2 \subseteq \mathbb{R}^{p_2}$, represent the mixture components and $0 < \pi < 1$ is the mixing probability. The null hypothesis of homogeneity can be written in different ways. We may set $\pi = 1$, which corresponds to $H_0 : f^0 = f_1(y;\theta_1)$, where f^0 represents the true unknown distribution, or alternatively, $\pi = 0$ and $H_0 : f^0 = f_2(y;\theta_2)$. If the two components, $f_1(y;\theta_1)$ and $f_2(y;\theta_2)$, are known, then the limiting distribution is a $\bar{\chi}^2(\omega, 1)$ with $\omega = (0.5, 0.5)$ (Lindsay, 1995, p. 75). Otherwise, for $f_1(y;\theta) = f_2(y;\theta)$ a third possibility arises: in this case homogeneity assumes that $H_0 : \theta_1 = \theta_2$. Whatever choice is made, some model parameters—θ_2 and θ_1, respectively, in the first two cases and π in the third—vanish under the null hypothesis.

In classical likelihood theory, the parameter which characterizes the true distribution is typically assumed to be a unique point θ^0 in the open subset $\Theta \subseteq \mathbb{R}^p$. The classical regularity conditions of Section 2.3 imply consistency of the maximum likelihood estimator $\hat{\theta}$, which combined with the existence of a quadratic approximation to the log-likelihood function $l(\theta)$ in an Euclidean $n^{-1/2}$-neighborhood of θ^0, ensure the standard limiting distributions of Section 2. As we have seen in Section 3, the failure of Condition 3 generally implies that the limiting distribution is truncated on its left to account for the fact that the maximum likelihood estimate can only fall on one side of the true parameter value. The failure of Condition 1 in addition implies that there is no value to which the maximum likelihood estimate of the indeterminate parameters can converge.

The first discussion of asymptotic theory for testing homogeneity of model (9) when all parameters are unknown goes back to Ghosh and Sen (1985). As pointed out in their paper, there is an additional major difficulty in dealing with finite mixture models: though the mixture itself may be identifiable, the parameters π, θ_1 and θ_2 may not be. For instance, for the simple mixture where $f_1(y;\theta) = f_2(y;\theta) = f(y;\theta)$, the equality

$$\begin{aligned} &\pi f(y;\theta_1) + (1-\pi) f(y;\theta_2) \\ &\qquad = \pi' f(y;\theta_1') + (1-\pi') f(y;\theta_2') \end{aligned}$$

holds for $\pi = \pi'$, $\theta_1 = \theta_1'$, $\theta_2 = \theta_2'$, but also for $1-\pi = \pi'$, $\theta_1 = \theta_2'$, $\theta_2 = \theta_1'$. That is, if the alternative hypothesis is true, there is a second set of parameters which gives rise to exactly the same distribution. Furthermore, under the null hypothesis of homogeneity the model is represented by the three curves $\pi = 1$, $\pi = 0$ and $\theta_1 = \theta_2$. As illustrated by Ghosh and Sen (1985), choosing an identifiable parameterisation doesn't solve the problem as then the density is no longer differentiable.

The first result derived by Ghosh and Sen (1985) characterizes the limiting distribution of the likelihood ratio statistic for strongly identifiable continuous mixtures. Write $f(y;\theta) = \pi f_1(y;\theta_1) + (1-\pi) f_2(y;\theta_2)$ with the convention that $\theta = (\pi, \theta_1, \theta_2)$. Strong identifiability holds if $f(y;\theta) = f(y;\theta')$ implies that $\pi = \pi'$, $\theta_1 = \theta_1'$ and $\theta_2 = \theta_2'$. Ghosh and Sen (1985) furthermore assume that Θ_1 is a closed bounded interval of $\mathbb{R}$, while $\Theta_2 \subseteq \mathbb{R}^p$. The likelihood ratio statistic for testing $H_0 : \pi = 0$ then converges to $W^2 I_{\{W>0\}}$, where $W = \sup_{\theta_1}\{Z(\theta_1)\}$ and $Z(\theta_1)$ is a zero mean Gaussian process on Θ_1 whose covariance function depends on the true value of the parameters (Ghosh and Sen, 1985, Theorem 2.1). As is well presented in Section 2 of their paper, this results from first approximating the likelihood ratio statistic by a quadratic expansion of the log-likelihood function with respect to the parameters present under the null hypothesis, π and θ_2 and then taking its supremum with respect to the non-identifable parameter θ_1. A similar result holds if the finite mixture is not strongly identifiable, such as when $f_1(y;\theta) = f_2(y;\theta)$, but a separatation condition between θ_1 and θ_2 of the form $||\theta_1 - \theta_2|| \geq \epsilon$ for a fixed quantity $\epsilon > 0$ is imposed so that H_0 is described by either $\pi = 0$ or $\pi = 1$; see Section 5 of Ghosh and Sen (1985). The authors furthermore take $0 \leq \pi \leq 0.5$ and require Θ_2 to be an open set containing the true value θ_2^0 and Θ_1 a closed set such that $\Theta_1 \cap \Theta_2 = \emptyset$. Indeed, these two conditions guarantee that the maximum likelihood estimate $(\hat{\pi}, \hat{\theta}_2)$ will fall with high probability into the $n^{-1/2}$-neighbourhood of $(0, \theta_2^0)$ and the null hypothesis is $H_0 : \pi = 0$. Removing this separation condition has become a challenging problem and many authors have addressed it. Hartigan (1985), for instance, showes that in the Gaussian case the likelihood ratio statistic diverges to infinity in probability if the mean parameters are unbounded. Ghosh and Sen (1985) also discuss the link to Bayesian testing and develop asymptotically locally minimax tests for some special cases; see Section 4 of their paper.

An example of the use of the likelihood ratio for homogeneity testing in discrete setting is Chernoff and Lander (1995). These authors study several versions of the two-component binomial mixture model motivated by a problem of interest in genetics. They show that the likelihood ratio statistic converges to the supremum of the square of a left-truncated zero-mean unit-variance Gaussian process with well behaved covariance function. Likelihood ratio tests for genetic linkage—which are tests of homogeneity in mixtures of binomial distributions—are furthermore studied by Lemdani and Pons (1997) for several classical models. Using ad hoc reparametrizations, the corresponding null distributions are shown to converge once again to the supremum of a squared left-truncated Gaussian processes under the null hypothesis.

Böhning et al. (1994) investigate numerically the asymptotic properties of the likelihood ratio statistic for testing homogeneity in the two-component mixture model (9) when the component distributions $f_k(y;\theta_k)$, $k = 1, 2$ belong to an exponential family and, more precisely, are binomial, Poisson, exponential or Gaussian with known common variance. They establish that, for sufficiently large sample sizes, the unknown null distribution is well approximated by a $\bar{\chi}^2(\omega, 1)$ which remains stable across the possible range of values for the parameters θ_1 and θ_2, but is model specific in the sense that the weigths ω depend on the model under consideration. Ciuperca (2002) considers homogeneity testing for the two-component mixture where $f_1(y;\theta)$ belongs to an exponential family and $f_2(y;\theta,\tau) = f_1(y-\tau;\theta)$ is a translation thereof by an unknown quantity $\tau \in \mathbb{R}$. The limiting distribution results in a fifty-fifty mixture of a point mass in zero and a random variable which diverges in probability to $+\infty$ even though the parameters are assumed to belong to a compact set. A theoretical motivation is given why this happens and how this result connects to the situation treated in Hartigan (1985). Liu et al. (2003) characterize the asymptotic behaviour of the likelihood ratio for testing homogeneity against a two-component gamma mixture with known shapes and where one of the rate parameters is known. They show that under the null hypothesis this asymptotic distribution agrees with the square of Davies's (1977) statistic for the Gaussian process test. Furthermore, if the unknown rate parameter belongs to an unbounded set, the likelihood ratio diverges to infinity in probability at a convergence rate of $\log(\log n)$, in accordance with Hartigan (1985).

Lemdani and Pons (1999) study the limiting distribution of the likelihood ratio statistic to test whether a known density f_1 is contaminated by another density f_2 of the same parametric family. Using this time a general reparametrization which ensures regularity properties, the likelihood ratio statistic is shown to converge again to the supremum of a squared left-truncated Gaussian process. The result is extended to the case where a mixture of K_0 known densities is contaminated by K_1 other ones of the same family—a situation which we will further treat in Section 5.3. Chen and Chen (2001b) consider the same setting than Böhning et al. (1994), though the components distributions are now allowed to belong to a generic parametric family. They show that under suitable conditions, which guarantee identifiability of the mixture and regularity of the component distributions $f_k(y;\theta_k)$, the likelihood ratio converges to the squared supremum of a left-truncated standard Gaussian process whose autocorrelation function is explicitly presented; see Sections 2 and 3 of their paper. Chen and Chen (2001b) furthermore recommend to use resampling to calculate the desired tail probabilities. The procedure is illustrated for three different component distributions, that is, the normal, binomial and Poisson.

A rather different route is taken in Chen et al. (2001). To overcome the two difficulties of asymptotic theory for mixture models—that is, the boundary problem and non-identifiability under the null hypothesis— they suggest to penalise the log-likelihood function

$$l(\pi,\theta;y) + c\log\{4\pi(1-\pi)\}, \tag{10}$$

where c is a constant chosen so as to suitably control the penalisation. As the authors point out, the penalisation term can be justified both, from the Bayesian perspective and referring to a conceptual auxiliary experiment. It furthermore guarantees that the maximum likelihood estimate of the mixing proportion $0 < \hat{\pi} < 1$ will not fall on the boundary of the parameter space and that the maximum likelihood estimates of all parameters are consistent under the null hypothesis $H_0 : \pi = 0$. Provided conditions 1–5 of their paper hold, the modified likelihood ratio statistic derived from (10) converges to a $\bar{\chi}^2(\omega, 1)$ with $\omega = (0.5, 0.5)$, that is, to a fifty-fifty mixture of a point mass in zero and a chi-squared distribution with one degree of freedom. Numerical assessment on Poisson and Gaussian mixtures reveals that their proposal well competes with alternative solutions especially with respect to power. In a later paper, Chen et al. (2008) derive the asymptotic distribution of the modified likelihood ratio test introduced in Chen et al. (2001) and of a further modification, called the iterative modified likelihood ratio test, for testing homogeneity against the alternative that the model is a two-component von Mises mixture where the two mean directions are both unknown. Both mentioned papers make suggestions of how to improve the accuracy of the asymptotic approximation in finite samples.

An overview of asymptotic results for testing homogeneity for model (9) together with the illustration of some new results such as the calculation of tail probabilities and asymptotic power, both in the case of a bounded and an unbounded parameter space, is provided in Garel (2007).

5.2 Gaussian mixtures

Theoretical results are particularly generious if the two-component model is a normal mixture. Goffinet et al. (1992) consider an i.i.d. sample from a d-dimensional random variable with density function

$$\pi\phi_d(y; \mu_1, \Sigma) + (1 - \pi)\phi_d(y; \mu_2, \Sigma),$$

with $0 \leq \pi \leq 1$ and $\phi_d(y; \mu, \Sigma)$ the d-dimensional normal density, $d \geq 1$, with mean $\mu \in \mathbb{R}^d$ and covariance matrix Σ. They derive the asymptotic distribution of the likelihood ratio statistic for testing the null hypothesis of homogeneity of the means, that is, $H_0 : \mu_1 = \mu_2$, assuming the mixing proportion π is known. Theorem 1 of Goffinet et al. (1992) treats the univarate case, while its bivariate extension is given in their Theorem 2. In short, for $d = 1$ the null distribution of the likelihood ratio converges to a χ^2_1 distribution if Σ is unknown and $\pi \neq 0.5$; otherwise it converges to a $\bar{\chi}^2(\omega, 1)$ with $\omega = (0.5, 0.5)$. The convergence rate depends on the value of the known mixing proportion π and is particularly slow if π is close to 0.5. If $d = 2$ the limiting distribution of the likelihood ratio for known Σ is

$$\frac{1}{2}\{\sup(0, T)\}^2, \quad \text{with} \quad T = Z + \sqrt{W},$$

where Z is the standard normal and W is an independent χ^2_2 distribution. This corresponds to a fifty-fifty mixture of a point mass in zero and the sum of a standard normal plus the square root of an independent χ^2_2. No result is given for $d = 2$ and Σ unknown. Polymenis and Titterington (1999) further analyse empirically the $d = 1$ scenarios treated by Goffinet et al. (1992) and give an heuristic explanation for the slow convergence. In particular, they propose to refer to the $\bar{\chi}^2(\tilde{\omega}, 1)$ distribution with suitably defined mixing proportions $\tilde{\omega}$ instead of the theoretical value $\omega = (0.5, 0.5)$ to improve the approximation of the null distribution of the likelihood ratio test in finite samples.

Chen and Chen (2001a) consider the slightly different univariate setting

$$\pi\phi(y; \mu_1, 1) + (1 - \pi)\phi(y; \mu_2, 1), \tag{11}$$

where $\phi(y; \mu, 1)$ is the univariate normal density with unit variance and unknown mean $\mu \in \mathbb{R}$. Differently from above, the mixing proportion π is unknown and the two means are bound to lie in a finite interval of the form $|\mu_1| \leq M$ and $|\mu_2| \leq M$ for positive M. They consider two situations: where either only one of the means is unknown, say μ_1, and the second, $\mu_2 = 0$, is set to zero, or where both location parameters are unknown. In both cases the asymptotic null distribution of the likelihood ratio statistics for testing homogeneity involves the supremum of a squared Gaussian random field. If both means are unknown and $\pi \leq 0.5$ to assure identifiability, the limiting distribution is $\{\sup_{|t| \leq M} Z(t)\}^2 + W$, where $Z(t)$, $t \in [-M, M]$, is a Gaussian process and W is an independent chi-squared random variable with one degree of freedom. The Gaussian process $Z(t)$ has zero mean and covariances

$$Cov\{Z(s), Z(t)\} = \frac{e^{st} - 1 - st}{\sqrt{(e^{s^2} - 1 - s^2)(e^{t^2} - 1 - t^2)}},$$

for $st \neq 0$, and $Cov\{Z(s), Z(t)\} = 0$ when $st = 0$; see Theorem 3 of Chen and Chen (2001a). If only one of the two means is unknown, the chi-squared term is absent and the expression of the covariance is slightly different and given in Theorem 2 of Chen and Chen (2001a). The generalization of these results to the two-component mixture model

$$\pi\phi(y; \mu_1, \sigma^2) + (1 - \pi)\phi(y; \mu_2, \sigma^2), \tag{12}$$

which now includes an unknown variance parameter $\sigma^2 > 0$, can be found in Chen and Chen (2003). They prove that the asymptotic distribution of the likelihood ratio for testing homogeneity is the supremum between a chi-squared random variable with two degrees of freedom and the square of a left-truncated Gaussian process with zero mean and unit variance. Again, the correlation structure of the process involved in the limiting distribution is presented explicitly; see their Theorem 2. Liu and Shao (2004) show that the likelihood ratio is asymptotically equivalent to the square of the supremum of the stochastic process studied in Theorem 1 of Bickel and Chernoff (1993). They furthermore find that the likelihood ratio diverges to $+\infty$ at the rate of $O\{\log(\log n)\}$ if the mean parameters are unbounded, as already conjectured by Hartigan (1985). Note that Hartigan's (1985) finding also results from Theorem 2 of Chen and Chen (2001a).

Qin and Smith (2004) consider the same model than Chen and Chen (2003) and the same restrictions on the mean parameters which are also given in Chen and Chen (2001a). Again identifiability is guaranteed by setting $\pi \leq 0.5$. In addition they assume that the mixing proportion satisfies the condition $\min(\pi, 1 - \pi) \geq \epsilon$ for some positive $\epsilon < \frac{1}{2}$. The likelihood ratio then follows asymptotically a fifty-fifty mixture of a χ^2_1 and a χ^2_2 distribution under the hypothesis of homogeneity.

The same results holds for the likelihood ratio based upon Chen's et al. (2001) modified log-likelihood (10) in addition to the fact that the second condition on the mixing proportion can be relaxed. Qin and Smith (2006) generalize their results on the asymptotic properties of the likelihood ratio statistic for testing homogeneity to a bivariate normal mixture model with known covariance matrix; see their Theorem 1. In practice, the limiting distribution must be found numerically though an approximation is provided in Section 4 of their paper. The corresponding two-sample problem is treated in Lei and Qin (2009).

Seven distinct cases of homogeneity testing using the likelihood ratio for the general two-component Gaussian mixture model

$$\pi\phi(y;\mu_1,\sigma_1^2)+(1-\pi)\phi(y;\mu_2,\sigma_2^2), \tag{13}$$

obtained by imposing different restrictions on the means μ_1 and μ_2 and the variances σ_1^2 and σ_2^2, are discussed in Garel (2001).

A different route was taken by a number of authors who look into the asymptotic null distribution of the likelihood ratio statistic for normal mixtures using simulation. Thode et al. (1988) consider testing the hypothesis that the sample is drawn from a normal distribution with unknown mean and unknown variance against the alternative that the sample comes from the two-component normal mixture (13) with $\mu_1 \neq \mu_2$ and common variance $\sigma_1^2 = \sigma_2^2 = \sigma^2$. All model parameters, the mixing proportion π included, are assumed to be unknown. Their extensive numerical investigation shows that the likelihood ratio statistic converges very slowly to a limiting distribution, if any exists, and is rather unstable even for sample sizes as large as $n = 1,000$. For very large sample sizes, the empirical distributions rather closely agree with the commonly assumed χ_2^2, while this limiting distribution may too liberal for small to moderate n. This gives little support to Hartigan's (1977) conjecture that the asymptotic distribution may lie between a χ_1^2 and a χ_2^2. An example of application to a study of population genetics is given motivated by the fact that these studies are typically of small to moderate sample sizes, which justifies the use of empirical techniques in place of asymptotic approximation. The distribution of the likelihood ratio under the alternative hypothesis (13) is investigated numerically in Mendell et al. (1991) for a wide range of mixing proportions π. The authors conjecture that the limiting distribution is a non-central chi-squared with 2 degrees of freedoms. Lo (2008) shows that the commonly believed χ^2 approximation for testing the null hypothesis of a homoscedastic normal mixture against the alternative that the data arise from a heteroscedastic model only works for samples sizes as large as $n = 2,000$ and component distributions which are well separated under the alternative. Furthermore, the restrictions of Hathaway (1985) need be imposed to ensure that the likelihood is bounded and to rule out spurious maxima under the alternative distribution. Otherwise, the author suggests to use parametric resampling.

5.3 Testing the number of components

Consider the general K-component mixture model

$$\sum_{k=1}^{K}\pi_k f_k(y;\theta_k), \quad K \geq 2, \tag{14}$$

where $f_k(y;\theta_k)$ are K probability density or mass functions indexed by $\theta_k \in \Theta_k \subseteq \mathbb{R}^{p_k}$ and $0 < \pi_k < 1$, $k = 1,\ldots,K$, with $\sum_{k=1}^{K}\pi_k = 1$. Developing a formal test for the null hypothesis that $H_0 : K = K_0$ against the alternative that the mixture includes $K > K_0$ components is a difficult task. Many routes have been taken, which include Wald type statistics derived from moment or alternative estimators, adaptation of model selection techniques and the use of simulation. For instance, using the findings of Vuong (1989), who develop likelihood ratio tests for non-nested models, Lo et al. (2001), claim that in the Gaussian case the likelihood ratio statistic based on the Kullback-Leibler information criterion converges under the null hypothesis to a weighted sum of independent chi-squared random variables with one degree of freedom. Jeffries (2003) disproves this result based on the fact that it requires conditions on the structure of the parameter space to hold which are generally not met when the null hypothesis of a K_0-component model holds. Oliveira-Brochado and Martins (2005) give a partial review of these techniques. Here we want to focus on the proper likelihood ratio test and its asymptotic distribution.

Using the inequalities on likelihood ratios developed in Gassiat (2002), Azaïs et al. (2006) provide the asymptotic distribution of the likelihood ratio statistic under the null hypothesis of a K_0-component model as well as under contiguous alternatives for a general mixture of parametric populations for a bounded parameter space. More precisely, if we define $\mathbb{K} = [-K, K]$ and let $\mathcal{F} = \{f_k, k \in \mathbb{K}\}$ be a parametric set of probability densities on $\mathbb{R}$, they consider the testing problem

$$H_0 : f^0 = f_0 \quad \text{against} \quad H_1 : f^0 : \pi f_0 + (1 - \pi) f_k,$$

with $k \in \mathbb{K}$ and $0 \leq \pi \leq 1$. In the particular case of Gaussian components, they prove that if the parameter space is unbounded, the likelihood ratio statistic won't be able to distinguish the null hypothesis from any contiguous alternative. A by-product of their paper is the characterization of the asymptotic properties of the likelihood ratio statistic for testing homogeneity of the means in the two-component normal mean mixture model of Section 5.2. In Azaïs et al. (2009) the same authors consider likelihood ratio testing for the general K-component model (14). In particular, they analyze two situations: testing homogeneity against any mixture, with application to Gaussian, Poisson and binomial distributions, and testing for the number of populations in a finite mixture with or without a structural parameter. A number of conditions are imposed that are proved to be almost necessary to avoid that the limiting distribution diverges to infinity. Kasahara and Shimotsu (2012) derive the asymptotic distribution of the likelihood ratio statistic for the simpler case of testing whether the mixture contains K_0 components or $K_0 + 1$ components. They furthermore propose a likelihood-based procedure for generally identifying the number of components. Chen and Kalbfleisch (2005) study a modification of the likelihood ratio statistic similar to the one proposed by Chen et al. (2001) to verify the hypothesis of a homogeneous model against the alternative of a Gaussian mixture of two or more components with a common and unknown variance. In particular, they show that the χ^2_2 distribution represents a stochastic upper bound to its limiting null distribution.

6 Change-point problems

A change-point problem arises whenever we want to identify a possible change in the probability distribution of the data, may these be a univariate or multivariate random sequence, time-dependent observations, or a sample of responses whose mean function depends on a number of covariates. A change in the data generating process generally affects the support of the random variable and/or implies that the log-likelihood function is no longer differentiable with respect to some values of the parameter. This typically leads to the failure of Conditions 2 and 4 of Section 2.3.

Change-point problems have been the subject of much intensive research owing to their widespread use in all those area where the constancy over time of random events is questioned. The theory has evolved over the past five decades to the extend that it is inconceivable to summarize all contributions. In particular, we won't tackle the problem of detecting a change in sequential analysis and quality control. A 240 pages long collection of bibliographic references for change-point problems in regression can be found in Khodadadi and Asgharian (2008). Lee (2010) presents the synopsis of the most recent literature up to his time of writing together with a comprehensive bibliography for the five types of change-point problems characterized by either a shift in the mean, a change in the variance, a switch in the regression slope, a change in the hazard rate or, more generally, in the distribution. For a book-length account of change-point problems with examples from medicine, genetics and finance, we refer the reader to the monograph by Chen and Gupta (2012).

The simplest situation of change-point problem occurs when we want to identify patterns in a sequence of random variables. A very early contribution is Page (1957) who considers the problem of testing the null hypothesis that all n observations in an independent sample come from the same population with probability distribution function $F(y;\theta)$ against the alternative that only the first τ, $0<\tau<n$ are generated from $F(y;\theta)$ while the remaining come from $F(y;\theta')$ with $\theta \neq \theta'$ and τ unknown. We will come back to this problem towards the end of the section. In a previous paper, Page (1955) considered the simpler situation of a change in the mean of a distribution, which has become one of the most frequently studied change-point problems. Under this set-up the model can generally be written as

$$Y_i = \eta_i + \varepsilon_i, \quad i = 1, \ldots, n, \tag{15}$$

where the ε_i's are independent zero-mean random errors. The function η_i is supposed to have been subjected to K changes in time, that is,

$$\begin{aligned} \eta_i &= \mu_1, \quad i \leq \tau_1, \\ &= \mu_2, \quad \tau_1 < i \leq \tau_2, \\ &\vdots \\ &= \mu_{K+1}, \quad \tau_K < i \leq n. \end{aligned}$$

Note that throughout the remainder of the section we will consider the observations in the order they appear. Both, the $K+1$ different mean values μ_k and the K change-points τ_k are supposed to be unknown. Two inferential problems are of interest: identifying the unknown number of changes and estimating where these occur and how large they are. Numerous inferential approaches have been advocated to deal with change-points. These include parametric and nonparametric techniques from both, the frequentist and the Bayesian framework. Here, we will again focus on testing problems based on the parametric likelihood ratio and its asymptotic distribution.

6.1 Mean-shift model

Hinkley (1970) considers model (15) with a single change-point τ, that is, for $K = 1$. In particular, borrowing from the theory of random walks he derives the asymptotic distribution of the maximum likelihood estimator of τ and of the likelihood ratio statistic for testing the null hypothesis $H_0 : \tau = \tau_0$, that is, that the change occurred at τ_0, when the errors $\varepsilon_i \sim N(0, \sigma^2)$ are centered normal variables with constant variance $\sigma^2 > 0$. The same model is considered in Sen and Srivastava (1975) and Hawkins (1977), though now the null hypothesis is of no mean shift in time. The first paper provides a Bayesian solution to the problem. Hawkins (1977), on the other hand, shows that for known σ^2 the likelihood ratio may be expressed as a function of the statistic

$$U = \max_{1 \leq \tau \leq n-1} \frac{|T_\tau|}{\sigma},$$

where

$$T_\tau^2 = \frac{\tau(n-\tau)}{n} (\bar{X}_\tau - \bar{X}'_\tau),$$

is the normalized between-groups sum of squares, $\bar{X}_\tau$ and $\bar{X}'_\tau$ being the sample means of the observations split at τ, respectively. The limiting null distribution of the corresponding test is proved to agree with the distribution of the maximum absolute value attained by a Gaussian process in discrete time having zero mean, unit variance and autocorrelation function given by expression (3.2) of Hawkins (1977). If σ^2 is unknown, Worsley (1979) provides the correct limiting distribution of the likelihood ratio statistic which now can be expressed as a function of the statistic

$$U = \max_{1 \leq \tau \leq n-1} (n-2)^{\frac{1}{2}} \frac{|T_\tau|}{S_\tau},$$

where S_τ^2 is the within-groups sum of squares of the observations split at τ. Tail probabilities are calculated using numerical techniques for samples sizes $n \leq 10$ and by Monte Carlo integration if $10 < n \leq 50$. An approximation to the asymptotic null distribution is provided using Bonferroni type inequalities. Hawkins (1992) generalizes these results to study eight procedures—which however do not include the likelihood ratio—for monitoring possible shifts in the mean vector or covariance matrix of an arbitrary multivariate distribution. Yao and Davis (1986) use results from the theory of Ornstein-Uhlenbeck process. In particular, they show how the likelihood ratio statistic may be suitably normalized so that it converges in distribution to the double exponential, or Gumbel, distribution. The asymptotic properties of this modified statistic are furthermore compared to those of a Bayesian test.

A comparison of various test statistics for detecting mean shifts in normal distributions, which also include the likelihood ratio, is given in James et al. (1987) and James et al. (1992) for, respectively, the univariate and multivariate case. The multivariate normal case is furthermore considered by Srivastava and Worsley (1986). These authors show that the likelihood ratio statistics agrees with the maximum of

Hotteling's statistic. They furthermore derive a conservative approximation for its null distribution using an improved Bonferroni inequality and show that the same statistic can be used to test for extra multinomial variation in a contingency table. Two interesting applications are presented which are the analysis of geological data previously studied by Chernoff (1973), and the study of the frequency patterns of pronouns in Shakespeare's dramas.

6.2 Changes in variance

Identifying a change in the variability of a distribution is of particular interest in the analysis of financial time series. Two examples of single or multiple change-point detection in the variance of a sequence of independent Gaussian random variables with known common mean are Hsu (1977) and Chen and Gupta (1997). The problem is however not treated using likelihood inference, which is the case for almost all contributions we are aware of.

6.3 Linear regression

Kim and Siegmund (1989) consider likelihood ratio testing for change-point detection in simple linear regression for the two situations where only the intercept is allowed to change and where both, the intercept and the slope, can change. Approximations for the tail probabilities are derived under reasonably general assumptions about the distribution of the independent variable. The extension to multiple linear regression is given in Kim (1994), while Kim and Cai (1993) study the robustness of the likelihood ratio pivot in simple linear regression. Andrews et al. (1996) determine a class of finite-sample optimal tests for the existence of a single or multiple changepoints at an unknown time in a normal linear multiple regression model with known variance. The power of several test statistics is compared using simulation. Luo et al. (1997) derive the asymptotic distribution of the likelihood ratio statistic for testing whether there is a lag in the effect of some covariates for right censored observations. Numerical examples illustrate how to implement the procedure. Most recently, Kelly (2015) considers three variants of the likelihood based statistics studied by Andrews and Ploberger (1994) for the general regression setting with time trend regressors. Critical values are obtained via simulation.

6.4 Lifetime data

Worsley (1983) derives the exact null and alternative distributions of the likelihood ratio for testing a change in a sequence of independent binomial random variables. The binomial case had already been considered by Hinkley and Hinkley (1970) and is further considered in Horváth (1989) who derive a number of limit theorems for the likelihood ratio. Its null distribution, in particular, is shown to converge to the Gumbel, or double exponential, distribution (see their Theorem 1). The study is further extended by Worsley (1986) to independent exponential family random variables, with particular emphasis on the exponential distribution. Loader (1992) test for the presence of a change point in non homogeneous Poisson processes. Large deviation techniques are used to approximate the significance level, and approximations for the power function are provided. A British coal mining accident data set is used to illustrate the methodology.

Borrowing from the theory of uniform quantile processes, Haccou et al. (1987) show that under the null hypothesis of no change in the rate parameter of an exponential distribution, the likelihood ratio statistic converges to an extreme value distribution. Worsley (1988) also considers survival data and in particular testing for a change in the hazard function. The likelihood ratio statistic is shown to be unbounded, but the exact null distribution of a suitably modified likelihood ratio test is provided. Modified likelihood ratio tests for the same setting are furthermore considered by Henderson (1990).

6.5 General distribution

Gombay and Horváth (1994) derive a likelihood ratio type statistic for testing whether there is a change in the parameter θ which indexes a general distribution $F(y;\theta)$. As such, the paper can be seen as the continuation of Page (1957). The authors prove that the distribution of the suitably centered and rescaled likelihood ratio statistic converges to a Gumbel distribution under the null hypothesis. Sadooghi-Alvandi et al. (2011) also consider change point detection in a general class of distributions. They derive the exact and asymptotic null distribution of the quasi Bayes and likelihood ratio tests. The techniques used stem from the theory of Brownian motion and bridge processes.

7 Discussion

This paper reviews the current literature on the asymptotic properties of the likelihood ratio statistic if the model is nonregular. Nonregularity can arise in many different ways, though all entail the failure of one, at times even two, regularity conditions. Many problems can be dealt with rather straightforwardly; other require rather sophisticated tools such as limit theorems and extreme value theory for stationary and non-stationary random fields. A wealth of contributions has been produced since the second half of the last century, though most of these are freestanding and scattered in time. We grouped them into three broad classes of problems—that is, boundary, indeterminate parameter and change-point problems—according to which conditions fail and the type of asymptotic arguments used.

Testing for a zero variance component in mixed effect models and constraint one-sided tests are two common examples of a boundary problem, which is the best studied nonregular case. The limiting distribution of the likelihood ratio can generally be traced back to a chi bar squared distribution with a suitable number of components and mixing weights which depend on the number of parameters on the boundary and on the design matrices in regression problems. This is also the only type of problem for which higher order asymptotic results are available.

The class of indeterminate parameter problems is far more heterogenous. Apart from finite mixtures, which we treated in a separate section, the remaining cases can be put under the two umbrellas of non-identifiable parameters and singular information matrix. The methodological difficulties increase as the limiting distributions depend on the parametric family and on the unknown parameters. If θ is scalar and we want to test homogeneity against a two-component mixture, the likelihood ratio converges to the supremum of a Gaussian process. For a larger number of mixture components and/or multidimensional θ, this becomes the supremum of a Gaussian random field. In these cases, simulation-based approaches are often preferred to obtain the required tail probabilities. An additional problem is that constraints need be imposed to guarantee identifiability of the mixture parameters. As outlined by Garel (2007), these may affect the parameter space, by bounding it or imposing suitable separation conditions among the parameters, or the alternative hypotheses which need be contiguous. We may furthermore constrain the mixing probability π so that $\min(\pi, 1-\pi) \geq \epsilon$ for some positive ϵ, or we may penalize the likelihood function, for instance by adding a term which is highly negative when π is close to 0 or 1, such as $\log\{\pi(1-\pi)\}$, to avoid that the maximum occurs around $\pi = 0$ or $\pi = 1$.

Change-point problems range from the simple situation of detecting an alteration in the regime of a random sequence to identifying a structural break in the hazard function of a lifetime distribution. Note that while in the second case the change-point can assume any positive value, in the first situation it is bound to vary in a discrete set. Furthermore, there may be situations where the likelihood ratio statistic for the unknown change-point is unbounded. Limit theorems for processes based on U-statistics (Csrgö and Horváth, 1997, Section 2.4) and extreme value theory for random processes play a central role.

At several occasions, and especially for the most intricate situations, the authors suggest to use resampling based techniques, such as parametric and nonparametric bootstrapping, to explore the finite-sample properties of likelihood based statistics; see Gombay and Horváth (1999) and Cheng (2017). Indeed, simulation may nowadays be used to establish the desired empirical distributions of the estimators and to compute approximations for the p-values obtained from the corresponding Wald type statistics. Methodological difficulties and prohibitive computational costs limit, however, this possibility to specific applications.

The review has focused almost exclusively on frequentist hypothesis testing using the likelihood ratio statistic. Maximum likelihood estimation for a class of nonregular cases, which include the three-parameter Weibull, the gamma, log-gamma and beta distributions, is considered in Smith (1985). A significant literature has grown since then, parts of which culminated in the book-length account of techniques for parameter estimation in non-standard settings by Cheng (2017). Most of the difficulties encountered in nonregular settings vanish if the model is analysed using Bayes' rule, though one has always to be cautious. See for instance the work by Ghosal and Samanta (1995); Bunke and Milhaud (1998); Ghosal (1999). An early Bayesian contribution to change-point detection is Gardner (1969); see Lai and Xing (2011) for a most recent one. Bayesian analysis of finite mixtures is treated in Richardson and Green (1997) and Nobile and Fearnside (2007). Ning, Pailden and Gupta (2012) proposes a nonparametric solutions to change-point detections in a random multivariate sequence. Further Bayesian and nonparametric contributions were mentioned in passing throughout the paper with suitable links to their frequentist counterparts.

The many contributions which have arisen since Smith's (1989) review proves that the interest in this type of models has not faded since they made their entrance back in the early '50s.

A Appendix

A.1 Edgeworth and saddlepoint expansions

Edgeworth and saddlepoint expansions play a central role in the derivation of higher order asymptotic results. Saddlepoint approximations were brought to statistics in the seminal paper by Daniels (1954). Edgeworth series date back longer in time (Cramér, 1937; Cornish and Fisher, 1937).

Let $Y_1, \ldots, Y_n$ be a random sample of a continuous one-dimensional random variable Y with cumulant generating function $\kappa_Y(t)$. Let ρ_m denote the mth standardized cumulant of Y and $S_n^* = (S_n - n\mu)/(\sqrt{n}\sigma)$ be the standardized version of $S_n = \sum_{i=1}^n Y_i$, where, as usual, $\mu = \mathrm{E}(Y) = \kappa'(0)$ and $\sigma^2 = \mathrm{Var}(Y) = \kappa''(0)$ denote the first and second cumulants of Y. The Edgeworth expansion for the density of S_n^* is

$$\begin{aligned} f_{S_n^*}(s) &= \phi(s)\left\{1 + \frac{\rho_3}{6n^{1/2}}H_3(s) + \right. \\ &+ \left. \frac{\rho_4}{24n}H_4(s) + \frac{\rho_3^2}{72n}H_5(s)\right\} + O(n^{-3/2}). \end{aligned} \tag{16}$$

Here, $\phi(s)$ is the standard normal density, while $H_m(\cdot)$ denotes the mth Hermite polynomial, defined by

$$(-1)^m d^m \phi_(s)/ds^m = H_m(s)\phi(s).$$

Integration of (16) yields the corresponding Edgeworth expansion of the distribution function of S_n^* (Barndorff-Nielsen and Cox, 1994, §4.2).

In general, Edgeworth expansions provide good approximations in the centre of the density, but can exhibit a series of practical disadvantages especially in the tails of the distribution. Saddlepoint approximations are far more accurate especially for small sample sizes. They require, however, the entire cumulant generating function to be known, while only the first few cumulants are necessary for Edgeworth approximations.

Saddlepoint approximations are derived by embedding the density of interest in a suitable exponential family, which is then replaced by its Edgeworth series but so as to guarantee that the tilted density has expected value equal to the observed one. In case of $S_n = \sum_{i=1}^n Y_i$, the saddlepoint expansion takes the form

$$f_{S_n}(s) = \frac{1}{\sqrt{2\pi n\kappa''(\tilde{t})}} e^{n\kappa_Y(\tilde{t}) - \tilde{t}s} \; \{1 + O(n^{-1})\}, \tag{17}$$

where $\tilde{t}$ is the so-called saddlepoint which guarantees that $n\kappa'(\tilde{t}) = s$. Lugannani and Rice (1980) integrated (17) so as to obtain the corresponding approximation of the distribution function of S_n.

Saddlepoint approximation can be obtained for any statistics which admit a cumulant generating function; see the review papers by Reid (1988, 1996, 2003) and the book length accounts by Jensen (1995) and Butler (2007). Davison and Wang (2002) discuss the interpretation of saddlepoint approximations for discrete data.

A.2 Asymptotic expansion of $(\hat{\theta} - \theta)$

Let $p = 1$ and $l(\theta)$ be the log-likelihood function for θ in a regular parametric model. Write $l_m = l_m(\theta) = d^m l(\theta)/d\theta^m$ for the derivative of order $m = 2, 3, \ldots$, of $l(\theta)$, while $u = u(\theta) = dl(\theta)/d\theta$ represents the score function. We start by expanding the likelihood equation around θ to give

$$0 = u(\hat{\theta}) = u + (\hat{\theta} - \theta)l_2 + \frac{1}{2}(\hat{\theta} - \theta)^2 l_3 + \frac{1}{6}(\hat{\theta} - \theta)^3 l_4 + \cdots ,$$

where $\hat{\theta}$ indicates the maximum likelihood estimate. Reordered, this expression gives an asymptotic expansion for $(\hat{\theta} - \theta)$ of the form

$$\hat{\theta} - \theta = j^{-1}u + \frac{1}{2}j^{-1}(\hat{\theta} - \theta)^2 l_3 + \frac{1}{6}j^{-1}(\hat{\theta} - \theta)^4 l_4 + \cdots , \qquad (18)$$

where j^{-1} is the inverse of the observed information $j = -l_2$. Next, iteratively substitute in the right-hand part of (18) $\hat{\theta} - \theta$ with its expansion and rearrange terms; this leads to

$$\hat{\theta} - \theta = j^{-1}u + \frac{1}{2}j^{-3}u^2 l_3 + \frac{1}{6}j^{-4}(l_4 + 3j^{-1}l_3^2)u^3 + \cdots \qquad (19)$$

To reorder the different terms in (19) according to their asymptotic order, we need to introduce the general notation

$$H_m = l_m - \nu_m, \quad \text{with} \quad \nu_m = E[l_m(\theta; Y)] \quad \text{for } m \geq 2. \qquad (20)$$

Note that while the score function $u(\theta)$ and H_m are of order $n^{1/2}$ under ordinary repeated sampling, ν_m is of order n. We further write $j = i\{1 - i^{-1}(i - j)\}$ and expand j^{-1} as

$$j^{-1} = i^{-1} + i^{-2}(i - j) + i^{-3}(i - j)^2 + \cdots , \qquad (21)$$

where $i = E[j(\theta; Y)]$ is the expected information. Now, inserting (21) into (19) and using notation (20), we may rewrite the asymptotic expansion of $(\hat{\theta} - \theta)$ to obtain

$$\hat{\theta} - \theta = i^{-1}u \overset{\bullet}{+} i^{-2}H_2 u + \frac{1}{2}i^{-3}u^2 \nu_3 \overset{\bullet}{+} O_p(n^{-3/2}), \qquad (22)$$

where a $\bullet$ is used used to mark a drop of $n^{-1/2}$. See Pace and Salvan (1997, Chapter 9) and Barndorff-Nielsen and Cox (1994, Chapter 5) for a detailed treatment.

Acknowledgements

This research was supported by University of Padova grant no. CPDA101912 *Large- and small-sample inference under non-standard conditions* (Progetto di Ricerca di Ateneo 2010).

References

Aitchison, J. and Silvey, S. D. (1960). Maximum likelihood procedures and associated tests of significance. *Journal of the Royal Statistical Society, Series B (Methodological)*, **22**, 154-171.

Andrews, D. W. K. (2001). Testing when a parameter is on the boundary of the maintained hypothesis. *Econometrica***69** 683–734.

Andrews, D. W. K. and Ploberger W. (1994). Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica*, **62**, 1383–1414.

Andrews, D. W. K., Lee, I. and Ploberger W. (1996). Optimal changepoint tests for normal linear regression. *Journal of Econometrics*, **70**, 9–38.

Azaïs, J.-M., Gassiat, É. and Mercadier, C. (2006). Asymptotic distribution and local power of the log-likelihood ratio test for mixtures: bounded and unbounded cases. *Bernoulli*, **12**, 775–799.

Azaïs, J.-M., Gassiat, É. and Mercadier, C. (2009). The likelihood ratio test for general mixture models with or without a structural parameter. *ESAIM: Probability and Statistics*, **13**, 301–327.

Azzalini, A. (1996). *Statistical Inference Based on the likelihood.* Chapman & Hall, London.

Barnabani, M. (2002). Wald-based approach with singular information matrix. *Workind Paper*, Department of Statistics "P. Fortunati", University of Florence.

Barndorff-Nielsen, O. E. (1983). On a formula for the distribution of the maximum likelihood estimator. *Biometrika*, **70**, 343–365.

Barndorff-Nielsen, O. E (1986). Inference on full or partial parameters based on the standardized signed log likelihood ratio. *Biometrika*, **73**, 307–322.

Barndorff-Nielsen, O. E. and Cox, D. R. (1994). *Inference and Asymptotics.* Chapman & Hall, London.

Bartlett, M. S. (1937). Properties of sufficiency and statistical tests. *Proceedings of the Royal Society London. Series A*, **160**, 268–282.

Bickel, P. and Chernoff, H. (1993). Asymptotic distribution of the likelihood ratio statistic in a prototypical non regular problem. In *Statistics and Probability: A Raghu Raj Bahadur Festschrift* (edited by J. K. Ghosh, S. K. Mitra, K. R. Parthasarathy and B. L. S. Prakasa), pp. 83–96. Wiley Eastern, New Dehli.

Böhning, D., Dietz, E., Schaub, R., Schlattmann, P. and Lindsay, B. G. (1994). The distribution of the likelihood ratio for mixture of densities from the one-parameter exponential family. *Annals of the Institute of Statistical Mathematics*, **46**, 373–388.

Brazzale, A. R. and Davison, A. C. (2008). Accurate parametric inference for small sample sizes. *Statistical Science*, **23**, 465–484.

Brazzale, A. R., Davison, A. C. and Reid, N. (2007). *Applied Asymptotics: Case Studies in Small Sample Statistics.*. Cambridge University Press, Cambridge.

Bunke, O. and Milhaud, X. (1998). Asymptotic behavior of Bayes estimates under possibly incorrect models. *The Annals of Statistics*, **26**, 617–644.

Butler, R. W. (2007). *Saddlepoint Approximations with Application.* Cambridge University Press, ??.

Chant, D. (1974). On asymptotic tests of composite hypotheses in non standard conditions. *Biometrika*, **61**, 291–298.

Chen, H. and Chen, J. (2001a). Large sample distribution of the likelihood ratio test for normal mixtures. *Statistics & Probability Letters*, **52**, 125–133.

Chen, H. and Chen, J. (2001b). The likelihood ratio test for homogeneity in finite mixture models. *The Canadian Journal of Statistics*, **29**, 201–215.

Chen, H. and Chen, J. (2003). Tests for homogeneity in normal mixtures in the presence of a structural parameter. *Statistica Sinica*, **13**, 351–365.

Chen, H., Chen, J. and Kalbfleisch, J. D. (2001). A modified likelihood ratio test for homogeneity in finite mixture models. *Journal of the Royal Statistical Society, Series B (Methodological)*, **63**, 19–29.

Chen, J. and Gupta, A. K. (1997). Testing and locating variance change points with application to stock prices. *Journal of American Statistical Association*, **92**, 739–747.

Chen, J. and Gupta, A. K. (2012). *Parametric Statistical Change Point Analysis with Application to Genetics, Medicine and Finance* (2nd ed.). Birkhäuser, Boston.

Chen, J. and Kalbfleisch, J. D. (2005). Modified likelihood ratio test in finite mixture models with a structural parameter. *Journal of Statistical Planning and Inference*, **129**, 93–107.

Chen, J., Li, P. and Fu, Y. (2008). Testing homogeneity in a mixture of von Mises distributions with a structural parameter. *The Canadian Journal of Statistics*, **36**, 129–142.

Cheng, R. (2017). *Non-Standard Parametric Statistical Inference.* Oxford University Press, New York.

CHERNOFF, H. (1954). On the distribution of the likelihood ratio. *The Annals of Mathematical Statistics*, **54**, 573–578.

CHERNOFF, H. (1973). The use of faces to represent points in k-dimensional space graphically. *Journal of the American Statistical Association*, **68**, 361–368.

CHERNOFF, H. and LANDER, E. (1995). Asymptotic distribution of the likelihood ratio test that a mixture of two binomials is a single binomial. *Journal of Statistical Planning and Inference*, **43**, 19–40.

CIUPERCA, G. (2002). Likelihood ratio statistic for exponential mixtures. *Annals of the Institute of Statistical Mathematics*, **54**, 585–594.

CLAESKENS, G., NGUTI, R. and JANSSEN, P. One-sided tests in shared frailty models. *Test*, **17**, 69–82.

CORNISH, E. A. and FISHER, R. A. (1937). Moments and cumulants in the specification of distributions. *International Statistical Review*, **5**, 307–322.

COX, D. R. (2006). *Principles of Statistical Inference.* Cambridge University Press, New York.

COX, D. R. and HINKLEY, D. V. (1974). *Theoretical Statistics.* Chapman & Hall, London.

CRAINICEANU, C. M. and RUPPERT, D. (2004). Likelihood ratio tests in linear mixed models with one variance component. *Journal of the Royal Statistical Society, Series B (Methodological)*, **66**, 165–185.

CRAMÉR, H. (1937). *Random Variables and Probability Distributions.* Cambridge University Press, Cambridge.

CRAMÉR, H. (1947). *Mathematical Methods of Statistics.* Princeton University Press, Princeton.

CSÖRGÖ, M. and HORVÁTH, L. (1997). *Limit Theorems in Change-point Analysis.* Wiley, New York.

DANIELS, H. E. (1954). Saddlepoint approximation in statistics. *The Annals of Mathematical Statistics*, **25**, 631–650.

DASGUPTA, A. (2008). *Asymptotic Theory of Statistics and Probability.* Springer-Verlag, New York.

DAVIES, R. B. (1977). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika*, **64**, 247–254.

DAVIES, R. B. (1987). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika*, **64**, 33–43.

DAVISON, A. C. (2003). *Statistical Models.* Cambridge University Press, Cambridge.

Davison, A. C. and Wang, S. (2002). Saddlepoint approximations as smoothers. *Biometrika*, **89**, 933–938.

del Castillo, J. and Lopez-Ratera, A. (2006). Saddlepoint approximation in exponential models with boundary points. *Biometrika*, **12**, 491–500.

Easton, G. S. and Ronchetti, E. M. (1986). General saddlepoint approximations, with applications to L statistics. *Journal of the American Statistical Association*, **81**, 420–430.

El-Helbawy, A. T. and Hassan, T. (1994). On the Wald, Lagrangian multiplier and likelihood ratio tests when the information matrix is singular. *Journal of the Italian Statistical Society*, **1**, 51–60.

Feder, P. I. (1968). On the distribution of the log likelihood ratio test statistic when the true parameter is near the boundaries of the hypothesis regions. *The Annals of Mathematical Statistics*, **39**, 2044–2055.

Fortunati, S., Gini, F., Greco, M., Farina, A., Graziano, A. and Giompapa, S. (2012). An identifiability criterion in the presence of random nuisance parameters. *Proceedings of the 20th European Signal Processing Conference* (EUSIPCO 2012). Bucharest, August 27–31, 2012.

Fraser, D. A. S. (1988). Normed likelihood as saddlepoint approximation. *Journal of Multivariate Analysis*, **27**, 181–193.

Fraser, D. A. S., Reid, N. and Wu, J. (1999). A simple general formula for tail probabilities for Frequentist and Bayesian inference. *Biometrika*, **86**, 249–264.

Gardner, L. A. (1969). On detecting change in the mean of normal variates. *Annals of Mathematical Statistics*, **40**, 116–126.

Garel, B. (2001). Likelihood ratio test for univariate Gaussian mixture. *Journal of Statistical Planning and Inference*, **96**, 325–350.

Garel, B. (2007). Recent asymptotic results in testing for mixtures. *Computational Statistics & Data analysis*, **51**, 5295–5304.

Gassiat, É. (2002). Likelihood ratio inequalities with applications to various mixtures. *Annales de l'Institut Henri Poincaré Probabilités et Statistiques*, **6**, 897–906.

Ghosal, S. (1999). Matching priors for non-regular cases. *Biometrika*, **86** 956–964.

Ghosal, S. and Samanta, T. (1995). Asymptotic behaviour of Bayes estimates and posterior distributions in multiparameter nonregular cases. *Mathematical Methods of Statistics*, **4**, 361-388.

Ghosh, K. J. and Sen, K. P. (1985). On the asymptotic performance of the log likelihood ratio statistic for the mixture model and related results. In *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer* (edited by L. LeCam, R. A. Olshen and C.-S. Cheng), Vol. II, pp. 789–806. Wadsworth Advanced Books & Software, Monterey.

GODAMBE, V. P. (1991). *Estimating Functions.* Oxford University Press, Oxford.

GOFFINET, B., LOISEL, P. and LAURENT, B. (1992). Testing in normal mixture models when the proportions are known. *Biometrika*, **79**, 842–846.

GOMBAY, E. and HORVÁTH, L. (1994). An application of the maximum likelihood test to the change-point problem. *Stochastic Processes and their Applications*, **50**, 161–171.

GOMBAY, E. and HORVÁTH, L. (1999). Change-points and bootstrap. *Environmetrics*, **10**, 725–736.

HACCOU, P., MEELIS, E. and VAN DE GEER, S. (1987). The likelihood ratio test for the change point problem for exponentially distributed random variables. *Stochastic Processes and their Applications*, **27**, 121–139.

HARTIGAN, J. A. (1977). Distribution problems in clustering. In *Classification and Clustering* (edited by J. van Ryzin), pp. 45-72. Academic Press, New York.

HARTIGAN, J. A. (1985). A failure of likelihood asymptotics for normal mixtures. In *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer* (edited by L. LeCam, R. A. Olshen and C.-S. Cheng), Vol. II, pp. 807–810. Wadsworth Advanced Books & Software, Monterey.

HATHAWAY, R. J. (1985). A constrained formulation of maximum-likelihood estimation for normal mixture distributions. *Annals of Statistics*, **13**, 795–800.

HAWKINS, D. M. (1977). Testing a sequence of observations for a shift in location. *Journal of the American Statistical Association*, **72**, 180–186.

HAWKINS, D. M. (1992). Detecting shifts in functions of multivariate location and covariance parameters. *Journal of Statistical Planning and Inference*, **33**, 233–244.

HENDERSON, R. (1990). A problem with the likelihood ratio test for a change-point hazard rate model. *Biometrika*, **77**, 835–843.

HINKLEY, D. V. (1970). Inference about the change-point in a sequence of random variables. *Biometrika*, **57**, 1–17.

HINKLEY, D. V. and HINKLEY, E. A. (1970). Inference about the change-point in a sequence of binomial random variables. *Biometrika*, **57**, 477–488.

HIRANO, K. and PORTER, J. R. (2003). Asymptotic efficiency in parametric structural models with parameter-dependent support. *Econometrica*, **71**, 1307–1338.

HORVÁTH, L. (1989). The limit distributions of the likelihood ratio and cumulative sum tests for a change in binomial probability. *Journal of Multivariate Analysis*, **31**, 148–159.

HSU, D. A. (1977). Tests for variance shifts at an unknown time point. *Applied Statistics*, **26**, 179–184.

HUBER, P. J and RONCHETTI, E. M. (2009). *Robust Statistics* (2nd ed.). John Wiley & Sons, Hoboken, New Jersey.

JEFFRIES, N. O. (2003). A note on "Testing the number of components in a normal mixture", *Biometrika*, **90**, 991–994.

JENSEN, J. L. (1995). *Saddlepoint Approximations.* Oxford University Press, Oxford.

JIN, F. and LEE, L.-F. (2017). Irregular N2SLS and LASSO estimation of the matrix exponential spatial specification model. *Journal of Econometrics.* URL: `https://doi.org/10.1016/j.jeconom.2018.06.005`.

JIN, F. and LEE, L.-F. (2018). Lasso maximum likelihood estimation of parametric models with singular information matrices. *Econometrics*, **6**, 8.

KASAHARA, H. and SHIMOTSU, K. (2015). Testing the number of components in finite mixture models. *Global COE Hi-Stat Discussion Paper Series*, **gd12-259**, Institute of Economic Research, Hitotsubashi University.

KELLY, G. E. (2015). Approximations to the p-values of tests for a change-point under non-standard conditions. *Journal of Statistical Computation and Simulation*, **86**, 1430–1449.

KHODADADI, A. and ASGHARIAN, M. (2008). Change-point problem and regression: an annotated bibliography. *COBRA Preprint Series*, Working Paper 44. URL: `https://biostats.bepress.com/cobra/art44`.

KIM, H.-J. Tests for a change-point in linear regression. *IMS Lecture Notes - Monograph Series* (edited by E. Carlstein, H.-G. Möller and D. Siegmud), **23**, pp. 170–176. IMS, Hayward.

KIM, H. J. and CAI, L. (1993). Robustness of the likelihood ratio test for a change in simple linear regression. *Journal of the American Statistical Association*, **88**, 864–871.

KIM, H. J. and SIEGMUND, D. (1989). The likelihood ratio test for a change-point in simple linear regression. *Biometrika*, **76**, 409–423.

KOENKER, R., CHERNOZHUKOV, V., HE, X. and PENG, L. (2017). *Handbook of Quantile Regression.* Chapman and Hall/CRC, Boca Raton, FL.

KOPYLEV, L. (2012). Constrained parameters in applications: Review of issues and approaches. *International Scholarly Research Network ISRN Biomathematics.* DOI:10.5402/2012/872956

KOPYLEV, L. and SINHA, B. (2011). On the asymptotic distribution of likelihood ratio test when parameters lie on the boundary. *Sankhyā B*, **73**, 20–41.

KUDÔ, A. (1963). A multivariate analogue of the one-sided test. *Biometrika*, **50**, 404–418.

James, B., James, K. L. and Siegmund, D. (1987). Tests for a change-point. *Biometrika*, **74**, 71–83.

James, B., James, K. L. and Siegmund, D. (1992). Asymptotic approximation for likelihood ratio test and confidence regions for a change point in mean of a multivariate normal distribution. *Statistica Sinica*, **2**, 69–90.

Lai, T. L. and Xing, H. (2011). A simple Bayesian approach to multiple change-points. *Statistica Sinica*, **21**, 539–569.

Lancaster, T. (2000). The incidental parameter problem since 1948. *Journal of Econometrics*, **95**, 391–413.

Lee, T.-S. (2010). Change-point problems: bibliography and review. *Journal of Statistical Theory and Practice*, **4**, 643–662.

Lei, Q. L. and Qin, Y. S. (2009). A modified likelihood ratio test for homogeneity in bivariate normal mixtures of two samples. *Journal of Systems Science and Complexity*, **22**, 460–468.

Lemdani, M. and Pons, O. (1997). Likelihood ratio tests for genetic linkage. *Statistics and Probability Letters*, **33**, 15–22.

Lemdani, M. and Pons, O. (1999). Likelihood ratio tests in contamination models. *Bernoulli*, **5**, 705–719.

Lindsay, B. G. (1995). *Mixture Models: Theory, Geometry and Applications.* Institute of Mathematical Statistics, Hayward.

Liu, X., Pasarica, C. and Shao, Y. (2003). Testing homogeneity in gamma mixture models. *Scandinavian Journal of Statistics*, **30**, 227–239.

Liu, X. and Shao, Y. (2003). Asymptotics for likelihood ratio tests under loss of identifiability. *The Annals of Statistics*, **31**, 807–832.

Liu, X. and Shao, Y. (2004). Asymptotics for likelihood ratio test in a two-component normal mixture model. *Journal of Statistical Planning and Inference*, **123**, 61–81.

Lo, Y. (2008). A likelihood ratio test of a homoscedastic normal mixture against a heteroscedastic normal mixture. *Statistics and Computing*, **18**, 233–240.

Lo, Y., Mendell, N. R. and Rubin, D. B. (2001). Testing the number of components in a normal mixture. *Biometrika*, **88**, 767–778.

Loader, C. R. (1992). A log-linear mModel for a Poisson process change point. *The Annals of Statistics*, **20**, 1391–1411.

Lugannani, R. and Rice, S. O. (1980). Saddlepoint approximation for the distribution of the sum of independent random variables. *Advances in Applied Probability*, **12**, 475–490.

LUO, X., TURNBULL, B. W. and CLARK, L. C. (1997). Likelihood ratio tests for a changepoint with wurvival data. *Biometrika*, **84**, 555-565.

MALLER, R.A. and ZHOU, X. (2003). Testing for individual heterogeneity in parametric models for event-history data. *Mathematical Methods of Statistics*, **12**, 276–304.

MENDELL, N. R., THODE, H. C. JR. and FINCH, S. J. (1991). The likelihood ratio test for the two-component normal mixture problem. *Biometrics*, **47**, 1143–1148.

MOLENBERGHS, G. and VERBEKE, G. (2007). Likelihood ratio, score and Wald tests in a constrained parameter set. *The American statistician*, **61**, 22–27.

MORAN, P. A. P. (1971). Maximum likelihood estimation in non standard conditions. *Mathematical Proceedings of the Cambridge Philosophical Society*, **70**, 441–50.

NEYMAN, J. and SCOTT, E. L. (1948). Consistent estimates based on partially consistent observations. *Econometrica*, **16**, 11–32.

NING, W., PAILDEN, J. and GUPTA, A. (2012). Empirical likelihood ratio test for the epidemic change model. *Journal of Data Science*, **10**, 107–127.

NOBILE, A. and FEARNSIDE, A. T. (2007). Bayesian finite mixtures with an unknown number of components: The allocation sampler. *Statistics & Computing*, **17**, 147–162.

OLIVEIRA-BROCHADO, A. and MARTINS, N. (2005). Assessing the number of components in mixture models: a review. *FEP Working Papers*, **194**, Universidade do Porto, Faculdade de Economia do Porto.

PACE, L. and SALVAN, A. (1997). *Principles of Statistical Inference: from a Neo-Fisherian Perspective.* World Scientific Publishing, Singapore.

PAGE, E. S. (1955). A test for a chance in a parameter occurring at an unknown point. *Biometrika*, **42**, 523–527.

PAGE, E. S. (1957). On problems in which a change in a parameter occurs at an unknown point. *Biometrika*, **44**, 248–252.

PAULINO, C. D. M. and PEREIRA, C. A. B. (1994). On identifiability of parametric statistical models. *Journal of the Italian Statistical Society*, **1**, 125–151.

POLYMENIS, A. and TITTERINGTON, D. M. (1999). A note on the distribution of the likelihood ratio statistic for normal mixture with known proportions. *Journal of Statistical Computation and Simulation*, **64**, 167–175.

PRAKASA RAO, B. L. S. (1992). *Identifiability in Stochastic Models: Characterization of Probability Distributions.* Academic Press, London.

QIN, Y. S. and SMITH, B. (2004). Likelihood ratio test for homogeneity in normal mixtures in the presence of a structural parameter. *Statistica Sinica*, **14**, 1165–1177.

QIN, Y. S. and SMITH, B. (2006). The likelihood ratio test for homogeneity in bivariate normal mixtures. *Journal of Multivariate Analysis*, **97**, 474–491.

REID, N. (1988). Saddlepoint methods and statistical inference (with Discussion). *Statistical Science*, **3**, 213–238.

REID, N. (1996). Likelihood and higher-order approximations to tail areas: a review and annotated bibliography. *Canadian Journal of Statistics*, **24**, 141–166.

REID, N. (2003). Asymptotics and the theory of inference. *The Annals of Statistics*, **31**, 1695–1731.

RICHARDSON, S. and GREEN, P. J. (1997). On Bayesian analysis of mixtures with an unknown number of components. *Journal of the Royal Statistical Society, Series B (Methodological)*, **59**, 731–792.

RITZ, C. and SKOVGAARD, IB M. (2005). Likelihood ratio tests in curved exponential families with nuisance parameters present only under the alternative. *Biometrika*, **92**, 507–517.

ROBERTSON, T., WRIGHT, F. T. and DYKSTRA, R. L. (1988). *Order Restricted Inference*. John Wiley & Sons, New York.

ROTHENBERG, T. J. (1971). Identification in parametric models. *Econometrica*, **39**, 577–591.

ROTNITZKY, A., COX, D. R., BOTTAI, M. and ROBINS, J. (2000). Likelihood-based inference with singular information matrix. *Bernoulli*, **6**, 243–284.

SADOOGHI-ALVANDI, S. M., NEMATOLLAHI, A. R. and HABIBI, R. (2011). Test procedures for change point in a general class of distributions, *Journal of Data Science*, **9**, 111–126.

SELF, S. G. and LIANG, K. (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under non standard conditions. *Journal of the American Statistical Association*, **82**, 605–610.

SEN, P. K. and SILVAPULLE, M. J. (2002). An appraisal of some aspects of statistical inference under inequality constraints. *Journal of Statistical Planning and Inference*, **107**, 3–43.

SEN, A. and SRIVASTAVA, M. S. (1975). Some one-sided tests for change in level. *Technometrics*, **17**, 61–64.

SEVERINI, T. (2000). *Likelihood Methods in Statistics*. Oxford University Press, Oxford.

Severini, T. (2004). A modified likelihood ratio statistic for some nonregular models. *Biometrika*, **91**, 603–612.

Shapiro, A. (1985). Asymptotic distribution of test statistics in the analysis of moment structures under inequality constraints. *Biometrika*, **72**, 133–144.

Shapiro, A. (1988). Towards a unified theory of inequality constrained testing in multivariate analysis. *International Statistical Review*, **56**, 49–62.

Silvapulle, M. J. and Sen, P. K. (2005). *Constrained Statistical Inference.* John Wiley & Sons, New York.

Silvey, S. D. (1959). The Lagragian multiplier test. *Ann. Math. Statist.*, **30**, 382–407.

Sinha, B., Kopylev, L. and Fox, J. (2012). Some new aspects of dose-response multi-stage models with applications. *Pakistan Journal of Statistics and Operation Research*, **8**, 441–478.

Smith, R. L. (1985). Maximum likelihood estimation in a class of non regular cases. *Biometrika*, **72**, 67–92.

Smith, R.L. (1989). A survey of non-regular problems. *Proceedings of the 47th Session of the International Statistical Institute*, Paris, August 29 – September 6, 1989, pp. 353–372.

Sørensen, H. (2008). Small sample distribution of the likelihood ratio effects in the random effects model. *Journal of Statistical Planning and Inference*, **138**, 1605–1614.

Song, R., Kosorok, M. R. and Fine, J. P. (2009). On asymptotically optimal tests under loss of identifiability in semiparametrics models. *The Annals of Statistics*, **37**, 2409–2444.

Srivastava, M.S. and Worsley, K.J. (1986). Likelihood ratio tests for a change in the multivariate mean. *Journal of the American Statistical Association*, **81**, 199–204.

Stram, D. O. and Lee, J. W. (1994). Variance components testing in the longitudinal mixed rffects model. *Biometrics*, **50**, 1171–1177.

Sun, H.-J. (1988). A FORTRAN subroutine for computing normal orthant probabilities of dimensions up to nine. *Communication in Statistics – Computation and Simulation*, **17**, 1097–1111.

Thode, H. C. Jr., Finch, S. J. and Mendell, N. R. (1988). Simulated percentage points for the null distribution of the likelihood ratio test for a mixture of two normals. *Biometrics*, **44**, 1195–1201.

Vu, H. T. V. and Zhou, S. (1997). Generalization of likelihood ratio tests under nonstandard conditions. *Ann. Statist.*, **25**, 897–916.

Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica*, **57**, 307–333.

Yao, Y.-C. and Davis, R.A. (1986). The asymptotic behavior of the likelihood ratio statistic for testing a shift in mean in a sequence of independent normal variates. *Sankhyiā A*, **48**, 339–353.

Wilks, S. S. (1938). The large sample distribution of the likelihood ratio for testing composite hypotheses. *The Annals of Mathematical Statistics*, **1**, 60–62.

Wolak, F. A. (1987). An exact test for multiple inequality and equality constraints in the linear regression model. *Journal of the American Statistical Association*, **82**, 782–793.

Worsley, K. J. (1979). On the likelihood ratio test for a shift in location of normal populations. *Journal of the American Statistical Association*, **74**, 365–367.

Worsley, K. J. (1983). The power of likelihood ratio and cumulative sum tests for a change in a binomial probability. *Biometrika*, **70**, 455–464.

Worsley, K. J. (1986). Confidence regions and test for a change point in a sequence of exponential family random variables. *Biometrika*, **73**, 91–104.

Worsley, K. J. (1988). Exact percentage points of the likelihood-ratio test for a change-point hazard-rate model. *Biometrics*, **44**, 259–263.

Zhang, D. and Lin, X. (2008). Variance component testing in generalized linear mixed models for longitudinal/clustered data and other related topics. In *Random Effect and Latent Variable Model Selection* (D. B. Dunson Editor), p. 19-36, Lecture Notes in Statistics, Springer-Veralg, New York.

Working Paper Series
Department of Statistical Sciences, University of Padua

You may order paper copies of the working papers by emailing wp@stat.unipd.it

Most of the working papers can also be found at the following url: http://wp.stat.unipd.it