The Korean Communications
in Statistics Vol. 9, No. 1, 2002
pp. 63-73

Estimating a Binomial Proportion with Bayes Estimated Imputed Conditional Means

Min Woong Shin[1] and Sang Eun Lee[2]

Abstract

The one of analytic imputation technique involving conditional means was mentioned by Schafer and Schenker(2000). And their derivations are based on asymptotic expansions of point estimator and their associated variance estimator, and the result of imputation can be thought of as first-order approximations to the estimators. Specially in this paper, we are presenting the method of estimating a Binomial proportion with Bayesian approach of imputed conditional means. That is, instead of using maximum likelihood(ML) estimator to estimate a Binomial proportion, in general, we use the Bayesian estimators and will show the result of estimated imputed conditional means.

Keywords : Missing data, Imputation, Bayesian estimator, Conditional means

1. Introduction

Techniques of nonresponse imputation are more paid attention to statisticians recently. Because the imputation gives the standard complete-data methods of analysis. In 1987, Little and Rubin stated that one drawback of imputation followed by the use of complete-data methods of analysis is that the resulting inference may be seriously misleading, because uncertainty due to missing data has not been addressed. Following that statement, general multiple imputation method was introduced by Rubin and the idea leads to several completed data. And these methods were pretty much based on the "repeated-imputation" methods. However Schafer and Schenker(2000) developed the analytic method to produce appropriate variance estimates with just a single, nonrandom imputation of predictive means for the missing values. 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

1) Professor Department of Statistics, Hankook University of Foreign Studies,
E-mail : mwshin@stat,hufs.ac.kr
2) Assistant Professor, Department of Applied and Information Statistics, Kyonggi University
E-mail : sanglee@stat.kyonggi.ac.kr

conditional mean imputation. In this paper we use the example of estimating a Binomial proportion with conditional mean imputation method in 2000 paper and in stead of using ML estimator , we applied the Bayesian estimator. In section 2, it will show the data pattern for imputation and section 3 gives the general conditional mean imputation. Finally in section 4 we have the theorem of the results for using Bayesian estimator of estimating Binomial proportion.

2. Description of data

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 and other variables $X_1, X_2, \ldots, X_p$ are completely observed. And let X denote the $n \times p$ matrix of observed data for $X_1, X_2, \ldots, X_p$ and 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 $r_0 = 1 - r_1$. We assume that r_0 is bounded away from 1 as $n \rightarrow \infty$.

2.2 Estimating with complete data

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

$$\hat{Q} = g(T_{X_1}, \ldots, T_{X_p}, T_y) \quad \text{---(1)}$$

where $T_{X_j} = n^{-1} \sum_{i=1}^{n} X_{ij}$, $j = 1, \ldots, p$, $T_y = n^{-1} \sum_{i=1}^{n} y_i$.

X_{ij} denotes the values of X_j for unit 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,

$$Q = g(ET_{X_1}, \ldots, ET_{X_p}, ET_y),$$

where $\hat{Q}$ can be thought of as a method - of -moments estimate of Q.
We assume that complete-data variance estimator U has the form

$$U = n^{-1}\left(\frac{\partial g(T)}{\partial T}\right)^T S\left(\frac{\partial g(T)}{\partial T}\right)$$

where $T=(T_{X_1},\ldots,T_{X_p},T_y)$ and $S=(n-1)^{-1}(Z^T Z - nTT^T)$, with $Z=(X,y)$. That is , U is the classical variance estimator for $\hat{Q}$ based on the sample covariance matrix and the δ-method.

Therefore,

$$U^{-1/2}(Q-\hat{Q}) \rightarrow N(0,1), \quad \text{as } n\rightarrow\infty.$$

2.3 Modeling missing data

In general(Rubin 1976), most procedures for handling missing data in surveys and elsewhere are based on an assumption of ignorability. The observed data, of course, provide no information to support or contradict this assumption; such support must come from a source external to the observed data. Therefore, every missing-data procedure must be based on some assumption that cannot be verified from (X, y_{obs}) alone.

Assuming that nonrespnse is ignorable and that a probability model for y_{mis} given (X, 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 (X, y_{obs}) under the assumed model for missing data. Also, let Γ denote an estimate of $V(\theta-\hat{\theta})$ based on (X, 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

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

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

and

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

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

$$\mu(\theta) = E(y_{mis}|X, Y_{obs}, \theta).$$

Conditional mean imputation can be efficient for point estimation of Q ; in fact, $\hat{Q}(X, Y_{obs}, \mu(\hat{\theta}))$ is a first order approximation to the "best" estimate of Q. Inference can be distorted with conditional mean imputation, however. With conditional means imputed for y_{mis} does not hold in general because is usually biased downward.

$$EU(X, y_{obs}, \mu(\hat{\theta})) < V(Q - \hat{Q}(X, y_{obs}, \mu(\hat{\theta}))\ .$$

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. The results are restricted to simple random samples form a infinite population.

3.2.1 Bayes interpretation

A Bayesian interpretation regards $\hat{Q}$ and U as fixed(given complete data) and Q as random. Regarding $\hat{Q}$ and U as approximate complete-data posterior mean and variance of Q, respectively,

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

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

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

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

$$\Gamma = V(\theta|X, y_{obs}).$$

Now, with incomplete data, however, inferences should be based on the posterior moments given only the data actually observed, $E(Q|X, y_{obs})$ and $V(Q|X, y_{obs})$. Note that

$$E(Q|X, y_{obs}) = E(\hat{Q}|X, y_{obs})$$

and

$$V(Q|X, y_{obs}) = V(\hat{Q}|X, y_{obs}) + E(U|X, y_{obs})$$

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.

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

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

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

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

and finally,

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

Also the point estimation for Q is following;

$$E(Q|X, y_{obs}) \approx \hat{Q}(X, y_{obs}, \mu(\hat{\theta})) \qquad ---(2)$$

and for variance estimation is as following;

$$V(Q|X, y_{obs}) \approx U(X, y_{obs}, \mu(\hat{\theta})) + C_1 + C_2 \qquad ---(3)$$

where

$$C_1 = 2(\frac{\partial g(\hat{T})}{\partial T_y})^2 n^{-1} \sum_{i \in mis} \sigma_i^2(\hat{\theta}) \qquad ---(4)$$

and

$$C_2 = (\frac{\partial g(\hat{T})}{\partial T_y})^2 D_{\mu}(\hat{\theta})^T \Gamma D_{\mu}(\hat{\theta}) \qquad ---(5)$$

4. Estimating Binomial Proportion

Suppose a simple random sample of size of n of binary(0-1) variables from infinite population, and let the estimated be $Q = p$, and inferences for p are based on the point estimate $\hat{Q} = \bar{y} = \frac{1}{n} \sum_i y_i$ and the variance of estimate is $U = \frac{1}{n(n-1)} \sum_i (y_i - \bar{y})$.

Suppose that Y is subject to missingness completely at random, so that y_{obs} is just a simple random sample from y

If the missing data, y_{mis} are modeled as a vector of i.i.d Bernoulli (θ) random variables, then $\mu_i(\theta)=\theta$ and $\sigma_i^2(\theta)=\theta(1-\theta)$.

4.1 Using ML estimation based on y_{obs}

Maximum likelihood estimators of θ and variance of θ are as following.

$$\hat{\theta}=\frac{1}{n_1}\sum_{i\in obs} y_i \quad \text{and} \quad \Gamma=\hat{\theta}(1-\hat{\theta})/n_1$$

and substitute the estimates to equation (2) - (5) then we have the below results.

$$\hat{Q}(X, y_{obs}, \mu(\hat{\theta}))=\bar{y}_1=\frac{1}{n_1}\sum_{i\in obs} y_i$$

$$U(X, y_{obs}, \mu(\hat{\theta}))\approx r_1^2\frac{1}{n_1(n_1-1)}\sum_{i\in obs}(y_i-\bar{y}_1)^2$$

$$C_1\approx 2r_1r_0\frac{1}{n_1(n_1-1)}\sum_{i\in obs}(y_i-\bar{y}_1)^2$$

and

$$C_2\approx r_0^2\frac{1}{n_1(n_1-1)}\sum_{i\in obs}(y_i-\bar{y}_1)^2$$

4.2 Using Bayes estimation based on y_{obs}

Now in this study we apply the Bayes estimates for θ and variance of θ.

Take the prior for θ as a beta distribution then

$$p(\theta)\propto\theta^{\alpha-1}(1-\theta)^{\beta-1} \quad ,0\le\theta\le 1$$

that is if $\theta \propto Be(\alpha, \beta)$, then the posterior of θ becomes a beta distribution with $\alpha + y_1,\ \beta + n_1 - x$

$$h(\theta | y_1) \propto \theta^{\alpha + y_1 - 1}(1-\theta)^{\beta + n_1 - x - 1}$$

where $y_1 = n\,\overline{y}_1$ is from observed data.

Now, Bayes estimator for θ is as following.

$$\hat{\theta}_B = E(\theta | y_1) = \frac{\alpha + y_1}{\alpha + \beta + n_1}$$

and

$$\Gamma_B = V(\theta | y_1) = \frac{(\alpha + y_1)(\beta + n_1 - y_1)}{(\alpha + \beta + n_1)^2(\alpha + \beta + n_1 + 1)}$$

Algebraic calculation yields as following.

$$\hat{Q}(X, y_{obs}, \mu(\hat{\theta})) = (\sum_{i \in obs} y_1 + n_0\,\overline{y}_B)/n$$
$$= r_1\,\overline{y}_1 + r_0\,\overline{y}_B$$

where $\overline{y}_B = \hat{\theta}_B = E(\theta | y_1) = \dfrac{\alpha + y_1}{\alpha + \beta + n_1}$

$$U(X, y_{obs}, \mu(\hat{\theta})) = \frac{\sum_{i \in obs}(y_i - \hat{Q})^2 + \sum_{i \in mis}(\hat{Q} - \hat{Q})^2}{n(n-1)}$$
$$= \frac{\sum_{i \in obs}(y_i - \hat{Q})^2}{n(n-1)}$$
$$= \frac{\sum_{i \in obs}(y_i - \hat{Q})^2}{n_1(n_1-1)} \times \frac{n_1(n_1-1)}{n(n-1)}$$
$$\approx r_1^2 \frac{1}{n_1(n_1-1)} \sum_{i \in obs}(y_i - \hat{Q})^2$$

$$C_1 = 2\left(\frac{\partial g(\hat{T})}{\partial T_y}\right)^2 n^{-2} \sum_{i \in mis} \sigma_i^2(\hat{\theta})$$
$$= 2 \cdot 1 \cdot n^{-2}\, n_0\, \hat{\theta}(1-\hat{\theta})$$
$$\approx 2 \cdot \frac{n_1}{n} \frac{n_0}{n} \frac{\sum_{i \in obs}(y_i - \hat{Q})}{n_1(n_1-1)}$$
$$\approx 2\, r_1\, r_0 \frac{1}{n_1(n_1-1)} \sum_{i \in obs}(y_i - \hat{Q})$$

$$C_2 = \left(\frac{\partial g(\hat{T})}{\partial T_y}\right)^2 D_\mu(\hat{\theta})^T \Gamma_B D_\mu(\hat{\theta})$$
$$= 1 \cdot r_0^2 \cdot \Gamma_B$$

where $D_\mu(\hat{\theta}) = \frac{n_0}{n} = r_0$

$$\Gamma_B = \frac{(\alpha + y_1)(\beta + n_1 - y_1)}{(\alpha+\beta+n_1)^2(\alpha+\beta+n_1+1)}$$

5. Summary

Here we state the summary as a theory :

If a simple random sample of size of n of binary(0-1) variables from infinite population, and that Y is subject to missingness completely at random, so that y_{obs} is just a simple random sample from y. If the missing data, y_{mis} are modeled as a vector of i.i.d Bernoulli (θ) random variables, then $\mu_i(\theta) = \theta$ and $\sigma_i^2(\theta) = \theta(1-\theta)$ and using Bayes estimator for θ then estimation of quantity $\hat{Q} = \bar{y} = \frac{1}{n}\sum_i y_i$ is as following.

$$E(\hat{Q}|X, Y_{obs}) \approx \bar{y}_1 + r_0\, \bar{y}_B$$

$$V(\hat{Q}|X, Y_{obs}) \approx r_1^2 \frac{1}{n_1(n_1-1)} \sum_{i \in obs} (y_i - \hat{Q})^2 + 2 r_1 r_0 \frac{1}{n_1(n_1-1)} \sum_{i \in obs} (y_i - \hat{Q})$$

$$+ r_0^2 \cdot \frac{(\alpha + \bar{y}_1)(\beta + n_1 - \bar{y}_1)}{(\alpha + \beta + n_1)^2 (\alpha + \beta + n_1 + 1)}$$

Finally, we have the results table using ML and Bayes Estimates.

	M.L.	***Bayesian***
$\hat{Q}(X, y_{obs}, \mu(\hat{\theta}))$	$\bar{y}_1 = \frac{1}{n_1} \sum_{i \in obs} y_i$	$r_1 \bar{y}_1 + r_0 (\frac{\alpha + \bar{y}_1}{\alpha + \beta + \bar{y}_1})$
$U(X, y_{obs}, \mu(\hat{\theta}))$	$r_1^2 \frac{1}{n_1(n_1-1)} \sum_{i \in obs} (y_i - \bar{y}_1)^2$	$r_1^2 \frac{1}{n_1(n_1-1)} \sum_{i \in obs} (y_i - \hat{Q})^2$
C_1	$2 r_1 r_0 \frac{1}{n_1(n_1-1)} \sum_{i \in obs} (y_i - \bar{y}_1)^2$	$2 r_1 r_0 \frac{1}{n_1(n_1-1)} \sum_{i \in obs} (y_i - \hat{Q})$
C_2	$r_0^2 \frac{1}{n_1(n_1-1)} \sum_{i \in obs} (y_i - \bar{y}_1)^2$	$1 \cdot r_0^2 \cdot \frac{(\alpha + \bar{y}_1)(\beta + n_1 - \bar{y}_1)}{(\alpha + \beta + n_1)^2 (\alpha + \beta + n_1 + 1)}$

where $\hat{Q} = \hat{Q}(X, y_{obs}, \mu(\hat{\theta}))$

From the above table, besides of changing the part of ML to Bayesian estimator, basically, the final results of both estimators have the same structure as what Schafer &.Schenker(2000) were insisted in their paper. Specially the variance terms are still efficient respect to the role of r_0 and r_1.

Consequently, equations of $U(X, y_{obs}, \mu(\hat{\theta}))$, C_1, C_2 show that the proportionate contributions to the correct variance estimate are approximately r_1^2, $2r_1 2r_0$, and r_0^2 respectively.

Reference

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

[2] Joseph L. Schafter & 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.

[5] R. J .K Little & Donald B. Rubin(1986). *Statistical Analysis with missing data* wiley & Sons

[2001년 7월 접수, 2001년 12월 채택]