# Bartlett correction to the likelihood ratio test for MCAR with two-step monotone sample

Nobumichi Shutoh

Graduate School of Maritime Sciences, Kobe University

email address: shutoh@maritime.kobe-u.ac.jp

Takahiro Nishiyama

Department of Business Administration, Senshu University

Masashi Hyodo

Department of Mathematical Sciences, Graduate School of Engineering, Osaka Prefecture University

*Assuming that two-step monotone missing data are drawn from a multivariate normal population, this paper derives the Bartlett-type correction to the likelihood ratio test for Missing Completely At Random (MCAR) which plays an important role in the statistical analysis of incomplete datasets. The advantages of our approach are confirmed in Monte Carlo simulations. Our correction drastically improved the accuracy of the type I error in Little's (1988) test for MCAR, and performed well even on moderate sample sizes.*

**Keywords and Phrases** Asymptotic expansion; Bartlett correction; Missing completely at random; Monotone missing data.

**Mathematics Subject Classification** 62H15; 62E20.

## 1 Introduction

When statistically analyzing missing data, the missing mechanism is important because it justifies or invalidates the application of the statistical method. Although specifying the missing mechanism in a likelihood function is a natural approach, misspecifying the missing mechanism leads to severe bias in the result. Even when the missing mechanism can be specified exactly, its parameters must be estimated along with the population parameters. These missing mechanism parameters are nuisance parameters.

To conduct a missing data analysis without specifying the missing mechanism, we must determine the ignorability of the missing mechanism. The ignorabolity condition holds if the Missing At Random (MAR) and parameter distinctness are both satisfied (for details, see Little and Rubin, 2002). Under ignorability, we can apply methods based on direct maximum likelihood. Typically, the estimators returned by direct maximum likelihood have no closed forms, implying that their exact ditribution cannot be theoretically obtained (see e.g., Srivastava and Carter, 1986). Kanda and Fujikoshi (1998) obtained closed forms and the exact distribution of the direct maximum likelihood estimators in monotone missing data, a special case that often manifests as dropout of the samples. However, the obtained estimators take more complicated forms than those of complete data. In the last two decades, researchers have developed direct likelihood methods for statistically analyzing monotone missing data with ignorability. As discussed in Hao and Krishnamoorthy (2001), Batsidis et al. (2006), and Tsukada (2014), most of these methods were developed for two-step monotone missing data under the following settings; for $j = 1, \ldots, N$, observe i.i.d. copies of $\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \Sigma)$ denoted by $\boldsymbol{x}'_j = (\boldsymbol{x}'_{1j},\ \boldsymbol{x}'_{2j})$. For $j = N_1 + 1, \ldots, N$, $\boldsymbol{x}_{2j}$ are missing from $N_2 \equiv N - N_1$ samples, where $\boldsymbol{x}_{2j}$ is a $(p - d)$-dimensional partitioned sample vector with $p > d > 0$.

More simply, for small sample sizes with missing data, we can apply statistical methods to complete datasets after listwise deletion, or simpler estimators based on pairwise deletion. However, in applying these methods to missing data, we must restrict the conditions of the missing mechanism, i.e., Missing Completely At Random (MCAR). To this end, we focus on testing the statistical inference for the satisfaction of MCAR. The classical MCAR test was pioneered by Little (1988). He developed a likelihood ratio test that asymptotically follows the chi-squared distribution under MCAR, which were implemented in statistical software. An alternative test, based on the generalized least squares criterion, was proposed by Kim and Bentler (2002). Recently, Li and Yu (2015) proposed an approximate test for MCAR under the nonnormal model. However, approximate tests for MCAR tend to fail at small sample sizes. For instance, the false rejection of the MCAR hypothesis in Little's test (i.e., type I error) is likely to increase on small datasets.

This paper considers a Bartlett-type correction of the likelihood ratio test pr-posed by Little (1988), which dramatically reduces the occurence of type I error without a complicated critical value. The correction is applied to two-step monotone missing data, for which various statistical methods have been developed. Because the test statistic depends not only on the ratio of determinants of Wishart matrices but also on the quadratic form of the difference of the sample mean vectors, we alternatively derive them by Nagao's (1973) perturbation method.

The remainder of this paper is organized as follows. Section 2 simplifies the test statistic derived by Little (1988) to a form useful for our purpose. Section 3 lists the auxiliary results and derives the main result of this paper. Section 4 demonstrates the advantages of our correction test in Monte Carlo simulations of small sample sizes. Conclusions are presented in Section 5, and the proofs are detailed in Appendix A.

## 2 Likelihood ratio test statistic for MCAR

This section derives the likelihood ratio test for MCAR. As shown in Li and Yu (2015) of Proposition 1, the MCAR test in this case reduces to the testing of

$$H: \boldsymbol{x}_j \overset{\text{i.i.d.}}{\sim} N_p(\boldsymbol{\mu}, \Sigma)\ (j = 1, \ldots, N_1) \text{ and } \boldsymbol{x}_{1j} \overset{\text{i.i.d.}}{\sim} N_d(\boldsymbol{\mu}_1, \Sigma_{11})\ (j = N_1 + 1, \ldots, N)$$

versus

$$A: \boldsymbol{x}_j \overset{\text{i.i.d.}}{\sim} N_p(\boldsymbol{\mu}, \Sigma)\ (j = 1, \ldots, N_1) \text{ and } \boldsymbol{x}_{1j} \overset{\text{i.i.d.}}{\sim} N_d(\boldsymbol{\nu}_1, \Gamma_{11})\ (j = N_1 + 1, \ldots, N).$$

At least one of the two equations $\boldsymbol{\mu}_1 = \boldsymbol{\nu}_1$ and $\Sigma_{11} = \Gamma_{11}$ is violated, where $\boldsymbol{\mu}$ and $\Sigma$ are decomposed as

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}.$$

Here, $\boldsymbol{\mu}_1$ is a $d(< p)$-dimensional subvector of $\boldsymbol{\mu}$ and $\Sigma_{11}$ is a $d \times d$ submatrix of $\Sigma$.

Little (1988) proposed the likelihood ratio test statistic $-2 \ln \Lambda$, where

$$\Lambda = \frac{\mathcal{L}_H(\widetilde{\boldsymbol{\mu}}, \widetilde{\Sigma}, \widetilde{\boldsymbol{\mu}}_1, \widetilde{\Sigma}_{11})}{\mathcal{L}_A(\widehat{\boldsymbol{\mu}}, \widehat{\Sigma}, \widehat{\boldsymbol{\nu}}_1, \widehat{\Gamma}_{11})}.$$

Let $\mathcal{L}_H(\widetilde{\boldsymbol{\mu}}, \widetilde{\Sigma}, \widetilde{\boldsymbol{\mu}}_1, \widetilde{\Sigma}_{11})$ and $\mathcal{L}_A(\widehat{\boldsymbol{\mu}}, \widehat{\Sigma}, \widehat{\boldsymbol{\nu}}_1, \widehat{\Gamma}_{11})$ be the likelihoods with maximum likelihood estimators (MLEs) under $H$ and $A$, respectively. The MLEs of $\boldsymbol{\mu}, \Sigma, \boldsymbol{\mu}_1, \Sigma_{11}$ under $H$, denoted by tildes placed over the parameters, were derived by Anderson and Olkin (1985). The MLEs of $\boldsymbol{\mu}, \Sigma, \boldsymbol{\nu}_1, \Gamma_{11}$ under $A$ are distinguished by hat symbols over the parameters.

In the assumed special case of the two-step monotone sample, we have

$$\begin{aligned} -2 \ln \Lambda &= q + N_1[\mathrm{tr}(S_F \widetilde{\Sigma}^{-1}) - p - \ln |S_F^{-1}| + \ln |\widetilde{\Sigma}^{-1}|] \\ &\quad + N_2[\mathrm{tr}(S_{L,11} \widetilde{\Sigma}_{11}^{-1}) - d - \ln |S_{L,11}^{-1}| + \ln |\widetilde{\Sigma}_{11}^{-1}|], \end{aligned} \tag{2.1}$$

where

$$
\begin{aligned}
q &= N_1(\overline{\boldsymbol{x}}_F - \widetilde{\boldsymbol{\mu}})'\widetilde{\Sigma}^{-1}(\overline{\boldsymbol{x}}_F - \widetilde{\boldsymbol{\mu}}) + N_2(\overline{\boldsymbol{x}}_{1L} - \widetilde{\boldsymbol{\mu}}_1)'\widetilde{\Sigma}_{11}^{-1}(\overline{\boldsymbol{x}}_{1L} - \widetilde{\boldsymbol{\mu}}_1), \\
\overline{\boldsymbol{x}}_F &= \begin{pmatrix} \overline{\boldsymbol{x}}_{1F} \\ \overline{\boldsymbol{x}}_{2F} \end{pmatrix}, \quad \overline{\boldsymbol{x}}_{\ell F} = \frac{1}{N_1}\sum_{j=1}^{N_1} \boldsymbol{x}_{\ell j}, \quad \overline{\boldsymbol{x}}_{1L} = \frac{1}{N_2}\sum_{j=N_1+1}^{N} \boldsymbol{x}_{1j}, \\
S_F &= \begin{pmatrix} S_{F,11} & S_{F,12} \\ S'_{F,12} & S_{F,22} \end{pmatrix}, S_{F,\ell m} = \frac{1}{N_1} W_{F,\ell m}, \ S_{L,11} = \frac{1}{N_2} W_{L,11}, \\
W_{F,\ell m} &= \sum_{j=1}^{N_1} (\boldsymbol{x}_{\ell j} - \overline{\boldsymbol{x}}_{\ell F})(\boldsymbol{x}_{mj} - \overline{\boldsymbol{x}}_{mF})', W_{L,11} = \sum_{j=N_1+1}^{N} (\boldsymbol{x}_{1j} - \overline{\boldsymbol{x}}_{1L})(\boldsymbol{x}_{1j} - \overline{\boldsymbol{x}}_{1L})', \\
\widetilde{\boldsymbol{\mu}} &= \begin{pmatrix} \widetilde{\boldsymbol{\mu}}_1 \\ \widetilde{\boldsymbol{\mu}}_2 \end{pmatrix}, \quad \widetilde{\boldsymbol{\mu}}_1 = \frac{N_1}{N}\overline{\boldsymbol{x}}_{1F} + \frac{N_2}{N}\overline{\boldsymbol{x}}_{1L}, \quad \widetilde{\boldsymbol{\mu}}_2 = \overline{\boldsymbol{x}}_{2F} - \widetilde{\Sigma}'_{12}\widetilde{\Sigma}_{11}^{-1}(\overline{\boldsymbol{x}}_{1F} - \widetilde{\boldsymbol{\mu}}_1), \\
\widetilde{\Sigma} &= \begin{pmatrix} \widetilde{\Sigma}_{11} & \widetilde{\Sigma}_{12} \\ \widetilde{\Sigma}'_{12} & \widetilde{\Sigma}_{22} \end{pmatrix}, \quad \widetilde{\Sigma}_{11} = \frac{N_1}{N} S_{F,11} + \frac{N_2}{N} S_{L,11}, \quad \widetilde{\Sigma}_{12} = \widetilde{\Sigma}_{11} S_{F,11}^{-1} S_{F,12}, \\
\widetilde{\Sigma}_{22} &= S_{F,22} + S'_{F,12} S_{F,11}^{-1}(\widetilde{\Sigma}_{11} - S_{F,11}) S_{F,11}^{-1} S_{F,12}
\end{aligned}
$$

for $\ell, m = 1, 2$.

In a complete dataset, the distribution of the likelihood ratio test statistic is usually invariant under an affine transformation $C\boldsymbol{X}$, where $C$ is a $p \times p$ nonsingular matrix. Unfortunately, such transformation invariance does not generally hold for two-step monotone sample. However, by restricting $C$, we can recover a similar property.

**Lemma 2.1.** *Suppose that $C$ is a $p \times p$ nonsingular matrix with the block decomposition:*

$$
C = \begin{pmatrix} C_{11} & \mathrm{O}_{12} \\ C_{21} & C_{22} \end{pmatrix},
$$

*where $C_{11}$, $C_{21}$, and $C_{22}$ are a $d \times d$ constant matrix, a $(p-d) \times d$ constant matrix, and a $(p-d) \times (p-d)$ constant matrix, respectively. $\mathrm{O}_{12}$ denotes a $d \times (p-d)$ matrix filled with zeros. Then, the distribution of the test statistic $-2 \ln \Lambda$ is invariant under the transformation $\boldsymbol{X} \mapsto C\boldsymbol{X}$.*

To simplify the form of the test statistic presented in (2.1), we state an auxiliary lemma. Furthermore, by matrix manipulations such as inverting the matrix via block decomposition (see e.g., Lemma 7 of Shutoh (2012)), we can simplify the form of the likelihood ratio test statistic presented in (2.1).

**Theorem 2.2.** *Suppose that we observe a two-step monotone sample from a multivariate normal distribution; that is, we draw* $\boldsymbol{x}_j$ $(j = 1, \ldots, N_1)$ *samples from a p-dimensional normal distribution and observe* $\boldsymbol{x}_{1j}$ $(j = N_1 + 1, \ldots, N)$ *on the first d characteristics of the same distribution. The likelihood ratio test statistic for H is then obtained as*

$$\begin{aligned} -2\ln\Lambda &= \boldsymbol{z}'\left(\frac{1}{N}(W_{F,11} + W_{L,11} + \boldsymbol{z}\boldsymbol{z}')\right)^{-1}\boldsymbol{z} + N\left\{\ln\left|\frac{1}{N}(W_{F,11} + W_{L,11} + \boldsymbol{z}\boldsymbol{z}')\right|\right. \\ &\qquad \left. + \operatorname{tr}[(W_{F,11} + W_{L,11})(W_{F,11} + W_{L,11} + \boldsymbol{z}\boldsymbol{z}')^{-1}] - d\right\} \\ &\quad -N_1\ln\left|\frac{1}{N_1}W_{F,11}\right| - N_2\ln\left|\frac{1}{N_2}W_{L,11}\right| \qquad (2.2) \end{aligned}$$

*where*

$$\boldsymbol{z} = \sqrt{\frac{N_1N_2}{N}}(\overline{\boldsymbol{x}}_{1F} - \overline{\boldsymbol{x}}_{1L}).$$

**Remark 2.3.** *The test statistic for H obtained by Theorem 2.2 is independent of* $\boldsymbol{x}_{2j}$ $(j = 1, \ldots, N_1)$.

## 3 Distribution of the test statistic and its Bartlett's type correction

This section derives the main result of this article, i.e., the Bartlett correction to the MCAR testing based on two-step monotone sample assumed in Section 1. The proof of this result relies heavily on the properties of the likelihood ratio test statistic described in Section 2. In particular, by Lemma 2.1, we can assume that $\Sigma = I_p$ holds without loss of generality.

For simplicity, we consider the distribution of $T$ obtained by replacing $N_1, N_2$ and $N$ with $n_1 = N_1 - 1$, $n_2 = N_2 - 1$ and $n = n_1 + n_2$ in the coefficients of (2.2), respectively:

$$\begin{aligned} T &= \boldsymbol{z}'\left(\frac{1}{n}(W_{F,11} + W_{L,11} + \boldsymbol{z}\boldsymbol{z}')\right)^{-1}\boldsymbol{z} + n\left\{\ln\left|\frac{1}{n}(W_{F,11} + W_{L,11} + \boldsymbol{z}\boldsymbol{z}')\right|\right. \\ &\qquad \left. + \operatorname{tr}[(W_{F,11} + W_{L,11})(W_{F,11} + W_{L,11} + \boldsymbol{z}\boldsymbol{z}')^{-1}] - d\right\} \\ &\quad -n_1\ln\left|\frac{1}{n_1}W_{F,11}\right| - n_2\ln\left|\frac{1}{n_2}W_{L,11}\right|. \end{aligned}$$

Note that $W_{F,11} \sim W_d(n_1, I_d), W_{L,11} \sim W_d(n_2, I_d), \boldsymbol{z} \sim N_d(\mathbf{0}, I_d)$, which are mutually independently distributed.

To obtain the asymptotic null distribution of $T$ in a large-sample asymptotic framework for two-step monotone sample:

$$n_1, n_2 \to \infty,\ \gamma_g = \frac{n_g}{n} \to c_g \in (0,1) \quad (g = 1,2), \tag{3.1}$$

we rewrite the Wishart matrices with

$$Y_1 \;=\; (y_{ij}^{(1)}) \;=\; \sqrt{\frac{n_1}{2}}\ln\left(\frac{1}{n_1}W_{F,11}\right),\ \ Y_2 \;=\; (y_{ij}^{(2)}) \;=\; \sqrt{\frac{n_2}{2}}\ln\left(\frac{1}{n_2}W_{L,11}\right).$$

The natural logarithm of matrices is defined in Nagao (1973). Furthermore, the symmetry of $Y_1$ and $Y_2$ holds by Lemma 2.1 of Nagao (1973). After some algebra, we obtain

$$T \;=\; T_0 + \frac{1}{\sqrt{n}}T_1 + \frac{1}{n}T_2 + \frac{1}{n\sqrt{n}}T_3 + O_p(n^{-2}),$$

where $t_0 = \boldsymbol{z}'\boldsymbol{z} + \mathrm{tr}\mathbb{Y}_2 - \mathrm{tr}\mathbb{Y}_1^2$,

$$\begin{aligned}
T_1 &= -\sqrt{2}\boldsymbol{z}'\mathbb{Y}_1\boldsymbol{z} + \frac{\sqrt{2}}{3}\mathrm{tr}\mathbb{Y}_3 - \sqrt{2}\mathrm{tr}\mathbb{Y}_1\mathbb{Y}_2 + \frac{2\sqrt{2}}{3}\mathrm{tr}\mathbb{Y}_1^3,\\
T_2 &= -\boldsymbol{z}'\mathbb{Y}_2\boldsymbol{z} - \frac{1}{2}(\boldsymbol{z}'\boldsymbol{z})^2 + 2\boldsymbol{z}'\mathbb{Y}_1^2\boldsymbol{z} + \frac{1}{6}\mathrm{tr}\mathbb{Y}_4 - \frac{1}{2}\mathrm{tr}\mathbb{Y}_2^2 - \frac{2}{3}\mathrm{tr}\mathbb{Y}_1\mathbb{Y}_3 + 2\mathrm{tr}\mathbb{Y}_1^2\mathbb{Y}_2 - \mathrm{tr}\mathbb{Y}_1^4,
\end{aligned}$$

$T_3$ is a homogeneous polynomial of degree 5 in terms of $(\boldsymbol{z}, Y_1, Y_2)$, and $\mathbb{Y}_i = \sum_g \gamma_g^{1-\frac{i}{2}} Y_g^i$ for $g = 1,2$ and $i = 1,2,3,4$, The subscript $g$ denotes the group of missing patterns in $\sum$ runs 1–2. As $\boldsymbol{z}$ and $\mathbb{Y}_i$'s are independently distributed, the characteristic function of $T$ is given by

$$\varphi(t) \equiv \mathrm{E}[\exp(itT)] = \mathrm{E}_{(Y_1,Y_2)}[\mathrm{E}_{\boldsymbol{z}}[\exp(itT)]].$$

The expectation with respect to $\boldsymbol{z}$ is described by the following formula:

$$\begin{aligned}
\mathrm{E}_{\boldsymbol{z}}[\exp(itT)] &= (1-2it)^{-\frac{d}{2}}\int_{\mathbb{R}^d}\exp(it\{\mathrm{tr}\mathbb{Y}_2 - \mathrm{tr}\mathbb{Y}_1^2\})\left[1 + \frac{(it)}{\sqrt{n}}T_1\right.\\
&\quad + \frac{1}{n}\left\{(it)T_2 + \frac{(it)^2}{2}T_1^2\right\} + \frac{1}{n\sqrt{n}}\left\{(it)T_3 + (it)^2T_1T_2\right.\\
&\quad \left.\left. + \frac{(it)^3}{6}T_1^3\right\} + O_p(n^{-2})\right]\phi(\boldsymbol{z};\mathbf{0},(t)_1 I_d)d\boldsymbol{z},
\end{aligned}$$

where $\phi(\boldsymbol{u};\boldsymbol{\eta},\Xi)$ denotes the probability density function of $\boldsymbol{u}$, which follows a multivariate normal distribution and takes parameters $(\boldsymbol{\eta},\Xi)$, and $(t)_i = (1-2it)^{-i}$.

Furthermore, using the results of Lemma 2 derived by Hyodo et al. (2015) and the technique stated in Section 7 of Nagao (1973), and extending the results in Section 2 of Nagao (1973), the asymptotic probability density function of $Y_g$ $(g=1,2)$ is obtained as

$$c^* \cdot \mathrm{etr}\left[\frac{1}{2}(n_g-d+1)\sqrt{\frac{2}{n_g}}Y_g - \frac{n_g}{2}e^{\sqrt{\frac{2}{n_g}}Y_g}\right] \tag{3.2}$$
$$\times\left[1+\frac{d-1}{2}\sqrt{\frac{2}{n_g}}\mathrm{tr}Y_g + \frac{1}{12n_g}\{(3d^2-6d+2)(\mathrm{tr}Y_g)^2 + d\,\mathrm{tr}Y_g^2\} + O_p(n^{-2})\right]$$

where $c^*$ is defined in formula (2.5) of Nagao (1973). The probability density function of $(Y_1,Y_2)$ is expressed as

$$\begin{aligned}\mathrm{E}_{(Y_1,Y_2)}[\mathrm{E}_{\boldsymbol{z}}[\exp(itT)]] &= (1-2it)^{-\frac{f}{2}}\int_{\mathbb{R}^{d(d+1)}}\left[1+\frac{1}{\sqrt{n}}\mathcal{A}_1+\frac{1}{n}\mathcal{A}_2\right.\\ &\qquad \left. +\frac{1}{n\sqrt{n}}\mathcal{A}_3 + O_p(n^{-2})\right]\phi(\boldsymbol{y};\boldsymbol{0},\Psi)d\boldsymbol{y}\end{aligned}$$

where $f=d(d+3)/2$, $\boldsymbol{y}'=(\boldsymbol{y}_1',\boldsymbol{y}_2')$, $\boldsymbol{y}_g'=(y_{11}^{(g)},\ldots,y_{pp}^{(g)},y_{12}^{(g)},\ldots,y_{p-1,p}^{(g)})$,

$$\begin{aligned}\Psi &= \mathrm{Cov}(y_{ij}^{(a)},y_{k\ell}^{(b)}) = \frac{1}{2}(\delta_{ik}\delta_{j\ell}+\delta_{i\ell}\delta_{jk})\{(t)_1(\delta_{ab}-\sqrt{\gamma_a\gamma_b})+\sqrt{\gamma_a\gamma_b}\},\\ \mathcal{A}_1 &= (it)\mathrm{E}_{\boldsymbol{z}}[T_1] - \frac{\sqrt{2}}{6}\sum_g\frac{\mathrm{tr}Y_g^3}{\sqrt{\gamma_g}},\\ \mathcal{A}_2 &= -\frac{\tilde{\gamma}}{24}d(2d^2+3d-1)+\frac{p_1}{12}\sum_g\frac{\mathrm{tr}Y_g^2}{\gamma_g}-\frac{1}{12}\sum_g\frac{\mathrm{tr}Y_g^4}{\gamma_g}-\frac{1}{12}\sum_g\frac{(\mathrm{tr}Y_g)^2}{\gamma_g}\\ &\quad +\frac{1}{36}\left\{\sum_g\frac{(\mathrm{tr}Y_g^3)^2}{\gamma_g}+\frac{2(\mathrm{tr}Y_1^3)(\mathrm{tr}Y_2^3)}{\sqrt{\gamma_1\gamma_2}}\right\}\\ &\quad +(it)\mathrm{E}_{\boldsymbol{z}}[T_2]+(it)\mathrm{E}_{\boldsymbol{z}}[T_1]\left(-\frac{\sqrt{2}}{6}\sum_g\frac{\mathrm{tr}Y_g^3}{\sqrt{\gamma_g}}\right)+\frac{(it)^2}{2}\mathrm{E}_{\boldsymbol{z}}[T_1^2],\end{aligned}$$

$\mathcal{A}_3$ is the sum of the homogeneous polynomials of degrees 1, 3, 5, 7, and 9 in the elements of $(Y_1,Y_2)$, $\delta_{ij}$ denotes the Kronecker delta, and $\tilde{\gamma}=\sum_g\gamma_g^{-1}$.

The expectations of $\mathcal{A}_1$ and $\mathcal{A}_3$ can now be calculated using the following auxiliary lemma:

**Lemma 3.1.** *Suppose that $z_1, \ldots, z_s$ are i.i.d. copies of a random variable $Z$ following a distribution that satisfies $\mathrm{E}[z_j^r] = 0$ if $r$ is an odd number and $\mathrm{E}[z_j^r] \neq 0$ otherwise, for $j = 1, \ldots, s$ and $r \in \mathbb{N}$. Furthermore, for $k \in \mathbb{N}$ and $i_j \in \mathbb{N}$ $(j = 1, \ldots, s)$, we define the set of integer partitions of $k$:*

$$P_k = \left\{ (i_1, \ldots, i_s) \middle| \sum_{j=1}^{s} i_j = k \right\}$$

*and another set*

$$E_k = \left\{ (i_1, \ldots, i_s) \middle| \mathrm{E}\left[ \prod_{j=1}^{s} z_j^{i_j} \right] = 0, \sum_{j=1}^{s} i_j = k \right\}.$$

*Then, if $k$ is odd, there exists $P_k = E_k$.*

Proof: If all of the $i_j$'s $(j = 1, \ldots, s)$ are even numbers, then $k$ is clearly even also. Note that $\mathrm{E}[\prod_{j=1}^{s} z_j^{i_j}] = 0$ holds if and only if $i_j$ is odd for at least one $j$.

By Lemma 3.1, we have

$$\int_{\mathbb{R}^{d(d+1)}} \mathcal{A}_1 \cdot \phi(\boldsymbol{y}; \boldsymbol{0}, \Psi) d\boldsymbol{y} = \int_{\mathbb{R}^{d(d+1)}} \mathcal{A}_3 \cdot \phi(\boldsymbol{y}; \boldsymbol{0}, \Psi) d\boldsymbol{y} = 0.$$

Finally, after applying the moments of the multivariate normal random variables to all terms in $\mathcal{A}_2$, we obtain the characteristic function $\varphi(t)$ and the Bartlett correction to $T$.

**Theorem 3.2.** *Under the large-sample asymptotic framework stated in (3.1), the characteristic function $\varphi(t)$ is expanded as*

$$\varphi(t) = (1 - 2it)^{-\frac{f}{2}} \left[ 1 - \frac{d}{24n} c(\tilde{\gamma}) \{1 - (1 - 2it)^{-1}\} \right] + O(n^{-2}),$$

*where $c(\tilde{\gamma}) = (2d^2 + 3d - 1)(\tilde{\gamma} - 1) + 6d$.*

**Corollary 3.3.** *Under the large-sample asymptotic framework stated in (3.1), the distribution function of*

$$T_B = \left( 1 - \frac{c(\tilde{\gamma})}{6n(d+3)} \right) T$$

*is expanded as $\Pr[T_B \leq x] = P_f(x) + O(n^{-2})$, where $P_f(x)$ denotes the distribution function of the chi-squared distribution with $f$ degrees of freedom.*

Theorem 3.2 and Corollary 3.3 are completed in the Appendix. Finally, we propose a test based on the Bartlett correction: reject $H$ if $T_B > \chi^2_f(\alpha)$, where $\alpha$ is the significance level, and $\chi^2_f(\alpha)$ is the upper $100\alpha\%$ point of the chi-squared distribution with $f$ degrees of freedom. In the next section, we demonstrate that $T_B$ better-controls the type I error than $-2\ln\Lambda$ in a simulation study.

## 4 Simulation study

In this section, the superiority of our test statistic $T_B$ over $T$ is demonstrated in Monte Carlo simulations of type I error correction for selected parameter values. For this purpose, we simulated the upper $100\alpha\%$ points of the test statistics $T$ and $T_B$, denoted by $T(\alpha)$ and $T_B(\alpha)$ respectively, under $H$. Here, $T(\alpha)$ and $T_B(\alpha)$ denote the $\lfloor r\cdot\alpha\rfloor$-th largest value of $T$'s and $T_B$'s, respectively, in $r$ replications. The attained significant levels (ASLs) of the test statistics $T$ and $T_B$ are respectively defined by

$$\mathrm{ASL}_T(\alpha) = \frac{\sharp[T > \chi^2_f(\alpha)]}{r}, \quad \mathrm{ASL}_B(\alpha) = \frac{\sharp[T_B > \chi^2_f(\alpha)]}{r}.$$

For each case in all simulations, we set $r = 1,000,000$ and varied $\alpha$ as $0.1, 0.05$ and $0.01$ (corresponding to $\lfloor r\cdot\alpha\rfloor = r\cdot\alpha = 100,000, 50,000$ and $10,000$).

In our first simulation study, we set $(p, d) = (4,1), (4,2), (4,3)$, and assumed equal sample sizes $N_1 = N_2 = 10, 15, 20, 25$. We evaluated the $\mathrm{ASL}_B(\alpha)$ and $T_B(\alpha)$ and compared them with $\mathrm{ASL}_T(\alpha)$ and $T(\alpha)$, respectively. The results are listed in Tables 1–9. In all cases, our test statistic $T_B$ outperformed $T$. Furthermore, the ASL of $T_B$ closely approximated $\alpha$ when the sample size is small. In particular, at smaller dimensionalities $d$, our proposed test clearly performed better than the $T$ statistic.

In our next simulation study, we set $(p, d) = (4, 2)$ and $N = 50$ and $100$, with $N_1/N_2 = 1$ and $1/4$. The $\mathrm{ASL}_B(\alpha)$ and $T_B(\alpha)$ were calculated similarly to the first cases, and the results are listed in Tables 10–12. Again, our test statistic $T_B$ consistently outperformed the $T$ statistic. However, when the sample sizes were unbalanced, the performances of both test statistics were degraded relative to their performances on equal sample sizes.

## 5 Conclusion

In this paper, we discussed the testing the statistical inference for the satisfaction of Missing Completely At Random (MCAR). If the missing mechanism is MCAR, we can apply statistical methods on the complete dataset after listwise deletion; otherwise, we can apply simpler estimators based on pairwise deletion. Therefore, MCAR plays a very important role in missing-data handling.

The classical MCAR test was derived by Little (1988), who considered a likelihood ratio test statistic that asymptotically distributed as chi-squared distribution under MCAR. However, approximate tests for MCAR perform poorly on small sample sizes. To resolve this problem, we proposed a Bartlett-type correction of the likelihood ratio test proposed by Little (1988), and applied it to two-step monotone missing data. Our proposed test drastically reduced the type I error without computing a complicated critical value. Furthermore, in Monte Carlo simulations, the size of the proposed test approximated the nominal significance level even for small sample sizes.

In conclusion, we recommend our proposed test for two-step monotone missing data. A test procedure based on the Bartlett-type correction will be developed for general-monotone missing data in future study.

## Acknowledgments

The authors would like to thank Dr. Tamae Kawasaki, Tokyo University of Science, for her help in numerical simulations. The research of the first two authors was supported in part by a Grant-in-Aid for Young Scientists (B) (16K17642, 26730020) from the Japan Society for the Promotion of Science. The third author was funded by Seed Grant Program for Junior Researchers of Osaka Prefecture University.

# A Proofs

## A.1 Proof of Theorem 2.2

Applying the following formula:

$$\widetilde{\Sigma} = \begin{pmatrix} I_d & O_{12} \\ \widetilde{\Sigma}_{12}'\widetilde{\Sigma}_{11}^{-1} & I_{p-d} \end{pmatrix} \begin{pmatrix} \widetilde{\Sigma}_{11} & O_{12} \\ O_{21} & \widetilde{\Sigma}_{22\cdot 1} \end{pmatrix} \begin{pmatrix} I_d & \widetilde{\Sigma}_{11}^{-1}\widetilde{\Sigma}_{12} \\ O_{21} & I_{p-d} \end{pmatrix}, \tag{A.1}$$

where $\widetilde{\Sigma}_{22\cdot 1} = \widetilde{\Sigma}_{22} - \widetilde{\Sigma}_{12}'\widetilde{\Sigma}_{11}^{-1}\widetilde{\Sigma}_{12}$, and applying the formula stated in (A.1) to $S_F$, we have

$$-\ln|S_F| + \ln|\widetilde{\Sigma}| \;=\; -\ln\left|\frac{1}{N_1}W_{F,11}\right| + \ln\left|\frac{1}{N}(W_{F,11} + W_{L,11} + \boldsymbol{z}\boldsymbol{z}')\right|. \tag{A.2}$$

Performing matrix inversion and $\widetilde{\Sigma}$ decomposition, we also obtain

$$\mathrm{tr}(S_F\widetilde{\Sigma}^{-1}) \;=\; \frac{N}{N_1}\mathrm{tr}[W_{F,11}(W_{F,11} + W_{L,11} + \boldsymbol{z}\boldsymbol{z}')^{-1}] + (p-d), \tag{A.3}$$

and

$$q \;=\; \boldsymbol{z}'\left(\frac{1}{N}(W_{F,11} + W_{L,11} + \boldsymbol{z}\boldsymbol{z}')\right)^{-1}\boldsymbol{z}. \tag{A.4}$$

Equations (A.2)–(A.4) complete the proof of Theorem 2.2.

## A.2 Proofs of Theorem 3.2 and Corollary 3.3

To prove the theorem, we first define

$$J_{(Y_1,Y_2)}[g(\boldsymbol{y})] = \int_{\mathbb{R}^{d(d+1)}} g(\boldsymbol{y}) \cdot \phi(\boldsymbol{y}; \boldsymbol{0}, \Psi) d\boldsymbol{y},$$

and

$$J_{\boldsymbol{z}}[h(\boldsymbol{y}, \boldsymbol{z})] = \int_{\mathbb{R}^d} h(\boldsymbol{y}, \boldsymbol{z}) \cdot \phi(\boldsymbol{z}; \boldsymbol{0}, (t)_1 I_d) d\boldsymbol{z},$$

where $g(\boldsymbol{y})$ is a function of the elements of $\boldsymbol{y}$ and $h(\boldsymbol{y}, \boldsymbol{z})$ is a function of the elements of $Y_1, Y_2$ and $\boldsymbol{z}$.

Using the result for multivariate normal random vectors, we obtain

$$
\begin{aligned}
J_{(Y_1,Y_2)}\left[\sum_g \frac{\mathrm{tr}Y_g^2}{\gamma_g}\right] &= \frac{d}{2}(d+1)\{(\tilde{\gamma}-2)(t)_1+2\},\\
J_{(Y_1,Y_2)}\left[\sum_g \frac{\mathrm{tr}Y_g^4}{\gamma_g}\right] &= \frac{d}{4}(2d^2+5d+5)\{(\tilde{\gamma}-3)(t)_2+2(t)_1+1\},\\
J_{(Y_1,Y_2)}\left[\sum_g \frac{(\mathrm{tr}Y_g)^2}{\gamma_g}\right] &= d\{(\tilde{\gamma}-2)(t)_1+2\},\\
J_{(Y_1,Y_2)}\left[\sum_g \frac{(\mathrm{tr}Y_g^3)^2}{\gamma_g}+\frac{2(\mathrm{tr}Y_1^3)(\mathrm{tr}Y_2^3)}{\sqrt{\gamma_1\gamma_2}}\right] &= \frac{3}{4}d(4d^2+9d+7)[(\tilde{\gamma}-4)(t)_3+3(t)_2+1]\\
&\quad +\frac{9}{2}d(d+1)^2\{-(t)_2+(t)_1\}.
\end{aligned}
$$

Therefore, we have

$$
\begin{aligned}
&J_{(Y_1,Y_2)}\Bigg[\frac{d}{12}\sum_g \frac{\mathrm{tr}Y_g^2}{\gamma_g}-\frac{1}{12}\sum_g \frac{\mathrm{tr}Y_g^4}{\gamma_g}-\frac{1}{12}\sum_g \frac{(\mathrm{tr}Y_g)^2}{\gamma_g}\\
&\qquad +\frac{1}{36}\left\{\sum_g \frac{(\mathrm{tr}Y_g^3)^2}{\gamma_g}+\frac{2(\mathrm{tr}Y_1^3)(\mathrm{tr}Y_2^3)}{\sqrt{\gamma_1\gamma_2}}\right\}\Bigg]\\
&= \frac{d}{48}(4d^2+9d+7)(\tilde{\gamma}-4)(t)_3-\frac{d}{48}(2d^2+5d+5)(\tilde{\gamma}-6)(t)_2\\
&\qquad +\frac{d}{24}(d-1)(d+2)(\tilde{\gamma}-1)(t)_1+\frac{d}{24}(3d^2+4d-3). \qquad (A.5)
\end{aligned}
$$

By Lemma 2 of Hyodo et al. (2015), we also have

$$
\begin{aligned}
J_{\boldsymbol{z}}[T_2] &= -(t)_1\mathrm{tr}\left(\sum_g Y_g^2\right)-\frac{1}{2}p_1(p_1+2)(t)_2+2(t)_1\mathrm{tr}\left[\left(\sum_g \sqrt{\gamma_g}Y_g\right)^2\right]\\
&\quad +\frac{1}{6}\mathrm{tr}\left(\sum_g \frac{1}{\gamma_g}Y_g^4\right)-\frac{1}{2}\mathrm{tr}\left[\left(\sum_g Y_g^2\right)^2\right]-\frac{2}{3}\mathrm{tr}\left[\left(\sum_g \sqrt{\gamma_g}Y_g\right)\left(\sum_g \frac{1}{\sqrt{\gamma_g}}Y_g^3\right)\right]\\
&\quad +2\mathrm{tr}\left[\left(\sum_g \sqrt{\gamma_g}Y_g\right)^2\left(\sum_g Y_g^2\right)\right]-\mathrm{tr}\left[\left(\sum_g \sqrt{\gamma_g}Y_g\right)^4\right].
\end{aligned}
$$

As the folowing relationships hold

$$
\begin{aligned}
J_{(Y_1,Y_2)}\left[\mathrm{tr}\left[\sum_g Y_g^2\right]\right] &= \frac{d}{2}(d+1)\{(t)_1+1\}, \\
J_{(Y_1,Y_2)}\left[\mathrm{tr}\left[\left(\sum_g \sqrt{\gamma_g}Y_g\right)^2\right]\right] &= \frac{d}{2}(d+1), \qquad (\mathrm{A.6}) \\
J_{(Y_1,Y_2)}\left[\mathrm{tr}\left[\sum_g \frac{1}{\gamma_g}Y_g^4\right]\right] &= \frac{d}{4}(2d^2+5d+5)\{(\tilde{\gamma}-3)(t)_2+2(t)_1+1\}, \\
J_{(Y_1,Y_2)}\left[\mathrm{tr}\left[\left(\sum_g Y_g^2\right)^2\right]\right] &= \frac{d}{4}(2d^2+5d+5)\{(t)_2+1\}+\frac{d}{2}(d+1)^2(t)_1, \\
J_{(Y_1,Y_2)}\left[\mathrm{tr}\left[\left(\sum_g \sqrt{\gamma_g}Y_g\right)\left(\sum_g \frac{1}{\sqrt{\gamma_g}}Y_g^3\right)\right]\right] &= \frac{d}{4}(2d^2+5d+5)\{(t)_1+1\}, \\
J_{(Y_1,Y_2)}\left[\mathrm{tr}\left[\left(\sum_g \sqrt{\gamma_g}Y_g\right)^2\left(\sum_g Y_g^2\right)\right]\right] &= \frac{d}{4}(2d^2+5d+5)+\frac{d}{4}(d+1)^2(t)_1,
\end{aligned}
$$

$$
J_{(Y_1,Y_2)}\left[\mathrm{tr}\left[\left(\sum_g \sqrt{\gamma_g}Y_g\right)^4\right]\right] = \frac{d}{4}(2d^2+5d+5),
$$

we can write

$$
\begin{aligned}
J_{(Y_1,Y_2)}[J_{\boldsymbol{z}}[T_2]] &= \left[\frac{d}{24}(2d^2+5d+5)(\tilde{\gamma}-6)-\frac{d}{2}(2d+3)\right](t)_2 \\
&\quad +\left[-\frac{d}{12}(2d^2+5d+5)+\frac{d}{4}(d+1)(d+3)\right](t)_1. \qquad (\mathrm{A.7})
\end{aligned}
$$

Further, by Lemma 2 of Hyodo et al. (2015), we can write

$$
\begin{aligned}
J_{\boldsymbol{z}}[T_1] &= -\sqrt{2}(t)_1\mathrm{tr}\left(\sum_g \sqrt{\gamma_g}Y_g\right)+\frac{\sqrt{2}}{3}\mathrm{tr}\left(\sum_g \frac{Y_g^3}{\sqrt{\gamma_g}}\right) \\
&\quad -\sqrt{2}\mathrm{tr}\left[\left(\sum_g \sqrt{\gamma_g}Y_g\right)\left(\sum_g Y_g^2\right)\right]+\frac{2\sqrt{2}}{3}\mathrm{tr}\left[\left(\sum_g \sqrt{\gamma_g}Y_g\right)^3\right],
\end{aligned}
$$

moreover, from

$$
\begin{aligned}
J_{(Y_1,Y_2)}\left[\mathrm{tr}\left(\sum_g \frac{1}{\sqrt{\gamma_g}}Y_g^3\right)\mathrm{tr}\left(\sum_g \sqrt{\gamma_g}Y_g\right)\right] &= \frac{3}{2}d(d+1)\{(t)_1+1\}, \qquad (\mathrm{A.8}) \\
J_{(Y_1,Y_2)}\left[\left[\mathrm{tr}\left(\sum_g \frac{1}{\sqrt{\gamma_g}}Y_g^3\right)\right]^2\right] &= \frac{3}{4}d(4d^2+9d+7)\{(\tilde{\gamma}-4)(t)_3+3(t)_2+1\} \\
&\quad +\frac{9}{2}d(d+1)^2\{-(t)_2+(t)_1\}, \qquad (\mathrm{A.9})
\end{aligned}
$$

$$
\begin{aligned}
J_{(Y_1,Y_2)}\Bigg[\mathrm{tr}\Bigg(\sum_g \frac{1}{\sqrt{\gamma_g}}Y_g^3\Bigg)\mathrm{tr}\Bigg[\Bigg(\sum_g \sqrt{\gamma_g}Y_g\Bigg)\Bigg(\sum_g Y_g^2\Bigg)\Bigg]\Bigg] &= \frac{3}{4}d(4d^2+9d+7)\{(t)_2+1\} \\
&\quad +\frac{3}{2}d(d+1)^2 \\
&\quad \times\{-(t)_2+2(t)_1\}, \qquad \text{(A.10)} \\
J_{(Y_1,Y_2)}\Bigg[\mathrm{tr}\Bigg(\sum_g \frac{1}{\sqrt{\gamma_g}}Y_g^3\Bigg)\mathrm{tr}\Bigg[\Bigg(\sum_g \sqrt{\gamma_g}Y_g\Bigg)^3\Bigg]\Bigg] &= \frac{3}{4}d(4d^2+9d+7) \\
&\quad +\frac{9}{4}d(d+1)^2(t)_1, \qquad \text{(A.11)}
\end{aligned}
$$

we obtain

$$
\begin{aligned}
J_{(Y_1,Y_2)}\Bigg[J_{\boldsymbol{z}}[T_1]\Bigg(-\frac{\sqrt{2}}{6}\sum_g \frac{\mathrm{tr}Y_g^3}{\sqrt{\gamma_g}}\Bigg)\Bigg] &= -\frac{p_1}{12}(4d^2+9d+7)(\tilde{\gamma}-4)(t)_3 \quad \text{(A.12)} \\
&\quad +\frac{d}{2}(d+1)\{(t)_2+(t)_1\}.
\end{aligned}
$$

Furthermore, again by Lemma 2 of Hyodo et al. (2015), we have

$$
\begin{aligned}
J_{\boldsymbol{z}}\left[\frac{1}{2}T_1^2\right] &= (t)_2\Bigg[\mathrm{tr}\Bigg(\sum_g \sqrt{\gamma_g}Y_g\Bigg)\Bigg]^2 + 2(t)_2\mathrm{tr}\Bigg[\Bigg(\sum_g \sqrt{\gamma_g}Y_g\Bigg)^2\Bigg] \\
&\quad +\frac{1}{9}\Bigg[\mathrm{tr}\Bigg(\sum_g \frac{Y_g^3}{\sqrt{\gamma_g}}\Bigg)\Bigg]^2 + \Bigg[\mathrm{tr}\Bigg[\Bigg(\sum_g \sqrt{\gamma_g}Y_g\Bigg)\Bigg(\sum_g Y_g^2\Bigg)\Bigg]\Bigg]^2 \\
&\quad +\frac{4}{9}\Bigg[\mathrm{tr}\Bigg[\Bigg(\sum_g \sqrt{\gamma_g}Y_g\Bigg)^3\Bigg]\Bigg]^2 - \frac{2}{3}(t)_1\mathrm{tr}\Bigg(\sum_g \sqrt{\gamma_g}Y_g\Bigg)\mathrm{tr}\Bigg(\sum_g \frac{Y_g^3}{\sqrt{\gamma_g}}\Bigg) \\
&\quad +2(t)_1\mathrm{tr}\Bigg(\sum_g \sqrt{\gamma_g}Y_g\Bigg)\mathrm{tr}\Bigg[\Bigg(\sum_g \sqrt{\gamma_g}Y_g\Bigg)\Bigg(\sum_g Y_g^2\Bigg)\Bigg] \\
&\quad -\frac{4}{3}(t)_1\mathrm{tr}\Bigg(\sum_g \sqrt{\gamma_g}Y_g\Bigg)\mathrm{tr}\Bigg[\Bigg(\sum_g \sqrt{\gamma_g}Y_g\Bigg)^3\Bigg] \\
&\quad -\frac{2}{3}\mathrm{tr}\Bigg(\sum_g \frac{Y_g^3}{\sqrt{\gamma_g}}\Bigg)\mathrm{tr}\Bigg[\Bigg(\sum_g \sqrt{\gamma_g}Y_g\Bigg)\Bigg(\sum_g Y_g^2\Bigg)\Bigg] \\
&\quad +\frac{4}{9}\mathrm{tr}\Bigg(\sum_g \frac{Y_g^3}{\sqrt{\gamma_g}}\Bigg)\mathrm{tr}\Bigg[\Bigg(\sum_g \sqrt{\gamma_g}Y_g\Bigg)^3\Bigg] \\
&\quad -\frac{4}{3}\mathrm{tr}\Bigg[\Bigg(\sum_g \sqrt{\gamma_g}Y_g\Bigg)\Bigg(\sum_g Y_g^2\Bigg)\Bigg]\mathrm{tr}\Bigg[\Bigg(\sum_g \sqrt{\gamma_g}Y_g\Bigg)^3\Bigg].
\end{aligned}
$$

Along with (A.6) and (A.8)–(A.11), the following relationships hold:

$$
\begin{aligned}
J_{(Y_1,Y_2)}\left[\left[\mathrm{tr}\left(\sum_g \sqrt{\gamma_g}Y_g\right)\right]^2\right] &= d, \\
J_{(Y_1,Y_2)}\left[\left\{\mathrm{tr}\left[\left(\sum_g \sqrt{\gamma_g}Y_g\right)\left(\sum_g Y_g^2\right)\right]\right\}^2\right] &= \frac{3}{4}d(4d^2+9d+7) \\
&\quad +\frac{d}{4}(2d^2+5d+5)(t)_2 \\
&\quad +\frac{3}{2}d(d+1)^2(t)_1, \\
J_{(Y_1,Y_2)}\left[\left\{\mathrm{tr}\left[\left(\sum_g \sqrt{\gamma_g}Y_g\right)^3\right]\right\}^2\right] &= \frac{3}{4}d(4d^2+9d+7), \\
J_{(Y_1,Y_2)}\left[\mathrm{tr}\left[\sum_g \sqrt{\gamma_g}Y_g\right]\mathrm{tr}\left[\left(\sum_g \sqrt{\gamma_g}Y_g\right)\left(\sum_g Y_g^2\right)\right]\right] &= \frac{3}{2}d(d+1)+\frac{d}{2}(d+1)(t)_1, \\
J_{(Y_1,Y_2)}\left[\mathrm{tr}\left[\sum_g \sqrt{\gamma_g}Y_g\right]\mathrm{tr}\left[\left(\sum_g \sqrt{\gamma_g}Y_g\right)^3\right]\right] &= \frac{3}{2}d(d+1), \\
J_{(Y_1,Y_2)}\left[\mathrm{tr}\left[\left(\sum_g \sqrt{\gamma_g}Y_g\right)\left(\sum_g Y_g^2\right)\right]\mathrm{tr}\left[\left(\sum_g \sqrt{\gamma_g}Y_g\right)^3\right]\right] &= \frac{3}{4}d(4d^2+9d+7) \\
&\quad +\frac{3}{4}d(d+1)^2(t)_1
\end{aligned}
$$

and therefore

$$
J_{(Y_1,Y_2)}\left[J_{\boldsymbol{z}}\left[\frac{T_1^2}{2}\right]\right] = \frac{d}{12}(4d^2+9d+7)(\tilde{\gamma}-4)(t)_3 + d(d+2)(t)_2. \quad \text{(A.13)}
$$

Combining (A.5), (A.7), (A.12) and (A.13) completes the proof of Theorem 3.2. The result of Corollary 3.3 is easily derived from Theorem 3.2 by a method similar to Fujikoshi et al. (2010).

## References


Anderson, T. W. and Olkin, I. (1985), Maximum-likelihood estimation of the parameters of a multivariate normal distribution, Linear Algebra Appl. **70**, 147–171.

Batsidis, A., Zografos, K. and Loukas, S. (2006), Errors in discrimination with monotone missing data from multivariate normal populations, Comput. Statist. Data Anal. **50**, 2600–2634.

Fujikoshi, Y., Ulyanov, V. V. and Shimizu, R. (2010), Multivariate Statistics High-Dimensional and Large-Sample Approximations, John Wiley & Sons, Inc., Hoboken, NJ., 221–223.

Hao, J. and Krishnamoorthy, K. (2001), Inferences on a normal covariance matrix and generalized variance with monotone missing data, J. Multivariate Anal. **78**, 62–82.

Hyodo, M., Shutoh, N., Nishiyama, T. and Pavlenko, T. (2015), Testing block-diagonal covariance structure for high-dimensional data, Stat. Neerl. **69**, 460–482.

Kanda, T. and Fujikoshi, Y. (1998), Some basic properties of the MLE's for a multivariate normal distribution with monotone missing data, Amer. J. Management. Sci. **18**, 161–190.

Kim, K. H. and Bentler, P. M. (2002), Tests of homogeneity of means and covariance matrices for multivariate incomplete data, Psychometrika **67**, 609–624.

Li, J. and Yu, Y. (2015), A nonparametric test of missing completely at random for incomplete multivariate data, Psychometrika **80**, 707–726.

Little, R. J. A. (1988), A test of missing completely at random for multivariate data with missing values, J. Amer. Statist. Assoc. **83**, 1198–1202.

Little, R. J. A. and Rubin, D. B. (2002), Statistical Analysis with Missing Data Second Edition, John Wiley & Sons, Inc., Hoboken, NJ., 117–120.

Nagao, H. (1973), On some test criteria for covariance matrix, Ann. Statist. **1**, 700–709.

Shutoh, N. (2012), An asymptotic approximation for EPMC in linear discriminant analysis based on monotone missing data, J. Statist. Plann. Infer. **142**, 110–125.

Srivastava, M. S. and Carter, E. M. (1986), The maximum likelihood method for non-response in sample survey, Survey Methodology **12**, 61–72.

Tsukada, S. (2014), Equivalence testing of mean vector and covariance matrix for multi-populations under a two-step monotone incomplete sample, J. Multivariate Anal. **132**, 183–196.

Table 1: Simulated upper percentiles and ASLs of $T$ and $T_B$ for $(p, d) = (4, 1)$ and $\alpha = 0.1$ ($\chi^2_f(\alpha) = 4.605$).

| $N_1$ | $N_2$ | $T(\alpha)$ | $T_B(\alpha)$ | $\mathrm{ASL}_T(\alpha)$ | $\mathrm{ASL}_B(\alpha)$ |
|---|---|---|---|---|---|
| 10 | 10 | 6.483 | 4.602 | 0.187 | 0.100 |
| 15 | 15 | 5.547 | 4.601 | 0.147 | 0.100 |
| 20 | 20 | 5.237 | 4.605 | 0.132 | 0.100 |
| 25 | 25 | 5.075 | 4.598 | 0.124 | 0.100 |

Table 2: Simulated upper percentiles and ASLs of $T$ and $T_B$ for $(p, d) = (4, 1)$ and $\alpha = 0.05$ ($\chi^2_f(\alpha) = 5.991$).

| $N_1$ | $N_2$ | $T(\alpha)$ | $T_B(\alpha)$ | $\mathrm{ASL}_T(\alpha)$ | $\mathrm{ASL}_B(\alpha)$ |
|---|---|---|---|---|---|
| 10 | 10 | 8.710 | 5.989 | 0.118 | 0.050 |
| 15 | 15 | 7.273 | 5.984 | 0.084 | 0.050 |
| 20 | 20 | 6.837 | 5.992 | 0.072 | 0.050 |
| 25 | 25 | 6.618 | 5.983 | 0.066 | 0.050 |

Table 3: Simulated upper percentiles and ASLs of $T$ and $T_B$ for $(p, d) = (4, 1)$ and $\alpha = 0.01$ ($\chi^2_f(\alpha) = 9.210$).

| $N_1$ | $N_2$ | $T(\alpha)$ | $T_B(\alpha)$ | $\mathrm{ASL}_T(\alpha)$ | $\mathrm{ASL}_B(\alpha)$ |
|---|---|---|---|---|---|
| 10 | 10 | 14.996 | 9.203 | 0.043 | 0.010 |
| 15 | 15 | 11.480 | 9.196 | 0.023 | 0.010 |
| 20 | 20 | 10.571 | 9.182 | 0.018 | 0.010 |
| 25 | 25 | 10.199 | 9.183 | 0.016 | 0.010 |

Table 4: Simulated upper percentiles and ASLs of $T$ and $T_B$ for $(p, d) = (4, 2)$ and $\alpha = 0.1$ ($\chi_f^2(\alpha) = 9.236$).

| $N_1$ | $N_2$ | $T(\alpha)$ | $T_B(\alpha)$ | $\text{ASL}_T(\alpha)$ | $\text{ASL}_B(\alpha)$ |
|---|---|---|---|---|---|
| 10 | 10 | 12.954 | 9.235 | 0.239 | 0.100 |
| 15 | 15 | 11.142 | 9.239 | 0.174 | 0.100 |
| 20 | 20 | 10.527 | 9.240 | 0.150 | 0.100 |
| 25 | 25 | 10.228 | 9.245 | 0.138 | 0.100 |

Table 5: Simulated upper percentiles and ASLs of $T$ and $T_B$ for $(p, d) = (4, 2)$ and $\alpha = 0.05$ ($\chi_f^2(\alpha) = 11.071$).

| $N_1$ | $N_2$ | $T(\alpha)$ | $T_B(\alpha)$ | $\text{ASL}_T(\alpha)$ | $\text{ASL}_B(\alpha)$ |
|---|---|---|---|---|---|
| 10 | 10 | 15.947 | 11.085 | 0.156 | 0.050 |
| 15 | 15 | 13.434 | 11.086 | 0.102 | 0.050 |
| 20 | 20 | 12.647 | 11.075 | 0.084 | 0.050 |
| 25 | 25 | 12.267 | 11.078 | 0.075 | 0.050 |

Table 6: Simulated upper percentiles and ASLs of $T$ and $T_B$ for $(p, d) = (4, 2)$ and $\alpha = 0.01$ ($\chi_f^2(\alpha) = 15.086$).

| $N_1$ | $N_2$ | $T(\alpha)$ | $T_B(\alpha)$ | $\text{ASL}_T(\alpha)$ | $\text{ASL}_B(\alpha)$ |
|---|---|---|---|---|---|
| 10 | 10 | 23.956 | 15.099 | 0.061 | 0.010 |
| 15 | 15 | 18.609 | 15.074 | 0.030 | 0.010 |
| 20 | 20 | 17.337 | 15.085 | 0.022 | 0.010 |
| 25 | 25 | 16.815 | 15.116 | 0.018 | 0.010 |

Table 7: Simulated upper percentiles and ASLs of $T$ and $T_B$ for $(p, d) = (4, 3)$ and $\alpha = 0.1$ ($\chi_f^2(\alpha) = 14.684$).

| $N_1$ | $N_2$ | $T(\alpha)$ | $T_B(\alpha)$ | $\text{ASL}_T(\alpha)$ | $\text{ASL}_B(\alpha)$ |
|---|---|---|---|---|---|
| 10 | 10 | 20.662 | 14.811 | 0.301 | 0.104 |
| 15 | 15 | 17.913 | 14.747 | 0.209 | 0.102 |
| 20 | 20 | 16.906 | 14.713 | 0.173 | 0.101 |
| 25 | 25 | 16.358 | 14.679 | 0.154 | 0.100 |

Table 8: Simulated upper percentiles and ASLs of $T$ and $T_B$ for $(p, d) = (4, 3)$ and $\alpha = 0.05$ ($\chi^2_f(\alpha) = 16.919$).

| $N_1$ | $N_2$ | $T(\alpha)$ | $T_B(\alpha)$ | $\text{ASL}_T(\alpha)$ | $\text{ASL}_B(\alpha)$ |
|---|---|---|---|---|---|
| 10 | 10 | 24.166 | 17.068 | 0.203 | 0.052 |
| 15 | 15 | 20.716 | 16.997 | 0.127 | 0.051 |
| 20 | 20 | 19.502 | 16.951 | 0.100 | 0.051 |
| 25 | 25 | 18.852 | 16.912 | 0.086 | 0.050 |

Table 9: Simulated upper percentiles and ASLs of $T$ and $T_B$ for $(p, d) = (4, 3)$ and $\alpha = 0.01$ ($\chi^2_f(\alpha) = 21.686$).

| $N_1$ | $N_2$ | $T(\alpha)$ | $T_B(\alpha)$ | $\text{ASL}_T(\alpha)$ | $\text{ASL}_B(\alpha)$ |
|---|---|---|---|---|---|
| 10 | 10 | 32.867 | 21.898 | 0.082 | 0.011 |
| 15 | 15 | 26.759 | 21.757 | 0.039 | 0.010 |
| 20 | 20 | 25.011 | 21.676 | 0.027 | 0.010 |
| 25 | 25 | 24.194 | 21.669 | 0.022 | 0.010 |

Table 10: Simulated upper percentiles and ASLs of $T$ and $T_B$ for $(p, d) = (4, 2)$ and $\alpha = 0.1$ ($\chi^2_f(\alpha) = 9.236$).

| $N_1$ | $N_2$ | $T(\alpha)$ | $T_B(\alpha)$ | $\text{ASL}_T(\alpha)$ | $\text{ASL}_B(\alpha)$ |
|---|---|---|---|---|---|
| 25 | 25 | 10.228 | 9.245 | 0.138 | 0.100 |
| 10 | 40 | 10.850 | 9.274 | 0.163 | 0.101 |
| 50 | 50 | 9.673 | 9.227 | 0.116 | 0.100 |
| 20 | 80 | 9.932 | 9.251 | 0.127 | 0.101 |

Table 11: Simulated upper percentiles and ASLs of $T$ and $T_B$ for $(p, d) = (4, 2)$ and $\alpha = 0.05$ ($\chi^2_f(\alpha) = 11.071$).

| $N_1$ | $N_2$ | $T(\alpha)$ | $T_B(\alpha)$ | $\text{ASL}_T(\alpha)$ | $\text{ASL}_B(\alpha)$ |
|---|---|---|---|---|---|
| 25 | 25 | 12.267 | 11.078 | 0.075 | 0.050 |
| 10 | 40 | 13.037 | 11.107 | 0.094 | 0.051 |
| 50 | 50 | 11.596 | 11.055 | 0.060 | 0.050 |
| 20 | 80 | 11.896 | 11.088 | 0.067 | 0.050 |

Table 12: Simulated upper percentiles and ASLs of $T$ and $T_B$ for $(p, d) = (4, 2)$ and $\alpha = 0.01$ ($\chi^2_f(\alpha) = 15.086$).

| $N_1$ | $N_2$ | $T(\alpha)$ | $T_B(\alpha)$ | $\mathrm{ASL}_T(\alpha)$ | $\mathrm{ASL}_B(\alpha)$ |
|---|---|---|---|---|---|
| 25 | 25 | 16.815 | 15.116 | 0.018 | 0.010 |
| 10 | 40 | 17.912 | 15.152 | 0.025 | 0.010 |
| 50 | 50 | 15.859 | 15.112 | 0.013 | 0.010 |
| 20 | 80 | 16.234 | 15.099 | 0.015 | 0.010 |