Journal of Computational and Applied Mathematics 333 (2018) 51–64

Contents lists available at ScienceDirect

Journal of Computational and Applied Mathematics

journal homepage: www.elsevier.com/locate/cam

Frequentist–Bayesian Monte Carlo test for mean vectors in high dimension

Ivair R. Silva [a,*], Edgard M. Maboudou-Tchao [b], Weslei L. de Figueiredo [c]

[a] Department of Statistics, Federal University of Ouro Preto, MG, Brazil
[b] Department of Statistics, University of Central Florida, Orlando, FL, USA
[c] Department of Mathematics, Federal University of Minas Gerais, MG, Brazil

ARTICLE INFO

Article history:
Received 12 August 2016

MSC 2010:
62L05
62L15
65C05

Keywords:
Inference in high dimension
Monte Carlo testing
Hotelling's test

ABSTRACT

Conventional methods for testing the mean vector of a P-variate Gaussian distribution require a sample size N greater than or equal to P. But, in high dimensional situations, that is when N is smaller than P, special and new adjustments are needed. Although Bayesian-empirical methods are well-succeeded for testing in high dimension, their performances are strongly dependent on the actual unknown covariance matrix of the Gaussian random vector. In this paper, we introduce a hybrid frequentist–Bayesian Monte Carlo test and prove that: (i) under the null hypothesis, the performance of the proposed test is invariant with respect to the real unknown covariance matrix, and (ii) the decision rule is valid, which means that, in terms of expected loss, the performance of the proposed procedure can always be made as good as the exact Bayesian test and, in terms of type I error probability, the method is always of α level for arbitrary $\alpha \in (0, 1)$.

1. Introduction

Hypothesis testing for the mean, μ, of a P-variate normal vector, $\tilde{X} = (X_1, \ldots, X_P)$, with hypotheses of the form $H_0 : \mu = \mu_0$ against $H_1 : \mu \neq \mu_0$, μ_0 known, is one of the most classical problems of the statistical multivariate analysis field. The test criterion is usually constructed in terms of a 'measure of evidence' in favor of (or against) the null hypothesis. The measure of evidence is a real-valued function of the sample composed by N observations of a random vector $\tilde{X}$. A quite peculiar problem is the case where the covariance matrix of $\tilde{X}$, say Σ, is unknown and the sample size, N, is smaller than the number P of variables in $\tilde{X}$. In this case, classical solutions, like e.g. the Hotelling's T^2 test, are not applicable [1]. Situations involving $P > N$ are broadly referred in the literature through the term 'high dimensional problems'.

Although many questions remain unsolved in this field, the study of high dimensional problems dates back to decades, and many solutions are due to efforts to solve model selection problems [2]. Among the pioneers, the nonparametric empirical Bayes approach developed by Robbins [3–5] is notorious. The idea is to take H_j as true if the posterior probability of H_j turns out to be greater than 0.5, for $j = 0, 1$. But the method is impractical when the exact calculation of the posterior distribution is not computable, which is not rare in practice. Motivated by the problems involving microarray experiments, [6] and [7] developed the nonparametric empirical Bayes method. For this last, the performance is greatly dependent on restrictive conditions over the choice of the prior distribution and over the actual structure of the covariance matrix Σ. Furthermore, according to [8], page 271, the nonparametric empirical Bayes method does not perform well even for large sample sizes.

* Corresponding author.
E-mail address: irs@ufmg.br (I.R. Silva).

https://doi.org/10.1016/j.cam.2017.10.022
0377-0427/

More recently, inspired by problems involving goodness-of-fit testing and functional data analysis, [9] introduced a multiple testing procedure, sensitive to the data, that complies with the asymptotic-consistency principle of [10]. The results of [9] are possible by assuming that the covariance matrix is known, although such assumption is rarely true in practice. Furthermore, the asymptotic-consistency is obtained by using non-informative priors, which is an inconvenient limitation in cases where usage of informative priors is desired or necessary. In the same direction, [11] suggests to use the Bayes Factor as a measure of evidence to perform the test with hypotheses $H_0 : \mu = \mathbf{0}$ versus arbitrary H_1. Like previous approaches, the prior distribution in the method of [11] cannot be chosen according to the analyst's uncertainty, but, instead, the user has to elicit a prior shape that favors a computable posterior distribution. Some elicited solutions follow multivariate Cauchy [12], smooth Cauchy [11], and mixture of Zellner's g-priors [13].

As a rule, the performances of former Bayesian methods are strongly dependent upon the unknown nuisance matrix, Σ. Also, they share the limitation of supporting performance evaluations under asymptotic arguments. But, as well stated by [14]:

'*The justification of asymptotic consistency will not help the naive user who should be more concerned with selection bias and the instability of the procedures*'.

Obviously, importance of large numbers properties is undeniable for statistical methodologies in general. But, asymptotic efficiency is not so useful when N is known to be small, which is usually the case in high dimensional problems. To solve this problem, the present paper introduces a hybrid methodology, which is so because it embraces frequentist and Bayesian reasoning in order to construct a method that is simultaneously sensitive to prior uncertainty and to sample variation. The solution is valid for any sample size and arbitrary prior distributions. The proposed method is valid for composite hypotheses of the form $H_0 : \mu = \mu_0$ against $H_1 : \mu \neq \mu_0$, and also for comparing G populations, i.e., hypotheses of the form $H_0 : \mu_1 = \cdots = \mu_G$ against $H_1 : \mu_i \neq \mu_j$ for some $i \neq j$, where μ_i, $i = 1, \ldots, G$, is the mean vector of population i. The Bayes Factor is the measure of evidence used to construct the decision rule. Following [15], a one-dimensional sufficient statistic is used to construct the posterior distribution.

Unfortunately, just like former methods, the shape of the posterior distribution of our method remains unknown. To resolve this problem, the test criterion is formulated through a Monte Carlo approach. For this, we prove that: (i) the performance of the proposed method in terms of type I error probability is invariant with respect to Σ, (ii) the decision rule is valid in practice, i.e., one can always calculate the required number of Monte Carlo simulations (m) for a true control of the expected loss in comparison to the exact Bayesian test, (iii) the method does not require a prior information about the nuisance matrix Σ, (iv) all results are valid for finite and even small sample sizes (N), and (v) arbitrary prior distributions can be used, that is, does not matter if priors are elicited for optimality in any sense or if they represent the actual user's uncertainty about μ. All results were obtained under analytical arguments.

The content of this paper is organized in the following way: next section establishes the main definitions and notation for tests based on the Bayes Factor and on loss functions. Section 3 introduces matrix operators that shall be hereinafter necessary. Section 4 introduces our Bayesian test in high dimension along with the demonstration of its main properties. Section 5 contextualizes applications of the proposed method for: control charts with heterogeneous variability, dependent longitudinal data, and multivariate sequential analysis for post-market drug and vaccine safety surveillance. Section 6 presents the main conclusions.

2. Bayes factor and loss functions

Under the conventional statistical inference theory, statistical hypotheses are two statements about a population parameter $\theta \in \mathbb{R}^K$, $K \in \mathbb{N}$. An usual format of the hypotheses is $H_0 : \theta \in \Theta_0$ against $H_1 : \theta \in \Theta_1$, where Θ_0 and Θ_1 are subsets of the parameter space, Θ, such that $\Theta_0 \cap \Theta_1 = \emptyset$. H_0 and H_1 are called the null hypothesis and the alternative hypothesis, respectively. A statistical hypothesis test is a rule that establishes for which values of the sample the hypothesis H_0 is taken as true and for which values the hypothesis H_1 is taken as true. The decision of taking H_i as true can be based on a random sample, $\mathbf{X} = (\tilde{X}_1, \ldots, \tilde{X}_N)'$, from the population, where $\tilde{X}_j$ is a P-variate random variable for each $j = 1, \ldots, N$. In practice, an arbitrary real-valued function of the sample, namely 'measure of evidence for H_i', $i = 0, 1$, is used to draw a decision. Small observed values for the measure of evidence for H_1 suggests that H_0 is to be taken as true and vice-verse.

Before defining a measure of evidence, it is necessary to define the empirical information for θ. Let $T(\mathbf{X})$ denote a sufficient statistic for θ based on the $(N \times P)$-dimensional random sample $\mathbf{X}$. $T(\mathbf{X})$ is also suggestively referred as the 'empirical information' for θ. $T(\mathbf{X})$ can be any mapping of the sample (not involving θ) from $\mathbb{R}^N$ to $\mathbb{R}^M$, with $M \leq N$. For example, a special but quite diffused option is to set the whole sample to be the empirical information, i.e., $T(\mathbf{X}) := \mathbf{X}$, which is the more common setting in the Bayesian practice and that generates a particular posterior distribution involving the likelihood function with respect to the observed vector $\mathbf{X} = \mathbf{x}_0$. As discussed by [15], another example is the case where $T(\mathbf{X})$ is a real-valued function, like a one-dimensional sufficient test statistic, for example.

Let $W(T(\mathbf{X}))$, or simply $W(T)$, denote the 'measure of evidence' for H_i, which is a real-valued function of $T(\mathbf{X})$. The function $W(T)$ is obtained by combining the empirical information with the 'prior' information. Before having a realization $\mathbf{x}_0$ of $\mathbf{X}$, the analyst might have insights, apart the data, of what would be the most, such as the less, plausible values for θ. This analyst's uncertainty can be measured through a probability density, called 'prior distribution' and here denoted by $\pi_\theta(y)$, $y \in \mathbb{R}^K$. Using the observed empirical information, $T(\mathbf{x}_0)$, the analyst can update his uncertainty about θ by using the Bayes' rule.