A CHI-SQUARE TYPE TEST FOR COVARIANCES

S. LEORATO

Dip. di Statistica, Probabilità e Statistiche Applicate, University of Rome "La Sapienza", P.le A. Moro 5, 00185 Roma

Tel.: *+39 0649910657* E-mail address: samantha.leorato@uniroma1.it

ABSTRACT. In this paper we propose a test procedure based on chi-square divergence, suitable to testing hypotheses on the covariances of a measure P, such as: $\int fg\,dP = \int f dP \int g dP$, f and g belonging to given classes of functions $\mathcal{H}$ and $\mathcal{K}$. The procedure enters in the range of minimum divergence statistics and relies on convexity and duality properties of the χ^2. We use the statistic χ_n^2 defined by Broniatowski and Leorato (2004) suitably adapted to the covariance constraints setting. Limiting properties of the test statistic are studied, including convergence in distribution under contiguous alternatives. The method is then applied to tests of independence between two random variables. In this case a Chernoff-type large deviation result under H_0 is also proved.

1991 Mathematics Subject Classification. Primary: 62G10. Secondary: 62G20, 60F17.

Key words and phrases. chi-square divergence, hypothesis testing, empirical processes, KMT property, tests of independence, large deviations.

1. Introduction

Let $X_1, \ldots, X_n$ be i.i.d. random variables with values on a metric space $\mathfrak{X}$ and with unknown probability law P. Set $\mathfrak{B} = \mathfrak{B}(\mathfrak{X})$ the Borel $\sigma-$algebra on $\mathfrak{X}$ and $\mathfrak{M}_1$ the set of all signed measures on $\mathfrak{B}$ integrating to one.

Assume we are given a subset Ω of $\mathfrak{M}_1$ and we are interested in testing the hypothesis:

$$H_0 : P \in \Omega \qquad \text{vs} \qquad H_1 : P \notin \Omega. \tag{1.1}$$

In typical situations, one of the most popular test procedures for this purpose is based on χ^2-divergence estimation. We recall that

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

and

$$\chi^2(Q,P) = \begin{cases} \int \left(\frac{dQ}{dP} - 1\right)^2 dP & \text{if } Q \text{ is abs. cont. w.r.to } P \\ \infty & \text{otherwise} \end{cases} \tag{1.3}$$

where $\frac{dQ}{dP}$ is the Radon-Nikodym derivative. The measure $Q^* \in \Omega$ which attains the minimum in (1.2), provided it exists, is called the *projection* (or χ^2-projection) of P onto Ω.

The estimation of χ^2, is typically based on refined partitions of the support $\mathfrak{X}$ and on the empirical measure P_n associated to a random sample $(X_1, \ldots, X_n)$, i.e. $P_n(x) = \frac{1}{n}\sum_{i=1}^n \mathbf{1}_{\{X_i \leq x\}}(X_i)$ (see f.i. Cressie and Read (1984), Györfi and Vajda (2001)...).

Broniatowski and Keziou (2003) proposed an estimation procedure, applied to a parametric context, which avoids partitioning based on the estimation of $\phi-$divergences (which include χ^2 as a particular case). See Liese and Vajda (1987) and references therein for the definition and main properties of $\phi-$divergences.

Broniatowski and Leorato (2004) extended the method to non-parametric setting, limitedly to χ^2 estimation, defining the estimator χ^2_n.

In this paper we focus on the test paradigm (1.1), when Ω is defined by covariance hypotheses, that is

$$\Omega = \{Q \in \mathfrak{M}_1 : Q\xi\zeta = Q\xi Q\zeta, Q\xi < \infty, Q\zeta < \infty, \xi \in \mathcal{H}, \zeta \in \mathcal{K}\} \tag{1.4}$$

where $\mathcal{H}$ and $\mathcal{K}$ are two classes of functions and where $Qf = \int f dQ$.

We propose a test procedure obtained by adapting the statistic χ^2_n mentioned above to the covariance setting (1.4).

In particular, we exploit the following quadratic form representation for χ^2_n (Proposition 3.1 of Broniatowski and Leorato (2004)):

$$\chi^2_n = (a_1 - P_n f_1, \ldots, a_k - P_n f_k)\, \Sigma_n^{-1} \begin{pmatrix} a_1 - P_n f_1 \\ \vdots \\ a_k - P_n f_k \end{pmatrix} \tag{1.5}$$

which holds if Ω is defined by a finite number of linear constraints, namely if

$$\Omega = \{Q \in \mathfrak{M}_1 : Qf_i = a_i,\ i = 1, \ldots, k\}, \tag{1.6}$$

for a given finite set of non-constant functions $\{f_1, \ldots, f_k\}$ and constants $a_i,\ i = 1, \ldots, k$.

The matrix Σ_n in (1.5) is the empirical covariance matrix of the random vector $(\sqrt{n}(P_n - P)f_i)_{1 \leq i \leq k}$.

An analogous representation can be proved for $\chi^2(\Omega, P)$:

$$\chi^2 = (a_1 - Pf_1, \ldots, a_k - Pf_k)\, \Sigma^{-1} \begin{pmatrix} a_1 - Pf_1 \\ \vdots \\ a_k - Pf_k \end{pmatrix}. \tag{1.7}$$

In (1.7), the covariance matrix Σ and the measure P are written in place of the empirical counterparts.

Remark 1. When Ω is an arbitrary subset of $\mathfrak{M}_1$, the definition of χ^2_n relies on the dual representation of χ^2-divergence, which is a consequence of the convexity of the mapping $Q \mapsto \chi^2(Q.P)$ (equation (2.7) in [6]), but (1.7) does not hold any more.

Remark 2. The reason why we assume $\Omega \subset \mathfrak{M}_1$ (although P is supposed to belong to the set of probability measures) is that, roughly speaking, dealing with a subset of non negative measures implies the introduction of inequality constraints in (1.6) which cause the failure of identity (1.7) (and (1.5), consequently). However, since $\chi^2(\Omega, P) = 0$ iff $P \in \Omega$, there is no restriction in assuming $\Omega \subset \mathfrak{M}_1$ for test purposes.

The class of tests (1.1) induced by sets of the form (1.4) includes many examples which are relevant in statistics.

Examples.

- *Test of independence.* Let (X, Y) be two r.v.'s with values in $[0,1]^2$ ($\mathfrak{X} = [0,1]^2$). We want to test whether X and Y are stochastically independent. It is enough to take Ω as in (1.4) and

$$\begin{aligned} \mathcal{H} &= \left\{(x,y) \mapsto \mathbf{1}_{(0,u]}(x),\ u \in [0,1)\right\} \\ \mathcal{K} &= \left\{(x,y) \mapsto \mathbf{1}_{(0,u]}(y),\ u \in [0,1)\right\}. \end{aligned} \tag{1.8}$$

In (1.8) both $\mathcal{H}$ and $\mathcal{K}$ are infinite dimensional. In these cases the estimator is defined using an approach by sieves.

This simple example will be studied more in detail in Section 4. We will show in particular, how the convergence results of the following sections can be improved once $\mathcal{H}$ and $\mathcal{K}$ are given.

- *Test on correlation coefficient.* Assume (X, Y) have values in $\mathfrak{X} = \mathbb{R}^2$. We are interested in testing the simple hypothesis on the correlation coefficient $\rho_{X,Y}$:

$$H_0 : \rho_{X,Y} = 0 \quad vs \quad H_1 : \rho_{X,Y} \neq 0.$$

The test paradigm can be clearly written in form (1.1), if Ω is induced by the functions $\mathcal{H} = \{(x,y) \mapsto x\}$ and $\mathcal{K} = \{(x,y) \mapsto y\}$.

A composite version of the test can also be written. Let us suppose that $Var(X) = Var(Y) = 1$. Then, if $\Omega = \bigcup_{0 \le a \le 1} \Omega_a$, with

$$\Omega_a = \left\{ Q \in \mathfrak{M}_1 : \int xy\, dQ(x,y) = \int x\, dQ_X(x) \int y\, dQ_Y(y) + a \right\}$$

we have that $\{H_0 : P \in \Omega \quad vs \quad H_1 : P \notin \Omega\}$ is equivalent to $\{H_0 : \rho_{X,Y} \ge 0 \quad vs \quad H_1 : \rho_{X,Y} < 0\}$.

- *Affine symmetry.* We consider now the hypothesis of affine symmetry between two r.v.'s (namely simultaneous independence and homogeneity of marginal laws). Assuming for simplicity $\mathfrak{X} = [0,1]^2$, we can write the test paradigm in form (1.1). We can take $\mathcal{H}$ and $\mathcal{K}$ as in (1.8), while

$$\begin{aligned} \Omega \quad = \quad & \left\{ Q \in \mathfrak{M}_1 : Q\mathbf{1}_{(0,u]}(X)\mathbf{1}_{(0,v]}(Y) = Q_X(u)Q_Y(v),\ (u,v) \in [0,1]^2 \right. \\ & \left. Q_X(u) = Q_Y(u),\ u \in [0,1] \right\} \end{aligned}$$

can be seen as the intersection of a subset induced by linear constraints (homogeneity) and another one induced by covariance constraints (independence)

□

We set $\overline{\mathcal{F}} := \mathcal{F} \cup \{\mathbf{1}\} = \overline{\mathcal{H}} \times \overline{\mathcal{K}}$, that is the class of products of functions in $\overline{\mathcal{H}} := \mathcal{H} \cup \{\mathbf{1}\}$ and $\overline{\mathcal{K}} := \mathcal{K} \cup \{\mathbf{1}\}$, where $\mathbf{1}$ is the function identically equal to 1.

Remark 3. Throughout the paper, we assume, without loss of generality, that $\overline{\mathcal{F}}$ is $P-$linearly independent, namely that none of its functions coincides, up to a $P-$null subset of $\mathfrak{X}$, with a linear combination of the other functions in $\mathcal{F}$, and this clearly occurs only if the same property holds for $\overline{\mathcal{H}}$ and $\overline{\mathcal{K}}$ too.

The set (1.4) is not a linear set of measures, indeed it is not even convex, since $Q_1, Q_2 \in \Omega$ does not imply $\alpha Q_1 + (1-\alpha)Q_2 \in \Omega$, for any $\alpha \in [0,1]$. Nevertheless, it is still possible to get a quadratic form representation for χ^2_n, by decomposing Ω into disjoint subsets.

To fix the ideas, we first consider the case when $\mathcal{H}$ and $\mathcal{K}$ have finite dimensions, say h and k respectively.

Assume that the means $P\xi_i,\ i = 1,\dots,h$ and $P\zeta_j,\ j = 1,\dots,k$ are known and are given by the vectors $\underline{r} = \{r_1,\dots,r_h\}$, $\underline{s} = \{s_1,\dots,s_k\}$.

In this case (1.4) writes

(1.9)
$$\Omega(\underline{r},\underline{s}) = \{Q \in M_1\ :\ Q\xi_i\zeta_j = r_i s_j,\ Q\xi_i = r_i,\ Q\zeta_j = s_j,\ i = 1,\dots,h,\ j = 1,\dots,k\}\,.$$

$\Omega(\underline{r},\underline{s})$ has the same linear structure of (1.6) with class of function

$$\mathcal{F} = \{\xi_1,\xi_2,\dots,\xi_h,\zeta_1,\dots,\zeta_k,\xi_1\zeta_1,\xi_1\zeta_2,\dots,\xi_i\zeta_j,\dots,\xi_h\zeta_k\}\,. \tag{1.10}$$

The dimension of $\mathcal{F}$ is $m := h + k + hk$.

For any $\underline{r}$ and $\underline{s}$ the existence of the projection of P to $\Omega(\underline{r},\underline{s})$ follows from Liese's existence theorem (see Liese (1975)), while (1.5) permits us to write:

$$\chi^2_n(\underline{r},\underline{s}) = \chi^2_n(\Omega(\underline{r},\underline{s}),P) = \underline{\nu}^{\mathcal{F}}_n(\underline{r},\underline{s})^T\,\Sigma_n^{-1}\underline{\nu}^{\mathcal{F}}_n(\underline{r},\underline{s}),$$

where

$$\begin{aligned}\underline{\nu}^{\mathcal{F}}_n(\underline{r},\underline{s})^T \quad := \quad \{&P_n\xi_1 - r_1,\dots,P_n\xi_h - r_h, P_n\zeta_1 - s_1,\dots,P_n\zeta_k - s_k,\\ &P_n\xi_1\zeta_1 - r_1 s_1,\dots,P_n\xi_i\zeta_j - r_i s_j,\dots,P_n\xi_h\zeta_k - r_h s_k\}\,.\end{aligned}$$

Throughout the paper we assume that the set (1.4) can be written as

$$\Omega = \bigcup_{(\underline{r},\underline{s})\in\Theta_{\mathcal{H}}\times\Theta_{\mathcal{K}}} \Omega(\underline{r},\underline{s})$$

where $\Theta_{\mathcal{H}} \subset \mathbb{R}^h$ and $\Theta_{\mathcal{K}} \subset \mathbb{R}^k$ are compact sets.

Such limitations seem necessary to our approach, and can be viewed as assumptions on the $P-$means of functions in $\mathcal{H}$ and $\mathcal{K}$.

Clearly, if $\mathcal{H}$ and $\mathcal{K}$ are subsets of B_b (i.e. the class of all bounded and measurable functions), then we can define $\Theta_{\mathcal{H}} = [-\mathsf{h}, \mathsf{h}]^h$ and $\Theta_{\mathcal{K}} = [-\mathsf{k}, \mathsf{k}]^k$, with

$$\mathsf{h} = \sup_{1\le i\le h} \sup_x |\xi_i(x)| \qquad \mathsf{k} = \sup_{1\le j\le k} \sup_x |\zeta_j(x)| \,.$$

Therefore,

$$\chi^2(\Omega, P) = \chi^2\left(\bigcup_{\underline{r},\underline{s}} \Omega(\underline{r},\underline{s}), P\right) = \inf_{\underline{r},\underline{s}} \chi^2(\Omega(\underline{r},\underline{s}), P). \tag{1.11}$$

The above condition can be generalized to infinite dimensional $\mathcal{H}$ and $\mathcal{K}$:

(C1) For any $\mathcal{H}_0 = \{\xi_1, \dots, \xi_h\} \subseteq \mathcal{H}$ and $\mathcal{K}_0 = \{\zeta_1, \dots, \zeta_k\} \subseteq \mathcal{K}$,

$$(P\xi_1, \dots, P\xi_h) \in \Theta_{\mathcal{H}_0} := \{(Q\xi_1, \dots, Q\xi_h) : Q \in \Omega\},$$

$$(P\zeta_1, \dots, P\zeta_k) \in \Theta_{\mathcal{K}_0} := \{(Q\zeta_1, \dots, Q\zeta_k) : Q \in \Omega\}$$

$\Theta_{\mathcal{H}_0}$ and $\Theta_{\mathcal{K}_0}$ are compact subsets of $\mathbb{R}^h$ and $\mathbb{R}^k$ and satisfy $\|\underline{r}\|^2 \le O(h)$, for every $\underline{r} \in \Theta_{\mathcal{H}_0}$ and $\|\underline{s}\|^2 \le O(k)$, for every $\underline{s} \in \Theta_{\mathcal{K}_0}$.

Then, we can rewrite the statistic χ^2_n as follows

$$\begin{aligned} \chi^2_n &= \inf_{(\underline{r},\underline{s})\in\Theta_{\mathcal{H}}\times\Theta_{\mathcal{K}}} \chi^2_n(\underline{r},\underline{s}) \\ &= \inf_{(\underline{r},\underline{s})\in\Theta_{\mathcal{H}}\times\Theta_{\mathcal{K}}} \underline{\nu}^{\mathcal{F}}_n(\underline{r},\underline{s})^T \,\Sigma_n^{-1} \underline{\nu}^{\mathcal{F}}_n(\underline{r},\underline{s}). \end{aligned} \tag{1.12}$$

The paper is structured as follows. In Section 2 we study the finite dimensional environment, prove consistency of the test statistic $n\chi^2_n$ and find asymptotic distribution under H_0.

Section 3 deals with infinite dimensional classes of functions and asymptotic results analogous to those of the previous section are obtained.

For the proofs of consistency and weak convergence of the test statistic under H_0, we will need the analogous results relative to the linear constraints case, for which we refer to Broniatowski and Leorato (2004).

In Section 4 we present a simple application to a test of independence between two r.v's. Here refined partitioning is induced by the set indexed structure of the classes of functions $\mathcal{H}$ and $\mathcal{K}$. A large deviation result of the Chernoff type is also proved.

The proofs of the main results are presented in the last Section.

Before closing this section let us define the vectors

$$\begin{aligned} \underline{\nu}^{\mathcal{F}}(\underline{r},\underline{s})^T \quad := \quad & \{P\xi_1 - r_1, \dots, P\xi_h - r_h, P\zeta_1 - s_1, \dots, P\zeta_k - s_k, \\ & \qquad P\xi_1\zeta_1 - r_1 s_1, \dots, P\xi_i\zeta_j - r_i s_j, \dots, P\xi_h\zeta_k - r_h s_k\} \end{aligned}$$

and

$$\underline{\gamma}_n^T := \underline{\gamma}_n(\mathcal{F})^T = \left\{\sqrt{n}(P_n - P)\xi_1, \dots, \sqrt{n}(P_n - P)\zeta_k, \dots, \sqrt{n}(P_n - P)\xi_i\zeta_j, \dots\right\}.$$

It clearly holds $\underline{\gamma}_n = \sqrt{n}\,(\underline{\nu}^{\mathcal{F}}_n(\underline{r},\underline{s}) - \underline{\nu}^{\mathcal{F}}(\underline{r},\underline{s}))$, for every $(\underline{r},\underline{s}) \in \Theta_{\mathcal{H}} \times \Theta_{\mathcal{K}}$.

2. Finite dimensional classes of functions

In this section we assume that Ω is given by (1.4) with $\mathcal{H} = \{\xi_1, \ldots, \xi_h\}$ and $\mathcal{K} = \{\zeta_1, \ldots, \zeta_k\}$, $h < \infty$ and $k < \infty$ and that Condition (C1) is satisfied. Moreover we assume

(C2) $\mathcal{H}$ and $\mathcal{K}$ have $P-$square integrable envelope functions H and K respectively.

Write $\chi^2(\underline{r}, \underline{s}) = \chi^2(\Omega(\underline{r}, \underline{s}), P)$ and $\chi^2 = \chi^2(\Omega, P)$ and write

$$\overline{\chi}_n^2(\underline{r}, \underline{s}) = \underline{\nu}_n^{\mathcal{F}}(\underline{r}, \underline{s})^T \Sigma^{-1} \underline{\nu}_n^{\mathcal{F}}(\underline{r}, \underline{s}) \tag{2.1}$$

and

$$\overline{\chi}_n^2 = \inf_{(\underline{r}, \underline{s}) \in \Theta_{\mathcal{H}} \times \Theta_{\mathcal{K}}} \overline{\chi}_n^2(\underline{r}, \underline{s}). \tag{2.2}$$

We define the vectors in $\Theta_{\mathcal{H}} \times \Theta_{\mathcal{K}}$

$$\begin{aligned}
(\underline{r}_0, \underline{s}_0) &= \arg \inf_{(\underline{r}, \underline{s}) \in \Theta_{\mathcal{H}} \times \Theta_{\mathcal{K}}} \chi^2(\underline{r}, \underline{s}), \\
(\underline{r}_n, \underline{s}_n) &= \arg \inf_{(\underline{r}, \underline{s}) \in \Theta_{\mathcal{H}} \times \Theta_{\mathcal{K}}} \chi_n^2(\underline{r}, \underline{s}) \\
(\overline{\underline{r}}_n, \overline{\underline{s}}_n) &= \arg \inf_{(\underline{r}, \underline{s}) \in \Theta_{\mathcal{H}} \times \Theta_{\mathcal{K}}} \overline{\chi}_n^2(\underline{r}, \underline{s}).
\end{aligned}$$

We recall also the definition of the algebraic norm of a matrix A which will be largely used in the following sections: $\|A\| = \sup_{\|x\| \leq 1} \frac{\|Ax\|}{\|x\|} = \sup_{\|x\|=1} \|Ax\| = |\lambda_m|$ where $|\lambda_m|$ is, in absolute value, the largest eigenvalue of A and $\|x\|$ is the Euclidean norm.

The first result, the proof of which is deferred to Section 5, concerns consistency of (1.12) as an estimate of $\chi^2(\Omega, P)$.

Theorem 2.1. *Let $\chi^2(\Omega, P) < \infty$. Then, if* (C1) *and* (C2) *hold, we have*

$$\lim_{n \to \infty} \left| \chi_n^2 - \chi^2(\Omega, P) \right| = 0, \quad P - a.s.$$

Note that $(\underline{r}_0, \underline{s}_0)$ and $(\underline{r}_n, \underline{s}_n)$ both exist because $\chi^2((\underline{r}, \underline{s}))$ and $\chi_n^2(\underline{r}, \underline{s})$ are continuous and differentiable functions of $(\underline{r}, \underline{s})$ on the compact and closed space $\Theta_{\mathcal{H}} \times \Theta_{\mathcal{K}} \subset \mathbb{R}^{h+k}$. Moreover, $\chi^2(\underline{r}, \underline{s})$ is strictly convex in $(\underline{r}, \underline{s})$ because it is a definite positive quadratic form and this implies that $(\underline{r}_0, \underline{s}_0)$ is uniquely defined while $(\underline{r}_n, \underline{s}_n)$ must be read as any of the (possibly many) vectors that achieve the infimum in (1.12). If $P \in \Omega$, then there exists a $(\underline{r}^*, \underline{s}^*) \in \Theta_{\mathcal{H}} \times \Theta_{\mathcal{K}}$ such that $P \in \Omega(\underline{r}^*, \underline{s}^*)$ and it is straightforward to see that

$$(\underline{r}_0, \underline{s}_0) = (\underline{r}^*, \underline{s}^*) = \{P\xi_1, \ldots, P\xi_h, P\zeta_1, \ldots, P\zeta_k\}. \tag{2.3}$$

Proposition 2.1. *Under* (C1) *and* (C2) *and if $\chi^2(\Omega, P) < \infty$, then*

$$\lim_{n \to \infty} |(\underline{r}_n, \underline{s}_n) - (\underline{r}_0, \underline{s}_0)| = 0 \quad P - a.s.. \tag{2.4}$$

Proof. The proof is an application of Corollary 3.2.3. in Van der Vaart and Wellner (1996). Set $M_n(\underline{r},\underline{s}) = -\chi_n^2(\underline{r},\underline{s})$, $M(\underline{r},\underline{s}) = -\chi^2(\underline{r},\underline{s})$ and $\Theta = \Theta_{\mathcal{H}} \times \Theta_{\mathcal{K}}$.

By the proof of Theorem 2.1, $\|M_n - M\|_\Theta \to 0$, while uniqueness of $(\underline{r}_0, \underline{s}_0)$ yields the condition $M(\underline{r}_0,\underline{s}_0) > \sup_{\{\Theta_{\mathcal{H}}\times\Theta_{\mathcal{K}}\}-\{G\}} M(\underline{r},\underline{s})$, for any open $G \subset \Theta$ *s.t.* $(\underline{r}_0,\underline{s}_0) \in G$. □

In order to get the asymptotic distribution of χ_n^2 under the null hypothesis we first study the limiting behaviour of the $(h+k)-$dimensional vector $\{(\underline{r}_n,\underline{s}_n) - (\underline{r}_0,\underline{s}_0)\}$. To do so, let us introduce, for every $(\underline{r},\underline{s}) \in \Theta_{\mathcal{H}}\times\Theta_{\mathcal{K}}$, the application $\underline{\eta} : \Theta_{\mathcal{H}}\times\Theta_{\mathcal{K}} \to \Theta_{\mathcal{F}}$:

$$\underline{\eta}(\underline{r},\underline{s}) = \{\eta_1,\ldots,\eta_m\} = \{r_1,\ldots,r_h,s_1,\ldots,s_k,r_1s_1,\ldots,r_is_j,\ldots,r_hs_k\}\,, \tag{2.5}$$

where $\mathcal{F}$ is defined by (1.10) and $\Theta_{\mathcal{F}}$ is closed and bounded in $\mathbb{R}^m$.

Denote by $J = J(\underline{r},\underline{s})$ the Jacobian of $\underline{\eta}$, that is, J is the $(h+k)\times m-$matrix:

$$J(\underline{r},\underline{s}) = \begin{pmatrix} I_h & \mathbf{0}_{h\times k} & \underline{e}_1\underline{s}^T & \underline{e}_2\underline{s}^T & \cdots & \underline{e}_h\underline{s}^T \\ \mathbf{0}_{k\times h} & I_k & r_1I_k & r_2I_k & \cdots & r_hI_k \end{pmatrix} \tag{2.6}$$

where I_k is the unit $k\times k-$matrix, $\underline{e}_j$ is the $j-$th column vector of I_k.

Remark 4. Note that

$$J_0J_0^T = \begin{pmatrix} (1+\|\underline{s}_0\|^2)I_h & \underline{r}_0\underline{s}_0^T \\ \underline{s}_0\underline{r}_0^T & (1+\|\underline{r}_0\|^2)I_k) \end{pmatrix},$$

where $J_0 = J(\underline{r}_0,\underline{s}_0)$. This implies

$$\begin{aligned} |\!|\!| J_0J_0^T - I |\!|\!| &= \sup_{\|\underline{z}\|=1} \underline{z}^T(J_0J_0^T - I)\underline{z} \\ &= \sup_{\|\underline{z}\|^2=\|\underline{x}\|^2+\|\underline{y}\|^2=1} \sum_{i=1}^h x_i^2\|\underline{s}_0\|^2 + \sum_{j=1}^k y_j^2\|\underline{r}_0\|^2 + 2\sum_{i=1}^h x_ir_{0,i}\sum_{j=1}^k y_js_{0,j} \\ &\le \sup_{\|\underline{x}\|^2+\|\underline{y}\|^2=1} \|\underline{x}\|^2\|\underline{s}_0\|^2 + \|\underline{y}\|^2\|\underline{r}_0\|^2 + 2\|\underline{x}\|\|\underline{s}_0\|\|\underline{y}\|\|\underline{r}_0\| \\ &\le (\|\underline{r}_0\| + \|\underline{s}_0\|)^2 \le O(h+k) = O(m^{1/2}) \qquad (2.7) \end{aligned}$$

by Condition (C1).

Analogously, choosing $x_i = \frac{r_{0,i}}{\sqrt{2}\|\underline{r}_0\|}$ and $y_j = \frac{s_{0,j}}{\sqrt{2}\|\underline{s}_0\|}$, J_0 can be bounded below by

$$\begin{aligned} |\!|\!| J_0J_0^T - I |\!|\!| &\ge \left(\|\underline{s}_0\|^2 \sum_{i=1}^h \frac{r_{0,i}^2}{2\|\underline{r}_0\|^2} + \|\underline{r}_0\|^2 \sum_{j=1}^k \frac{s_{0,j}^2}{2\|\underline{s}_0\|^2} + 2\sum_{i=1}^h \frac{r_{0,i}^2}{\sqrt{2}\|\underline{r}_0\|} \sum_{j=1}^k \frac{s_{0,j}^2}{\sqrt{2}\|\underline{s}_0\|} \right) \\ &= \frac{1}{2}\left(\|\underline{r}_0\|^2 + \|\underline{s}_0\|^2 + 2\|\underline{r}_0\|\|\underline{s}_0\|\right) = O(h+k). \qquad (2.8) \end{aligned}$$

We are now able to write $\underline{\eta}$ in terms of $(\underline{r},\underline{s})$ and J_0, as is shown in the next lemma, the proof of which is omitted.

Lemma 2.1. *For every* $(\underline{r},\underline{s}) \in \Theta_{\mathcal{H}}\times\Theta_{\mathcal{K}}$,

$$\{\underline{\eta}(\underline{r},\underline{s}) - \underline{\eta}(\underline{r}_0,\underline{s}_0)\}^T = \{(\underline{r},\underline{s}) - (\underline{r}_0,\underline{s}_0)\}^T J_0 + \underline{\mathsf{a}}(\underline{r},\underline{s})^T\,, \tag{2.9}$$

with

$$\underline{\mathsf{a}}(\underline{r},\underline{s})^T = \{\overbrace{0,\dots,0}^{h+k},(r_1-r_{0,1})(s_1-s_{0,1})\dots,(r_i-r_{0,i})(s_j-s_{0,j}),\dots\}. \tag{2.10}$$

It follows easily from (2.10), by writing $\underline{\mathsf{a}}_n = \underline{\mathsf{a}}(\underline{r}_n,\underline{s}_n)$, that

$$\|\underline{\mathsf{a}}_n\|^2 = \|\underline{r}_n - \underline{r}_0\|^2\|\underline{s}_n - \underline{s}_0\|^2 \le \frac{1}{2}\|(\underline{r}_n,\underline{s}_n)-(\underline{r}_0,\underline{s}_0)\|^4. \tag{2.11}$$

For brevity's sake, we will write, from now on, $\underline{\eta}_n := \underline{\eta}(\underline{r}_n,\underline{s}_n)$, $\underline{\eta}_0 := \underline{\eta}(\underline{r}_0,\underline{s}_0)$, $\underline{\tau}_n := \sqrt{n}\,\{(\underline{r}_n,\underline{s}_n)-(\underline{r}_0,\underline{s}_0)\}$ and finally $\underline{\nu}_0^{\mathcal{F}} := \underline{\nu}^{\mathcal{F}}(\underline{r}_0,\underline{s}_0)$. We remark that, if $P\in\Omega$, it follows from (2.3) that

$$\underline{\eta}_0 = \{P\xi_1,\dots,P\xi_h,P\zeta_1,\dots,P\zeta_k,P\xi_1\zeta_1,\dots,P\xi_h\zeta_k\}.$$

Then, if H_0 holds, we can write, by Lemma 2.1

$$\begin{aligned}
\chi_n^2 &= (\underline{\eta}_n-\underline{\eta}_0)^T\Sigma^{-1}(\underline{\eta}_n-\underline{\eta}_0) - \frac{2}{\sqrt{n}}(\underline{\eta}_n-\underline{\eta}_0)^T\Sigma^{-1}\underline{\gamma}_n + n^{-1}\underline{\gamma}_n^T\Sigma^{-1}\underline{\gamma}_n\\
&\quad+(\underline{\eta}_n-\underline{\eta}_0)^T\left(\Sigma_n^{-1}-\Sigma^{-1}\right)(\underline{\eta}_n-\underline{\eta}_0) - \frac{2}{\sqrt{n}}(\underline{\eta}_n-\underline{\eta}_0)^T\left(\Sigma_n^{-1}-\Sigma^{-1}\right)\underline{\gamma}_n\\
&\quad+n^{-1}\underline{\gamma}_n^T\left(\Sigma_n^{-1}-\Sigma^{-1}\right)\underline{\gamma}_n\\
&= n^{-1}\left[\underline{\tau}_n^T J_0\Sigma^{-1}J_0^T\underline{\tau}_n - 2\underline{\tau}_n^T J_0\Sigma^{-1}\underline{\gamma}_n + \underline{\gamma}_n^T\Sigma^{-1}\underline{\gamma}_n\right] + B_n + C_n \qquad (2.12)
\end{aligned}$$

where

$$\begin{aligned}
B_n &= n^{-1}\underline{\tau}_n^T J_0\Sigma^{-1/2}\left(\Sigma^{1/2}\Sigma_n^{-1}\Sigma^{1/2}-I\right)\Sigma^{-1/2}J_0^T\underline{\tau}_n\\
&\quad+n^{-1}\underline{\gamma}_n^T\Sigma^{-1/2}\left(\Sigma^{1/2}\Sigma_n^{-1}\Sigma^{1/2}-I\right)\Sigma^{-1/2}\underline{\gamma}_n\\
&\quad-2n^{-1}\underline{\gamma}_n^T\Sigma^{-1/2}\left(\Sigma^{1/2}\Sigma_n^{-1}\Sigma^{1/2}-I\right)\Sigma^{-1/2}J_0^T\underline{\tau}_n, \qquad (2.13)
\end{aligned}$$

and

$$\begin{aligned}
C_n &= 2n^{-1/2}\underline{\tau}_n^T J_0\Sigma^{-1}\underline{\mathsf{a}}_n + \underline{\mathsf{a}}_n^T\Sigma^{-1}\underline{\mathsf{a}}_n - 2n^{-1/2}\underline{\mathsf{a}}_n^T\Sigma^{-1}\underline{\gamma}_n\\
&\quad+\underline{\mathsf{a}}_n^T\Sigma^{-1/2}\left(\Sigma^{1/2}\Sigma_n^{-1}\Sigma^{1/2}-I\right)\Sigma^{-1/2}\underline{\mathsf{a}}_n\\
&\quad+2n^{-1/2}\left(\underline{\tau}_n^T J_0-\underline{\gamma}_n\right)\Sigma^{-1/2}\left(\Sigma^{1/2}\Sigma_n^{-1}\Sigma^{1/2}-I\right)\Sigma^{-1/2}\underline{\mathsf{a}}_n. \qquad (2.14)
\end{aligned}$$

Negligibility of B_n and C_n follows straightforwardly from (2.4), (2.11) taking into account (5.6).

Theorem 2.2. *Let $P\in\Omega$. Then*

$$\underline{\tau}_n = \left[J_0\Sigma^{-1}J_0^T\right]^{-1}J_0\Sigma^{-1}\underline{\gamma}_n + o_P(1). \tag{2.15}$$

Proof. A more general result, from which (2.15) can be extracted, is proved in Theorem 3.2. □

By using Theorem 2.2, (2.12) becomes:

$$\begin{aligned}
n\chi_n^2 &= \underline{\gamma}_n^T\Sigma_n^{-1}\underline{\gamma}_n - \underline{\gamma}_n^T\Sigma^{-1}J_0^T\left(J_0\Sigma^{-1}J_0^T\right)^{-1}J_0\Sigma^{-1}\underline{\gamma}_n + o_P(1)\\
&= \underline{\gamma}_n^T\Sigma^{-1/2}\left\{I-\Sigma^{-1/2}J_0^T\left(J_0\Sigma^{-1}J_0^T\right)^{-1}J_0\Sigma^{-1/2}\right\}\Sigma^{-1/2}\underline{\gamma}_n + o_P(1). \qquad (2.16)
\end{aligned}$$

In (2.16) we have used $nB_n = o_P(1)$ and $nC_n = o_P(1)$, which also results from the proof of Theorem 3.2.

Theorem 2.3. *If $P \in \Omega$ then $n\chi_n^2$ converges weakly to a chi − square distributed r.v., with degrees of freedom $d = hk$.*

Proof. The matrix

$$I - \mathbf{P} = I - \Sigma^{-1/2} J_0^T \left(J_0 \Sigma^{-1} J_0^T \right)^{-1} J_0 \Sigma^{-1/2} \tag{2.17}$$

is idempotent with trace $tr\{I - \mathbf{P}\} = h + k + hk - tr\{J_0 \Sigma^{-1} J_0^T \left(J_0 \Sigma^{-1} J_0^T \right)^{-1}\} = hk$.

It is well known that, if $\underline{y} \sim N(0, I)$, then $\underline{y}^T (I - \mathbf{P}) \underline{y}$ is a chi-square distributed r.v. with degrees of freedom equal to $tr\{I - \mathbf{P}\}$. Multidimensional CLT for $\underline{\gamma}_n$ completes the proof. □

3. Infinite dimensional classes of functions.

We now consider the case where $\mathcal{H}$ and $\mathcal{K}$ are infinite dimensional classes of functions satisfying Conditions (C1) and (C2). In this case we can't write the quadratic form representation (1.7) as such. The method used to adapt χ_n^2 to an infinite number of covariance constraints is the same of Broniatowski and Leorato (2004) and is based on the approximation of Ω by sieves.

We therefore consider two sequences of finite dimensional classes of functions $\{\mathcal{H}_n\}_{n \geq 1}$ and $\{\mathcal{K}_n\}_{n \geq 1}$ such that the following condition is fulfilled:

(C3) For every n, $\mathcal{H}_n$ and $\mathcal{K}_n$ have finite dimensions h_n and k_n respectively, with $\lim_{n\to\infty} h_n = \infty$, $\lim_{n\to\infty} k_n = \infty$, such that

$$\mathcal{H}_n \subseteq \mathcal{H}_{n+1} \subseteq \ldots \subseteq \mathcal{H}, \quad \lim_{n\to\infty} \mathcal{H}_n = \cup_{n=1}^{\infty} \mathcal{H}_n \subseteq \mathcal{H}, \quad cl\,\{\cup_{n=1}^{\infty} \mathcal{H}_n\} = \mathcal{H},$$

$$\mathcal{K}_n \subseteq \mathcal{K}_{n+1} \subseteq \ldots \subseteq \mathcal{K}, \quad \lim_{n\to\infty} \mathcal{K}_n = \cup_{n=1}^{\infty} \mathcal{K}_n = \mathcal{K}, \quad cl\,\{\cup_{n=1}^{\infty} \mathcal{K}_n\} = \mathcal{K}.$$

Remark 5. Lower semicontinuity of the function $\chi^2(\cdot, P) \to \mathbb{R}^+$ (see Liese and Vajda (1987)), implies that, if Λ_n is a sequence of subsets of $\mathfrak{M}_1$, converging monotonically to $\Lambda \subset \mathfrak{M}_1$ then

$$\lim_{n\to\infty} \chi^2(\Lambda_n, P) = \chi^2(\Lambda, P).$$

Furthermore, monotone convergence of the sequence $\mathcal{F}_n$ to $\mathcal{F}$ (Condition (C3)) corresponds to monotone convergence of the sequence of linear sets of measures, that is, $\Lambda_n = \{Q \in \mathfrak{M} : \; Qf = 0, f \in \mathcal{F}_n\}$ decreases to $\Lambda = \{Q \in \mathfrak{M} : \; Qf = 0, f \in \mathcal{F}\}$. Indeed Conditions (C1) and (C3) imply also monotone convergence for the non-increasing sequence $\{\Omega_n\}_{n \geq 1}$ given by

$$\Omega_n = \{Q \in \mathfrak{M} : Q\xi\zeta = Q\xi Q\zeta, \; \xi \in \mathcal{H}_n, \; \zeta \in \mathcal{K}_n\}, \quad n \geq 1, \tag{3.1}$$

for which it holds, $\Omega_n \supseteq \Omega_{n+1}$, for every n, and $\lim_{n\to\infty} \Omega_n = \cap_{n=1}^{\infty} \Omega_n = \Omega$. This guarantees that $\lim_{n\to\infty} \chi^2(\Omega_n, P) = \chi^2(\Omega, P)$, and implies also convergence in *variation* of the respective projections (see also Teboulle and Vajda (1993)).

Let us define, for every $(\underline{r},\underline{s}) \in \Theta_{\mathcal{H}_n} \times \Theta_{\mathcal{K}_n}$

$$\begin{aligned}\Omega_n(\underline{r},\underline{s}) = \{Q : Q\xi_i\zeta_j = r_i s_j,\ Q\xi_i = r_i,\ Q\zeta_j = s_j,\ \xi_i \in \mathcal{H}_n,\ \zeta_j \in \mathcal{K}_n, \\ i = 1,\dots,h_n,\ j = 1,\dots,k_n\}.\end{aligned} \tag{3.2}$$

As in Section 2, we can write:

$$\chi_n^2 = \inf_{(\underline{r},\underline{s})\in\Theta_{\mathcal{H}_n}\times\Theta_{\mathcal{K}_n}} \chi_n^2(\underline{r},\underline{s}) \tag{3.3}$$

with $\chi_n^2(\underline{r},\underline{s}) = \underline{\nu}_n^{\mathcal{F}_n}(\underline{r},\underline{s})^T \Sigma_n^{-1} \underline{\nu}_n^{\mathcal{F}_n}(\underline{r},\underline{s})$, and

$$\chi_{(n)}^2 = \chi^2(\Omega_n, P) = \inf_{(\underline{r},\underline{s})\in\Theta_{\mathcal{H}_n}\times\Theta_{\mathcal{K}_n}} \chi_{(n)}^2(\underline{r},\underline{s}) \tag{3.4}$$

where $\chi_{(n)}^2(\underline{r},\underline{s}) = \underline{\nu}^{\mathcal{F}_n}(\underline{r},\underline{s})^T \Sigma_{(n)}^{-1} \underline{\nu}^{\mathcal{F}_n}(\underline{r},\underline{s})$. The collection $\mathcal{F}_n$, indexing $\underline{\nu}^{\mathcal{F}_n}$ and $\underline{\nu}_n^{\mathcal{F}_n}$ in (3.3) and (3.4), is given by $\mathcal{F}_n = \left\{\overline{\mathcal{H}}_n \times \overline{\mathcal{K}}_n\right\} - \{\mathbf{1}\}$. In order to shorten the notation, we write $\Sigma := \Sigma_{(n)}$ (λ_1 for its minimum eigenvalue) and $\underline{\gamma}_n := \underline{\gamma}_n(\mathcal{F}_n)$ and we will often write h and k instead of h_n and k_n respectively.We also introduce, in analogy with Section 2,

$$\begin{aligned}(\underline{r}_n,\underline{s}_n) &= \arg\inf_{(\underline{r},\underline{s})\in\Theta_{\mathcal{H}_n}\times\Theta_{\mathcal{K}_n}} \chi_n^2(\underline{r},\underline{s}) \\ (\underline{r}_0,\underline{s}_0) &= \arg\inf_{(\underline{r},\underline{s})\in\Theta_{\mathcal{H}_n}\times\Theta_{\mathcal{K}_n}} \chi_{(n)}^2(\underline{r},\underline{s}) \\ (\overline{\underline{r}}_n,\overline{\underline{s}}_n) &= \arg\inf_{(\underline{r},\underline{s})\in\Theta_{\mathcal{H}_n}\times\Theta_{\mathcal{K}_n}} \overline{\chi}_n^2(\underline{r},\underline{s}).\end{aligned} \tag{3.5}$$

Theorem 3.1. *Let $\mathcal{H}$ and $\mathcal{K}$ be two $P-$Donsker classes of functions such that Conditions* (C1) *and* (C2) *hold. Assume that there exist two approximating sequences for $\mathcal{H}$ and $\mathcal{K}$, $\{\mathcal{H}_n\}_{n\geq 1}$ and $\{\mathcal{K}_n\}_{n\geq 1}$ respectively, satisfying Condition* (C3)*, with dimensions $h_n \to \infty$ and $k_n \to \infty$, such that:*

$$\lim_{n\to\infty} \lambda_1^{-1}(h + k + hk)n^{-1/2} = 0. \tag{3.6}$$

Then, $\left|\chi_n^2 - \chi^2\right| \to 0 \quad P-a.s.$.

Proof. Write $\left|\chi_n^2 - \chi^2\right| \leq \left|\chi_n^2 - \chi_{(n)}^2\right| + \left|\chi_{(n)}^2 - \chi^2\right|$.

Condition (C3) and Remark 5 imply that it is enough to prove $\lim_{n\to\infty}\left|\chi_n^2 - \chi_{(n)}^2\right| = 0$ $P-$a.s..

Equations (5.1), (5.2) and (5.3), together with (5.4) (see the Proof of Theorem 2.1 in Section 5) and the opposite inequality imply then the result, provided that $n^{-1/2}m^{1/2} = o_P(1)$ and that

$$\lim_{n\to\infty}\left\|\Sigma^{1/2}\Sigma_n^{-1}\Sigma^{1/2} - I\right\| = \lim_{n\to\infty}\left\|\Sigma_n^{1/2}\Sigma^{-1}\Sigma_n^{1/2} - I\right\| = 0.$$

The last two limits follow if $\lim_{n\to\infty}\lambda_1^{-1}mn^{-1/2} = \lim_{n\to\infty}\lambda_{1,n}^{-1}mn^{-1/2} = 0$ (and these are sufficient $n^{-1/2}m^{1/2} = o_P(1)$). Indeed, any of the two limits implies the other one, therefore (3.6) is sufficient for consistency. The proof of the above claim is deferred to Remark 7. □

The following results are generalizations to Proposition 2.1 and Theorem 2.2 to the infinite dimensional setting and are necessary to prove the weak convergence under H_0.

A slight adjustment on the proof of Lemma 3.2.1 in van der Vaart and Wellner (1996) leads to

Lemma 3.1. *Let $\hat{M}_n$ be a sequence of stochastic processes and M_n a sequence of deterministic functions indexed on the (sequence of) parametric spaces Θ_n and continuous for every n.*

Suppose that $\|\hat{M}_n - M_n\|_{\Theta_n} = o_P(1)$ for $n \to \infty$ and that $M_n(\theta_n^*) > \sup_{\theta \in G} M_n(\theta)$ for every $G \subset \Theta_n$ such that $\theta_n^* \notin G$.

Then each sequence $\hat{\theta}_n^*$ such that $\hat{M}_n(\hat{\theta}_n^*) > \sup_{\Theta_n} \hat{M}_n(\theta) - o_P(1)$ satisfies $\|\hat{\theta}_n^* - \theta_n^*\| \to 0$ $P-$a.s..

Choosing $\hat{M}_n = -\chi_n^2$, $M_n = -\chi_{(n)}^2$, $\Theta_n = \Theta_{\mathcal{H}_n} \times \Theta_{\mathcal{K}_n}$ and under the same conditions of Theorem 3.1, we then get $\|(\underline{r}_n, \underline{s}_n) - (\underline{r}_0, \underline{s}_0)\| = o_P(1)$.

The following theorem implies, as a particular case, result (2.15). Its proof is postponed to the last section.

Theorem 3.2. *Suppose that $\mathcal{H}$ and $\mathcal{K}$ are $P-$Donsker classes such that Conditions* (C1) *and* (C2) *hold. Assume that there exist two sequences of finite classes of functions $\{\mathcal{H}_n\}_{n\geq 1}$ and $\{\mathcal{K}_n\}_{n\geq 1}$ which satisfy Condition* (C3) *for $\mathcal{H}$ and $\mathcal{K}$ respectively.*

Let moreover the sequences of dimensions $h_n \to \infty$ and $k_n \to \infty$ satisfy:

$$\lim_{n\to\infty} \frac{m^{5/4}}{\lambda_1} n^{-1/2} = \lim_{n\to\infty} \frac{(hk)^{5/4}}{\lambda_1} n^{-1/2} = 0 \qquad P - a.s.. \tag{3.7}$$

Finally let $\lambda_m = o_P(m^{1/2})$, $\frac{\lambda_m}{\lambda_1} = O_P(m)$ where λ_m is the largest eigenvalue of Σ.

Then, if $P \in \Omega$

$$\underline{\tau}_n = [J_0 \Sigma^{-1} J_0^T]^{-1} J_0 \Sigma^{-1} \underline{\gamma}_n + o_P\left(m^{1/4}\right). \tag{3.8}$$

By replacing (3.8) into (2.12), we get, under H_0,

$$n\chi_n^2 = \underline{\gamma}_n^T \Sigma^{-1/2} (I - \mathbf{P}) \Sigma^{-1/2} \underline{\gamma}_n + nB_n + nC_n + o_P\left(\left\| \mathbf{P} \Sigma^{-1/2} \underline{\gamma}_n \right\|^2 \right). \tag{3.9}$$

LLN yields that $n\chi_n^2 = O_P(hk)$. Then, in order to obtain useful asymptotic results, we study the convergence in distribution of the standardized test statistic $\frac{n\chi_n^2 - hk}{\sqrt{2hk}}$. Theorem 3.3 below proves that, under the appropriate conditions over the sequence of dimensions h and k, the test statistic converges weakly to a standard normal r.v..

We first recall some useful definitions and results.

A class $\mathcal{F}$ is said to be *pre-Gaussian* if there exists a version $\omega_P^0(.)$ 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 $\|H\|_{\mathcal{F}} = \sup_{f\in\mathcal{F}} |H(f)|$. Let δ_n be a decreasing sequence such that for some $a > 0$ we have $\delta_n = o(n^{-a})$.

Definition 1. A class of functions $\mathcal{F}$ is *Komlós-Major-Tusnády* with respect to P and with rate δ_n (i.e. $\mathcal{F} \in KMT(\delta_n, P)$) *iff* it is pre-Gaussian and there exists a version $\omega_n(\cdot)$ of $P-$Brownian bridges such that, for every $t > 0$:

$$P\left\{\sup_{f\in\mathcal{F}}\left|\sqrt{n}(P_n - P)f - \omega_n(f)\right| \geq \delta_n(t + b\log n)\right\} \leq ce^{-\beta t} \tag{3.10}$$

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

Remark 6. Dating back to Komlós *et al.* pioneering works of (1975), concerning inequality (3.10) for the classical empirical process, there is a wide literature on the subject. Most of the results concern generalizations to set indexed classes of functions, such as spheres in $\mathbb{R}^d$, quadrants in $\mathbb{R}^d$ or VC sets (see for instance Borisov (1982) and Massart (1989)).

¿From Borel-Cantelli Lemma and (3.10) if follows that

$$\sup_{f\in\mathcal{F}}\left|\gamma_n(f) - \omega_n^0(f)\right| = O_P\left(\delta_n \log n\right). \tag{3.11}$$

We will use the fact that a $KMT-$class is also $P-$Donsker. With all this at hand we are now able to prove the weak convergence result for χ_n^2

Theorem 3.3. *Suppose that all conditions of Theorem 3.2 are fulfilled and that*

$$\lim_{n\to\infty} n^{-1/2} m^{3/2} \lambda_1^{-1} = 0 \quad P - a.e. \tag{3.12}$$

Assume further that the class $\mathcal{F}$ defined by (1.10) is a $KMT(\delta_n; P)$ class, with

$$\lim_{n\to\infty} \delta_n \log n \, m^{1/2} \lambda_1^{-1/2} = 0 \quad P - a.e. \tag{3.13}$$

Then, if $P \in \Omega$,

$$\frac{n\chi_n^2 - hk}{\sqrt{2hk}} \to N(0,1). \tag{3.14}$$

Proof. The proof is deferred to Section 5. □

Remark 7 (Data dependent choice of the number of classes.). Conditions (3.6), (3.7), (3.12) and (3.13) depend implicitly on the quantities $P\xi\zeta$, $\xi \in \mathcal{H}, \zeta \in \mathcal{K}$, due to λ_1.

We show here that λ_1 can be replaced by $\lambda_{n,1}$, the minimum eigenvalue of the empirical covariance matrix Σ_n and, following the lines of Conti and Scanu (1998), we give a method for choosing the number of classes through the sample.

We claim that, if (3.6) holds, with λ_1 replaced by $\lambda_{n,1}$, then $\lim_{n\to\infty} \frac{\lambda_{n,1}}{\lambda_1} = \lim_{n\to\infty} \frac{\lambda_1}{\lambda_{n,1}} = 1$.

In fact, by the inequalities

$$\inf_{\|x\|=1} x^T(\Sigma - \Sigma_n)x \leq \inf_{\|x\|=1} x^T \Sigma x - \inf_{\|x\|=1} x^T \Sigma_n x \leq \sup_{\|x\|=1} x^T(\Sigma - \Sigma_n)x$$

we have

$$\inf_{\|x\|=1} x^T(\Sigma - \Sigma_n)x + \lambda_{n,1} \leq \lambda_1 \leq \lambda_{n,1} + \sup_{\|x\|=1} x^T(\Sigma - \Sigma_n)x. \tag{3.15}$$

Moreover

$$cmn^{-1/2}\inf_{g\in\mathcal{F}^2}|\gamma_n(g)| \le \inf_{\|x\|=1} x^T(\Sigma-\Sigma_n)x \le \|\Sigma-\Sigma_n\| \le Cmn^{-1/2}\sup_{g\in\mathcal{F}^2}|\gamma_n(g)|,$$

for some constants c and C. That is (3.15) becomes

$$1+O_P\left(mn^{-1/2}\lambda_{n,1}^{-1}\right) \le \lambda_1\lambda_{n,1}^{-1} \le 1+O_P\left(mn^{-1/2}\lambda_{n,1}^{-1}\right).$$

The above inequalities imply that, if $mn^{-1/2}\lambda_{n,1}^{-1}=o_P(1)$ then also (3.6) holds.

On the other hand, repeating the same reasoning with λ_1 and $\lambda_{n,1}$ exchanged, we conclude that (3.6) implies $\lim_{n\to\infty} mn^{-1/2}\lambda_{n,1}^{-1}=0$ and $\frac{\lambda_{n,1}}{\lambda_1}\to 1$.

It therefore follows that, if $\lambda_{n,1}^{-1}m^{3/2}n^{-1/2}\to 0$ then (3.12) holds and $\lim_{n\to\infty}\delta_n \log n\lambda_{n,1}^{-1/2}m^{1/2}=0$ implies (3.13).

¿From the application presented in the next section, it emerges that sometimes condition (3.13) alone is sufficient to prove the consistency and convergence of the test statistic under H_0, and that (3.13) yields

$$\lim_n \lambda_1\lambda_{n,1}^{-1} = \lim_n \lambda_1\lambda_{n,1}^{-1} = \lim_n \frac{np_{i,j}}{N_{i,j}} = \lim_n \frac{N_{i,j}}{np_{i,j}} = 1.$$

Write now

$$m_n = h_n k_n = \begin{cases} 1 & n=1 \\ m_{n-1} & if\ m\in\mathbb{N},\ \varphi(m)\notin V_n \\ \max\left\{m\in\mathbb{N}:\ \varphi(m)\in V_n\right\}, & \end{cases} \tag{3.16}$$

where

$$\varphi(m)=\min\left\{\lambda_{n,1}m^{-3/2}, m^{-1/2}\lambda_{n,1}^{1/2}\frac{n^{-1/2}}{\delta_n\log n}\right\}$$

and $V_n=\left[(n+1)^{-1/2+\varepsilon}, n^{-1/2+\varepsilon}\right)$ for some $0<\varepsilon<1/2$.

Then it is easy to see that m_n is the higher sequence of cells for which the required conditions hold (see Conti and Scanu (1998)).

3.1. **Distribution under contiguous alternatives.** The asymptotic distribution of a test statistic under contiguous alternatives is necessary to study efficiency of the test procedure in terms of its Pitman ARE.

We consider the contiguous alternatives defined by the following model:

$$\underline{\rho}-\underline{\eta}_P = n^{-1/2}\underline{\varepsilon} \tag{3.17}$$

where $\underline{\rho}$ is the vector

$$\underline{\rho}=\{P\xi_1, P\xi_2,\ldots,P\xi_h, P\zeta_1,\ldots,P\zeta_k,\ldots,P\xi_i\zeta_j,\ldots,P\xi_h\zeta_k\}$$

and $\underline{\eta}_P=\underline{\eta}(\underline{r}_P,\underline{s}_P)$, with

$$(\underline{r}_P,\underline{s}_P)=\{P\xi_1,\ldots,P\xi_h,P\zeta_1,\ldots,P\zeta_k\}.$$

In (3.17) $\underline{\varepsilon}$ is the $m-$dimensional vector (with values depending on n)

$$\underline{\varepsilon}=\{0\ldots,0,\varepsilon_{1,1},\ldots,\varepsilon_{i,j},\ldots,\varepsilon_{h,k}\},$$

with the constraints $\varepsilon_{i,\cdot} = \sum_{j=1}^{k} \varepsilon_{i,j} = \sum_{i=1}^{h} \varepsilon_{i,j} = \varepsilon_{\cdot,j} = 0$, for every i, j, that is

$$\underline{\varepsilon} = \{\varepsilon_{1,\cdot}, \ldots, \varepsilon_{\cdot,k}, \varepsilon_{1,1}, \ldots, \varepsilon_{i,j}, \ldots, \varepsilon_{h,k}\}.$$

We want to examine the asymptotic distribution of χ_n^2 when (3.17) holds and determine the sequence $\underline{\varepsilon}$ which guarantees convergence of the standardized χ_n^2 to a $N(0,1)$ r.v..

The asymptotic distribution of χ_n^2, suitably normalized, can be obtained by decomposing χ_n^2 into a sum similar to (2.16) but much more complicated by the fact that the term $(\underline{\eta}_0 - \underline{\rho})$ does not cancel out:

$$\begin{aligned} n\chi_n^2(\underline{r}, \underline{s}) \quad = \quad & n\left(\underline{\eta}(\underline{r}, \underline{s}) - \underline{\eta}_0\right)^T \Sigma_n^{-1} \left(\underline{\eta}(\underline{r}, \underline{s}) - \underline{\eta}_0\right) + 2n\left(\underline{\eta}(\underline{r}, \underline{s}) - \underline{\eta}_0\right)^T \Sigma_n^{-1} \left(\underline{\eta}_0 - \underline{\eta}_P\right) \\ & + \left(\underline{\eta}_0 - \underline{\eta}_P\right)^T \Sigma_n^{-1} \left(\underline{\eta}_0 - \underline{\eta}_P\right) - 2\sqrt{n}\left(\underline{\eta}(\underline{r}, \underline{s}) - \underline{\eta}_0\right)^T \Sigma_n^{-1}(\underline{\gamma}_n + \underline{\varepsilon}_n) \\ & -2\sqrt{n}\left(\underline{\eta}_0 - \underline{\eta}_P\right)^T \Sigma_n^{-1}(\underline{\gamma}_n + \underline{\varepsilon}_n) \\ & + \left(\underline{\gamma}_n + \underline{\varepsilon}_n\right)^T \Sigma_n^{-1} \left(\underline{\gamma}_n + \underline{\varepsilon}_n\right). \end{aligned} \tag{3.18}$$

In this setting, the following convergence result holds true, the proof of which is sketched in Section 5.

Theorem 3.4. *Suppose P satisfies (3.17), and that $\underline{\varepsilon}_n^T \Sigma^{-1/2}(I - \mathbf{P})\Sigma^{-1/2}\underline{\varepsilon}_n = O_P((hk)^{1/2})$, then under the same conditions of Theorem 3.3,*

$$\frac{n\chi_n^2 - \mu(\underline{\varepsilon})}{\sqrt{2hk}} \to N(0,1). \tag{3.19}$$

4. A test for independence

Assume Ω is the set of all probability measures on the product space $\mathfrak{X} \times \mathfrak{Y} = [0,1] \times [0,1]$.

Consider the classes of functions $\mathcal{H}$ and $\mathcal{K}$

$$\begin{aligned} \mathcal{H} &= \left\{\mathbf{1}_{u,1}(x,y),\ u \in [0,1],\ (x,y) \in [0,1]^2\right\} \\ \mathcal{K} &= \left\{\mathbf{1}_{1,v}(x,y),\ v \in [0,1],\ (x,y) \in [0,1]^2\right\}, \end{aligned} \tag{4.1}$$

where

$$\mathbf{1}_{u,v}(x,y) := \mathbf{1}_{u,1}(x,y)\mathbf{1}_{1,v}(x,y) = \begin{cases} 1 & \text{if } 0 \le x \le u,\ 0 \le y \le v \\ 0 & \text{otherwise} \end{cases}$$

Then we can write

$$\Omega = \{Q \in \mathfrak{M}_1 :\ Qfg - QgQf = 0,\ \forall\, f \in \mathcal{H},\ g \in \mathcal{K}\}.$$

Since the classes $\mathcal{H}$ and $\mathcal{K}$ are infinite, according to the procedure outlined in the previous section, we need to build up their approximating sequences.

For $n \ge 1$, $h = h_n \ge 1$, $k = k_n \ge 1$, we choose the sets $\mathcal{U}_n = \left\{u_1^{(n)}, \ldots, u_h^{(n)}\right\}$ and $\mathcal{V}_n = \left\{v_1^{(n)}, \ldots, v_k^{(n)}\right\}$, with $0 < u_1^{(n)} < \ldots < u_i^{(n)} < \ldots < u_h^{(n)} < 1 = u_{h+1}^{(n)}$ and $0 < v_1^{(n)} < \ldots < v_j^{(n)} < \ldots < v_k^{(n)} < 1 = v_{k+1}^{(n)}$, such that:

$$\lim_{n\to\infty} h = \infty, \quad \lim_{n\to\infty} k = \infty; \tag{4.2}$$

$$(4.3)\qquad \begin{aligned} 0 &< \lim_{n\to\infty} \inf_{1\le i\le h} h(u_i^{(n)} - u_{i-1}^{(n)}) \le \lim_{n\to\infty} \sup_{1\le i\le h} h(u_i^{(n)} - u_{i-1}^{(n)}) < \infty \\ 0 &< \lim_{n\to\infty} \inf_{1\le j\le k} k(v_j^{(n)} - v_{j-1}^{(n)}) \le \lim_{n\to\infty} \sup_{1\le j\le k} k(v_j^{(n)} - v_{j-1}^{(n)}) < \infty \end{aligned}$$

and such that

$$(4.4)\qquad \begin{aligned} &\mathcal{U}_n \subseteq \mathcal{U}_{n+1} \subseteq \ldots \subseteq \cup_{n=1}^{\infty}\mathcal{U}_n = \lim_{n\to\infty} \mathcal{U}_n = \mathcal{U} \text{ and } cl\{\mathcal{U}\} = [0,1]; \\ &\mathcal{V}_n \subseteq \mathcal{V}_{n+1} \subseteq \ldots \subseteq \cup_{n=1}^{\infty}\mathcal{V}_n = \lim_{n\to\infty} \mathcal{V}_n = \mathcal{V} \text{ and } cl\{\mathcal{V}\} = [0,1]. \end{aligned}$$

Then for the sequences

$$\begin{aligned} \{\mathcal{H}_n\}_{n\ge1} &= \{\mathbf{1}_{u_i,1}(x,y),\ u_i \in \mathcal{U}_n\}_{n\ge1} \\ \{\mathcal{K}_n\}_{n\ge1} &= \{\mathbf{1}_{1,v_j}(x,y),\ v_j \in \mathcal{V}_n\}_{n\ge1} \end{aligned}$$

Condition (C3) holds.

Write $\overline{\mathcal{F}} = \overline{\mathcal{H}} \times \overline{\mathcal{K}}$ following (1.10), and

$$(4.5)\qquad \mathcal{F}_n = \left\{\mathbf{1}_{u_i,v_j},\ (u_i,v_j) \in \mathcal{U}_n \times \mathcal{V}_n\right\}$$

and finally define χ_n^2 by formula (3.3) using (4.5).

The class $\overline{\mathcal{F}}$ is a KMT class for P with a rate $\delta_n = O(n^{-1/2}\log n)$ if P belongs to Ω (Tusnady (1977)).

Theorem 3.3 implies then that $\frac{\chi_n^2 - hk}{\sqrt{2hk}}$ is asymptotically normally distributed (under the null hypothesis) if condition (3.8) and the two conditions below are satisfied

$$\begin{aligned} &\lim_{n\to\infty} \left(\frac{hk}{\lambda_1 n}\right)^{1/2} (\log n)^2 = 0, \quad P - a.e.; \\ &\lambda_m \le o_P(m^{1/2}), \quad \frac{\lambda_m}{\lambda_1} = O_P(m). \end{aligned}$$

In order to explicit the rate conditions over h and k, we need an estimate of the eigenvalues λ_1 and λ_m.

To this extent, we consider the classes of increment functions (indicator functions of disjoint intervals):

$$(4.6)\qquad \begin{aligned} \overline{\Delta\mathcal{H}}_n &= \left\{\mathbf{1}_{i,\cdot} = \mathbf{1}_{u_i,1} - \mathbf{1}_{u_{i-1},1},\ i = 1\ldots,h+1\right\} \\ \overline{\Delta\mathcal{K}}_n &= \left\{\mathbf{1}_{\cdot,j} = \mathbf{1}_{1,v_j} - \mathbf{1}_{1,v_{j-1}},\ j = 1\ldots,k+1\right\} \end{aligned}$$

and define the class $\Delta\mathcal{F}_n$ by $\overline{\Delta\mathcal{F}}_n = \overline{\Delta\mathcal{H}}_n \times \overline{\Delta\mathcal{K}}_n$ $\left(\Delta\mathcal{F}_n = \overline{\Delta\mathcal{F}}_n - \{\mathbf{1}\}\right)$.

In dealing with these classes of functions it is convenient to endow the vector $\overline{\Delta\mathcal{F}}_n$ with the ordering that associates, to the $\{(i-1)(k+1)+j\}-$th position, the function $\mathbf{1}_{i,j} = \mathbf{1}_{i,\cdot}\mathbf{1}_{\cdot,j}$, $1 \le i \le h+1$, $1 \le j \le k+1$.

Let $\Delta\Sigma$ be the covariance matrix of $\underline{\gamma}_n(\Delta\mathcal{F}_n)$. $\Delta\Sigma$ is easier than Σ to deal with. Moreover, it can be seen that the two matrices are linked by

$$(4.7)\qquad \Sigma = \mathbf{M}\,(\Delta\Sigma)\,\mathbf{M}^{\mathbf{T}}$$

where $\mathbf{M}$ is the block matrix

$$\mathbf{M} = \begin{pmatrix} M_{k+1,k+1} & \mathbf{0} & \dots & \mathbf{0} \\ M_{k+1,k+1} & M_{k+1,k+1} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ M_{k,k+1} & M_{k,k+1} & \dots & M_{k,k}. \end{pmatrix} \tag{4.8}$$

In the above formula $M_{k,k}$ is the inferior triangular unit matrix of dimension k while $M_{k,k+1}$ is obtained from $M_{k+1,k+1}$ by cutting off the $(k+1)-$th row.

An analogous relation holds for $\underline{\nu}^{\mathcal{F}_n}(\underline{r},\underline{s})$ and $\underline{\nu}_n^{\mathcal{F}_n}(\underline{r},\underline{s})$:

$$\underline{\nu}^{\mathcal{F}_n}(\underline{r},\underline{s}) = \mathbf{M}^{\mathbf{T}}\underline{\nu}^{\Delta\mathcal{F}_n}(\Delta\underline{r},\Delta\underline{s}), \quad \underline{\nu}_n^{\mathcal{F}_n}(\underline{r},\underline{s}) = \mathbf{M}^{\mathbf{T}}\underline{\nu}_n^{\Delta\mathcal{F}_n}(\Delta\underline{r},\Delta\underline{s}), \tag{4.9}$$

$(\Delta\underline{r},\Delta\underline{s})$ being the vector of increments

$$\{r_1, r_2 - r_1, \dots, r_h - r_{h-1}, s_1, \dots, s_k - s_{k-1}\},$$

with $\sum_{i=1}^{h+1}(r_i - r_{i-1}) = \sum_{j=1}^{k+1}(s_j - s_{j-1}) = 1, \;\; r_0 = s_0 = 0.$

The proof of (4.7) and (4.9) is only a matter of algebra and it is omitted.

Set $p_{i,j} = P\mathbf{1}_{i,j}$, for every $i \le h+1, \; j \le k+1$. Null hypothesis $P \in \Omega$ implies then $p_{i,j} = p_{i,\cdot}p_{\cdot,j}$. Write also $N_{i,j}$ for the frequency observed in the $(i,j)-$th cell and $N_{i,\cdot}, \; N_{\cdot,j}$ for the marginal frequencies.

For $\Delta\Sigma$ and $\Delta\Sigma_n$ (its empirical counterpart) we have the following result:

Lemma 4.1. *Let T be the diagonal matrix with $\{(i-1)(k+1)+j\}-$th diagonal component equal to $p_{i,j}$, $\mathbf{U}$ the block matrix with $(i,l)-$th block equal to*

$$U^{i,l} = \begin{cases} \left\{\sqrt{p_{i,j_1}p_{l,j_2}}\right\}_{j_1,j_2 \le k+1} & \textit{for } 1 \le i,l \le h \\ \left\{\sqrt{p_{i,j_1}p_{l,j_2}}\right\}_{j_1 \le k+1, \; j_2 \le k} & \textit{for } 1 \le i \le h, \; l = h+1 \\ \left\{\sqrt{p_{i,j_1}p_{l,j_2}}\right\}_{j_1 \le k, \; j_2 \le k+1} & \textit{for } 1 \le l \le h, \; i = h+1 \\ \left\{\sqrt{p_{i,j_1}p_{l,j_2}}\right\}_{j_1 \le k,; j_2 \le k} & \textit{for } i = h+1, \; l = h+1 \end{cases}$$

Write T_n and $\mathbf{U}_n$ for their corresponding empirical versions. Then

$$\Delta\Sigma = T^{1/2}(I - \mathbf{U})T^{1/2} \tag{4.10}$$

and

$$\Delta\Sigma_n = T_n^{1/2}(I - \mathbf{U}_n)T_n^{1/2}. \tag{4.11}$$

In particular, under H_0, (4.10) writes

$$\Delta\Sigma = T^{1/2}(I - D^{1/2}\tilde{\mathbf{U}}D^{1/2})T^{1/2} \tag{4.12}$$

where D is the diagonal block matrix with $i-th$ block equal to $D_i = \{p_{i,\cdot} \cdot I_{k+1}\}$ for $i \le k$, ($D_{k+1} = \{p_{k+1,\cdot} \cdot I_k\}$), while the $(i,l)-$th block of $\tilde{\mathbf{U}}$ is $\tilde{\mathbf{U}}^{i,l} = \left\{\sqrt{p_{\cdot,j_1}p_{\cdot,j_2}}\right\}_{j_1,j_2}$

Proof. We prove (4.12) only. The proof of the other identities can be obtained in a similar way.

Let $u = (i_1 - 1)(k+1) + j_1$, $v = (i_2 - 1)(k+1) + j_2$. Then $(u, v)-$th component of $\Delta\Sigma$ is equal to:

(4.13)
$$P\mathbf{1}_{i_1,j_1}\mathbf{1}_{i_2,j_2} - P\mathbf{1}_{i_1,j_1}P\mathbf{1}_{i_2,j_2} = \begin{cases} p_{i,j}(1-p_{i,j}) & \text{if } i_1 = i_2 = i,\ j_1 = j_2 = j \\ -p_{i_1,\cdot}p_{\cdot,j_1}p_{i_2,\cdot}p_{\cdot,j_2} & \text{otherwise} \end{cases}$$

On the other hand, the $(i_1, i_2)-$th block of $D^{1/2}\tilde{\mathbf{U}}D^{1/2}$ is $\sqrt{p_{i_1,\cdot}p_{i_2,\cdot}}\tilde{\mathbf{U}}^{i_1,i_2}$. Then the $(i_1, i_2)-$th block of $T^{1/2}(I - D^{1/2}\tilde{\mathbf{U}}D^{1/2})T^{1/2}$ is:

$$\left\{\sqrt{p_{i_1,\cdot}p_{\cdot,j}}\left(\delta_{i_1,i_2}I - \sqrt{p_{i_1,\cdot}p_{i_2,\cdot}}\sqrt{p_{\cdot,j}p_{\cdot,l}}\right)\sqrt{p_{i_2,\cdot}p_{\cdot,l}}\right\}_{j,l} \tag{4.14}$$

where indexes j and l vary from 1 to $k+1$ or to k according to i_1 and i_2 be less then or equal to h respectively and $\delta_{i,j}$ is the Kronecker delta function.

Hence, $(j_1, j_2)-$th component of (4.14) equals

$$\sqrt{p_{i_1,\cdot}p_{\cdot,j_1}}\sqrt{p_{i_2,\cdot}p_{\cdot,j_2}}\left(\delta_{i_1,i_2}\delta_{j_1,j_2} - \sqrt{p_{i_1,\cdot}p_{\cdot,j_1}}\sqrt{p_{i_2,\cdot}p_{\cdot,j_2}}\right)$$

which coincides with (4.13). □

¿From Lemma 4.1 we have

Corollary 4.1.

$$\Delta\Sigma^{-1} = T^{-1/2}\left(I + \frac{1}{p_{h+1,k+1}}\mathbf{U}\right)T^{-1/2}; \tag{4.15}$$

$$\Delta\Sigma_n^{-1} = T_n^{-1/2}\left(I + \frac{1}{N_{h+1,k+1}/n}\mathbf{U}_n\right)T_n^{-1/2}. \tag{4.16}$$

Moreover,

$$\lambda_1 \geq p_{h+1,k+1} \inf_{\substack{1 \leq i \leq h \\ 1 \leq j \leq k}} p_{i,j} \tag{4.17}$$

$$\lambda_m \leq \sup_{\substack{1 \leq i \leq h \\ 1 \leq j \leq k}} p_{i,j}. \tag{4.18}$$

Proof. The proof relies on series expansion of $((I - \mathbf{U}) = \left(\sum_{i=0}^{\infty}\mathbf{U}^i\right)$ and on the identity $\mathbf{U}^2 = (1 - p_{h+1,k+1})\mathbf{U}$ and is omitted since it can be derived by arguments similar to those in the proof of Lemma 3.10 of Broniatowski and Leorato (2004). □

Corollary 4.1 can be used to write χ_n^2 in explicit form. It is in fact possible to prove (see Lemma 3.14 in Broniatowski and Leorato (2004)) that

$$\chi_n^2(\underline{r}, \underline{s}) = \sum_{i=1}^{h+1}\sum_{j=1}^{k+1}\frac{\left(\frac{N_{i,j}}{n} - \Delta r_i\Delta s_j\right)^2}{\frac{N_{i,j}}{n}}; \tag{4.19}$$

$$\chi_{(n)}^2(\underline{r}, \underline{s}) = \sum_{i=1}^{h+1}\sum_{j=1}^{k+1}\frac{(p_{i,j} - \Delta r_i\Delta s_j)^2}{p_{i,j}} \quad \text{and} \quad \overline{\chi}_n^2 = \inf_{\Delta\underline{r},\Delta\underline{s}}\sum_{i=1}^{h+1}\sum_{j=1}^{k+1}\frac{(N_{i,j}/n - \Delta r_i\Delta s_j)^2}{p_{i,j}}.$$

In other words, χ_n^2 is the minimum modified chi-square test statistic under the constraints $\sum_{i=1}^{h+1}\Delta r_i = \sum_{j=1}^{k+1}\Delta s_j = 1$, $\Delta r_i \geq 0$, $\forall\, i$, $\Delta s_j \geq 0$, $\forall\, j$.

The sufficient conditions of Theorem 3.3 can be weakened once the class $\mathcal{F}$ is given. In particular, for $\mathcal{F}$ indexed by cells in $[0,1]^2$, as in the present example, condition (3.6) is stronger than (3.13) and is not necessary in order to attain the convergence result for $(n\chi_n^2 - hk)/\sqrt{2hk}$. The following Theorem proves this assertion.

Theorem 4.1. *If* $P \in \Omega$ *and*

$$n^{-1/2}(\log n)^2\sqrt{hk}\left(\inf_{i\leq h+1} p_{i,\cdot} \inf_{j\leq k+1} p_{\cdot,j}\right)^{-1/2} = o_P(1), \tag{4.20}$$

then

$$\frac{n\chi_n^2 - hk}{\sqrt{2hk}} \to N(0,1).$$

Proof. The proof is postponed to Section 5. □

Remark 8. The statistics χ_n^2 and $\overline{\chi}_n^2$ are closely related to X_n^2 and Y_n^2 proposed in Conti and Scanu (1998) for testing the independence hypothesis for lattice distributions. However it is easy to see that conditions (5) of Conti and Scanu i.e.

$$\frac{n^\delta}{\sqrt{h+1}}\min_{1\leq i\leq h+1} p_{i,\cdot} \to \infty \qquad \frac{n^\delta}{\sqrt{k+1}}\min_{1\leq j\leq k+1} p_{\cdot,j} \to \infty$$

for some $0 < \delta < 1/4$, imply (4.20).

4.1. **Large deviations under the null.** We now show a large deviation result of the Chernoff-type under the null hypothesis, for the test statistic defined by (4.19).

The main instrument used for the proof is Lemma 1 in Beirlant *et al.* (2001). We recall the definition of the Kullback-Leibler divergence for discrete distributions $Q = (q_1, \ldots, q_m)$ and $P = (p_1, \ldots, p_m)$:

$$I(Q,P) = \sum_{i=1}^m q_i \log\left(\frac{q_i}{p_i}\right).$$

The result of Theorem 4.2 below is indeed the same of Theorem 5 in Beirlant *et al.* (2001). However, we propose an alternative proof which avoids the use of the doubtful lower bound for $\inf_{\{Q:\chi_n^2(Q,P)\geq\varepsilon\}} I(Q,P)$ applied therein.

Theorem 4.2. *Let* $P = \left(\frac{1}{m_n}, \ldots, \frac{1}{m_n}\right)$ *be the uniform distribution and* $P_n = \left(\frac{N_1}{n}, \ldots, \frac{N_{m_n}}{n}\right)$ *the empirical measure, restricted to the sequence of partitions* $\mathfrak{A}_n = (A_1^{(n)}, \ldots, A_{m_n}^{(n)})$ *of the support* $\mathfrak{X}$,

$$\mathsf{X}_n^2 = \chi_n^2(P, P_n) = \sum_{i=1}^{m_n} \frac{\left(\frac{1}{m_n} - \frac{N_i}{n}\right)^2}{\frac{N_i}{n}}.$$

Then, if $\frac{m_n^2}{n}\log n \to 0$ *we have that*

$$\lim_{n\to\infty} \frac{m_n}{n} \log Pr\left\{\mathsf{X}_n^2 > \varepsilon\right\} = -1 \tag{4.21}$$

for all $\varepsilon > 0$.

Proof. The proof is deferred to next Section. □

As a corollary of Theorem 4.2 we first derive a large deviation result for the statistic

$$\hat{\chi}_n^2 = \sum_{i=1}^{h_n} \sum_{j=1}^{k_n} \frac{\left(\frac{1}{h_n k_n} - \frac{N_{i,j}}{n}\right)^2}{N_{i,j}/n} 1_{N_{i,j}>0}, \tag{4.22}$$

which corresponds to the case where the marginals of P are known (and cells are arranged in order to have uniformity).

Corollary 4.2. *Assume* $\frac{(h_n k_n)^2}{n} \log n \to 0$. *Then*

$$\lim_{n\to\infty} \frac{h_n k_n}{n} \log Pr\left\{\hat{\chi}_n^2 \geq \varepsilon\right\} = -1 \tag{4.23}$$

for all $\varepsilon > 0$.

Proof. It is enough to repeat the proof of Theorem 4.2 with $h_n k_n$ instead of m_n. In this case $Q_0 = (q_{1,1}, q_{1,2}, \ldots, q_{h_n,k_n}) = \left(0, \frac{1}{h_n k_n - 1}, \ldots, \frac{1}{h_n k_n - 1}\right)$ and Q^* is similarly defined from (5.28). □

We can finally write down the Chernoff-type result for χ_n^2:

Corollary 4.3. *Let* $\frac{h_n k_n}{n} \log n \to 0$. *Then*

$$\lim_{n\to\infty} \frac{(h_n k_n)^2}{n} \log Pr\left\{\chi_n^2 \geq \varepsilon\right\} = -1 \tag{4.24}$$

for all $\varepsilon > 0$.

Proof. We consider the set

$$\Gamma = \left\{ Q : \inf_{\underline{r},\underline{s}} \sum_{i,j} \frac{(r_i s_j - q_{i,j})^2}{q_{i,j}} \geq \varepsilon \right\}.$$

Then, using again Lemma 1 in Beirlant *et al.* (2001) and the condition $\frac{(h_n k_n)^2}{n} \log n \to 0$, we have

$$\lim_{n\to\infty} \frac{h_n k_n}{n} \log Pr\left\{\chi_n^2 \geq \varepsilon\right\} = -\lim_{n\to\infty} \inf_{Q\in\Gamma} I(Q,P). \tag{4.25}$$

For the lower bound of (4.25) we take the distribution Q^0 defined above. For the upper bound, we consider the set

$$\Gamma_p = \left\{ Q : \sum_{i,j} \frac{\left(\frac{1}{h_n k_n} - q_{i,j}\right)^2}{q_{i,j}} \geq \varepsilon \right\} \supseteq \Gamma.$$

Then the result follows by Corollary 4.2. □

5. Proofs

Proof of Theorem 2.1. We have

$$0 \leq \left|\chi_n^2 - \chi^2\right| \leq \left|\chi_n^2 - \overline{\chi}_n^2\right| + \left|\overline{\chi}_n^2 - \chi^2\right| \tag{5.1}$$

We note that

$$\begin{aligned}
\overline{\chi}_n^2 - \chi^2 &\leq \overline{\chi}_n^2(\underline{r}_0, \underline{s}_0) - \chi^2(\underline{r}_0, \underline{s}_0) \\
&= n^{-1/2}(\underline{\nu}_n^{\mathcal{F}}(\underline{r}_0, \underline{s}_0) + \underline{\nu}^{\mathcal{F}}(\underline{r}_0, \underline{s}_0))^T \Sigma^{-1} \underline{\gamma}_n \\
&\leq 2n^{-1/2} \left\| \underline{\nu}^{\mathcal{F}}(\underline{r}_0, \underline{s}_0))^T \Sigma^{-1/2} \right\| \left\| \Sigma^{-1/2} \underline{\gamma}_n \right\| + n^{-1} \left\| \underline{\gamma}_n^T \Sigma^{-1/2} \right\|^2 \\
&= O_P\left(n^{-1/2}\sqrt{\chi^2} m^{1/2} + n^{-1} m\right) \qquad (5.2)
\end{aligned}$$

where $\|x\|^2 = \sum_{i=1}^m x_i^2$ is the Euclidean norm.

In the last step we have applied Chebyshev inequality to the r.v. $\underline{\gamma}_n^T \Sigma^{-1} \underline{\gamma}_n$, which, by CLT, converges to a chi-squared r.v. with $m = [(h+1)(k+1) - 1]$ degrees of freedom.

Since Σ is definite positive the above inequality implies:

$$0 \leq \overline{\chi}_n^2 \leq \chi^2 \left(1 + O_P\left(n^{-1/2} m^{1/2} \left(\chi^2\right)^{-1/2}\right)\right) + O_P(n^{-1} m).$$

For the opposite inequality we have:

$$\begin{aligned}
\chi^2 - \overline{\chi}_n^2 &\leq n^{-1/2} 2 \left\| \underline{\nu}_n^{\mathcal{F}}(\overline{r}_n, \overline{s}_n)^T \Sigma^{-1/2} \right\| \left\| \underline{\gamma}_n^T \Sigma^{-1/2} \right\| + n^{-1} \left\| \underline{\gamma}_n^T \Sigma^{-1/2} \right\|^2 \\
&= O_P\left(n^{-1/2} \sqrt{\overline{\chi}_n^2} m^{1/2} + n^{-1} m\right) = o_P(1). \qquad (5.3)
\end{aligned}$$

By putting together (5.2) and (5.3) we obtain $\left|\overline{\chi}_n^2 - \chi^2\right| = o_P(1)$.

It remains to prove that the first term in (5.1) is negligible.

It then follows that

$$\begin{aligned}
\chi_n^2 - \overline{\chi}_n^2 &\leq \chi_n^2(\overline{r}_n, \overline{s}_n) - \overline{\chi}(\overline{r}_n, \overline{s}_n) = \underline{\nu}_n^{\mathcal{F}}(\overline{r}_n, \overline{s}_n)^T \left(\Sigma_n^{-1} - \Sigma^{-1}\right) \underline{\nu}_n^{\mathcal{F}}(\overline{r}_n, \overline{s}_n) \\
&= \underline{\nu}_n^{\mathcal{F}}(\overline{r}_n, \overline{s}_n)^T \Sigma^{-1/2} \left(\Sigma^{1/2} \Sigma_n^{-1} \Sigma^{1/2} - I\right) \Sigma^{-1/2} \underline{\nu}_n^{\mathcal{F}}(\overline{r}_n, \overline{s}_n) \\
&\leq \overline{\chi}_n^2 \left\|\left\| \Sigma^{1/2} \Sigma_n^{-1} \Sigma^{1/2} - I \right\|\right\|. \qquad (5.4)
\end{aligned}$$

Σ_n is positive semi-definite and this allows us to write, by (5.4), that $0 \leq \chi_n^2 \leq \overline{\chi}_n^2 \left(1 + O_P\left(\|\|\Sigma^{1/2} \Sigma_n^{-1} \Sigma^{1/2} - I\|\|\right)\right)$.

For the opposite inequality, we easily get $\overline{\chi}_n^2 - \chi_n^2 \leq \chi_n^2 \left\|\left\| \Sigma_n^{1/2} \Sigma^{-1} \Sigma_n^{1/2} - I \right\|\right\|$.

It follows that, if both norms $\|\|\Sigma^{1/2} \Sigma_n^{-1} \Sigma^{1/2} - I\|\|$ and $\left\|\left\|\Sigma_n^{1/2} \Sigma^{-1} \Sigma_n^{1/2} - I\right\|\right\|$ are $o_P(1)$, then we are done.

By some algebraic manipulations and using Taylor expansion for $f(x) = (1+x)^{-1}$ we have

$$\Sigma_n^{-1} = \Sigma^{-1/2} \left[I + \sum_{h=1}^{\infty} \left(\Sigma^{-1/2}(\Sigma - \Sigma_n)\Sigma^{-1/2}\right)^h\right] \Sigma^{-1/2}. \tag{5.5}$$

and

$$
\begin{aligned}
\left|\!\left|\!\left|\Sigma^{1/2}\Sigma_n^{-1}\Sigma^{1/2}-I\right|\!\right|\!\right| &\le \sum_{h=1}^{\infty} |\!|\!|\Sigma-\Sigma_n|\!|\!|^h \left|\!\left|\!\left|\Sigma^{-1/2}\right|\!\right|\!\right|^{2h} \\
&\le \frac{m}{\lambda_1}\left(\sup_{\substack{0\le i\le h\\0\le j\le k}}\left|(P_n-P)\xi_i^2\zeta_j^2\right| + 2\sup_{\substack{0\le i\le h\\0\le j\le k}}\left|(P_n-P)\xi_i\zeta_j\right|\sqrt{PH^2PK^2}\right)(1+o_P(1))
\end{aligned}
\tag{5.6}
$$

where $\lambda_1\le\lambda_2\le\ldots\le\lambda_m$ are the eigenvalues of Σ and where we have set $\xi_0=\mathbf{1}$ and $\zeta_0=\mathbf{1}$. The right hand side of (5.6) converges to zero $P-$a.s. from multidimensional LLN and by Condition (C2).

By repeating the same arguments used above, Σ_n and Σ exchanged, one can easily obtain

$$
\begin{aligned}
&\left|\!\left|\!\left|\Sigma_n^{1/2}\Sigma^{-1}\Sigma_n^{1/2}-I\right|\!\right|\!\right| \\
&\le m\lambda_{1,n}^{-1}\left(\sup_{i,j}\left|(P_n-P)\xi_i^2\zeta_j^2\right| + 2\sqrt{PH^2PK^2}\sup_{i,j}\left|(P_n-P)\xi_i\zeta_j\right|\right)\to 0
\end{aligned}
\tag{5.7}
$$

by LLN, where $\lambda_{1,n}$ denotes the smallest eigenvalue of Σ_n. □

Proof of Theorem 3.2. By (2.17), $tr(\mathbf{P})=rank(\mathbf{P})=h+k$ and $J_0\Sigma^{-1/2}\mathbf{P}=J_0\Sigma^{-1/2}$.

Define the sequence $(\underline{\tilde{r}}_n,\underline{\tilde{s}}_n)\in\Theta_{\mathcal{H}_n}\times\Theta_{\mathcal{K}_n}$ such that

$$
\sqrt{n}\left\{(\underline{\tilde{r}}_n,\underline{\tilde{s}}_n)-(\underline{r}_0,\underline{s}_0)\right\}^T J_0\Sigma^{-1/2}=\underline{\tilde{\tau}}_n^T J_0\Sigma^{-1/2}=\underline{\gamma}_n^T\Sigma^{-1/2}\mathbf{P}. \tag{5.8}
$$

We have already noticed that $\left|\underline{\tilde{\tau}}_n^T J_0\Sigma^{-1}J_0^T\underline{\tilde{\tau}}_n\right|=\left|\underline{\gamma}_n^T\Sigma^{-1/2}\mathbf{P}\Sigma^{-1/2}\underline{\gamma}_n\right|=O_P(h+k)$.

It holds also $\left|\underline{\tau}_n^T J_0\Sigma^{-1}J_0^T\underline{\tau}_n\right|=O_P(h+k)$. In fact, since (under H_0) $\chi_n^2(\underline{r}_0,\underline{s}_0)=n^{-1}\underline{\gamma}_n^T\sigma_n^{-1}\underline{\gamma}_n$ and using (2.12), we have

$$
\begin{aligned}
0 &\le \chi_n^2(\underline{r}_0,\underline{s}_0)-\chi_n^2(\underline{r}_n,\underline{s}_n)\\
&= -n^{-1}\left(\underline{\tau}_n^T J_0\Sigma^{-1}J_0^T\underline{\tau}_n-2\underline{\tau}_n^T J_0\Sigma^{-1}\underline{\gamma}_n\right)\\
&\quad -\left(B_n-n^{-1}\underline{\gamma}_n^T\Sigma^{-1/2}\left(\Sigma^{1/2}\Sigma_n^{-1}\Sigma^{1/2}-I\right)\Sigma^{-1/2}\underline{\gamma}_n\right)-C_n\\
&= -n^{-1}\left(\underline{\tau}_n^T J_0\Sigma^{-1}J_0^T\underline{\tau}_n-2\underline{\tau}_n^T J_0\Sigma^{-1}J_0^T\underline{\tilde{\tau}}_n\right)\\
&\quad -\left(B_n-n^{-1}\underline{\gamma}_n^T\Sigma^{-1/2}\left(\Sigma^{1/2}\Sigma_n^{-1}\Sigma^{1/2}-I\right)\Sigma^{-1/2}\underline{\gamma}_n\right)-C_n
\end{aligned}
$$

with B_n and C_n given by (2.13) and (2.14).

Complete the square and move it left hand side to get

$$
\begin{aligned}
&n^{-1}\left(\underline{\tau}_n-\underline{\tilde{\tau}}_n\right)^T J_0\Sigma^{-1}J_0^T\left(\underline{\tau}_n-\underline{\tilde{\tau}}_n\right)\\
&\le n^{-1}\underline{\tilde{\tau}}_n^T J_0\Sigma^{-1}J_0^T\underline{\tilde{\tau}}_n-C_n-\left(B_n-n^{-1}\underline{\gamma}_n^T\Sigma^{-1/2}\left(\Sigma^{1/2}\Sigma_n^{-1}\Sigma^{1/2}-I\right)\Sigma^{-1/2}\underline{\gamma}_n\right).
\end{aligned}
\tag{5.9}
$$

We can bound

$$\begin{aligned}
&\left|B_n - n^{-1}\underline{\gamma}_n^T\Sigma^{-1/2}\left(\Sigma^{1/2}\Sigma_n^{-1}\Sigma^{1/2}-I\right)\Sigma^{-1/2}\underline{\gamma}_n\right| \\
&\quad = \left|n^{-1}\underline{\tau}_n^T J_0\Sigma^{-1/2}\left(\Sigma^{1/2}\Sigma_n^{-1}\Sigma^{1/2}-I\right)\Sigma^{-1/2}J_0^T\underline{\tau}_n\right. \\
&\qquad\qquad \left. -2n^{-1}\underline{\tau}_n^T J_0\Sigma^{-1/2}\left(\Sigma^{1/2}\Sigma_n^{-1}\Sigma^{1/2}-I\right)\Sigma^{-1/2}\underline{\gamma}_n\right| \\
&\quad \le n^{-1}\left(\left|\underline{\tau}_n^T J_0\Sigma^{-1}J_0^T\underline{\tau}_n\right| + 2\left\|\underline{\tau}_n^T J_0\Sigma^{-1/2}\right\|\left\|\Sigma^{-1/2}\underline{\gamma}_n\right\|\right)\left\|\Sigma^{1/2}\Sigma_n^{-1}\Sigma^{1/2}-I\right\| \\
&\quad \le O_P\left(m^{5/4}n^{-1/2}\lambda_1^{-1}\right)n^{-1}O_P\left(\left(\left\|\underline{\tau}_n^T J_0\Sigma^{-1/2}\right\| + \left\|\tilde{\underline{\tau}}_n^T J_0\Sigma^{-1/2}\right\|\right)^2\right) \qquad (5.10)
\end{aligned}$$

and

$$\begin{aligned}
|C_n| &= \left|2n^{-1/2}\underline{\tau}_n^T J_0\Sigma^{-1}\underline{\mathsf{a}}_n + \underline{\mathsf{a}}_n^T\Sigma^{-1}\underline{\mathsf{a}}_n - 2n^{-1/2}\underline{\mathsf{a}}_n^T\Sigma^{-1}\underline{\gamma}_n\right| \\
&\quad + o_P\left(2n^{-1/2}\underline{\tau}_n^T J_0\Sigma^{-1}J_0^T\underline{\mathsf{a}}_n + \underline{\mathsf{a}}_n^T\Sigma^{-1}\underline{\mathsf{a}}_n - 2n^{-1/2}\underline{\mathsf{a}}_n^T\Sigma^{-1}\underline{\gamma}_n\right) \qquad (5.11)
\end{aligned}$$

To explain the last step in (5.10) we remark that

$$\begin{aligned}
\left\|\Sigma^{-1/2}\underline{\gamma}_n\right\| &= \frac{\left\|\Sigma^{-1/2}\underline{\gamma}_n\right\|}{\left\|\mathbf{P}\Sigma^{-1/2}\underline{\gamma}_n\right\|}\left\|\tilde{\underline{\tau}}_n^T J_0\Sigma^{-1/2}\right\| \\
&= (\text{by LLN}) = O_P\left(m^{1/4}\right)\left\|\tilde{\underline{\tau}}_n^T J_0\Sigma^{-1/2}\right\|
\end{aligned}$$

and thus, using (5.6) and $P-$Donsker property for $\mathcal{F}$,

$$\begin{aligned}
&\left(\left|\underline{\tau}_n^T J_0\Sigma^{-1}J_0^T\underline{\tau}_n\right| + 2\left\|\underline{\tau}_n^T J_0\Sigma^{-1/2}\right\|\left\|\Sigma^{-1/2}\underline{\gamma}_n\right\|\right)\left\|\Sigma^{1/2}\Sigma_n^{-1}\Sigma^{1/2}-I\right\| \\
&\le O_P(m^{1/4})\left(\left\|\underline{\tau}_n^T J_0\Sigma^{-1/2}\right\|^2 + 2\left\|\underline{\tau}_n^T J_0\Sigma^{-1/2}\right\|\left\|\tilde{\underline{\tau}}_n^T J_0\Sigma^{-1/2}\right\|\right)O_P\left(m\lambda_1^{-1}n^{-1/2}\right)
\end{aligned}$$

It is not difficult to deduce, by (2.10) and (2.11), that $n\left\|\underline{\mathsf{a}}_n\right\| \le O_P\left(\left\|\underline{\tau}_n^T J_0\right\|^2\right)$. In fact

$$\begin{aligned}
\left\|J_0 j_0^T\right\|^2 &= n\Big[(1+\|\underline{s}_0\|^2)\|\underline{r}_n-\underline{r}_0\|^2 + (1+\|\underline{r}_0\|^2)\|\underline{s}_n-\underline{s}_0\|^2 \\
&\qquad\qquad + 2\sum_i r_{0,i}(r_{n,i}-r_{0,i})\sum_j s_{0,j}(s_{n,j}-s_{0,j})\Big] \\
&\ge n\left(\|\underline{r}_n-\underline{r}_0\|^2 + \|\underline{s}_n-\underline{s}_0\|^2\right) + n\left(\|\underline{r}_0\|\|\underline{s}_n-\underline{s}_0\| + \|\underline{s}_0\|\|\underline{r}_n-\underline{r}_0\|\right)^2 \\
&\ge 2n\|\underline{r}_n-\underline{r}_0\|\|\underline{s}_n-\underline{s}_0\|
\end{aligned}$$

Then we can bound (5.11) by

$$\begin{aligned}
|C_n| &= O_P\left(n^{-1/2}\underline{\mathsf{a}}_n^T\Sigma^{-1}\underline{\gamma}_n\right) \\
&\le O_P\left(n^{-1}n^{-1/2}\left\|\underline{\tau}_n^T J_0\right\|^2\left\|\Sigma^{-1/2}\right\|\left\|\Sigma^{-1/2}\underline{\gamma}_n\right\|\right) \\
&\le O_P\left(n^{-1}\left(\underline{\tau}_n^T J_0\Sigma^{-1}J_0^T\underline{\tau}_n\right)n^{-1/2}m^{1/2}\lambda_1^{-1/2}\left\|\Sigma\right\|\right) \\
&= O_P\left(n^{-1}\left(\underline{\tau}_n^T J_0\Sigma^{-1}J_0^T\underline{\tau}_n\right)\right)O_P\left(n^{-1/2}m^{1/2}\lambda_1^{-1/2}\lambda_m\right). \qquad (5.12)
\end{aligned}$$

Note that in (5.12) we have used the inequality:

$$\frac{\left\|\underline{r}_n^T J_0\right\|^2}{\left\|\underline{r}_n^T J_0\Sigma^{-1/2}\right\|^2} = \frac{\underline{r}_n^T J_0 J_0^T \underline{r}_n}{\underline{r}_n^T J_0\Sigma^{-1} J_0^T \underline{r}_n} \leq \sup_x \frac{x^T J_0 J_0^T x}{x^T J_0 \Sigma^{-1} J_0^T x} = \sup_y \frac{y^T y}{y^T \Sigma^{-1} y} = \sup_x \frac{x^T \Sigma x}{x^T x}.$$

By inserting (5.10) and (5.12) into (5.9) we obtain

$$\begin{aligned}
0 &\leq n^{-1}\left(\left\|\underline{r}_n^T J_0\Sigma^{-1/2}\right\| - \left\|\underline{\tilde{r}}_n^T J_0\Sigma^{-1/2}\right\|\right)^2 \\
&\quad \times\left(1 + O_P\left(m^{5/4}\lambda_1^{-1}n^{-1/2}\right) + O_P\left(n^{-1/2}\lambda_m\lambda_1^{-1/2}m^{1/2}\right)\right) \\
&\leq n^{-1}\left(\underline{r}_n - \underline{\tilde{r}}_n\right)^T J_0\Sigma^{-1}J_0^T\left(\underline{r}_n - \underline{\tilde{r}}_n\right)(1 + o_P(1)) \\
&\leq n^{-1}\underline{\tilde{r}}_n^T J_0\Sigma^{-1}\underline{\tilde{r}}_n = O_P\left(n^{-1}m^{1/2}\right) \qquad (5.13)
\end{aligned}$$

where negligibility of the two $O_P(\cdot)$ terms above follows by (3.7) and by $n^{-1/2}\lambda_m\lambda_1^{-1/2}m^{1/2} = o_P\left(m^{5/4}\lambda_1^{-1}n^{-1/2}\right)$.

Write now

$$\begin{aligned}
0 &\geq n^{1/2}\left(\chi_n^2(\underline{r}_n, \underline{s}_n) - \chi_n^2(\underline{\tilde{r}}_n, \underline{\tilde{s}}_n)\right) \\
&= n^{-1/2}\left(\underline{r}_n - \underline{\tilde{r}}_n\right)^T J_0\Sigma^{-1}J_0^T\left(\underline{r}_n - \underline{\tilde{r}}_n\right) + n^{1/2}(B_n - \tilde{B}_n) + n^{1/2}(C_n - \tilde{C}_n), \qquad (5.14)
\end{aligned}$$

where $\tilde{B}_n$ and $\tilde{C}_n$ are obtained from (2.13) and (2.14) replacing $(\underline{r}_n, \underline{s}_n)$ with $(\underline{\tilde{r}}_n, \underline{\tilde{s}}_n)$.

By the same reasoning made above we obtain

$$\begin{aligned}
&\left|B_n - \tilde{B}_n\right| \\
&= \frac{1}{n}\Big|\underline{r}_n^T J_0\Sigma^{-1/2}\Big(\Sigma^{1/2}\Sigma_n^{-1}\Sigma^{1/2} - I\Big)\Sigma^{-1/2}J_0^T\underline{r}_n - 2\underline{r}_n^T J_0\Sigma^{-1/2}\Big(\Sigma^{1/2}\Sigma_n^{-1}\Sigma^{1/2} - I\Big)\Sigma^{-1/2}\underline{\gamma}_n \\
&\quad - \underline{\tilde{r}}_n^T J_0\Sigma^{-1/2}\Big(\Sigma^{1/2}\Sigma_n^{-1}\Sigma^{1/2} - I\Big)\Sigma^{-1/2}J_0^T\underline{\tilde{r}}_n + 2\underline{\tilde{r}}_n^T J_0\Sigma^{-1/2}\Big(\Sigma^{1/2}\Sigma_n^{-1}\Sigma^{1/2} - I\Big)\Sigma^{-1/2}\underline{\gamma}_n\Big| \\
&\leq O_P\left(m^{5/4}n^{-1/2}\lambda_1^{-1}\right)O_P\left(m^{1/2}n^{-1}\right)
\end{aligned}$$

and $\left|C_n - \tilde{C}_n\right| \leq |C_n| + |\tilde{C}_n| \leq o_P\left(\frac{1}{n}\left(\underline{r}_n^T J_0\Sigma^{-1}J_0^T\underline{r}_n + \underline{\tilde{r}}_n^T J_0\Sigma^{-1}J_0^T\underline{\tilde{r}}_n\right)\right) = o_P\left(n^{-1}m^{1/2}\right)$.

By the two bounds above we get

$$\Sigma^{-1/2}J_0^T\underline{r}_n = \Sigma^{-1/2}J_0^T\underline{\tilde{r}}_n + o_P\left(m^{1/4}\right) = \mathbf{P}\Sigma^{-1/2}\underline{\gamma}_n + o_P\left(\left\|\Sigma^{-1/2}J_0^T\underline{r}_n\right\|\right),$$

which yields the result. □

Proof of Theorem 3.3. Let ω^0 be a version of the $P-$Brownian bridge satisfying the strong invariance principle on $\mathcal{F}$ and $\underline{\omega}_n^0$ its restriction to $\mathcal{F}_n$.

Then we can write

$$\begin{aligned}\frac{n\chi_n^2 - hk}{\sqrt{2hk}} &= \frac{\underline{\gamma}_n{}^T \Sigma^{-1/2}[I - \Sigma^{-1/2} J_0^T (J_0\Sigma^{-1}J_0^T)^{-1} J_0 \Sigma^{-1/2}]\Sigma^{-1/2}\underline{\gamma}_n - hk}{\sqrt{2hk}} \\ &\quad + n\,\frac{B_n + C_n}{\sqrt{2hk}} \\ &= \frac{\underline{\omega}_n^0{}^T \Sigma^{-1/2}(I - \mathbf{P})\Sigma^{-1/2}\underline{\omega}_n^0 - hk}{\sqrt{2hk}} \\ &\quad -2(2hk)^{-1/2}\underline{\omega}_n^0{}^T \Sigma^{-1/2}(I - \mathbf{P})\Sigma^{-1/2}(\underline{\omega}_n^0 - \underline{\gamma}_n) \\ &\quad +(2hk)^{-1/2}(\underline{\omega}_n^0 - \underline{\gamma}_n)^T \Sigma^{-1/2}(I - \mathbf{P})\Sigma^{-1/2}(\underline{\omega}_n^0 - \underline{\gamma}_n) \\ &\quad +n(2hk)^{-1/2}(B_n + C_n) \\ &\stackrel{d}{=} \frac{\sum_{i=1}^{hk} Z_i^2 - hk}{\sqrt{2hk}} + D_n + E_n + n(2hk)^{-1/2}(B_n + C_n) \qquad (5.15)\end{aligned}$$

where $Z_i \sim N(0,1)$. Thus by CLT the first term converges weakly to the desired limit. It remains then to prove that the other terms are negligible.

By (3.8) and (5.10), we write

$$\begin{aligned}n(hk)^{-1/2}B_n &= (hk)^{-1/2}\underline{\gamma}_n^T \Sigma^{-1/2}(I - \mathbf{P})\left(\Sigma^{1/2}\Sigma_n^{-1}\Sigma^{1/2} - I\right)(I - \mathbf{P})\Sigma^{-1/2}\underline{\gamma}_n \\ &= O_P\left(\left\|\Sigma^{1/2}\Sigma_n^{-1}\Sigma^{1/2} - I\right\|\right)(hk)^{-1/2} O_P\left(\underline{\gamma}_n^T \Sigma^{-1/2}(I - \mathbf{P})\Sigma^{-1/2}\underline{\gamma}_n\right) \\ &= O_P\left(m^{3/2}\lambda_1^{-1} n^{-1/2}\right) = o_P(1),\end{aligned}$$

while (5.11) and (5.13) yield $nC_n(hk)^{-1/2} = o_P\left(m\lambda_1^{-1}n^{-1/2}\right)$.

The KMT property implies $E_n = o_P(D_n)$ and we have

$$\begin{aligned}D_n &\le 2(2hk)^{-1/2}\sqrt{\underline{\omega}_n^0 \Sigma^{-1/2}(I - \mathbf{P})\Sigma^{-1/2}\underline{\omega}_n^0} \\ &\quad \times\sqrt{(\underline{\omega}_n^0 - \underline{\gamma}_n)\Sigma^{-1/2}(I - \mathbf{P})\Sigma^{-1/2}(\underline{\omega}_n^0 - \underline{\gamma}_n)} \\ &\le 2(2hk)^{-1/2} O_P\left((hk)^{1/2}\right)\left\|\underline{\omega}_n^0 - \underline{\gamma}_n\right\| \left\|\Sigma^{-1/2}\right\| \|I - \mathbf{P}\|^{1/2} \\ &\le m^{1/2}\sup_{f\in\mathcal{F}}\left|\omega_n^0(f) - \gamma_n(f)\right| \lambda_1^{-1/2} O_P(1) \qquad (5.16)\end{aligned}$$

which follows by $\|I - \mathbf{P}\| = 1$ and by $\left\|\underline{\omega}_n^0 - \underline{\gamma}_n\right\| \le m^{1/2}\sup_{f\in\mathcal{F}}\left|\omega_n^0(f) - \gamma_n(f)\right|$.

□

Proof of Theorem 3.4 (sketch). We deduce, by (2.9) and (2.10), the relation

$$\left(\underline{\eta}_P - \underline{\eta}_0\right)^T = \left\{(\underline{r}_P, \underline{s}_P) - (\underline{r}_0, \underline{s}_0)\right\}^T J_0 + \underline{\mathsf{a}}_P.$$

Set $\underline{\tau}_P = \sqrt{n}\left\{(\underline{r}_P, \underline{s}_P) - (\underline{r}_0, \underline{s}_0)\right\}$.

By (3.18) we can then write, after some calculations,

$$\begin{aligned}n\chi_n^2 &= \left(\underline{\tau}_n - \underline{\tau}_P\right)^T J_0\Sigma^{-1}J_0^T\left(\underline{\tau}_n - \underline{\tau}_P\right) - 2\left(\underline{\tau}_n - \underline{\tau}_P\right)^T J_0\Sigma^{-1}J_0^T\left(\underline{\gamma}_n + \underline{\varepsilon}_n\right) \\ &\quad +2\left(\underline{\gamma}_n + \underline{\varepsilon}_n\right) J_0\Sigma^{-1}J_0^T\left(\underline{\gamma}_n + \underline{\varepsilon}_n\right) + n\,B_{n,P} + n\,C_{n,P}. \qquad (5.17)\end{aligned}$$

Similarly to (5.10) and (5.11), we have

$$\begin{aligned} n\,B_{n,P} &= \left(\underline{\tau}_n - \underline{\tau}_P\right)^T J_0\Sigma^{-1/2}\left(\Sigma^{1/2}\Sigma_n^{-1}\Sigma^{1/2} - I\right)\Sigma^{-1/2}\left(\underline{\tau}_n - \underline{\tau}_P\right) \\ &\quad -2\left(\underline{\tau}_n - \underline{\tau}_P\right)^T J_0\Sigma^{-1/2}\left(\Sigma^{1/2}\Sigma_n^{-1}\Sigma^{1/2} - I\right)\Sigma^{-1/2}\left(\underline{\gamma}_n + \underline{\varepsilon}_n\right) \\ &\quad + \left(\underline{\gamma}_n + \underline{\varepsilon}_n\right)^T \Sigma^{-1/2}\left(\Sigma^{1/2}\Sigma_n^{-1}\Sigma^{1/2} - I\right)\Sigma^{-1/2}\left(\underline{\gamma}_n + \underline{\varepsilon}_n\right) \end{aligned}$$

and

$$\begin{aligned} n\,C_{n,P} &= \left(\underline{\mathsf{a}}_n - \underline{\mathsf{a}}_P\right)^T \Sigma^{-1}\left(\underline{\mathsf{a}}_n - \underline{\mathsf{a}}_P\right) - 2\left(\underline{\mathsf{a}}_n - \underline{\mathsf{a}}_P\right)^T \Sigma^{-1}\left(\underline{\tau}_n - \underline{\tau}_P\right) \\ &\quad + o_P\left(\left(\underline{\mathsf{a}}_n - \underline{\mathsf{a}}_P\right)^T \Sigma^{-1}\left(\underline{\mathsf{a}}_n - \underline{\mathsf{a}}_P\right) - 2\left(\underline{\mathsf{a}}_n - \underline{\mathsf{a}}_P\right)^T \Sigma^{-1}\left(\underline{\tau}_n - \underline{\tau}_P\right)\right) \end{aligned}$$

(where the second term is $o_P(\cdot)$, provided that $\left|\!\left|\!\left|\Sigma^{1/2}\Sigma_n^{-1}\Sigma^{1/2} - I\right|\!\right|\!\right| = o_P(1)$).

Set $(\underline{\tilde{r}}_n, \underline{\tilde{s}}_n)$ a $(h+k)-$th dimensional vector satisfying

$$\underline{\tilde{\tau}} - \underline{\tau}_P = \sqrt{n}\left\{(\underline{\tilde{r}}_n, \underline{\tilde{s}}_n) - (\underline{r}_P, \underline{s}_P)\right\} = \left(J_0\Sigma^{-1}J_0^T\right)^{-1} J_0\Sigma^{-1}\left(\underline{\gamma}_n + \underline{\varepsilon}_n\right).$$

Then, from (5.17) we get,

$$\begin{aligned} 0 &\leq \sqrt{n}\left\{\chi_n^2(\underline{\tilde{r}}_n, \underline{\tilde{s}}_n) - \chi_n^2\right\} \\ &= -\left\{\left(\underline{\tau}_n - \underline{\tau}_P\right)^T\left(J_0\Sigma^{-1}J_0^T\right) - \left(\underline{\gamma}_n + \underline{\varepsilon}_n\right)^T\Sigma^{-1}J_0^T\right\}\left(J_0\Sigma^{-1}J_0^T\right)^{-1} \\ &\qquad \times\left\{\left(J_0\Sigma^{-1}J_0^T\right)\left(\underline{\tau}_n - \underline{\tau}_P\right) - J_0\Sigma^{-1}\left(\underline{\gamma}_n + \underline{\varepsilon}_n\right)\right\} \\ &\quad + n(\tilde{B}_{n,P} - B_{n,P}) + n(\tilde{C}_{n,P} - C_{n,P}) \end{aligned} \tag{5.18}$$

Repeating the same reasoning exploited in the proof of Theorem 3.2 it is easy to find

$$O\left(\left|\underline{\tau}_n^T J_0\Sigma^{-1}J_0^T\,\underline{\tau}_n\right|\right) = O\left(\left|\underline{\tilde{\tau}}_n^T J_0\Sigma^{-1}J_0^T\,\underline{\tilde{\tau}}_n\right|\right).$$

By taking the derivative of

$$\chi^2(\underline{r}, \underline{s}) = \left(\underline{\eta}(\underline{r}, \underline{s}) - \underline{\eta}_P\right)^T \Sigma^{-1}\left(\underline{\eta}(\underline{r}, \underline{s}) - \underline{\eta}_P\right) - 2n^{-1/2}\underline{\varepsilon}_n^T\Sigma^{-1}\left(\underline{\eta}(\underline{r}, \underline{s}) - \underline{\eta}_P\right) + \underline{\varepsilon}_n^T\Sigma^{-1}\underline{\varepsilon}_n$$

with respect to $(\underline{r}, \underline{s})$, we see that the infimum is achieved in $(\underline{r}_0, \underline{s}_0)$ if and only if

$$\underline{\tau}_P = \sqrt{n}\left\{(\underline{r}_P, \underline{s}_P) - (\underline{r}_0, \underline{s}_0)\right\} = \left(J_0\Sigma^{-1}J_0^T\right)^{-1} J_0\Sigma^{-1}\left(\underline{\varepsilon}_n + \underline{\mathsf{a}}_P\right).$$

It thus follows that $\underline{\tau}_n^T J_0\Sigma^{-1}J_0^T\,\underline{\tau}_n = O_P\left((hk)^{1/2}\right)$.

Inequalities analogous to (5.11) for $C_{n,P}$ and $\tilde{C}_{n,P}$ allow to conclude that those terms are negligible, using the fact that $\|\underline{\mathsf{a}}_P\| = O_P\left(n^{-1/2}\|\underline{\tau}_P^T J_0\|^2\right)$.

Negligibility of $n\left|B_{n,P} - \tilde{B}_{n,P}\right|$ can be also derived adapting the arguments used in Theorem 3.2.

We can then write

$$\left(\underline{\tau}_n - \underline{\tau}_P\right) = \left(\underline{\tilde{\tau}}_n - \underline{\tau}_P\right) + o_P\left(m^{1/4}n^{-1/2}\right).$$

The above display yields

$$n\chi_n^2 = \left(\underline{\gamma}_n + \underline{\varepsilon}_n\right)^T \Sigma^{-1/2}(I - \mathbf{P})\Sigma^{-1/2}\left(\underline{\gamma}_n + \underline{\varepsilon}_n\right).$$

Write $\mu(\underline{\varepsilon})$ and $\sigma(\underline{\varepsilon})$ for the mean and variance of $n\chi_n^2$ conditionally to (3.17):

$$\mu(\underline{\varepsilon}) = kh + \underline{\varepsilon}^T\Sigma^{-1/2}(I - \mathbf{P})\Sigma^{-1/2}\underline{\varepsilon};$$

$$\sigma(\underline{\varepsilon}) = 2hk.$$

In order to prove (3.19), as was done in the Proof of Theorem 3.3, we now split the quantity at left hand side of (3.19) into several components:

$$\begin{aligned}
\frac{n\chi_n^2 - \mu(\underline{\varepsilon})}{\sqrt{2hk}} &= \frac{\left(\underline{\omega}_n^0 + \underline{\varepsilon} - \underline{\omega}_n^0 + \underline{\gamma}_n\right)^T \Sigma^{-1/2}(I-\mathbf{P})\Sigma^{-1/2}\left(\underline{\omega}_n^0 + \underline{\varepsilon} - \underline{\omega}_n^0 + \underline{\gamma}_n\right) - \mu(\underline{\varepsilon})}{\sqrt{2hk}} \\
&\quad + n(hk)^{-1/2}\left(B_{n,P} + C_{n,P}\right) \\
&= \frac{\left(\underline{\omega}_n^0 + \underline{\varepsilon}\right)^T \Sigma^{-1/2}(I-\mathbf{P})\Sigma^{-1/2}\left(\underline{\omega}_n^0 + \underline{\varepsilon}\right) - hk - \underline{\varepsilon}^T\Sigma^{-1/2}(I-\mathbf{P})\Sigma^{-1/2}\underline{\varepsilon}}{\sqrt{2hk}} \\
&\quad + n(hk)^{-1/2}\left(B_{n,P} + C_{n,P}\right) \\
&\quad + D_n + E_n - 2\frac{\left(\underline{\omega}_n^0 - \underline{\gamma}_n\right)^T\Sigma^{-1/2}(I-\mathbf{P})\Sigma^{-1/2}\underline{\varepsilon}}{\sqrt{2hk}}
\end{aligned}$$

where D_n and E_n coincide with those in (5.15). It follows by the same arguments used in Theorem 3.3 that the term $(hk)^{-1/2}\left(B_{n,P} + C_{n,P}\right)$ is $o_P(n^{-1})$, as well as the terms in the last line of the above display, because of (3.13) and $\underline{\varepsilon}_n^T\Sigma^{-1/2}(I-\mathbf{P})\Sigma^{-1/2}\underline{\varepsilon}_n = O_P((hk)^{1/2})$ and also because of KMT property for $\mathcal{F}$.

The first term is distributed according to a non-central chi-square law, with degrees of freedom equal to hk, and noncentrality coefficient equal to $\underline{\varepsilon}^T\Sigma^{-1/2}(I-\mathbf{P})\Sigma^{-1/2}\underline{\varepsilon}$. Then, using standard CLT we get (3.19). □

Proof of Theorem 4.1. We have

$$\frac{n\chi_n^2 - hk}{\sqrt{2hk}} = \frac{n\overline{\chi}_n^2 - hk}{\sqrt{2hk}} + (2hk)^{-1/2} n(\chi_n^2 - \overline{\chi}_n^2). \tag{5.19}$$

It is enough to show that the first term in right hand side of (5.19) converges weakly to a standard normal r.v. while the second term is negligible.

First we prove convergence of $\frac{n\overline{\chi}_n^2 - hk}{\sqrt{2hk}}$.

By $\overline{\chi}_n^2 = \inf_{\underline{r},\underline{s}} \underline{\nu}_n^{\mathcal{F}_n}(\underline{r},\underline{s})^T\Sigma^{-1}\underline{\nu}_n^{\mathcal{F}_n}(\underline{r},\underline{s})$ and following the lines of the proof of Theorem 3.2, we can write:

$$\begin{aligned}
0 &\geq \sqrt{n}\left(\overline{\chi}_n^2(\overline{\underline{r}}_n, \overline{\underline{s}}_n) - \overline{\chi}_n^2(\tilde{\underline{r}}_n, \tilde{\underline{s}}_n)\right) \\
&= n^{-1/2}(\overline{\underline{\tau}}_n - \tilde{\underline{\tau}}_n)^T J_0\Sigma^{-1}J_0^T(\overline{\underline{\tau}}_n - \tilde{\underline{\tau}}_n) + n^{1/2}\overline{C}_n
\end{aligned}$$

where $\overline{C}_n = 2n^{-1/2}\overline{\underline{\tau}}_n^T J_0\Sigma^{-1}\overline{\underline{\mathsf{a}}}_n - 2n^{-1/2}\underline{\gamma}_n^T\Sigma^{-1}\overline{\underline{\mathsf{a}}}_n + \overline{\underline{\mathsf{a}}}_n^T\Sigma^{-1}\overline{\underline{\mathsf{a}}}_n$, $\overline{\underline{\mathsf{a}}}_n = \underline{\mathsf{a}}_n(\overline{\underline{r}}_n, \overline{\underline{s}}_n)$ and $\overline{\underline{\tau}}_n = \sqrt{n}\left((\overline{\underline{r}}_n, \overline{\underline{s}}_n) - (\underline{r}_0, \underline{s}_0)\right)$.

Observe that

$$\|\overline{C}_n\| \leq O_P\left(n^{-1/2}m^{1/2}\lambda_1^{-1/2}\lambda_m\right) O_P\left(n^{-1}\overline{\underline{\tau}}_n^T J_0\Sigma^{-1}J_0^T\overline{\underline{\tau}}_n\right)$$

and that

$$n^{-1/2}m^{1/2}\lambda_1^{-1/2}\lambda_m \leq n^{-1/2}m^{1/2}\left(p_{h+1,k+1}\inf_{i,j} p_{i,j}\right)^{-1/2}.$$

It then follows that $\overline{\underline{\tau}}_n = \tilde{\underline{\tau}}_n + o_P\left(\|\overline{\underline{\tau}}_n\|\right)$, which, by performing a decomposition analogous to (2.12) for $\overline{\chi}_n^2$ yields:

$$n\overline{\chi}_n^2 = \underline{\gamma}_n^T \Sigma^{-1/2}(I - \mathbf{P})\Sigma^{-1/2}\underline{\gamma}_n + n\overline{C}_n + o_P\left(m^{1/2}\right). \tag{5.20}$$

Then, following the notation of (5.15), we write

$$\frac{n\overline{\chi}_n^2 - hk}{\sqrt{2kh}} \stackrel{d}{=} \sum_{i=1}^{hk} \frac{Z_i^2 - 1}{\sqrt{2hk}} + D_n + E_n + m^{-1/2} n\overline{C}_n + o_P(1). \tag{5.21}$$

Condition (4.20) implies that D_n and E_n are $o_P(1)$, as shown in (5.16), while the first term converges to a standard normal r.v. by CLT.

As to $n\overline{C}_n$, we have already pointed out that $n\overline{C}_n = o_P\left(m^{1/2}\right) = o_P\left(\underline{\overline{\tau}}_n^T J_0 \Sigma^{-1} J_0^T \underline{\overline{\tau}}_n\right)$.

We now have to prove that the last term in (5.19) is negligible.

To this aim, we exploit representations (4.19) and (4) to obtain the following bounds:

$$\begin{aligned}
\frac{n}{\sqrt{hk}}\left(\chi_n^2 - \overline{\chi}_n^2\right) &\leq \frac{n}{\sqrt{hk}}\left(\chi_n^2(\underline{\overline{r}}_n, \underline{\overline{s}}_n) - \overline{\chi}_n^2(\underline{\overline{r}}_n, \underline{\overline{s}}_n)\right) \\
&= \frac{n}{hk}\sqrt{hk}\sum_{i=1}^{h+1}\sum_{j=1}^{k+1}\left(\frac{N_{i,j}}{n} - \Delta\overline{r}_{n,i}\Delta\overline{s}_{n,j}\right)^2\left(\frac{n}{N_{i,j}} - \frac{1}{p_{i,j}}\right) \\
&\leq \frac{n\overline{\chi}_n^2}{hk}\sqrt{hk}\max_{i,j}\left(\frac{np_{i,j}}{N_{i,j}} - 1\right);
\end{aligned} \tag{5.22}$$

$$\begin{aligned}
\frac{n}{\sqrt{hk}}\left(\chi_n^2 - \overline{\chi}_n^2\right) &= \frac{n}{\sqrt{hk}}\left(\min_{\underline{r},\underline{s}} \chi_n^2(\underline{r}, \underline{s}) - \min_{\underline{r},\underline{s}} \overline{\chi}_n^2(\underline{r}, \underline{s})\right) \\
&\geq \frac{n}{\sqrt{hk}} \min_{\underline{r},\underline{s}}\left(\chi_n^2(\underline{r}, \underline{s}) - \overline{\chi}_n^2(\underline{r}, \underline{s})\right) \\
&\geq \frac{n}{hk}\sqrt{hk}\min_{i,j}\left(\frac{np_{i,j}}{N_{i,j}} - 1\right)\overline{\chi}_n^2.
\end{aligned} \tag{5.23}$$

Taking into account that $n\overline{\chi}_n^2 = O_P(hk)$ and inequalities (5.22) and (5.23), it follows that if we prove $\sqrt{hk}\max_{i,j}\left|\frac{p_{i,j}}{N_{i,j}/n} - 1\right| \to 0$, then we are done.

We use inequality 10.3.2 p.415 in Shorack and Wellner (1986): for every i, j,

$$P\left\{\frac{N_{i,j}}{np_{i,j}} \geq x\right\} \leq \exp\left\{-np_{i,j}h(x)\right\}, \quad x \geq 1 \tag{5.24}$$

$$P\left\{\frac{np_{i,j}}{N_{i,j}} \geq x\right\} \leq \exp\left\{-np_{i,j}h\left(\frac{1}{x}\right)\right\} \quad x \geq 1 \tag{5.25}$$

with $h(x) = x\log x - x + 1 \geq 0$ for every $x \neq 1$. The steps are the same as in the proof of Theorem 3.15 of Broniatowski and Leorato (2004) and therefore are omitted.

□

Proof of Theorem 4.2. The proof follows the lines of Theorem 3 in Beirlant *et al.* (2001). Define the set $\Gamma = \left\{Q : \chi_n^2(P, Q) \geq \varepsilon\right\}$ and let $\mathfrak{L}_n$ be the set of measures having support on $(A_1^{(n)}, \dots, A_{m_n}^{(n)})$. By using Lemma 1 in Beirlant *et al.* (2001), we can write

$$\left| \frac{m_n}{n} \log Pr\left\{\chi_n^2 \geq \varepsilon\right\} + m_n \inf_{Q \in \Gamma \cap \mathfrak{L}_n} I\left(Q, P\right) \right| \leq \frac{m_n^2}{n} \log(n+1),$$

which can be rewritten as

$$\begin{aligned} \lim_{n\to\infty} \frac{m_n}{n} \log Pr\left\{\chi_n^2 \geq \varepsilon\right\} &= -\lim_{n\to\infty} \inf_{Q\in\Gamma\cap\mathfrak{L}_n} I(Q,P) \\ &= -\lim_{n\to\infty} \inf_{Q\in\Gamma} I(Q,P). \end{aligned} \tag{5.26}$$

For the upper bound of (5.26) we consider the distribution (24) in Beirlant *et al.* (2001):

$$Q_0 = \left(0, \frac{1}{m_n - 1}, \dots, \frac{1}{m_n - 1}\right) \tag{5.27}$$

which yields (see Beirlant *et al.* (2001))

$$m_n \inf_{Q\in\Gamma} I(Q,P) \leq m_n I(Q_0, P) \to 1.$$

For the converse inequality, we consider the distribution

$$\overline{Q} = \left(c, \frac{1-c}{m_n - 1}, \dots, \frac{1-c}{m_n - 1}\right), \tag{5.28}$$

where c is such that $\chi^2(P, \overline{Q}) = \varepsilon$, that means, $c = \frac{2+\varepsilon m_n - \sqrt{\varepsilon^2 m_n^2 + 4\varepsilon m_n - 4\varepsilon}}{2m_n(1+\varepsilon)}$.

Let $Q^* = \arg\inf_{Q\in\Gamma} I(Q,P)$ and assume that $Q^* = (q_1, \dots, q_{m_n})$, with $0 \leq q_1 \leq q_2, \leq \dots \leq q_{m_n}$. We want to prove that $Q^* = \overline{Q}$.

(a) $q_1 > 0$. This part follows without modifications from part (a) of Theorem 3 in Beirlant *et al.* (2001).

(b) There exist $1 \leq r_n < m_n$ such that $q_1 = \dots = q_{r_n} < q_{r_n+1} = \dots q_{m_n}$. Hypothesis (b) fails if either $q_1 = \dots = q_{m_n} = \frac{1}{m_n}$ or if there exist $1 \leq r_n < s_n \leq m_n$ for which $q_1 < q_{r_n} < q_{s_n}$. The first case gives $\chi^2(P, Q^*) = 0$, then $Q^* \notin \Gamma$. To show that the second case also leads to a contradiction, we proceed as follows: suppose, for simplicity $q_1 < q_2 < q_3$ (e.g. $r_n = 2$ and $s_n = 3$). We now build up a new distribution $\tilde{Q}$ coinciding with Q^* except for

$$\tilde{q}_1 = q_1 - \delta, \quad \tilde{q}_2 = q_2 + \delta t, \quad \tilde{q}_3 = q_3 - \delta(t-1),$$

where δ and t ($0 < \delta < q_1$, $0 < t < 1 + \frac{q_3}{q_1}$) are chosen such that $\chi^2(P, Q^*) = \chi^2(P, \tilde{Q})$. This means that

$$t = \frac{q_2^2}{q_1^2} \frac{q_3^2 - q_1^2}{q_3^2 - q_2^2} + o(1) \tag{5.29}$$

for $\delta \to 0$. Then

$$\begin{aligned} I(\tilde{Q}, P) &= (q_1 - \delta)\log((q_1 - \delta)m_n) + (q_2 + \delta t)\log((q_2 + \delta t)m_n) \\ &\quad + (q_3 - \delta(t-1))\log((q_3 - \delta(t-1))m_n) \\ &= I(Q^*, P) + \delta\left(t \log q_2 - \log q_1 - (t-1)\log q_3\right) + o(\delta). \end{aligned}$$

Using (5.29) we get

$$\begin{aligned} I(\tilde{Q},P)-I(Q^*,P) &= \frac{\delta}{2}\left(\frac{q_2^2}{q_1^2}\frac{q_3^2-q_1^2}{q_3^2-q_2^2}(\log q_2^2-\log q_3^2)-(\log q_1^2-\log q_3^2)\right)+o(\delta) \\ &= \frac{\delta}{2}\frac{1}{x(1-y)}\left(y(1-x)\log y - x(1-y)\log x\right)+o(\delta) \end{aligned}$$

with $0<x=\frac{q_1^2}{q_3^2}<y=\frac{q_2^2}{q_3^2}<1$. Therefore, since the function $\frac{x}{1-x}\log x$ is decreasing for x in $[0,1]$ we have $I(\tilde{Q},P)-I(Q^*,P)\le 0$ for δ small enough, which contradicts the hypothesis that Q^* reaches the infimum.

(c) $r_n=1$ (for all but finitely many n). From point (b) we have that

$$Q^*=\left(\frac{q_1}{r_n},\dots,\frac{q_1}{r_n},\frac{1-q_1}{m_n-r_n},\dots,\frac{1-q_1}{m_n-r_n}\right)$$

and that

$$I(Q^*,P)=q_1\log\frac{q_1 m_n}{r_n}+(1-q_1)\log\frac{(1-q_1)m_n}{m_n-r_n} \tag{5.30}$$

while for $\overline{Q}$, with $q_1=c$, we have

$$I(\overline{Q},P)=q_1\log q_1 m_n+(1-q_1)\log\frac{(1-q_1)m_n}{m_n-1}.$$

Some easy calculations permit us to write (5.30) as

$$I(Q^*,P)=I(\overline{Q},P)+I(Q^*,\overline{Q})+\frac{r_n-1}{r_n}q_1\log\left(1+\frac{1-q_1m_n}{q_1(m_n-1)}\right). \tag{5.31}$$

Therefore it is enough to take $q_1\le\frac{1}{m_n}$ to conclude

$$I(Q^*,P)-I(\overline{Q},P)\ge I(Q^*,\overline{Q})\ge 0.$$

Indeed, since

$$c=\frac{2+\varepsilon m_n-\sqrt{\varepsilon^2m_n^2+4\varepsilon m_n-4\varepsilon}}{2m_n(1+\varepsilon)}\le\frac{2+\varepsilon m_n-\sqrt{\varepsilon^2 m_n^2}}{2m_n(1+\varepsilon)}=\frac{1}{m_n(1+\varepsilon)}<\frac{1}{m_n},$$

we are able to say that the distribution (5.27) must be the one which attains the infimum in (5.26)

(d) It remains now to prove that $m_nI(\overline{Q},P)\to 1$. We use that fact that the function

$$I(q)=q\log qm_n+(1-q)\log\frac{(1-q)m_n}{m_n-1}$$

is monotone decreasing for $q\in(0,\frac{1}{m_n})$. Since $c\in[0,\frac{1}{m_n}]$ and, by using the following upper and lower bounds for the function $f(x)=\sqrt{1+x}$ (x near 0):

$$1+\frac{x}{2}-\frac{x^2}{4}\le f(x)\le 1+\frac{x}{2},$$

when $x=\frac{4(m_n-1)}{\varepsilon m_n^2}$, we get

$$c\ge\frac{2+\varepsilon m_n-\varepsilon m_n\left(1+\frac{4(m_n-1)}{2\varepsilon m_n^2}\right)}{2m_n(1+\varepsilon)}=\frac{1}{m_n^2(1+\varepsilon)}, \tag{5.32}$$

and

$$1-c\ge 1-\frac{1+2(m_n-1)^2/(m_n^3\varepsilon)}{m_n^2(1+\varepsilon)}. \tag{5.33}$$

Putting (5.32) and (5.33) into $I(\overline{Q}, P)$, we finally have

$$
\begin{aligned}
I(\overline{Q}, P) &\geq \frac{1}{m_n^2(1+\varepsilon)} \log \frac{1}{m_n(1+\varepsilon)} + \left(1 - \frac{1+O(\frac{1}{m_n})}{m_n^2(1+\varepsilon)}\right) \\
&\qquad \times \log \left(\frac{m_n}{m_n - 1} - \frac{1 + \frac{2(m_n-1)}{\varepsilon m_n^2}}{m_n^2(1+\varepsilon)} \right) \\
&= -\frac{\log((1+\varepsilon)m_n)}{m_n^2(1+\varepsilon)} + \left(1 + o\left(\frac{1}{m_n}\right)\right) \log \left(1 + \frac{1}{m_n - 1} + o\left(\frac{1}{m_n}\right)\right) \\
&\geq \left(1 - o\left(\frac{1}{m_n}\right)\right) \left(\frac{1}{m_n - 1} - \frac{1}{2(m_n-1)^2} + o\left(\frac{1}{m_n}\right)\right) + o\left(\frac{1}{m_n}\right),
\end{aligned}
$$

which entails

$$
m_n \inf_{Q \in \Gamma} I(Q, P) = m_n I(\overline{Q}, P) \geq 1 + o(1).
$$

□

Acknowledgements

The author is grateful to Prof. M. Broniatowski for having inspired the subject of the paper.

References

[1] J. Beirlant, L. Devroye, L. Györfi and I. Vajda. Large deviations of divergence measures on partitions. *J. Statist. Plann. Inference* **93**(2001), no. 1-2, 1–16.

[2] P.J. Bickel, Ya. Ritov and J.A. Wellner. Efficient estimation of linear functionals of a probability measure P with known marginal distributions. *Ann. Statist.* **19**(1991), no. 3, 1316–1346.

[3] I.S. Borisov. An approximation of empirical fields. *Nonparametric statistical inference*, Vol. I, II (Budapest, 1980), 77–87, Colloq. Math. Soc. János Bolyai, **32**, North-Holland, Amsterdam, 1982.

[4] M. Broniatowski. Estimation of the Kullback-Leibler divergence. *Math. Methods Statist.* **12**(2003), no. 4, 391–409 (2004).

[5] M. Broniatowski and A. Keziou. Parametric estimation and tests through divergences, Submitted, 2003.

[6] M. Broniatowski and S. Leorato. An estimation method for the chi-square divergence with application to test of hypotheses. Submitted, 2004.

[7] N. Cressie and T.R.C. Read. Multinomial goodness-of-fit tests. *J. Roy. Statist. Soc. Ser. B* **46**(1984), no. 3, 440–464.

[8] P.L. Conti and M. Scanu. Testing for independence in lattice distributions. *Math. Methods Statist.* **7**(1998), no. 4, 429–444 (1999).

[9] A. Dembo and O. Zeitouni. *Large deviations techniques and applications.* Jones and Bartlett Publishers, Boston, MA, 1993.

[10] L. Györfi and I. Vajda. A class of modified Pearson and Neyman statistics. *Statist. Decisions* **19**(2001), no. 3, 239–252.

[11] T. Inglot, W.C.M. Kallenberg and T. Ledwina. Asymptotic behavior of some bilinear functionals of the empirical process. *Math. Methods Statist.* **2**(1993), no. 4, 316–336.

[12] J. Komlós, P. Major and G. Tusnády. An approximation of partial sums of independent RV's and the sample DF. I. *Z. Wahrschein. und Verw. Gebiete* **32**(1975), 111–131.

[13] F. Liese and I. Vajda. *Convex statistical distances.* Teubner-Texte zur Mathematik [Teubner Texts in Mathematics], 95. BSB B. G. Teubner Verlagsgesellschaft, Leipzig, 1987.

[14] P. Massart. Strong approximation for multivariate empirical and related processes, via KMT constructions. *Ann. Probab.* **17**(1989), no. 1, 266–291.

[15] D, Morales, L. Pardo and I. Vajda. Some new statistics for testing hypotheses in parametric models. *J. Multivariate Anal.* **62**(1997), no. 1, 137–168

[16] 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.

[17] M. Teboulle and I. Vajda. Convergence of best ϕ-entropy estimates. *IEEE Trans. Inform. Theory* **39**(1993), no. 1, 297–301.

[18] G. Tusnády. A remark on the approximation of the sample DF in the multidimensional case. *Period. Math. Hungar.* **8**(1977), no. 1, 53–55.

[19] A.W. van der Vaart and J.A. Wellner. *Weak convergence and empirical processes. With applications to statistics.* Springer Series in Statistics. Springer-Verlag, New York, 1996.