IOWA STATE UNIVERSITY
**Digital Repository**

Statistics Publications

Statistics

2020

# Semiparametric fractional imputation using Gaussian mixture models for handling multivariate missing data

Hejian Sang
*Google Inc*

Jae Kwang Kim
*Iowa State University*, jkim@iastate.edu

Danhyang Lee
*University of Alabama - Tuscaloosa*

# Semiparametric fractional imputation using Gaussian mixture models for handling multivariate missing data

## Abstract

Item nonresponse is frequently encountered in practice. Ignoring missing data can lose efficiency and lead to misleading inference. Fractional imputation is a frequentist approach of imputation for handling missing data. However, the parametric fractional imputation of Kim (2011) may be subject to bias under model misspecification. In this paper, we propose a novel semiparametric fractional imputation method using Gaussian mixture models. The proposed method is computationally efficient and leads to robust estimation. The proposed method is further extended to incorporate the categorical auxiliary information. The asymptotic model consistency and √n- consistency of the semiparametric fractional imputation estimator are also established. Some simulation studies are presented to check the finite sample performance of the proposed method.

## Keywords

## Disciplines

Statistical Methodology | Statistical Models

## Comments

# Semiparametric fractional imputation using Gaussian mixture models for handling multivariate missing data

Hejian Sang[*]

Google Inc

and

Jae Kwang Kim

Department of Statistics, Iowa State University,Ames

September 18, 2018

**Abstract**

Item nonresponse is frequently encountered in practice. Ignoring missing data can lose efficiency and lead to misleading inference. Fractional imputation is a frequentist approach of imputation for handling missing data. However, the parametric fractional imputation of Kim (2011) may be subject to bias under model misspecification. In this paper, we propose a novel semiparametric fractional imputation method using Gaussian mixture models. The proposed method is computationally efficient and leads to robust estimation. The proposed method is further extended to incorporate the categorical auxiliary information. The asymptotic model consistency and $\sqrt{n}$- consistency of the semiparametric fractional imputation estimator are also established. Some simulation studies are presented to check the finite sample performance of the proposed method.

*Keywords:* Item nonresponse; Robust estimation; Variance estimation.

1

# 1  Introduction

Missing data is frequently encountered in survey sampling, clinical trials and many other areas. Imputation can be used to handle item nonresponse and several imputation methods have been developed in the literature. Motivated from a Bayesian perspective, Rubin (1996) proposed multiple imputation to create multiple complete data sets. Alternatively, under the frequentist framework, fractional imputation (Kim and Fuller, 2004; Kim, 2011) makes one complete data with multiple imputed values and their corresponding fractional weights. Little and Rubin (2002) and Kim and Shao (2013) provided comprehensive overviews of the methods for handling missing data.

For multivariate missing data with arbitrary missing patterns, valid imputation methods should preserve the correlation structure in the imputed data. Judkins et al. (2007) proposed an iterative hot deck imputation procedure that is closely related to the data augmentation algorithm of Tanner and Wong (1987) but they did not provide variance estimation. Other imputation procedures for multivariate missing data include the multiple imputation approaches of Raghunathan et al. (2001) and Murray and Reiter (2016), and parametric fractional imputation of Kim (2011). The approaches of Judkins et al. (2007) and Raghunathan et al. (2001) are based on conditionally specified models and the imputation from the conditionally specified model is generically subject to the model compatibility problem (Chen, 2010; Liu et al., 2013; Bartlett et al., 2015). Conditional models for the different missing patterns calculated directly from the observed patterns may not be compatible with each other. The parametric fractional imputation uses the joint distribution to create imputed values and does not suffer from model compatibility problems.

Note that parametric imputation requires correct model specification. Nonparametric imputation methods, such as kernel regression imputation (Cheng, 1994; Wang and Chen, 2009), are robust but may be subject to the curse of dimensionality. Hence, it is important, often critical, to develop a unified, robust and efficient imputation method that can be used for general purpose estimation. The proposed semiparametric method

fills in this important gap by considering a more flexible model for imputation.

In this paper, to achieve robustness against model misspecification, we develop an imputation procedure based on Gaussian mixture models. Gaussian mixture model is a very flexible model that can be used to handle outliers, heterogeneity and skewness. McLachlan and Peel (2004) and Bacharoglou (2010) argued that any continuous distribution can be approximated by a finite Gaussian mixture distribution. The proposed method using Gaussian mixture model makes a nice compromise between efficiency and robustness. It is semiparametric in the sense that the number of mixture components is chosen automatically from the data. The computation for parameter estimation in our proposed method is based on EM algorithm and its implementation is relatively simple and efficient.

We note that Di Zio et al. (2007) also proposed to use Gaussian mixture model to impute missing data. However, variance estimation and choice of the mixture component are not discussed in Di Zio et al. (2007). Elliott and Stettler (2007) and Kim et al. (2014) introduced the multiple imputation using mixture models. Instead of multiple imputation, we use fractional imputation for general-purpose estimation. We provide a completely theoretical justification for consistency of the proposed imputation method. The variance estimator and the model selection for the number of mixture component are also carefully discussed and demonstrated in numerical studies. The proposed method is further extended to handle mixed type data including categorical variable in Section 5. By allowing the proportion vector of mixture component to depend on categorical auxiliary variable, the proposed fractional imputation using Gaussian mixture models can incorporate the observed categorical variables and provide a very flexible tool for imputation.

The paper is structured as follows. The setup of the problem is introduced and a short review of fractional imputation are presented in Secction 2. In Section 3, the proposed semiparametric method and its algorithm for implementation are introduced. Some asymptotic results are presented in Section 4. In Section 5, the proposed method is further extended to handle mixed type data. Some numerical studies and a real data application are presented to show the performance of the proposed method

in Section 6 and Section 7, respectively. In Section 8, some concluding remarks are made. The technical derivations and proof are presented in Appendix.

## 2 Basic Setup

Consider a $p$-dimensional vector of study variable $Y = (y_1, y_2, \cdots, y_p)$. Suppose that $\{Y_1, Y_2, \cdots, Y_n\}$ are $n$ independent and identically distributed realizations of the random vector $Y$. In this paper, we use the upper case to represent vector or matrix and the lower case to denote the elements within vector or matrix. Assume that we are interested in estimating parameter $\theta \in \Theta$, which is defined through $E\{U(\theta; Y)\} = 0$, where $U(\cdot; Y)$ is the estimating function of $\theta$. With no missingness, a consistent estimator of $\theta$ can be obtained by the solution to

$$\frac{1}{n} \sum_{i=1}^{n} U(\theta; Y_i) = 0. \tag{1}$$

To avoid unnecessary details, we assume that the solution to (1) exists uniquely almost everywhere.

However, due to missingness, the estimating equation in (1) cannot be applied directly. To formulate the multivariate missingness problem, we further define the response indicator vector $R = (r_1, r_2, \cdots, r_p)$ for $Y = (y_1, y_2, \cdots, y_p)$ as

$$r_j = \begin{cases} 1 & \text{if } y_j \text{ is observed} \\ 0 & \text{otherwise,} \end{cases}$$

where $j = 1, 2, \cdots, p$. We assume that the response mechanism is missing at random in the sense of Rubin (1976). We decompose $Y = (Y_{obs}, Y_{mis})$, where $Y_{obs}$ and $Y_{mis}$ represent the observed and missing parts of $Y$, respectively. Thus, the missing-at-random assumption is described as

$$\text{pr}\{R = (r_1, r_2, \cdots, r_p) \mid Y_{obs}, Y_{mis}\} = \text{pr}\{R = (r_1, r_2, \cdots, r_p) \mid Y_{obs}\}, \tag{2}$$

for any $r_j \in \{0, 1\}$, $j = 1, 2, \cdots, p$.

Under the missing-at-random assumption, a consistent estimator of $\theta$ can be obtained by solving the following estimating equation:

$$\frac{1}{n} \sum_{i=1}^{n} E\{U(\theta; Y_i) \mid Y_{i,obs}\} = 0, \tag{3}$$

4

where it is understood that $E\left\{U(\theta; Y_i) \mid Y_{i,obs}\right\} = U(\theta; Y_i)$ if $Y_{i,obs} = Y_i$. To compute the conditional expectation in (3), the parametric fractional imputation (PFI) method of Kim (2011) can be developed. To apply the parametric fractional imputation, we can assume that the random vector $Y$ follows a parametric model in that $F_0(Y) \in \{F_\zeta(Y) : \zeta \in \Omega\}$. In the parametric fractional imputation, $M$ imputed values for $Y_{i,mis}$, say $\{Y_{i,mis}^{*(1)}, Y_{i,mis}^{*(2)}, \cdots, Y_{i,mis}^{*(M)}\}$ are generated from a proposal distribution with the same support of $F_0(Y)$ and are assigned with fractional weights, say $\{w_{i1}^*, w_{i2}^*, \cdots, w_{iM}^*\}$, so that a consistent estimator of $\theta$ can be obtained by solving

$$\sum_{i=1}^{n} \sum_{j=1}^{M} w_{ij}^* U(\theta; Y_{i,obs}, Y_{i,mis}^{*(j)}) = 0,$$

where the fractional weights are constructed to satisfy $\sum_{j=1}^{M} w_{ij}^* U(\theta; Y_{i,obs}, Y_{i,mis}^{*(j)}) \cong E\left\{U(\theta; Y_i) \mid Y_{i,obs}\right\}$ as closely as possible, with $\sum_{j=1}^{M} w_{ij}^* = 1$. In Kim (2011), importance sampling idea is used to compute the fractional weights.

However, for multivariate missing data, it is not easy to find a joint distribution family $\{F_\zeta(Y) : \zeta \in \Omega\}$ correctly. If the joint distribution family $\{F_\zeta(Y) : \zeta \in \Omega\}$ is misspecified, the parametric fractional imputation can lead to biased inference. All aforementioned concerns motivate us to consider a more robust fractional imputation method using Gaussian mixture models, which cover a wider class of parametric models.

# 3   Proposed method

We assume that the random vector $Y$ follows a Gaussian mixture model

$$f(Y; \alpha, \zeta) = \sum_{g=1}^{G} \alpha_g f(Y; \zeta_g), \tag{4}$$

where $G$ is the number of mixture component, $\alpha_g \in (0, 1)$ is the mixture proportion satisfying $\sum_{g=1}^{G} \alpha_g = 1$, and $f(\cdot; \zeta_g)$ is the density function of multivariate normal distribution with parameter $\zeta_g = \{\mu_g, \Sigma_g\}$. Here, we consider the same values of $\Sigma_g = \Sigma$ across all components to get a parsimonious model. The proposed joint model in (4) can be easily extended to use the group-dependent variance $\Sigma_g$, as in Di Zio et al. (2007).

5

To formulate the proposal, define the group indicator vector $Z = (z_1, z_2, \cdots, z_G)$, where $z_g = 1$ and $z_j = 0$ for all $j \neq g$, if sample unit belongs to the $g$-th group. Note that $Z$ is a latent variable with parameter $\mathrm{pr}(z_g = 1) = \alpha_g$, satisfying $\sum_{g=1}^{G} \alpha_g = 1$. We assume that the Gaussian mixture model in (4) satisfies the strong first-order identifiability assumption (Chen, 1995; Liu and Shao, 2003; Chen and Khalili, 2008), where the first-order derivatives of $f(Y; \alpha, \zeta)$ respect to all parameters are linearly independent. Using $Z$ variable, we can express $f(Y) = \sum_{g=1}^{G} \mathrm{pr}(z_g = 1) f(Y \mid z_g = 1)$, which leads to the marginal distribution in (4).

To handle item nonresponse, we propose to use the fractional imputation method to impute the missing values. Note that, the joint predictive distribution of $(Y_{mis}, Z)$ given $Y_{obs}$ can be written as $f(Y_{mis}, Z \mid Y_{obs}) = f(Z \mid Y_{obs}) f(Y_{mis} \mid Y_{obs}, Z)$, which implies that the prediction model for $Y_{mis}$ is

$$f(Y_{mis} \mid Y_{obs}) = \sum_{g=1}^{G} \mathrm{pr}(z_g = 1 \mid Y_{obs}) f(Y_{mis} \mid Y_{obs}, z_g = 1). \tag{5}$$

The first part in (5) can be obtained by

$$\mathrm{pr}(z_g = 1 \mid Y_{obs}) = \frac{f(Y_{obs} \mid z_g = 1) \alpha_g}{\sum_{g=1}^{G} f(Y_{obs} \mid z_g = 1) \alpha_g},$$

where $Y_{obs} \mid (z_g = 1)$ is normal. The second part of (5), which is $Y_{mis} \mid (Y_{obs}, z_g = 1)$, is also normal. Therefore, the EM algorithm for the proposed fractional imputation using Gaussian mixture models (FIGMM) can be described as follows:

*I-step*: To generate $Y_{i,mis}^*$ from $f(Y_{i,mis} \mid Y_{i,obs}; \alpha^{(t)}, \zeta^{(t)})$ in (5), we use the following two-step method:

*Step 1*: Compute

$$p_{ig}^{(t)} = \frac{f(Y_{i,obs} \mid z_{ig} = 1; \zeta_g^{(t)}) \alpha_g^{(t)}}{\sum_{g=1}^{G} f(Y_{i,obs} \mid z_{ig} = 1; \zeta_g^{(t)}) \alpha_g^{(t)}},$$

where $f(Y_{i,obs} \mid z_{ig} = 1; \zeta_g)$ is the marginal density of $Y_{i,obs}$ derived from $(Y_{i,obs}, Y_{i,mis}) \mid (z_{ig} = 1) \sim N(\mu_g, \Sigma)$. Generate $(M_{i1}^{(t)}, M_{i2}^{(t)}, \cdots, M_{iG}^{(t)}) \sim$ Multinomial$(M; \mathrm{P}_i^{(t)})$, where $\mathrm{P}_i^{(t)} = (p_{i1}^{(t)}, \cdots, p_{iG}^{(t)})$ with $M_{ij}^{(t)} > 0$.

*Step 2*: For each $g = 1, 2, \cdots, G$, we generate $M_{ig}^{(t)}$ independent realizations of $Y_{i,mis}^*$, say $\{Y_{i,mis}^{*(gj)}, j = 1, 2, \cdots, M_{ig}^{(t)}\}$, from the conditional distribution $f(Y_{i,mis} \mid Y_{i,obs}, z_{ig} = 1; \zeta_g^{(t)})$, which is also normal.

*W-step*: Compute the fractional weights for $Y_{i,mis}^{*(gj)}$ as $w_{igj(t)}^* = p_{ig}^{(t)}/M_{ig}^{(t)}$. Note that $\sum_{g=1}^{G} \sum_{j=1}^{M_{ig}^{(t)}} w_{igj(t)}^* = 1$, for each $i = 1, 2, \cdots, n$. Using $(w_{igj(t)}^*, Y_{i,mis}^{*(gj)})$, we can compute

$$Q^*(\alpha, \zeta \mid \alpha^{(t)}, \zeta^{(t)}) = \sum_{i=1}^{n} \sum_{g=1}^{G} \sum_{j=1}^{M_{ig}^{(t)}} w_{igj(t)}^* \left\{ \log \alpha_g + \log f(Y_i^{*(gj)} \mid z_{ig} = 1; \zeta_g) \right\}, \quad (6)$$

where $Y_i^{*(gj)} = (Y_{i,obs}, Y_{i,mis}^{*(gj)})$. If $\delta_i = 1$, then $Y_i^{*(gj)} = Y_i$.

*M-step*: Update the parameters by maximizing (6) with respect to $(\alpha, \zeta)$. The solution is

$$\alpha_g^{(t+1)} = \frac{1}{n} \sum_{i=1}^{n} \frac{M_{ig}^{(t)}}{M}; \qquad \mu_g^{(t+1)} = \frac{\sum_{i=1}^{n} \sum_{j=1}^{M_{ig}^{(t)}} w_{igj(t)}^* Y_i^{*(gj)}}{\sum_{i=1}^{n} \sum_{j=1}^{M_{ig}^{(t)}} w_{igj(t)}^*};$$

$$\Sigma^{(t+1)} = \frac{1}{n} \sum_{i=1}^{n} \sum_{g=1}^{G} \sum_{j=1}^{M_{ig}^{(t)}} w_{igj(t)}^* \left( Y_i^{*(gj)} - \mu_g^{(t+1)} \right) \left( Y_i^{*(gj)} - \mu_g^{(t+1)} \right)^{\mathrm{T}}.$$

Repeat *I-step* to *M-step* until the convergence is achieved. Then, the final estimator, say $\hat{\theta}_{FIGMM}$, of $\theta$ can be obtained by solving the fractionally imputed estimating equation, given by

$$\frac{1}{n} \sum_{i=1}^{n} \sum_{g=1}^{G} \sum_{j=1}^{M_{ig}} w_{igj}^* U(\theta; Y_i^{*(gj)}) = 0, \quad (7)$$

where $w_{igj}^*$ are the final fractional weights and $M_{ig}$ are the final imputation sizes for group $g$, after convergence of the EM algorithm.

**Remark 1** *We now briefly discuss variance estimation of* $\hat{\theta}_{FIGMM}$. *To estimate the variance of* $\hat{\theta}_{FIGMM}$, *replication methods, such as jackknife, can be used. First note that, the fractional weight assigned to* $Y_i^{*(gj)}$ *is* $w_{igj}^* = \hat{p}_{ig} M_{ig}^{-1} := \hat{p}_{ig} \hat{\pi}_{2j|ig}$, *where* $\hat{p}_{ig}$ *is obtained from*

$$\hat{p}_{ig} = \frac{f(Y_{i,obs} \mid z_{ig} = 1; \hat{\zeta}_g) \hat{\alpha}_g}{\sum_{g=1}^{G} f(Y_{i,obs} \mid z_{ig} = 1; \hat{\zeta}_g) \hat{\alpha}_g}. \quad (8)$$

*Thus, the $k$-th replicate of $w^*_{igj}$ can be obtained by*

$$w^{*(k)}_{igj} = \hat{p}^{(k)}_{ig}\hat{\pi}^{(k)}_{2j|ig}, \tag{9}$$

*where $\hat{p}^{(k)}_{ig}$ is obtained from (8) using $\hat{\zeta}^{(k)}$ and $\alpha^{(k)}_g$, the $k$-th replicate of $\hat{\zeta}$ and $\hat{\alpha}_g$ respectively, and*

$$\hat{\pi}^{(k)}_{2j|ig} \propto \frac{f(Y^{*(gj)}_{i,mis} \mid Y_{i,obs}, z_{ig} = 1; \hat{\zeta}^{(k)}_g)}{f(Y^{*(gj)}_{i,mis} \mid Y_{i,obs}, z_{ig} = 1; \hat{\zeta}_g)}$$

*and $\sum_{j=1}^{M_{ig}} \hat{\pi}^{(k)}_{2j|ig} = 1$. The calculation of $\hat{\pi}^{(k)}_{2j|ig}$ is based on the idea of importance sampling. Construction of replicate fractional weights using importance sampling idea has been used in Berg et al. (2016).*

*The replicate parameter estimates $(\hat{\alpha}^{(k)}, \hat{\zeta}^{(k)})$ are computed by maximizing*

$$l^{(k)}_{obs}(\alpha, \zeta) = \sum_{i=1}^{n} w^{(k)}_i \log f_{obs}(Y_{i,obs}; \alpha, \zeta) \tag{10}$$

*respect to $(\alpha, \zeta)$, where $f_{obs}(Y_{i,obs}; \alpha, \zeta) = \sum_{g=1}^{G} \alpha_g f(Y_{i,obs} \mid z_{ig} = 1; \zeta_g)$, and $w^{(k)}_i$ is the $k$-th replicate of $w_i = n^{-1}$. The maximizer of $l^{(k)}_{obs}(\alpha, \zeta)$ in (10) can be obtained by applying the same EM algorithm using replicate weights and replicate fractional weights in the W-step and M-step. There is no need to repeat I-step. Variance estimation for $\hat{\theta}_{FIGMM}$ can be obtained by computing the $k$-th replicate of $\hat{\theta}_{FIGMM}$ from*

$$\sum_{i=1}^{n} w^{(k)}_i \sum_{g=1}^{G} \sum_{j=1}^{M_{ig}} w^{*(k)}_{igj} U(\theta; Y^{*(gj)}_i) = 0. \tag{11}$$

*For example, the jackknife variance estimator of $\hat{\theta}_{FIGMM}$ can be obtained by*

$$\widehat{\mathrm{var}}\left(\hat{\theta}_{FIGMM}\right) = \frac{n-1}{n} \sum_{i=1}^{n} \left(\hat{\theta}^{(k)}_{FIGMM} - \hat{\theta}_{FIGMM}\right)^2,$$

*where $\hat{\theta}^{(k)}_{FIGMM}$ is the $k$-th replicate of $\hat{\theta}_{FIGMM}$ obtained from (11).*

## 4   Asymptotic theory

In our proposed fractional imputation method using Gaussian mixture models in §3, we have assumed that the size of mixture components, $G$, is known. In practice , $G$ is

often unknown and we need to estimate it from the sample data. If $G$ is larger than necessary, the proposed mixture model may be subject to overfitting and increase its variance. If $G$ is small, then the approximation of the true distribution cannot provide accurate prediction due to its bias. Hence, we can allow the model complexity parameter $G$ to depend on the sample size $n$, say $G = G(n)$. The choice of $G$ under complete data has been well explored in the literature. The popular methods are based on Bayesian information criterion (BIC) and Akaike's information criterion (AIC). See Wallace and Dowe (1999), Windham and Cutler (1992), Schwarz et al. (1978), Fraley and Raftery (1998), Keribin (2000) and Dasgupta and Raftery (1998). The alternative way of using SCAD penalty (Fan and Li, 2001) is studied in Chen and Khalili (2008) and Huang et al. (2017).

In this paper, we consider using the Bayesian information criterion to select $G$. Under multivariate missingness, we do not have the complete log-likelihood function. Thus, we use the observed log-likelihood function to serve the role of the complete log-likelihood function in computing the information criterion, in the sense that

$$\text{BIC}(G) = -2 \sum_{i=1}^{n} \log \left\{ \sum_{g=1}^{G} \hat{\alpha}_g f(Y_{i,obs} \mid z_{ig} = 1; \hat{\zeta}_g) \right\} + (\log n)\phi(G), \qquad (12)$$

under the assumption of $\Sigma_g = \Sigma$, where $(\hat{\alpha}, \hat{\zeta})$ are the estimators obtained from the proposed method and $\phi(G)$ is a monotone increasing function of $G$. In (12), $\phi(G) = G + Gp$ if ignoring constant terms. However, our model selection framework and theoretical results can be directly applied to any general penalty function $\phi(G)$. Using Gaussian mixture models, the observed log-likelihood function is expressed as a closed-form.

In this section, we first establish the consistency of model selection using (12) under the Gaussian mixture model assumption. After that, we establish some asymptotic results when the Gaussian mixture model assumption is violated.

To establish the first part, assume that $\{Y_1, Y_2, \cdots, Y_n\}$ is a random sample from $f_0(Y) = \sum_{g=1}^{G^o} \alpha_g^o f(Y; \zeta_g^o)$, where $(G^o, \alpha^o, \zeta^o)$ are true parameter values. For $\zeta_g^o = (\mu_g^o, \Sigma^o)$, we need the following regularity assumptions:

($A1$) The mean vectors for each mixture component is bounded uniformly, in

9

the sense of $\|\mu_g^o\| \le C_1$, for $g = 1, 2, \cdots, G^o$.

($A2$) $\|\Sigma^o\| \le C_2$. Furthermore, $\Sigma^o$ is nonsingular.

The first assumption means the first moment is bounded. Assumption ($A2$) is to make sure that $\Sigma^0$ is bounded and nonsingular. Both assumptions are commonly used.

To establish the model consistency, we furthermore make the additional assumptions on the response mechanism:

($A3$) The response rate for $y_j$ is bounded below from 0, say $\lim_n n^{-1} \sum_{i=1}^n r_{ij} > C_3$ , for $j = 1, 2, \cdots, p$, where $C_3 > 0$ is a constant.

($A4$) The response mechanism satisfies the mising-at-random condition in (2).

The following theorem shows that the true number of mixture components can be selected by minimizing $\mathrm{BIC}(G)$ in (12) consistently.

**Theorem 4.1** *Assume the true density $f_0$ is the Gaussian mixture model, satisfying (A1)–(A2). Let $\hat{G}$ be the minimizer of $\mathrm{BIC}(G)$ in (12). Under assumptions (A3)–(A4), we have*

$$\mathrm{pr}(\hat{G} = G^o) \to 1,$$

*as $n \to \infty$, where $G^o$ is the true number of mixture components.*

The proof of Theorem 4.1 is shown in the Supplementary Material. Theorem 4.1 states that minimizing $\mathrm{BIC}(G)$ consistently selects the true mixture components under the assumption that the true distribution is in the Gaussian mixture model.

Now, in the second scenario, the true distribution does not necessary belong to the class of Gaussian mixture models, Thus, we first establish the following lemma to measure how well Gaussian mixture model can approximate the arbitrary density function. We furthermore make additional assumptions about the true density function $f_0$. Use $E_0$ to denote the expectation respect to $f_0$.

($A5$) Assume $f_0(Y)$ is continuous with $\int Y^2 f_0(Y) < \infty$.

$(A6)$ Assume $E_0 \{\partial f(Y)/\partial \alpha\} < \infty$ and $E_0 \{\partial f(Y)/\partial \mu\} < \infty$, where $f(Y) = \sum_{g=1}^{G} \alpha_g f(Y; \mu_g, \Sigma)$. Moreover, assume $E_0 \{f(Y)^{-2}\} < \infty$.

Assumption $(A5)$ is satisfied for any continuous random variable with bounded second moments. Assumption $(A6)$ is true for any finite Gaussian mixture model and $f_0$ has a valid moment generating function.

**Lemma 4.2** *Under assumptions (A5)–(A6) and missing at random, for any $\epsilon > 0$, there exist $G = \epsilon^{-\gamma}$, such that*

$$\|f_0 - \hat{f}\|_1 = O(\epsilon), \tag{13}$$

$$\mathrm{var}(f_0 - \hat{f}) = O(\epsilon^{-\gamma} n^{-1}), \tag{14}$$

*with probability one, where $\hat{f}(Y) = \sum_{g=1}^{G} \hat{\alpha}_g f(Y; \hat{\mu}_g, \hat{\Sigma})$ is obtained from the proposed method in §3, $\gamma > 0$ and $\|f_0 - \hat{f}\|_1 = \int |f_0(Y) - \hat{f}(Y)| f_0(Y) \mathrm{d}Y$.*

The proof of Lemma 4.2 is presented in the Supplementary Material. If $f_0$ is a density function of the Gaussian mixture model, then $\gamma = 0$ and by Theorem 4.1, our proposed BIC$(G)$ can select the true model consistently. For any $f_0$ satisfies $(A5)$–$(A6)$ and is not a finite Gaussian mixture model, the bias can goes to 0 as $G \to \infty$ from (13). The variance will increase as $G \to \infty$ from (14) for fixed $n$. There is a trade-off between bias and variance for the divergence case $(\gamma > 0, G \to \infty)$.

Using Lemma 4.2, we can further establish the $\sqrt{n}$-consistency of $\hat{\theta}_{FIGMM}$. The following assumptions are the sufficient conditions to obtain the $\sqrt{n}$-consistency.

$(A7)$ $E_0 \{U^2(\theta; Y_i)\} < \infty$.

$(A8)$ $\gamma \in (0, 2)$.

$(A9)$ $\epsilon = O(n^{-1/(2-\Delta)})$, for any $\Delta \in (0, 2)$.

**Theorem 4.3** *Under assumptions (A5)–(A9), $\gamma + \Delta < 2$ and MAR, we have*

$$\frac{1}{n} \sum_{i=1}^{n} \sum_{g=1}^{G} \sum_{j=1}^{M_{ig}} w_{igj}^* U(\theta; Y_i^{*(gj)}) \cong J_1 + o_p(n^{-1/2}), \tag{15}$$

where $J_1 = n^{-1} \sum_{i=1}^{n} E_0 \{U(\theta; Y_i) \mid Y_{i,obs}\}$, if $M = \sum_{i,g} \{M_{ig}\} \to \infty$. Furthermore, we have

$$\sqrt{n}(\hat{\theta}_{FIGMM} - \theta_0) \to N(0, \Sigma), \tag{16}$$

for some $\Sigma$ which is positive definite and $\theta_0$ satisfies $E_0 \{U(\theta_0; Y)\} = 0$.

The proof of (15) is shown in the Supplementary Material and (16) can be directly derived from (15). From Theorem 4.3, we have $G = O(n^{\gamma/(2-\Delta)}) = o(n) \to \infty$ with the rate smaller than $n$. Thus, even under non-Gaussian mixture families, our proposed method still enjoys $\sqrt{n}$-consistency.

## 5 Extension

In Section 3, we assume that $Y$ is fully continuous. However, in practice, categorical variables can be used to build imputation models. We extend our proposed method to incorporate the categorical variable as a covariate in the model.

To introduce the proposed method, we first introduce the conditional Gaussian mixture model. Suppose that $(X, Y)$ is a random vector where $X$ is discrete and $Y$ is continuous. We further assume that $X$ is always observed. To obtain the conditional Gaussian mixture model, we assume that $Z$ satisfies

$$f(Y \mid X, Z) = f(Y \mid Z), \tag{17}$$

in the sense that $Z$ is a partition of the sample such that $Y$ is homogeneous within each group defined by $Z$. Furthermore, we assume that $f(Y \mid z_g = 1)$ follows a Gaussian distribution. Combining these assumptions, we have the following conditional Gaussian mixture model

$$f(Y \mid X) = \sum_{g=1}^{G} \alpha_g(X) f(Y \mid z_g = 1), \tag{18}$$

where $\alpha_g(X) = \text{pr}(z_g = 1 \mid X)$ and $f(Y \mid z_g = 1)$ is the density function of the normal distribution with parameter $\zeta_g = \{\mu_g, \Sigma\}$. We also assume that the identifiability conditions in (18) hold.

12

Using the argument similar to (5), the predictive model of $Y_{i,mis}$ under (17) can be expressed as

$$f(Y_{i,mis} \mid Y_{i,obs}, X_i) = \sum_{g=1}^{G} \mathrm{pr}(z_{ig} = 1 \mid Y_{i,obs}, X_i) f(Y_{i,mis} \mid Y_{i,obs}, z_{ig} = 1), \qquad (19)$$

where $f(Y_{i,mis} \mid Y_{i,obs}, z_{ig} = 1)$ can be derived from $(Y_{i,obs}, Y_{i,mis}) \mid (z_{ig} = 1) \sim N(\mu_g, \Sigma)$. The posterior probability of $z_{ig} = 1$ given the observed data is

$$\mathrm{pr}(z_{ig} = 1 \mid Y_{i,obs}, X_i) = \frac{f(Y_{i,obs} \mid z_{ig} = 1)\mathrm{pr}(z_{ig} = 1 \mid X_i)}{\sum_{g=1}^{G} f(Y_{i,obs} \mid z_{ig} = 1)\mathrm{pr}(z_{ig} = 1 \mid X_i)}.$$

Therefore, the proposed fractional imputation using conditional Gaussian mixture models can be summarized as follows:

*I-step*: Creating $M$ imputed values of $Y_{i,mis}$ from (19) can be described in the following two steps.

*Step 1*: For each $g = 1, 2, \cdots, G$, given the current parameter values $(\alpha_g^{(t)}, \zeta_g^{(t)})$, the posterior probabilities of $z_{ig} = 1$ given $(Y_{i,obs}, X_i)$ can be obtained from

$$p_{ig}^{(t)} = \frac{f(Y_{i,obs} \mid z_{ig} = 1; \zeta_g^{(t)})\alpha_g^{(t)}(X_i)}{\sum_{g=1}^{G} f(Y_{i,obs} \mid z_{ig} = 1; \zeta_g^{(t)})\alpha_g^{(t)}(X_i)}.$$

*Step 2*: Generate $M$ imputed values of $Y_{i,mis}$ following the same procedure of *I-step* in §3.

*W-step*: Update the fractional weights for $Y_i^{*(gj)} = (Y_{i,obs}, Y_{i,mis}^{*(gj)})$ as $w_{igj(t)}^* = p_{ig}^{(t)}/M_{ig}^{(t)}$, for $j = 1, 2, \cdots M_{ig}^{(t)}$ and $\sum_{g=1}^{G} M_{ig}^{(t)} = M$. Note that $\sum_{g=1}^{G} \sum_{j=1}^{M_{ig}^{(t)}} w_{igj(t)}^* = 1$.

*M-step*: Update the parameter values by maximizing

$$Q^*(\alpha, \zeta \mid \alpha^{(t)}, \zeta^{(t)}) = \sum_{i=1}^{n} \sum_{g=1}^{G} \sum_{j=1}^{M_{ig}^{(t)}} w_{igj(t)}^* \left\{ \log \alpha_g(X_i) + \log f(Y_i^{*(gj)} \mid z_{ig} = 1; \zeta_g) \right\},$$

respect to $(\alpha, \zeta)$.

Repeat *I-step* to *M-step* iteratively until convergence is achieved. The final estimator of $\theta$ can be obtained by solving the fractionally imputed estimating equation in (7). Note that the proposed method builds the proportion vector of mixture components into a function of auxiliary variable and assumes that the mixture components share the same mean and variance structure. Thus, the proposed method can borrow information across different $X$ values. Moreover, the auxiliary information is incorporated to build a more flexible class of joint distributions.

# 6 Numerical Studies

We consider two simulation studies to evaluate the performance of the proposed methods. The first simulation study is used to check the performance of the proposed imputation method using Gaussian mixture models under multivariate continuous variables. The second simulation study considers the case of multivariate mixed categorical and continuous variables. To save space, we only present the first simulation study. The second simulation study is presented in the Supplementary Material.

In the first simulation study, we consider the following models for generating $Y_i = (Y_{i1}, Y_{i2}, Y_{i3})$.

1. *M1*: A mixture distribution with density $f(Y) = \sum_{g=1}^{3} \alpha_g f_g(Y)$, where $(\alpha_1, \alpha_2, \alpha_3) = (0\cdot3, 0\cdot3, 0\cdot4)$ and $f_g(Y)$ is a density function for multivariate normal distribution with mean $\mu_g$ and variance

$$\Sigma(\rho) = \begin{pmatrix} 1 & \rho & \rho^2 \\ \rho & 1 & \rho \\ \rho^2 & \rho & 1 \end{pmatrix}.$$

   Let $\rho = 0\cdot7$ and $\mu_1 = (-3, -3, -3, ), \mu_2 = (1, 1, 1), \mu_3 = (5, 5, 5)$.

2. *M2*: Use the same model as *M2* except for $f_1(Y)$, where $f_1(Y)$ is a product of the density for the exponential distribution with rate parameter 1.

3. *M3*: $Y_{i1} = 1 + e_{i1}, Y_{i2} = 0\cdot5Y_{i1} + e_{i2}$ and $Y_{i3} = Y_{i2} + e_{i3}$, where $e_{i1}, e_{i2}, e_{i3}$ are independently generated from $N(0, 1)$, Gamma$(1, 1)$ and $\chi_1^2$ distributions, respectively.

4. *M4*: Generate $(Y_{i1}, Y_{i2})$ independently from a Gaussian distribution with mean $(1, 2)$ and variance

$$\begin{pmatrix} 1 & 0{\cdot}5 \\ 0{\cdot}5 & 1 \end{pmatrix}.$$

   Let $Y_{i3} = Y_{i2}^2 + e_{i3}$, where $e_{i3} \sim N(0, 1)$.

In *M1*, a Gaussian mixture model with $G = 3$ is used to generate the samples. A non-Gaussian mixture distribution is used in *M2* to check the robustness of the imputation methods. *M3* and *M4* are used to check the performance of the imputation methods under skewness and nonlinearity, respectively.

The size for each realized sample is $n = 500$. Once the complete sample is obtained, for $y_{ij}, j = 2, 3$, we select 25% of the sample independently to make missingness with the selection probabilities equal to $\pi_{ij}$, where $\text{logit}(\pi_{i2}) = -0{\cdot}8 + 0{\cdot}4y_{i1}$, $\text{logit}(\pi_{i3}) = 0{\cdot}4 - 0{\cdot}8y_{i1}$, and $\text{logit}(u) = \exp(u)/\{1 + \exp(u)\}$. Since we assume $y_{i1}$ are fully observed, the response mechanism is missing at random.

The overall missing rate is approximate 55%. For each realized incomplete samples, we apply the following methods:

[*Full*]: As a benchmark, we use the full samples to estimate parameters. Confidence intervals with 95% coverage rates are constructed using Wald method.

[*CC*]: Only use the complete cases to estimate parameters and construct confidence intervals.

[*MICE*]: Apply multivariate imputation by chained equations (Buuren and Groothuis-Oudshoorn, 2011). The predictive mean matching is used as a default. The variance estimators are obtained using Rubin's formula and confidence intervals are constructed using Wald method.

[*PFI*]: Parametric fractional imputation method of Kim (2011), where we assume the joint distribution is a multivariate normal distribution.

[*SFI*]: The proposed semiparametric fractional imputation method using Gaussian mixture models, where the number of components $G$ is selected using the
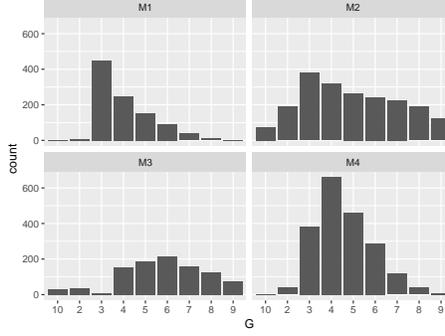
Figure 1: The histograms of selected $G$ values using the proposed BIC.

BIC in (12). The confidence interval is implemented using the variance estimator presented in Remark 1.

The parameters of interest are the population means and the population proportions. For $Y = (y_1, y_2, y_3)$, define $\theta_2 = E(y_2), \theta_3 = E(y_3)$. The parameters of population proportions are defined as $P_2 = \mathrm{pr}(y_2 < c_2), P_3 = \mathrm{pr}(y_3 < c_3)$, where $(c_2, c_3) = (-2, -2)$ for $M1$ and $M2$, $(c_2, c_3) = (2, 3)$ for $M3$, $(c_2, c_3) = (2, 5)$ for $M4$. The simulation is repeated for $B = 2,000$ times.

To evaluate the above methods, the relative mean square error (RMSE) is defined as

$$\text{RMSE} = \frac{\text{MSE}_{\text{method}}}{\text{MSE}_{\text{Full}}} \times 100,$$

where $\text{MSE}_{\text{method}}$ is the mean squared error of the parameters of the current method and $\text{MSE}_{\text{Full}}$ is the mean squared error of the parameters of using full samples. The simulation results of RMSE and average coverage probability are presented in Table 1. The histograms of selected $G$ values using the proposed BIC method are shown in Figure 1.

Table 1 presents the relative mean squared errors and the corresponding average coverage probabilities from the above simulation study of size $B = 2,000$. Under $M1$, the joint model is a Gaussian mixture. The proposed method obtains almost the same RMSEs for estimating $\theta_2, \theta_3$ and $P_2$ with MICE and PFI, but outperforms MICE and PFI for estimating $P_3$. Moreover, coverage probabilities of MICE and PFI are less than 95% in estimating $\theta_3$ and $P_3$. SFI uniformly achieves approximate 95%

16

Table 1: The relative mean squared errors and coverage probabilities (in parentheses) of the five methods computed from a Monte Carlo simulation of size $B = 2,000$. Full, the method uses full samples; CC, the method only uses complete cases; MICE, multiple imputation uses chain equations; PFI, parametric fractional imputation; SFI, the proposed semiparametric fractional imputation method uses Gaussian mixture models.

| Model | Method | $\theta_2$ | $\theta_3$ | $P_2$ | $P_3$ |
|-------|--------|-----------|-----------|-----------|-----------|
| *M1* | Full | 100(94·9) | 100(95·0) | 100(94·5) | 100(94·9) |
| | CC | 3233(2·9) | 3010(3·5) | 2639(4·5) | 2407(5·6) |
| | MICE | 100(95·2) | 105(93·9) | 102(94·7) | 145(90·6) |
| | PFI | 100(94·6) | 104(93·3) | 102(94·0) | 141(89·1) |
| | SFI | 100(94·6) | 106(95·0) | 102(95·3) | 134(95·3) |
| *M2* | Full | 100(94·1) | 100(94·6) | 100(94·1) | 100(94·5) |
| | CC | 2841(3·0) | 2863(3·6) | 2284(4·8) | 2233(5·7) |
| | MICE | 102(94·4) | 109(93·7) | 103(93·8) | 189(86·3) |
| | PFI | 102(93·8) | 108(94·1) | 102(93·3) | 180(83·8) |
| | SFI | 106(95·1) | 107(95·4) | 103(93·8) | 177(92·8) |
| *M3* | Full | 100(94·7) | 100(94·7) | 100(95·2) | 100(95·1) |
| | CC | 157(77·9) | 155(85·6) | 145(77·0) | 145(86·6) |
| | MICE | 100(94·8) | 100(93·5) | 117(92·5) | 128(90·4) |
| | PFI | 100(94·6) | 100(95·2) | 117(89·9) | 127(89·7) |
| | SFI | 97(95·1) | 83(92·8) | 106(93·8) | 91(94·3) |
| *M4* | Full | 100(94·9) | 100(95·2) | 100(94·3) | 100(95·2) |
| | CC | 386(81·8) | 353(86·2) | 317(86·4) | 308(86·6) |
| | MICE | 110(94·8) | 114(95·2) | 128(93·5) | 207(86·7) |
| | PFI | 108(94·8) | 111(94·2) | 129(91·6) | 197(84·4) |
| | SFI | 126(95·0) | 124(94·7) | 135(94·5) | 139(94·2) |

coverage probabilities. Thus, we can conclude that MICE and PFI are biased, due to model misspecification. The histogram in Figure 1 shows that most of selected G values are 3, which is the true number of mixture components.

In *M2*, instead of Gaussian mixture models, one component is the exponential distribution. Table 1 shows that SFI outperforms MICE and PFI in term of RMSEs for estimating proportions and obtains similar performance for estimating means. The coverage rates for MICE and PFI for $P_3$ are poor.

The joint distribution *M3* is a skewed distribution. From Tables 1, we can see that SFI outperforms MICE and PFI uniformly. Furthermore, SFI provides better coverage probabilities than MICE and PFI for $P_2$ and $P_3$.

The joint distribution in *M4* has a nonlinear mean structure of $Y_{i3}$. Under M4, SFI obtains the much smaller RMSE than MICE and PFI in estimating $P_4$, but larger RMSE in estimating $\theta_2, \theta_3, P_3$. However, SFI achieves consistent confidence intervals with approximate 95% coverage probabilities. MICE and PFI are biased in interval estimation and coverage probabilities are much less than 95%. Overall, the performance of SFI is much better than MICE or PFI in terms of coverage probabilities.

Interestingly, the imputed estimators are sometimes more efficient than the full sample estimators. This phenomenon, called superefficiency (Meng, 1994), can happen when the method-of-moment is used in the full sample estimator. Yang and Kim (2016) give a rigorous theoretical justification for this phenomenon.

# 7 Application

In this section, we apply the proposed method in §3 to a synthetic data that mimics monthly retail trade survey data at U.S. Census Bureau. The synthetic monthly retail trade survey data was made for contest in a conference and can be found in http://www.portal-stat.admin.ch/ices5/imputation-contest/. The sampling scheme is a stratified simple random sample without replacement sample with six strata: one certain (take-all) and five non-certainty strata. The sample sizes are computed using Neyman allocation. An overview of the monthly retail trade survey data is presented in Figure 2.

The overall response rate is approximately 71%. Current month sales and inventories are subject to missingness. From Figure 2, we can find that this monthly retail trade data are highly skewed. From the normal quantile-quantile plot, normality assumption is violated and there exist three extreme outliers.

To impute current month sales and inventories, we applied the proposed fractional imputation method and MICE. After implementation, MICE failed to converge due to high correlations among the survey items. Therefore, we only present the final analysis results using the proposed method. The final results are shown in Table 2. Comparing with the true population statistics, provided by U.S. Census Bureau, we can see that our proposed fractional imputation method with Gaussian mixture
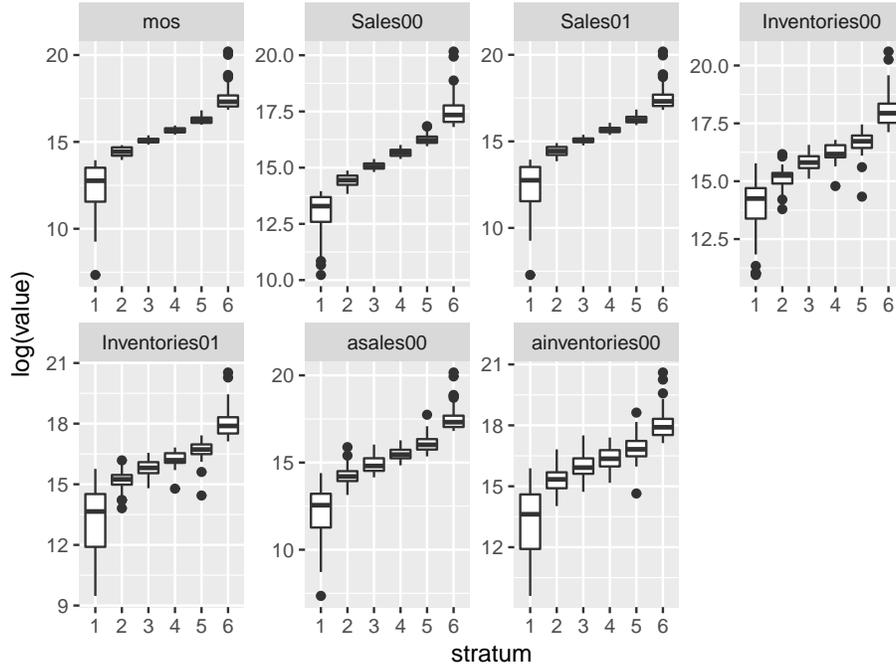
Figure 2: Overview for log-values of the monthly retail trade survey data:"mos" is frame measure of size; "Sales00" denotes current month sales for unit (subject to missing); "asales00" is current month administrative data value for sales; "Sales01" means prior month sales for unit; "Inventories00" is current month inventories for unit (Subject to missing); "ainventories00" is current month administrative data value for inventories; "Inventories01" is prior month inventories for unit.

Table 2: Imputation results for the monthly retail trade survey data. Parameter estimation, 95% confidence interval and true values are presented. Estimate, the estimators provided by the proposed method; Truth, the true parameter values from U.S. Census Bureau.

| Parameter | Estimate | 95% Confidence interval | Truth |
|---|---|---|---|
| Mean of Sales00 ($\times 10^{-6}$) | 2·28 | (2·10, 2·46) | 2·30 |
| Mean of Inventories00 ($\times 10^{-6}$) | 4·76 | (4·42, 5·10) | 4·81 |
| Correlation of Sales00 and Inventories00 | 0·97 | (0·94, 0·99) | 0·97 |

models works well to preserve the correlation structure and handle skewness and outliers. In Table 2, we can see that all 95% confidence intervals contain their true values.

# 8 Discussion

Fractional imputation has been proposed as a tool for frequentist imputation, as an alternative to multiple imputation. Multiple imputation using Rubin's formula can be biased when the model is uncongenial or the point estimator is not self-efficient (Meng, 1994; Yang and Kim, 2016). In this paper, we have proposed a semiparametric fractional imputation method using Gaussian mixture models to handle arbitrary multivariate missing data. The proposed method automatically selects the size of mixture components and provides a unified framework for robust imputation. Even if the group size $G$ increases with the sample size $n$, the resulting estimator enjoys $\sqrt{n}$-consistency. We have also extended the proposed method to incorporate categorical auxiliary variable. The flexible model assumption and efficient computation are the main advantages of our proposed method. An extension of the proposed method to survey data is a topic of future research. An R software package for the proposed method is under development.

# References

Bacharoglou, A. (2010). Approximation of probability distributions by convex mixtures of Gaussian measures. *Proceedings of the American Mathematical Society 138*(7), 2619–2628.

Bartlett, J. W., S. R. Seaman, I. R. White, J. R. Carpenter, and A. D. N. Initiative (2015). Multiple imputation of covariates by fully conditional specification: accommodating the substantive model. *Statistical Methods in Medical Research 24*(4), 462–487.

Berg, E., J.-K. Kim, and C. Skinner (2016). Imputation under informative sampling. *Journal of Survey Statistics and Methodology 4*(4), 436–462.

Buuren, S. and K. Groothuis-Oudshoorn (2011). MICE: Multivariate imputation by chained equations in R. *Journal of Statistical Software 45*(3), 1–67.

Chen, H. Y. (2010). Compatibility of conditionally specified models. *Statistics & Probability Letters 80*(7-8), 670–677.

Chen, J. (1995). Optimal rate of convergence for finite mixture models. *The Annals of Statistics*, 221–233.

Chen, J. and A. Khalili (2008). Order selection in finite mixture models with a nonsmooth penalty. *Journal of the American Statistical Association 103*(484), 1674–1683.

Cheng, P. E. (1994). Nonparametric estimation of mean functionals with data missing at random. *Journal of the American Statistical Association 89*(425), 81–87.

Dasgupta, A. and A. E. Raftery (1998). Detecting features in spatial point processes with clutter via model-based clustering. *Journal of the American Statistical Association. 93*(441), 294–302.

Di Zio, M., U. Guarnera, and O. Luzi (2007). Imputation through finite gaussian mixture models. *Computational Statistics & Data Analysis 51*(11), 5305–5316.

Elliott, M. R. and N. Stettler (2007). Using a mixture model for multiple imputation in the presence of outliers: the healthy for lifeproject. *Journal of the Royal Statistical Society: Series C (Applied Statistics) 56*(1), 63–78.

Fan, J. and R. Li (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association. 96*(456), 1348–1360.

Fraley, C. and A. E. Raftery (1998). How many clusters? which clustering method? answers via model-based cluster analysis. *The di2007imputationurnal 41*(8), 578–588.

Huang, T., H. Peng, and K. Zhang (2017). Model selection for Gaussian mixture models. *Statistica Sinica 27*(1), 147–169.

Judkins, D., T. Krenzke, A. Piesse, Z. Fan, and W.-C. Haung (2007). Preservation of skip patterns and covariate structure through semi-parametric whole question-naire imputation. In *Proceedings of the Section on Survey Research Methods of the American Statistical Association*, pp. 3211–3218.

Keribin, C. (2000). Consistent estimation of the order of mixture models. *Sankhya A*, 49–66.

Kim, H. J., J. P. Reiter, Q. Wang, L. H. Cox, and A. F. Karr (2014). Multiple imputation of missing or faulty values under linear constraints. *Journal of Business & Economic Statistics 32*(3), 375–386.

Kim, J. K. (2011). Parametric fractional imputation for missing data analysis. *Biometrika 98*(1), 119–132.

Kim, J. K. and W. Fuller (2004). Fractional hot deck imputation. *Biometrika 91*(3), 559–578.

Kim, J. K. and J. Shao (2013). *Statistical Methods for Handling Incomplete Data*. CRC Press.

Little, R. J. and D. B. Rubin (2002). *Statistical Analysis with Missing Data*. John Wiley & Sons.

Liu, J., A. Gelman, J. Hill, Y.-S. Su, and J. Kropko (2013). On the stationary distribution of iterative imputations. *Biometrika 101*(1), 155–173.

Liu, X. and Y. Shao (2003). Asymptotics for likelihood ratio tests under loss of identifiability. *The Annals of Statistics 31*(3), 807–832.

McLachlan, G. and D. Peel (2004). *Finite Mixture Models*. John Wiley & Sons.

Meng, X.-L. (1994). Multiple-imputation inferences with uncongenial sources of input. *Statistical Science 9*(4), 538–558.

Murray, J. S. and J. P. Reiter (2016). Multiple imputation of missing categorical and continuous values via Bayesian mixture models with local dependence. *Journal of the American Statistical Association 111*(516), 1466–1479.

Raghunathan, T. E., J. M. Lepkowski, J. Van Hoewyk, and P. Solenberger (2001). A multivariate technique for multiply imputing missing values using a sequence of regression models. *Survey Methodology 27*(1), 85–96.

Rubin, D. B. (1976). Inference and missing data. *Biometrika 63*(3), 581–592.

Rubin, D. B. (1996). Multiple imputation after 18+ years. *Journal of the American Statistical Association. 91*(434), 473–489.

Schwarz, G. et al. (1978). Estimating the dimension of a model. *The Annals of Statistics 6*(2), 461–464.

Tanner, M. A. and W. H. Wong (1987). The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association. 82*(398), 528–540.

Wallace, C. S. and D. L. Dowe (1999). Minimum message length and kolmogorov complexity. *The Computer Journal 42*(4), 270–283.

Wang, D. and S. X. Chen (2009). Empirical likelihood for estimating equations with missing values. *The Annals of Statistics 37*(1), 490–517.

Windham, M. P. and A. Cutler (1992). Information ratios for validating mixture analyses. *Journal of the American Statistical Association. 87*(420), 1188–1192.

Yang, S. and J. K. Kim (2016). A note on multiple imputation for method of moments estimation. *Biometrika 103*(1), 244–251.