

Inference for the Mean of Large p Small n Data: a Finite-Sample High-Dimensional Generalization of Hotelling's Theorem

Piercesare Secchi, Aymeric Stamm, Simone Vantini

▶ **To cite this version:**

Piercesare Secchi, Aymeric Stamm, Simone Vantini. Inference for the Mean of Large p Small n Data: a Finite-Sample High-Dimensional Generalization of Hotelling's Theorem. Electronic Journal of Statistics , 2013, 7, pp.2005-31. 10.1214/13-EJS833 . inserm-00858214

HAL Id: inserm-00858214

https://inserm.hal.science/inserm-00858214

Submitted on 4 Sep 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Electronic Journal of Statistics
Vol. 7 (2013) 2005–2031
ISSN: 1935-7524
DOI: 10.1214/13-EJS833

Inference for the mean of large p small n data: A finite-sample high-dimensional generalization of Hotelling's theorem

Piercesare Secchi,

MOX – Department of Mathematics "Francesco Brioschi", Politecnico di Milano
Piazza Leonardo da Vinci, 32, 20133, Milano, Italy
e-mail: piercesare.secchi@polimi.it

Aymeric Stamm

IRISA, University of Rennes I
UMR CNRS-6074, Campus de Beaulieu, F-35042 Rennes, France
e-mail: aymeric.stamm@irisa.fr

and

Simone Vantini

MOX – Department of Mathematics "Francesco Brioschi", Politecnico di Milano
Piazza Leonardo da Vinci, 32, 20133, Milano, Italy
e-mail: simone.vantini@polimi.it

Abstract: We provide a generalization of Hotelling's Theorem that enables inference (*i*) for the mean vector of a multivariate normal population and (*ii*) for the comparison of the mean vectors of two multivariate normal populations, when the number p of components is larger than the number n of sample units and the (common) covariance matrix is unknown. In particular, we extend some recent results presented in the literature by finding the (finite-n) p-asymptotic distribution of the Generalized Hotelling's T^2 enabling the inferential analysis of large-p small-n normal data sets under mild assumptions.

AMS 2000 subject classifications: 62H15, 62H10.
Keywords and phrases: Large p small n data, High-dimensional data, Hotelling's Theorem, Inference for the mean, MANOVA.

Received August 2012.

Contents

1 Introduction 2006
2 p-asymptotic generalized Hotelling's theorem 2008
3 Inference for the mean of large p small n data 2011
3.1 Connections between the generalized Hotelling's T^2 and the student's t test statistic 2013
3.2 Invariance properties of the generalized Hotelling's T^2 2013

3.3 On the p-asymptotic law of the generalized Hotelling's T^2 . . . 2014
3.4 p-asymptotic confidence region and hypothesis test for the mean of a normal population when $p \gg n$ . . . 2015
3.5 p-asymptotic pooled confidence region and hypothesis test for comparing the means of two normal populations when $p \gg n$. 2018
4 Simulation study . . . 2020
4.1 Comparison between the p-asymptotic generalized Hotelling's test and the identity-matrix generalized Hotelling's test . . . 2021
4.2 Comparison between the p-asymptotic generalized Hotelling's test and the n-p-asymptotic generalized Hotelling's test . . . 2024
4.3 Comparison between the p-asymptotic generalized Hotelling's test and the Pesarin-Salmaso's test . . . 2025
5 Discussion . . . 2027
A Some useful properties of the Moore-Penrose generalized inverse . . 2028
Acknowledgements . . . 2030
References . . . 2030

1. Introduction

The advent and development of high precision data acquisition technologies in active fields of research (e.g., medicine, engineering, climatology, economics), that are able to capture real-time and/or spatially-referenced measures, have provided the scientific community with large amount of data that challenge the classical approach to data analysis.

Data sets are indeed increasingly becoming characterized by a number of random variables that is much larger than the number of sample units (large p small n data sets) in contrast to the "familiar" data sets where the number of sample units is often much larger than the number of random variables (small p large n data sets). This makes many classical inferential tools (e.g. Hotelling's Theorem) almost useless in many fields at the forefront of scientific research and raises the demand for new inferential tools able to efficiently deal with this new kind of data.

The work of Srivastava (2007) is pioneering in this direction. In it, a generalization of the Hotelling's Theorem is proposed: a generalized T^2 test statistic is found and its distribution law is computed for $p \geq n$ under the assumptions of normality and of proportionality of the covariance matrix to the identity matrix with the proportionality constant unknown (Theorem 2.2, Srivastava (2007)); this assumption implies the independence among components (and among univariate test statistics as well), enabling also other classical inference procedures. We shall show that without relying on the latter assumption, it is possible to generalize this work in a much less stringent framework. In Srivastava (2007), some inferential results not depending on strong assumptions on the covariance structure are presented as well, but, being asymptotic in both p and n, they are not suitable to perform inferential statistical analysis of large p small n data (Theorem 2.3, Srivastava (2007)).

Other methods to deal with the analysis of large p small n data are objects of statistical investigation. Approaches based on the joint use of univariate test statistics for each component to build multivariate inference procedures have already appeared in the literature also under the assumption of dependence among components and thus among univariate test statistics as well. These approaches rely on the correction of each univariate significance (or confidence) levels such that the global significance (or confidence) level approximates a desired value. In particular, we can distinguish between a priori corrections, based on widely valid theoretical results (e.g., Bonferroni correction), and a posteriori corrections, based on the empirical distribution of the univariate p-values (e.g., Benjamini and Yekutieli (2001), Storey (2003)).

Permutation tests provide a further alternative approach to the inference for large p small n data. Permutation tests are inferential procedures that are conditional (the focus is on the sample rather than on the population), and distribution-free (no strong assumption about the population distribution law is necessary). Pesarin and Salmaso (2010, 2009) and Hall and Keilegorn (2007) recently proposed the use of permutation tests in the framework of multivariate and functional data analysis (an extreme case of large p small n data, Ramsay and Silverman (2005), Ferraty and Vieu (2006)).

Similarly to Srivastava (2007) and differently from the other works cited above, our proposal is Hotelling-inspired. In particular, to overcome the impossibility of treating large p small n data by means of a classical model-based approach, our strategy focuses on the random "variability space explored by the data", i.e., the space generated by the first $n-1$ principal components. In this reduced space, the proposed analysis is almost classical with the important distinction that the randomness of this data-dependent reduced space is fully taken into account. A pairwise comparison between our new inferential procedure and the other approaches presented above is of sure interest. Thus, in the present work, the theoretical and empirical comparison with the inferential approach proposed in Srivastava (2007) and with the one proposed in Pesarin and Salmaso (2010, 2009) is carried out. These are two approaches to the same problem that are very close in spirit and far from ours, respectively.

The paper is outlined as follows: in Section 2, after the introduction of the probabilistic framework, a generalized version of the Hotelling's Theorem for $p \to \infty$ is proposed; part of its proof is reported in Appendix A. In Section 3, the previous generalized version of the Hotelling's Theorem is used for the inference for the mean vector of a p-variate normal population (and for the difference of the mean vectors of two p-variate normal populations) when the number p of components is far larger than the number n of sample units; a theoretical comparison with the classical Hotelling's Theorem is here undertaken. In Section 4, by means of MC simulations, our new inferential procedure is compared with the two presented in Srivastava (2007) and in Pesarin and Salmaso (2010, 2009). In the final Section, some hints for possible future investigations are reported; in particular, a conjecture supported both by the results of MC simulations and by its theoretical consistency with our generalization of the Hotelling's Theorem and with the results presented in Srivastava (2007) is proposed.

2. *p*-asymptotic generalized Hotelling's theorem

The classical approach to inference for the mean μ_p of a p-variate normal random vector with unknown full rank covariance matrix Σ_p relies on a famous corollary of the Hotelling's Theorem that holds when the number n of sample units is larger than the number p of random vector components.

Theorem 1 (Hotelling's Theorem)**.** *For $m \geq 1$ and $p \geq 1$, assume that:*

(i) $\mathbf{X} \sim N_p(\mu_p, \Sigma_p)$*;*
(ii) $W \sim Wishart_p(\Sigma_p, m)$*;*
(iii) $\mathbf{X}$ *and* W *are independent.*

Then, for $m \geq p$:

$$\frac{m-p+1}{p}(\mathbf{X}-\mu_p)'W^{-1}(\mathbf{X}-\mu_p) \sim F(p, m-p+1).$$

Corollary 2 (Hotelling's T^2 Distribution Law)**.** *For $n \geq 2$ and $p \geq 1$, assume that:*

(i') $\{\mathbf{X}_i\}_{i=1,\ldots,n} \sim iid\ N_p(\mu_p, \Sigma_p)$.

Then, for $n > p$:

$$\frac{(n-p)n}{(n-1)p}(\overline{\mathbf{X}}-\mu_p)'S^{-1}(\overline{\mathbf{X}}-\mu_p) \sim F(p, n-p),$$

with $\overline{\mathbf{X}}$ and S being the sample mean and the sample covariance matrix, respectively.

The quantity $n(\overline{\mathbf{X}}-\mu_p)'S^{-1}(\overline{\mathbf{X}}-\mu_p)$ is known as Hotelling's T^2 due to its analogy with the squared of the univariate Student's t test statistic. Corollary 2 makes possible the development of inferential tools for the mean value of a p-variate normal random vector (e.g. confidence ellipsoidal regions or hypothesis testing) when the number n of sample units is larger than the number p of random vector components; there are no assumptions on the covariance matrix Σ_p that is only required to be positive definite. Proofs of Theorem 1 and Corollary 2 can be found, for instance, in Anderson (2003).

Theorem 1 and Corollary 2 become useless in applications where the covariance matrix is unknown and the number p of random vector components is larger than the number $n-1$, with n being the number of sample units. Indeed, in these cases, T^2 is not defined since S is not invertible because $rank(S) = \min(n-1, p)$ a.s. Analogously to Srivastava (2007), we decide to suitably generalize the inverse of S in order to obtain a suitable generalization of T^2. We considered the *Moore-Penrose Generalized Inverse* (Rao and Mitra (1971)) of a rectangular matrix, whose general definition can be found in Appendix A, since it always exists, it is unique, and it is equal to the inverse matrix when the latter is squared and invertible. Moreover, in the special case of a squared real positive

semi-definite matrix A, the Moore-Penrose generalized inverse A^+ can be proved (see Appendix A) to be equal to:

$$A^+ = \sum_{i:\lambda_i \neq 0} \frac{1}{\lambda_i} \mathbf{e}_i \mathbf{e}_i',$$

with $\{\lambda_i\}_{i=1,\ldots,p}$ and $\{\mathbf{e}_i\}_{i=1,\ldots,p}$ being the eigenvalues and eigenvectors of A, respectively.

We now present a generalization of Hotelling's Theorem that can be used to make inference for the mean of a multivariate normal random vector when the sample size n is finite, the number of components p goes to infinity, and the covariance matrix is unknown.

Theorem 3 (p-asymptotic Generalized Hotelling's Theorem)**.** *For $m \geq 1$ and $p \geq 1$, assume that:*

(i) $\mathbf{X} \sim N_p(\mu_p, \Sigma_p)$*;*
(ii) $W \sim Wishart_p(\Sigma_p, m)$*;*
(iii) $\mathbf{X}$ *and* W *are independent;*
(iv) $0 < \overline{\sigma} = \lim_{p\to\infty} \frac{tr(\Sigma_p)}{p} < +\infty$, $0 < \overline{\sigma^2} = \lim_{p\to\infty} \frac{tr(\Sigma_p^2)}{p} < +\infty$, *and* $0 < \overline{\sigma^4} = \lim_{p\to\infty} \frac{tr(\Sigma_p^4)}{p} < +\infty$.

Then, for $p \to \infty$:

$$\frac{\overline{\sigma}^2}{\overline{\sigma^2}} p(\mathbf{X} - \mu_p)' W^+ (\mathbf{X} - \mu_p) \xrightarrow{\mathcal{D}} \chi^2(m).$$

The proof of Theorem 3 is based on the p-asymptotic distribution of three auxiliary random matrices Y, L, H that provide alternative useful representation of the random matrix W appearing in (ii). In particular, Y is a $p \times m$ random matrix whose m columns are independent $N_p(\mathbf{0}, \Sigma_p)$, i.e. we may represent $W = YY'$ since the two random matrices have the same law; L is a random diagonal matrix whose diagonal elements are the m non-zero ordered eigenvalues of W; and H is a $m \times p$ random matrix whose rows are the corresponding m eigenvectors (i.e. $W = H'LH$ with $HH' = I_m$). Random matrices Y, L, and H exist almost surely since W is a Wishart random matrix with m degrees of freedom. Also the diagonal matrix $\Lambda_p = diag(\lambda_1(\Sigma_p), \ldots, \lambda_p(\Sigma_p))$ with $\lambda_1(\Sigma_p) \geq \cdots \geq \lambda_p(\Sigma_p) > 0$ being the ordered eigenvalues of Σ_p always exists thanks to the positive definiteness of Σ_p.

Theorem 3 relies, among others, on the fact that under assumptions (ii) and (iv) of the same theorem: $\frac{Y'Y}{p} \xrightarrow[p\to\infty]{\mathcal{P}} \overline{\sigma} I_m$, $\frac{L}{p} \xrightarrow[p\to\infty]{\mathcal{P}} \overline{\sigma} I_m$, and $H\Lambda_p H' \xrightarrow[p\to\infty]{\mathcal{D}} \frac{\overline{\sigma^2}}{\overline{\sigma}} I_m$. These p-asymptotic convergences are presented in Lemma A.2 (a), (b), and (d) of Srivastava (2007); their proofs rely on Chebyshev's Inequality, Prokhorov's Theorem, Slutsky's Theorem, and on algebraic relations supported by the properties of Moore-Penrose general inverses.

Proof of Theorem 3. Let us define two auxiliary matrices $A = (H\Lambda_p H')^{-1/2}$ and $\mathbf{Z} = AH(\mathbf{X} - \mu_p)$. The conditional distribution of $\mathbf{Z}$ given H is

$$\mathbf{Z}|H \sim N_m(\mathbf{0}_m, AH\Lambda_p H'A) = N_m(\mathbf{0}_m, I_m),$$

since $\mathbf{X}$ is distributed as $N_p(\mu_p, \Lambda_p)$. The conditional distribution of $\mathbf{Z}$ given H does not depend on H: therefore, $\mathbf{Z}$ and H are independent and

$$\mathbf{Z} \sim N_m(\mathbf{0}_m, I_m) \quad \text{while} \quad \mathbf{Z}'\mathbf{Z} \sim \chi^2(m).$$

Thanks to Proposition 9 in Appendix A, the following equalities in distribution hold:

$$\begin{aligned} & p(\mathbf{X} - \mu_p)' W^+ (\mathbf{X} - \mu_p) \quad = \\ = \quad & p(\mathbf{X} - \mu_p)' H' L^{-1} H (\mathbf{X} - \mu_p) \quad = \\ = \quad & p\mathbf{Z}'(ALA)^{-1}\mathbf{Z} \quad = \qquad (2.1) \\ = \quad & \left[\mathbf{Z}' \left(H\Lambda_p H'\right)^{1/2} \left(\frac{L}{p}\right)^{-1/2}\right] \left[\left(\frac{L}{p}\right)^{-1/2} \left(H\Lambda_p H'\right)^{1/2} \mathbf{Z}\right]. \qquad (2.2) \end{aligned}$$

Because of Lemma A.2 in Srivastava (2007) and the continuity of the maps $B \mapsto B^{-1/2}$ and $B \mapsto B^{1/2}$ over the set of positive definite matrices, we have that:

$$\begin{aligned} \left(\tfrac{L}{p}\right)^{-1/2} \quad & \xrightarrow[p\to\infty]{\mathcal{P}} \quad (\overline{\sigma})^{-1/2} I_m, \\ \left(H\Lambda_p H'\right)^{1/2} \quad & \xrightarrow[p\to\infty]{\mathcal{D}} \quad \left(\tfrac{\overline{\sigma^2}}{\overline{\sigma}}\right)^{1/2} I_m. \end{aligned}$$

Thus, Slutsky's Theorem (e.g., Serfling (2002)) implies that:

$$\left(\frac{L}{p}\right)^{-1/2} \left(H\Lambda_p H'\right)^{1/2} \mathbf{Z} \xrightarrow[p\to\infty]{\mathcal{D}} \left(\frac{\overline{\sigma^2}}{\overline{\sigma}^2}\right)^{1/2} \mathbf{Z}.$$

Finally, since the Euclidean squared norm function on $\mathbb{R}^m$ is continuous,

$$p(\mathbf{X} - \mu_p)' W^+ (\mathbf{X} - \mu_p) \xrightarrow[p\to\infty]{\mathcal{D}} \frac{\overline{\sigma^2}}{\overline{\sigma}^2} \mathbf{Z}'\mathbf{Z},$$

and thus

$$\frac{\overline{\sigma}^2}{\overline{\sigma^2}} p(\mathbf{X} - \mu_p)' W^+ (\mathbf{X} - \mu_p) \xrightarrow[p\to\infty]{\mathcal{D}} \chi^2(m). \qquad \square$$

Remarks about Theorem 3

1. The proof of Theorem 3 recalls the proof of Theorem 2.3 in Srivastava (2007) where a n-p-asymptotic result is proven (i.e., a result for both n and p going to infinity). That result is reported in Section 3 (paragraph Remarks about Corollary 4). The key point that allows us to obtain a (finite-n) p-asymptotic result for the distribution of $p(\mathbf{X}-\mu_p)'W^+(\mathbf{X}-\mu_p)$

is the use of the distributional equality (2.2) in place of the distributional equality (2.1), which is instead used in Srivastava (2007). Indeed, this new representation of the random quantity $p(\mathbf{X}-\mu_p)'W^+(\mathbf{X}-\mu_p)$ (identity (2.2)) makes unnecessary Lemma 2.1 of Srivastava (2007) (where an n-asymptotic result is proven) for identifying the distribution of $p(\mathbf{X}-\mu_p)'W^+(\mathbf{X}-\mu_p)$, leaving space for a finite-n result.

2. Note that in Theorem 3 the practical importance of "$p\to\infty$" (i.e., p-asymptoticity) is very general. For instance we might consider the situation where we add extra components to a random normal vector $\mathbf{X}$ and infinite extra rows and columns to a random Wishart matrix W. This is for instance the case of discrete-time series when time goes to infinity, or micro-array expressions when the number of genes goes to infinity. But "$p\to\infty$" can also be relevant in more complex situations where a sequence of random vectors $\mathbf{X}$ and of random matrices W of increasing dimensionality is investigated without any "nesting" property as p increases. This is for instance the case of sequential finite-dimensional representations of functional data by means of sequential non-necessarily nested basis (e.g. B-splines) whose dimension goes to infinity.
3. It is easy to prove that if the eigenvalues of Σ_p are uniformly bounded away from 0 and $+\infty$, i.e.:

   $(iv')\ \exists\ \underline{\lambda},\overline{\lambda}:\forall p, 0<\underline{\lambda}\le\lambda_1(\Sigma_p)\le\cdots\le\lambda_p(\Sigma_p)\le\overline{\lambda}<+\infty,$

   and at least one of the limits $\lim_{p\to\infty}\frac{tr(\Sigma_p)}{p}$, $\lim_{p\to\infty}\frac{tr(\Sigma_p^2)}{p}$ or $\lim_{p\to\infty}\frac{tr(\Sigma_p^4)}{p}$ exists, assumption (iv) of Theorem 3 is satisfied.

3. Inference for the mean of large p small n data

Provided that one can evaluate the ratio $\overline{\sigma}^2/\overline{\sigma^2}$ (this issue is tackled in subsection 3.3), Theorem 3 can be straightforwardly used to make inference for the mean of a multivariate normal distribution when the number p of components is larger than the number n of sample units. Indeed, its natural consequence is the following:

Corollary 4 (Generalized Hotelling's T^2 p-asymptotic distribution law)**.** *For $n\ge 2$ and $p\ge 1$, assume that:*

(i') $\{\mathbf{X}_i\}_{i=1,\dots,n}\sim iid\ N_p(\mu_p,\Sigma_p)$*;*

(iv) $0<\overline{\sigma}=\lim_{p\to\infty}\frac{tr(\Sigma_p)}{p}<+\infty$, $0<\overline{\sigma^2}=\lim_{p\to\infty}\frac{tr(\Sigma_p^2)}{p}<+\infty$, *and* $0<\overline{\sigma^4}=\lim_{p\to\infty}\frac{tr(\Sigma_p^4)}{p}<+\infty$.

Then, for $p\to\infty$:

$$\frac{\overline{\sigma}^2}{\overline{\sigma^2}}\frac{np}{n-1}(\overline{\mathbf{X}}-\mu_p)'S^+(\overline{\mathbf{X}}-\mu_p)\xrightarrow{\mathcal{D}}\chi^2(n-1),$$

where $\overline{\mathbf{X}}$ and S are the sample mean and the sample covariance matrix, respectively.

Proof. It is a direct application of Theorem 3 since $\sqrt{n}(\bar{\mathbf{X}} - \mu_p) \sim N_p(\mathbf{0}_p, \Sigma_p)$, $(n-1)S \sim Wishart_p(\Sigma_p, n-1)$, and they are independent. □

The random quantity

$$T^2 = n(\overline{\mathbf{X}} - \mu_p)' S^+ (\overline{\mathbf{X}} - \mu_p) \tag{3.1}$$

can be naturally denoted as Generalized Hotelling's T^2 since it is defined for any n and p such that $n \geq 2$ and $p \geq 1$ and coincides with the classical Hotelling's $T^2 = n(\overline{\mathbf{X}} - \mu_p)' S^{-1} (\overline{\mathbf{X}} - \mu_p)$ when $p < n$. Despite the simplicity of this generalization, important differences occur between the new framework $p \geq n$ and the classical framework $p < n$. These differences involve:

- the connection between T^2 and the univariate Student's t test statistic (subsection 3.1);
- the invariance properties of T^2 (subsection 3.2);
- the distribution law of T^2 (subsection 3.3);
- the geometrical characteristics of the confidence regions and of the critical regions for the mean that can be derived from Corollaries 2 and 4 (subsection 3.4).

Remarks about Corollary 4

1. The p-asymptotic distribution law of T^2 strongly depends on assumptions (i') and (iv) involving the normal distribution of the observations and the p-asymptotic behavior of the covariance matrix, respectively. In particular, while the latter assumption could probably be relaxed (this issue is still under investigations by the authors), the former assumption cannot since there is no Central Limit Theorem for $p \to \infty$ providing that $\sqrt{n}(\bar{\mathbf{X}} - \mu_p)$ is approximately normal and $(n-1)S$ is approximately a Wishart.
2. Corollary 4 (and similarly Corollary 5) generalizes to the finite-sample framework the result presented in Theorem 2.3 of Srivastava (2007) where, under the same assumptions (i') and (iv), the distribution of a transformation of T^2 is presented both for p and $n \to \infty$. In detail, Theorem 2.3 of Srivastava (2007) states that, under assumptions (i') and (iv), for p and $n \to \infty$:
$$\left(\frac{n-1}{2}\right)^{1/2} \left(\frac{p}{(n-1)^2} \frac{\overline{\sigma}^2}{\overline{\sigma^2}} T^2 - 1\right) \xrightarrow{\mathcal{D}} N(0,1).$$
In the light of Corollary 4, Theorem 2.3 of Srivastava (2007) turns out to be a special case which is obtained when $n \to \infty$ and the n-asymptotic normal approximation of the $\chi^2(n-1)$ distribution is used.
Even if our result covers a wider variety of real applications, it does not cover the analysis of small p small n data with $p > n$. This latter scenario is still an open issue except for the case of homoscedastic and independent components. This case is indeed fully developed in Srivastava (2007) where Theorem 2.2 states that under assumption (i') and $\Sigma_p = \gamma \mathbf{I}_p$, for $p \geq n$:
$$\frac{p-n+2}{(n-1)^2} T^2 \sim F(n-1, p-n+2).$$

3.1. Connections between the generalized Hotelling's T^2 and the student's t test statistic

Student's t statistic comes natural in multivariate statistics if the $\mathbb{R}^p$-representations of the n sample units are projected along a certain direction $\mathbf{a} \in \mathbb{R}^p \setminus \{\mathbf{0}_p\}$. Along this direction, the usual Student's t statistic can be computed:

$$t_{\mathbf{a}} = \sqrt{n}\frac{\mathbf{a}'(\bar{\mathbf{X}} - \mu_p)}{\sqrt{\mathbf{a}' S \mathbf{a}}} \sim t(n-1),$$

and univariate inference can be carried on along that direction.

Note that for all $\mathbf{a} \in \mathbb{R}^p \setminus \{\mathbf{0}_p\}$, $t_{\mathbf{a}}$ is almost surely defined. Indeed, since $ker(S)$ has null Lebesgue measure on $\mathbb{R}^p$ and since $\mathbf{X}_i$ with $i = 1, \ldots, n$ are absolutely continuous random variables with respect to the same measure, the probability that $ker(S) \ni \mathbf{a}$ is equal to zero.

The maximization lemma of quadratic forms points out a strong relation between T^2 defined in (3.1) and the univariate $t_{\mathbf{a}}$. Indeed, one can show that

$$T^2 = \max_{\mathbf{a} \in Im(S) \setminus \{\mathbf{0}_p\}} t_{\mathbf{a}}^2. \tag{3.2}$$

This means that making multivariate inference using T^2 at a certain confidence (significance) level is formally the same as making simultaneous univariate inference along any direction belonging to the "variability space explored by the data" (i.e. any direction $\mathbf{a} \in Im(S) \setminus \{\mathbf{0}_p\}$), while controlling the overall joint confidence (significance) level, ignoring all orthogonal directions (i.e. any direction $\mathbf{a} \in ker(S) \setminus \{\mathbf{0}_p\}$).

Note that $\mathbf{R}^p = Im(S) \oplus ker(S)$ for $n \geq 2$ and $p \geq 1$, with $Im(S) = \mathbf{R}^p$ and $ker(S) = \{\mathbf{0}_p\}$ almost surely if and only if $n > p$. Thus, when $n > p$, T^2 can be more simply defined as $\max_{\mathbf{a} \in \mathbb{R}^p \setminus \{\mathbf{0}_p\}} t_{\mathbf{a}}^2$; this is actually the most common way through which T^2 is introduced in the classical framework $n > p$. In the general framework the latter definition does not hold since $t_{\mathbf{a}}^2$ is not uniformly bounded in $\mathbf{R}^p \setminus \{\mathbf{0}_p\}$ when $n \leq p$.

3.2. Invariance properties of the generalized Hotelling's T^2

The Generalized Hotelling's T^2 is invariant under similarity transformations of the components (affine transformation preserving angles), i.e. affine transformation $A \bullet + \mathbf{b}$ such that $A \in \mathbb{R}^{p \times p}$ with $A = aO$, where $a \in \mathbb{R}_0^+$ and O is an orthogonal matrix, and $\mathbf{b} \in \mathbb{R}^p$. Indeed, under these assumptions we have that:

$$\begin{aligned} & n((A\overline{\mathbf{X}} + \mathbf{b}) - (A\mu_p + \mathbf{b}))'(ASA')^+((A\overline{\mathbf{X}} + \mathbf{b}) - (A\mu_p + \mathbf{b})) = \\ = \quad & n(\overline{\mathbf{X}} - \mu_p)' S^+ (\overline{\mathbf{X}} - \mu_p). \end{aligned}$$

The previous result relies on the fact, proven in Appendix A, that $(ASA')^+ = (A')^{-1} S^+ A^{-1}$ only for $A = aO$, where $a \in \mathbb{R}_0^+$ and O is an orthogonal matrix.

Similarity transformations are also those transformations that do not affect assumption (iv) of Theorem 3 nor the value of the constant $\overline{\sigma}^2/\overline{\sigma^2}$. On the contrary, they might affect the sparsity of both the covariance matrix Σ_p and the mean vector μ_p. Thus, as pointed out by the referees, no extra standard assumption on sparsity is required to carry on inference based on the Generalized Hotelling'd T^2. Indeed for instance, given a covariance structure, it is always possible by means of suitable orthogonal transformations to obtain inferentially equivalent scenarios in which teh covariance matrix is very sparse (i.e., diagonal) or completely full. thus, sparsity of the covariance matrix is not an issue for inference based on the Generalized Hotelling'd T^2.

Finally, it is easy to show that for $n > p$, T^2 is invariant under the wider class of affine transformations of the components, i.e. transformations $A \bullet + \mathbf{b}$ with $A \in \mathbb{R}^{p\times p}$ invertible and $\mathbf{b} \in \mathbb{R}^p$. This is due to the fact that $(ASA')^{-1} = (A')^{-1}S^{-1}A^{-1}$ for any invertible A. Lehmann and Romano (2005) proved that invariance under generic affinity transformations cannot be achieved in the framework $p \geq n$.

3.3. On the p-asymptotic law of the generalized Hotelling's T^2

For Corollary 4 to have some impact for inferential purposes, the constant $\overline{\sigma}^2/\overline{\sigma^2}$ needs to be known or at least efficiently estimated. Two cases may occur in practical situations when Σ_p is not known:

(a) the constant $\overline{\sigma}^2/\overline{\sigma^2}$ is known even if Σ_p is not completely known. This case may occur when partial knowledge of Σ_p is available;
(b) the constant $\overline{\sigma}^2/\overline{\sigma^2}$ is not known and it thus needs to be estimated.

Case (a) covers a few practical situations. For instance, the constant $\overline{\sigma}^2/\overline{\sigma^2}$ is known if the unknown covariance matrix Σ_p is equal to $\Sigma_p = \widetilde{\Sigma}_p + \gamma V_p$, with $\widetilde{\Sigma}_p$ an unknown positive definite (or even semi-definite) matrix such that $\lim_{p\to\infty} tr(\widetilde{\Sigma}_p) < +\infty$, γ an unknown positive constant, and V_p a known positive definite matrix satisfying (iv'). Indeed, in this case it can be proven that $\overline{\sigma}^2/\overline{\sigma^2} = \overline{\sigma_v}^2/\overline{\sigma_v^2}$ where $\overline{\sigma_v} = \lim_{p\to\infty} \frac{tr(V_p)}{p}$ and $\overline{\sigma_v^2} = \lim_{p\to\infty} \frac{tr(V_p^2)}{p}$; the proof comes straightforward once it is noticed that, without loss of generality, V_p can be assumed to be diagonal. A covariance matrix of this form occurs for instance in all applications where the observed p-variate random vectors are assumed to be generated by the sum of two independent terms: a structural term whose variability is concentrated on a finite number of components (or even infinite but with finite total variance) and a zero-mean nuisance term (due to background noise or measurement errors) satisfying (iv) acting on all components. If the covariance matrix of the nuisance term is assumed to be proportional to the identity matrix (as it often happens), in this case we have $\overline{\sigma}^2/\overline{\sigma^2} = 1$. This assumption may hold for instance in genetics, where long arrays of genes are observed on a small number of patients, the variability of the array can indeed be assumed to be generated by two independent terms, an informative variability

concentrated on a reduced number of positively/negatively correlated genes and a nuisance homoscedastic error variability acting independently on each gene. Spectral data presents another situation where the latter assumption may hold; indeed, spectral data are characterized by the presence of nuisance background variability along the entire set of observed frequencies plus a series of independent sources of variability at some specific frequencies (bands) associated to the spectral firms of different molecules.

Case (b) is the case where the information about the covariance structure is sufficient to know that Σ_p satisfies (iv), but not sufficient to know the value of the constant $\overline{\sigma}^2/\overline{\sigma^2}$. For instance, referring to the previous examples, we might know that V_p has a block structure with blocks $\ell \times \ell$ all equal to an unknown positive definite matrix B. In this case we know that $\overline{\sigma}$ and $\overline{\sigma^2}$ are for sure positive and finite without knowing their actual values $\overline{\sigma} = tr(B)/\ell$ and $\overline{\sigma^2} = tr(B^2)/\ell$.

In this second case having a good estimate of the constant $\overline{\sigma}^2/\overline{\sigma^2}$ becomes of primary importance. First of all, one can rely on some natural bounds to the constant based on the ratio between the maximal and minimal eigenvalue of Σ_p. For instance, $1-(\lambda_{max}/\lambda_{min}-1)^2 \leq \overline{\sigma}^2/\overline{\sigma^2} \leq 1$ useful when all eigenvalues are known to be similar (i.e., Σ_p is well-conditioned) or $(\lambda_{min}/\lambda_{max})^2 \leq \overline{\sigma}^2/\overline{\sigma^2} \leq 1$ useful when some eigenvalues are known to be very different (i.e., Σ_p is ill-conditioned).

A better estimate of $\overline{\sigma}^2/\overline{\sigma^2}$ can be obtained by using some estimates for $\frac{tr(\Sigma_p)}{p}$ and $\frac{tr(\Sigma_p^2)}{p}$. Indeed for $p \to \infty$ these quantities converge by definition to $\overline{\sigma}$ and $\overline{\sigma^2}$, and thus any unbiased estimator for $\frac{tr(\Sigma_p)}{p}$ (or $\frac{tr(\Sigma_p^2)}{p}$) for a given p is also a p-asymptotic unbiased estimator for $\overline{\sigma}$ (or $\overline{\sigma^2}$). The following estimators (introduced in Srivastava (2005)), defined for all $n \geq 3$ and $p \geq 1$, can be proven to satisfy this property:

$$\begin{array}{rcl} \hat{\overline{\sigma}}_p & := & \frac{trS}{p}, \ and \\ \hat{\overline{\sigma^2}}_p & := & \frac{(n-1)^2}{(n-2)(n+1)}\left[\frac{trS^2}{p} - \frac{1}{n-1}\frac{(trS)^2}{p}\right]. \end{array} \tag{3.3}$$

3.4. p-asymptotic confidence region and hypothesis test for the mean of a normal population when $p \gg n$

Corollary 4 turns out to be a useful tool for the construction of confidence regions and hypothesis tests for the mean in all practical situations where the number p of random vector components is far larger than the number n of sample units or even virtually infinite (e.g., functional data) and data can be assumed normally distributed.

A p-asymptotic **Confidence Region** for the mean μ_p can be defined as follows:

$$CR_{1-\alpha}(\mu_p) := \left\{ \mathbf{m}_p \in \mathbf{R}^p : \frac{\overline{\sigma}^2}{\overline{\sigma^2}} \frac{np}{n-1} (\mathbf{m}_p - \overline{\mathbf{X}})' S^+ (\mathbf{m}_p - \overline{\mathbf{X}}) \leq \chi^2_\alpha(n-1) \right\}, \tag{3.4}$$

with $\chi^2_\alpha(n-1)$ being the upper α-quantile of a $\chi^2(n-1)$ random variable and $1-\alpha$ being the p-asymptotic confidence level.

Equivalently, a p-asymptotic **Hypothesis Test** for $H_0 : \mu_p = \mu_{0p}$ versus $H_1 : \mu_p \neq \mu_{0p}$ with p-asymptotic significance level α has the following rejection region:

$$\text{Reject } H_0 \text{ in favor of } H_1 \text{ if:}$$

$$\frac{\overline{\sigma}^2}{\overline{\overline{\sigma}^2}} \frac{np}{n-1} (\overline{\mathbf{X}} - \mu_{0p})' S^+ (\overline{\mathbf{X}} - \mu_{0p}) > \chi^2_\alpha(n-1). \tag{3.5}$$

The confidence region $CR_{1-\alpha}(\mu_p)$ is not of practical use for graphical purposes since a clear visual representation of it is not straightforward due to the large value of p. Similarly to the traditional multivariate framework, univariate projections of the confidence region along some directions (i.e. T^2-simultaneous confidence intervals) can give a rough idea about the location and shape of the confidence region, providing – in the case of rejection of H_0 – also some help in detecting the directions that have taken to the rejection of H_0. From Corollary 4 and characterization property 3.2 we have that for $p \to \infty$

$$\mathbb{P}\left(\mathbf{a}'\boldsymbol{\mu_p} \in \left[\mathbf{a}'\bar{\mathbf{X}} \pm \sqrt{\frac{\mathbf{a}'S\mathbf{a}}{n}} \sqrt{\frac{\overline{\sigma^2}}{\overline{\overline{\sigma}^2}} \frac{n-1}{p} \chi^2_\alpha(n-1)} \right], \forall \mathbf{a} \in Im(S) \right) \to 1-\alpha.$$

Thus, given a direction $\mathbf{a} \in Im(S) \setminus \{0\}$, the corresponding **$T^2$-simultaneous confidence interval** with p-asymptotic family-wise confidence $1-\alpha$ can be defined as follows:

$$\mathbf{a}'\boldsymbol{\mu_p} \in \left[\mathbf{a}'\bar{\mathbf{X}} \pm \sqrt{\frac{\mathbf{a}'S\mathbf{a}}{n}} \sqrt{\frac{\overline{\sigma^2}}{\overline{\overline{\sigma}^2}} \frac{n-1}{p} \chi^2_\alpha(n-1)} \right].$$

If $\mathbf{a} \notin ImS \setminus \{0\}$, then the corresponding T^2-simultaneous confidence interval is not bounded, i.e, equal to $[-\infty, +\infty]$.

Confidence region (3.4) and rejection region of test (3.5) present some peculiar features that are worth a little discussion.

Because S^+ is positive semi-definite, the confidence region $CR_{1-\alpha}(\mu_p)$ – which for $n > p$ is an ellipsoid subset of $\mathbb{R}^p$ – turns out to be a cylinder in $\mathbb{R}^p$ generated by the orthogonal extension in $ker(S)$ of an $n-1$-dimensional ellipsoid contained in $Im(S)$. As illustrative examples, three confidence regions for the mean vector when $p = 3$ and $n = 2, 3, 4$, respectively, are reported in Figure 1. In particular, as shown by the analytic expression of the generalized T^2-simultaneous confidence intervals, $CR_{1-\alpha}(\mu_p)$ is bounded in all directions belonging to the random space $Im(S)$. These directions are easily identifiable since the first $n-1$ sample principal components provide an orthonormal basis for $Im(S)$.

Due to the non-null dimension of the random space $ker(S)$ and to the orthogonality between $ker(S)$ and $Im(S)$, we have that the statistic $\frac{\overline{\sigma}^2}{\overline{\overline{\sigma}^2}} \frac{np}{n-1} (\overline{\mathbf{X}} -$

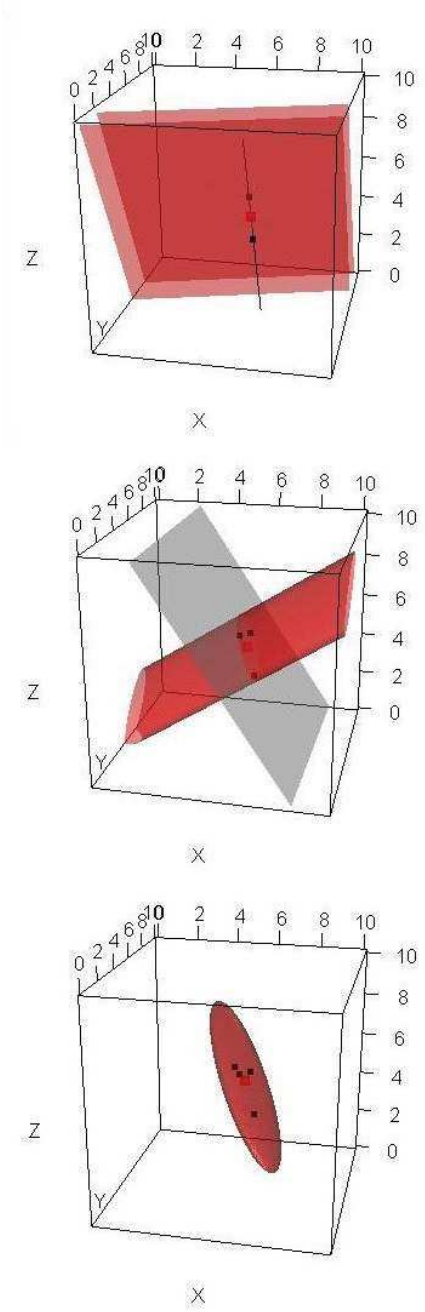


FIG 1. *Examples of confidence regions for the mean vector when* $p = 3$ *and* $n = 2, 3, 4$*, respectively. Data points in black, sample means in red, confidence regions in red,* $Im(S)$ *in gray.*

$\mu_{0p})' S^{+}(\overline{\mathbf{X}} - \mu_{0p})$ in the hypothesis test (3.5) does not change if μ_{0p} is replaced by $\mu_{0p} + \mathbf{m}_{ker(S)}$ with $\mathbf{m}_{ker(S)}$ being any vector belonging to $ker(S)$. This implies that H_0 might not be rejected even for values of the sample mean $\overline{\mathbf{X}}$ that are "really very far" from μ_{0p} in some direction within $ker(S)$. This is not surprising, because the use of S^{+} implies an exclusive focus on the space $Im(S)$ (the variability space explored by the data), neglecting all $p - n + 1$ directions associated to $ker(S)$ (the space orthogonal to the variability space explored by the data).

3.5. *p*-asymptotic pooled confidence region and hypothesis test for comparing the means of two normal populations when $p \gg n$

Theorem 3 can also be used to tackle the problem of comparing the means of two normal populations when the number p of components is larger than the number n of sample units. Indeed, under the same assumptions of the classical multivariate analysis of variance, we have that:

Corollary 5 (Generalized Pooled Hotelling's T^2_{pooled} p-asymptotic distribution law)**.** *For $n_a \geq 1$, $n_b \geq 1$, and $p \geq 1$, assume that:*

(i") $\{\mathbf{X}_{ai}\}_{i=1,\dots,n_a} \sim iid\ N_p(\mu_{pa}, \Sigma_p)$, $\{\mathbf{X}_{bi}\}_{i=1,\dots,n_b} \sim iid\ N_p(\mu_{pb}, \Sigma_p)$ *and the two finite sequences are independent;*

(iv) $0 < \overline{\sigma} = \lim_{p\to\infty} \frac{tr(\Sigma_p)}{p} < +\infty$, $0 < \overline{\sigma^2} = \lim_{p\to\infty} \frac{tr(\Sigma_p^2)}{p} < +\infty$, *and* $0 < \overline{\sigma^4} = \lim_{p\to\infty} \frac{tr(\Sigma_p^4)}{p} < +\infty$.

Then, for $n_a + n_b \geq 3$ and $p \to \infty$:

$$\frac{\overline{\sigma}^2}{\overline{\sigma^2}} \frac{p}{n_a + n_b - 2} \left(\frac{1}{n_a} + \frac{1}{n_b}\right)^{-1} \cdot$$
$$\cdot \left((\overline{\mathbf{X}}_\mathbf{a} - \overline{\mathbf{X}}_\mathbf{b}) - (\mu_{pa} - \mu_{pb})\right)' S^{+}_{pooled} \left((\overline{\mathbf{X}}_\mathbf{a} - \overline{\mathbf{X}}_\mathbf{b}) - (\mu_{pa} - \mu_{pb})\right) \xrightarrow{\mathcal{D}} \chi^2(n_a + n_b - 2),$$

where $\overline{\mathbf{X}}_\mathbf{a}$ and $\overline{\mathbf{X}}_\mathbf{b}$ are the two sample means, and S_{pooled} is the pooled sample covariance matrix.

Proof. It is another direct application of Theorem 3, since $(\frac{1}{n_a} + \frac{1}{n_b})^{-1/2}((\overline{\mathbf{X}}_\mathbf{a} - \overline{\mathbf{X}}_\mathbf{b}) - (\mu_{pa} - \mu_{pb})) \sim N_p(\mathbf{0}_p, \Sigma_p)$, $(n_a + n_b - 2)S_{pooled} = (n_a - 1)S_a + (n_b - 1)S_b \sim Wishart_p(\Sigma_p, n_a + n_b - 2)$, and they are independent. □

It is natural to denote the following quantity as Generalized Pooled Hotelling's T^2_{pooled}:

$$T^2_{pooled} = \left(\tfrac{1}{n_a} + \tfrac{1}{n_b}\right)^{-1} \left((\overline{\mathbf{X}}_\mathbf{a} - \overline{\mathbf{X}}_\mathbf{b}) - (\mu_{pa} - \mu_{pb})\right)' S^{+}_{pooled} \left((\overline{\mathbf{X}}_\mathbf{a} - \overline{\mathbf{X}}_\mathbf{b}) - (\mu_{pa} - \mu_{pb})\right). \tag{3.6}$$

Indeed, it is defined for any n_a, n_b, and p such that $n_a + n_b \geq 3$, $n_a \geq 1$, $n_b \geq 1$ and $p \geq 1$ and coincides with the classical definition of Pooled Hotelling's T^2_{pooled} when $p \leq n_a + n_b - 2$. The similarities and the differences between the framework $p > n_a + n_b - 2$ and the classical framework $p \leq n_a + n_b - 2$ are analogous to the ones presented in Section 3 for the Generalized Hotelling's T^2.

In particular, also in the two-population framework we obtain a confidence region to estimate the difference of the two means and a rejection region to test the difference of the two means.

A p-asymptotic **Confidence Region** for difference of the means $\mu_{pa} - \mu_{pb}$ can be defined as follows:

$$CR_{1-\alpha}(\mu_{pa} - \mu_{pb}) = \left\{ \Delta\mathbf{m_p} \in \mathbf{R}^p : \frac{\overline{\sigma}^2}{\overline{\sigma^2}} \frac{p}{n_a + n_b - 2} \left(\frac{1}{n_a} + \frac{1}{n_b} \right)^{-1} \cdot \right. \tag{3.7}$$
$$\left. \cdot \left(\overline{\mathbf{X}}_\mathbf{a} - \overline{\mathbf{X}}_\mathbf{b} - \Delta\mathbf{m_p} \right)' S^+_{pooled} \left(\overline{\mathbf{X}}_\mathbf{a} - \overline{\mathbf{X}}_\mathbf{b} - \Delta\mathbf{m_p} \right) \leq \chi^2_\alpha(n_a + n_b - 2) \right\},$$

with $1 - \alpha$ being the p-asymptotic confidence level.

Equivalently, a p-asymptotic **Hypothesis Test** for $H_0 : \mu_{pa} - \mu_{pb} = \Delta\mu_{0p}$ versus $H_1 : \mu_{pa} - \mu_{pb} \neq \Delta\mu_{0p}$ with p-asymptotic significance level α has the following rejection region:

Reject H_0 in favor of H_1 if:

$$\frac{\overline{\sigma}^2}{\overline{\sigma^2}} \frac{p}{n_a + n_b - 2} \left(\frac{1}{n_a} + \frac{1}{n_b} \right)^{-1} \cdot \tag{3.8}$$
$$\cdot \left(\overline{\mathbf{X}}_\mathbf{a} - \overline{\mathbf{X}}_\mathbf{b} - \Delta\mu_{0p} \right)' S^+_{pooled} \left(\overline{\mathbf{X}}_\mathbf{a} - \overline{\mathbf{X}}_\mathbf{b} - \Delta\mu_{0p} \right)) > \chi^2_\alpha(n_a + n_b - 2).$$

Also the analytical expression of the T^2_{pooled}-simultaneous confidence intervals for the difference of the means comes naturally.

Given a direction $\mathbf{a} \in Im(S_{pooled}) \setminus \{0\}$, the corresponding T^2_{pooled}**-simultaneous confidence interval** with p-asymptotic family-wise confidence $1 - \alpha$ can be defined as follows:

$$\mathbf{a}'(\boldsymbol{\mu_{pa}} - \boldsymbol{\mu_{pb}}) \in \left[\mathbf{a}'(\overline{\mathbf{X}}_\mathbf{a} - \overline{\mathbf{X}}_\mathbf{b}) \pm \sqrt{\left(\frac{1}{n_a} + \frac{1}{n_b} \right) \mathbf{a}' S \mathbf{a}} \sqrt{\frac{\overline{\sigma^2}}{\overline{\sigma}^2} \frac{n_a + n_b - 2}{p} \chi^2_\alpha(n_a + n_b - 2)} \right].$$

If $\mathbf{a} \notin Im(S_{pooled}) \setminus \{0\}$, then the corresponding T^2_{pooled}-simultaneous confidence interval is not bounded, i.e, equal to $[-\infty, +\infty]$.

Also in the two population framework, the unknown constant $\overline{\sigma}^2/\overline{\sigma^2}$ can be estimated by means of the following p-asymptotically unbiased and consistent estimators:

$$\begin{array}{rcl} \hat{\overline{\sigma}}_p & := & \frac{trS_{pooled}}{p}, \ and \\ \hat{\overline{\sigma^2}}_p & := & \frac{(n_a+n_b-2)^2}{(n_a+n_b-3)(n_a+n_b)} \left[\frac{trS^2_{pooled}}{p} - \frac{1}{n_a+n_b-2} \frac{(trS_{pooled})^2}{p} \right]. \end{array} \tag{3.9}$$

Recently, Srivastava and Yanagihara (2010) and Pigoli et al. (2012) proposed tests for the equality of two covariance matrices in the large p small n data framework, enabling to check for the homoscedasticity assumption which the previous results rely on.

4. Simulation study

In this section, we estimate, by means of MC simulations, the power and the actual level of significance of the new test, presented in (3.8); from now on we will refer to it as the p-asymptotic Generalized Hotelling's test, being based on the (finite-n) p-asymptotic distribution of the Generalized Hotelling's T^2. In particular, we estimate the probability of rejecting the null hypothesis $H_0 : \mu_a = \mu_b$ in favor of the alternative hypothesis $H_1 : \mu_a \neq \mu_b$ in twelve different cases and for increasing values of the number p of components ranging between 2^0 and 2^{10} (i.e., 1 and 1024) and $n_a = n_b = 10$:

Simulation Scenario	$\mu_a = \mathbf{0}$ $\mu_b = \mathbf{0}$	$\mu_a = \mathbf{0}$ $\mu_b = 0.4 \cdot \mathbf{1}$	$\mu_a = \mathbf{0}$ $\mu_b = 0.4 \cdot (\sqrt{p} \quad 0 \quad \cdots \quad 0)'$
$\Sigma = I$	I0	I1	I2
$\Sigma = D$	D0	D1	D2
$\Sigma = R$	R0	R1	R2
$\Sigma = S$	S0	S1	S2
$\Sigma = L$	L0	L1	L2

where I is the identity matrix; D is a diagonal matrix whose diagonal alternatively assumes the values 0.5 and 1.5; R is a block matrix whose blocks are equal to the matrix $\left(\begin{smallmatrix} 1 & 0.5 \\ 0.5 & 1 \end{smallmatrix}\right)$; S is a block matrix whose blocks are equal to the matrix $\left(\begin{smallmatrix} 1 & -0.5 \\ -0.5 & 1 \end{smallmatrix}\right)$. Covariance matrices R and S can be simply obtained from D by means of an orthogonal transformation: 45° anticlockwise and 45° clockwise pairwise rotations, respectively; L is a diagonal matrix whose diagonal alternatively assumes the values 0.001 and 1.999. The values for n_a and n_b are the same used in the simulation study presented in Pesarin and Salmaso (2010, 2009). The values for μ_a and μ_b used in cases I0, D0, R0, S0, and L0 are the ones under null hypothesis; the ones used in cases I1, D1, R1, S1, and L1 are the same used in the simulation study presented in Pesarin and Salmaso (2010, 2009); the ones used in cases I2, D2, R2, S2, and L2 investigate an alternative hypothesis scenario where (for the same value of the non-centrality parameter $||\mu_a - \mu_b||$) the mean difference between the two populations is concentrated on just one component (i.e., cases I2, D2, R2, S2, and L2) and not uniformly spread over all components (i.e., cases I1, D1, R1, S1, and L1). The value for Σ used in cases I0, I1, and I2 are once again the same used in Pesarin and Salmaso (2010, 2009), while the value used in the remaining cases are meant to provide less trivial situations where assumption (iv) still holds. Note that in cases I0, I1, and I2 the constant $\overline{\sigma}^2/\overline{\sigma^2} = 1$, in cases L0, L1, and L2 the constant $\overline{\sigma}^2/\overline{\sigma^2} \approx 1/2$, while in the remaining ones $\overline{\sigma}^2/\overline{\sigma^2} = 4/5$. All simulations have been performed twice: either assuming the constant $\overline{\sigma}^2/\overline{\sigma^2}$ known, to avoid confounding effects due to its estimation; and estimating the constant $\overline{\sigma}^2/\overline{\sigma^2}$ by means of (3.9), to evaluate the bias due to its estimation. Results of the latter group of simulations are briefly discussed at the end of subsection 4.2. The rest of this session refers on the former group of simulations.

In details, for each case and for each value of the number p of components, 1000 synthetic data sets have been randomly generated according to the corresponding model and, for each one of these, the p-asymptotic Generalized

Hotelling's test has been performed at a nominal level of significance $\alpha = 0.05$. The relative number of times the null hypothesis has been rejected provides the estimate of either the actual level of significance of the p-asymptotic Generalized Hotelling's test (cases I0, D0, R0, S0, and L0) or its power (all remaining cases). The same data sets have been also used to perform three other tests recently appeared in the literature: the one presented in Pesarin and Salmaso (2010, 2009), the one presented in Theorem 2.2 of Srivastava (2007), and the one presented in Theorem 2.3 of Srivastava (2007)). In this section we will refer to them as the Pesarin-Salmaso's test, the Identity-matrix Generalized Hotelling's test, and the n-p-asymptotic Generalized Hotelling's test, respectively.

Analogously to the p-asymptotic Generalized Hotelling's test, also the n-p-asymptotic and the Identity-matrix Generalized Hotelling's tests are based on the generalized T^2_{pooled}. The three tests differ for the distribution used to build the corresponding rejection region: the p-asymptotic Generalized Hotelling's test uses a rejection region built from its p-asymptotic distribution under the assumption (iv) (equation 3.8), the n-p-asymptotic Generalized Hotelling's test uses a rejection region built from its n-p-asymptotic distribution under the assumption (iv) (Theorem 2.3 of Srivastava (2007)), and the Identity-matrix Generalized Hotelling's test uses a rejection region built from its exact distribution under the assumption of independent and homoscedastic components (Theorem 2.2 of Srivastava (2007)).

The Pesarin-Salmaso's test is not a model-based test but a permutation test; the implementation used here is the same used in the simulation study presented in Pesarin and Salmaso (2010, 2009), i.e., the statistic used is actually a random weighted sum of the p univariate Student's t^2_{pooled} that can be written as $(\overline{\mathbf{X}}_\mathbf{a} - \overline{\mathbf{X}}_\mathbf{b} - \Delta\mu_0)' S^{-1}_{diag} (\overline{\mathbf{X}}_\mathbf{a} - \overline{\mathbf{X}}_\mathbf{b} - \Delta\mu_0)$, and its conditional distribution over the values observed within each data set is estimated by sampling 1000 random permutations of the $n_a + n_b = 20$ p-dimensional vectors making each data set. S_{diag} is the diagonal matrix whose diagonal elements are the p (non-pooled) sample variances of the p components.

The results of the simulation study are summarized in Figure 2. For completeness, in the cases in which $p \leq n_a + n_b - 2$, a classical Hotelling's test has been implemented.

In subsections 4.1, 4.2, and 4.3, the p-asymptotic Generalized Hotelling's test is compared with the Identity-matrix Generalized Hotelling's test, the n-p-asymptotic Generalized Hotelling's, and the Pesarin-Salmaso's test, respectively.

4.1. Comparison between the p-asymptotic generalized Hotelling's test and the identity-matrix generalized Hotelling's test

In case I0, where H_0 is true and the hypotheses supporting the Identity-matrix Generalized Hotelling's test hold, the observed rate of rejection of the Identity-matrix Generalized Hotelling's test clearly matches its nominal level of significance 5%; on the contrary, in cases D0, R0, S0, and L0, where H_0 is still true

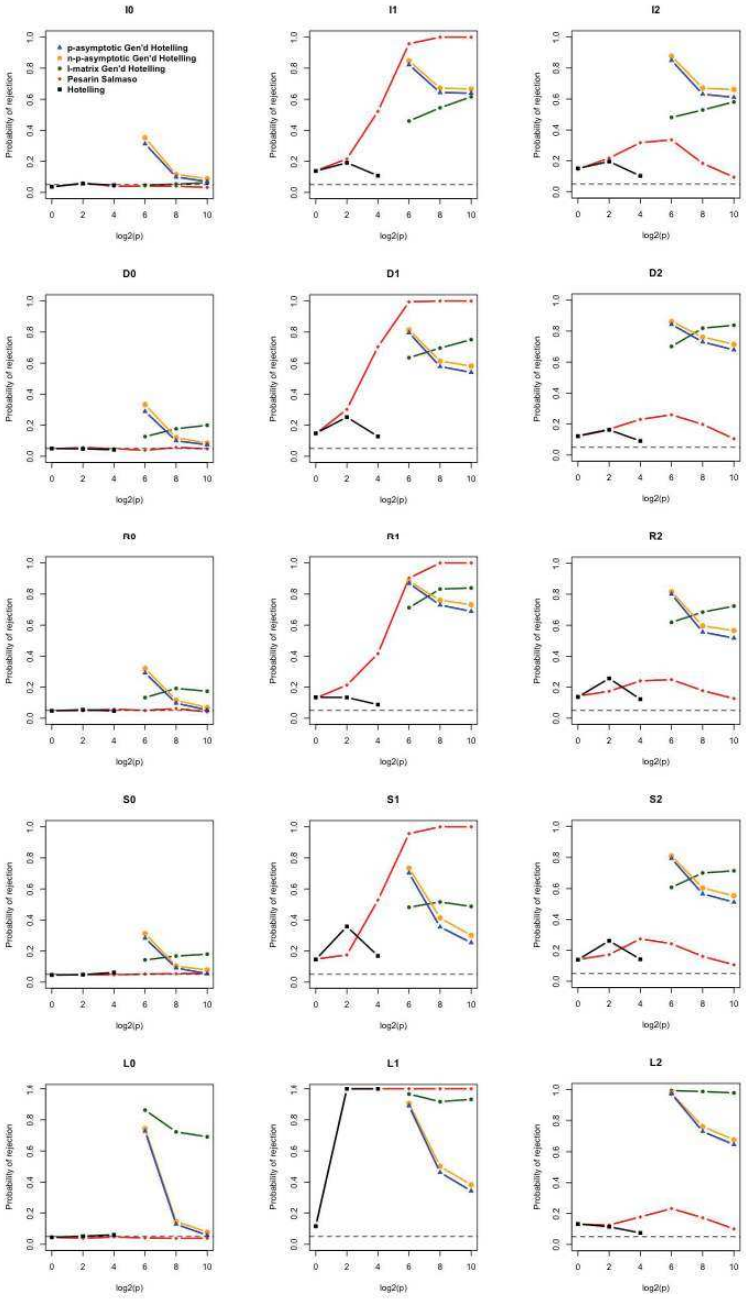


FIG 2. *MC-estimates of the probability of rejecting* $H_0 : \mu_a = \mu_b$ *for different values of the number* p *of components. Each plot is associated to a different generative model (title) and each line to a different test (legend).*

but the hypotheses supporting the Identity-matrix Generalized Hotelling's test do not hold, the observed rate of rejection of the Identity-matrix Generalized Hotelling's test significantly exceeds its nominal level of significance providing a strongly non-conservative test.

The assumptions which the p-asymptotic Generalized Hotelling's test is based on, hold instead for all cases, indeed for p "large enough" (in these cases 1024 seems to be a large enough value for p) the observed rate of rejection matches the nominal level of significance 5%. The almost identical patterns shown for cases D0, R0, and S0 confirm the invariance of the p-asymptotic Generalized Hotelling's test under orthogonal transformations of the components. Further simulations, not reported here, for different values of n_a and n_b show a quicker (slower) convergence to the nominal level of significance for smaller (larger) values of n_a and n_b. For instance, for $n_a = n_b = 2$ (i.e. the smallest sample size we tested), the nominal value is already reached for $p = 64$. Mind the fact that, though small sample sizes increase the reliability of the p-asymptotic Generalized Hotelling's test, they of course also reduce the power of the same test, as expected.

Fortunately, the same simulations also suggest that the convergence rate is independent from the value of the constant $\overline{\sigma}^2/\overline{\sigma^2}$. This fact enables an a-priori empirical measure, for a given sample size, of the minimal number p of random vector components (or given p, of the maximal sample size) that is necessary to make the p-asymptotic Generalized Hotelling's test reliable.

Talking about the power under the alternative hypothesis $\Delta\mu = 0.4 \cdot \mathbf{1}$, in case I1 the superiority of the p-asymptotic Generalized Hotelling's test is just apparent and due to the mismatch between its actual and its nominal level of significance for too small values of p. For large value of p (i.e. values for which the actual level of significance reaches its nominal value), the powers of the two tests appear almost identical confirming the p-asymptotical inferential equivalence of the two under the more stringent hypotheses of the Identity-matrix Generalized Hotelling's test.

In cases D1, R1, S1, and L1, the mismatch between the actual and the nominal level of significance completely affects the Identity-matrix Generalized Hotelling's test providing meaningless power curves for this test. The only values of interest in these plots are the estimated powers of the p-asymptotic Generalized Hotelling's test for $p = 1024$, that is the only case in which the nominal level of significance equals the actual one. Different values of that power are achieved in the four cases. In particular, a comparison of the p-asymptotic Generalized Hotelling's test and the classical Hotelling's test across these four cases points out an opposite behavior of the two: the power of the p-asymptotic Generalized Hotelling's test is higher when the power of the classical Hotelling's test is lower and viceversa. More in detail, the power of the p-asymptotic Generalized Hotelling's test is enhanced (reduced) and the power of the classical Hotelling's test reduced (enhanced) for alternative hypotheses providing a difference of the means with large (small) components in the directions of important (in terms of eigenvalues) principal components and small (large) components in the directions of minor (in terms of eigenvalues) principal components. Indeed, classical

Hotelling's test is based on the Mahalanobis distance (induced by the inverse of the sample covariance matrix) between the sample difference of the means and the H_0 difference of the means; thus, in the classical Hotelling's test, the effect of differences occurring in the direction of the minor sample principal components is enlarged with respect to similar differences occurring in the direction of the important sample principal components. On the contrary, in the framework of the p-asymptotic Generalized Hotelling's test the directions associated to the minor principal components are expected to be close to the directions detected by $ker(S)$ and thus any difference in these directions have a high probability to be annihilated by the Mahalanobis semi-distance induced by the generalized inverse of the sample covariance matrix used.

4.2. Comparison between the p-asymptotic generalized Hotelling's test and the n-p-asymptotic generalized Hotelling's test

The actual significance level and power functions of the n-p-asymptotic Generalized Hotelling's test recall the ones of the p-asymptotic Generalized Hotelling's test except for a permanent bias observed in all simulated cases due to the use of the n-asymptotic approximation in a finite-n scenario. Indeed, differently from the p-asymptotic Generalized Hotelling's test, the actual significance level of the n-p-asymptotic Generalized Hotelling's test does not seem to converge to the nominal significance level of 5% for p going to infinity, providing in the large p small n case non-conservative inference. Given the actual (finite-n) p-asymptotic distribution of the Generalized Hotelling's T^2 under the null hypothesis (equation 3.8) it is straightforward to compute the actual p-asymptotic significance level of the n-p-asymptotic Generalized Hotelling's test, which does not depend on the covariance matrix. For convenience of the reader, in the right panel of Figure 3, for the nominal significance level $\alpha = 1\%, 5\%$, and 10%, the p-asymptotic actual significance level of the n-p-asymptotic Generalized Hotelling's test is reported as the sample size $n_a + n_b$ grows from 3 to 1024. As one can see the rate of convergence to the nominal value is quite slow, providing non-conservative inference in most real-world scenarios. In particular, the highest bias is observed in correspondence of small values of the nominal significance level and small sample sizes strongly discouraging the use of the n-p-asymptotic approximation on the finite-n framework. In the left panel of Figure 3 the same plot is reported for the one population test.

Both the n-p-asymptotic and the p-asymptotic Generalized Hotelling's test require the knowledge of the constant $\overline{\sigma}^2/\overline{\sigma^2}$. If the constant is not known, one can perform both tests estimating the constant by means of estimators (3.9). In this case, all simulations, as expected, point out a positive bias for both tests (which indeed do not take into account the variability introduced by the estimators (3.9)). In details, for $p = 1024$, the actual level of significance of the p-asymptotic Generalized Hotelling's test is estimated to be close to 7% in all simulated cases. This bias is expected to decrease/increase as the sample sizes increase/decrease. Moreover, by comparing cases I0, D0, and L0, the bias does not seem to be affected by the condition number of Σ_p.

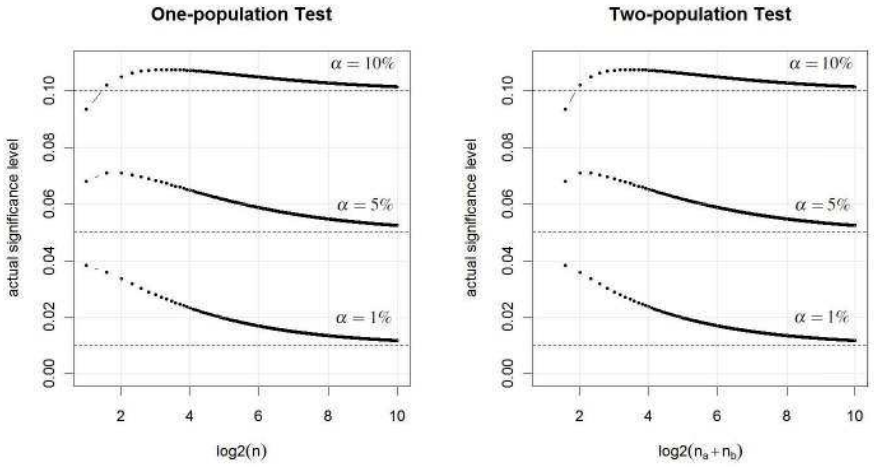


FIG 3. *p-asymptotic actual significance level of the n-p-asymptotic Generalized Hotelling's test as function of the sample size for nominal significance level* $\alpha = 1\%, 5\%,$ *and* 10%. *One-population test on the left and two-population test on the right.*

4.3. Comparison between the p-asymptotic generalized Hotelling's test and the Pesarin-Salmaso's test

The p-asymptotic Generalized Hotelling's test has also been compared with the Pesarin-Salmaso's test (Pesarin and Salmaso (2010, 2009)) by means of MC simulations. The aim of this comparison is to see to what extent the model-based approach, pioneered by Srivastava (2007) and further developed in this work, can compete with another promising and less traditional approach to the analysis of large p small n data: multivariate permutation test (Pesarin and Salmaso (2010, 2009)).

The Pesarin-Salmaso's test presents some very interesting features (proven in Pesarin and Salmaso (2010, 2009)): first of all it does not require the normality of data (test for multivariate normality is still an open problem); secondly, its actual level of significance resembles the nominal level in all simulated scenarios (cases I0, D0, R0, S0, and L0) and for any value of p (i.e., it is not p-asymptotic). The Pesarin-Salmaso's test also presents a few drawbacks due to the discrete nature of the permutational distribution, to the factorial growth of the number of permutation with respect to the sample size, and the non parametric nature of the permutational distribution. Indeed, given sample sizes n_a and n_b, we have in general $(n_a + n_b)!$ possible permutations of data associated – under the null hypothesis – to $\frac{(n_a+n_b)!}{n_a!\ n_b!}$ equiprobable different values of the test statistic.

The discrete nature of the permutational distribution is particulary evident for small sample sizes: in these cases, test randomization becomes mandatory in

order to maintain a certain level of significance α; for instance, for $n_a = n_b = 2$, only 24 permutations are possible, the support of the permutational distribution is reduced to just six values, and thus the only non-randomized bilateral tests that can be performed are the ones carried out with level of significance $\alpha = 1/3$ or $\alpha = 2/3$. If the sample size grows, this issue becomes less relevant from a practical point of view, but on the meantime the number of possible permutations quickly increases making mandatory the use of an approximated permutational distribution based on a randomly selected subset of permutations; for instance, for $n_a = n_b = 10$, the number of possible permutations already exceeds 10^{18} and the support of the permutational distribution is made of more than 10^5 values.

On the whole, the fact that (in most cases) the statistical conclusions might change across different runs of the same analysis and of the same data set makes the permutation-based analysis non replicable. Though the randomness induced by the approximating permutational distribution is just due to computational limits and it can be overcome by increasing the size of the random subset of permutations, on the contrary, the randomness induced by the discrete nature of permutational distribution is not due to computational limits but is intrinsic to this approach and thus non reducible.

Finally, due to the non parametric nature of the permutational distribution, the computation, in high dimension, of confidence regions and of T^2-simultaneous confidence intervals by means of the permutational approach is still an open issue.

A comparison of the estimated power functions of the two tests (cases I1, D1, R1, S1, L1, I2, D2, R2, S2, and L2) shows: a substantial equivalence of the statistical power of the Pesarin-Salmaso's test and of the classical Hotelling's test in the univariate case (i.e., $p = 1$); a predominance of one of the two depending on the number p of random vector components and on the scenario for the other values of p; this alternate predominance is even more pronounced for large values of p where in cases I1, D1, R1, S1, and L1 the Pesarin-Salmaso's test performs much better than the p-asymptotic Generalized Hotelling's test while in cases I2, D2, R2, S2, and L2 the Pesarin-Salmaso's test performs much worse than the p-asymptotic Generalized Hotelling's test.

To this purpose, it is interesting to compare the power functions of the scenarios I1 and I2. Indeed, from a geometric point of view the scenario I2 is obtained from scenario I1 by means of an orthogonal transformation (i.e., a rotation such that the vector $\mu_b - \mu_a$, lying along the bisector of the first 2^p-ant in case I1, lies along the first axis in case I2). On the one hand, one can appreciate the invariance in terms of power of all tests that are based on the Generalized Hotelling's T^2_{pooled} and on the other hand the dramatic drop of the power from nearly 100% to slightly more than 5% (i.e., the level of the test) of the Pesarin-Salmaso's test for large values of p. The unstable behavior of the Pesarin-Salmaso's test is probably due to the used test statistic (i.e., direct combination of univariate t-statistics) rather than on the permutational approach itself. Anyway among all test statistics suggested in Pesarin and Salmaso (2010, 2009) the "direct combination" one is probably the one closest to the Generalized Hotelling's T^2_{pooled} and

thus the one that should provide more similar inferential properties. Detecting, in the large p small n data framework, scenarios for the alternative hypothesis in which the supremacy of one of the two approaches can be taken for granted is still a matter of investigation.

5. Discussion

In this paper we dealt with the problem of making inference for the mean vector of a p-variate normal random vector when the sample size is too small to enable the use of Hotelling's Theorem. The problem of making inference for the difference of the mean vectors of two p-variate normal random vectors when the sample sizes are too small is discussed as well. In particular, we provided a generalization of the Hotelling's Theorem for p going to ∞ and sample size remaining finite, based on the notion of Moore-Penrose generalized inverse of the sample covariance matrix that holds under weak assumptions guaranteeing the existence and non degeneracy of the corresponding limit statistic and distribution. Together with the theorem, we provided also explicit formulas to perform hypothesis test and to build confidence regions and T^2-simultaneous confidence intervals.

We tested our results by means of MC simulations performed for different values of the means, of the covariance matrix, of the sample size, and of the number p of random vector components. Simulations confirm our theoretical results and moreover have been used to estimate the statistical power of the test and of the rate of convergence to the p-asymptotic framework. Some interesting cases are presented in the paper and critically discussed in comparison with three other approaches presented in the literature: the Identity-matrix Generalized Hotelling's test (Theorem 2.2, Srivastava (2007)), the n-p-asymptotic Generalized Hotelling's test (Theorem 2.3, Srivastava (2007)), and the Pesarin-Salmaso's permutation test (Pesarin and Salmaso (2010, 2009)).

Theoretical and simulation results suggest $\frac{(tr(\Sigma_p)/p)^2}{tr(\Sigma_p^2)/p}$ (i.e., the finite version of $\overline{\sigma}^2/\overline{\sigma^2}$) to be the right constant to correct the Identity-matrix Generalized Hotelling's test statistic to make it distributed as a $F(n_a+n_b-2, p-n_a-n_b+3)$ (i.e., the distribution that T^2 is known to follow when the covariance matrix is proportional to the identity matrix). Among these results we have: the p-asymptotic inferential equivalence of the Identity-matrix Generalized Hotelling's test and of the p-asymptotic Generalized Hotelling's test under the more stringent hypothesis of the former one; the n-asymptotic inferential equivalence of the n-p-asymptotic Generalized Hotelling's test and of the p-asymptotic Generalized Hotelling's test; and finally, the independence, from the ratio $\overline{\sigma}^2/\overline{\sigma^2}$, of the rate of convergence of the p-asymptotic Generalized Hotelling's test to its nominal level of significance observed in the simulations.

If this were true (further simulations not reported here seem to confirm it), we could describe within a unique framework the classical Hotelling's test, Identity-matrix Generalized Hotelling's test, n-p-asymptotic Generalized Hotelling's test, and the p-asymptotic Generalized Hotelling's test, as follows:

Conjecture 6 (Generalized Hotelling's T^2 distribution law)**.** *For $n \geq 2$ and $p \geq 1$, assume that:*

(i') $\{\mathbf{X}_i\}_{i=1,\ldots,n} \sim iid\ N_p(\mu_p, \Sigma_p)$.

Then:

$$\frac{tr(\mathbf{C}_p)^2}{tr(\mathbf{C}_p^2)} \frac{\nu_2}{(n-1)p\nu_1} T^2 \sim F(\nu_1, \nu_2),$$

with $\nu_1 = \min(n-1, p)$, $\nu_2 = |(n-1)-p|+1$, *and* $\mathbf{C}_p = \begin{cases} \mathbf{I}_p & for\ p \leq n-1 \\ \Sigma_p & for\ p > n-1 \end{cases}$.

Conjecture 7 (Generalized Pooled Hotelling's T^2_{pooled} distribution law)**.** *For $n_a \geq 1$, $n_b \geq 1$, $n_a + n_b \geq 3$, and $p \geq 1$, assume that:*

(i") $\{\mathbf{X}_{ai}\}_{i=1,\ldots,n_a} \sim iid\ N_p(\mu_{pa}, \Sigma_p)$, $\{\mathbf{X}_{bi}\}_{i=1,\ldots,n_b} \sim iid\ N_p(\mu_{pb}, \Sigma_p)$, *and the two finite sequences are independent.*

Then:

$$\frac{tr(\mathbf{C}_p)^2}{tr(\mathbf{C}_p^2)} \frac{\nu_2}{(n_a+n_b-2)p\nu_1} T^2_{pooled} \sim F(\nu_1, \nu_2),$$

with $\nu_1 = \min(n_a + n_b - 2, p)$, $\nu_2 = |(n_a + n_b - 2) - p| + 1$, *and* $\mathbf{C}_p = \begin{cases} \mathbf{I}_p & for\ p \leq n_a + n_b - 2 \\ \Sigma_p & for\ p > n_a + n_b - 2 \end{cases}$.

The previous conjectures are proven to be correct when $p \leq n-1$ and $p \leq n_a + n_b - 2$ respectively (Hotelling's Theorem), when $\Sigma_p = \mathbf{I}_p$ (Theorem 2.2, Srivastava (2007)) and when $p \to +\infty$ (Corollaries 4 and 5). The generic Σ_p finite p case is at the moment just supported by our simulations and by its consistence with the classical Hotelling's test, with the Identity-matrix Hotelling's test, and with the p and n-p-asymptotic Generalized Hotelling's tests when restricted conditions are posed. Its general proof is still under investigation.

Appendix A: Some useful properties of the Moore-Penrose generalized inverse

Many results presented in the paper rely on properties related to the Moore-Penrose inverse of positive semi-definite sample covariance matrices. In this appendix, all these properties are recalled.

Definition 1. Let A be an $q \times r$ matrix. The Moore-Penrose inverse of A, denoted by A^+, is the unique $r \times q$ matrix such that

1. $AA^+A = A$;
2. $A^+AA^+ = A^+$;
3. $(A^+A)^\star = A^+A$;
4. $(AA^+)^\star = AA^+$.

The first two properties let A^+ be a generalized inverse of A. The last two properties confer to A^+ its uniqueness. The symbol "$\star$" indicates the conjugate of a matrix. For our purposes, all matrices will have real entries and thus, it is equivalent to the symbol "$'$" indicating the transposed matrix.

The proof of the uniqueness of A^+ can be found, for instance, in Rao and Mitra (1971).

Moreover, it can be proven, by means of simple computations, that if A is a $p \times p$ symmetric matrix with real entries with rank $m \leq p$, then $A^+ = \sum_{i=1}^{m} \lambda_i^{-1} \mathbf{e}_i \mathbf{e}_i'$, with $\lambda_1, \ldots, \lambda_m$ being the m non-zero eigenvalues of A and $\mathbf{e}_1, \ldots, \mathbf{e}_m$ the corresponding eigenvectors. An immediate consequence of this result is that, if A is of full-rank, then $A^+ = A^{-1}$.

Hereby, we report some results necessary to the proof of Theorem 3.

Proposition 8. *Let A be a $\ell \times m$ matrix and B be an $m \times n$ matrix. If*

- *A has orthonormal columns, i.e, $A'A = I_m$; or,*
- *B has orthonormal rows, i.e, $BB' = I_m$; or,*
- *A is of full column rank m and B is of full row rank m,*

then, we have

$$(AB)^+ = B^+ A^+$$

The proof can be found in Rao and Mitra (1971).

Proposition 9. *Let A be a $\ell \times m$ matrix. Two particular cases are of interest:*

- *if A is of full column rank m, then $A'A$ is invertible and we get*

$$A^+ = (A'A)^{-1} A'$$

- *if A is of full row rank ℓ, then AA' is invertible and we get*

$$A^+ = A'(AA')^{-1}$$

The proof can be found in Rao and Mitra (1971).

Proposition 10. *With the same notations defined in the proof of Theorem 3, we have*

$$W^+ = H' L^{-1} H.$$

Proof. We first have

$$W^+ = (H'LH)^+ = [(H'L)\, H]^+$$

H has orthonormal rows since $HH' = I_m$. Thus, Proposition 8 holds and we have

$$W^+ = H^+ (H'L)^+$$

Now, we focus on the product $H'L$. H' has orthonormal columns since H has orthonormal rows. Therefore, once again, Proposition 8 holds and we obtain

$$W^+ = H^+ L^+ (H')^+$$

We now observe that

- L is invertible $\Rightarrow L^+ = L^{-1}$;
- H is of full row rank $\Rightarrow H^+ = H'(HH')^{-1}$ thanks to Proposition 9 (first part); then, since $HH' = I_m$, we obtain $H^+ = H'$;
- H' is of full column rank $\Rightarrow (H')^+ = (HH')^{-1}H$ thanks to Proposition 9 (second part); then, since $HH' = I_m$, we obtain $(H')^+ = H$.

and it ends the proof. □

Acknowledgements

This research is partly funded by the Large P Small n Project – Politecnico di Milano. All MC simulations and figures have been obtained by means of the statistical programming environment R (R Development Core Team, 2012). Finally, the authors want to thank Prof. Marco Fuhrman (Dept. of Mathematics "F. Brioschi" – Politecnico di Milano) for his help in detailing some technical aspects of the work.

References

Anderson, T. W. (2003). *An Introduction to Multivariate Statistical Analysis*, 3rd revised ed. *Wiley Series in Probability and Statistics*. John Wiley and Sons Inc. MR1990662

Benjamini, Y. and Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *Ann. Statist.* **29** 1165–1188. MR1869245 (2002i:62135)

Ferraty, F. and Vieu, P. (2006). *Nonparametric functional data analysis. Springer Series in Statistics*. Springer, New York. MR2229687 (2007b:62008)

Hall, P. and Keilegorn, I. V. (2007). Two-sample tests in functional data analysis starting from discrete data. *Statistica Sinica* **17** 1511–1531. MR2413533

Lehmann, E. L. and Romano, J. P. (2005). *Testing Statistical Hypotheses*, 3rd ed. *Springer Texts in Statistics*. Springer, New York. MR2135927 (2006m:62005)

Pesarin, F. and Salmaso, L. (2009). Advances in Permutation Testing Approach Working Paper No. 6, Department of Management and Engineering, University of Padua.

Pesarin, F. and Salmaso, L. (2010). *Permutation Tests for Complex Data: Theory, Applications and Software*. Wiley Series in Probability and Statistics, Chichester.

Pigoli, D., Aston, J. A. D., Dryden, I. L. and Secchi, P. (2012). Distances and Inference for Covariance Functions Technical Report No. 35/2012, MOX – Department of Mathematics, Politecnico di Milano.

Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis*, 2nd ed. Springer New York. MR2168993

Rao, C. R. and Mitra, S. K. (1971). *Generalized Inverse of Matrices and Its Applications. Wiley Series in Probability and Statistics*. John Wiley and Sons Inc. MR0338013

Serfling, R. J. (2002). *Approximation Theorems of Mathematical Statistics. Wiley series in probability and mathematical statistics* **413**. Wiley. MR0595165

Srivastava, M. S. (2005). Some tests concerning the covariance matrix in high dimensional data. *Journal of the Japan Statistical Society* **35** 251–272. MR2328427

Srivastava, M. S. (2007). Multivariate theory for analyzing high dimensional data. *Journal of Japan Statistical Society* **37** 53–86. MR2392485

Srivastava, M. S. and Yanagihara, H. (2010). Testing the equality of several covariance matrices with fewer observations than the dimension. *J. Multivariate Anal.* **101** 1319–1329. MR2609494

Storey, J. D. (2003). The positive false discovery rate: A Bayesian interpretation and the q-value. *Ann. Statist.* **31** 2013–2035. MR2036398 (2004k:62055)

R Development Core Team (2012). R: A Language and Environment for Statistical Computing R Foundation for Statistical Computing, Vienna, Austria.