

Non-identifiability and testability of missing mechanisms in incomplete two-way contingency tables

Yousung Park^a, Seung Mo Oh^a, Tae Yeon Kwon^{1,b}

^aDepartment of Statistics, Korea University, Korea;

^bDepartment of International Finance, Hankuk University of Foreign Studies, Korea

Abstract

We showed that any missing mechanism is reproduced by EMAR or MNAR with equal fit for observed likelihood if there are non-negative solutions of maximum likelihood equations. This is a generalization of Molenberghs *et al.* (2008) and Jeon *et al.* (2019). Nonetheless, as MCAR becomes a nested model of MNAR, a natural question is whether or not MNAR and MCAR are testable by using the well-known three statistics, LR (Likelihood ratio), Wald, and Score test statistics. Through simulation studies, we compared these three statistics. We investigated to what extent the boundary solution affect testing MCAR against MNAR, which is the only testable pair of missing mechanisms based on observed likelihood. We showed that all three statistics are useful as long as the boundary proximity is far from 1.

Keywords: MNAR, MCAR, identifiability, observed likelihood, boundary proximity

1. Introduction

Traditionally, a missing mechanism may be classified as Missing completely at random (MCAR), Missing at random (MAR), and Missing not at random (MNAR) as defined by Little (1988); Little and Rubin (2019). This taxonomy of missing mechanisms can be applied to two-way contingency tables with item missings but not with any unit missing, under which Molenberghs *et al.* (2008); showed that any MNAR is uniquely reproduced by a MAR with equal fit in observed likelihood, implying that MAR is not testable against MNAR. Ibrahim *et al.* (2008) extended the missing mechanisms of MCAR, MAR, and MNAR to those of MCAR, Extended missing at random (EMAR), and MNAR so that such a taxonomy can be applied to two-way tables including not only item missing but unit missing. They also showed that EMAR is not testable against MNAR and proposed alternative criteria to determine whether or not the missing mechanism is EMAR.

Baker *et al.* (1992) reparameterized missing cell probabilities by which their missing mechanisms are specified in a loglinear model. Following this reparameterization, Jeon *et al.* (2019) specified the missing mechanisms of unit and item missing in the pattern mixture model of Little (1994), selection model of Little (2008), and loglinear model of Baker *et al.* (1992). In this paper, using the observed likelihood defined with such a reparameterization, we first show that based on the observed likelihood in an incomplete two-way, it is impossible to test H_0 : EMAR vs. H_a : MNAR or MCAR and H_0 :

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (NRF-2018R1C1B5043739) and this work was supported by Hankuk University of Foreign Studies Research Fund.

¹ Corresponding author: Department of International Finance, Hankuk University of Foreign Studies, 81 Mohyeon-myeon, Oedae-ro Cheoin-gu, Yongin-si Gyeonggi-do, Republic of Korea. E-mail: tykwon@hufs.ac.kr

Published 31 May 2021 / journal homepage: <http://csam.or.kr>

© 2021 The Korean Statistical Society, and Korean International Statistical Society. All rights reserved.

Table 1: $I \times J \times 2 \times 2$ Table including item and unit missingness

		$R_2 = 1$			$R_2 = 2$
		$Y_2 = 1$	\dots	$Y_2 = J$	
$R_1 = 1$	$Y_1 = 1$	z_{1111}	\dots	z_{1J11}	z_{1+12}
	\vdots	\vdots	\ddots	\vdots	\vdots
	$Y_1 = I$	z_{I111}	\dots	z_{IJ11}	z_{I+12}
$R_1 = 2$		z_{+121}	\dots	z_{+J21}	z_{++22}

MNAR vs. H_a : EMAR or MCAR. Park *et al.* (2014) proposed boundary solution conditions at which some missing cell probabilities are estimated to be zeros in a MNAR missing mechanism. Jeon *et al.* (2019) showed that the missingness in a two-way table is a EMAR if the boundary solution conditions are met. However, there still remains a problem of identification between MNAR and MCAR. Different from other missing mechanisms under which a precise imputation procedures are required in advance, the missing can be ignored under MCAR. As MCAR is a nested model of MNAR in the reparameterization of Baker *et al.* (1992), MCAR is only testable against MNAR by using the well-known three statistics, LR, Wald, and score defined by observed likelihood. Although both MCAR and MNAR missing mechanisms are out of the boundary condition of Jeon *et al.* (2019), we are not sure that the three statistics work in testing MCAR against MNAR even when the boundary condition is barely avoided. We conducted a simulation to examine the performance of LR, Wald and Score test statistics for testing H_0 : MCAR vs. H_a : MNAR, the only testable pair of missingness based on the observed likelihood. We found that they are useful as long as the boundary proximity is far from 1. The paper is composed of 4 sections. In Section 2, we show the reproducibility of observed likelihood by MNAR and/or EMAR no matter what missing mechanisms the observed likelihood is constructed from. In Section 3, after defining a boundary proximity from the boundary solution condition, we test MCAR against MNAR by using the LR, Wald, and Score tests. Simulation studies are carried out to see to what extent the boundary proximity affect the suitability and reliability of the three traditional tests. Finally, in Section 4, we close with conclusions and limitations of the study.

2. Reproducibility of observed likelihood

A two-way table with missing data for two categorical variables Y_1 and Y_2 can be summarized as below in Table 1.

The variable Y_1 has I categories and the variable Y_2 has J categories, and R_1 and R_2 indicate the missing state of Y_1 and Y_2 , respectively. If $R_i = 1$, it means that Y_i has been observed. If $R_i = 2$, Y_i has not been observed for $i = 1, 2$. In this contingency table, we have completely observed cell counts denoted by z_{ij11} when $R_1 = R_2 = 1$, supplemental margins only on Y_1 denoted by z_{i+12} when $R_1 = 1$ and $R_2 = 2$, supplemental margins only on Y_2 denoted by z_{+j21} when $R_1 = 2$ and $R_2 = 1$, and the count of missing units denoted by z_{++22} when $R_1 = R_2 = 2$.

Under a multinomial distribution, the observed likelihood function of the two-way table shown in Table 1 which include both item and unit missings is expressed as given by

$$L = \prod_{i=1}^I \prod_{j=1}^J \pi_{ij11}^{z_{ij11}} \prod_{i=1}^I \pi_{i+12}^{z_{i+12}} \prod_{j=1}^J \pi_{+j21}^{z_{+j21}} \pi_{++22}^{z_{++22}},$$

where $\pi_{ij11} = P(Y_1 = i, Y_2 = j, R_1 = 1, R_2 = 1)$, $\pi_{+j21} = \sum_{i=1}^I P(Y_1 = i, Y_2 = j, R_1 = 2, R_2 = 1)$, $\pi_{i+12} = \sum_{j=1}^J P(Y_1 = i, Y_2 = j, R_1 = 1, R_2 = 2)$, and $\pi_{++22} = \sum_{i=1}^I \sum_{j=1}^J P(Y_1 = i, Y_2 = j, R_1 = 2, R_2 = 2)$ with fixed $N = \sum_{i,j} z_{ij11} + \sum_i z_{i+12} + \sum_j z_{+j21} + z_{++22}$.

We reparameterize the missing cell probabilities of π_{ij21} and π_{ij12} by

$$\alpha_{ij} = \frac{\pi_{ij21}}{\pi_{ij11}} \quad \text{and} \quad \beta_{ij} = \frac{\pi_{ij12}}{\pi_{ij11}}, \tag{2.1}$$

so that we have $\pi_{ij21} = \alpha_{ij}\pi_{ij11}$ and $\pi_{ij12} = \beta_{ij}\pi_{ij11}$. Note that $\alpha_{ij} = \alpha_{..}, \alpha_{i.}, \alpha_{.j}$ for MCAR, MNAR, and EMAR Y_1 , respectively, and $\beta_{ij} = \beta_{..}, \beta_{i.}, \beta_{.j}$ for MCAR, EMAR, and MNAR Y_2 , respectively as shown in Jeon *et al.* (2019). Then the observed log likelihood can be written as

$$\begin{aligned} \log L = & \sum_{i=1}^I \sum_{j=1}^J z_{ij11} \log(\pi_{ij11}) + \sum_{i=1}^I z_{i+12} \log \left(\sum_{j=1}^J \beta_{ij} \pi_{ij11} \right) \\ & + \sum_{j=1}^J z_{+j21} \log \left(\sum_{i=1}^I \alpha_{ij} \pi_{ij11} \right) + z_{++22} \log (\pi_{++22}). \end{aligned} \tag{2.2}$$

The reparameterization of (2.1) requires the maximum likelihood estimates of π_{ij11} , α_{ij} , and β_{ij} in (2.2) which satisfy

$$\begin{aligned} \pi_{+j21} &= \sum_{i=1}^I \alpha_{ij} \pi_{ij11} \quad \text{for } j = 1, \dots, J, \\ \pi_{i+12} &= \sum_{j=1}^J \beta_{ij} \pi_{ij11} \quad \text{for } i = 1, \dots, I. \end{aligned} \tag{2.3}$$

We then have the following main result.

Theorem 1. *The observed likelihood given in (2.2) under any missing mechanisms of Y_1 and Y_2 is reproduced by a combination of EMAR and MNAR with equal fit if there are unique non-negative solutions of α_{ij} and β_{ij} satisfying the system of equations of (2.3).*

A combination of EMAR and MNAR in Theorem 1 includes a pair of only EMARs and of MNARs, implying that Theorem 1 is a generalization of Molenberghs *et al.* (2008) and Jeon *et al.* (2019)’s results as they allowed only a pair of EMARs for reproducing equal fit in observed likelihood. Theorem 1 also indicates that EMAR and MNAR missing mechanisms are not identifiable and are not testable against the other missing mechanisms including MCAR.

Since Y_1 and Y_2 are MCAR when $\alpha_{ij} = \alpha_{..}$ and $\beta_{ij} = \beta_{..}$, respectively, the observed likelihood under MCAR is nested in those under MNAR and EMAR, implying that MCAR can be testable against MNAR or EMAR by using the likelihood ratio, Wald, and Score tests. However, testing MCAR against MNAR may be affected by the boundary solution problem as some $\alpha_{i.}$ and $\beta_{.j}$ are forced to zero when MNAR suffers from a boundary solution. When a boundary solution occurs in MNAR, Rubin *et al.* (1995); Jeon *et al.* (2019) suggested that EMAR produced better estimates of missing cell than the true MNAR as EMAR has no boundary solution. Thus, our interest is on testing MCAR against MNAR when no boundary solution occurs but close-to-boundary solution occurs.

3. Simulation studies

Jeon *et al.* (2019) proposed a new criterion to distinguish EMAR from the other missing mechanisms as follows. The missing mechanism of Y_1 is EMAR if there is a pair of j and j' satisfying C_1 , and that

of Y_2 is EMAR if there is a pair of i and i' satisfying C_2 .

$$\begin{aligned} C_1 : \omega_{jj'}^+ &< \omega_{jj'}^{\min} \quad \text{or} \quad \omega_{jj'}^+ > \omega_{jj'}^{\max}, \\ C_2 : \omega_{i'i'}^+ &< \omega_{i'i'}^{\min} \quad \text{or} \quad \omega_{i'i'}^+ > \omega_{i'i'}^{\max}, \end{aligned}$$

where

$$\begin{aligned} \omega_{i'i'}^+ &= \frac{\pi_{i+12}}{\pi_{i'+12}}, \quad \omega_{i'i'}^{\max} = \max_j \frac{\pi_{ij11}}{\pi_{i'j11}}, \quad \text{and} \quad \omega_{i'i'}^{\min} = \min_j \frac{\pi_{ij11}}{\pi_{i'j11}} \quad \text{for } i \neq i'; \\ \omega_{jj'}^+ &= \frac{\pi_{+j21}}{\pi_{+j'21}}, \quad \omega_{jj'}^{\max} = \max_i \frac{\pi_{ij11}}{\pi_{ij'11}}, \quad \text{and} \quad \omega_{jj'}^{\min} = \min_i \frac{\pi_{ij11}}{\pi_{ij'11}} \quad \text{for } j \neq j'. \end{aligned}$$

Therefore, the missing mechanisms of Y_1 and Y_2 are MCAR or MNAR when the conditions C_1 and C_2 are violated. Note that these conditions are, unfortunately, the necessary conditions for MNAR not to fall on a boundary solution, implying that there exists a MNAR suffering from a boundary solution even though it violates the condition C_1 and C_2 . Any test to differentiate MCAR from MNAR is meaningless in such cases because some of π_{ij21} , π_{ij12} , and π_{ij22} are, by force, made equal to zero solely due to mathematical restrictions. In practice, EMAR is used when applying a MNAR to missing cells suffers from a boundary solution as discussed before.

One of our interests is if and how the distance of $\omega_{jj'}^+$ from the boundaries $\omega_{jj'}^{\min}$ and $\omega_{jj'}^{\max}$ in C_1 and that of $\omega_{i'i'}^+$ from the corresponding boundaries of $\omega_{i'i'}^{\min}$ and $\omega_{i'i'}^{\max}$ in C_2 affect the performance of LR, Wald, Score tests based on the observed likelihood. We call $\min(\omega_{jj'}^{\max}/\omega_{jj'}^+, \omega_{jj'}^+/\omega_{jj'}^{\min})$ and $\min(\omega_{i'i'}^{\max}/\omega_{i'i'}^+, \omega_{i'i'}^+/\omega_{i'i'}^{\min})$ the boundary proximity to the boundary. The closer the boundary proximity is to 1 the closer the solution for α_i and/or β_j under MNAR is to 0 (i.e., a boundary solution).

The simulations are carried out to compare the three statistics, LR, Wald, and Score, for testing MCAR against MNAR in terms of the significance level $\alpha = 0.05$ and the power of each test statistic. We assume that Y_1 is MCAR or MNAR and Y_2 is EMAR and known so that we focus on only the missing mechanism of Y_1 for simplicity of discussion. It is straightforward to show that when Y_1 is MCAR and Y_2 is EMAR, the maximum likelihood (ML) estimates maximizing the observed likelihood of (2.2) are given by

$$N\hat{\pi}_{ij11} = \frac{z_{ij11}z_{+j+1}z_{+++11}}{z_{+j11}z_{+++1}}, \quad \hat{\alpha}_{..} = \frac{z_{++21}}{z_{+++1}}, \quad \hat{\beta}_i = \frac{z_{i+12}}{N\hat{\pi}_{i+11}}.$$

When Y_1 is MNAR and Y_2 is EMAR, the ML estimates are

$$\hat{\pi}_{ij11} = \frac{z_{ij11}}{N}, \quad \hat{\alpha}_i \text{ satisfying } \sum_i z_{ij11}\hat{\alpha}_i = z_{+j21}, \quad \hat{\beta}_i = \frac{z_{i+12}}{z_{z_{i+11}}}.$$

Using these ML estimates, we test H_0 : MCAR vs. H_a : MNAR to check type 1 error and to examine the power of LR, Wald, and Score test statistics under three scenarios of boundary proximities. Table 2 summarizes the three scenarios with different degrees of MNAR for $2 \times 2 \times 2 \times 2$ and $3 \times 3 \times 2 \times 2$ contingency tables. $\alpha_1 = \alpha_2$ and $\alpha_1 = \alpha_2 = \alpha_3$ are equivalent to MCAR in Table 2. The boundary proximities defined by $\min(w_{jj'}^{\max}/w_{jj'}^+, w_{jj'}^+/\omega_{jj'}^{\min})$ are $S1 \sim S3$ for $2 \times 2 \times 2 \times 2$ and $S4 \sim S6$ for $3 \times 3 \times 2 \times 2$ contingency tables.

Scenario $S1$ (and $S4$) is the furthest from the boundary solution, whereas $S3$ (and $S6$) is the closest. The other simulation factors are sample sizes from 5,000 and 10,000 and missing rates (item missing

Table 2: Boundary proximity for degrees of MNAR

I, J	Scenario	$\alpha_1 : \alpha_2 = 1:1$	1:2	1:3	
2	S1	8.18	5.65	4.16	
	S2	2.33	1.78	1.55	
	S3	1.40	1.23	1.17	
I, J	Scenario	$\alpha_1 : \alpha_2 : \alpha_3 = 1:1:1$	1:1:2	1:1:3	1:2:3
3	S4	8.00	4.89	3.69	3.85
	S5	2.33	1.87	1.64	1.71
	S6	1.34	1.35	1.35	1.45

Table 3: **Type 1 Error.** The probabilities rejecting $H_0: Y_1$ is MCAR with nominal $\alpha = 5\%$ when Y_1 is simulated under MCAR (%)

Scenario	Missing rate (%)	Sample size = 5000			Sample size = 10000		
		LR	Wald	Score	LR	Wald	Score
S1	5	5.21	5.38	5.36	5.17	5.27	5.27
	10	5.23	5.28	5.28	4.60	4.64	4.64
	15	4.90	4.96	4.96	5.06	5.07	5.07
S2	5	4.96	5.11	5.09	5.43	5.49	5.49
	10	5.54	5.60	5.59	4.91	4.93	4.93
	15	5.16	5.20	5.19	5.02	5.03	5.03
S3	5	4.12	4.21	4.20	5.02	5.08	5.08
	10	5.24	5.28	5.28	5.12	5.15	5.15
	15	4.96	4.97	4.97	5.16	5.20	5.20
S4	5	4.94	5.30	5.29	5.02	5.13	5.13
	10	4.82	4.95	4.94	5.40	5.52	5.52
	15	5.19	5.30	5.25	5.00	5.00	5.00
S5	5	4.89	5.15	5.12	4.77	4.88	4.88
	10	5.45	5.72	5.71	4.98	5.03	5.03
	15	5.09	5.18	5.17	4.91	4.93	4.93
S6	5	3.31	3.39	3.37	4.14	4.19	4.18
	10	4.08	4.16	4.14	4.88	5.05	5.05
	15	4.51	4.54	4.52	5.21	5.27	5.27

rate) of Y_1 from 5% to 15% with a fixed 5% of Y_2 missing and 2% of unit missing. The reason for considering large-sized samples is to secure a sufficient number of missing values to identify the missing mechanism, which are between 250 and 1,500. Each simulation combination is repeated 10,000 times.

Table 3 shows type 1 errors that are probabilities rejecting $H_0: Y_1$ is MCAR when MCAR is true with the nominal level $\alpha = 0.05$. Except for S6 with missing rate = 5% and sample size = 5,000, type 1 errors of all tests are well maintained near 5%. The type errors are close to 5% as sample size increases from 5,000 to 10,000.

Table 3 S4 and S5 present powers of the three statistics for three degrees of boundary proximities in different sample sizes, missing rates and the strength of MNAR. As with the maintenance of the significance level, no noticeable difference in power was found in the three test statistics. The boundary proximity has a profound effect on the powers of test statistics. The powers are rapidly reduced as the boundary proximity approaches 1 for all the three statistics (i.e., as S1 \rightarrow S2 \rightarrow S3 and S4 \rightarrow S5 \rightarrow S6).

Table 3 S4 and S5 also show that the power increases as sample size increases, missing percentage increases, or the degree of MNAR increases as desired. However, the powers decrease as the number of levels in 2-way tables, I , increases by comparing the powers of Table 4 with those of Table 5 as

Table 4: Power of Test. The probabilities rejecting MCAR against MNAR in $2 \times 2 \times 2 \times 2$ (%)

Sample size	Scenario	Missing rate (%)	1:2			1:3		
			LR	WALD	SCORE	LR	WALD	SCORE
5000	S1	5	98.8	98.7	98.7	100	100	100
		10	100	100	100	100	100	100
		15	100	100	100	100	100	100
	S2	5	53.7	54.1	54.1	87.3	87.5	87.5
		10	80.9	81.1	81.1	99.0	99.0	99.0
		15	91.8	91.9	91.9	99.9	99.9	99.9
	S3	5	11.2	11.5	11.4	18.8	19.4	19.4
		10	25.4	25.5	25.5	45.3	45.6	45.5
		15	35.1	35.2	35.2	60.7	60.8	60.8
10000	S1	5	100	100	100	100	100	100
		10	100	100	100	100	100	100
		15	100	100	100	100	100	100
	S2	5	82.7	82.9	82.9	99.3	99.3	99.3
		10	98.2	98.2	98.2	100	100	100
		15	99.8	99.8	99.8	100	100	100
	S3	5	27.1	27.3	27.3	49.0	49.2	49.2
		10	47.4	47.5	47.5	76.9	76.9	76.9
		15	61.0	61.0	61.0	89.6	89.6	89.6

Table 5: Power of Test. The probabilities rejecting MCAR against MNAR in $3 \times 3 \times 2 \times 2$ (%)

Sample size	Scenario	Missing rate (%)	1:1:2			1:1:3			1:2:3		
			LR	WALD	SCORE	LR	WALD	SCORE	LR	WALD	SCORE
5000	S4	5	92.1	91.5	91.5	100	100	100	98.0	98.1	98.1
		10	99.7	99.7	99.7	100	100	100	100	100	100
		15	100	100	100	100	100	100	100	100	100
	S5	5	30.1	30.2	30.2	64.5	64.0	64.0	38.6	39.7	39.7
		10	51.2	50.8	50.7	90.6	90.4	90.4	66.4	67.0	67.0
		15	68.0	67.6	67.6	97.5	97.3	97.3	82.5	82.9	82.9
	S6	5	6.9	7.0	6.9	14.2	14.8	14.7	5.9	6.0	6.0
		10	12.5	12.6	12.6	29.9	30.4	30.4	8.0	8.1	8.1
		15	17.5	17.5	17.5	40.3	40.9	40.8	9.4	9.4	9.4
10000	S4	5	99.8	99.8	99.8	100	100	100	100	100	100
		10	100	100	100	100	100	100	100	100	100
		15	100	100	100	100	100	100	100	100	100
	S5	5	54.7	54.2	54.2	91.5	91.1	91.1	68.6	69.2	69.2
		10	82.7	82.2	82.2	99.7	99.7	99.7	93.1	93.2	93.2
		15	94.3	94.2	94.2	100	100	100	98.7	98.8	98.8
	S6	5	13.18	13.2	13.2	31.9	32.3	32.3	8.3	8.3	8.3
		10	22.2	22.3	22.3	56.2	56.6	56.6	11.8	11.6	11.6
		15	31.5	31.6	31.6	71.9	72.3	72.3	15.0	14.9	14.9

the larger the number of levels, the higher possibility the boundary proximity close to 1. Since the boundary proximity close to 1 implies that there are the levels of MNAR Y_1 corresponding to α_i close to zero and the levels of MNAR Y_2 corresponding to β_j close to zero, Such levels in incompletely observed data should be rare events. As seen in Table 5, in particular, the very low powers of the three statistics for the boundary proximities close to 1 indicates that one must be careful to accept MCAR if there is no prior information that no level of missing data is rare event.

4. Conclusion

We showed that the LR, Wald, and Score test statistics based on the observed likelihood in an incomplete two-way table are not applicable to test H_0 : EMAR vs. H_a : the other missing mechanisms and H_0 : MNAR vs. H_a : the other missing mechanism except for H_0 : MCAR vs. H_a : MNAR as the observed likelihood constructed from any missing mechanism is copied by MNAR and EMAR with equal fit. Fortunately, since the boundary condition provided by Jeon *et al.* (2019) can be used to identify EMAR from MNAR or MCAR, MNAR is applied first when missing imputation. If a boundary solution problem occurs, we adopt EMAR as a missing mechanism, whereas we adopt MNAR or MCAR if not.

The LR, Wald, and Score test statistics for testing MCAR against MNAR, the only testable pair of missing mechanisms based on the observed likelihood, is useful as long as the boundary proximity is far from 1. When the boundary proximity is close to 1, however, one should take a careful caution to interpret the test result as the test powers of the three test statistics are very low in such a case. This requires a test method free from the boundary proximity to identify between MCAR and MNAR.

Acknowledgement

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (NRF-2018R1C1B5043739). This work was supported by Hankuk University of Foreign Studies Research Fund.

Appendix:

Proof of Theorem 1

Suppose that the missing combination is EMAR Y_1 and EMAR Y_2 for $\alpha_{ij} = \alpha_{.j}$ and $\beta_{ij} = \beta_{i.}$. Then, the inner equations of (2.3) are $\pi_{i+12} = \beta_{i.}\pi_{i+11}$ for $i = 1, \dots, I$ and $\pi_{+j21} = \alpha_{.j}\pi_{+j11}$ for $j = 1, \dots, J$ to have $\alpha_{.j} = \pi_{+j21}/\pi_{+j11}$ and $\beta_{i.} = \pi_{i+12}/\pi_{i+11}$.

Since these $\alpha_{.j}$ and $\beta_{i.}$ are non-negative and the number of parameters, π_{ij11} , $\alpha_{.j}$, and $\beta_{i.}$, are the same as the number of observations -1 , the degrees of freedom equal zero. Thus, we can obtain ML estimates equal to observed cell proportions: $\hat{\pi}_{ij11} = z_{ij11}/N$, $\hat{\alpha}_{.j} = z_{+j21}/z_{+j11}$, and $\hat{\beta}_{i.} = z_{i+12}/z_{i+11}$. This produces the same observed likelihood with equal fit.

When at least one of missing mechanism of Y_1 and Y_2 is MNAR, the level of Y_1 , I , should be the same as the level of Y_2 , J for the unique solutions of equations (1.1). Because the degrees of freedom equal zero when $I = J$, we can obtain the solutions of the equations by setting the observed proportions equal to the corresponding cell probabilities. That is, when Y_1 is MNAR,

$$\frac{z_{+j21}}{N} = \sum_{i=1}^I \frac{\alpha_{i.} z_{ij11}}{N}, \quad j = 1, \dots, J, \tag{A.1}$$

only if the solution $\alpha_{i.}$ s are all non-negative. Otherwise, the equality in (A.1) does not hold as negative $\alpha_{i.}$ s should be set to zero due to its non-negative constraint. These zero settings produce different ML estimates of π_{ij11} and π_{+j21} from those in (A.1), which is called a boundary solution problem.

Similarly, for MNAR Y_2 , non-negative ML estimates of $\beta_{.j}$ satisfying

$$\frac{z_{i+12}}{N} = \sum_{j=1}^J \frac{\beta_{.j} z_{ij11}}{N}.$$

Therefore, regardless of which mechanisms the observed likelihood is realized from, the observed likelihood can be reproduced by EMAR Y_1 and MNAR Y_2 , MNAR Y_1 and EMAR Y_2 , and MNAR Y_1 , and MNAR Y_2 with equal fit if $I = J$ and non-negative solutions of α_i and β_j in the inner equations of (2.3). This completes the proof.

References

- Baker SG, Rosenberger WF, and Dersimonian R (1992). Closed-form estimates for missing counts in two-way contingency tables, *Statistics in Medicine*, **11**, 643–657.
- Ibrahim JG, Zhu H, and Tang N (2008). Model selection criteria for missing-data problems using the em algorithm, *Journal of the American Statistical Association*, **103**, 1648–1658.
- Jeon S, Kwon TY, and Park Y (2019). Variable-based missing mechanism for an incomplete contingency table with unit missingness, *Statistics & Probability Letters*, **146**, 90–96.
- Little RJ (1988). A test of missing completely at random for multivariate data with missing values, *Journal of the American statistical Association*, **83**, 1198–1202.
- Little RJ (1994). A class of pattern-mixture models for normal incomplete data, *Biometrika*, **81**, 471–483.
- Little RJ (2008). Selection and pattern-mixture models, *Longitudinal Data Analysis*(pp.409-431), Chapman and Hall, London.
- Little RJ and Rubin DB (2019). *Statistical Analysis with Missing Data*, **793**, Wiley Blackwell, England.
- Molenberghs G, Beunckens C, Sotito C, and Kenward MG (2008). Every missingness not at random model has a missingness at random counterpart with equal fit, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **70**, 371–388.
- Park Y, Kim D, and Kim S (2014). Identification of the occurrence of boundary solutions in a contingency table with nonignorable nonresponse, *Statistics & Probability Letters*, **93**, 34–40.
- Rubin DB, Stern HS, and Vehovar V (1995). Handling “don’t know” survey responses: the case of the slovenian plebiscite, *Journal of the American Statistical Association*, **90**, 822–828.

Received January 22, 2021; Revised March 2, 2021; Accepted March 26, 2021