Biometrika (????), **??**, ?, *pp*. 1–14

Printed in Great Britain

Parametric fractional imputation for missing data analysis

BY JAE KWANG KIM

Department of Statistics, Iowa State University, Ames, Iowa 50011, U.S.A.

jkim@iastate.edu

SUMMARY

Parametric fractional imputation is proposed as a general tool for missing data analysis. Using fractional weights, the observed likelihood can be approximated by the weighted mean of the imputed data likelihood. Computational efficiency can be achieved using the idea of importance sampling and calibration weighting. The proposed imputation method provides efficient parameter estimates for the model parameters specified in the imputation model and also provides reasonable estimates for parameters that are not part of the imputation model. Variance estimation is discussed and results from a limited simulation study are presented.

Some key words: EM algorithm, Importance sampling, Item nonresponse, Monte Carlo EM, Multiple imputation.

1. INTRODUCTION

Suppose that $y_1, \ldots, y_n$ are the observations for a probability sample selected from a finite population, where the finite population values are independent realisations of a random variable Y with a p-dimensional distribution $F_0(y) \in \{F_\theta(y); \theta \in \Omega\}$. Suppose that, under complete response, a parameter $\eta_g = E\{g(Y)\}$ is unbiasedly estimated by

$$\hat{\eta}_g = \sum_{i=1}^{n} w_i g(y_i) \tag{1}$$

for some function $g(y_i)$ with sampling weights w_i. Under simple random sampling, the sampling weight is $1/n$ and the sample can be regarded as a random sample from an infinite population with distribution $F_0(y)$.

Under nonresponse, one can replace (1) with

$$\hat{\eta}_{gR} \equiv \sum_{i=1}^{n} w_i E\left\{g(y_i) \mid y_{i,\text{obs}}\right\}, \tag{2}$$

where $y_{i,\text{obs}}$ and $y_{i,\text{mis}}$ denote the observed part and missing part of y_i, respectively. To simplify the presentation, we assume the sampling mechanism and the response mechanism are ignorable in the sense of Rubin (1976). To compute the conditional expectation in (2), we need a correct specification of the conditional distribution of $y_{i,\text{mis}}$ given $y_{i,\text{obs}}$. The conditional expectation in (2) depends on θ_0, where θ_0 is the true parameter value corresponding to F_0. That is, $E\left\{g(y_i) \mid y_{i,\text{obs}}\right\} = E\left\{g(y_i) \mid y_{i,\text{obs}}, \theta_0\right\}$.

To compute the conditional expectation in (2), a Monte Carlo approximation based on the imputed data can be used. Thus, one can interpret imputation as a Monte Carlo approximation of the conditional expectation given the observed data. Imputation is very attractive in practice because, once the imputed data are created, the data analyst does not need to know the conditional distribution in (2). Monte Carlo methods for approximating the conditional expectation in (2) can be placed in two classes. One is the Bayesian approach, where the imputed values are generated from the posterior predictive distribution of $y_{i,\text{mis}}$ given $y_{\text{obs}} = (y_{i,\text{obs}}; i = 1, \ldots, n)$:

$$f\left(y_{i,\text{mis}} \mid y_{\text{obs}}\right) = \int f\left(y_{i,\text{mis}} \mid \theta, y_{\text{obs}}\right) f\left(\theta \mid y_{\text{obs}}\right) d\theta. \tag{3}$$

This is essentially the approach used in multiple imputation as proposed by Rubin (1987). The other is the frequentist approach, where the imputed values are generated from the conditional distribution $f(y_{i,\text{mis}} \mid y_{\text{obs}}, \hat{\theta})$ and $\hat{\theta}$ is an estimated value for θ.

In the Bayesian approach to imputation, the convergence to a stable posterior predictive distribution (3) is difficult to check (Gelman et al., 1996). Also, the variance estimator used in multiple imputation is not consistent for some estimated parameters. For examples, see Wang & Robins (1998) and Kim et al. (2006).

The frequentist approach for imputation has received less attention than the Bayesian imputation. One notable exception is Wang & Robins (1998) who studied the asymptotic properties of multiple imputation and a parametric frequentist imputation procedure. They considered the estimated parameter $\hat{\theta}$ to be given, and did not discuss parameter estimation.

We consider frequentist imputation given a parametric model for the original distribution. Using the idea of importance sampling, we propose a frequentist imputation method that can be implemented with fractional imputation, discussed in Fay (1996) and Kim & Fuller (2004), where fractional imputation was presented as a nonparametric imputation method in the context of survey sampling and the parameters of interest are of descriptive nature. The proposed fractional imputation, called parametric fractional imputation, is also applicable in an analytic setting where interest lies in the model parameters of the superpopulation model. The parametric fractional imputation method can be modified to reduce Monte Carlo error and can be used to simplify the Monte Carlo implementation of the EM algorithm.

2. FRACTIONAL IMPUTATION

As discussed in § 1, we consider an approximation for the conditional expectation in (2) using fractional imputation. In fractional imputation, $M > 1$ imputed values for $y_{i,\text{mis}}$, say $y_{i,\text{mis}}^{*(1)}, \ldots, y_{i,\text{mis}}^{*(M)}$, are generated and assigned fractional weights, $w_{i1}^{*}, \ldots, w_{iM}^{*}$, so that

$$\sum_{j=1}^{M} w_{ij}^{*} g(y_{ij}^{*}) = E\{g\left(y_i\right) \mid y_{i,\text{obs}}, \hat{\theta}\}, \tag{4}$$

where $y_{ij}^{*} = (y_{i,\text{obs}}, y_{i,\text{mis}}^{*(j)})$, holds at least approximately for large M, where $\hat{\theta}$ is a consistent estimator of θ_0. A popular choice for $\hat{\theta}$ is the pseudo maximum likelihood estimator, where $\hat{\theta}$ is the θ that maximizes the pseudo log-likelihood function. That is,

$$\hat{\theta} = \arg\max_{\theta \in \Omega} \sum_{i=1}^{n} w_i \log\left\{f_{\text{obs}(i)}\left(y_{i,\text{obs}}; \theta\right)\right\}, \tag{5}$$

where $f_{\text{obs}(i)}\left(y_{i,\text{obs}}; \theta\right) = \int f\left(y_i; \theta\right) dy_{i,\text{mis}}$ is the marginal density of $y_{i,\text{obs}}$. A computationally simple method of finding the pseudo maximum likelihood estimator will be discussed in § 4.

Condition (4) applied to $g(y_i) = c$ implies that

$$\sum_{j=1}^{M} w_{ij}^* = 1 \tag{6}$$

for all i. Given fractionally imputed data satisfying (4) and (6), the parameter η_g can be estimated by

$$\hat{\eta}_{FI,g} = \sum_{i=1}^{n} \sum_{j=1}^{M} w_i w_{ij}^* g\left(y_{ij}^*\right). \tag{7}$$

The imputed estimator (7) is obtained by applying the formula (1) using y_{ij}^* as the observations with weights $w_i w_{ij}^*$. For a single parameter $\eta_g = E\left\{g\left(Y\right)\right\}$, any fractional imputation satisfying (4) provides a consistent estimator of η_g. For general purpose estimation, the g-function defining η_g is unknown at the time of imputation (Fay, 1992). To create fractional imputation for categorical data with a finite number of possible values for $y_{i,mis}$, we take the possible values as the imputed values and compute the conditional probability of $y_{i,\text{mis}}$ as

$$p\left(y_{i,\text{mis}}^{*(j)} \mid y_{i,\text{obs}}, \hat{\theta}\right) = \frac{f(y_{ij}^*; \hat{\theta})}{\sum_{k=1}^{M_i} f(y_{ik}^*; \hat{\theta})},$$

where $f\left(y_i; \theta\right)$ is the joint density of y_i evaluated at θ and M_i is the number of possible values of $y_{i,\text{mis}}$. The choice of $w_{ij}^* = p(y_{i,\text{mis}}^{*(j)} \mid y_{i,\text{obs}}, \hat{\theta})$ satisfies (4) and (6). Fractional imputation for categorical data using $w_{ij}^* = p(y_{i,\text{mis}}^{*(j)} \mid y_{i,\text{obs}}, \hat{\theta})$, which is close in spirit to the expectation-maximisation by weighting method of Ibrahim (1990), is discussed in Kim & Rao (2009).

For a continuous random variable y_i, condition (4) can be approximately satisfied using importance sampling, where $y_{i,\text{mis}}^{*(1)}, \ldots, y_{i,\text{mis}}^{*(M)}$ are independently generated from a distribution with density $h\left(y_{i,\text{mis}}\right)$ which has the same support as $f\left(y_{i,\text{mis}} \mid y_{i,\text{obs}}, \theta\right)$ for all $\theta \in \Omega$. The corresponding fractional weights are

$$w_{ij0}^* = w_{ij0}^*(\hat{\theta}) = C_i \frac{f(y_{i,\text{mis}}^{*(j)} \mid y_{i,\text{obs}}; \hat{\theta})}{h(y_{i,\text{mis}}^{*(j)})}, \tag{8}$$

where C_i is chosen to satisfy (6). If $h\left(y_{i,\text{mis}}\right) = f(y_{i,\text{mis}} \mid y_{i,\text{obs}}, \hat{\theta})$ is used, $w_{ij0}^* = M^{-1}$.

REMARK 1. *Under mild conditions, $\bar{g}_i^* = \sum_{j=1}^{M} w_{ij0}^* g(y_{ij}^*)$ with w_{ij0}^* in (8) converges to $\bar{g}_i(\hat{\theta}) \equiv E\{g\left(y_i\right) \mid y_{i,\text{obs}}, \hat{\theta}\}$ with probability 1, as $M \to \infty$. The approximate variance is σ_i^2/M, where*

$$\sigma_i^2 = E\left[\left\{g\left(y_i\right) - \bar{g}_i(\hat{\theta})\right\}^2 \frac{f(y_{i,\text{mis}} \mid y_{i,\text{obs}}, \hat{\theta})}{h\left(y_{i,\text{mis}}\right)} \mid y_{i,\text{obs}}, \hat{\theta}\right].$$

The $h\left(y_{i,mis}\right)$ that minimizes σ_i^2 is

$$h^*\left(y_{i,\text{mis}}\right) = f\left(y_{i,\text{mis}} \mid y_{i,\text{obs}}, \hat{\theta}\right) \times \frac{\left|g\left(y_i\right) - \bar{g}_i(\hat{\theta})\right|}{E\left\{\left|g\left(y_i\right) - \bar{g}_i(\hat{\theta})\right| \mid y_{i,\text{obs}}, \hat{\theta}\right\}}.$$

When the g-function is unknown, $h\left(y_{i,\text{mis}}\right) = f(y_{i,mis} \mid y_{i,\text{obs}}, \hat{\theta})$ is a reasonable choice in terms of statistical efficiency. Other choices of $h\left(y_{i,\text{mis}}\right)$ can have better computational efficiency in some situations.

For public access data, a large number of imputed values is not desirable. We propose an approximation with a small imputation size, say $M = 10$. To describe the procedure, let $y_{i,\text{mis}}^{*(1)}, \ldots, y_{i,\text{mis}}^{*(M)}$ be independently generated from a distribution with density $h\left(y_{i,\text{mis}}\right)$. Given the imputed values, it remains to compute the fractional weights that satisfy (4) and (6) as closely as possible. The proposed fractional weights are computed in two steps. In the first step, the initial fractional weights are computed by (8). In the second step, the initial fractional weights are adjusted to satisfy (6) and

$$\sum_{i=1}^{n}\sum_{j=1}^{M} w_i w_{ij}^* s(\hat{\theta}; y_{ij}^*) = 0, \tag{9}$$

where $s\left(\theta; y\right) = \partial \log f\left(y; \theta\right)/\partial\theta$ is the score function of θ. Adjusting the initial weights to satisfy a constraint is often called calibration. As can be seen in § 3, constraint (9) makes the resulting imputed estimator $\hat{\eta}_{FI,g}$ in (7) fully efficient for a linear function of θ.

To construct the fractional weights satisfying (6) and (9), regression weighting or empirical likelihood weighting can be used. For example, in the regression weighting, the fractional weights are

$$w_{ij}^* = w_{ij0}^* - \left(\sum_{i=1}^{n} w_i \bar{s}_i^*\right)^{\mathrm{T}} \left\{\sum_{i=1}^{n}\sum_{j=1}^{M} w_i w_{ij0}^* \left(s_{ij}^* - \bar{s}_i^*\right)^{\otimes 2}\right\}^{-1} w_{ij0}^* \left(s_{ij}^* - \bar{s}_i^*\right), \tag{10}$$

where w_{ij0}^* is the initial fractional weight (8) using importance sampling, $\bar{s}_i^* = \sum_{j=1}^{M} w_{ij0}^* s_{ij}^*$, $B^{\otimes 2} = BB^{\mathrm{T}}$, and $s_{ij}^* = s(\hat{\theta}; y_{ij}^*)$. Here, M need not be large.

If the distribution belongs to exponential family of the form

$$f\left(y;\theta\right) = \exp\left\{t\left(y\right)^{\mathrm{T}}\theta + \phi\left(\theta\right) + A\left(y\right)\right\},$$

then (9) can be obtained from $\sum_{i=1}^{n}\sum_{j=1}^{M} w_i w_{ij}^* \{t(y_{ij}^*) + \dot{\phi}(\hat{\theta})\} = 0$, where $\dot{\phi}\left(\theta\right) = \partial\phi\left(\theta\right)/\partial\theta$. In this case, calibration can be used only for complete sufficient statistics.

3. Asymptotic Results

In this section, we discuss some asymptotic properties of the fractionally imputed estimator (7). We consider two types of fractionally imputed estimators. One is obtained by using the initial fractional weights in (8) and the other is obtained by using the calibrated fractional weights of (10). The imputed estimator $\hat{\eta}_{FI,g}$ in (7) is a function of n and M, where n is the sample size and M is the number of imputed values for each missing value. Thus, we use $\hat{\eta}_{g0,n,M}$ and $\hat{\eta}_{g1,n,M}$ to denote the imputed estimator (7) using the initial fractional weights in (8) and the imputed estimator using the calibration fractional weights in (10), respectively. The following theorem presents some asymptotic properties of the fractionally imputed estimators. The proof is presented in Appendix A.

THEOREM 1. *Under some regularity conditions stated in Appendix A,*

$$\left(\hat{\eta}_{g0,n,M} - \eta_g\right)/\sigma_{g0,n,M} \to N\left(0,1\right) \tag{11}$$

and

$$\left(\hat{\eta}_{g1,n,M} - \eta_g\right)/\sigma_{g1,n,M} \to N\left(0,1\right) \tag{12}$$

in distribution, as $n \to \infty$*, for each* $M > 1$*, where*

$$\sigma^2_{g0,n,M} = \text{var}\left[\sum_{i=1}^{n} w_i \left\{\bar{g}^*_i\left(\theta_0\right) + K_1^{\text{T}}\bar{s}_i\left(\theta_0\right)\right\}\right],$$

$$\sigma^2_{g1,n,M} = \text{var}\left[\sum_{i=1}^{n} w_i \left\{\bar{g}^*_i\left(\theta_0\right) + K_1^{\text{T}}\bar{s}_i\left(\theta_0\right) + B^{\text{T}}\left(\bar{s}_i\left(\theta_0\right) - \bar{s}^*_i\left(\theta_0\right)\right)\right\}\right],$$

$\bar{g}^*_i\left(\theta\right) = \sum_{j=1}^{M} w^*_{ij0}\left(\theta\right) g(y^*_{ij})$, $\bar{s}_i\left(\theta\right) = E_\theta\left\{s\left(\theta; y_i\right) \mid y_{i,\text{obs}}\right\}$, $\bar{s}^*_i(\theta) = \sum_{j=1}^{M} w^*_{ij0}\left(\theta\right) s(\theta; y^*_{ij})$, $B = \left\{I_{\text{mis}}\left(\theta_0\right)\right\}^{-1} I_{g,\text{mis}}\left(\theta_0\right)$, *and* $K_1 = \left\{I_{\text{obs}}\left(\theta_0\right)\right\}^{-1} I_{g,\text{mis}}\left(\theta_0\right)$. *Here,* $I_{\text{obs}}\left(\theta\right) = E\left\{-\sum_{i=1}^{n} w_i \partial \bar{s}_i(\theta)/\partial\theta\right\}$, $I_{g,\text{mis}}\left(\theta\right) = E\left[\sum_{i=1}^{n} w_i \left\{s\left(\theta; y_i\right) - \bar{s}_i\left(\theta\right)\right\} g\left(y_i\right)\right]$, *and* $I_{\text{mis}}\left(\theta\right) = E\left[\sum_{i=1}^{n} w_i \left\{s\left(\theta; y_i\right) - \bar{s}_i\left(\theta\right)\right\}^{\otimes 2}\right]$.

In Theorem 1,

$$\sigma^2_{g0,n,M} = \sigma^2_{g1,n,M} + B^{\text{T}}\text{var}\left\{\sum_{i=1}^{n} w_i\left(\bar{s}^*_i - \bar{s}_i\right)\right\} B$$

and the last term represents the reduction in the variance of the fractionally imputed estimator of η_g due to the calibration in (9). Thus, $\sigma^2_{g0,n,M} \geq \sigma^2_{g1,n,M}$ with equality for $M = \infty$. Clayton et al. (1998) and Robins & Wang (2000) proved results similar to (11) for the special case of $M = \infty$.

To consider variance estimation, let

$$\hat{V}\left(\hat{\eta}_g\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} \Omega_{ij} g\left(y_i\right) g\left(y_j\right)$$

be a consistent estimator for the variance of $\hat{\eta}_g = \sum_{i=1}^{n} w_i g\left(y_i\right)$ under complete response, where Ω_{ij} are coefficients. Under simple random sampling, $\Omega_{ij} = -1/\left\{n^2\left(n-1\right)\right\}$ for $i \neq j$ and $\Omega_{ii} = 1/n^2$.

For large M, using the results in Theorem 1, a consistent estimator for the variance of $\hat{\eta}_{FI,g}$ in (7) is

$$\hat{V}\left(\hat{\eta}_{FI,g}\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} \Omega_{ij} \bar{e}^*_i \bar{e}^*_j, \tag{13}$$

where $\bar{e}^*_i = \bar{g}^*_i(\hat{\theta}) + \hat{K}_1^{\text{T}}\bar{s}^*_i(\hat{\theta}) = \sum_{j=1}^{M} w^*_{ij0}\hat{e}^*_{ij}$, $\hat{e}^*_{ij} = g(y^*_{ij}) + \hat{K}_1^{\text{T}} s(\hat{\theta}; y^*_{ij})$ and

$$\hat{K}_1 = \left\{\sum_{i=1}^{n} w_i \bar{s}^*_i(\hat{\theta})\bar{s}^*_i(\hat{\theta})^{\text{T}}\right\}^{-1} \sum_{i=1}^{n}\sum_{j=1}^{M} w_i w^*_{ij}\left\{s(\hat{\theta}; y^*_{ij}) - \bar{s}^*_i\right\} g\left(y^*_{ij}\right).$$

For moderate size M, the expected value of variance estimator (13) can be written

$$E\left\{\hat{V}\left(\hat{\eta}_{FI,g}\right)\right\} = E\left\{\sum_{i=1}^{n}\sum_{j=1}^{n} \Omega_{ij}\bar{e}_i\bar{e}_j\right\} + E\left\{\sum_{i=1}^{n}\sum_{j=1}^{n} \Omega_{ij}\text{cov}_I\left(\bar{e}^*_i, \bar{e}^*_j\right)\right\},$$

where $\bar{e}_i = E_I\left(\bar{e}_i^*\right)$ and the subscript I is used to denote the expectation with respect to the imputation mechanism generating $y_{i,mis}^{*(j)}$ from $h\left(y_{i,mis}\right)$. If the imputed values are generated independently, $\text{cov}_I(\bar{e}_i^*, \bar{e}_j^*) = 0$ for $i \neq j$ and, using the argument in Remark 1, $\text{var}_I\left(\bar{e}_i^*\right)$ can be estimated by $\hat{V}_{Ii,e} \equiv \sum_{j=1}^{M} (w_{ij0}^*)^2 (\hat{e}_{ij}^* - \bar{e}_i^*)^2$. Thus, an unbiased estimator for $\sigma_{g0,n,M}^2$ is

$$\hat{\sigma}_{g0,n,M}^2 = \sum_{i=1}^{n}\sum_{j=1}^{n} \Omega_{ij}\bar{e}_i^*\bar{e}_j^* - \sum_{i=1}^{n} \Omega_{ii}\hat{V}_{Ii,e} + \sum_{i=1}^{n} w_i^2 \hat{V}_{Ii,g},$$

where $\hat{V}_{Ii,g} = \sum_{j=1}^{M} (w_{ij0}^*)^2 (g_{ij}^* - \bar{g}_i^*)^2$. The estimator of $\sigma_{g1,n,M}^2$ in (12) can be derived in a similar manner.

Variance estimation with fractionally imputed data can be also performed using the replication method described in Appendix 2.

4. MAXIMUM LIKELIHOOD ESTIMATION

In this section, we propose a computational method for obtaining the pseudo maximum likelihood estimator in (5). The pseudo maximum likelihood estimator reduces to the usual maximum likelihood estimator if the sampling design is simple random sampling with $w_i = 1/n$. With missing data, the pseudo maximum likelihood estimator of θ_0 can be obtained by

$$\hat{\theta} = \arg\max_{\theta \in \Omega} \sum_{i=1}^{n} w_i E\left\{\log f\left(y_i; \theta\right) \mid y_{i,\text{obs}}\right\}. \quad (14)$$

For $w_i = 1/n$, Dempster et al. (1977) proved that the maximum likelihood estimator in (14) is equal to (5). They proposed using the EM algorithm, computing the solution iteratively by defining $\hat{\theta}_{(t+1)}$ to be the solution to

$$\hat{\theta}_{(t+1)} = \arg\max_{\theta \in \Omega} \sum_{i=1}^{n} w_i E\left\{\log f\left(y_i; \theta\right) \mid y_{i,\text{obs}}, \hat{\theta}_{(t)}\right\}, \quad (15)$$

where $\hat{\theta}_{(t)}$ is the estimate of θ obtained at the t-th iteration. To compute the conditional expectation in (15), Monte Carlo implementation of the EM algorithm of Wei & Tanner (1990) can be used.

In the Monte Carlo EM method, independent draws of $y_{i,\text{mis}}$ are generated from the conditional distribution $f(y_{i,\text{mis}} \mid y_{i,\text{obs}}, \hat{\theta}_{(t)})$ for each t to approximate the conditional expectation in (15). The Monte Carlo EM method requires heavy computation because the imputed values are re-generated for each iteration t. Also, generating imputed values from $f(y_{i,\text{mis}} \mid y_{i,\text{obs}}, \hat{\theta}_{(t)})$ can be computationally challenging since it often requires an iterative algorithm such as the Metropolis-Hastings algorithm for each EM iteration. To avoid re-generating values from the conditional distribution at each step, we propose the following algorithm for parametric fractional imputation:

[Step 0] Obtain an initial estimator $\hat{\theta}_{(0)}$ of θ and set $h(y_{i,\text{mis}}) = f(y_{i,\text{mis}} \mid y_{i,\text{obs}}, \hat{\theta}_{(0)})$.

[Step 1] Generate M imputed values, $y_{i,\text{mis}}^{*(1)}, \ldots, y_{i,\text{mis}}^{*(M)}$, from $h\left(y_{i,\text{mis}}\right)$.

[Step 2] With the current estimate of θ, denoted by $\hat{\theta}_{(t)}$, compute the fractional weights by $w_{ij(t)}^* = w_{ij0}(\hat{\theta}_{(t)})$, where $w_{ij0}(\hat{\theta})$ is defined in (8).

[Step 3] (Optional) If $w^*_{ij(t)} > C/M$ for some $i = 1, \ldots, n$ and $j = 1, \ldots, M$, then set $h\left(y_{i,\text{mis}}\right) = f(y_{i,\text{mis}} \mid y_{i,\text{obs}}, \hat{\theta}_{(t)})$ and go to Step 1. Increase M if necessary.

[Step 4] Find $\hat{\theta}_{(t+1)}$ that maximises over $\theta \in \Omega$ the quantity

$$Q^*\left(\theta \mid \hat{\theta}_{(t)}\right) = \sum_{i=1}^{n}\sum_{j=1}^{M} w_i w^*_{ij(t)} \log f\left(y^*_{ij};\theta\right) \tag{16}$$

over .

[Step 5] Set $t = t + 1$ and go to Step 2. Stop if $\hat{\theta}_{(t)}$ meets the convergence criterion.

In Step 0, the initial estimator $\hat{\theta}_{(0)}$ can be the maximum likelihood estimator obtained by using only the respondents. Step 1 and Step 2 correspond to the E-step of the EM algorithm. Step 3 can be used to control the variation of the fractional weights and to avoid extremely large fractional weights. The threshold C/M in Step 3 guarantees that no individual fractional weight exceeds C times of the average of the fractional weights. In Step 4, the value of θ that maximizes $Q^*(\theta \mid \hat{\theta}_{(t)})$ in (16) can be obtained by solving

$$\sum_{i=1}^{n}\sum_{j=1}^{M} w_i w^*_{ij(t)} s\left(\theta; y^*_{ij}\right) = 0, \tag{17}$$

where $s\left(\theta; y\right)$ is the score function of θ. Thus, the solution can be obtained by applying the complete sample score equation to the fractionally imputed data. Equation (17) can be called the imputed score equation using fractional imputation. Unlike the Monte Carlo EM method, the imputed values are not changed for each iteration, only the fractional weights are changed.

REMARK 2. *In Step 2, fractional weights can be computed by using the joint density with the current parameter estimate $\hat{\theta}_{(t)}$. Note that $w^*_{ij(0)}(\theta)$ in (8) can be written*

$$w^*_{ij(0)}(\theta) = \frac{f(y^{*(j)}_{i,\text{mis}} \mid y_{i,\text{obs}},\theta)/h(y^{*(j)}_{i,\text{mis}})}{\sum_{k=1}^{M} f(y^{*(k)}_{i,\text{mis}} \mid y_{i,\text{obs}},\theta)/h(y^{*(k)}_{i,\text{mis}})} = \frac{f(y^*_{ij};\theta)/h(y^{*(j)}_{i,\text{mis}})}{\sum_{k=1}^{M} f(y^*_{ik};\theta)/h(y^{*(k)}_{i,\text{mis}})}, \tag{18}$$

which does not require the marginal density in computing the conditional distribution. Only the joint density is needed.

Given the M imputed values, $y^{*(1)}_{i,\text{mis}}, \ldots, y^{*(M)}_{i,\text{mis}}$, generated from $h\left(y_{i,\text{mis}}\right)$, the sequence of estimators $\{\hat{\theta}_{(0)}, \hat{\theta}_{(1)}, \ldots\}$ can be constructed using importance sampling. The following theorem presents some convergence properties of the sequence of the estimators.

THEOREM 2. *Let $Q^*(\theta \mid \hat{\theta}_{(t)})$ be the weighted log likelihood function (16) based on fractional imputation. If*

$$Q^*(\hat{\theta}_{(t+1)} \mid \hat{\theta}_{(t)}) \geq Q^*(\hat{\theta}_{(t)} \mid \hat{\theta}_{(t)}) \tag{19}$$

then

$$l^*_{\text{obs}}(\hat{\theta}_{(t+1)}) \geq l^*_{\text{obs}}(\hat{\theta}_{(t)}), \tag{20}$$

where $l^*_{\text{obs}}(\theta) = \sum_{i=1}^{n} w_i \log\{f^*_{\text{obs}(i)}(y_{i,\text{obs}};\theta)\}$ and

$$f^*_{\text{obs}(i)}(y_{i,\text{obs}};\theta) = \frac{\sum_{j=1}^{M} f(y^*_{ij};\theta)/h(y^{*(j)}_{i,\text{mis}})}{\sum_{j=1}^{M} 1/h(y^{*(j)}_{i,\text{mis}})}.$$

Proof. By (18) and using Jensen's inequality,

$$\begin{aligned} l^*_{\text{obs}}(\hat\theta_{(t+1)}) - l^*_{\text{obs}}(\hat\theta_{(t)}) &= \sum_{i=1}^n w_i \log \sum_{j=1}^M w^*_{ij(t)} \frac{f(y^*_{ij};\hat\theta_{(t+1)})}{f(y^*_{ij};\hat\theta_{(t)})} \\ &\geq \sum_{i=1}^n \sum_{j=1}^M w_i w^*_{ij(t)} \log \frac{f(y^*_{ij};\hat\theta_{(t+1)})}{f(y^*_{ij};\hat\theta_{(t)})} \\ &= Q^*(\hat\theta_{(t+1)} \mid \hat\theta_{(t)}) - Q^*(\hat\theta_{(t)} \mid \hat\theta_{(t)}). \end{aligned}$$ □

Therefore, (19) implies (20).

Note that $l^*_{\text{obs}}(\theta)$ is an imputed version of the observed pseudo log-likelihood based on the the M imputed values, $y^{*(1)}_{i,\text{mis}}, \ldots, y^{*(M)}_{i,\text{mis}}$. Thus, by Theorem 2, the sequence $l^*_{\text{obs}}(\hat\theta_{(t)})$ is monotonically increasing and, under the conditions stated in Wu (1983), the convergence of $\hat\theta_{(t)}$ to a stationary point follows for fixed M. Theorem 2 does not hold for the sequence obtained from the Monte Carlo EM method for fixed M, because the imputed values are re-generated for each E-step of the Monte Carlo EM method, and convergence is very hard to check for the Monte Carlo EM (Booth & Hobert, 1999).

REMARK 3. *Sung & Geyer (2007) considered a Monte Carlo maximum likelihood method that directly maximizes* $l^*_{\text{obs}}(\theta)$*. Computing the value of* θ *that maximizes* $Q^*(\theta \mid \hat\theta_{(t)})$ *is easier than computing the value of* θ *that maximizes* $l^*_{\text{obs}}(\theta)$*.*

5. SIMULATION STUDY

In a simulation study, $B = 2,000$ Monte Carlo samples of size $n = 200$ were independently generated from an infinite population with $x_i \sim N(2,1)$, $y_{1i} \mid x_i \sim N(\beta_0 + \beta_1 x_i, \sigma_{ee})$, where $(\beta_0, \beta_1, \sigma_{ee}) = (1, 0{\cdot}7, 1)$, $y_{2i} \mid (x_i, y_{1i}) \sim \text{Ber}(p_i)$, $\log\{p_i/(1-p_i)\} = \phi_0 + \phi_1 x_i + \phi_2 y_{1i}$, $(\phi_0, \phi_1, \phi_2) = (-3, 0{\cdot}5, 0{\cdot}7)$, $\delta_{i1} \mid (x_i, y_i, z_i) \sim \text{Ber}(\pi_i)$, $\log\{\pi_i/(1-\pi_i)\} = 0{\cdot}5 x_i$, and $\delta_{i2} \mid (x_i, y_i, z_i, \delta_{i1}) \sim \text{Ber}(0{\cdot}7)$. The variables x_i, δ_{i1}, and δ_{i2} are always observed. Variable y_{1i} is observed if $\delta_{i1} = 1$ and is not observed if $\delta_{i1} = 0$. Variable y_{2i} is observed if $\delta_{i2} = 1$ and is not observed if $\delta_{i2} = 0$. The overall response rate for y_1 is about 72%.

We are interested in estimating four parameters: the marginal mean of y, $\eta_1 = E(y_1)$; the marginal mean of y_2, $\eta_2 = E(y_2)$; the slope for the regression of y_1 on x, $\eta_3 = \beta_1$; and the proportion of y_1 less than 3, $\eta_4 = \text{pr}(y_1 < 3)$. Under complete response, η_1, η_2, and η_3 are computed by the maximum likelihood method and the proportion η_4 is estimated by

$$\hat\eta_{4,n} = \frac{1}{n} \sum_{i=1}^n I(y_{1i} < 3). \tag{21}$$

Under nonresponse, four imputed estimators were computed: the parametric fractional imputation estimator using w^*_{ij0} in (8) with $M = 100$; the calibration fractional imputation estimator using the regression weighting method in (10) with $M = 10$; and two multiple imputation estimators with $M = 100$ and $M = 10$, respectively. In fractional imputation, M imputed values of y_{1i} were independently generated by $y^*_{1ij} \sim N(\hat\beta_{0(0)} + \hat\beta_{1(0)} x_i, \hat\sigma_{ee(0)})$, where $(\hat\beta_{0(0)}, \hat\beta_{1(0)}, \hat\sigma_{ee(0)})$ is the initial regression parameter estimator computed from the respondents of y_1. Also, M imputed values of y_{2i} were independently generated by $y^*_{2ij} \mid (x_i, y^*_{1ij}) \sim \text{Ber}(\hat p_{ij(0)})$, where

Table 1. *Monte Carlo standardised variances of the imputed estimators*

Imputation method	η_1	η_2	η_3	η_4
FI ($M = 100$)	129	137	150	110
MI ($M = 100$)	129	136	150	110
CFI ($M = 10$)	129	137	150	110
MI ($M = 10$)	132	138	156	111

FI, fractional imputation; CFI, calibration fractional imputation; MI, multiple imputation.

$\log\{\hat{p}_{ij(0)}/(1-\hat{p}_{ij(0)})\} = \hat{\phi}_{0(0)} + \hat{\phi}_{1(0)}x_i + \hat{\phi}_{2(0)}y^*_{1ij}$ and $(\hat{\phi}_{0(0)}, \hat{\phi}_{1(0)}, \hat{\phi}_{2(0)})$ is the initial coefficient for the logistic regression of y_{2i} on $(1, x_i, y^*_{1ij})$ obtained by solving the imputed score equation for (ϕ_0, ϕ_1, ϕ_2) using the respondents for y_2 only. For each imputed value, we assign the fractional weight

$$w^*_{ij(t)} \propto \frac{f_1\left(y_{1i}^{*(j)} \mid x_i, \hat{\theta}_{1(t)}\right) f_2\left(y_{2i}^{*(j)} \mid x_i, y_{1i}^{*(j)}, \hat{\theta}_{2(t)}\right)}{f_1\left(y_{1i}^{*(j)} \mid x_i, \hat{\theta}_{1(0)}\right) f_2\left(y_{2i}^{*(j)} \mid x_i, y_{1i}^{*(j)}, \hat{\theta}_{2(0)}\right)}, \tag{22}$$

where $f_1(y_1 \mid x, \theta_1)$ denotes the conditional distribution of y_1 given x evaluated at $\theta_1 = (\beta_0, \beta_1, \sigma_{ee})$ and

$$f_2(y_2 \mid x, y_1, \theta_2) = \begin{cases} \text{pr}\,(y_2 = 1 \mid x, y_1, \theta_2) & \text{if } y_2 = 1 \\ \text{pr}\,(y_2 = 0 \mid x, y_1, \theta_2) & \text{if } y_2 = 0, \end{cases}$$

with $\theta_2 = (\phi_0, \phi_1, \phi_2)$. In (22), the parameter estimates $\hat{\theta}_{1(t)}$ and $\hat{\theta}_{2(t)}$ were obtained by the maximum likelihood method using the fractionally imputed data with fractional weight $w^*_{ij(t-1)}$. In Step 3 of the fractional imputation for maximum likelihood in § 4, $C = 5$ was used. In the calibration fractional imputation method, $M = 10$ values were randomly selected from $M_1 = 100$ initial fractionally imputed values by systematic sampling with selection probability proportional to w^*_{ij0} in (8). The regression fractional weights were then computed by (10). In Step 5, the convergence criterion was $\|\hat{\theta}_{(t+1)} - \hat{\theta}_{(t)}\| < 10^{-9}$. In multiple imputation, the imputed values are generated from the posterior predictive distribution iteratively using Gibbs sampling with 100 iterations.

All the point estimators are nearly unbiased and are not listed here. The standardised variances of the four imputed estimators are presented in Table 1. The standardised variance in Table 1 was computed by dividing the variance of each estimator by that of the complete sample estimator. The simulation results in Table 1 show that the fractional imputed estimator and the multiple imputation estimator have similar properties for $M = 100$. The calibration fractional imputation estimator is more efficient than the multiple imputation estimator for $M = 10$ because it uses extra information in the imputed score functions.

In addition to point estimators, variance estimators were also computed for each Monte Carlo sample. We used the linearised variance estimator (13) for fractional imputation. For multiple imputation, we used the variance formula of Rubin (1987). Table 2 presents the Monte Carlo relative biases for the variance estimators. The simulation error for the relative bias of the variance estimators reported in Table 2 is less than 1%. Table 2 shows that the proposed linearisation method provides good estimates for the variance of the fractional imputation estimators. The multiple imputation variance estimators are essentially unbiased for η_1, η_2, and η_3 which ap-

Table 2. *Relative biases of the variance estimators (%)*

Imputation method	var($\hat{\eta}_1$)	var($\hat{\eta}_2$)	var($\hat{\eta}_3$)	var($\hat{\eta}_4$)
FI ($M = 100$)	1·0	−1·1	−2·6	−3·2
MI ($M = 100$)	1·3	−0·6	−1·4	12·2
CFI ($M = 10$)	0·9	−2·1	−2·9	−1·6
MI ($M = 10$)	0·4	0·1	−2·3	12·7

FI, fractional imputation; CFI, calibration fractional imputation; MI, multiple imputation.

pear in the imputation model. For variance estimation of the proportion, the multiple imputation variance estimator shows significant bias (12·7% for $M = 10$ and 12·2% for $M = 100$). The multiple imputation method in this simulation is congenial for the estimators of η_1, η_2 and η_3, but it is not congenial for the estimator (21) of η_4. See Meng (1994) and Appendix 3.

6. Concluding Remarks

Parametric fractional imputation is proposed as a method of creating a complete data set with fractionally imputed data. Parameter estimation with fractionally imputed data can be implemented using existing software treating the imputed values as observed. The data provider, who has good information for model development, can use an imputation model to construct the fractionally imputed data with replicated fractional weights for variance estimation. No information beyond the data set is required for analysis.

If parametric fractional imputation is used to construct the score function, the solution to the imputed score equation is very close to the maximum likelihood estimator for the parameters in the model. Parametric fractional imputation yields consistent estimates for parameters that are not part of the imputation model. For example, in the simulation study, parametric fractional imputation computed from a normal model provides direct estimates for the cumulative distribution function. Thus, the proposed imputation method is useful when the parameters of interest are unknown at the time of imputation. Variance estimation can be performed using a linearisation method or a replication method. Variance estimation for parametric fractional imputation, unlike multiple imputation, does not require the congeniality condition of Meng (1994).

The proposed fractional imputation is applicable when the response mechanism is nonignorable and the response mechanism is specified. Also, parametric fractional imputation can be used with data from a large scale survey sample obtained by a complex sampling design. These topics are beyond the scope of this paper and will be presented elsewhere. Some computational issues such as the convergence criteria for the EM algorithm using fractional imputation are also topics for future research.

Acknowledgement

The research was partially supported by a Cooperative Agreement between the US Department of Agriculture Natural Resources Conservation Service and Iowa State University. The author wishes to thank professor Wayne Fuller, three anonymous referees, and the associate editor for their very helpful comments.

Supplementary Material

More details of the simulation setup, including the program codes, are available at http://jkim.public.iastate.edu/fi.html.

Appendix 1

Assumptions and proof for Theorem 1

We consider a regular parametric family $\{f(y;\theta);\theta \in \Omega\}$, where Ω is in a finite dimensional Euclidean space. Assume that the true parameter θ_0 lies in the interior of Ω. Define $\bar{S}(\theta) = \sum_{i=1}^{n} w_i E\left\{s(\theta;y_i) \mid y_{i,\text{obs}},\theta\right\}$ and $\bar{\eta}_g(\theta) = \sum_{i=1}^{n} w_i E\left\{g(y_i) \mid y_{i,\text{obs}},\theta\right\}$. We assume the following conditions:

(C1) The solution $\hat{\theta}$ in (5) is unique and satisfies $n^{1/2}(\hat{\theta}-\theta_0) = O_p(1)$.
(C2) The partial derivatives of $\bar{S}(\theta)$ and $\bar{\eta}_g(\theta)$ exist and are continuous around θ_0 almost everywhere.
(C3) The partial derivative of $\bar{S}(\theta)$ satisfies

$$\|\partial\bar{S}(\theta)/\partial\theta - E\left\{\partial\bar{S}(\theta)/\partial\theta\right\}\| \to 0$$

in probability, uniformly in θ and $E\left\{\partial\bar{S}(\theta)/\partial\theta\right\}$ is continuous and nonsingular at θ_0. Also, the partial derivative of $\bar{\eta}_g(\theta)$ satisfies

$$\|\partial\bar{\eta}_g(\theta)/\partial\theta - E\left\{\partial\bar{\eta}_g(\theta)/\partial\theta\right\}\| \to 0$$

in probability, uniformly in θ and $E\left\{\partial\bar{\eta}_g(\theta)/\partial\theta\right\}$ is continuous at θ_0.
(C4) There exists a positive d such that $E\left\{g(Y)^{2+d}\right\} < \infty$ and $E\left\{S_j(\theta_0)^{2+d}\right\} < \infty$ where $S_j(\theta) = \partial\log f(y;\theta)/\partial\theta_j$ for $j = 1,\ldots,p$ and θ_j is the j-th element of θ.

Condition (C1) is a standard condition and will be satisfied in most cases. Conditions (C2) and (C3) provide some conditions about the partial derivatives of the estimator computed from the conditional expectation. Note that $E\left\{\partial\bar{S}(\theta)/\partial\theta\right\} = -I_{\text{obs}}(\theta)$ and $E\left\{\partial\bar{\eta}_g(\theta)/\partial\theta\right\} = I_{g,\text{mis}}(\theta)$, which are defined in Theorem 1. Condition (C4) is the moment conditions for the central limit theorem.

Proof of Theorem 1. Define a class of estimators

$$\tilde{\eta}_{g0,n,K}(\theta) = \sum_{i=1}^{n}\sum_{j=1}^{M} w_i w_{ij0}^{*}(\theta)\, g\left(y_{ij}^{*}\right) + K^{\mathrm{T}} \sum_{i=1}^{n} w_i E\left\{s(\theta;y_i) \mid y_{i,\text{obs}},\theta\right\}$$

indexed by K. Note that, by (5), we have $\sum_{i=1}^{n} w_i E\{s(\hat{\theta};y_i) \mid y_{i,\text{obs}},\hat{\theta}\} = 0$ and $\tilde{\eta}_{g0,n,K}(\hat{\theta}) = \hat{\eta}_{g0,n,M}$ for any K. According to Theorem 2.13 of Randles (1982), we have

$$\tilde{\eta}_{g0,n,K}(\hat{\theta}) - \tilde{\eta}_{g0,n,K}(\theta_0) = o_p\left(n^{-1/2}\right),$$

if

$$E\left\{\frac{\partial}{\partial\theta}\tilde{\eta}_{g0,n,K}(\theta_0)\right\} = 0 \qquad \text{(A.1)}$$

is satisfied. Using

$$\sum_{i=1}^{n}\sum_{j=1}^{M} w_i \left\{\frac{\partial}{\partial\theta} w_{ij0}^{*}(\theta)\right\} g\left(y_{ij}^{*}\right) = \sum_{i=1}^{n}\sum_{j=1}^{M} w_i w_{ij0}^{*}(\theta)\left\{s\left(\theta;y_{ij}^{*}\right) - \bar{s}_i^{*}(\theta)\right\} g\left(y_{ij}^{*}\right),$$

the choice of $K = \{I_{\text{obs}}(\theta_0)\}^{-1} I_{g,\text{mis}}(\theta_0) = K_1$ in Theorem 1 satisfies (A.1) and thus (11) follows.

To show (12), consider

$$\tilde{\eta}_{g1,n,K}(\theta) = \sum_{i=1}^{n}\sum_{j=1}^{M} w_i w_{ij}^*(\theta)\, g\left(y_{ij}^*\right) + K^{\mathrm{T}} \sum_{i=1}^{n} w_i E\left\{s(\theta; y_i) \mid y_{i,\text{obs}}, \theta\right\}.$$

Using (10),

$$\sum_{i=1}^{n}\sum_{j=1}^{M} w_i w_{ij}^*(\theta)\, g\left(y_{ij}^*\right) = \sum_{i=1}^{n}\sum_{j=1}^{M} w_i w_{ij0}^*(\theta)\, g\left(y_{ij}^*\right) + \left[\sum_{i=1}^{n} w_i E\left\{s(\theta; y_i) \mid y_{i,\text{obs}}, \theta\right\} - \sum_{i=1}^{n}\sum_{j=1}^{M} w_i w_{ij0}^*(\theta)\, s\left(\theta; y_{ij}^*\right)\right]^{\mathrm{T}} \hat{B}_g(\theta),$$

where

$$\hat{B}_g(\theta) = \left[\sum_{i=1}^{n}\sum_{j=1}^{M} w_i w_{ij0}^*(\theta)\left\{s_{ij}^*(\theta) - \bar{s}_i^*(\theta)\right\}^{\otimes 2}\right]^{-1} \sum_{i=1}^{n}\sum_{j=1}^{M} w_i w_{ij0}^*(\theta)\left\{s_{ij}^*(\theta) - \bar{s}_i^*(\theta)\right\} g\left(y_{ij}^*\right).$$

After some algebra, it can be shown that the choice of $K = K_1$ in Theorem 1 also satisfies $E\{\partial \tilde{\eta}_{g1,n,K}(\theta_0)/\partial\theta\} = 0$ and, by Randles (1982) again,

$$\tilde{\eta}_{g1,n,K}(\hat{\theta}) - \tilde{\eta}_{g1,n,K}(\theta_0) = o_p\left(n^{-1/2}\right).$$

APPENDIX 2

Replication variance estimation

Under complete response, let $w_i^{[k]}$ be the k-th replication weight for unit i. Assume that the replication variance estimator

$$\hat{V}_n = \sum_{k=1}^{L} c_k \left(\hat{\eta}_g^{[k]} - \hat{\eta}_g\right)^2,$$

where c_k is the factor associate with replication k, L is the number of replication, $\hat{\eta}_g = \sum_{i=1}^{n} w_i g(y_i)$ and $\hat{\eta}_g^{[k]} = \sum_{i=1}^{n} w_i^{[k]} g(y_i)$, is consistent for the variance of $\hat{\eta}_g$. For replication with the calibration method of (9), we consider the following steps for creating replicated fractional weights.

[Step 1] Compute $\hat{\theta}^{[k]}$, the k-th replicate of $\hat{\theta}$, using fractional weights.

[Step 2] Using the $\hat{\theta}^{[k]}$ computed from Step 1, compute the replicated fractional weights by

$$\sum_{i=1}^{n}\sum_{j=1}^{M} w_i^{[k]} w_{ij}^{*[k]} s\left(\hat{\theta}^{[k]}; y_{ij}^*\right) = 0, \tag{A.2}$$

using the regression weighting technique.

Equation (A.2) is the calibration equation for the replicated fractional weights. For any estimator of the form (7), the replication variance estimator is constructed as

$$\hat{V}\left(\hat{\eta}_{FI,g}\right) = \sum_{k=1}^{L} c_k \left(\hat{\eta}_{FI,g}^{[k]} - \hat{\eta}_{FI,g}\right)^2$$

where $\hat{\eta}_{FI,g}^{[k]} = \sum_{i=1}^{n}\sum_{j=1}^{M} w_i^{[k]} w_{ij}^{*[k]} g(y_{ij}^*)$ and $w_{ij}^{*[k]}$ is computed from (A.2).

In general, Step 1 can be computationally problematic since $\hat{\theta}^{[k]}$ is often computed from the iterative algorithm (16) for each replicate. Thus, we consider an approximation for $\hat{\theta}^{[k]}$ using Taylor linearisation of $0 = \bar{S}^{[k]}(\hat{\theta}^{[k]})$ around $\hat{\theta}$. The one-step approximation is

$$\hat{\theta}^{[k]} \cong \hat{\theta} + \left[\hat{I}^{[k]}_{\text{obs}}\left(\hat{\theta}\right)\right]^{-1} \bar{S}^{[k]}\left(\hat{\theta}\right),$$

where

$$\hat{I}^{[k]}_{\text{obs}}(\theta) = \sum_{i=1}^{n}\sum_{j=1}^{M} w_i^{[k]} w_{ij}^{*}\left\{-\frac{\partial}{\partial\theta} s\left(\theta; y_{ij}^{*}\right)\right\} - \sum_{i=1}^{n}\sum_{j=1}^{M} w_i^{[k]} w_{ij}^{*}\left\{s\left(\theta; y_{ij}^{*}\right) - \bar{s}_i^{*}(\theta)\right\}^{\otimes 2}$$

and $\bar{S}^{[k]}(\theta) = \sum_{i=1}^{n} w_i^{[k]} \bar{s}_i^{*}(\theta)$. The $\hat{I}^{[k]}_{\text{obs}}(\theta)$ is a replicated version of the observed information matrix proposed by Louis (1982).

APPENDIX 3

A note on multiple imputation for a proportion

Assume that we have a random sample of size n with observations (x_i, y_i) obtained from a bivariate normal distribution. The parameter of interest is a proportion, for example, $\eta = \text{pr}(y \le 3)$. An unbiased estimator of η is

$$\hat{\eta}_n = \frac{1}{n}\sum_{i=1}^{n} I(y_i \le 3). \tag{A.3}$$

Note that $\hat{\eta}_n$ is unbiased but has larger variance than the maximum likelihood estimator

$$\int_{-\infty}^{3} \phi\left(\frac{y - \hat{\mu}_y}{\hat{\sigma}_{yy}}\right) dy, \tag{A.4}$$

where $\phi(y)$ is the density of the standard normal distribution and $(\hat{\mu}_y, \hat{\sigma}_{yy})$ is the maximum likelihood estimator of (μ_y, σ_{yy}).

For simplicity, assume that the first $r(< n)$ elements have both x_i and y_i responding, but the last $n - r$ elements have x_i observed and y_i missing. In this situation, an efficient imputation method such as

$$y_i^{*} \sim N\left(\hat{\beta}_0 + x_i\hat{\beta}_1, \hat{\sigma}_e^2\right) \tag{A.5}$$

can be used, where $\hat{\beta}_0$, $\hat{\beta}_1$ and $\hat{\sigma}_e^2$ can be computed from the respondents. In multiple imputation, the parameter estimates are generated from a posterior distribution given the observations. Under the imputation mechanism (A.5), the imputed estimator of μ_2 of the form $\hat{\mu}_{2,I} = n^{-1}\left(\sum_{i=1}^{r} y_i + \sum_{i=r+1}^{n} y_i^{*}\right)$ satisfies

$$\text{var}(\hat{\mu}_{2,FE}) = \text{var}(\bar{y}_n) + \text{var}(\hat{\mu}_{2,FE} - \bar{y}_n), \tag{A.6}$$

where $\hat{\mu}_{2,FE} = E_I(\hat{\mu}_{2,I})$. Condition (A.6) is the congeniality condition of Meng (1994).

Now, for $\eta = \text{pr}(y \le 3)$, the imputed estimator of η based on $\hat{\eta}_n$ in (A.3) is

$$\hat{\eta}_I = \frac{1}{n}\left\{\sum_{i=1}^{r} I(y_i \le 3) + \sum_{i=r+1}^{n} I(y_i^{*} \le 3)\right\}. \tag{A.7}$$

The expected value of $\hat{\eta}_I$ over the imputation mechanism is

$$\begin{aligned} E_I(\hat{\eta}_I) &= \frac{1}{n}\left\{\sum_{i=1}^{r} I(y_i \le 3) + \sum_{i=r+1}^{n} \text{pr}(y_i \le 3 \mid x_i, \hat{\theta})\right\} \\ &= \hat{\eta}_{FE} + \frac{1}{n}\sum_{i=1}^{r}\{I(y_i \le 3) - \text{pr}(y_i \le 3 \mid x_i, \hat{\theta})\}, \end{aligned}$$

where $\hat{\eta}_{FE} = n^{-1}\sum_{i=1}^{n} \text{pr}(y_i \leq 3 \mid x_i, \hat{\theta})$. For the proportion, $\hat{\eta}_{FE} \neq E_I(\hat{\eta}_I)$ and so the congeniality condition does not hold. In fact,

$$\text{var}\{E_I(\hat{\eta}_I)\} < \text{var}(\hat{\eta}_n) + \text{var}\{E_I(\hat{\eta}_I) - \hat{\eta}_n\}$$

and the multiple imputation variance estimator overestimates the variance of $\hat{\eta}_I$ in (A.7). If the maximum likelihood estimator (A.4) is used, then

$$\text{var}(\hat{\eta}_{FE}) = \text{var}(\hat{\eta}_n) + \text{var}(\hat{\eta}_{FE} - \hat{\eta}_n)$$

and the multiple imputation variance estimator will be approximately unbiased.

REFERENCES

BOOTH, J. G. & HOBERT, J. P. (1999). Maximizing generalized linear models with an automated Monte Carlo EM algorithm. *J. R. Statist. Soc. B* **61**, 625–685.

CLAYTON, D., SPIEGELHALTER, D., DUNN, G. & PICKLES, A. (1998). Analysis of longitudinal binary data from multiphase sampling. *J. R. Statist. Soc. B* **60**, 71–87.

DEMPSTER, A. P., LAIRD, N. M. & RUBIN, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *J. R. Statist. Soc. B* **39**, 1–37.

FAY, R. E. (1992). When are inferences from multiple imputation valid? In *Proc. Survey Res. Meth. Sect.* Alexandria, VA: American Statistical Association.

FAY, R. E. (1996). Alternative paradigms for the analysis of imputed survey data. *J. Am. Statist. Assoc.* **91**, 490–498.

GELMAN, A., MENG, X. L. & STERN, H. (1996). Posterior predictice assessmemt of model fitness via realized discrepancies (with discussion). *Statist. Sinica* **6**, 733–807.

IBRAHIM, J. G. (1990). Incomplete data in generalized linear models. *J. Am. Statist. Assoc.* **85**, 765–769.

KIM, J. K., BRICK, M. J., FULLER, W. A. & KALTON, G. (2006). On the bias of the multiple imputation variance estimator in survey sampling. *J. R. Statist. Soc. B* **68**, 509–521.

KIM, J. K. & FULLER, W. (2004). Fractional hot deck imputation. *Biometrika* **91**, 559–578.

KIM, J. K. & RAO, J. N. K. (2009). Unified approach to linearization variance estimation from survey data after imputation for item nonresponse. *Biometrika* **96**, 917–932.

LOUIS, T. A. (1982). Finding the observed information matrix when using the EM algorithm. *J. R. Statist. Soc. B* **44**, 226–233.

MENG, X. L. (1994). Multiple-imputation inferences with uncongenial sources of input (with discussion). *Statist. Sci.* **9**, 538–573.

RANDLES, R. H. (1982). On the asymptotic normality of statistics with estimated parameters. *Ann. Statist.* **10**, 462–474.

ROBINS, J. M. & WANG, N. (2000). Inference for imputation estimators. *Biometrika* **87**, 113–124.

RUBIN, D. B. (1976). Inference and missing data. *Biometrika* **63**, 581–590.

RUBIN, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys*. Wiley.

SUNG, Y. J. & GEYER, C. J. (2007). Monte Carlo likelihood inference for missing data models. *Ann. Statist.* **35**, 990–1011.

WANG, N. & ROBINS, J. M. (1998). Large-sample theory for parametric multiple imputation procedures. *Biometrika* **85**, 935–948.

WEI, G. C. & TANNER, M. A. (1990). A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithms. *J. Am. Statist. Assoc.* **85**, 699–704.

WU, C. F. J. (1983). On the convergence properties of the EM algorithm. *Ann. Statist.* **11**, 95–103.

[*Received January* ????. *Revised July* ????]