An estimation method for the chi-square divergence with application to test of hypotheses

Michel Broniatowski, Samantha Leorato

AN ESTIMATION METHOD FOR THE CHI-SQUARE DIVERGENCE WITH APPLICATION TO TEST OF HYPOTHESES

M. BRONIATOWSKI[1] AND S. LEORATO[2]

Abstract. We propose a new definition of the chi-square divergence between distributions. Based on convexity properties and duality, this version of the χ^2 is well suited both for the classical applications of the χ^2 for the analysis of contingency tables and for the statistical tests for parametric models, for which it has been advocated to be robust against *inliers*.

We present two applications in testing. In the first one we deal with tests for finite and infinite numbers of linear constraints, while, in the second one, we apply χ^2-methodology for parametric testing against contamination.

1. Introduction

The χ^2 distance is commonly used for categorized data. For the continuous case, optimal grouping pertaining to the χ^2 criterion have been proposed by various authors; see f.i. [5], [18], [12]. These methods are mainly applied for tests, since they may lead to some bias effect for estimation.

This paper introduces a new approach to the χ^2, inserting its study inside the range of divergence-based methods and presenting a technique avoiding grouping for estimation and test. Let us first introduce some notation.

Let M_1 denote the set of all probability measures on $\mathbb{R}^d$ and M the set of all signed measures on $\mathbb{R}^d$ with total mass 1. For $P \in M_1$ and $Q \in M$, introduce the χ^2 distance between P and Q by

$$\chi^2(Q,P) = \begin{cases} \int \left(\frac{dQ-dP}{dP}\right)^2 dP & Q \text{ is a.c. w.r.t. } P \\ \infty & \text{otherwise.} \end{cases} \tag{1.1}$$

For Ω a subset of M denote

$$\chi^2(\Omega,P) = \inf_{Q\in\Omega} \chi^2(Q,P), \tag{1.2}$$

with $\inf_{\{\emptyset\}} = \infty$.

When the infimum in (1.2) is reached at some measure Q^* which belongs to Ω, then Q^* is the *projection* of P to Ω. Also the role of the class of measures M will appear later, in connection with the possibility to obtain easily Q^* through usual

1991 *Mathematics Subject Classification.* 62F03, 62F10,62F30.

Key words and phrases. chi-square divergence, hypothesis testing, linear constraints, marginal distributions, contamination models, Fenchel-Legendre transform, inliers.

optimization methods, which might be quite difficult when we consider subsets Ω in M_1.

For a problem of test such as $H_0 : P \in \Omega$ vs $H_1 : P \notin \Omega$, the test statistic will be an estimate of $\chi^2(\Omega, P)$, which equals 0 under H_0, since in that case $P = Q^*$. Therefore, under H_0, there is no restriction when considering Ω a subset of M.

The χ^2 distance belongs to the so-called $\phi-$divergences, defined through

$$\phi(Q,P) = \begin{cases} \int \varphi\left(\frac{dQ}{dP}\right) dP & \text{when } Q \ll P \\ \infty & \text{otherwise} \end{cases} \tag{1.3}$$

where φ is a convex function defined on $\mathbb{R}^+$ satisfying $\varphi(1) = 0$. This class of discrepancy measures between probability measures has been introduced by I. Csiszár [10], and the monograph by F. Liese and I. Vajda [20] provides their main properties.

The extension of $\phi-$divergences when Q is assumed to be in M is presented in [9], in the context of parametric estimation and tests.

The class of minimum $\phi-$divergence test statistics include, within the others, the *loglikelihood ratio test*.

For this class it is a matter of fact that first order efficiency is not a useful criterion of discrimination. A notion of robustness against model contamination is found in Lindsay [21] (for estimators) and in Jimenez and Shao [16] (for test procedures), which gives an instrument to compare the tests associated to different divergences. Although their argument deals with finite support models, it may help as a benchmark for more general situations.

By these papers it emerges that the *minimum Hellinger distance* test provides a reasonable compromise between robustness against model contaminations induced by outliers and by inliers.

However, when the model might be subject to inlier contaminations only (namely missing data problems), as will be advocated in the present paper for contamination models, then minimum χ^2-divergence test behaves better than both minimum Hellinger distance and loglikelihood ratio tests, in terms of their *residual adjustment functions* (RAF), because (we refer to [16] for the notation)

$$\left|\frac{A_{\chi^2}(-1)}{A_{LR}(-1)}\right| = \frac{1}{2} < 1 \quad \text{and} \qquad \left|\frac{A_{\chi^2}(-1)}{A_{HD}(-1)}\right| = \frac{1}{4} < 1.$$

Formula (1.1) is not suitable for statistical purposes as such. Indeed, suppose that we are interested in testing wether P is in some class Ω of distributions with absolutely continuous component. Let $X = (X_1, \ldots, X_n)$ be an i.i.d. sample with unknown distribution P. Assume that $P_n := \frac{1}{n}\sum_{i=1}^n \delta_{X_i}$, the empirical measure pertaining to X, is the only information available on P, where δ_x is the Dirac measure at point x. Then, for all $Q \in \Omega$, the χ^2 distance between Q and P_n is infinite. Therefore no plug-in technique can lead to a definite statistic in this usual case.

Our approach solves this difficulty and is based on the "dual representation " for the χ^2 divergence, which is a consequence of the convexity of the mapping $Q \longmapsto \chi^2(Q,P)$, plus some regularity property; this will be set in Section 2, together with conditions under which P has a χ^2-projection on Ω. We will also provide an estimate for the function $\frac{dQ^*}{dP}$ which indicates the local changes induced on P by the projection operator.

In some cases it is possible to replace Ω by $\Omega \cap \Lambda_n$ where Λ_n is the set of all measures in M whose support is X, when this intersection is not void, as happens when Ω is defined for example through moment conditions. This approach is called the Generalized Likelihood paradigm (see [24] and references therein), and we will develop in Section 3 a complete study pertaining to such case when handling the χ^2 divergence, in the event that Ω is defined through linear constraints, namely when

$$\Omega = \left\{ Q \in M \text{such that} \int f(x) dQ(x) = 0 \right\} \tag{1.4}$$

for some $\mathbb{R}^k-$valued function f defined on $\mathbb{R}^d$. In this case the projection Q^* has a very simple form and its estimation results as the solution of a linear system of equations, which motivates the choice of χ^2 criterion for tests of the form $H_0 : P \in \Omega$ with Ω as in (1.4). As is shown in Section 3, by Theorem 2.2 the constrained problem is in fact reduced to an unconstrained one.

Also for the problem of testing whether P belongs to Ω our results include the asymptotic distribution of the test statistics under any P in the alternative, proving consistency of the procedure, a result that is not addressed in the current literature on Generalized Likelihood.

In Section 3 we will apply the above results to the case of a test of fit, where $\Omega = \{P_0\}$ is a fixed p.m.

When $\Omega \cap \Lambda_n$ is void some smoothing technique has been proposed, following [2], substituting P_n by some regularized version; see [23]. In those cases we have chosen not to make any smoothing, exploiting the dual representation in a parametric context. Section 4 addresses this approach through the study of contamination models, for a composite problem, when the contamination modifies a distribution with unknown parameter.

2. The definition of the estimator

2.1. **Some properties of χ^2-distance.** We will consider sets Ω of signed measures with total mass 1 that integrate some class of functions Φ. The choice of Φ depends on the context as seen below. Let

$$M_\Phi := \left\{ Q \in M \text{ such that } \int |\varphi| \, d|Q| < \infty, \text{ for all } \varphi \in \Phi \right\}. \tag{2.1}$$

We first consider sufficient conditions for the existence of Q^*, the projection of P on Ω. We introduce the following notation.

Let $\mathbf{\Phi} = \Phi \cup \mathcal{B}_b$, where $\mathcal{B}_b$ is the class of all measurable bounded functions on $\mathbb{R}^d$. Let $\tau_{\mathbf{\Phi}}$ be the coarsest topology on M which makes all mappings $Q \longmapsto \int \varphi dQ$ continuous for all $\varphi \in \mathbf{\Phi}$. When $\mathbf{\Phi}$ is restricted to $\mathcal{B}_b$, the $\tau_{\mathbf{\Phi}}$ topology turns out to be the usual $\tau-$ topology (see e.g. [13]).

Assume that for all functions φ in Φ there exists some positive ε with

$$\int \varphi^{2+\varepsilon} dP < \infty.$$

Whenever Ω is a closed set in M_Φ equipped with the $\tau_{\mathbf{\Phi}}$ topology and $\chi^2(\Omega, P)$ is finite, then, as a consequence of Theorem 2.3 in [9], P has a projection in Ω. Moreover, when Ω is convex, uniqueness is achieved.

In statistical applications the set Ω is often defined through some statistical functional; for example, let Ω defined as in (1.4). In this case $\Phi := \{f\}$ and Ω is closed by the very definition of Φ; therefore the choice of the class of functions Φ is intimately connected with the set Ω under consideration. As seen in Section 4, and as developed in [9] also when Ω is a subset of some parametric family of distributions, the class Φ can be defined with respect to Ω.

We first provide a characterization of the χ^2-projection of a p.m. P on some set Ω in M.

Let $\mathcal{D}$ denote the domain of the divergence for fixed P, namely

$$\mathcal{D} = \left\{Q \in M \text{ such that } \chi^2(Q, P) < \infty\right\}.$$

We have (see [8], Theorem 2.6)

Theorem 2.1. *Let Ω be a subset of M. Then*

(1) *If there exists some Q^* in Ω such that for all Q in $\Omega \cap \mathcal{D}$,*

$$q^* \in L_1(Q) \text{ and } \int q^* dQ^* \leq \int q^* dQ$$

where $q^* = \frac{dQ^*}{dP}$, then Q^* is the χ^2-projection of P on Ω

(2) *If Ω is convex and P has projection Q^* on Ω, then, for all Q in Ω, q^* belongs to $L_1(P)$ and $\int q^* dQ^* \leq \int q^* dQ$.*

Many statistically relevant problems in estimation and testing pertain to models defined by linear constraints (Empirical Likelihood paradigm and others). Section 3 is devoted to this case. We therefore present a characterization result for the χ^2-projection on sets of measures defined by linear constraints.

Let Φ be a collection (finite or infinite, countable or not) of real valued functions defined on $\mathbb{R}^d$, which we assume to contain the function 1. Let Ω a subset of M be defined by

$$\Omega = \left\{Q \in M \text{ such that } \int g dQ = 0 \text{ for all } g \text{ in } \Phi - \{1\}\right\}.$$

Denote $< \Phi >$ the linear span of Φ.

We then have the following result (see [8]):

Theorem 2.2. (1) *P has a projection Q^* in Ω iff Q^* belongs to Ω and for all $Q \in \Omega$, $q^* \in L_1(Q)$ and $\int q^* dQ^* \leq \int q dQ^*$.*

(2) *If q^* belongs to $< \Phi >$ and Q^* belongs to Ω, then Q^* is the projection of P on Ω.*

(3) *If P has projection Q^* on Ω, the q^* belongs to $\overline{< \Phi >}$, the closure of Φ in $L_1(Q^*)$.*

Remark 2.3. The above result only provides a partial answer to the characterization of the projections. Let P be the uniform distribution on $[0,1]$. The set $M_1(P)$ of all p.m.'s absolutely continuous with respect to P is a closed subset of M_Φ, when $\Phi := \{x \mapsto x\} \cup \{x \mapsto 1\}$. Let $\Omega := \left\{Q \in M_1(P) : \int x dQ(x) = \frac{1}{4}\right\}$. Then P has a projection on Ω and $\frac{dQ^*}{dP}(x) 1_{\{q^*>0\}}(x) = c_0 + c_1 x$, with $q^* = \frac{dQ^*}{dP}$. The support of Q^* is strictly included in $[0,1]$. Otherwise we obtain $c_0 = \frac{5}{2}$ and $c_1 = -3$, a contradiction, since then Q^* is not a probability measure.

2.2. **An alternative version of the χ^2.** The χ^2 distance defined on M for fixed P in M_1 through $\chi^2(Q,P) = \int \left(\frac{dQ}{dP} - 1\right)^2 dP$ is a convex function; as such it is the upper envelope of its support hyperplanes. The first result, which is Proposition 2.1 in [9], provides the description of the hyperplanes in M_Φ.

Proposition 2.4. *Equip M_Φ with the $\tau_{\mathbf{\Phi}}-$topology. Then M_Φ is a Hausdorff locally convex topological space. Further, the topological dual space of M_Φ is the set of all mappings $Q \mapsto \int f dQ$ when f belongs to $< \mathbf{\Phi} >$.*

Proposition 2.3 in [9] asserts that the χ^2 distance defined on M for fixed P in M_1 is l.s.c. in $(M_\Phi, \tau_{\mathbf{\Phi}})$. We can now state the duality lemma.

Define on $< \mathbf{\Phi} >$, the Fenchel-Legendre transform of $\chi^2(\cdot, P)$

$$T(f,P) := \sup_{Q \in M_\Phi} \int f dQ - \chi^2(Q,P). \tag{2.2}$$

We have

Lemma 2.5. *The function $Q \longmapsto \chi^2(Q,P)$ admits the representation*

$$\chi^2(Q,P) = \sup_{f \in < \mathbf{\Phi} >} \int f dQ - T(f,P). \tag{2.3}$$

Standard optimization techniques yield

$$T(f,P) = \int f dP + \frac{1}{4} \int f^2 dP$$

for all $f \in < \mathbf{\Phi} >$, see e.g. [1], Chapter 4.

The function $f^* = 2\left(\frac{dQ^*}{dP} - 1\right)$ is the supremum in (2.3) as can be seen through classical convex optimization procedures.

We now consider a subclass $\mathcal{F}$ in $< \mathbf{\Phi} >$ and we assume:

(C1) f^* belongs to $\mathcal{F}$.

Therefore

$$\chi^2(Q,P) = \sup_{f\in\mathcal{F}} \int f dQ - T(f,P)$$

which we call the *dual representation* of the χ^2.

This can be restated as follows: let

$$m_f(x) := \int f dQ - \left(f(x) + \frac{1}{4} f^2(x) \right).$$

Then

$$\chi^2(Q,P) = \sup_{f\in\mathcal{F}} \int m_f(x) dP(x). \tag{2.4}$$

Hence we have

$$\chi^2(\Omega,P) = \inf_{Q\in\Omega} \sup_{f\in\mathcal{F}} \int m_f(x) dP(x). \tag{2.5}$$

In the case when Ω is defined through a finite number of linear constraints, say

$$\Omega = \left\{ Q \in M : \int f_i(x) dQ(x) = a_i,\ 1 \le i \le k \right\},$$

when P has a projection Q^* on Ω and $supp\{Q^*\}$ is known to coincide with that of P, then we may choose $\mathcal{F}$ as the linear span of $\{1, f_1, \ldots, f_k\}$ and (2.5) turns out to be a parametric unconstrained optimization problem, since, by Theorem 2.2 (3)

$$\chi^2(\Omega,P) = \sup_{c_0,c_1,\ldots,c_k} c_0 + \sum_{i=1}^k c_i a_i - T\left(c_0 + \sum_{i=1}^k c_i f_i, P \right).$$

In some other cases we may have a complete description of all functions $\frac{dQ}{dP}$ when Q belongs to Ω. A typical example is when P and Q belong to parametric families.

2.3. **The estimator χ^2_n.** Let us now present the estimate of $\chi^2(\Omega,P)$.

Together with an i.i.d. sample $X_1,\ldots,X_n$ with common unknown distribution P, define the estimate of $\chi^2(Q,P)$ through

$$\chi^2_n(Q,P) := \sup_{f\in\mathcal{F}} \int m_f(x) dP_n(x) \tag{2.6}$$

a plug-in version of (2.4).

We also define the estimate of $\chi^2(\Omega,P)$ through

$$\chi^2_n(\Omega,P) := \inf_{Q\in\Omega} \sup_{f\in\mathcal{F}} \int m_f(x) dP_n(x). \tag{2.7}$$

These estimates may seem cumbersome. However, in the case when we are able to reduce the class $\mathcal{F}$ to a reasonable degree of complexity, these estimates perform quite well and can be used for testing $P \in \Omega$ against $P \notin \Omega$. This will be made clear in the last two sections which serve as examples for the present approach.

In some cases it is possible to commute the sup and the inf operators in (2.5), which turns out to become

$$\chi^2(\Omega,P) = \sup_{f\in\mathcal{F}} \inf_{Q\in\Omega} \int f dQ - T(f,P), \tag{2.8}$$

in which the inf operator acts only on the linear functional $\int f dQ$.

Also, when (2.8) holds, we may define an estimate of $\chi^2(\Omega, P)$ through

$$\overline{\chi}_n^2(\Omega, P) = \sup_{f\in\mathcal{F}} \inf_{Q\in\Omega} \int f dQ - T(f, P_n). \tag{2.9}$$

When (2.8) holds, it is quite easy to get the limit properties of $\overline{\chi}_n^2$.

Indeed, by (2.8) and (2.9)

$$\overline{\chi}_n^2(\Omega, P) - \chi^2(\Omega, P) = \left(\sup_{f\in\mathcal{F}} \inf_{Q\in\Omega} \int f dQ - T(f, P_n)\right) - \left(\sup_{f\in\mathcal{F}} \inf_{Q\in\Omega} \int f dQ - T(f, P)\right).$$

Now define

$$\phi_R(f) := \inf_{Q\in\Omega} \int f dQ - T(f, R) = \inf_{Q\in\Omega} \int f dQ - \int \left(f + \frac{1}{4}f^2\right) dR,$$

a concave function of f.

When $\mathcal{F}$ is compact in a topology for which ϕ_R is uniformly continuous for all R in M_1, then a sufficient condition for the a.s. convergence of $\overline{\chi}_n^2(\Omega, P)$ to $\chi^2(\Omega, P)$ is

$$\lim_{n\to\infty} \sup_{f\in\mathcal{F}} |\phi_{P_n}(f) - \phi_P(f)| = 0 \ \ a.s.$$

which in turn is

$$\lim_{n\to\infty} \sup_{f\in\mathcal{F}} \left| \int \left(f + \frac{1}{4}f^2\right) dP_n - \int \left(f + \frac{1}{4}f^2\right) dP \right| = 0 \ \ a.s.$$

This clearly holds when the class of functions $\left\{\left(f + \frac{1}{4}f^2\right),\ f \in \mathcal{F}\right\}$ satisfies the functional Glivenko-Cantelli (GC) condition (see [25]).

The limit distribution of the statistic $\overline{\chi}_n^2(\Omega, P)$ under $H1$, i.e. when P does not belong to Ω, can be obtained under the following hypotheses, following closely the proof of Theorem 3.6 in [7], where a similar result is proved for the Kullback-Leibler divergence estimate.

Assume

(C2) P has a unique projection Q^* on Ω.

(C3) The class $\mathcal{F}$ is compact in the sup-norm.

(C4) The class $\left\{f + \frac{1}{4}f^2,\ f \in \mathcal{F}\right\}$ is a functional Donsker class.

We then have

Theorem 2.6. *Under $H1$, assume that* (C1)–(C4) *hold. The asymptotic distribution of*

$$\sqrt{n}\left(\overline{\chi}_n^2(\Omega, P) - \chi^2(\Omega, P)\right)$$

is that of $B_P(g^*)$, where $B_P(\cdot)$ is the $P-$Brownian bridge defined on $\mathcal{F}$, and $g^* = -f^* - \frac{1}{4}f^{*2}$.

Therefore $\sqrt{n}\left(\overline{\chi}_n^2(\Omega, P) - \chi^2(\Omega, P)\right)$ has an asymptotic centered normal distribution with variance $E_P\left(\left(f^* + \frac{1}{4}f^{*2}\right)^2(X)\right) - E_P\left(\left(-f^* - \frac{1}{4}f^{*2}\right)(X)\right)^2$, where X has law P.

The asymptotic distribution of $\overline{\chi}^2_n$ under $H0$, i.e. when P belongs to Ω, cannot be obtained in a general frame and must be derived accordingly to the context.

In the next Sections we develop two applications of the above statements. In the first one we consider sets Ω defined by an infinite number of linear constraints. We approximate Ω through some sieve technique and provide consistent test for $H_0 : P \in \Omega$. We specialize this problem to the two sample test for paired data. So, in this first application, we basically use the representation of the projection Q^* of P on linear sets as described through Theorem 2.2. In this first range of applications we will project P_n on the non void set $\Omega \cap \Lambda_n$.

The second application deals with parametric models and test for contamination. We obtain a consistent test for the case when Ω is a set of parametrized distributions F_θ for θ in $\Theta \subset \mathbb{R}^d$. The test is

$$H0 : \ P \in \Omega = \{F_\theta, \, \theta \in \Theta\}, \text{i.e. } \lambda = 0 \quad \text{vs} \quad H1 : P \in \{(1-\lambda)F_\theta + \lambda R, \, \lambda \neq 0, \, \theta \in \Theta\}.$$

In this example we project P_n on a set of absolutely continuous distributions and we make use of the minimax assumption (2.8) which we prove to hold.

3. Test of a set of linear constraints

Let $\mathcal{F}$ be a countable family of real-valued functions defined on $\mathbb{R}^d$, $\{a_i\}_{i=1}^{\infty}$ a real sequence and

$$\Omega := \left\{ Q \in M \text{ such that } \int f_i dQ(x) = a_i, \ i \geq 1 \right\} \tag{3.1}$$

We assume that Ω is not void. In accordance with the previous section we assume that the function $f_0 := 1$ belongs to $\mathcal{F}$ with $a_0 = 1$.

Let $X_1, \ldots, X_n$ be an i.i.d. sample with common distribution P.

We intend to propose a test for $H_0 : P \in \Omega$ vs $H_1 : P \notin \Omega$.

We first consider the case when $\mathcal{F}$ is a finite collection of functions, and next extend our results to the infinite case.

For notational convenience we write Pf for $\int f dP$ whenever defined.

3.1. Finite number of linear constraints.

Consider the set Ω defined in (3.1) with $card\{\mathcal{F}\} = k$. Introduce the estimate of $\chi^2(\Omega, P)$ through

$$\chi^2_n(\Omega, P) = \inf_{Q \in \Omega \cap \Lambda_n} \chi^2(Q, P_n). \tag{3.2}$$

Embedding the projection device in $M \cap \Lambda_n$ instead of $M_1 \cap \Lambda_n$ yields to a simple solution for the optimum in (3.2), since no inequality constrains will be used. Also the topological context is simpler than as mentioned in the previous section since the projection of P_n belongs to $\mathbb{R}^n$. When developed in $M_1 \cap \Lambda_n$ this approach is known as the Generalized Likelihood (GEL) paradigm (see [11]). Our approach differs from the lattest through the use of the dual representation (2.7), which provides consistency of the test procedure.

It is readily checked that

$$\chi_n^2(\Omega, P) = \overline{\chi}_n^2(\Omega, P)$$

The set $\Omega \cap \Lambda_n$ is a convex closed subset in $\mathbb{R}^n$. When the projection of P_n on $\Omega \cap \Lambda_n$ exists uniqueness therefore holds. In the next section we develop various properties of our estimates, which are based on the duality formula (2.6).

The next subsections provide all limit properties of $\chi_n^2(\Omega, P)$.

3.1.1. *Notation and basic properties.* Let Q_0 be any fixed measure in Ω. By (2.5)

$$\begin{aligned}\chi^2(\Omega, P) &= \sup_{f \in <\mathcal{F}>} (Q_0 - P) f - \frac{1}{4} P f^2 \\ &= \sup_{a_0, a_1, \ldots, a_k} \sum_{i=1}^{k} a_i (Q_0 - P) f_i - \frac{1}{4} P \left(\sum_{i=1}^{k} a_i f_i + a_0 \right)^2 . \end{aligned} \tag{3.3}$$

since, for Q in Ω and for all f in $\mathcal{F}$, $Qf = Q_0 f$ and

$$\chi_n^2 = \sup_{a_0, a_1, \ldots, a_k} \sum_{i=1}^{k} a_i (Q_0 - P_n) f_i - \frac{1}{4} P_n \left(\sum_{i=1}^{k} a_i f_i + a_0 \right)^2 .$$

The infinite dimensional optimization problem in (2.5) thus reduces to a $(k+1)-$dimensional one, much easier to handle.

We can write the chi-square and χ_n^2 through a quadratic form.

Define the vectors $\underline{\nu}_n$ e $\underline{\nu}$ by

$$\begin{aligned} \underline{\nu}_n{}' &= \underline{\nu}(\mathcal{F}, P_n)' = \{(Q_0 - P_n) f_1, \ldots, (Q_0 - P_n) f_k\} \\ \underline{\nu}' &= \underline{\nu}(\mathcal{F}, P)' = \{(Q_0 - P) f_1, \ldots, (Q_0 - P) f_k\} \\ \underline{\gamma}_n &= \underline{\gamma}_n(\mathcal{F}) = \sqrt{n} \{(P_n - P) f_1, \ldots, (P_n - P) f_k\} = \sqrt{n} (\underline{\nu} - \underline{\nu}_n) . \end{aligned} \tag{3.4}$$

Let S be the covariance matrix of $\underline{\gamma}_n$. Write S_n for the empirical version of S, obtained substituting P by P_n in all entries of S.

Proposition 3.1. *Let Ω be as in* (3.1) *and let card $\{\mathcal{F}\}$ be finite. We then have*

(i) $\chi_n^2 = \underline{\nu}_n' S_n^{-1} \underline{\nu}_n$

(ii) $\chi^2(\Omega, P) = \underline{\nu}' S^{-1} \underline{\nu}$

Proof. (i) Differentiating the function in (3.3) with respect to a_s, $s = 0, 1, \ldots, k$ yields

$$a_0 = -\sum_{i=1}^{k} a_i P_n f_i \tag{3.5}$$

for $s = 0$, while for $s > 0$

$$(Q_0 - P_n) f_s = \frac{1}{2} \left(a_0 P_n f_s + \sum_{i=1}^{k} a_i P_n f_i f_s \right) . \tag{3.6}$$

Substituting (3.5) in the last display,

$$(Q_0 - P_n) f_s = \frac{1}{2} \sum_{i=1}^{k} a_i (P_n f_i f_s - P_n f_s P_n f_i) ,$$

i.e.

$$2\underline{\nu}_n = S_n \underline{a} \tag{3.7}$$

where $\underline{a}' = \{a_1, a_2, \ldots, a_k\}$.

Set $f_n^* = \arg\max_{<\mathcal{F}>}(Q_0 - P_n)f - \frac{1}{4}P_n f^2$. For every $h \in <\mathcal{F}>$, $(Q_0 - P_n)\, h - \frac{1}{2}P_n h f_n^* = 0$. Set $h := f_n^*$ to obtain $(Q_0 - P_n)\, f_n^* = \frac{1}{2}P_n (f_n^*)^2$.

It then follows, using (3.5) e (3.7),

$$\chi_n^2 = \left[(Q_0 - P_n)\, f_n^* - \frac{1}{4}P_n (f_n^*)^2\right] = \frac{1}{4}P_n (f_n^*)^2 =$$

$$= \frac{1}{4}P_n \left(\sum_{i=1}^k a_i f_i - \sum_{i=1}^k a_i P_n f_i\right)^2 =$$

$$= \frac{1}{4}\underline{a}' S_n \underline{a} = \underline{\nu}_n' S_n^{-1} \underline{\nu}_n.$$

(ii) The proof is similar to the above one.

□

3.1.2. *Almost sure convergence.* Call an envelope for $\mathcal{F}$ a function F such that $|f| \leq F$ for all f in $\mathcal{F}$.

Theorem 3.2. *Assume that $\chi^2(\Omega, P)$ is finite. Let $\mathcal{F}$ be a finite class of functions as in (3.1) with an envelope function F such that $PF^2 < \infty$.*
Then $\left|\chi_n^2 - \chi^2(\Omega, P)\right| \to 0$, $P - a.s.$

Proof. From Proposition 3.1,

$$\left|\chi_n^2 - \chi^2(\Omega, P)\right| = \left|\underline{\nu}_n' S_n^{-1} \underline{\nu}_n - \underline{\nu} S^{-1} \underline{\nu}\right|$$

$$= \left|\underline{\nu}_n' \left(S_n^{-1} - S^{-1}\right) \underline{\nu}_n\right| + \left|\underline{\nu}_n' S^{-1} \underline{\nu}_n - \underline{\nu}' S^{-1} \underline{\nu}\right|.$$

For $\underline{x}$ in $\mathbb{R}^k$ denote $\|\underline{x}\|$ the euclidean norm. Over the space of matrices $k \times k$ introduce the algebraic norm $|\!|\!| A |\!|\!| = \sup_{\|\underline{x}\| \leq 1} \frac{\|A\underline{x}\|}{\|\underline{x}\|} = \sup_{\|\underline{x}\|=1} \|A\underline{x}\|$. All entries of A satisfy $|a(i,j)| \leq |\!|\!| A |\!|\!|$. Moreover, if $|\lambda_1| \leq |\lambda_2| \leq \ldots \leq |\lambda_k|$ are the eigenvalues of A, $|\!|\!| A |\!|\!| = |\lambda_k|$. Observe further that, if for all (i,j), $|a(i,j)| \leq \varepsilon$, then, for any $\underline{x} \in \mathbb{R}^k$, such that $\|\underline{x}\| = 1$, $\|A\underline{x}\|^2 = \sum_{i=1}^k \left(\sum_j a(i,j)\, x_j\right)^2 \leq \sum_i \sum_j a(i,j)^2 \|\underline{x}\|^2 \leq k^2\varepsilon^2$, i.e. $|\!|\!| A |\!|\!| \leq k\varepsilon$.

For the first term in the RHS of the above display

$$A := \underline{\nu}_n{}' \left(S_n^{-1} - S^{-1}\right) \underline{\nu}_n = \underline{\nu}_n{}' S^{-1/2} \left(S^{1/2} S_n^{-1} S^{1/2} - I\right) S^{-1/2} \underline{\nu}_n$$

$$\leq \left\|\underline{\nu}_n{}' S^{-1/2}\right\| \left|\!\left|\!\left| S^{1/2} S_n^{-1} S^{1/2} - I \right|\!\right|\!\right| \leq cost.\ k \left|\!\left|\!\left| S^{1/2} S_n^{-1} S^{1/2} - I \right|\!\right|\!\right|.$$

Hence if $B := |\!|\!| S^{1/2} S_n^{-1} S^{1/2} - I |\!|\!|$ tends to 0 a.s., so does A.

First note that

$$S_n^{-1} = (S + S_n - S)^{-1} = S^{-1/2} \left(I + S^{-1/2}(S_n - S) S^{-1/2}\right)^{-1} S^{-1/2} =$$

$$= S^{-1/2} \left[I + \sum_{h=1}^{\infty} \left(S^{-1/2}(S - S_n) S^{-1/2}\right)^h\right] S^{-1/2}.$$

Hence

$$S^{1/2}S_n^{-1}S^{1/2} - I = \sum_{h=1}^{\infty} \left(S^{-1/2} (S - S_n) S^{-1/2} \right)^h,$$

which entails

$$\left\| S^{1/2}S_n^{-1}S^{1/2} - I \right\| = \left\| \sum_{h=1}^{\infty} \left(S^{-1/2} (S - S_n) S^{-1/2} \right)^h \right\| \leq \sum_{h=1}^{\infty} \|S - S_n\|^h \left\| S^{-1/2} \right\|^{2h}$$
$$= O_P \left(\lambda_1^{-1} k \sup_{i,j} |s_n(i,j) - s(i,j)| \right),$$

where λ_1 is the smallest eigenvalue of S.

Since

$$C := \sup_{i,j} |s_n(i,j) - s(i,j)| \leq \sup_{i,j} |(P_n - P) f_i f_j| + \sup_i |(P_n - P) f_i| \, |(P_n + P) F| \tag{3.8}$$

the LLN implies that C tends to 0 a.s. which in turn implies that B tends to 0.

Now consider the second term. $\left| \underline{\nu}_n{}' S^{-1} \underline{\nu}_n - \underline{\nu}' S^{-1} \underline{\nu} \right| = \left| (\underline{\nu}_n + \underline{\nu})' S^{-1} \left(n^{-1/2} \underline{\gamma}_n \right) \right|$ tends to 0 by LLN. □

3.1.3. *Asymptotic distribution of the test statistic.* Write

$$n\chi_n^2 = \sqrt{n}\underline{\nu}_n' S^{-1} \sqrt{n}\underline{\nu}_n + \sqrt{n}\underline{\nu}_n' \left(S_n^{-1} - S^{-1} \right) \sqrt{n}\underline{\nu}_n.$$

We then have

Theorem 3.3. *Let Ω be defined by (3.1) and $\mathcal{F}$ be a finite class of linearly independent functions with envelope function F such that $PF^2 < \infty$. Set $k = card\{\mathcal{F}\}$. Then, under $H0$,*

$$n\chi_n^2 \xrightarrow{d} chi(k)$$

where $chi(k)$ denotes a chi-square distribution with k degrees of freedom.

Proof. For P in Ω, $\sqrt{n}\underline{\nu}_n = \underline{\gamma}_n$. Therefore $n\chi_n^2 = \underline{\gamma}_n{}' S^{-1} \underline{\gamma}_n + \sqrt{n}\underline{\nu}_n{}' \left(S_n^{-1} - S^{-1} \right) \sqrt{n}\underline{\nu}_n$.

By continuity of the mapping $h\left(\underline{y}\right) = \underline{y}' S^{-1} \underline{y}$, $\underline{\gamma}_n' S^{-1} \underline{\gamma}_n$ has a limiting $chi(k)$ distribution.

It remains to prove that the second term is negligible. Indeed again from

$$\left(\sqrt{n}\underline{\nu}_n \right)' \left(S_n^{-1} - S^{-1} \right) \left(\sqrt{n}\underline{\nu}_n \right) \leq cst.\, k \left\| S^{1/2}S_n^{-1}S^{1/2} - I \right\|$$

it is enough to show that $\left\| S^{1/2}S_n^{-1}S^{1/2} - I \right\|$ is $o_P(1)$. This follows from (3.8). □

The asymptotic behavior of χ_n^2 under $H1$ is captured by

$$\sqrt{n} \left(\chi_n^2 - \chi^2 \right) = -2\underline{\gamma}_n' S^{-1} \underline{\nu} + \sqrt{n}\underline{\nu}' S^{-1/2} \left(S^{1/2}S_n^{-1}S^{1/2} - I \right) S^{-1/2} \underline{\nu}$$
$$- 2\underline{\gamma}_n' S^{-1/2} \left(S^{1/2}S_n^{-1}S^{1/2} - I \right) S^{-1/2} \underline{\nu} + n^{-1/2} \underline{\gamma}_n' S_n^{-1} \underline{\gamma}_n. \tag{3.9}$$

This proves that the test based on $n\chi_n^2$ is asymptotically consistent.

3.2. **Infinite number of linear constraints, an approach by sieves.** In various cases Ω is defined through a countable collection of linear constraints. An example is presented in Section 3.3. Suppose thus that Ω is defined as in (3.1), with $\mathcal{F}$ an infinite class of functions

$$\mathcal{F} = \left\{ f_\alpha : \mathbb{R}^d \to \mathbb{R}, \quad \alpha \in A \right\}$$

where $A \subseteq \mathbb{R}$ is a countable set of indices and $card\,(\mathcal{F}) = card\,(A) = \infty$. Thus $\Omega = \{Q \in M : Qf = Q_0 f,\ f \in \mathcal{F}\}$, for some Q_0 in M.

Assume that the projection Q^* exists in Ω. Then, by Theorem 2.2

$$f^* \in cl_{L_1(Q^*)} \left(< \mathcal{F} > \right).$$

We approximate $\mathcal{F}$ through a suitable increasing sequence of classes of functions $\mathcal{F}_n$, with finite cardinality $k = k(n)$ increasing with n. Each $\mathcal{F}_n$ induces a subset Ω_n included in Ω.

Define therefore $\{\mathcal{F}_n\}_{n\geq 1}$ such that

$$\mathcal{F}_n \subseteq \mathcal{F}_{n+1} \subset \mathcal{F}, \text{ for all } \quad n \geq 1 \tag{3.10}$$

$$\mathcal{F} = \bigcup_{n\geq 1} \mathcal{F}_n \tag{3.11}$$

and

$$\Omega_n = \{Q : Qf = Q_0 f, \quad f \in \mathcal{F}_n\}.$$

We thus have $\Omega_n \supseteq \Omega_{n+1}$, $n \geq 1$ and $\Omega = \bigcap_{n\geq 1} \Omega_n$.

The idea of determining the projection of a measure P on a set Ω through an approximating sequence of sets -or sieve- has been introduced in this setting in [27].

Theorem 3.4 (Teboulle-Vajda, 1993)**.** *With the above notation, define Q_n^* as the projection of P on Ω_n. Suppose that the above assumptions on $\{\Omega_n\}_{n\geq 1}$ hold and that $\Omega_n \supseteq \Omega$ for each $n \geq 1$. Then*

$$\lim_{n\to\infty} \|f^* - f_n^*\|_{L_1(P)} = \lim_{n\to\infty} \left\| \frac{dQ^*}{dP} - \frac{dQ_n^*}{dP} \right\|_{L_1(P)}. \tag{3.12}$$

By Scheffe's Lemma this is equivalent to $\lim_{n\to\infty} d_{var}(Q_n^*, Q^*) = 0$ where $d_{var}(Q, P) := \sup_{A\in\mathcal{B}(\mathbb{R}^d)} |Q(A) - P(A)|$ is the variation distance between the p.m's P and Q. When $\sup_{f\in\mathcal{F}} \sup_x f(x) < \infty$ then (3.12) implies

$$\lim_{k\to\infty} \chi^2(\Omega_n, P) = \chi^2(\Omega, P) \tag{3.13}$$

The above result states that we can build a sequence of estimators of $\chi^2(\Omega, P)$ letting $k = k(n)$ grow to infinity together with n. Define

$$\chi^2_{n,k} = \sup_{f\in\mathcal{F}_n} (Q_0 - P_n) f - \frac{1}{4} P_n f^2.$$

In the following section we consider conditions on $k(n)$ entailing the asymptotic normality of the suitably normalized sequence of estimates $\chi_{n,k}$ when P belongs to Ω, i.e. under $H0$.

3.2.1. *Convergence in distribution under H0.* As a consequence of Theorem 3.3, $n\chi^2_{n,k}$ tends to infinity with probability 1 as $n \to \infty$.

We consider the statistics

$$\frac{n\chi^2_{n,k} - k}{\sqrt{2k}} \tag{3.14}$$

which will be seen to have a nondegenerate distribution as $k(n)$ tends to infinity together with n.

As in [14] and [15], the main tool of the proof of the asymptotic normality of (3.14) relies on the strong approximation of the empirical processes. We briefly recall some useful notions.

Definition 3.5. A class of functions $\mathcal{F}$ is *pregaussian* if there exists a version $B^0_P(.)$ of $P-$Brownian bridges uniformly continuous in $\ell^\infty(\mathcal{F})$, with respect to the metric $\rho_P(f,g) = (Var_P|f-g|)^{1/2}$, where $\ell^\infty(\mathcal{F})$ is the Banach space of all functionals $H : \mathcal{F} \to \mathbb{R}$ uniformly bounded and with norm $\|H\|_{\mathcal{F}} = \sup_{f\in\mathcal{F}} |H(f)|$.

For some $a > 0$, let δ_n be a decreasing sequence with $\delta_n = o(n^{-a})$.

Definition 3.6. A class of functions $\mathcal{F}$ is *Komlós-Major-Tusnády* (KMT) with respect to P, with rate δ_n $(\mathcal{F} \in KMT(\delta_n; P))$ iff it is pregaussian and there exists a version $B^0_n(.)$ of $P-$Brownian bridges such that for any $t > 0$ it holds

$$\Pr\left\{\sup_{f\in\mathcal{F}} \left|\sqrt{n}(P_n - P)f - B^0_n(f)\right| \geq \delta_n(t + b\log n)\right\} \leq ce^{-\theta t}, \tag{3.15}$$

where the positive constants b, c and θ depend on $\mathcal{F}$ only.

We refer to [4], [22], [6], and [17] for examples of classical and useful classes of KMT classes, together with calculations of rates; we will use the fact that a KMT class is also a Donsker class.

From (3.15) and Borel-Cantelli lemma it follows that

$$\sup_{f\in\mathcal{F}} \left|\gamma_n(f) - B^0_n(f)\right| = O(\delta_n \log n) \tag{3.16}$$

a.s. where, with the same notation as in the finite case (see (3.4)), $\gamma_n(f) = \sqrt{n}(P_n - P)f$ is the empirical process indexed by $f \in \mathcal{F}$.

Let $\{\mathcal{F}_n\}_{n\geq 1}$ be a sequence of classes of linearly independent functions satisfying (3.10).

For any n, set $\underline{\gamma}_{n,k} = \underline{\gamma}_n(\mathcal{F}_n)$ (resp. $\underline{B}^0_{n,k}$) the $k-$dimensional vector resulting from the projection of the empirical process γ_n (resp. of the $P-$Brownian bridge B^0_n) defined on $\mathcal{F}$ to the subset $\mathcal{F}_n$. Then, if $\mathcal{F}_n = \left\{f_1^{(n)}, \ldots, f_k^{(n)}\right\}$, $\underline{\gamma}_{n,k} = \left\{\gamma_n(f_1^{(n)}), \ldots, \gamma_n(f_k^{(n)})\right\}$ and $\underline{B}^0_{n,k} = \left\{B^0_n\left(f_1^{(n)}\right), \ldots, B^0_n\left(f_k^{(n)}\right)\right\}$. Denote S_k the covariance matrix of the vector $\underline{\gamma}_{n,k}$ and $S_{n,k}$ its empirical covariance matrix. Let $\lambda_{1,k}$ be the smallest eigenvalue of S_k.

Theorem 3.7. *Let $\mathcal{F}$ have an envelope F and be $KMT(\delta_n; P)$ for some sequence $\delta_n \downarrow 0$. Define further a sequence $\{\mathcal{F}_n\}_{n\geq 1}$ of classes of linearly independent functions satisfying* (3.10).

Moreover, let k satisfy

$$\lim_{n\to\infty} k(n) = \infty$$
$$\lim_{n\to\infty} \lambda_{1,k}^{-1/2} k^{1/2} \delta_n \log n = 0 \tag{3.17}$$
$$\lim_{n\to\infty} \lambda_{1,k}^{-1} k^{3/2} n^{-1/2} = 0. \tag{3.18}$$

Then under $H0$

$$\frac{n\chi_{n,k}^2 - k}{\sqrt{2k}} \xrightarrow{d} N(0,1).$$

Proof. By Proposition 3.1

$$\begin{aligned}\frac{n\chi_{n,k}^2 - k}{\sqrt{2k}} &= \frac{\left(\underline{B}_{n,k}^0\right)' S_k^{-1} \underline{B}_{n,k}^0 - k}{\sqrt{2k}} + 2(2k)^{-1/2} \left(\underline{B}_{n,k}^0\right)' S_k^{-1} \left(\underline{\gamma}_{n,k} - \underline{B}_{n,k}^0\right) \\ &+ (2k)^{-1/2} \left(\underline{\gamma}_{n,k} - \underline{B}_{n,k}^0\right)' S_k^{-1} \left(\underline{\gamma}_{n,k} - \underline{B}_{n,k}^0\right) \\ &+ (2k)^{-1/2} \underline{\gamma}_{n,k}{}' \left(S_{n,k}^{-1} - S_k^{-1}\right) \underline{\gamma}_{n,k} \\ &= A + B + C + D.\end{aligned}$$

The first term above can be written

$$A = \frac{\left(\underline{B}_{n,k}^0\right)' S_k^{-1} \underline{B}_{n,k}^0 - k}{\sqrt{2k}} = \frac{\sum_{i=1}^k \left(Z_i^2 - EZ_i^2\right)}{\sqrt{kVarZ_i^2}}$$

which converges weakly to the standard normal distribution by the CLT applied to the i.i.d.standard normal r.v's Z_i.

As to the term C it is straightforward that $C = o(B)$. From the proof of Theorem 3.3, D goes to zero if $\lambda_{1,k}^{-1} k^{1/2} \left(\sup_{i,j} |s_{n,k}(i,j) - s_k(i,j)|\right) \left\|\underline{\gamma}'_{n,k} S_k^{-1/2}\right\|^2 = o_P(1)$. Since, using (3.18) and (3.10), $\sup_{i,j} |s_{n,k}(i,j) - s_k(i,j)| \leq \sup_{f,g\in\mathcal{F}} |(P_n - P) fg|$ $+ \sup_{f\in\mathcal{F}} |(P_n - P) f| |(P_n + P) F| = O_P\left(n^{-1/2}\right)$, and considering that $\left\|\underline{\gamma}'_{n,k} S_k^{-1/2}\right\|^2 = O_P(k)$, we are done.

For B, $\left|\left(\underline{B}_{n,k}^0\right)' S_k^{-1} \left(\underline{\gamma}_{n,k} - \underline{B}_{n,k}^0\right)\right| = \left|\left(\underline{B}_{n,k}^0\right)' S_k^{-1/2} S_k^{-1/2} \left(\underline{\gamma}_{n,k} - \underline{B}_{n,k}^0\right)\right|$ $\leq \left\|\left(\underline{B}_{n,k}^0\right)' S_k^{-1/2}\right\| \left\|S_k^{-1/2} \left(\underline{\gamma}_{n,k} - \underline{B}_{n,k}^0\right)\right\| = \sqrt{\sum_{i=1}^k Z_i^2} \left\|S_k^{-1/2} \left(\underline{\gamma}_{n,k} - \underline{B}_{n,k}^0\right)\right\|$ where, as used in A, Z_i^2 are i.i.d. with a χ^2 distribution with 1 df. Hence $\sqrt{\sum_{i=1}^k Z_i^2} = O_P\left(k^{1/2}\right)$. Further

$$\begin{aligned}\left\|S_k^{-1/2} \left(\underline{\gamma}_{n,k} - \underline{B}_{n,k}^0\right)\right\| &\leq \left\|S_k^{-1/2}\right\| \cdot \left\|\underline{\gamma}_{n,k} - \underline{B}_{n,k}^0\right\| \\ &\leq \lambda_{1,k}^{-1/2} k^{1/2} \sqrt{\frac{1}{k} \sum_{i=1}^k \left(\underline{\gamma}_{n,k}(f_i) - \underline{B}_{n,k}^0(f_i)\right)^2} \\ &\leq \lambda_{1,k}^{-1/2} k^{1/2} \sup_{f\in\mathcal{F}} \left|\gamma_n(f) - B_n^0(f)\right|\end{aligned}$$

from which $B = O_P\left(\lambda_{1,k}^{-1/2} k^{1/2}\delta_n \log n\right) = o_P(1)$ if (3.17) holds. We have used the fact that P belongs to Ω in the last evaluation of B. □

Remark 3.8. Under $H1$, using the relation $\underline{\nu}_n = \underline{\nu} - n^{-1/2}\underline{\gamma}_{n,k}$, we can write

$$\begin{aligned}\frac{n\chi^2_{n,k} - k}{\sqrt{2k}} &= (2k)^{-1/2}\left(\underline{\gamma}'_{n,k} S^{-1}_{n,k}\underline{\gamma}_{n,k} - k\right) + (2k)^{-1/2}\left(n\underline{\nu}' S^{-1}_{n,k}\underline{\nu} - 2\sqrt{n}\underline{\gamma}'_{n,k} S^{-1}_{n,k}\underline{\nu}\right)\\ &= (2k)^{-1/2}\left(n\underline{\nu}' S^{-1}_{n,k}\underline{\nu} - 2\sqrt{n}\underline{\gamma}'_{n,k} S^{-1}_{n,k}\underline{\nu}\right) + O_P(1)\end{aligned}$$

where the $O_P(1)$ term captures $(2k)^{-1/2}\left(\underline{\gamma}'_{n,k} S^{-1}_{n,k}\underline{\gamma}_{n,k} - k\right)$ that coincides with the test statistic $\frac{n\chi^2_{n,k}-k}{\sqrt{2k}}$ under $H0$. We can bound the first term from below by

$$\begin{aligned}&(2k)^{-1/2} n\underline{\nu}' S_k^{-1}\underline{\nu}\left(1 - O_P\left(\left\|S_k^{1/2} S^{-1}_{n,k} S_k^{1/2} - I\right\|\right)\right)\\ &\quad - (2n/k)^{1/2}\left\|\underline{\gamma}'_{n,k} S_k^{-1/2}\right\|\left\|\underline{\nu}' S_k^{-1/2}\right\|\left(1 + O_P\left(\left\|S_k^{1/2} S^{-1}_{n,k} S_k^{1/2} - I\right\|\right)\right)\\ &= O_P\left(nk^{-1/2}\right) - O_P(n^{1/2}).\end{aligned}$$

Hence, if (3.17) and (3.18) are satisfied then the test statistic is asymptotically consistent also for the case of an infinite number of linear constraints.

In both conditions (3.17) and (3.18) the value of $\lambda_{1,k}$ appears, which cannot be estimated without any further hypothesis on the structure of the class $\mathcal{F}$. However, for concrete problems, once defined $\mathcal{F}$ it is possible to give bounds for λ_1, depending on k. This is what will be shown in the last section, for a particular class of goodness of fit tests.

3.3. **Application: testing marginal distributions.** Let P be an unknown distribution on $\mathbb{R}^d$ with density bounded by below. We consider goodness-of-fit tests for the marginal distributions $P_1, ..., P_d$ of P on the basis of an i.i.d. sample $(X_1, ..., X_n)$.

Let thus $Q_1^0, \ldots, Q_d^0$ denote d distributions on $\mathbb{R}$. The null hypothesis writes $H0$: $P_j = Q_j^0$ for $j = 1, ..., d$. That is to say that we simultaneously test goodness-of-fit of the marginal laws $P_1, \ldots, P_d$ to the laws $Q_1^0, \ldots, Q_d^0$. Through the transform $P'(y_1, \ldots, y_d) = P\left(\left(Q_1^0\right)^{-1}(y_1), \ldots, \left(Q_d^0\right)^{-1}(y_d)\right)$ we can restrict the analysis to the case when all p.m's have support $[0,1]^d$ and marginal laws uniform in $[0,1]$ under H0. So without loss of generality we write Q_0 for the uniform distribution on $[0,1]$.

P.J. Bickel, Y. Ritov and J.A. Wellner [3] focused on the estimation of linear functionals of the probability measure subject to the knowledge of the marginal laws in the case of r.v.'s with a.c. distribution, letting the number of cells grow to infinity.

Define the class

$$\mathcal{F} := \left\{1_{u,j} : [0,1]^d \longrightarrow \{0,1\},\ j = 1, \ldots, d,\ \ u \in [0,1]\right\}$$

where $1_{u,j}(x_1, \ldots, x_d) = \begin{cases} 1, & x_j \le u \\ 0, & x_j > u \end{cases}$.

Let Ω be the set of all p.m's on $[0,1]^d$ with uniform marginals, i.e

$$\Omega = \left\{ Q \in M_1\left([0,1]^d\right) \quad \text{such that} \quad Qf = \int_{[0,1]^d} f(x)dx\,, \quad f \in \mathcal{F} \right\}. \tag{3.19}$$

This set Ω has the form of (3.1) , where $\mathcal{F}$ is the class of characteristic functions of intervals, which is a KMT class with rate $\delta_n = n^{-1/2d}$ ($\delta_n = \sqrt{n}$ if $d = 2$) ; see [6].

We now build the family $\mathcal{F}_n$ satisfying (3.10) and (3.11).

Let $m = m(n)$ tend to $+\infty$ with n. Let $0 < u_1 < \ldots < u_m < 1$ and $\left\{\mathcal{U}^{(n)}\right\}$ be the $m \cdot d$ points in $[0,1]^d$ with coordinates in $\{u_1, \ldots, u_m\}$.

Let $\mathcal{F}_n$ denote the class of characteristic functions of the $d-$dimensional rectangles $[\underline{0}, \underline{u}]$ for $\underline{u} \in \mathcal{U}^{(n)}$. Hence $card\{\mathcal{F}_n\} = k = m \cdot d$.

Namely,

$$\mathcal{F}_n = \left\{ 1_{u_i,j} : [0,1]^d \to \{0,1\}\,,\ j = 1, \ldots, d,\ \ u_i \in (0,1)\,,\ u_i < u_{i+1}\,,\ i = 1, \ldots, m \right\}, \tag{3.20}$$

which satisfies $\mathcal{F}_n \subseteq \mathcal{F}_{n+1}$ for all $n \geq 1$ (i.e. (3.10)) and $\mathcal{F} = \bigcup_{n\geq 1} \mathcal{F}_n$ (i.e. (3.11)).

The sequence $\{\mathcal{F}_n\}_{n\geq 1}$ and the class $\mathcal{F}$ satisfy conditions of Theorem 3.7: $\mathcal{F}$ (and consequently each $\mathcal{F}_n$) has envelope function $F = 1$ and $RF^h = 1$, for all R in $M_1\left([0,1]^d\right)$ and h in $\mathbb{R}$.

In order to establish a lower bound for $\lambda_{1,k}$, the smallest eigenvalue of S_k, we will impose that the volumes of the cells in the grid defined by the $u_i^{(n)}$ do not shrink too rapidly to 0. Suppose that the intervals $(u_i, u_{i+1}]$ are such that

$$0 < \liminf_{n\to\infty} \min_{i=1,\ldots,m-1} k\,(u_{i+1} - u_i) \leq \limsup_{n\to\infty} \max_{i=1,\ldots,m-1} k\,(u_{i+1} - u_i) < \infty. \tag{3.21}$$

Remark 3.9. Condition for the sequence $\mathcal{F}_n$ to converge to $\mathcal{F}$ coincides with (F2) and (F3) in [3].

We first obtain an estimate for the eigenvalue $\lambda_{1,k}$. The final result of this step is stated in Lemma 3.11 below.

Let P belong to Ω. Let us then write the matrix S_k. We have $P1_{u_i,j} = Q_0 1_{u_i,j} = u_i$ for $i = 1, ..., m$ and $j = 1, ..., d$. Set $P1_{u_i,j}1_{u_l,h} = P\,(X_j \leq u_i,\, X_h \leq u_l)$, for every $h, j = 1, ..., d$ and $l, i = 1, \ldots, m$. When $j = h$ then $P1_{u_i,j}1_{u_l,j} = P\,(X_j \leq u_i \wedge u_l) = u_i \wedge u_l$.

Consider for the vector of functions f_j the following ordering

$$\left(f_1, \ldots, f_m, f_{m+1}, \ldots, f_{2m}, \ldots, f_{(d-1)m+1}, \ldots, f_{dm},\right) = \left(1_{u_1,1} \ldots, 1_{u_m,1}, 1_{u_1,2}, \ldots, 1_{u_m,d}\right).$$

The generic term of S_k writes

$$s_k\,(u,v) = s_k\,((j-1)m + i, (h-1)m + l) = P1_{u_i,j}1_{u_l,h} - P1_{u_i,j}P1_{u_l,h}$$

$$= \begin{cases} u_i - u_i^2 \quad , & \text{if } \ j = h,\ i = l \\ P(X_j \leq u_i \wedge u_l) - u_i u_l \quad , & \text{if } \ j = h,\ i \neq l \\ P(X_j \leq u_i, X_h \leq u_l) - u_i u_l\,(i,l) - p_i p_l & \text{if } \ j \neq h \end{cases}$$

We make use of the class of functions

$$\begin{aligned}\mathcal{F}_n^\delta &= \left\{ f_{j\cdot i} - f_{j(i-1)},\ i = 1, \dots, m,\ j = 1 \dots, d,\ \ f_h \in \mathcal{F}_n,\ f_0 = 0 \right\} \\ &= \left\{ 1_{A_i^j},\ i = 1 \dots, m,\ j = 1, \dots, d \right\}.\end{aligned}$$

In the above display the $\{A_i^j\}_{i=1,\dots,m}$ describe the partition of $[0,1]$, the support of the marginal distribution P_j, induced by the vector $\{u_i\}_{i \le m}$. Namely we have for every $j = 1, \dots, d$, $A_i^j \cap A_l^j = \emptyset$, $i \neq l$, $\cup_{i=1}^{m+1} A_i^j = [0,1]$, with $A_{m+1}^j = [u_m, 1]$, for all j.

Set S_k^δ the covariance matrix of the vector $\underline{\gamma}_n^\delta = \underline{\gamma}_n(\mathcal{F}^\delta)$ and consider the vectors $\underline{\nu}^\delta$ and $\underline{\nu}_n^\delta$ defined as in (3.4).

S_k^δ has $((j-1)m + i, (h-1)m + l)-$th component equal to $P_{A_i^j A_l^h} - P_{A_i^j} P_{A_l^h}$, which is

$$\begin{cases} p_i - p_i^2, & \text{if } \ j = h,\ i = l \\ -p_i p_l, & \text{if } \ j = h,\ i \neq l \\ P(u_{i-1} \le X_j < u_i, u_{l-1} \le X_h < u_l) - p_i p_l, & \text{if } \ j \neq h \end{cases}$$

where we have written $p_i = P(u_{i-1} \le X_j < u_i) = u_i - u_{i-1}$, for all $j = 1, \dots, d$.

χ^2 (and χ_n^2) can be written using $\mathcal{F}_n^\delta$ instead of $\mathcal{F}_n$:

$$\chi^2(\Omega, P) = \underline{\nu}' S_k^{-1} \underline{\nu} = \left(\underline{\nu}^\delta\right)' \left(S_k^\delta\right)^{-1} \left(\underline{\nu}^\delta\right).$$

Let M be the diagonal $d-$block matrix with all diagonal blocks equal to the unit inferior triangular $(m \times m)$ matrix. Then $\underline{\nu} = M \underline{\nu}^\delta$.

On the other hand, after some algebra it can be checked that $S_k = M S_k^\delta M'$.

Thus $\underline{\nu}^{\delta\prime} (S_k^\delta)^{-1} \underline{\nu}^\delta = \underline{\nu}' (M')^{-1} M' S_k^{-1} M (M)^{-1} \underline{\nu} = \chi^2$. Similar arguments yield $\chi_n^2 = \underline{\nu}_n^{\delta\prime} (S_{n,k}^\delta)^{-1} \underline{\nu}_n^\delta$.

The matrix M has all eigenvalues equal to one. This allows us to write, for $\lambda_{1,\delta}$ the minimum eigenvalue of S_k^δ:

$$\lambda_{1,k} \le \min_x \frac{x' S_k x}{\|x\|^2} \le \min_y \frac{y' S_k^\delta y}{\|y\|^2} \max_x \frac{\|Mx\|^2}{\|x\|^2} = \lambda_{1,\delta} \le \min_y \frac{y' S_k y}{\|y\|^2} \max_x \frac{\|M^{-1} x\|^2}{\|x\|^2} = \lambda_{1,k}.$$

We will now consider the covariance matrix of $\underline{\gamma}_n^\delta$ under H0, when the underlying distribution is Q_0, i.e. the uniform distribution on $[0,1]^d$. Denote this matrix S_k^0. We then have

Lemma 3.10. *If $P \in \Omega$, then*

(i) $S_k^0 = D^{1/2}(I - V) D^{1/2}$, where D and V are both diagonal block matrices with diagonal blocks equal to diag $\{p_i\}_{i=1,\dots,m}$ and to $U = \left\{\sqrt{p_i p_l}\right\}_{i=1,\dots,m,\, l=1,\dots,m}$ respectively.

(ii) The $(m \times m)$ matrix U has eigenvalues equal to

$$\lambda_U = \begin{cases} (1 - p_{m+1}) = \sum_{i=1}^m p_i & \text{with cardinality } 1 \\ 0 & \text{with cardinality } m - 1. \end{cases}$$

Moreover $(I - U)^{-1} = (I + \frac{1}{p_{m+1}} U)$.

(iii) For any eigenvalue λ of S_k^0 it holds

$$p_{m+1} \min_{1\le i\le m} p_i \le \lambda \le \max_{1\le i\le m} p_i.$$

Proof. (i) This can easily be checked through some calculation.

(ii) First notice that

$$U^2 = (1-p_{m+1})U \tag{3.22}$$

Formula (3.22) implies that at least one eigenvalue equals $(1-p_{m+1})$. On the other hand, summing up all diagonal entries in U we get $trace(U) = \sum_{i=1}^m p_i = 1-p_{m+1}$. This allows us to conclude that there can be only one eigenvalue equal to $1-p_{m+1}$ while the other must be zero.

For the second statement, by Taylor expansion of $(1-x)^{-1}$, $(I-U)^{-1} = I + \sum_{h=1}^\infty U^h$.

Then, using recursively (3.22), $(I-U)^{-1} = I+U\sum_{h=1}^\infty (1-p_{m+1})^h = U+\frac{1}{p_{m+1}}U$.

(iii) For any eigenvalue λ of S_k^0 we have:

$$\lambda \le \lambda_{k,k} = \left|\!\left|\!\left| S_k^0 \right|\!\right|\!\right| \le \left|\!\left|\!\left| D^{1/2} \right|\!\right|\!\right|^2 \left|\!\left|\!\left| (I-V) \right|\!\right|\!\right| = \max_{1\le i\le m} p_i \left(1 - \inf_{\|\underline{x}\|=1} \underline{x}' V \underline{x}\right) = \max_{1\le i\le m} p_i$$

where for the last identity we have used the fact that the eigenvalues of V coincide with the eigenvalues of U with order multiplied by d.

For the opposite inequality consider

$$\begin{aligned}\lambda^{-1} \le \lambda_{1,k}^{-1} &= \left|\!\left|\!\left| S_k^{0\,-1} \right|\!\right|\!\right| \le \left|\!\left|\!\left| D^{-1} \right|\!\right|\!\right| \left|\!\left|\!\left| \left(I + \frac{1}{p_{m+1}} V\right) \right|\!\right|\!\right| \\ &= \left(\max_{1\le i\le m} p_i^{-1}\right)\left(1 + \frac{1}{p_{m+1}}(1-p_{m+1})\right) = \left(\min_{1\le i\le m} p_i\right)^{-1} p_{m+1}^{-1}.\end{aligned}$$

□

Lemma 3.11. *Suppose that P has density on $[0,1]^d$ bounded from below by $\alpha > 0$. Then the smallest eigenvalue of S_k^δ, and consequently $\lambda_{1,k}$, is bounded below by $p_{m+1} \min_{1\le i\le m} p_i$.*

Proof. Write $s_k^\delta(u,v)$ for the (u,v) −th element of S_k^δ. We have, for $P \in \Omega$, i.e. if $Pf = Q_0 f$, for every $f \in \mathcal{F}_n^\delta$:

$$\begin{aligned}s_k^\delta((j-1)m+i,(h-1)m+l) = s_k^\delta(u,v) &= Pf_u f_v - Pf_u Pf_v = \\ &= P(f_u - Q_0 f_u)(f_v - Q_0 f_v) = P\left(\overline{f_u f_v}\right)\end{aligned}$$

where $\overline{f_u} = f_u - Q_0 f_u$. For each vector $\underline{a} \in \mathbb{R}^{d\cdot m}$ it holds then

$$\begin{aligned}\underline{a}' S_k^\delta \underline{a} = \sum_{u=1}^{dm}\sum_{v=1}^{dm} a_u a_v P\left(\overline{f_u f_v}\right) &= P\left(\sum_{u=1}^{dm} a_u \overline{f_u}\right)^2 \\ = \int_{[0,1]^d} \left(\sum_{u=1}^{dm} a_u \overline{f_u}\right)^2 dP \ge \alpha \int_{[0,1]^d} \left(\sum_{u=1}^{dm} a_u \overline{f_u}\right)^2 dQ_0 &= \\ = \alpha\, \underline{a}' \left\{Q_0(f_u - Q_0 f_u)(f_v - Q_0 f_v)\right\}_{u,v} \underline{a} &= \alpha\, \underline{a}' S_k^0 \underline{a}\end{aligned}$$

On the other hand the preceding inequality implies

$$\inf_{\underline{a}} \frac{\underline{a}' S_k^\delta \underline{a}}{\|\underline{a}\|^2} \geq \alpha \inf_{\underline{a}} \frac{\underline{a}' S_k^0 \underline{a}}{\|\underline{a}\|^2} \tag{3.23}$$

that is a lower bound for the smallest eigenvalue of S_k^δ depending on the smallest eigenvalue of S_k^0.

Apply Lemma 3.10 (iii) to get the lower bound for λ_1. □

Remark 3.12. Existence of $\alpha > 0$ such that the density of P in $[0,1]^d$ is bounded below by α seems necessary for this kind of approach; see assumption (P3) in [3].

From Theorem 3.7 and using (3.21) in order to evaluate $p_{m+1} \min_{1\leq i\leq m} p_i$, together with the fact that the class $\mathcal{F}$ is KMT with rate $\delta_n = n^{-1/2}$ we obtain

Theorem 3.13. *Let (3.21) hold. Assume that P belongs to Ω defined by (3.19) and has a density bounded below by some positive number. Let further $k = d \cdot m(n)$ be a sequence such that $\lim_{n\to\infty} k = \infty$ and $\lim_{n\to\infty} k^{7/2} n^{-1/2} = 0$*

Then $\frac{n\chi^2_{n,k} - k}{\sqrt{2k}} = \frac{n\underline{\gamma}'_{n,k} S^{-1}_{n,k} \underline{\gamma}_{n,k} - k}{\sqrt{2k}}$ has limiting normal standard distribution.

In the last part of this Section we intend to show that conditions in Theorem 3.13 can be weakened for small values of d. When $d = 2$ the rate for $k = 2m$ is achieved when condition (3.17) holds.

We consider the case when $d = 2$; for larger values of d, see Remark 3.16.

In order to make the notation more clear, define $p_{i,j}$ and $N_{i,j}$, respectively, $P(A_i^1 \times A_j^2)$ and $nP_n(A_i^1 \times A_j^2)$, where the events A_i^h, $h = 1, 2$, $i = 1, \ldots, m$ are as above. The marginal distributions will be denoted $p_{i,\cdot} = p_{\cdot,i} = p_i$ (since H0 holds), and the empirical marginal distributions by $N_{i,\cdot}/n$ and $N_{\cdot,i}/n$.

Turning back to the proof of Theorem 3.7 we see that condition (3.18) is used in order to ensure that $\underline{\gamma}'_{n,k}(S^{-1}_{n,k} - S^{-1}_k)\underline{\gamma}_{n,k}$ goes to 0 in probability as n tends to infinity, while condition (3.17) implies the convergence of $\frac{\underline{\gamma}'_{n,k} S_k^{-1} \underline{\gamma}_{n,k} - 2m}{\sqrt{4m}}$ to the standard normal distribution.

Let

$$\mathcal{Q} = \Bigg\{ Q \in M_1([0,1]^2) : \sum_{i=1}^{m+1} q_{i,j} = p_{\cdot,j} = q^0_{\cdot,j} = u_{j+1} - u_j,\ j = 1, \ldots, m+1; \\ \sum_{j=1}^{m+1} q_{i,j} = p_{i,\cdot} = q^0_{i,\cdot} = u_{i+1} - u_i,\ i = 1, \ldots, m+1 \Bigg\},$$

where $q^0_{i,j} = Q^0(A_i^1 \times A_j^2) = (u_{i+1} - u_i)(u_{j+1} - u_j)$.

Lemma 3.14. *When $P \in \Omega$, it holds*

$$n\chi^2_{n,k} = \min_{Q\in\mathcal{Q}} \sum_{i=1}^{m+1} \sum_{j=1}^{m+1} \frac{(nq_{i,j} - N_{i,j})^2}{N_{i,j}} \mathbb{I}_{N_{i,j}>0} \tag{3.24}$$

$$\underline{\gamma}'_{n,k} S_k^{-1} \underline{\gamma}_{n,k} = \min_{Q\in\mathcal{Q}} \sum_{i=1}^{m+1} \sum_{j=1}^{m+1} \frac{(nq_{i,j} - N_{i,j})^2}{np_{i,j}} \tag{3.25}$$

Proof. We prove (3.24), since the proof of (3.25) is similar. Following [3] the RHS in (3.24) is

$$\sum_{i=1}^{m+1}\sum_{j=1}^{m+1} N_{i,j}(a_i+b_j)^2,$$

where the vectors a and $b \in \mathbb{R}^{m+1}$ are solutions of the equations

$$a_i\frac{N_{i,\cdot}}{n} = p_i - \frac{N_{i,\cdot}}{n} - \sum_{j=1}^{m+1} b_j\frac{N_{i,j}}{n}, \quad i=1,\dots,m+1,$$

$$b_j\frac{N_{\cdot,j}}{n} = p_j - \frac{N_{\cdot,j}}{n} - \sum_{i=1}^{m+1} a_i\frac{N_{i,j}}{n}, \quad j=1,\dots,m+1.$$

Let $\underline{a} = (\tilde{a}_1,\dots,\tilde{a}_m,\tilde{b}_1,\dots,\tilde{b}_m)$ be the coefficients in equation (3.7). Making use of equations (3.5) and (3.6) we obtain, using the class $\mathcal{F}_n^\delta$ in place of $\mathcal{F}_n$ in the definition of $\chi^2_{n,k}$,

$$\begin{aligned}
\widetilde{a}_i &= 2\left(a_i - a_{m+1}\right),\ i=1,\dots,m \\
\widetilde{b}_j &= 2\left(b_j - b_{m+1}\right),\ j=1,\dots,m \\
\widetilde{a}_0 &= 2\left(a_{m+1}+b_{m+1}\right).
\end{aligned} \tag{3.26}$$

From the proof of Proposition 3.1 we get, setting $\delta_{i,j}=1$ for $i=j$ and 0 otherwise,

$$\begin{aligned}
\chi^2_{n,k} &= \frac{1}{4}\underline{a}' S_{n,k}\underline{a} \\
&= \frac{1}{4n}\sum_{i=1}^{m}\sum_{j=1}^{m}\Big[\tilde{a}_i\tilde{a}_j\left(N_{i,\cdot}\delta_{i,j} - N_{i,\cdot}N_{j,\cdot}/n\right) + \tilde{b}_i\tilde{b}_j\left(N_{\cdot,i}\delta_{i,j} - N_{\cdot,i}N_{\cdot,j}/n\right) \\
&\qquad +2\tilde{a}_i\tilde{b}_j\left(N_{i,j} - N_{i\cdot}N_{\cdot,j}/n\right)\Big],
\end{aligned}$$

which, using (3.26) and after some algebra yields

$$n\chi^2_{n,k} = \sum_{i=1}^{m+1}\sum_{j=1}^{m+1} N_{i,j}(a_i+b_j)^2.$$

□

We now can refine Theorem 3.13.

Theorem 3.15. *Let (3.21) hold. Assume that $P \in \Omega$ satisfies the condition in Lemma 3.11 for some $\alpha > 0$.*

Let $m(n)$ be such that $\lim_{n\to\infty} m = \infty$ and $\lim_{n\to\infty} m^{3/2}n^{-1/2}\log n = 0$.

Then, under $H0$,

$$\frac{n\chi^2_{n,k} - 2m}{\sqrt{4m}} \longrightarrow N(0,1).$$

Proof. It is enough to prove $\frac{n\chi^2_{n,k} - \underline{\gamma}'_{n,k}S_k^{-1}\underline{\gamma}_{n,k}}{\sqrt{4m}} = o_P(1)$.

Denote $\hat{P}$ and $\overline{P}$ the minimizers of (3.24) and (3.25) in $\mathcal{Q}$. Let $\hat{p}_{i,j}$ and $\overline{p}_{i,j}$ denote the respective probabilities of cells.

We write

$$n\chi^2_{n,k} - \underline{\gamma}'_{n,k} S_k^{-1} \underline{\gamma}_{n,k} \leq \sum_{i=1}^{m+1}\sum_{j=1}^{m+1} \left(N_{i,j} - n\overline{p}_{i,j}\right)^2 \left(\frac{1}{N_{i,j}} - \frac{1}{np_{i,j}}\right)$$
$$\leq \max_{i,j}\left(\frac{np_{i,j}}{N_{i,j}} - 1\right) n\underline{\gamma}'_{n,k} S_k^{-1} \underline{\gamma}_{n,k}$$

and

$$n\chi^2_{n,k} - \underline{\gamma}'_{n,k} S_k^{-1} \underline{\gamma}_{n,k} \geq \min_{Q\in\mathcal{Q}} \sum_{i=1}^{m+1}\sum_{j=1}^{m+1} \frac{\left(N_{i,j} - nq_{i,j}\right)^2}{np_{i,j}} \left(\frac{np_{i,j}}{N_{i,j}} - 1\right)$$
$$\geq -\max_{i,j}\left|\frac{np_{i,j}}{N_{i,j}} - 1\right| \underline{\gamma}'_{n,k} S_k^{-1} \underline{\gamma}_{n,k}.$$

Whenever

$$\sqrt{m}\max_{i,j}\left|\frac{np_{i,j}}{N_{i,j}} - 1\right| \xrightarrow{P} 0 \tag{3.27}$$

holds, then the above inequalities yield $\frac{n\chi^2_{n,k} - \underline{\gamma}'_{n,k} S_k^{-1} \underline{\gamma}_{n,k}}{\sqrt{4m}} = o_P\left(\frac{\underline{\gamma}'_{n,k} S_k^{-1} \underline{\gamma}_{n,k}}{m}\right) = o_P\left(\frac{\underline{\gamma}'_{n,k} S_k^{-1} \underline{\gamma}_{n,k} - 2m}{\sqrt{4m}} \frac{2}{\sqrt{m}} + 1\right) = o_P(1)$, which proves the claim.
We now prove (3.27). We proceed as in Lemma 2 in [3], using inequalities (10.3.2) in [26]. Let $B_n \sim Bin(n,p)$. Then, for $t > 1$,

$$\Pr\left(\frac{np}{B_n} \geq t\right) \leq \exp\left\{-np\, h\left(1/t\right)\right\} \quad \text{and} \quad \Pr\left(\frac{B_n}{np} \geq t\right) \leq \exp\left\{-np\, h\left(t\right)\right\}, \tag{3.28}$$

where $h(t) = t\log t - t + 1$ is a positive function.

Since $N_{i,j} \sim Bin(n, p_{i,j})$,

$$Pr\left\{\max_{i,j}\left(\frac{np_{i,j}}{N_{i,j}} - 1\right) \geq \frac{t}{\sqrt{m}}\right\} \leq \sum_{i=1}^{m+1}\sum_{j=1}^{m+1} Pr\left\{\frac{np_{i,j}}{N_{i,j}} \geq \frac{t}{\sqrt{m}} + 1\right\}$$
$$\leq \sum_{i=1}^{m+1}\sum_{j=1}^{m+1} \exp\left\{-np_{i,j} h\left(1/\left(1 + t/\sqrt{m}\right)\right)\right\}$$

(by (3.21) and by$p_{i,j} > \alpha p_{i,.} p_{.,j}$) $\leq (m+1)^2 \exp\left\{-c\alpha \frac{n}{\log n}(\log n) m^{-2} h\left(1/\left(1 + t/\sqrt{m}\right)\right)\right\}$.

For $x = 1 + \varepsilon$, $h(x) = O(\varepsilon^2)$. Therefore, using (3.17) with $k = 2m$, for every $M > 0$ there exists n large enough that

$$\alpha c \frac{n}{\log n} m^{-2}\, h\left(1 + \frac{-t/\sqrt{m}}{1 + t/\sqrt{m}}\right) \geq M$$

and consequently $Pr\left\{\max_{i,j}\left(\frac{np_{i,j}}{N_{i,j}} - 1\right) \geq \frac{t}{\sqrt{m}}\right\}$ goes to 0.

To get convergence to zero of $Pr\left\{\max_{i,j}\left(1 - \frac{np_{i,j}}{N_{i,j}}\right) \geq \frac{t}{\sqrt{m}}\right\} = Pr\left\{\max_{i,j}\frac{N_{i,j}}{np_{i,j}} \geq \frac{1}{1 - t/\sqrt{m}}\right\}$, the second inequality in (3.28) is used in a similar way. □

Remark 3.16. The preceding arguments carry over to the case $d > 2$ and yield to the condition $\lim_n m^{d+1/2} n^{-1/2} \log n = 0$. However for $d \geq 6$ this ultimate condition is stronger than (3.18).

4. Application: a contamination model

Let $\mathcal{P}_\theta$ an identifiable class of densities on $\mathbb{R}$. A contamination model typically writes

$$p(x) = (1-\lambda) f_\theta(x) + \lambda r(x) \tag{4.1}$$

where λ is supposed to be close to zero and $r(x)$ is a density on $\mathbb{R}$ which represents the distribution of the contaminating data.

An example is when $f_\theta(x) = \theta e^{-\theta x}$, $x > 0$ and $r(x)$ is a Pareto type distribution, say

$$r(x) := r_{\gamma,\nu}(x) = \gamma \nu^\gamma (x)^{-(\gamma+1)}, \tag{4.2}$$

with $x > \nu$ and $\gamma > 1,\ \nu > 1$.

Such a case corresponds to a proportion λ of outliers generated by the density $r_{\gamma,\nu}$.

We test contamination when we have at hand a sample $X_1, \ldots, X_n$ of i.i.d. r.v.'s with unknown density function $p(x)$ as in (4.1). We state the test paradigm as follows.

Let $H0$ denote the composite null hypothesis $\lambda = 0$, i.e.

$$H0:\ p(x) = f_{\theta_0}(x),\ \theta_0 \in \Theta$$

versus

$$H1:\ p(x) = (1-\lambda) f_\theta(x) + \lambda r(x)$$

$$\text{for some } \theta \in \Theta \text{ and with } \lambda \neq 0.$$

Such problems have been addressed in the recent literature; see [19] and references therein. We assume identifiability, stating that, under $H1$, λ, θ and r are uniquely defined. This assumption holds for example when $f_\theta(x) = \theta e^{-\theta x}$ and $r(x)$ is like in (4.2).

For test problems pertaining to λ we embed $p(x)$ in the class of density functions of signed measures with total mass 1, allowing to belong to Λ_0 an open interval that contains 0.

In order to present the test statistic, we first consider a simplified version of the problem above.

Assume that $\theta_0 = \alpha$ is fixed, i.e. $\Theta = \{\alpha\}$. We consider the hypotheses

$$H0:\ p(x) = f_\alpha(x)$$

versus

$$H1:\ p(x) = (1-\lambda) f_\alpha(x) + \lambda r(x), \qquad \text{with } \lambda \neq 0.$$

In this case $\Omega = \{f_\alpha\}$ and the null hypothesis H0 is simple.

For this problem the χ^2 approach appears legitimate. From the discussion in Section 1 the χ^2 criterion is robust against inliers. A contamination model as (4.1) captures the outlier contamination through the density r. As such the test statistic does not need to have any robustness property against those, since they are included

in the model. At the contrary, missing data might lead to advocate in favour of $H1$ unduly. Therefore the test statistic should be robust versus such cases.

By the necessary inclusion $f^* = 2\left(\frac{q^*}{p} - 1\right) \in \mathcal{F}$ we define

$$\mathcal{F} = \mathcal{F}_\alpha = \left\{ g = 2\left(\frac{f_\alpha}{(1-\lambda)f_\alpha + \lambda r} - 1\right) \quad \text{such that} \quad \int |g| f_\alpha < \infty,\ \lambda \in \Lambda_0 \right\}. \tag{4.3}$$

Following (2.9)

$$\chi_n^2(f_\alpha, p) = \sup_{g \in \mathcal{F}_\alpha} \int g f_\alpha - T(g, P_n). \tag{4.4}$$

Example 4.1. Consider the case $f_\alpha(x) = \alpha e^{-\alpha x}$ and $r(x) = \gamma \nu^\gamma(x)^{-(\gamma+1)}$ for some γ fixed, $x > \nu$.

Then $\mathcal{F}_\alpha = \left\{ 2\left(\frac{\alpha e^{-\alpha x}}{(1-\lambda)\alpha e^{-\alpha x} + \lambda r_{\gamma,\nu}} - 1\right),\ \lambda \in \Lambda_0 \text{ such that } \int \frac{\alpha^2 e^{-2\alpha x}}{(1-\lambda)\alpha e^{-\alpha x} + \lambda r_{\gamma,\nu}(x)} dx < \infty \right\}$.

Let us now turn back to composite hypothesis.

Let Ω be defined by

$$\Omega = \{ q(x) = f_\alpha(x), \alpha \in \Theta \}.$$

We can write

$$\mathcal{F}_\alpha = \left\{ g(\theta, \lambda, \alpha) = 2\left(\frac{f_\alpha}{(1-\lambda)f_\theta + \lambda r} - 1\right) : \int |g| f_\alpha < \infty, \lambda \in \Lambda_0, \theta \in \Theta \right\}$$

and

$$\chi^2(\Omega, P) = \inf_{\alpha \in \Theta} \sup_{g \in \mathcal{F}_\alpha} \int g f_\alpha - T(g, P).$$

The supremum is to be found over a class of functions $\mathcal{F}_\alpha$ which changes with α.

Denote Δ_α the subset of (Θ, Λ_0) which parametrizes $\mathcal{F}_\alpha$.

Example 4.2 (Continued). We assume $\Theta = [\underline{\alpha}, \overline{\alpha}]$, which corresponds, in our example, to the restriction of the expected value of P (under $H0$) to the finite interval $[\frac{1}{\overline{\alpha}}, \frac{1}{\underline{\alpha}}]$.

Therefore

$$\chi_n^2(\Omega, P) = \inf_{\underline{\alpha} \le \alpha \le \overline{\alpha}} \sup_{(\theta,\lambda) \in \Delta_\alpha} \int 2\left(\frac{\alpha e^{-\alpha x}}{(1-\lambda)\theta e^{-\theta x} + \lambda r_\gamma(x)} - 1\right) \alpha e^{-\alpha x} dx - T(g(\theta, \lambda, \alpha); P_n). \tag{4.5}$$

The supremum in (4.5) is evaluated over a set which changes with α.

In accordance with the discussion in Section 2 we may define

$$\begin{aligned} \mathcal{F} &= \left\{ g(\theta, \lambda, \beta) = 2\left(\frac{\beta e^{-\beta x}}{(1-\lambda)\theta e^{\theta x} + \lambda r(x)} - 1\right) : \int \frac{\alpha\beta e^{-(\alpha+\beta)x}}{(1-\lambda)\theta e^{-\theta x} + \lambda r} dx < \infty, (\alpha, \theta, \beta) \in \Theta^3, \lambda \in \Lambda_0 \right\} \\ &\subseteq \{ g(\theta, \lambda, \beta) : \lambda \in \Lambda_0,\ (\theta, \beta) \in \Gamma \}, \end{aligned} \tag{4.6}$$

a class not depending upon α.

The resulting test statistic would be then

$$\chi_n^2(\Omega, P) = \inf_{\alpha \in \Theta} \sup_{(\theta,\beta) \in \Gamma,\ \lambda \in \Lambda_0} \int g(\theta, \lambda, \beta) \alpha e^{-\alpha x} dx - T(g(\theta, \lambda, \beta); P_n) \tag{4.7}$$

and the supremum in (4.7) is determined on a set that does not depend on α.

The use of (4.5) is proposed by M. Broniatowski and A. Keziou [9]. Also in our context it is easy to see that (4.5) is preferable to (4.7), in the sense that it reduces considerably the computational complexity of the problem, from a subset of $\{(\theta,\lambda,\beta)\in\Theta\times\Lambda_0\times\Theta\}$ to a subset of $\{(\lambda,\theta)\in\Lambda_0\times\Theta\}$.

We first derive the asymptotic distribution of the test statistic χ^2_n under $H1$; in order to use Theorem 2.6 we commute the inf and sup operators in (4.5) through the following Lemma 4.3.

Assume

(A1) Θ is compact.

(A2) For all α in Θ, Δ_α is compact.

Condition (A2) is verified in our example due to the compactness of the interval Θ and to the distribution of the outliers.

Lemma 4.3. *Let*

$$\Theta_{(\theta_1,\lambda,\theta_2)}=\{\alpha\in\Theta:\ (\theta_1,\lambda,\theta_2)\in\Delta_\alpha\}.$$

Under (A1) and (A2),

$$\begin{aligned}&\inf_{\alpha\in\Theta}\sup_{(\theta_1,\lambda,\theta_2)\in\Delta_\alpha}\int g(\theta_1,\lambda,\theta_2)f_\alpha-T(g(\theta_1,\lambda,\theta_2);P) \qquad (4.8)\\ &\quad=\sup_{(\theta_1,\theta_2,\lambda)\in\Theta^2\times\Lambda_0}\inf_{\alpha\in\Theta_{(\theta_1,\lambda\theta_2)}}\int g(\theta_1,\lambda,\theta_2)f_\alpha-T(g(\theta_1,\lambda,\theta_2);P).\end{aligned}$$

Proof. For $\Theta_{(\theta_1,\lambda\theta_2)}$ defined as above we have

$$\begin{aligned}&\inf_{\alpha\in\Theta}\sup_{(\theta_1,\lambda,\theta_2)\in\Delta_\alpha}\int g(\theta_1,\lambda,\theta_2)f_\alpha-T(g(\theta_1,\lambda,\theta_2);P) \qquad (4.9)\\ &\quad\le\sup_{(\theta_1,\theta_2,\lambda)\in\Theta^2\times\Lambda_0}\inf_{\alpha\in\Theta_{(\theta_1,\lambda\theta_2)}}\int g(\theta_1,\lambda,\theta_2)f_\alpha-T(g(\theta_1,\lambda,\theta_2);P).\end{aligned}$$

On the other hand,

$$\begin{aligned}\sup_{\theta_1,\lambda,\theta_2}\int gf_\alpha-T(g;P)&=\sup_{\theta_1,\lambda,\theta_2}\left\{\int 2\frac{f_{\theta_2}}{(1-\lambda)f_{\theta_1}+\lambda r}\frac{f_\alpha}{p}pdx-\int\left(\frac{f_{\theta_2}}{(1-\lambda)f_{\theta_1}+\lambda r}\right)^2pdx+1\right\}\\ &=\sup_{\theta_1,\lambda,\theta_2}-\int\left(\frac{f_{\theta_2}}{(1-\lambda)f_{\theta_1}+\lambda r}-\frac{f_\alpha}{p}\right)^2pdx+\int\left(\frac{f_\alpha}{p}-1\right)^2pdx\\ &\le\int\left(\frac{f_\alpha}{p}-1\right)^2pdx=\chi^2(f_\alpha,p)\end{aligned}$$

and equality holds if $(\theta_1,\theta_2,\lambda)$ are such that $\frac{f_\alpha}{p}=\frac{f_{\theta_2}}{(1-\lambda)f_{\theta_1}+\lambda r}$ (identifiability allows to find such $(\theta_1,\theta_2,\lambda)$ for every $\alpha\in\Theta$, and for every contaminated measure p).

Also we have

$$\begin{aligned}&\sup_{(\theta_1,\lambda,\theta_2)}\inf_{\alpha}\int gf_\alpha-T(g;P)\\ &\quad=\sup_{(\theta_1,\lambda,\theta_2)}\left\{-\int\left(\frac{f_{\theta_2}}{(1-\lambda)f_{\theta_1}+\lambda r}-\frac{f_{\alpha^*}}{p}\right)^2pdx+\int\left(\frac{f_{\alpha^*}}{p}-1\right)^2pdx\right\}\\ &\quad=\chi^2(f_{\alpha^*},p),\end{aligned}$$

for some α^* in $\Theta_{(\theta_1,\lambda\theta_2)}$.

We thus get

$$\sup_{(\theta_1,\lambda,\theta_2)} \inf_{\alpha} \int gf_\alpha - T(g;P) = \chi^2(f_{\alpha^*},p) \geq \chi^2(\Omega,p) = \inf_{\alpha} \sup_{(\theta_1,\lambda,\theta_2)} \int gf_\alpha - T(g;P),$$

which, by (4.9), concludes the proof. □

Theorem 2.6 implies consistency of χ^2_n as an estimator of χ^2 and convergence in distribution of $\sqrt{n}\left(\chi^2_n - \chi^2\right)$ to a normally distributed r.v. with mean zero and variance given by $P\left(\left(-g^* - \frac{1}{4}g^{*2}\right)^2\right) - \left(P\left(-g^* - \frac{1}{4}g^{*2}\right)\right)^2$, under $H1$.

The asymptotic distribution under the null hypothesis can be found subject to the choice of the parametric class $\{f_\alpha\}$ and of the density r, as can be deduced by Theorem 3.5 in [9]. Following their Theorem 3.5, which holds for composite hypothesis testing in a parametric environment, the test statistic $n\chi^2_n$ converges weakly, under $H0$, to a chi-squared distribution with degrees of freedom depending on the dimension of the parameter space Θ and on the cardinality of the constraints induced by $P \in \Omega$.

In the following, we focus on definition (4.5) for $\chi^2_n(\Omega, P)$.

The null hypothesis reduces the space $\Theta \times \Lambda_0$ to $\Theta \times \{0\}$.

Theorem 3.5 in [9] implies that the degree of freedom d of the limiting chi-squared distribution equals the number of parameters of P under $H0$. In the following we assume $d = 1$, as in Example 4.1.

Let $h(\theta, \lambda; x) = (1-\lambda)f_\theta(x) + \lambda r(x)$.

Checking conditions (C.12)-(C.15) in [9] yields:

Theorem 4.4. *Under $H0$, with $P = P_{\theta_0}$, assume that*

(i) *The class of contaminated densities $\{h(\theta,\lambda), \theta \in \Theta, \lambda \in \Lambda_0\}$ is $P_{\theta_0}-$identifiable;*

(ii) *The class of functions $\left\{\frac{h(\alpha,\nu)}{h(\theta,\lambda)}, \theta \in \Theta_\alpha, \lambda \in \Lambda_0, \alpha \in \Theta, |\nu| < \varepsilon\right\}$ is $P_{\theta_0}-GC$ for some ε small enough;*

(iii) *The densities f_θ are differentiable up to the second order in some neighborhood $V(\theta_0)$ of θ_0 and $F_\theta(x) = \int_{-\infty}^{x} f_\theta(u)du$ is differentiable with respect to θ;*

(iv) *There exists a neighborhood V of $(\theta_0, 0, \theta_0, 0)$ such that, for every $(\theta, \lambda, \alpha, \nu) \in V$ we have*

$$\frac{f_\alpha}{h(\theta,\lambda)} \leq H_1(x), \qquad \frac{\ddot{f}_\alpha}{h(\theta,\lambda)} \leq H_3(x),$$
$$\frac{\dot{f}_\alpha}{h(\theta,\lambda)} \leq H_2(x), \qquad \frac{r}{h(\theta,\lambda)} \leq H_4(x),$$

where each of the functions H_j $(j = 1,2,3,4)$ is square integrable w.r. to the density $h(\alpha,\nu)$ and is in $L_4(P_{\theta_0})$.

Then, $n\chi^2_n$ converges to a chi-squared distributed r.v. with degree of freedom equal to 1.

Acknowledgements

This work was supported by *Progetto Ateneo di Padova* coordinated by Prof. G. Celant.

References

[1] D. Aze. *Elements D' Analyse Convexe et Variationnelle*. Ellipses, 1997.

[2] R. Beran. Minimum hellinger distance estimates for parametric models. *Annals of Statistics*, 5:445–463, 1977.

[3] P.J. Bickel, Y. Ritov, and J.A Wellner. Efficient estimation of linear functionals of a probability measure p with known marginal distributions. *Ann. Stat.*, 19:1316–1346, 1991.

[4] I.S. Borisov. On the accuracy of the approximation of empirical random fields. *Th Prob. Appl.*, 26:632–633, 1981.

[5] D Bosq. Sur une classe de tests qui contient le test du chi-2. *Publications de l'ISUP*, 23:1–16, 1980.

[6] J. Bretagnolle and P. Massart. Hungarian construction from the nonasymptotic viewpoint. *Ann. Prob.*, 17:239–256, 1989.

[7] M. Broniatowski. Estimation through kullback-leibler divergence. *to appear in Mathematical Methods of Statistics*, 2003.

[8] M Broniatowski and A Keziou. Estimation and tests for models satisfying linear constraints with unknown parameters. 2003, submitted.

[9] Michel Broniatowski and Amor Keziou. Parametric estimation and testing through divergences, (*Submitted*). 2003.

[10] I. Csiszar. Information type measures of difference of probability distributions and indirect observations. *Studia Sci. Math. Hungar.*, 2:229–318, 1967.

[11] S. G. Donald, G. W. Imbens, and W.K. Newey. Empirical likelihood estimation and consistent tests with conditional moment restrictions. *J. Econometrics*, 117(1):55–93, 2003.

[12] P. Grinwood and M. S. Nikulin. Some remarks with respect to the application of tests of chi-square type. *Zap. Nauchn. Sem. Leningrad. Otdel. Mat. Inst. Steklov. (LOMI)*, 158(Probl. Teor. Veroyatn. Raspred. X):49–71, 170, 1987.

[13] P. Groeneboom, J. Oosterhooff, and F.H. Ruymgaart. Large deviation theorem for empirical probability measures. *Annals of Probability*, 7:553–586, 1979.

[14] Kallenberg W.C.M. Inglot, T. and T. Ledwina. Asymptotic behaviour of some bilinear functionals of the empirical process. *Math. Meth. Stat.*, 2:316–336, 1993.

[15] T. Inglot and T. Ledwina. On probabilities of excessive deviations for kolmogorov-smirnov, cramer-von mises and chi-square statistics. *Ann. Stat.*, 18:1491–1495, 1990.

[16] R. Jimenez and Y. Shao. On robustness and efficiency of minimum divergence estimators. *Test*, 10,2:241–248, 2001.

[17] V.I. Koltchinskii. Komlos-major-tusnady approximation for the general empirical preocess and haar expansions of classes of functions. *J. Theoret. Prob.*, 7:73–118, 1994.

[18] H. O. Lancaster. *The chi-squared distribution*. John Wiley & Sons Inc., New York, 1969.

[19] M Lemdani and O Pons. Likelihood tests in contamination models. *Bernouilli*, 5,4:705–719, 1999.

[20] F. Liese and I. Vajda. *Convex Statistical Distances*. BSB Teubner, Leipzig, 1987.

[21] B.G. Lindsay. Efficiency versus robustness: The case of minimum hellinger distance and related methods. *Annals of Statistics*, 22:1081–1114, 1994.

[22] P. Massart. Strong approximation for multivariate empirical and related processes, via KMT reconstructions. *Ann. Prob.*, 17:266–291, 1989.

[23] D. Morales, L. Pardo, and I. Vajda. Some new statistics for testing hypotheses in parametric models. *Journal of Multivariate Analysis*, 62:137–168, 1997.

[24] W Newey and R Smith. Higher order properties of GMM and generalized empirical likelihood estimators. *Cemmap Working Paper CWP04/03*, 2003.

[25] D Pollard. *Convergence of stochastic processes*. Springer Series in Statistics. Springer-Verlag, New York, 1984.

[26] G. R. Shorack and J.A. Wellner. *Empirical processes with applications to statistics*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, 1986.

[27] M. Teboulle and I. Vajda. Convergence of best phi-entropy estimates. *IEEE Trans. Inform. Theory*, 39:297–301, 1993.

[1]Université de REIMS and LSTA, Université PARIS 6, 175 Rue du Chevaleret 75013 PARIS, FRANCE, [2]Dip. di Statistica, Probabilitá e Statistiche Applicate, University of Rome "La Sapienza", P.le A. Moro, 5 00185 Roma

E-mail address: [2]samantha.leorato@uniroma1.it