Proceedings of the Autumn Conference in 2001
and the 30th Anniversary of the Korean Statistical Society

Approximate moments of a variance estimate with imputed conditional means

Woo Ram Kang, Min Woong Shin[1] Sang Eum Lee[2]

Abstract

Schafer and Shenker(2000) mentioned the one of analytic imputation technique involving conditional means. We derive an approximate moments of a variance estimate with imputed conditional means.

Keywords : *Missing data, Imputation, Conditional means,*

1. Introduction

The method to produce variance estimates with just a single, nonrandom imputation of predictive means for the missing values was developed by Schafer and Schenker(2000). And the method was based on the asymptotic expansions of point estimators and their associated variance estimators and produces a first-order approximation to Rubin's repeated-imputation inference with a infinite number of imputations and mentioned the conditional mean imputation.

2. Setup and Assumptions

2.1 Pattern of missing data

Suppose a simple random sample of size n observational units from a population size N. Sometimes a single variable Y is missing y denote the $n\times 1$ vector of Y values. Then y can be consisted of observed and missing components, y_{obs} and y_{mis} , with size of n_1 and $n_0 = n - n_1$The rate of observed data is $r_1 = n_1/n$ and the rate of missing data is $r_0 = 1 - r_1$

2.2 Estimating with complete data

1) Department of Statistics, Hankuk University of Foreign Studies, Kyounggi-do, Korea, 449-791.
2) Department of Applied Information Statistics, Kyonggi University, Kyounggi-do, Korea, 442-760

Let Q denote a scalar quantity to be interested If the data were complete, a point estimate be noted as $\hat{Q}=\hat{Q}(y)$, and estimate of variance for $\hat{Q}$, note as $U=U(y)$ The point estimator $\hat{Q}$ that we consider are smooth functions of linear statistics. Let

$$\hat{Q}=g(T_y) \tag{2.1}$$

where $T_y=n^{-1}\sum_{i=1}^{n} y_i$.

y_i denotes the values of Y(observed or missing) for unit i and g is smooth and well-behaved. , the estimated Q will be the expectations of the linear statistics,

$$Q=g(ET_y) \tag{2.2}$$

where $\hat{Q}$ can be thought of as a method-of-moments estimate of Q.

2.3 Modeling missing data

Assuming that nonrespnse is ignorable and that a probability model for y_{mis} given y_{obs} has been correctly specified. A typical specification for this model will include unknown estimable parameters, θ.

Let $\hat{\theta}$ denote an estimate of θ based on y_{obs} under the assumed model for missing data. Also, let Γ denote an estimate of $V(\theta-\hat{\theta})$based on y_{obs}.

Therefore, if $\hat{\theta}$ may be a maximum likelihood(ML) estimate, and Γ may be the inverse of the observed information matrix evaluated at $\hat{\theta}$. We assume that $\Gamma=O(n^{-1})$and that

$$\Gamma^{-1/2}(\theta-\hat{\theta}) \rightarrow N(0,I)$$

where I denotes the identity matrix.

Let mis denote the indices of i such that $y_{mis} \in mis$. Then for all $i, i' \in mis$, we assume that

$$E(y_i \mid y_{obs}, \theta)=\mu_i(\theta),$$

$$V(y_i \mid y_{obs}, \theta)=\sigma_i^2(\theta),$$

$$Cov(y_i, y_{i'} \mid y_{obs}, \theta)=0$$

where μ_i and σ_i^2 are function of θ.

3. Conditional Mean Imputation

3.1 Conditional mean imputation

By Little and Rubin(1986), let $\mu(\theta)$ denote the vector with elements $\mu_i(\theta)$, $i \in mis$; that is

$$\mu(\theta)=E(y_{mis} \mid Y_{obs}, \theta) \tag{3.1}$$

Conditional mean imputation can be efficient for point estimation of Q ; in fact, $\hat{Q}(Y_{obs}, \mu(\hat{\theta}))$ is a first order approximation to the "best" estimate of Q.

3.2 Corrected analysis methods for conditional mean imputation

This is the method of drawing inference for Q from a data set which the missing values of Y have been replaced by conditional means. This method(Schafer & Schenker 2000) can be considered a linear approximation to a full analysis using multiple-imputed data.

3.2.1 Bayes interpretation

A Bayesian interpretation regards $\hat{Q}$ and U as fixed(given complete data) and Q as random.

$$\hat{Q} = E(Q \mid y_{obs},\ \theta),$$

$$U = V(Q \mid y_{obs},\ \theta).$$

Bayesian interpretation of $\hat{\theta}$ and Γ is posterior moments of θ given the observed data,

$$\hat{\theta} = E(\theta \mid y_{obs}),$$

$$\Gamma = V(\theta \mid y_{obs}).$$

Note that by Rubin(1987, p43)

$$\begin{aligned} E(Q \mid y_{obs}) &= E[E(Q \mid y_{obs}, y_{mis}) \mid y_{obs})] \\ &= E[\hat{Q}(y_{obs}, y_{mis}) \mid y_{obs}] \\ &= E(\hat{Q} \mid y_{obs}) \end{aligned}$$

and

$$V(Q \mid y_{obs}) = V(\hat{Q} \mid y_{obs}) + E(U \mid y_{obs})$$

where

$$\begin{aligned} \hat{Q} &= \hat{Q}(y_{obs}, y_{mis}) = E(Q \mid y_{obs}, y_{mis}) \\ U &= V(Q \mid y_{obs}, y_{mis}) \end{aligned}$$

To obtain approximate posterior moments of Q, then, we need only to approximate the mean and variance of $\hat{Q}$ and the mean of U over the predictive distribution of y_{mis}.

Here are the approximate moments of $\hat{Q}$ and U.

$$E(\hat{Q} \mid y_{obs}) = \hat{Q}(y_{obs}, \mu(\hat{\theta})) + O_p(n^{-1}) \tag{3.2}$$

$$V(\hat{Q} \mid y_{obs}) = \left(\frac{\partial g(\hat{T})}{\partial T_y}\right)^2 n^{-2} \sum_{i \in mis} \sigma_i^2(\hat{\theta}) + \left(\frac{\partial g(\hat{T})}{\partial T_y}\right)^2 D_\mu(\hat{\theta})^T \Gamma D_\mu(\hat{\theta}) + O_p(n^{-3/2}) \tag{3.3}$$

where $\hat{T}$ is for the complete-data statistics T with $\mu(\hat{\theta})$ substituted for y_{mis},

and where $D_\mu(\theta) = n^{-1} \sum_{i \in mis} \left(\frac{\partial \mu_i(\theta)}{\partial \theta}\right)$ and finally,

$$E(U \mid y_{obs}) = U(y_{obs}, \mu(\hat{\theta})) + \left(\frac{\partial g(\hat{T})}{\partial T_y}\right)^2 n^{-2} \sum_{i \in mis} \sigma_i^2(\hat{\theta}) + O_p(n^{-3/2})$$

Also the point estimation for Q is following;

$$E(Q \mid y_{obs}) \approx \hat{Q}(y_{obs}, \mu(\hat{\theta})) \tag{3.4}$$

and for variance estimation is as following;

$$V(Q|\,y_{obs}) \approx U(y_{obs},\,\mu(\hat{\theta})) + C_1 + C_2 \tag{3.5}$$

where

$$C_1 = 2\left(\frac{\partial g(\hat{T})}{\partial T_y}\right)^2 n^{-1} \sum_{i \in mis} \sigma_i^2(\hat{\theta}) \tag{3.6}$$

and

$$C_2 = \left(\frac{\partial g(\hat{T})}{\partial T_y}\right)^2 D_\mu(\hat{\theta})^T \Gamma D_\mu(\hat{\theta}) \tag{3.7}$$

4. Approximate moments of $\hat{Q}$ and U

Under the assumptions outlined in section 2

$$E[\,\hat{Q}(y_{obs},\,y_{mis})|\,y_{obs}] \;=\; \hat{Q}(y_{obs},\,\mu(\hat{\theta})) + O_p(n^{-1}) \tag{4.1}$$

We derive the approximate moments of $\hat{Q}$ and U.

Functions of y_{obs} $(e.g.,\ \hat{\theta})$are considered to be fixed, whereas functions of y_{mis} or θ are considered to be random.

Note first that the only random argument of $\hat{Q} = g(T)$is T_y.

We can write

$$\begin{aligned} T_y - \hat{T}_y \;&=\; n^{-1} \sum_{i \in mis} (y_i - \mu_i(\hat{\theta})) \\ &=\; n^{-1} \sum_{i \in mis} \varepsilon_i + n^{-1} \sum_{i \in mis} (\mu_i(\theta) - \mu_i(\hat{\theta})) \end{aligned} \tag{4.2}$$

where the $\varepsilon_i \;=\; y_i - \mu_i(\theta)$are independent random variables with mean 0 and variance $\sigma_i^2(\theta)$.

Thus the first term in (4.2) in $O_p(n^{-\frac{1}{2}})$ For the second term, note that

$$\begin{aligned} \mu_i(\theta) - \mu_i(\hat{\theta}) \;&=\; \left(\frac{\partial \mu_i(\hat{\theta})}{\partial \theta}\right)^T (\theta - \hat{\theta}) + O_p(n^{-1}) \\ &= O_p(n^{-\frac{1}{2}}) \end{aligned} \tag{4.3}$$

because the number of elements in mis in O(n), the second term in (4.2) is also $O_p(n^{-\frac{1}{2}})$ and thus

$$T_y \;=\; \hat{T}_y + O_p(n^{-\frac{1}{2}}) \tag{4.4}$$

To establish (4.1) , espand $\hat{Q} \;=\; g(T)$in a Taylor series about $T_y = \hat{T}_y$

$$\begin{aligned} \hat{Q}(y_{obs},\,y_{mis}) - \hat{Q}(y_{obs},\,\mu(\hat{\theta})) \;&=\; g(T) - g(\hat{T}) \\ &= \left(\frac{\partial g(\hat{T}(\theta))}{\partial T_y}\right)(T_y - \hat{T}_y(\theta)) + O_p(n^{-1}) \end{aligned}$$

$\hat{Q}(y_{obs},\,\mu(\hat{\theta}))$ is fixed and $E(T_y) \;=\; \hat{T}_y(\theta) \;+\; O_p(n^{-1})$

Therefore

$$E(\hat{Q}|y_{obs}) - E[\hat{Q}(y_{obs}, \mu(\hat{\theta}))|y_{obs}] = E(\hat{Q}|y_{obs}) - \hat{Q}(y_{obs}, \mu(\hat{\theta})) = O_p(n^{-1})$$

Note that $\hat{\theta}$ are considered to be fixed.
That is,

$$E(\hat{Q}|y_{obs}) = \hat{Q}(y_{obs}, \mu(\hat{\theta})) + O_p(n^{-1})$$

$$V(\hat{Q}|y_{obs}) = \left(\frac{\partial g(\hat{T}(\theta))}{\partial T_y}\right)^2 n^{-2} \sum_{i \in mis} \sigma_i^2(\theta) + \left(\frac{\partial g(\hat{T}(\theta))}{\partial T_y}\right)^2 D_\mu(\hat{\theta})^T \Gamma D_\mu(\hat{\theta}) + O_p(n^{-\frac{1}{2}}) \quad (4.5)$$

To establish (4.5), write

$$V(\hat{\theta}) = EV(\hat{\theta}|\theta) + VE(\hat{\theta}|\theta)$$

Let

$$\hat{T}_y(\theta) = n^{-1}[\sum_{i \in obs} y_i + \sum_{i \in mis} \mu_i(\theta)]$$

so that $\hat{T}_y(\hat{\theta}) = \hat{T}_y$ and let $\hat{T}(\theta) = (\hat{T}_y(\theta)).^T$

Note that $\hat{T}$ is shorthand for the complete-data statistic T calculated with $\mu(\hat{\theta})$ substituted for y_{mis} .

For any fixed θ, $T_y - \hat{T}_y(\theta)$ has mean zero and variance $n^{-2} \sum_{i \in mis} \sigma_i^2(\theta)$

Note that

$$T_y - \hat{T}_y(\theta) = n^{-1}[\sum_{i \in mis} y_i - \sum_{i \in mis} \mu_i(\theta)]$$

has variance $n^{-2} \sum_{i \in mis} \sigma_i^2(\theta)$

Thus the expansion

$$\hat{Q}(y_{obs}, y_{mis}) - \hat{Q}(y_{obs}, \mu(\hat{\theta})) = \left(\frac{\partial g(\hat{T}(\theta))}{\partial T_y}\right)(T_y - \hat{T}_y(\theta)) + O_p(n^{-1})$$

imlpies that

$$V[(\hat{Q}(y_{obs}, y_{mis}) - \hat{Q}(y_{obs}, \mu(\theta)))|\theta] = V[\hat{Q}(y_{obs}, y_{mis})] = \left(\frac{\partial g(\hat{T}(\theta))}{\partial T_y}\right)^2 n^{-2} \sum_{i \in mis} \sigma_i^2(\theta) + O_p(n^{-\frac{3}{2}}) \quad (4.6)$$

The second them, $\hat{Q}(y_{obs}, \mu(\theta))$ is fixed for given θ.

Notice that the leading term in (4.6) is of order n^{-1}. Expanding $nV(\hat{Q}|\theta)$ about $\theta = \hat{\theta}$ leads to

$$EV(\hat{Q}|\theta) = \left(\frac{\partial g(\hat{T}(\theta))}{\partial T_y}\right)^2 n^{-2} \sum_{i \in mis} \sigma_i^2(\theta) + O_p(n^{-\frac{3}{2}}) \quad (4.7)$$

where $\hat{Q}(y_{obs}, y_{mis})$.

Also, for any fixed θ

$$E(\hat{Q}|\theta)-E(\hat{Q}(y_{obs},\mu(\theta))) = E(\hat{Q}|\theta)-g(\hat{T}(\theta)) = O_p(n^{-1})$$
$$E(\hat{Q}|\theta) = g(\hat{T}(\theta))+O_p(n^{-1})$$

where $\hat{Q}(y_{obs}, y_{mis})$ and $\hat{T}(\theta)=\hat{Q}(y_{obs}, \mu(\theta))$

Note that $\hat{Q}(y_{obs}, \mu(\theta))$ is fixed for any fixed θ.

Expanding this expression for $E(\hat{Q}|\theta)$ about $\theta=\hat{\theta}$ gives

$$VE(\hat{Q}|\theta) = \left(\frac{\partial g(T(\hat{\theta}))}{\partial\theta}\right)^T \Gamma\left(\frac{\partial g(\hat{T}(\hat{\theta}))}{\partial\theta}\right) + O_p(n^{-\frac{3}{2}})$$

Note that

$$\begin{aligned} E(\hat{Q}|\theta)-E(\hat{Q}|\hat{\theta}) &= g(\hat{T}(\theta))-g(\hat{T}(\hat{\theta})) \\ &= \frac{\partial g(\hat{T}(\hat{\theta}))}{\partial\theta}(\theta-\hat{\theta})+O_p(n^{-1}) \end{aligned} \tag{4.8}$$

Since $E(\hat{Q}|\hat{\theta})$ is fixed, (4.8) is derived.

$$\begin{aligned} V[E(\hat{Q}|\theta)-E(\hat{Q}|\hat{\theta})] &= V[E(\hat{Q}|\theta)] \\ &= \left(\frac{\partial g(\hat{T}(\hat{\theta}))}{\partial\theta}\right)^T \Gamma\left(\frac{\partial g(\hat{T}(\hat{\theta}))}{\partial\theta}\right)+O_p(n^{-\frac{3}{2}}) \end{aligned}$$

But by the chain rule,

$$\begin{aligned} \frac{\partial g(\hat{T}(\theta))}{\partial\theta} &= \frac{\partial g(\hat{T}(\theta))}{\partial T_y}\frac{\partial(\hat{T}_y(\theta))}{\partial\theta} \\ &= \frac{\partial g(\hat{T}(\theta))}{\partial T_y} n^{-1}\sum_{i\in mis}\left(\frac{\partial\mu_i(\theta))}{\partial\theta}\right) \end{aligned}$$

so

$$VE(\hat{Q}|\theta) = \left(\frac{\partial g(\hat{T})}{\partial T_y}\right)^2 D_\mu(\hat{\theta})^T\Gamma D_\mu(\hat{\theta}) + O_p(n^{-\frac{3}{2}}) \tag{4.9}$$

We may use Bayes estimate instead of MLE for θ. If $\theta\propto Be(\alpha, \beta)$ then we may calculate α and β with empirical Bayes estimate.

5. Discussion

Schafer and Schenker (2000) presented the method estimating a binomial proportion and the missing data y_{mis} are modeled as a vector of i.i.d. *Bernoulli*(θ) random variables. We may consider hierarchical Bayes estimate for θ. And we may use Markov Chain Monte Carlo algorithm to obtain the posterior distribution of θ.

Reference

[1] Donald B. Rubin(1987). *Multiple Imputation for Nonresponse in Survey.* Wiley & Sons

[2] Joseph L. Schafer & Nathaniel Schenker(2000) " Inference With Imputed Conditioned Means" *JASA Vol. 95 No. 449*

[3] Lee, P. M(1988) *Bayesian Statistics: An Introduction.* University of York. English

[4] Press, S. J.(1982) *Applied Multivariate Analysis: Using Bayesian and Frequentist Methods of Inference.* R.E. Krieger Publishing Co.