Ann Inst Stat Math (2011) 63:1141–1163
DOI 10.1007/s10463-010-0270-0

An optimal approach for hypothesis testing in the presence of incomplete data

Albert Vexler · Sergey Tarima

Received: 9 July 2007 / Revised: 21 July 2009 / Published online: 10 February 2010

Abstract The adverse effect of small sample sizes, excessive nonresponse rate, and high dimensionality on likelihood ratio test statistic can be reduced by integrating with respect to a prior distribution. If information regarding the prior is too general (for example, only a parametric family can be specified), this distribution can be chosen from a principle of the most powerful testing. We propose the integrated most powerful test in the presence of missing data. This test can be used as a viable alternative to the maximum likelihood.

Keywords Parametric hypothesis testing · Most powerful test · Likelihood ratio · Missing data · Maximum likelihood

1 Introduction

There are several conditions when maximum likelihood ratio (MLR) tests provide unstable results. Among these possibilities we highlight only three: small sample size, high dimensionality, and the presence of missing data. Small sample size and high dimensionality are tied together in many genetic data (genom-wide data for few subjects), longitudinal data sets (several subjects are observed over time), and in a variety pilot medical projects. In some cases these two are aggravated by incomplete observations. Saint Loise data set published in Little and Rubin (2002) analyzed in this manuscript absorbed all these problems.

A. Vexler (✉)
Department of Biostatistics, The State University of New York at Buffalo, Buffalo, USA
e-mail: avexler@buffalo.edu

S. Tarima
Division of Biostatistics, Medical College of Wisconsin, Milwaukee, USA

Statistical analyses of such data are usually based on parametric models. Parametric assumptions allow to decrease dimensionality of a problem and make MLR to be a most popular technique for hypothesis testing.

Many researchers have already suggested different solutions for this or similar problems. Laud and Ibrahim (1995) considered Bayesian approaches for model selection. Sequential change-point detection (a special case of hypothesis testing) literature assigns a prior distribution on test parameters improving change-point detection algorithms (e.g., Pollak 1978). An interesting solution for high dimensional hypothesis testing was suggested by Goeman et al. (2006). Following empirical Bayesian model, test parameter is assumed to come from a parametric family and the initial testing problem is substituted by a dual hypothesis on the hyperparameters of this assumed family.

In this manuscript, we indicate a problem of MLR testing, suggest to use a prior distribution on test parameters, and derive an optimal test statistic in the presence of missing data. In some cases the proposed below test statistic is a special case of Bayes factor (e.g., Marden 2000). In general it mixes both the Bayes factor and ML method for generating a test statistic. If this prior is not known we suggest to choose this distribution from the principle of the most powerful testing.

The layout is as follows. Section 2 goes over maximum likelihood ratio testing and shows that the maximum likelihood ratio does not retain optimal properties except for the setting of the Neyman-Pearson lemma. Section 3 formally considers the use of the integrated likelihood ratio test statistic, several simple examples, and handling nuisance parameters. Section 4 suggests hypothesis testing procedure for ignorable missing data. MAR and MCAR simulations for a normal bivariate case are given in Sect. 5. An application is presented in Sect. 6 for illustrative purposes. A short summary in Section 7 concludes the paper. Mathematically intensive derivations for missing data are moved to Appendix.

2 Statement of problem

We focus on a classical hypothesis testing problem restricted to nonrandomized tests. Let $X_1, \ldots, X_n$ be independent and identically distributed random variables with a probability density function $f(x|\theta), \theta \in \Omega \subset \mathcal{R}, x \in \mathcal{X} \subset \mathcal{R}$. A simple null hypothesis $H : \theta = \theta_0$ is to be tested versus a composite alternative $K : \theta \in \Omega_K \equiv \Omega \setminus \theta_0$. Let P_{θ_0} denote the distribution of the sequence $X_1, \ldots, X_n$ under H and P_υ denote its distribution under K with $\theta = \upsilon$. Likewise, let E_{θ_0} and E_υ denote expectation under P_{θ_0} and P_υ, respectively.

As follows from the fundamental lemma of Neyman and Pearson, if $\Omega_K \equiv \theta_1$ (simple hypothesis) then the likelihood ratio statistics

$$T_n^{NP} = \prod_{i=1}^{n} \frac{f(X_i|\theta_1)}{f(X_i|\theta_0)} \tag{1}$$

is the most powerful unbiased test, i.e. for a pre-specified significance level α there exists a threshold value k such that $P_{\theta_0}(T_n^{NP} > k) = \alpha$ and the power $P_{\theta_1}(T_n^{NP} > k)$

is the highest among all other tests, where $P_\theta(A) = \int_A \prod_{i=1}^n f(x_i|\theta)\mathrm{d}x_i,\ A \subset \mathcal{X}^n$. In our case Ω_K is not a single value set.

Lehmann (1986) considered hypothesis testing problem from the prospective of optimal decision making. The optimal decision assures the smallest long term loss, in other words, the smallest risk. A risk associated with a decision $\delta = I(T_n > k)$, $I(\cdot)$ is an indicator function, based on a test statistic T_n is defined as

$$bP_{\theta_0}(T_n > k) + \int_{\Omega_K} a(\theta)P_\theta(T_n < k)\mathrm{d}\Theta(\theta), \tag{2}$$

where $a(\theta)$ represents loss associated with accepting H when $\theta \in \Omega_K$, b - loss from rejecting H when $\theta = \theta_0$, Θ is a probability distribution from $\mathcal{P}$, a set of all probability measures on a σ-algebra of all measurable subsets of Ω. It is often assumed that $a(\theta)$ is independent of θ and α is fixed, then the risk minimization problem is simplified to minimizing

$$\int_{\Omega_K} P_\theta(T_n < k)\mathrm{d}\Theta(\theta). \tag{3}$$

If Θ is specified, we have an unambiguous criterium for finding the optimal test statistic, denote T_n^0.

Maximum likelihood ratio test statistic. In many practical cases the importance of imposing Θ is disregarded and $\Omega_K \equiv \theta_1$ is assumed, even when it is not true. If $\Omega_K \equiv \theta_1$, we are in the settings of the Neyman and Pearson lemma, $P_{\theta_1}(T_n < k)$ minimization corresponds to maximization of power $P_{\theta_1}(T_n > k)$ and $T_n^0 = T_n^{NP}$ is the optimal solution. If we assume that $\Omega_K \equiv \theta_1$ we are not necessarily aware of the actual value of θ_1, then the *maximum likelihood estimator* of T_n^{NP}

$$\hat{T}_n^{ML} = \sup_\theta \prod_{i=1}^n \frac{f(X_i|\theta)}{f(X_i|\theta_0)} \tag{4}$$

is usually used. The estimator $\hat{T}_n^{ML}$ yields an asymptotically uniformly or locally most powerful test (e.g., Choi et al. 1996). To formalize certain conclusions in this article, hereafter (sub)martingales under the hypothesis $H(K)$ will be denoted as $H(K)$-(sub)martingales. For a finite sample size, $\hat{T}_n^{ML} \geq T_n^{NP}$ (i.e. the estimator (4) of T_n^{NP} is biased), and moreover

Proposition 1 *The likelihood ratio T_n^{NP} is an H-martingale, whereas its estimator $\hat{T}_n^{ML}$ is an H-submartingale with respect to $\{X_1, \ldots, X_n\}$. Under K, T_n^{NP} is a submartingale.*

Proof It is clear that, by definition (1)

$$E\left(T_n^{NP}\middle|X_1, \ldots, X_{n-1}\right) = T_{n-1}^{NP} E\frac{f(X_n|\theta_1)}{f(X_n|\theta_0)}, \tag{5}$$

where

$$E_{\theta_0}\frac{f(X_n|\theta_1)}{f(X_n|\theta_0)} = \int \frac{f(x_n|\theta_1)}{f(x_n|\theta_0)} f(x_n|\theta_0)\mathrm{d}x_n = 1 \quad \text{and}$$

$$E_{\theta_1}\frac{f(X_n|\theta_1)}{f(X_n|\theta_0)} = E_{\theta_1}\left[\frac{f(X_n|\theta_0)}{f(X_n|\theta_1)}\right]^{-1} \geq \left[E_{\theta_1}\frac{f(X_n|\theta_0)}{f(X_n|\theta_1)}\right]^{-1}$$
$$\geq \left[\int \frac{f(x_n|\theta_0)}{f(x_n|\theta_1)} f(x_n|\theta_1)\mathrm{d}x_n\right]^{-1} = 1.$$

That is, T_n^{NP} is either an H-martingale or K-submartingale with respect to $\{X_1, \ldots, X_n\}$.

Define the maximum likelihood estimator

$$\hat{\theta}_n = \arg\max_{\theta} \prod_{i=1}^{n} f(X_i|\theta), \quad \hat{\theta}_0 = \theta_0.$$

Equation (4) yields

$$\hat{T}_n^{ML} = \prod_{i=1}^{n} \frac{f(X_i|\hat{\theta}_n)}{f(X_i|\theta_0)} \geq \prod_{i=1}^{n} \frac{f(X_i|\hat{\theta}_{n-1})}{f(X_i|\theta_0)} = \hat{T}_{n-1}^{ML} \frac{f(X_n|\hat{\theta}_{n-1})}{f(X_n|\theta_0)},$$

where $\hat{\theta}_{n-1}$ is independent of X_n. Then,

$$E_{\theta_0}\left(\frac{f(X_n|\hat{\theta}_{n-1})}{f(X_n|\theta_0)}\Big| X_1, \ldots, X_{n-1}\right) = \int_{\mathcal{R}} \frac{f(x_n|\hat{\theta}_{n-1})}{f(x_n|\theta_0)} f(x_n|\theta_0)\mathrm{d}x_n = 1.$$

In a manner similar to (5), we conclude that $\hat{T}_n^{ML}$ is an H-submartingale. □

Following Proposition 1, the MLR test does not preserve the H-martingale property and $P_{\theta_0}\{\hat{T}_n^{ML} > k\} \geq P_{\theta_0}\{T_n^{NP} > k\}$, where the test based on T_n^{NP} is most powerful. Thus, the test based on $\hat{T}_n^{ML}$ generally is not optimal.

3 Integrated likelihood ratio test statistic

In the literature on sequential change point detection, for preserving a martingale property of detection policy, the use of another test statistic has been suggested, e.g. Krieger et al. (2003). The well known Robert–Shiryaev test statistic was integrated with respect to Θ. We adopted and applied this approach to the likelihood ratio test statistic, then

$$T_n(\Theta) = \int_{\Omega_K} \prod_{i=1}^{n} \frac{f(X_i|\theta)}{f(X_i|\theta_0)} \mathrm{d}\Theta(\theta) \tag{6}$$

defines the *integrated likelihood ratio* (ILR) test statistic. We emphasis that hereafter the dependence on Θ (e.g. $T_n(\Theta)$) defines dependence on the functional form of Θ and not on its argument θ. The problem of an unknown θ_1 is substituted by a problem of choosing a reasonable Θ. We can pretend that $\theta_1 \sim \Theta$, however this assumption is unnecessary for applying the test statistic (6).

For example, following Krieger et al. (2003), we can consider two forms of Θ. If there exists a suspicion that main mass of the alternative choices of θ is concentrated at one side from θ_0, then $\Theta(\theta)$ can be chosen as

$$\left(\Phi\left(\frac{\theta-\mu}{\sigma}\right) - \Phi\left(\frac{-\mu}{\sigma}\right)\right)^{+} \Big/ \Phi\left(\frac{\mu}{\sigma}\right), \tag{7}$$

if there is no such suspicion we can use

$$\frac{1}{2}\Phi\left(\frac{\theta-\mu}{\sigma}\right) + \frac{1}{2}\Phi\left(\frac{\theta+\mu}{\sigma}\right). \tag{8}$$

In Eqs. (7) and (8), Φ denotes the standard normal distribution function, μ and σ have to be defined by a researcher, for example, $\mu = 0$ and $\sigma = 1$ as suggested in Krieger et al. (2003). Figures 1 and 2 plot functions (7) and (8) for different choices of their arguments.

Remark The considered approach does not involve estimation of an unknown parameter, as it happened with $\hat{T}_n^{ML}$, and can be applied when estimation of an unknown

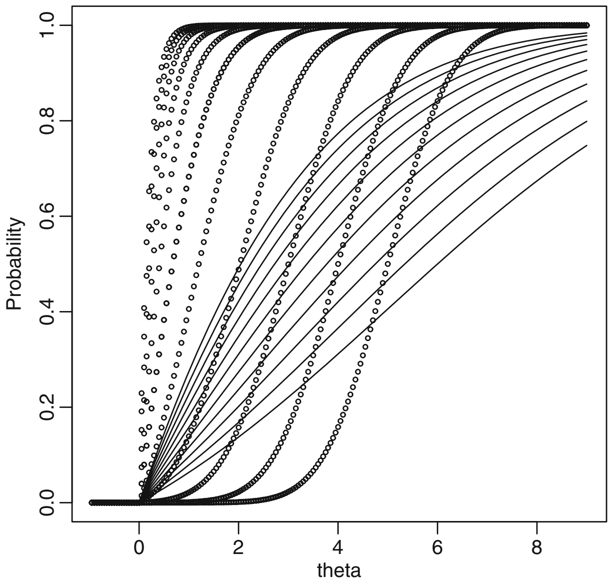

Fig. 1 Cumulative distribution functions defined by (7) for μ changing from -5 to 5. *Points* correspond to $\sigma = 1$ and *lines* to $\sigma = 5$

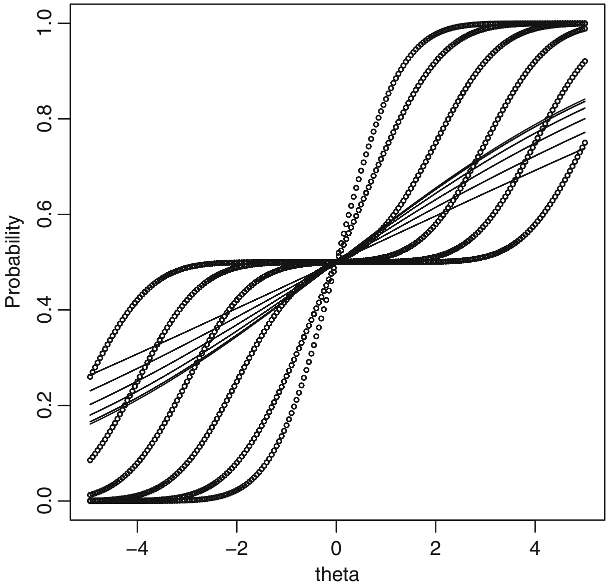

Fig. 2 Cumulative distribution functions defined by (8) for μ changing from -5 to 5. *Points* correspond to $\sigma = 1$ and *lines* to $\sigma = 5$

parameter is expected to be difficult, e.g. the sample size n is relatively small or the dataset suffers from an excessive nonresponse rate.

We propose to chose Θ from the principle of the most powerful testing. Note, this choice does not depend on $X_1, \ldots, X_n$. To this end, we present

Proposition 2 *The definition* (6) *preserves the H-martingale property of the test statistic $T_n(\Theta)$. In addition, for any dichotomous decision rule δ on a given sample ($\delta = 1$ rejects H) (6) is the integrated most powerful test statistic with respect to Θ, i.e.*

$$\int_{\Omega_K} \left(P_\theta \{T_n(\Theta) > k(\Theta)\} - k(\Theta) P_{\theta_0} \{T_n(\Theta) > k(\Theta)\}\right) d\Theta(\theta)$$
$$\geq \int_{\Omega_K} \left(P_\theta \{\delta \text{ rejects } H\} - k(\Theta) P_{\theta_0} \{\delta \text{ rejects } H\}\right) d\Theta(\theta). \quad (9)$$

Proof Since by (6)

$$E_{\theta_0}(T_n(\Theta)|X_1, \ldots, X_{n-1}) = \int_{\Omega_K} \prod_{i=1}^{n-1} \frac{f(X_i|\theta)}{f(X_i|\theta_0)} E_{\theta_0} \frac{f(X_n|\theta)}{f(X_n|\theta_0)} d\Theta(\theta) = T_{n-1}(\Theta),$$

the test statistic (6) is an H-martingale with respect to $\{X_1, \ldots, X_n\}$.

For any $a, b \in \mathcal{R}$ an elementary inequality

$$(a - b)(I(a > b) - \delta) \geq 0 \quad (10)$$

holds when $\delta \in [0, 1]$. With $a = T_n(\Theta)$, $b = k$, and an arbitrary dichotomous decision $\delta = \delta(X_1, \ldots, X_n)$, the inequality (10) can be written as

$$\left(\int_{\Omega_K} \prod_{i=1}^{n} \frac{f(X_i|\theta)}{f(X_i|\theta_0)} \mathrm{d}\Theta(\theta) - k\right) I\left(T_n(\Theta) > k\right) \geq \left(\int_{\Omega_K} \prod_{i=1}^{n} \frac{f(X_i|\theta)}{f(X_i|\theta_0)} \mathrm{d}\Theta(\theta) - k\right) \delta. \tag{11}$$

Since

$$E_{\theta_0} \prod_{i=1}^{n} \frac{f(X_i|\theta)}{f(X_i|\theta_0)} \delta = \int_{\mathcal{X}^n} \prod_{i=1}^{n} \frac{f(x_i|\theta)}{f(x_i|\theta_0)} \delta(x_1, \ldots, x_n) \prod_{i=1}^{n} f(x_i|\theta_0) \prod_{i=1}^{n} \mathrm{d}x_i = E_\theta \delta, \tag{12}$$

taking E_{θ_0} on each side of (11) completes the proof. □

Thus, for a given Θ, the test based on (6) is the integrated most powerful. This is clearly seen if we define $k(\Theta)$ from $P_{\theta_0}\{T_n(\Theta) > k(\Theta)\} = P_{\theta_0}\{\delta \text{ rejects } H\} = \alpha$ for all δ, then, the inequality (9) states that the integrated power

$$\int_{\Omega_K} P_\theta \left\{T_n(\Theta) > k(\Theta)\right\} \mathrm{d}\Theta(\theta)$$

of the test $\{T(\Theta) > k(\Theta)\}$ is the highest among all size α tests. An integrated risk, similar to (6), with respect to a known prior of an alternative parameter has been proposed by Lorden (1967) for sequential testing procedures. Pollak (1978) has shown an optimality of a sequential procedure based on (6) type statistics.

If there are no specific recommendations on Θ, the Θ with the highest integrated power, as follows from Proposition 2, is

$$\Theta^0 = \arg\max_{\Theta \in \mathcal{P}} \int_{\Omega_K} \left(P_\theta \left\{T_n(\Theta) > k(\Theta)\right\} - k(\Theta) P_{\theta_0} \left\{T_n(\Theta) > k(\Theta)\right\}\right) \mathrm{d}\Theta(\theta). \tag{13}$$

The solution Θ^0 is not necessarily unique in $\mathcal{P}$, we denote the set of these solutions by $\mathcal{P}^0 \subset \mathcal{P}$. The search of $\Theta^0 \in \mathcal{P}^0$ is not a trivial task and it seems reasonable to confine possible choices to a class of, for example, parametric distributions, denote $\mathcal{Q} \subset \mathcal{P}$. If $\mathcal{P}^0 \cap \mathcal{Q}$ is empty, we will not be able to find the Θ^0 defined by (13), however, the larger $\mathcal{Q}$ the more likely $\Theta^0 \in \mathcal{Q}$ or the closer (in terms of the highest integrated power) the chosen distribution will be to the class $\mathcal{P}^0$.

Note, if in (2) $a(\theta)$ is constant for all $\theta \in \Omega_K$ and T_n is defined by (6) then Θ^0 brings minimum to (2).

Requiring $P_{\theta_0}\{T_n(\Theta) > k(\Theta)\} = P_{\theta_0}\{\delta \text{ rejects } H\} = \alpha$ we restrict the choices of Θ to those which satisfy the unbiasedness condition. Then, the inequality (9) becomes

$$\int_{\Omega_K} P_\theta \{T_n(\Theta) > k(\Theta)\} \, \mathrm{d}\Theta(\theta) \geq \int_{\Omega_K} P_\theta \{\delta \text{ rejects } H\} \, \mathrm{d}\Theta(\theta) \tag{14}$$

and the optimal Θ is written as

$$\Theta^0 = \arg\max_{\Theta \in \mathcal{P}} \int_{\Omega_K} P_\theta \{T_n(\Theta) > k(\Theta)\} \, \mathrm{d}\Theta(\theta). \tag{15}$$

Interesting to see that the distribution (15) brings a minimum to (3) as well.

In this manuscript we examine practical aspects of the test, where $\mathcal{P}$ is limited to parametrical families.

3.1 Examples

Example 1 Let $H : X_i \sim N(0, 1)$ be tested versus $K : X_i \sim N(\theta_1, 1)$, $\theta_1 \in (-\infty, 0) \cup (0, \infty)$. If $\Theta = N(\mu, \sigma^2)$ with known μ and σ,

$$\begin{aligned} T_n(\Theta) = T_n(\mu, \sigma) &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} \exp\left[-\sum_{i=1}^n \frac{(X_i - \theta)^2}{2} + \sum_{i=1}^n \frac{X_i^2}{2}\right] \\ &\quad \times \exp\left[-\frac{(\theta - \mu)^2}{2\sigma^2}\right] \mathrm{d}\theta \\ &= \frac{1}{\sqrt{\sigma^2 n + 1}} \exp\left[\frac{\sigma^2(\mu\sigma^{-2} + \sum_{i=1}^n X_i)^2}{2(\sigma^2 n + 1)} - \frac{\mu^2}{2\sigma^2}\right]. \end{aligned} \tag{16}$$

Since a normally distributed $\sum_{i=1}^n X_i$ represents a sufficient statistic for $T_n(\mu, \sigma)$, the test is identical to one-sample z-test, which is solely based on $\sum_{i=1}^n X_i$ as well.

Example 2 Testing H versus K from Example 1 we assume that, Θ belongs to a symmetric mixture of two normal distributions, i.e.

$$\Theta \in \mathcal{Q} = \left\{0.5\, N(\mu, \sigma^2) + 0.5\, N(-\mu, \sigma^2) \middle| \mu \in \mathcal{R}, \sigma = \sigma_0 > 0\right\}.$$

Then, the family of the test statistics (6) defined on $\mathcal{Q}$ is

$$\begin{aligned} T_n(\Theta) = T_n(\mu, \sigma) &= \frac{0.5}{\sqrt{\sigma^2 n + 1}} \exp\left[\frac{\sigma^2(\mu/\sigma^2 + \sum_{i=1}^n X_i)^2}{2(\sigma^2 n + 1)} - \frac{\mu^2}{2\sigma^2}\right] \\ &\quad + \frac{0.5}{\sqrt{\sigma^2 n + 1}} \exp\left[\frac{\sigma^2(-\mu/\sigma^2 + \sum_{i=1}^n X_i)^2}{2(\sigma^2 n + 1)} - \frac{\mu^2}{2\sigma^2}\right]. \end{aligned}$$

In the class $\mathcal{Q}$, the integrated most powerful solution $\Theta^0 = N(\mu_0, \sigma^2)$ is defined by $\mu_0 = \arg\max_a U(a)$,where

$$U(a) = \frac{1}{2\sqrt{2\pi\sigma^2}} \int_{\mathcal{R}} \left[P\left\{ T_n(\theta,\sigma) > k \middle| \sum_{i=1}^n X_i \sim N(n\theta, n) \right\} \right. \\ \left. \times \left(\exp\left[-\frac{(\theta-a)^2}{2\sigma^2} \right] + \exp\left[-\frac{(\theta+a)^2}{2\sigma^2} \right] \right) \right] \mathrm{d}\theta \\ - kP\left\{ T_n(0,\sigma) > k \middle| \sum_{i=1}^n X_i \sim N(0,n) \right\}.$$

Example 3 If $\mathcal{Q} = \{Unif\ [0,b], b > 0\}$ then $\Theta^0 = Unif\ [0, b_0]$, where

$$b_0 = \arg\max_{b>0} \frac{1}{b} \int_0^b P_\theta \left\{ \frac{1}{b} \int_0^b \prod_{i=1}^n \frac{f(X_i|u)}{f(X_i|\theta)} \mathrm{d}u > k \right\} \\ - kP_{\theta_0} \left\{ \frac{1}{b} \int_0^b \prod_{i=1}^n \frac{f(X_i|u)}{f(X_i|\theta_0)} \mathrm{d}u > k \right\} \mathrm{d}\theta.$$

3.2 Integrated likelihood ratio with nuisance parameters

Many practical applications of hypothesis testing retain nuisance parameters η in addition to the parameters of interest θ. Under the null $X_i \sim f(x|\theta_0, \eta)$, where η is unknown, $i = 1, \ldots, n$. If alternative is true, $X_i \sim f(x|\theta, \eta)$, where $\theta \in \Omega_K$ and η is unknown again. We need some assumptions about η. And we decided to consider two main categories of them: where η is not random (frequentist point of view) and where a prior distribution on η can be assumed.

1. Let η be defined in an space Φ such that (θ, η) is defined on $\Omega_K \times \Phi$. So, we keep the same η under both H and K. Then, we can use

$$T_n^1(\Theta|\hat{\eta}) = \int_{\Omega_K} \prod_{i=1}^n \frac{f(X_i|\theta, \hat{\eta})}{f(X_i|\theta_0, \hat{\eta})} \mathrm{d}\Theta(\theta), \tag{17}$$

   where $\hat{\eta}$ is an estimator of η.

In some cases, the $X_1, \ldots, X_n$ has an invariant structure, which means that the influence of η can be eliminated (e.g., Yakir 1998). The invariant structure in change point literature is closely related to the presence of an ancillary complement in a sufficient statistic. For example, if $X_1, \ldots, X_n \sim N(\theta, \eta)$ then the distribution of the ancillary statistic $S_n^{-1} \sum_{i=1}^n X_i$ does not depend on η, where S_n is an unbiased estimate of η. Then, the optimal test for this example is based on $S_n^{-1} \sum_{i=1}^n X_i$.

2. Let $\Theta(\theta, \eta)$ be a pre-assumed distribution of interest and $\Theta(\theta, \eta) = \Theta_1(\theta)\Theta_2(\eta)$. This assumption allows us to avoid undesirable associations between the η and θ, and keep the same distribution of η under both H and K.

   We redefine the estimator (6) by

$$T_n(\Theta|\hat{\eta}) = \int_{\Omega_K \times \Phi} \prod_{i=1}^{n} \frac{f(X_i|\theta, \eta)}{f(X_i|\theta_0, \hat{\eta})} \mathrm{d}\Theta(\theta, \eta), \tag{18}$$

   where $\hat{\eta}$ is an estimator of η.

There are many different choices for this estimator: the maximum likelihood, under H; the maximum likelihood, under K; minimum variance unbiased estimator, under H; minimum variance unbiased estimator, under K; pooled estimators, which take into account that η does not change from H to K; and many others. For example, in Sect. 6 we calculated an EM estimator (there are missing values) of variance covariance matrix under an assumption that it does not change from H to K. This EM estimator is an analogue of a pooled estimator in MANOVA and ANOVA models. We are NOT trying to analyze which estimator of η is the best one. However, Proposition 3 presents a theoretical result conditional on a chosen estimator of η.

Note, the estimator (18) can also be used if θ and η are not independent. The distribution Θ is a joint distribution of η and θ. Then, $\hat{\eta}$ is an estimator of η under H. However, the interpretation of results of this testing usually is not straightforward.

Obviously, at a proper convergence rate of $\hat{\eta}$ to η, the estimator $T_n(\Theta|\hat{\eta})$ converges to the optimal test $T_n(\Theta)$, as $n \to \infty$. Alternatively, we can formulate

Proposition 3 *Let $\hat{\eta}(X_1, \ldots, X_n)$ be an estimator of η and for any dichotomous $\delta = \delta(X_1, \ldots, X_n)$*

$$\begin{aligned} &\hat{P}\left\{\delta \text{ rejects } H \middle| X_i \sim f(x|\theta_0, \hat{\eta}), i = 1, \ldots, n\right\} \\ &:= \int \prod_{i=1}^{n} f\left(x_i|\theta_0, \hat{\eta}(x_1, \ldots, x_n)\right) I\left(\delta(x_1, \ldots, x_n) \text{ rejects } H\right) \prod_{i=1}^{n} \mathrm{d}x_i = \alpha, \end{aligned} \tag{19}$$

then (18) is the integrated most powerful test statistic with respect to Θ.

Proof By virtue of the inequality (10)

$$\left(\int_{\Omega_K} \prod_{i=1}^{n} \frac{f(X_i|\theta, \eta)}{f(X_i|\theta_0, \hat{\eta})} \mathrm{d}\Theta(\theta, \eta) - k\right)\left(I\left(T_n(\Theta|\hat{\eta}) > k\right) - \delta\right) \geq 0,$$

or

$$\begin{aligned} &\left(\int_{\Omega_K} \prod_{i=1}^{n} \frac{f(X_i|\theta, \eta)}{f(X_i|\theta_0, \eta)} d\Theta(\theta, \eta) - k \prod_{i=1}^{n} \frac{f(X_i|\theta_0, \hat{\eta})}{f(X_i|\theta_0, \eta)}\right) \times \left(I\left(T_n(\Theta|\hat{\eta}) > k\right) - \delta\right) \\ &\geq 0. \end{aligned} \tag{20}$$

Applying $E_{\theta_0,\eta}$ to the elements of (20) we have

$$\begin{aligned}
&E_{\theta_0,\eta}\int_{\Omega_K}\prod_{i=1}^{n}\frac{f(X_i|\theta,\eta)}{f(X_i|\theta_0,\eta)}\delta(X_1,\ldots,X_n)\mathrm{d}\Theta(\theta,\eta)\\
&\quad=\int_{\mathcal{R}^n}\int_{\Omega_K}\prod_{i=1}^{n}\frac{f(x_i|\theta,\eta)}{f(x_i|\theta_0,\eta)}\delta(x_1,\ldots,x_n)\mathrm{d}\Theta(\theta,\eta)\prod_{i=1}^{n}f\,(x_i|\theta_0,\eta)\prod_{i=1}^{n}\mathrm{d}x_i\\
&\quad=\int_{\Omega_K}\int_{\mathcal{R}^n}\prod_{i=1}^{n}f(x_i|\theta,\eta)\delta(x_1,\ldots,x_n)\prod_{i=1}^{n}\mathrm{d}x_i\mathrm{d}\Theta(\theta,\eta)\\
&\quad=\int_{\Omega_K}\left[E_{\theta,\eta}\delta\right]\mathrm{d}\Theta(\theta,\eta)=E\left(E_{\theta,\eta}\delta|K\right)\\
&\quad=E\left(E_{\theta,\eta}\left(\delta\text{ rejects }H|X_i\sim f(x|\theta,\eta),i=1,\ldots,n\right)|K\right)\\
&\quad=E\left(\delta\text{ rejects }H|K\right)=P\left(\delta\text{ rejects }H|K\right),
\end{aligned}\tag{21}$$

by analogy,

$$\begin{aligned}
E_{\theta_0,\eta}\int_{\Omega_K}\prod_{i=1}^{n}\frac{f(X_i|\theta_0,\hat\eta)}{f(X_i|\theta_0,\eta)}\delta\,\mathrm{d}\Theta(\theta,\eta)&=\int_{\Omega_K}\prod_{i=1}^{n}\int\frac{f(x_i|\theta_0,\hat\eta)}{f(x_i|\theta_0,\eta)}\delta(x_1,\ldots,x_n)\\
&\quad\times\prod_{i=1}^{n}f(x_i|\theta_0,\eta)\prod_{i=1}^{n}\mathrm{d}x_i\,\mathrm{d}\Theta(\theta,\eta)=\alpha.
\end{aligned}\tag{22}$$

The assumption (19) was used in (22) for obtaining α. Then, the inequality (20) is simplified to

$$P\left(T_n(\Theta|\hat\eta)\text{ rejects }H|K\right)\geq P\left(\delta\text{ rejects }H|K\right)\tag{23}$$

which completes the proof. □

Remark 1 Proposition 3 makes an important extension on the nuisance parameters case. It says that for a fixed size of the test (for example, $\alpha = 5\%$) and a pre-determined estimating method of the unknown nuisance parameter η (for example, maximum likelihood estimator $\hat\eta$) the test statistic (18) provides the most powerful test when possible alternatives are described by Θ.

Comparing power of the tests at the same size is a common practice. We decided to use the assumption (19) as an additional condition for comparing different tests. Shortly, we compare tests at given α and η.

Details to Remark 1. The assumption (19) is not the classic type I error $P_H\{\delta\text{ rejects }H\}$, however, when η is unknown, the practical sense of comparing decision rules with fixed $P_H\{\delta\text{ rejects }H\}$, which is a function of η (or $\hat\eta$ for a plug-in estimator), is unclear.

Suppose we have two tests for testing H' versus K', say,

$$(A)\colon\ \text{reject } H' \text{ iff statistic } L > C^L \quad \text{and} \quad (B)\colon\ \text{reject } H' \text{ iff statistic } D > C^D,$$

where C^L and C^D are thresholds. When comparing tests (A) and (B) in the presence of unknown parameters under H' and K', the suggestion "to specify the type I errors of (A) and (B) and contrast the powers of these tests" is problematic. The thresholds C^L and C^D cannot be chosen from $P_{H'}\{L > C_\alpha^L\} = P_{H'}\{D > C_\alpha^D\} = \alpha$, because $P_{H'}\{\cdot\}$ is a function of η. Even if we assume that C_α^L and C_α^D are known (or evaluated) at H', the power comparisons, $P_{K'}\{L > C_\alpha^L\}$ versus $P_{K'}\{D > C_\alpha^D\}$, is an arduous task, $P_{K'}\{\cdot\}$ depends on unknown η and θ_1. Alternatively, we suggest to estimate η at H' and confine statistical inference to

$$\hat{P}_{H'}\{H' \text{ is rejected}\big|\hat{\eta}\} = \int \prod_{i=1}^{n} f\left(x_i|\theta_0, \hat{\eta}(x_1,\ldots,x_n)\right)\, I\left(H' \text{ is rejected}\right) \prod_{i=1}^{n} \mathrm{d}x_i.$$

Proposition 3 formalizes these ideas and allows comparing (18) with the other test statistics (in the context of the integrated power of testing), where η is estimated by $\hat{\eta}$, for example a maximum likelihood estimator would be $\hat{\eta} = \arg\max_a \prod_{i=1}^n f_0(X_i|\theta_0, a)$ and the assumption (19) holds.

Remark 2 Following Lehmann (1986), we should restrict δ to $\sup_\eta P_{\theta_0,\eta}\{\delta \text{ rejects } H\} \le \alpha'$ to keep the Type I Error under control. For the MLE $\hat{\eta}$ and any η

$$\begin{aligned}\alpha &= \int \prod_{i=1}^{n} f\left(x_i|\theta_0, \hat{\eta}(x_1,\ldots,x_n)\right)\, I\left(\delta(x_1,\ldots,x_n) \text{ rejects } H\right) \prod_{i=1}^{n} \mathrm{d}x_i \\ &\ge \int \prod_{i=1}^{n} f\left(x_i|\theta_0, \eta\right)\, I\left(\delta(x_1,\ldots,x_n) \text{ rejects } H\right) \prod_{i=1}^{n} \mathrm{d}x_i = P_{\theta_0,\eta}\left\{\delta \text{ rejects } H\right\}.\end{aligned}$$

Hence, the first type error is never greater than α.

Remark 3 Note that, since by the definition (18) the event $(T_n(\Theta|\hat{\eta}) > k)$ corresponds to $(\prod_{i=1}^n f(x_i|\theta_0, \hat{\eta}(x_1,\ldots,x_n)) \le \int_{\Omega_K} \prod_{i=1}^n f(X_i|\theta,\eta) d\Theta(\theta,\eta)/k)$, we have the inequality

$$\alpha = \int \prod_{i=1}^{n} f\left(x_i|\theta_0, \hat{\eta}(x_1,\ldots,x_n)\right)\, I\left(T_n(\Theta|\hat{\eta}) > k\right) \prod_{i=1}^{n} \mathrm{d}x_i \le 1/k.$$

4 Integrated likelihood ratio in the presence of missing data

The dataset in Sect. 6 contains fair amount of missing data. Due to Little and Rubin (2002) likelihood based inference can be substituted by inference based on observed data likelihood if missing data mechanism is ignorable, which means that missing data are missing at random (MAR) and $(\theta, \psi) \in \Omega_K \times \Psi$, where θ and ψ are the model and missing data parameters. Employing the idea of observed data likelihood, the test statistic $T_n(\Theta)$ can be easily modified to the form acceptable for using with missing data. In a case when there are no missing data all X_i are available and the full data likelihood is

$$L_{\text{full}}(\theta|X_1,\ldots,X_n)=\prod_{i=1}^{n} f(X_i|\theta).$$

If missing data mechanism incurred some missing data, then $X_i = (X_{i,\text{obs}}, X_{i,\text{mis}})$, where $X_{i,\text{obs}}$ denotes the observed part and $X_{i,\text{mis}}$ define missing components, the values of missing components is usually denoted by *NA* (not available), $X_{i,\text{mis}} = NA$. This partitioning into the observed and missing parts allows to define the observed data likelihood

$$L_{\text{obs}}(\theta|X_1,\ldots,X_n)=\prod_{i=1}^{n}\int f\left(X_{i,\text{obs}}, X_{i,\text{mis}}|\theta\right) \mathrm{d}X_{i,\text{mis}}. \tag{24}$$

The notation from Little and Rubin (2002) is used in (24). Then the test statistic (6) is generalized to

$$T_{n,\text{obs}}(\Theta)=\int_{\Omega_K}\frac{L_{\text{obs}}(\theta|X_1,\ldots,X_n)}{L_{\text{obs}}(\theta_0|X_1,\ldots,X_n)}\mathrm{d}\Theta(\theta). \tag{25}$$

Proposition 4 *The test based on* $T_{n,\text{obs}}(\Theta)$ *is the integrated most powerful.*

Proof Following the pattern of the proof of Proposition 2 with $L_{\text{obs}}(\theta|X_1,\ldots,X_n)$ instead of the full data likelihood ($\prod_i f(X_i|\theta)$), which is not available in the presence of missing data, the proposition is easily proved. □

Example 4 Let $Z_1,\ldots,Z_n$ be independent and identically distributed random variables from a bivariate normal with an unknown vector of means μ and known variance covariance matrix Σ. We want to test $H:\mu\equiv\mathbf{0}$ versus $K:\mu\in\mathcal{R}^2\backslash\mathbf{0}$.

We assume that missing data mechanism is ignorable and compare the following approaches.

First, a maximum likelihood statistics can be calculated on complete data (both components of Z_i are observed). However, we definitely loosing information from incomplete observations.

Second, a maximum likelihood statistics can be calculated on the basis of observed data likelihood.

Third, the test statistic (25) with an assumed prior on alternative values of θ can be calculated. Again observed data likelihood will be used.

Maximum likelihood statistic based on complete data likelihood. On complete data $\hat{T}_n^{ML}$, where n defines the number of complete observations, becomes

$$\hat{T}^{ML}_{n,\text{comp}} = \prod_{i=1}^{n} \frac{\exp\left[-0.5(Z_i-\hat{\mu})^T\Sigma^{-1}(Z_i-\hat{\mu})\right]}{\exp\left[-0.5Z_i^T\Sigma^{-1}Z_i\right]}$$
$$= \exp\left[0.5n\hat{\mu}^T\Sigma^{-1}\hat{\mu}\right]. \tag{26}$$

Then, H is rejected if $\{\hat{T}^{ML}_{n,\text{comp}} > k\}$ or $\{n\hat{\mu}^T\Sigma^{-1}\hat{\mu} > 2\log k = k'\}$. Under H, $n\hat{\mu}^T\Sigma^{-1}\hat{\mu} \sim \chi^2_2$ and for an α-size test $P(\chi^2_2 > k') = \alpha$ or $k' = (\chi^2_2)^{-1}(1-\alpha)$, where χ^2_2 is a random variable distributed as chi-square with 2 d.f. and $(\chi^2_2)^{-1}$ defines inverse chi-square distribution.

Maximum observed data likelihood test statistic. In this case, in addition to complete observations the likelihood uses marginal densities obtained on partially observed components. Without lose of generality we assume that after n complete observations there were m_1 observations with only the first component observed and m_2 with the second only. Then,

$$\hat{T}^{ML}_{n,\text{obs}} = \prod_{i=1}^{n} \frac{\exp\left[-0.5(Z_i-\hat{\mu})^T\Sigma^{-1}(Z_i-\hat{\mu})\right]}{\exp\left[-0.5Z_i^T\Sigma^{-1}Z_i\right]} \prod_{i=n+1}^{n+m_1} \frac{\exp\left[-0.5\sigma_{11}^{-1}(Z_{i1}-\tilde{\mu}_1)^2\right]}{\exp\left[-0.5\sigma_{11}^{-1}Z_{i1}^2\right]}$$
$$\times \prod_{i=n+m_1+1}^{n+m_1+m_2} \frac{\exp\left[-0.5\sigma_{22}^{-1}(Z_{i2}-\tilde{\mu}_2)^2\right]}{\exp\left[-0.5\sigma_{22}^{-1}Z_{i2}^2\right]}$$
$$= \exp\left[0.5n\hat{\mu}^T\Sigma^{-1}\hat{\mu} + 0.5m_1\sigma_{11}^{-1}\tilde{\mu}_1^2 + 0.5m_2\sigma_{22}^{-1}\tilde{\mu}_2^2\right], \tag{27}$$

where $\tilde{\mu}_1(\tilde{\mu}_2)$ is the sample mean for $\mu_1(\mu_2)$ based on $Z_{n+1},\ldots,Z_{n+m_1}$ $(Z_{n+m_1+1},\ldots,Z_{n+m_1+m_2})$. The rejection region $\{\hat{T}^{ML}_{n,\text{obs}} > k\}$ is represented as

$$\left\{n\hat{\mu}^T\Sigma^{-1}\hat{\mu} + m_1\sigma_{11}^{-1}\tilde{\mu}_1^2 + m_2\sigma_{22}^{-1}\tilde{\mu}_2^2 > 2\log k = k'\right\}.$$

Since $m_1\sigma_{11}^{-1}\tilde{\mu}_1^2$ and $m_2\sigma_{22}^{-1}\tilde{\mu}_2^2$ are distributed as χ^2_1, the $k' = (\chi^2_4)^{-1}(1-\alpha)$ for α-size tests.

Integrated likelihood ratio test statistic. We consider the integrated likelihood ratio test statistic on observed data likelihood with $\Theta = N(\mathbf{b}, \Sigma_{\mathbf{b}})$. Then,

$$T_{n,\text{obs}}(\Theta) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \prod_{i=1}^{n} \frac{\exp\left[-0.5(Z_i-\mathbf{a})^T\Sigma^{-1}(Z_i-\mathbf{a})\right]}{\exp\left[-0.5Z_i^T\Sigma^{-1}Z_i\right]} \prod_{i=n+1}^{n+m_1} \frac{\exp\left[-0.5\sigma_{11}^{-1}(Z_{i1}-a_1)^2\right]}{\exp\left[-0.5\sigma_{11}^{-1}Z_{i1}^2\right]}$$
$$\times \prod_{i=n+m_1+1}^{n+m_1+m_2} \frac{\exp\left[-0.5\sigma_{22}^{-1}(Z_{i2}-a_2)^2\right]}{\exp\left[-0.5\sigma_{22}^{-1}Z_{i2}^2\right]} \frac{\exp\left[-0.5\left(\mathbf{a}-\mathbf{b}\right)^T\Sigma_{\mathbf{b}}^{-1}\left(\mathbf{a}-\mathbf{b}\right)\right]}{2\pi|\Sigma_{\mathbf{b}}|} da_1\, da_2$$
$$= \exp\left[-0.5\mathbf{b}^T\Sigma_{\mathbf{b}}^{-1}\mathbf{b}\right]\frac{|\Psi|}{|\Sigma_{\mathbf{b}}|}\exp\left[0.5\tau^T\Psi\tau\right]. \tag{28}$$

The equality (28) is a special case of a K-dimensional formula presented in Appendix A1. In Eq. (28)

$$\Psi^{-1} = n\Sigma^{-1} + \Sigma_{\mathbf{b}}^{-1} + \begin{pmatrix} \sigma_{11}m_1^{-1} & 0 \\ 0 & \sigma_{22}m_2^{-1} \end{pmatrix}^{-1}$$

and all sample information is incorporated in

$$\tau = \tau\left(\hat{\mu}_1, \hat{\mu}_2, \tilde{\mu}_1, \tilde{\mu}_2\right) = \mathbf{b}^T\Sigma_{\mathbf{b}}^{-1} + n\hat{\mu}^T\Sigma^{-1} + \left(\tilde{\mu}_1\ \tilde{\mu}_2\right)\begin{pmatrix} \sigma_{11}m_1^{-1} & 0 \\ 0 & \sigma_{22}m_2^{-1} \end{pmatrix}^{-1}$$

a two-dimensional sufficient statistic for $T_{n,\mathrm{obs}}(\Theta)$. Easy to see that

$$\tau = \mathbf{b}^T\Sigma_{\mathbf{b}}^{-1} + \hat{\mu}^T\mathrm{Var}^{-1}\left(\hat{\mu}\right) + \left(\tilde{\mu}_1\ \tilde{\mu}_2\right)\mathrm{Var}^{-1}\left(\tilde{\mu}_1\ \tilde{\mu}_2\right).$$

The τ is normal and under the null hypothesis $E(\tau) = \mathbf{b}^T\Sigma_{\mathbf{b}}^{-1}$ and

$$\mathrm{Var}\,(\tau) = \mathrm{Var}^{-1}\left(\hat{\mu}\right) + \mathrm{Var}^{-1}\left(\tilde{\mu}_1\ \tilde{\mu}_2\right).$$

Then, the test

$$\left\{\exp\left[-0.5\mathbf{b}^T\Sigma_{\mathbf{b}}^{-1}\mathbf{b}\right]\frac{|\Psi|}{|\Sigma_{\mathbf{b}}|}\exp\left[0.5\tau^T\Psi\tau\right] > k\right\}$$

corresponds to

$$\left\{\tau^T\mathrm{Var}^{-1}(\tau)\tau > k'\right\},\ \tau^T\mathrm{Var}^{-1}(\tau)\tau \sim NC\chi_2^2(\lambda),$$

where $NC\chi_2^2(\lambda)$ is a non-central chi-square distribution with 2 d.f. and $\lambda = \mathbf{b}^T\Sigma_{\mathbf{b}}^{-1}\mathbf{b}$. From $P\{\tau^T\mathrm{Var}^{-1}(\tau)\tau > k'\} = \alpha$ find $k' = (NC\chi_2^2(\lambda))^{-1}(1-\alpha)$.

Remark Integrated likelihood ratio statistic based on complete data likelihood, denote $T_{n,\mathrm{comp}}(\Theta)$, is a special case of $T_{n,\mathrm{obs}}(\Theta)$, where m_1 and m_2 are forced to be zeroes.

5 Simulation studies

Let (X, Y) be a bivariate normal random variable with $\mu = (\mu_x, \mu_y)$ and $\Sigma = \begin{pmatrix} 1.1 & 0.9 \\ 0.9 & 1.4 \end{pmatrix}$. Two missing data mechanisms are considered: (1) MCAR with $p_x = p_y = 0.5$, where p_x (p_y) defines the probability that X (Y) is missing, (2) MAR with $p_x = 0$ and $p_y = \exp(X)/(1 + \exp(X))$.

Simulation study design. For each experiment we (1) drew a sample of 50 bivariate observations, (2) generated missing values, (3) calculated $T_{n,\mathrm{comp}}^{ML}$, $T_{n,\mathrm{obs}}^{ML}$, and $T_{n,\mathrm{obs}}(\Theta)$, where $\Theta = N(0, \Sigma)$ is assumed (in Table 2 the matrix Σ was multiplied

Table 1 Monte-Carlo power of the test based on $T^{ML}_{n,\text{comp}}$, $T^{ML}_{n,\text{obs}}$, and $T_{n,\text{obs}}(\Theta)$ at different choices of μ, missing data are MCAR

$\mu = (\mu_1, \mu_2)$	$T^{ML}_{n,\text{comp}}$	$T^{ML}_{n,\text{obs}}$	$T_{n,\text{obs}}(\Theta)$
(0.0, 0.0)	0.0505	0.0502	0.0495
(0.0, 0.1)	0.0618	0.0624	0.0667
(0.1, 0.1)	0.0615	0.0678	0.0754
(0.2, 0.3)	0.1095	0.1671	0.2211
(0.0, 0.4)	0.3269	0.3570	0.4551
(0.4, 0.4)	0.2171	0.4068	0.5160
(0.0, 1.0)	0.9630	0.9904	0.9959
(1.0, 1.0)	0.7649	0.8944	0.8960
$\mu \sim N(\mathbf{0}, \Sigma)$	0.7863	0.8498	0.8787
$\mu \sim N(\mathbf{0.4}, \Sigma)$	0.8074	0.8701	0.8939
$\mu \sim N(\mathbf{1}, \Sigma)$	0.8539	0.9044	0.9230

Table 2 Monte-Carlo power of the test based on $T^{ML}_{n,\text{comp}}$, $T^{ML}_{n,\text{obs}}$, and $T_{n,\text{obs}}(\Theta)$ at different choices of μ, missing data are MAR

$\mu = (\mu_1, \mu_2)$	$T^{ML}_{n,\text{comp}}$	$T^{ML}_{n,\text{obs}}$	$T_{n,\text{obs}}(\Theta)$
(0.0, 0.0)	0.4615	0.7306	0.0446
(0.0, 0.1)	0.4834	0.7401	0.0735
(0.1, 0.1)	0.2981	0.7412	0.0862
(0.2, 0.3)	0.2383	0.8339	0.2724
(0.0, 0.4)	0.8351	0.9241	0.5847
(0.4, 0.4)	0.0712	0.9466	0.6879
(0.0, 1.0)	0.9998	1.0000	0.9998
(1.0, 1.0)	0.2641	1.0000	1.0000
$\mu \sim N(\mathbf{0}, \Sigma)$	0.8487	0.9847	0.9232
$\mu \sim N(\mathbf{0.4}, \Sigma)$	0.8164	0.9831	0.9219
$\mu \sim N(\mathbf{1}, \Sigma)$	0.7983	0.9849	0.9437

by 10), (4) tested the null $H : \mu \equiv \mathbf{0}$ versus $K : \mu \in \mathcal{R}^2\backslash\mathbf{0}$, and (5) repeated the experiment 10, 000 times for confident comparisons of the results. Such simulation studies were preformed for two different missing data mechanisms and the following choices of μ : (0, 0), (0, 0.1), (0.1, 0.1), (1, 1), and (0, 4), or $\mu \sim N(0, \Sigma)$. The results of the simulation studies are presented in Tables 1 and 2.

The $\Theta = N(0, \Sigma)$ selected for the simulation study is an appropriate solution for detecting small deviations from the null hypothesis, for example (0.4, 0.4).

6 Application

Little and Rubin (2002) published "St. Louis Risk Research Data". Data on 69 families with two children were collected. Experts classified these families into three risk groups: (1) a normal group of control families (27 families), (2) a moderate-risk group where one patient was diagnosed with a psychiatric illness (24 families), (3) a high-risk group where one parent was diagnosed as having schizophrenia or an affective mental disorder (last 18 families). The authors compared the data on continuous variables R_1, V_1, R_2, and V_2, where R_c and V_c were the results of reading and verbal tests for the c^{th} child, $c = 1, 2$.

Normal group was compared with a merged group of moderate and high risk families. Making emphasis on data augmentation technique authors presented 95% posterior probability intervals of mean differences for each of the R_1, V_1, R_2, and V_2. Three of these intervals (mean differences for V_1, V_2, and R_1) were entirely positive providing evidence that the results for verbal and reading tests are better for the control group of normal families comparing with the moderate-high risk group. Later in their book the authors also used MANOVA assumptions for this data, the results were similar.

We decided to compare the same two groups on the basis of the integrated most powerful testing proposed in this manuscript. Following the MANOVA assumptions, we considered a multivariate normal model with mean change under alternatives.

To be able to apply the hypothesis testing consistent with the proposed formulas we need (1) to know a four dimensional mean μ under the null and a 4×4 variance covariance matrix Σ (in MANOVA settings Σ is the same under the null and alternatives) and (2) have a good prior guess on four dimensional group means μ^1 and μ^2.

As it is usually done in practice, we estimated μ, μ^1, μ^2, and Σ by their observed data MLE. To obtain these estimates we used EM-algorithm. Note that freely available software (for example, package NORM in R) does not allow, or we simply did not find how to do it, to apply EM-algorithm to normal data with different group means and the same variance. However, its development and implementation is not difficult. We found

$$
\begin{aligned}
\hat{\mu}^1_{em} &= (115.5, 142.4, 108.7, 128.9), \\
\hat{\mu}^2_{em} &= (103.3, 114.2, 101.9, 109.0), \\
\hat{\Sigma}_{em} &= \begin{pmatrix} 234.8 & 327.0 & 93.9 & 255.5 \\ 327.0 & 717.4 & 178.0 & 506.0 \\ 93.9 & 178.0 & 204.2 & 300.5 \\ 255.5 & 506.0 & 300.5 & 871.9 \end{pmatrix},
\end{aligned}
$$

where $\hat{\mu}^1_{em}$ and $\hat{\mu}^2_{em}$ are EM estimators of group means and $\hat{\Sigma}_{em}$ is an EM-analogue of a pooled estimator. Since $\hat{\Sigma}_{em}$ estimates Σ under null as well, we developed EM under null given $\Sigma = \hat{\Sigma}_{em}$ and found

$$
\hat{\mu}_{em} = (107.8, 125.5, 104.6, 117.2).
$$

Table 3 P-values to reject H at different choices of $\mathbf{b}$ and $\Sigma_{\mathbf{b}}$

$\mathbf{b}^1$	$\mathbf{b}^2$	$\Sigma = \Sigma_{\mathbf{b}^1} = \Sigma_{\mathbf{b}^2}$	P-value
$\hat{\mu}_{em}^1 - \hat{\mu}_{em}$	$\hat{\mu}_{em}^2 - \hat{\mu}_{em}$	$\hat{\Sigma}_{em}$	0.0026
$\hat{\mu}_{em}^1 - \hat{\mu}_{em}$	$\hat{\mu}_{em}^2 - \hat{\mu}_{em}$	$0.8\hat{\Sigma}_{em}$	0.0005
$\hat{\mu}_{em}^1 - \hat{\mu}_{em}$	$\hat{\mu}_{em}^2 - \hat{\mu}_{em}$	$1.2\hat{\Sigma}_{em}$	0.0081
$\hat{\mu}_{em}^1 - \hat{\mu}_{em} + 5$	$\hat{\mu}_{em}^2 - \hat{\mu}_{em}$	$\hat{\Sigma}_{em}$	0.0240
$\hat{\mu}_{em}^1 - \hat{\mu}_{em} + 5$	$\hat{\mu}_{em}^2 - \hat{\mu}_{em}$	$0.8\hat{\Sigma}_{em}$	0.0082
$\hat{\mu}_{em}^1 - \hat{\mu}_{em} + 5$	$\hat{\mu}_{em}^2 - \hat{\mu}_{em}$	$1.2\hat{\Sigma}_{em}$	0.0493
$\hat{\mu}_{em}^1 - \hat{\mu}_{em}$	$\hat{\mu}_{em}^2 - \hat{\mu}_{em} + 5$	$\hat{\Sigma}_{em}$	0.0028
$\hat{\mu}_{em}^1 - \hat{\mu}_{em}$	$\hat{\mu}_{em}^2 - \hat{\mu}_{em} + 5$	$0.8\hat{\Sigma}_{em}$	0.0005
$\hat{\mu}_{em}^1 - \hat{\mu}_{em}$	$\hat{\mu}_{em}^2 - \hat{\mu}_{em} + 5$	$1.2\hat{\Sigma}_{em}$	0.0086

Table 4 P-values to reject H at different choices of $\mathbf{b}$ and $\Sigma_{\mathbf{b}}$

$\mathbf{b}^1$	$\mathbf{b}^2$	$\Sigma \sim$	$\Sigma_{\mathbf{b}^1} = \Sigma_{\mathbf{b}^2}$	P-value
$\hat{\mu}_{em}^1 - \hat{\mu}_{em}$	$\hat{\mu}_{em}^2 - \hat{\mu}_{em}$	$W^{-1}(15, \hat{\Sigma}_{em})$	$\hat{\Sigma}_{em}$	0.0000
$\hat{\mu}_{em}^1 - \hat{\mu}_{em}$	$\hat{\mu}_{em}^2 - \hat{\mu}_{em}$	$W^{-1}(6, \hat{\Sigma}_{em})$	$\hat{\Sigma}_{em}$	0.0028
$\hat{\mu}_{em}^1 - \hat{\mu}_{em}$	$\hat{\mu}_{em}^2 - \hat{\mu}_{em}$	$W^{-1}(5, \hat{\Sigma}_{em})$	$\hat{\Sigma}_{em}$	0.0032
$\hat{\mu}_{em}^1 - \hat{\mu}_{em}$	$\hat{\mu}_{em}^2 - \hat{\mu}_{em}$	$W^{-1}(5, 0.8\hat{\Sigma}_{em})$	$0.8\hat{\Sigma}_{em}$	0.0099
$\hat{\mu}_{em}^1 - \hat{\mu}_{em}$	$\hat{\mu}_{em}^2 - \hat{\mu}_{em}$	$W^{-1}(5, 1.2\hat{\Sigma}_{em})$	$1.2\hat{\Sigma}_{em}$	0.0001
$\hat{\mu}_{em}^1 - \hat{\mu}_{em} + 5$	$\hat{\mu}_{em}^2 - \hat{\mu}_{em}$	$W^{-1}(5, \hat{\Sigma}_{em})$	$\hat{\Sigma}_{em}$	0.0021
$\hat{\mu}_{em}^1 - \hat{\mu}_{em} + 5$	$\hat{\mu}_{em}^2 - \hat{\mu}_{em}$	$W^{-1}(5, 0.8\hat{\Sigma}_{em})$	$0.8\hat{\Sigma}_{em}$	0.0095
$\hat{\mu}_{em}^1 - \hat{\mu}_{em} + 5$	$\hat{\mu}_{em}^2 - \hat{\mu}_{em}$	$W^{-1}(5, 1.2\hat{\Sigma}_{em})$	$1.2\hat{\Sigma}_{em}$	0.0013
$\hat{\mu}_{em}^1 - \hat{\mu}_{em}$	$\hat{\mu}_{em}^2 - \hat{\mu}_{em} + 5$	$W^{-1}(5, \hat{\Sigma}_{em})$	$\hat{\Sigma}_{em}$	0.0016
$\hat{\mu}_{em}^1 - \hat{\mu}_{em}$	$\hat{\mu}_{em}^2 - \hat{\mu}_{em} + 5$	$W^{-1}(5, 0.8\hat{\Sigma}_{em})$	$0.8\hat{\Sigma}_{em}$	0.0079
$\hat{\mu}_{em}^1 - \hat{\mu}_{em}$	$\hat{\mu}_{em}^2 - \hat{\mu}_{em} + 5$	$W^{-1}(5, 1.2\hat{\Sigma}_{em})$	$1.2\hat{\Sigma}_{em}$	0.0001

We subtracted $\hat{\mu}_{em}$ from our data because $H : \mu \equiv \mathbf{0}$. Then, the test statistic $T_n^1(\Theta|\Sigma = \Sigma_{\mathbf{b}^s} = \hat{\Sigma}_{em}, \mathbf{b}^s = \hat{\mu}_{em}^s - \hat{\mu}_{em}, s = 1, 2)$ is asymptotically χ_4^2 distributed. This test statistic is a special case of a more general formula (31) derived in Appendix.

For different choices of $\mathbf{b}^s$ and $\Sigma_{\mathbf{b}^s}$ the test statistic defines different P-values for rejecting H. We varied $\mathbf{b}^s$ and $\Sigma_{\mathbf{b}^s}$ to explore sensitivity of the hypothesis testing to misspecification of Θ. The results are in Table 3.

An *alternative* approach is to assume that Σ is a random variable distributed as inverse Wishart $W^{-1}(df, (df + 1)\hat{\Sigma}_{em})$, where df is a degrees of freedom (the larger df the smaller will be the variability of randomly generated matrices). It is not clear which df should be used when Σ is estimated with missing data. We tried several different values (see Table 4 for some of them) and decided to stop at $df = 5$.

Then, the test statistics is represented not as a four dimensional integral with respect to Θ but as a fourteen dimensional with respect to both Θ and W^{-1}. This integral is difficult to take analytically, but it can be efficiently estimated by Monte Carlo simulations. See Table 4 for the results (10,000 for estimating simulations were used). This approach continue using $\hat{\mu}_{em}$, $\hat{\Sigma}_{em}$ for formulating the null and for describing alternatives, even though it is done differently from above.

An obvious conclusion after comparing Tables 3 and 4 is that assigning an inverse Wishart distribution on a nuisance parameter provides more stable results. More stable to changes in estimates of μ_1, μ_2, and Σ, where sampling variability is one of the sources of these possible changes. This conclusion aligns with the overall strategy of this manuscript showing that a reasonable prior may improve hypothesis testing.

7 Conclusion

The likelihood ratio test statistic was modified by integrating its numerator with respect to a distribution of the test parameter Θ. This modification provides the highest integrated power if the long term behavior is actually described by Θ. Handling nuisance parameters is investigated as well.

When unknown parameters are multidimensional or the number of observations is small, we believe the proposed test statistics is a reasonable solution in contrast with MLR.

In the presence of missing data, the proposed test statistic can be used with an observed data likelihood. As shown by a MCAR simulation study the power of the considered test is higher comparing with the MLR tests. Another simulation shows that a researcher may expect more stable and reasonable power when missing data mechanism is MAR.

The application of the proposed method to the Saint Louis data set illustrates the use of these tests for multidimensional data with missing values.

Acknowledgments Authors thank the editor and unknown to us referees for their comments which helped to improve our manuscript.

Appendix

A1 Integrated likelihood ratio test statistic in the presence of missing data (one sample case)

We consider a one-sample test for testing a null hypothesis that an i.i.d. sample $\mathbf{Z}_1, \ldots, \mathbf{Z}_n$ came from a K-dimensional normal distribution $N_K(\mathbf{0}, \Sigma)$, Σ is known, versus an alternative that the K-dimensional mean is different from $\mathbf{0}$ defined in $\mathcal{R}^K$. Assume that Θ is $N_K(\mathbf{b}, \Sigma_\mathbf{b})$ and missing data mechanism is ignorable. Missing data can appear due to a pre-specified design or by chance in observational studies. In a hypothetical case a sample can be divided into P non-overlapping subsamples (for simplicity we assumed that the subsample sizes are fixed) with respect to different patterns of missing data, denote these subsamples by $\mathbf{Z}_{1,p}, \ldots, \mathbf{Z}_{n_p,p}$, $\sum_{p=0}^{P} n_p = n$. The value n_0 defines the size of the complete data subsample. In Σ_p (or $\hat{\mu}_p$) the subscript p specifies that the components of Σ ($\hat{\mu}$) associated with missing components in the p's missing data pattern are set to zero producing in this way Σ_p ($\hat{\mu}_p$). The inverse of Σ_p is calculated by a Moon–Penrose generalized inverse method, so, zero rows and columns are continue to be zeros after the inverse, and dimensions of all Σ_p are K. In this notation $\Sigma = \Sigma_0$ and $\hat{\mu}_0 = \hat{\mu}$. Also, $\hat{\mu} = n_0^{-1}\sum_{i=1}^{n_0}\mathbf{Z}_{i,0}$ and $\tilde{\mu}_p = n_p^{-1}\sum_{i=1}^{n_p}\mathbf{Z}_{i,p}$. Since the test statistic in this case is built on the assumptions that Σ, $\mathbf{b}$, and $\Sigma_{\mathbf{b}}$ are known, it will be denoted by $T_{n,\mathrm{obs}}(\Theta|\Sigma, \mathbf{b}, \Sigma_{\mathbf{b}})$. Then,

$$
\begin{aligned}
T_{n,obs}\left(\Theta|\Sigma, \mathbf{b}, \Sigma_{\mathbf{b}}\right) &= \int_{\mathcal{R}^K} \prod_{p=0}^{P}\prod_{i=1}^{n_p} \frac{\exp\left[-0.5(\mathbf{Z}_{i,p}-\mathbf{a})^T\Sigma_p^{-1}(\mathbf{Z}_{i,p}-\mathbf{a})\right]}{\exp\left[-0.5\mathbf{Z}_{i,p}^T\Sigma_p^{-1}\mathbf{Z}_{i,p}\right]} \\
&\quad \times \exp\left[-0.5\left(\mathbf{a}-\mathbf{b}\right)^T\Sigma_{\mathbf{b}}^{-1}\left(\mathbf{a}-\mathbf{b}\right)\right](2\pi|\Sigma_{\mathbf{b}}|)^{-K/2}\,\mathbf{da} \\
&= \int_{\mathcal{R}^K} \exp\sum_{i=1}^{n_0}\left[-0.5(\mathbf{Z}_i-\mathbf{a})^T\Sigma^{-1}(\mathbf{Z}_i-\mathbf{a}) + 0.5\mathbf{Z}_i^T\Sigma^{-1}\mathbf{Z}_i\right] \\
&\quad \times\exp\sum_{p=1}^{P}\sum_{i=1}^{n_p}\left[-0.5(\mathbf{Z}_{i,p}-\mathbf{a})^T\Sigma_p^{-1}(\mathbf{Z}_{i,p}-\mathbf{a}) + 0.5\mathbf{Z}_{i,p}^T\Sigma_p^{-1}\mathbf{Z}_{i,p}\right] \\
&\quad \times\exp\left[-0.5\mathbf{a}^T\Sigma_{\mathbf{b}}^{-1}\mathbf{a}+\mathbf{b}^T\Sigma_{\mathbf{b}}^{-1}\mathbf{a} - 0.5\mathbf{b}^T\Sigma_{\mathbf{b}}^{-1}\mathbf{b}\right](2\pi|\Sigma_{\mathbf{b}}|)^{-K/2}\,\mathbf{da} \\
&= \int_{\mathcal{R}^K} \exp\left[-0.5n_0\mathbf{a}^T\Sigma^{-1}\mathbf{a} + n_0\hat{\mu}^T\Sigma^{-1}\mathbf{a}\right] \\
&\quad \times\exp\sum_{p=1}^{P}\left[-0.5n_p\mathbf{a}^T\Sigma_p^{-1}\mathbf{a}+n_p\tilde{\mu}_p^T\Sigma_p^{-1}\mathbf{a}\right] \\
&\quad \times \exp\left[-0.5\mathbf{a}^T\Sigma_{\mathbf{b}}^{-1}\mathbf{a} + \mathbf{b}^T\Sigma_{\mathbf{b}}^{-1}\mathbf{a} - 0.5\mathbf{b}^T\Sigma_{\mathbf{b}}^{-1}\mathbf{b}\right](2\pi|\Sigma_{\mathbf{b}}|)^{-K/2}\,\mathbf{da} \\
&= \int_{\mathcal{R}^K} \exp\left[-0.5\mathbf{a}^T\left(\Sigma_{\mathbf{b}}^{-1} + \sum_{p=0}^{P} n_p\Sigma_p^{-1}\right)\mathbf{a}\right] \\
&\quad \times \exp\left[\left(\mathbf{b}^T\Sigma_{\mathbf{b}}^{-1} + n_0\hat{\mu}^T\Sigma^{-1} + \sum_{p=1}^{P} n_p\tilde{\mu}_p^T\Sigma_p^{-1}\right)\mathbf{a}\right] \\
&\quad \times \exp\left[-0.5\mathbf{b}^T\Sigma_{\mathbf{b}}^{-1}\mathbf{b}\right](2\pi|\Sigma_{\mathbf{b}}|)^{-K/2}\,\mathbf{da} \\
&= \exp\left[-0.5\mathbf{b}^T\Sigma_{\mathbf{b}}^{-1}\mathbf{b}\right]\frac{|\Psi|^{K/2}}{|\Sigma_{\mathbf{b}}|^{K/2}} \\
&\quad \times \int_{\mathcal{R}^K} \exp\left[-0.5\mathbf{a}^T\Psi^{-1}\mathbf{a} + \tau^T\mathbf{a}\right]\frac{\mathbf{da}}{(2\pi|\Psi|)^{K/2}} \\
&= \exp\left[-0.5\mathbf{b}^T\Sigma_{\mathbf{b}}^{-1}\mathbf{b} + 0.5\tau^T\Psi\tau\right]\frac{|\Psi|^{K/2}}{|\Sigma_{\mathbf{b}}|^{K/2}},
\end{aligned} \tag{29}
$$

where

$$
\Psi^{-1} = \Sigma_{\mathbf{b}}^{-1} + \sum_{p=0}^{P} n_p\Sigma_p^{-1}
$$

and all sample information is incorporated in

$$\tau = \mathbf{b}^T \Sigma_{\mathbf{b}}^{-1} + n_0 \hat{\mu}^T \Sigma^{-1} + \sum_{p=1}^{P} n_p \tilde{\mu}_p^T \Sigma_p^{-1}$$

a K-dimensional sufficient statistic for $T_{n,\mathrm{obs}}(\Theta|\Sigma, \mathbf{b}, \Sigma_{\mathbf{b}})$. Easy to see that $E\tau = \mathbf{b}^T \Sigma_{\mathbf{b}}^{-1}$, $\mathrm{Var}(\tau) = \Sigma^{-1} + \sum_{p=1}^{P} \Sigma_p^{-1}$, the test is written as $\{\tau^T \mathrm{Var}^{-1} \tau > k'\}$, and $\tau^T \mathrm{Var}^{-1}(\tau)\tau \sim NC\chi_K^2(\lambda)$, where $NC\chi_K^2(\lambda)$ is a non-central chi-square distribution with K d.f. and $\lambda = \mathbf{b}^T \Sigma_{\mathbf{b}}^{-1} \mathbf{b}$. From $P\{\tau^T \mathrm{Var}^{-1}(\tau)\tau > k'\} = \alpha$ find the cutoff point $k' = (NC\chi_K^2(\lambda))^{-1}(1-\alpha)$.

We assumed above that the matrix Σ is known, which is not true in many real life problems. If Σ is not available, a straightforward solution is to use its unbiased estimate, for example, $\hat{\Sigma}$ on available cases. Then, the test will be based on

$$\hat{\tau} = \mathbf{b}^T \Sigma_{\mathbf{b}}^{-1} + n_0 \hat{\mu}^T \hat{\Sigma}^{-1} + \sum_{p=1}^{P} n_p \tilde{\mu}_p^T \hat{\Sigma}_p^{-1},$$

however, the distribution of $\hat{\tau}^T \mathrm{Var}^{-1}(\hat{\tau})\hat{\tau}$ is not a non-central chi-square anymore, only asymptotically.

The matrix $\hat{\Sigma}$ is an estimator of Σ based on available cases. Each variance (or covariance) of Σ is estimated on a different number of available cases (or available pairs of cases for estimating covariances). This fact does not allow us to find a good close form distribution of $\hat{\Sigma}$ (like Wishart in multivariate analysis), and consequently, the distribution of $\hat{\tau}$ is not easily expressible.

In this case a nonparametric bootstrap procedure can be used (e.g., Efron and Tibshirani 1993). It allows to estimate the distribution of $\hat{\tau}^T \mathrm{Var}^{-1}(\hat{\tau})\hat{\tau}$ under H ($\hat{\tau} = n_0 \hat{\mu}^T \hat{\Sigma}^{-1} + \sum_{p=1}^{P} n_p \tilde{\mu}_p^T \hat{\Sigma}_p^{-1}$) and find an estimate of k'. This approach is also not without drawbacks. If sample size is small, the resampling bootstrap procedure may face inability to estimate some components of $\hat{\Sigma}$ at some iterations. Simple exclusion of these iterations from the resampling procedure may increase bias associated with sample sample bootstraps. Another problem is that for some iterations the matrix $\hat{\Sigma}$ may be not invertible, which also leads to inability to estimate $\hat{\tau}$ on these bootstrap iterations.

A2 Integrated likelihood ratio test statistic in the presence of missing data (case of multiple samples)

Here we consider a case of multiple samples for testing a null hypothesis that all the S i.i.d. samples $\mathbf{Z}_1^s, \ldots, \mathbf{Z}_n^s$ $(s = 1, \ldots, S)$ came from a K-dimensional normal distribution $N_K(\mathbf{0}, \Sigma)$ versus an alternative that there exists at least one sample $\mathbf{Z}_1^{s'}, \ldots, \mathbf{Z}_n^{s'}$ with mean different from $\mathbf{0}$. Assume that missing data mechanism is ignorable and every subsample can have its own prior describing a set of possible alternatives. Denote these priors by $\Theta^s \sim N_K(\mathbf{b}^s, \Sigma_{\mathbf{b}}^s)$. In a simplified version, all the Θ^s can be all the same.

By analogy with the one-sample case, different patterns of missing data define P^s different subsamples $(\mathbf{Z}_{1,p}^s, \ldots, \mathbf{Z}_{n_p^s,p}^s)$ for each of the S samples, $\sum_{p=0}^{P^s} n_p^s = n^s$, where P^s stands for the number of missing data patterns in the s's sample. Then,

$$
\begin{aligned}
&T_{n,\mathrm{obs}}(\Theta|\Sigma, \Sigma_{\mathbf{b}^s}, \mathbf{b}^s, s = 1, \ldots, S) \\
&= \int_{\mathcal{R}^K} \prod_{s=1}^{S} \prod_{p=0}^{P^s} \prod_{i=1}^{n_p^s} \exp\left[-0.5(\mathbf{Z}_{i,p}^s - \mathbf{a}^s)^T \Sigma_p^{-1} (\mathbf{Z}_{i,p}^s - \mathbf{a}^s) + 0.5 \left(\mathbf{Z}_{i,p}^s\right)^T \Sigma_p^{-1} \mathbf{Z}_{i,p}^s\right] \\
&\quad \times \exp\left[-0.5 \left(\mathbf{a}^s - \mathbf{b}^s\right)^T \Sigma_{\mathbf{b}^s}^{-1} \left(\mathbf{a}^s - \mathbf{b}^s\right)\right] (2\pi|\Sigma_{\mathbf{b}^s}|)^{-K/2} \, \mathbf{da}^s. \qquad (30)
\end{aligned}
$$

Following the algebra of (29), Eq. (30) transforms to

$$
\begin{aligned}
&T_{n,\mathrm{obs}}(\Theta|\Sigma, \Sigma_{\mathbf{b}^s}, \mathbf{b}^s, s = 1, \ldots, S) \\
&= \prod_{s=1}^{S} \exp\left[-0.5 \left(\mathbf{b}^s\right)^T \Sigma_{\mathbf{b}^s}^{-1} \mathbf{b}^s + 0.5 \left(\tau^s\right)^T \Psi^s \tau^s\right] \left|\Psi^s\right|^{K/2} |\Sigma_{\mathbf{b}^s}|^{-K/2}, \qquad (31)
\end{aligned}
$$

where

$$
\left(\Psi^s\right)^{-1} = \Sigma_{\mathbf{b}^s}^{-1} + \sum_{p=0}^{P^s} n_p^s \Sigma_p^{-1}
$$

and

$$
\tau^s = \left(\mathbf{b}^s\right)^T \Sigma_{\mathbf{b}^s}^{-1} + n_0^s \left(\hat{\mu}^s\right)^T \Sigma^{-1} + \sum_{p=1}^{P^s} n_p^s \left(\tilde{\mu}_p^s\right)^T \Sigma_p^{-1}.
$$

Hence, the test depends on $\tau = \frac{1}{\sqrt{S}} \sum_{s=1}^{S} \tau^s$. Under H, τ is a normal random variable with $E\tau = \frac{1}{\sqrt{S}} \sum_{s=1}^{S} (\mathbf{b}^s)^T \Sigma_{\mathbf{b}^s}^{-1}$ and $\mathrm{Var}(\tau) = \Sigma^{-1} + \frac{1}{S} \sum_{s=1}^{S} \sum_{p=1}^{P^s} \Sigma_p^{-1}$.

As it happened in the one-sample case, the test is based on $\tau^T \mathrm{Var}^{-1}(\tau)\tau \sim NC\chi_K^2(\lambda)$ with K d.f. and a non-centrality parameter $\lambda = (E\tau)^T \mathrm{Var}^{-1}(\tau) E\tau$.

From $P\{\tau^T \mathrm{Var}^{-1}(\tau)\tau > k'\} = \alpha$ find the cutoff point $k' = (NC\chi_K^2(\lambda))^{-1}(1-\alpha)$.

If Σ is not known, its unbiased estimate can be used, for example, a pooled estimator based on the available case estimators of Σ from different groups. Then, the non-central chi-square is reached only asymptotically. Another solution would be a bootstrap.

References

Choi, S., Hall, W. J., Schick, A. (1996). Asymptotically uniformly most powerfull tests in parametric and semiparametric models. *The Annals of Statistics, 24*, 841–861.

Efron, B., Tibshirani, J. (1993). *An introduction to the bootstrap* (p. 436). New York: Chapman & Hall.
Goeman, J. J., Van De Geer, S. A., Van Houwelingen, H. C. (2006). Testing against a high dimensional alternative. *Journal of the Royal Statistical Society B, 68*, 477–493.
Krieger, A. M., Pollak, M., Benjamin, Y. (2003). Surveillance of a simple linear regression. *Journal of the Americal Statistical Association, 98*, 456–469.
Laud, P., Ibrahim, J. (1995). Predictive model selection. *Journal of the Royal Statistical Society B, 57*, 247–262.
Lehmann, E. L. (1986). *Testing statistical hypotheses.* New York: Springer.
Little, R. J. A., Rubin, D. B. (2002). *Statistical analysis with missing data.* New Jersey: Wiley-Interscience.
Lorden, G. (1967). Integrated risk of asymptotically Bayes sequential tests. *The Annals of Mathematical Statistics, 38*, 1399–1422.
Marden, J. I. (2000). Hypothesis testing: From p values to Bayes factors. *Journal of the Americal Statistical Association, 95*, 1316–1320.
Pollak, M. (1978). Optimality and almost optimality of mixture stopping rules. *The Annals of Statistics, 6*, 910–916.
Yakir, B. (1998). On the average run length to false alarm in surveillance problems which possess an invariant structure. *The Annals of Statistics, 26*, 1198–1214.